Oracle® Retail AI Foundation Cloud Services: AI Foundation

User Guide

Release 24.1.101.0
F91880–02
February 2024

Oracle Retail AI Foundation Cloud Services: AI Foundation User Guide, Release 24.1.101.0

F91880–02

Copyright © 2024, Oracle and/or its affiliates.

Primary Author: Judith Meskill

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software, software documentation, data (as defined in the Federal Acquisition Regulation), or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs (including any operating system, integrated software, any programs embedded, installed, or activated on delivered hardware, and modifications of such programs) and Oracle computer documentation or other Oracle data delivered to or accessed by U.S. Government end users are "commercial computer software," "commercial computer software documentation," or "limited rights data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, reproduction, duplication, release, display, disclosure, modification, preparation of derivative works, and/or adaptation of i) Oracle programs (including any operating system, integrated software, any programs embedded, installed, or activated on delivered hardware, and modifications of such programs), ii) Oracle computer documentation and/or iii) other Oracle data, is subject to the rights and limitations specified in the license contained in the applicable contract. The terms governing the U.S. Government's use of Oracle cloud services are defined by the applicable contract for such services. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle®, Java, MySQL and NetSuite are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Inside are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Epyc, and the AMD logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

Contents

Send Us Your Comments

Preface

Audience xiv
Documentation Accessibility xiv
Related Documents xiv
Customer Support xv
Improved Process for Oracle Retail AI Foundation Documentation Corrections xv
Oracle Retail AI Foundation Documentation on the Oracle Help Center xv
Conventions xvi

1 Getting Started

About Oracle Retail AI Foundation Cloud Services 1-1
Administration 1-1
Oracle Retail AI Foundation Dashboard 1-1
Process 1-3
Overview of the User Interface 1-8
- Process Train 1-8
- View Menu 1-8
- Embedded Help 1-8
- Process Indicator 1-8
- Search 1-8
- Icons 1-8
- Buttons 1-10

Browser Settings 1-10
Supported Characters for Text Entry 1-11
Histograms 1-11

2 Customer Decision Trees

Introduction 2-1

Overview of CDT Process 2-1
Overview Tab 2-2
CDT Generation Stage Status 2-2
Aggregate Statistics 2-3
Calculation Report 2-4
CDT Score Distribution Histogram 2-6
Generate CDT Tab 2-8
Data Setup 2-8
Category Selections 2-9
Time Interval Setup 2-11
Data Filtering 2-11
Filter Setup 2-12
Data Filtering Summary 2-13
Data Filtering Histograms 2-14
Calculation 2-15
Version Setup 2-16
Category Attributes Setup 2-17
Calculation Report 2-19
CDT Score Distribution Histogram 2-20
Evaluation 2-21
Pruning Setup 2-21
Pruning Results 2-22
Evaluation Report 2-23
Evaluation Histograms 2-24
Escalation 2-25
Setup Escalation 2-26
Escalation Report 2-27
Completion of Process 2-28
Manage CDTs Tab 2-28
Browse by Category or Version 2-30
Search 2-30
Create CDT 2-30
Deleting a CDT 2-31
Using the CDT Editor 2-31
Add a Child 2-31
Editing Nodes 2-32
Deleting Nodes 2-32
Copying/Pasting in Customer Decision Tree Editor 2-32
Viewing a Customer Decision Tree 2-32
Comparing Two CDTs 2-33

3 Demand Transference

Introduction 3-1
Overview of DT Process 3-1
Overview Tab 3-1
Generation Stage Status 3-2
Aggregate Statistics 3-3
Calculation Report 3-4
Generate Models Tab 3-5
Data Setup 3-5
Category Selections 3-5
Data Filtering 3-7
Filter Setup 3-8
Data Filter Summary 3-8
Data Filtering Histograms 3-9
Similarity Calculation 3-11
Version Setup 3-11
Category Attribute Setup 3-12
Transaction-Based Similarity Availability Per Category 3-14
Similarity Display 3-14
Elasticity Calculation 3-15
Calculation Report 3-16
Pruning Report 3-17
Assortment Elasticity Histogram 3-17
Escalation 3-18
Setup Escalation 3-19
Escalation Report 3-20
Completion of Process 3-21
Manage Models Tab 3-21
Time Interval Setup 3-22
Calculate Substitutable Demand Percentages 3-23
Browsing and Searching 3-23
Escalation Report 3-24
Histogram 3-25

4 Advanced Clustering

Introduction 4-1
Features 4-1
Overview of Advanced Clustering Process 4-2
Cluster Criteria Overview Tab 4-2
Clustering Criteria 4-3

Generate Store Clusters Tab 4-4
Cluster Criteria Stage 4-4
All Cluster Criteria 4-5
All Cluster Criteria Filter 4-6
All Cluster Criteria Summary 4-6
All Cluster Criteria Scenario List 4-7
Cluster Criteria 4-7
Effective Period 4-9
Summarization 4-9
Source Time Period 4-10
Contextual Area 4-10
Explore Data 4-12
Summary 4-12
View Stores 4-13
Product Performance 4-13
Contextual Area 4-14
Cluster Setup Stage 4-17
Summary 4-17
Scenario Definition Section 4-18
Cluster Results Stage 4-22
Summary 4-22
Scenario Results Section 4-23
Export to Excel 4-27
Scenario List 4-28
Scenario Compare 4-28
Scenario System Recommendations 4-29
Scenario Optimality 4-30
Scenario Rank 4-30
Outlier Indicator 4-30
Cluster Scores 4-30
New Stores and Stores with a Poor History 4-31
Insights Stage 4-31
Summary 4-33
Contextual Information 4-33
Manage Store Clusters Tab 4-33
Contextual Information 4-34
Scenario Compare 4-37
Attribute Analysis 4-37
Cluster Overrides 4-38
Rename Cluster 4-38
Create New Manual Cluster 4-39

Delete Cluster 4-40
Move Stores to Clusters 4-40
Cluster Review, Approve, and Adjust 4-40
Nested vs. Mixed Attribute 4-41
Nested 4-41
Mixed Attributes 4-42

5 Customer Segmentation

Introduction 5-1
Features 5-1
Overview of the Customer Segmentation Process 5-2
Customer Segmentation Overview Tab 5-2
Segmentation Criteria 5-3
Version List Tab 5-4
Version Actions 5-4
Version Criteria Pop-up 5-5
Source Time Period 5-6
Top Categories 5-7
Generate Customer Segments Tab 5-7
Segment Criteria Stage 5-8
All Segment Criteria 5-8
All Segment Criteria Filter 5-9
All Segment Criteria Summary 5-10
All Segment Criteria Scenario List 5-11
Segment Criteria 5-11
Effective Period 5-12
Summarization 5-13
Source Time Period 5-13
Contextual Area 5-13
Explore Data 5-14
Summary 5-15
Attribute Mining 5-15
Attribute Importance 5-15
Attribute Correlation 5-17
Filter 5-17
Correlation Matrix 5-18
Contextual Area 5-18
Segment Setup Stage 5-19
Scenario Definition Tab 5-20
Scenario List Tab 5-23

Scenario Compare Tab 5-24
Segment Results Stage 5-24
Summary 5-25
Scenario Results Section 5-25
Scenario List 5-29
Scenario Compare 5-30
Scenario System Recommendations 5-31
Scenario Optimality 5-31
Scenario Rank 5-32
Outlier Indicator 5-32
Segment Rules 5-32
Segment Scores 5-33
New Customers 5-33
Generate Store Profile 5-33
Segment Insights Stage 5-33
Customer Segment Analysis 5-34
Category Analysis 5-34
Opportunity Analysis Tab 5-35
Summary 5-37
Cross Segment Spend and Margin Analysis BI 5-37
RFM Analysis BI 5-38
Customer Segment vs. All Customer Sales Trends BI 5-39
Customer Attribute Analysis BI 5-39
Category Analysis BI 5-40
Trend Category Analysis BI 5-41

6 Attribute Extraction

Introduction 6-1
Overview Tab 6-1
Edit Labels Tab 6-2
Annotation Tab 6-3
Unigraphs 6-3
Assigning User Labels and Reviewing Machine Labels 6-3
Finding New Values 6-5
Digraphs 6-6
Assigning User Labels to Digraphs 6-6
Errors Tab 6-7
Normalization Tab 6-8
Results Tab 6-11

Status 6-11

7 Affinity Analysis

Introduction 7-1
User Interface 7-2
Assortment Recommender 7-5
Using Assortment Recommender Results 7-9
Market Basket Analysis 7-10
Using Market Basket Rules 7-15
Affinity Analysis 7-16
Affinity Analysis Overview 7-16
Compare Results 7-24
Create Custom Run 7-26
Manage Affinity Analysis 7-28
Compare Results 7-29

8 Size Profiles

Overview 8-1
Navigation 8-1
All Runs 8-2
Creating a Run 8-2
Information Tab 8-3
Size Ranges Tab 8-4
Parameters Tab 8-5
Reviewing a Run 8-9
Run Output 8-9
Size Profiles Tab 8-10
Input Tab 8-12
Submitted Profiles 8-12
Approved Profiles 8-13
Submitted Profiles Table 8-14
Assigned Tasks 8-15
Assigned Tasks Table 8-16
Approved Profiles 8-17
Useful Tables and Views 8-18

9 Innovation Workbench

Process 9-1
Workspace List 9-1

SQL Workshop 9-2
Application Builder 9-6
Manage Notebook 9-6
Features 9-6

10 Oracle Digital Assistance

11 Control and Tactical Center

Strategy & Policy Management 11-1
Manage Business Rules 11-2
Alert Management 11-3
Rules Management 11-4
Promotion Management 11-4
Manage Forecast Configurations 11-8
Overview 11-9
Setup 11-9
Create a Run Type 11-11
Data Aggregation 11-12
Add Multiple Run Types 11-13
Delete Run Type 11-14
Test 11-15
Create What-If Run 11-16
Duplicate Run 11-20
Delete Run 11-20
Approve Demand Parameters 11-21
Approve Base Demand and Forecast 11-21
Manage 11-21
Promotion Effects 11-22
Forecast Configuration — Promotion Features 11-22
Forecast Configuration — Promotion Model Settings 11-23
Map 11-25
Manage Master Data 11-26
Manage Credential Stores 11-28
Attribute Extraction and Attribute Binning 11-29

12 Forecasting Methods in AIF

Some Forecasting Concepts 12-1
Exponential Smoothing 12-1

Automatic Method Selection 12-2
Seasonal Regression Analysis 12-2
Bayesian Analysis 12-2
Forecasting Methologies 12-3
Automatic Exponential Smoothing Method 12-3
Sales & Promo Method 12-4
Life Cycle Method 12-6
Hierarchical Bayesian Method 12-7

Glossary

Send Us Your Comments

> **Note:**
>
> To view the Table of Contents Click the TOC Icon

Oracle Retail AI Foundation User Guide, Release 24.1.101.0

Oracle welcomes customers' comments and suggestions on the quality and usefulness of this document.

Your feedback is important, and helps us to best meet your needs as a user of our products. For example:

- Are the implementation steps correct and complete?
- Did you understand the context of the procedures?
- Did you find any errors in the information?
- Does the structure of the information help you with your tasks?
- Do you need different information or graphics? If so, where, and in what format?
- Are the examples correct? Do you need more examples?

If you find any errors or have any other suggestions for improvement, then please tell us your name, the name of the company who has licensed our products, the title and part number of the documentation and the chapter, section, and page number (if available).

> **Note:**
>
> Before sending us your comments, you might like to check that you have the latest version of the document and if any concerns are already addressed. To do this, access the Online Documentation available on the Oracle Help Center It contains the most current Documentation Library plus all documents revised or released recently.

Send your comments to us using the electronic mail address: retail-doc_us@oracle.com

Please give your name, address, electronic mail address, and telephone number (optional).

If you need assistance with Oracle software, then please contact your support representative or Oracle Support Services.

If you require training or instruction in using Oracle software, then please contact your Oracle local office and inquire about our Oracle University offerings. A list of Oracle offices is available on our Web site at http://www.oracle.com.

Preface

This guide describes the Oracle Retail AI Foundation Cloud Services user interface. It provides step-by-step instructions to complete most tasks that can be performed through the application.

Audience

This User Guide is intended for retailers and analysts.

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Related Documents

For more information, see the following documents in the Oracle Retail AI Foundation Cloud Services documentation set:

- *Oracle Retail AI Foundation Cloud Services Administration Guide*
- *Oracle Retail AI Foundation Cloud Services Implementation Guide*
- *Oracle Retail AI Foundation Cloud Services Security Guide*
- *Oracle Retail AI Foundation Cloud Services Release Notes*
- *Oracle Retail Analytics and Planning Cloud Services Data Interface*
- *Oracle Retail AI Foundation Cloud Services User Guide*
- *Oracle Retail AI Foundation Cloud Services Assortment and Space Optimization User Guide*
- *Oracle Retail Inventory Planning Optimization Cloud Service-Inventory Optimization User Guide*
- *Oracle Retail Cloud Service Lifecycle Pricing Optimization User Guide*

Customer Support

To contact Oracle Customer Support, access My Oracle Support at the following URL:

https://support.oracle.com

When contacting Customer Support, please provide the following:

- Product version and program/module name
- Functional and technical description of the problem (include business impact)
- Detailed step-by-step instructions to re-create
- Exact error message received
- Screen shots of each step you take

Improved Process for Oracle Retail AI Foundation Documentation Corrections

To more quickly address critical corrections to Oracle Retail AI Foundation documentation content, Oracle Retail AI Foundation documentation may be republished whenever a critical correction is needed. For critical corrections, the republication of an Oracle Retail AI Foundation document may at times not be attached to a numbered software release; instead, the Oracle Retail AI Foundation document will simply be replaced on the Oracle Help Center Web site, or, in the case of Data Models, to the applicable My Oracle Support Documentation container where they reside.

Oracle Retail AI Foundation documentation is available on the Oracle Help Center at the following URL:

https://docs.oracle.com/en/industries/retail/index.html

An updated version of the applicable Oracle AI Foundation document is indicated by Oracle part number, as well as print date (month and year). An updated version uses the same part number, with a higher-numbered suffix. For example, part number E123456-02 is an updated version of a document with part number E123456-01.

If a more recent version of a document is available, that version supersedes all previous versions.

Oracle Retail AI Foundation Documentation on the Oracle Help Center

Oracle Retail AI Foundation product documentation is available on the following web site:

https://docs.oracle.com/en/industries/retail/index.html

(Data Model documents are not available through Oracle Help Center. You can obtain these documents through My Oracle Support.)

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
italic	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
`monospace`	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.

1
Getting Started

This chapter provides an overview of Oracle Retail AI Foundation Cloud Services.

About Oracle Retail AI Foundation Cloud Services

The Oracle Retail AI Foundation Cloud Services is an analytical product that consists of the following modules: Customer Decision Tree Science Cloud Service, Demand Transference Science Cloud Service, Advanced Clustering Cloud Service, Customer Segmentation, Lifecycle Pricing Optimization Cloud Service, and Assortment and Space Optimization Cloud Service. You may have access to all of these modules or to a subset of them.

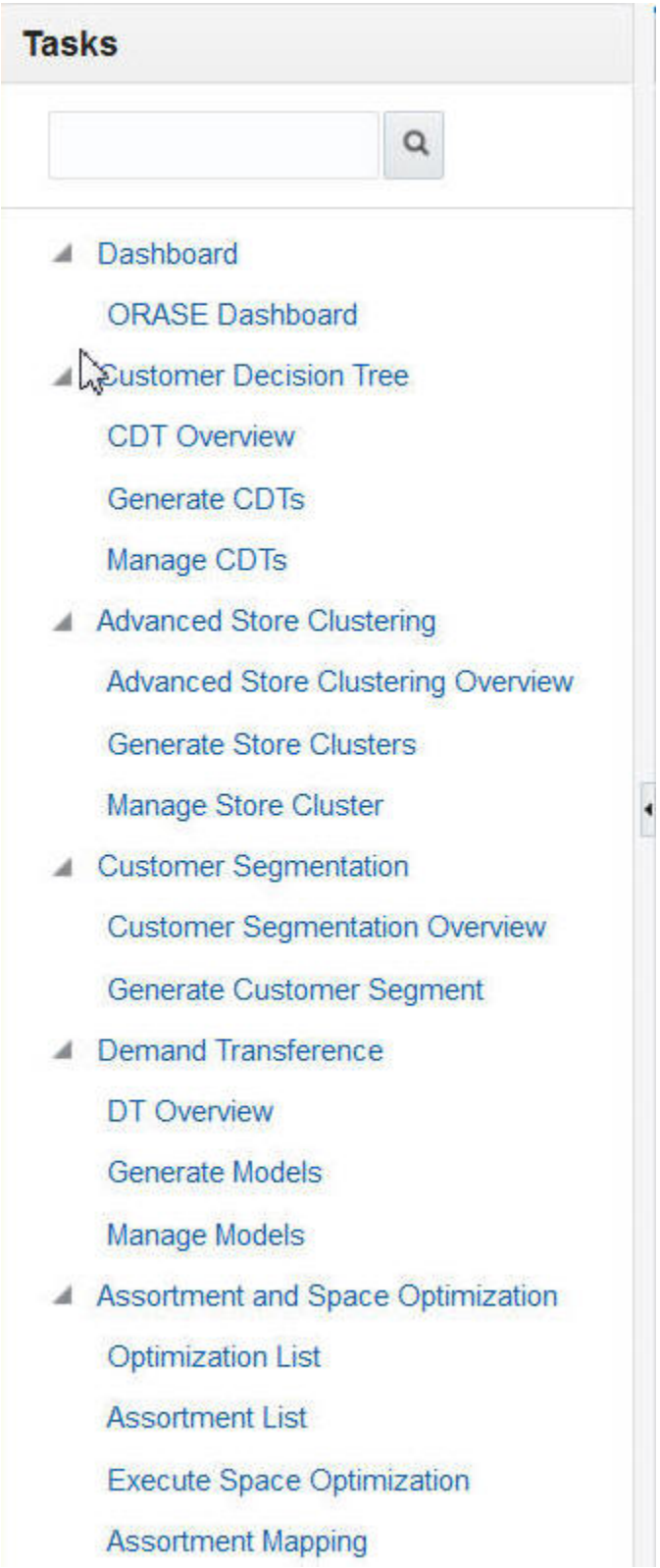
Administration

For information about the administration of Oracle Retail AI Foundation Cloud Services, see *Oracle Retail AI Foundation Cloud Services Administration Guide*.

Oracle Retail AI Foundation Dashboard

The dashboard provides access to all Oracle Retail AI Foundation modules. The main tasks associated with each module are listed in the panel on the left, as shown in the following figure. Each module has its own tab.

Figure 1-1 Oracle Retail AI Foundation Tasks



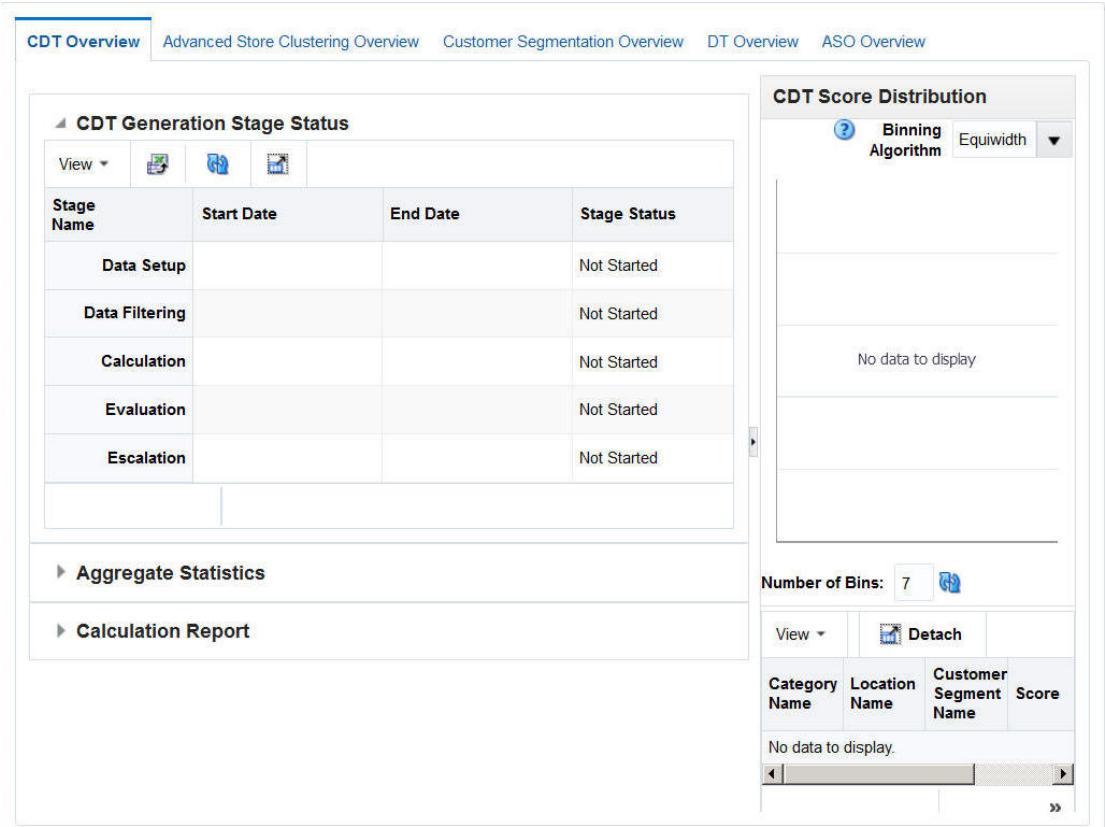
Process

This guide contains a chapter for each of the modules with detailed instructions. In general, each module contains a set of tabs, including an overview tab and tabs to use the module functionality.

Customer Decision Tree

The Customer Decision Tree Overview tab is shown in following figure.

Figure 1-2 CDT Overview Tab



When you use the CDT Cloud Service module, you follow this general iterative process to create and manage CDTs:

- Setup. You define the domain of calculation by selecting the categories and time frame for the calculation of the CDTs.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable answers.
- Calculation. You calculate a specific version of a CDT. Different versions of the CDT based on different configurations can be calculated and compared.
- Evaluation. You examine the results of the calculation and determine the reliability and accuracy of the answers. You can prune inaccurate or unreliable results.

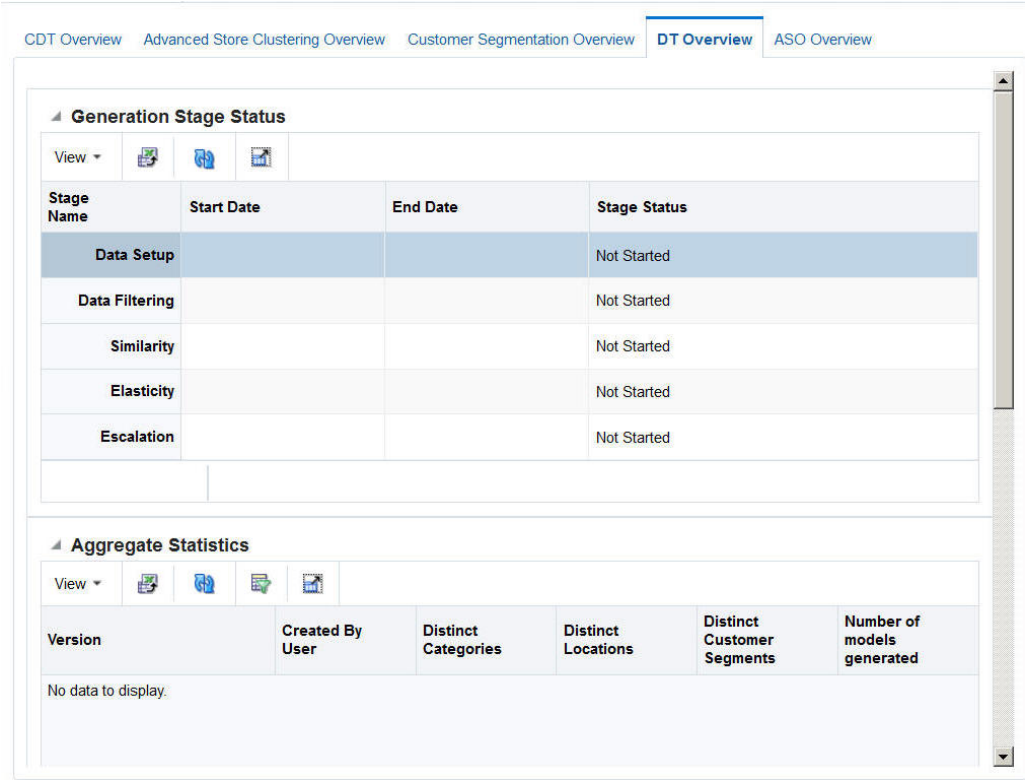
- Escalation. When you are satisfied with the results, you can set the escalation path to fill in the holes for partitions whose CDTs were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then, you can set a version of a CDT as complete.

Within each stage, you set the required parameters. In some stages, once you do that you click **Run** to initiate the process. To determine when a run is complete, go to the Overview tab and click the **Refresh** icon in order to see the current status of each stage. If you make changes to a stage, you must re-run that stage and all the relevant stages that follow it as the results of those stages are made invalid by the changes you made.

Demand Transference

The Demand Transference Overview tab is shown in the following figure..

Figure 1-3 DT Overview



When you use the DT Cloud Service module, you follow this general iterative process to create and manage DT models:

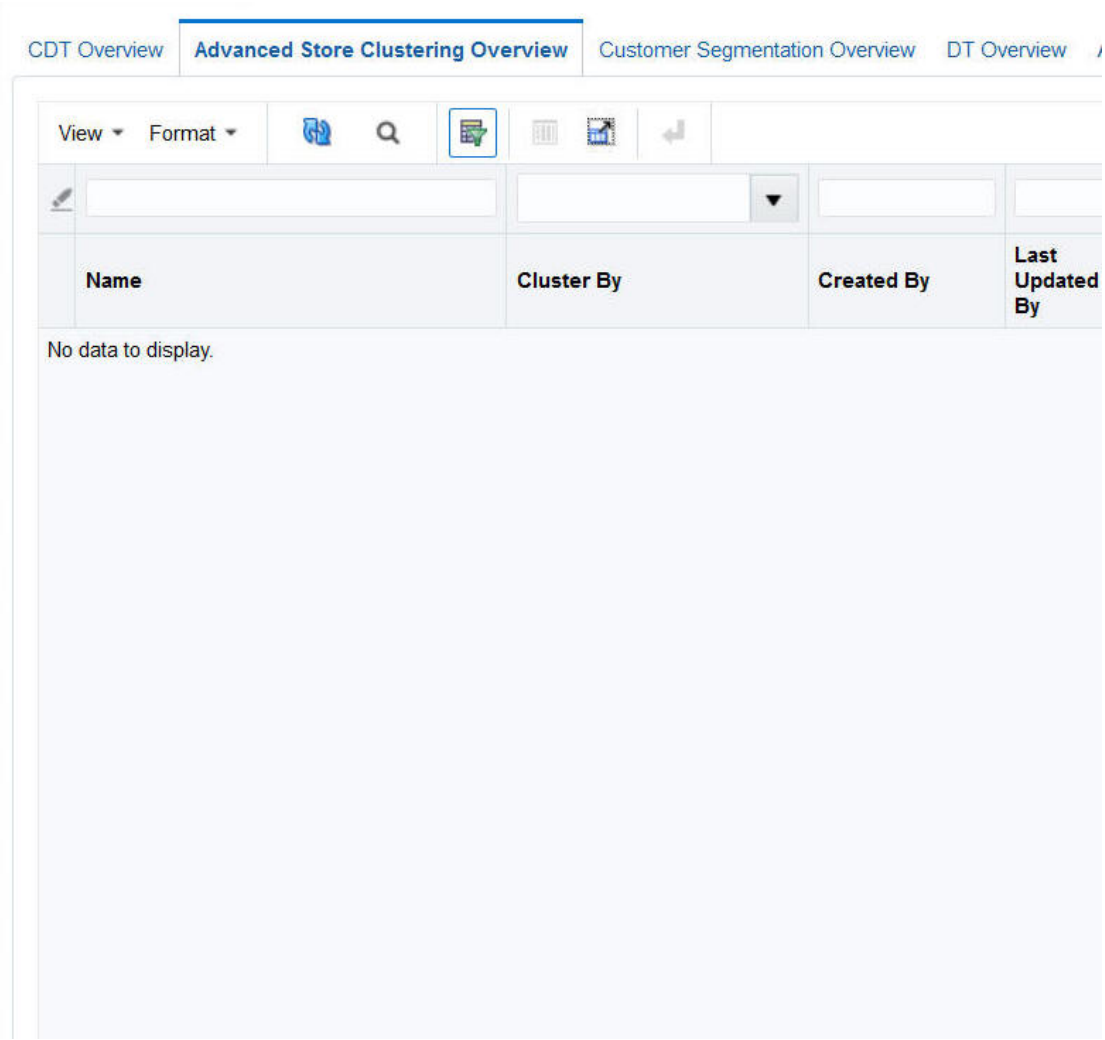
- Data Setup. You define the categories to be used in the DT calculation.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable results.
- Similarity Calculation. You calculate similarities and assess the results of the calculation.

- Elasticity Calculation. You calculate the assortment elasticities and assess the results in terms of substitutable demand, which is the percentage of demand of a SKU that is retained when the SKU is deleted from the stores where it is selling.
- Escalation. When you are satisfied with your results, you can set the escalation path to fill in the holes for partitions whose DT models were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then you can set a version of the DT model as complete.
- Manage Models. Use this tab to set time intervals for evaluating your results and to override the value for the maximum substitutable demand percentage.

Advanced Clustering

The Advanced Clustering Overview tab is shown in the following figure.

Figure 1-4 Advanced Clustering Overview Tab



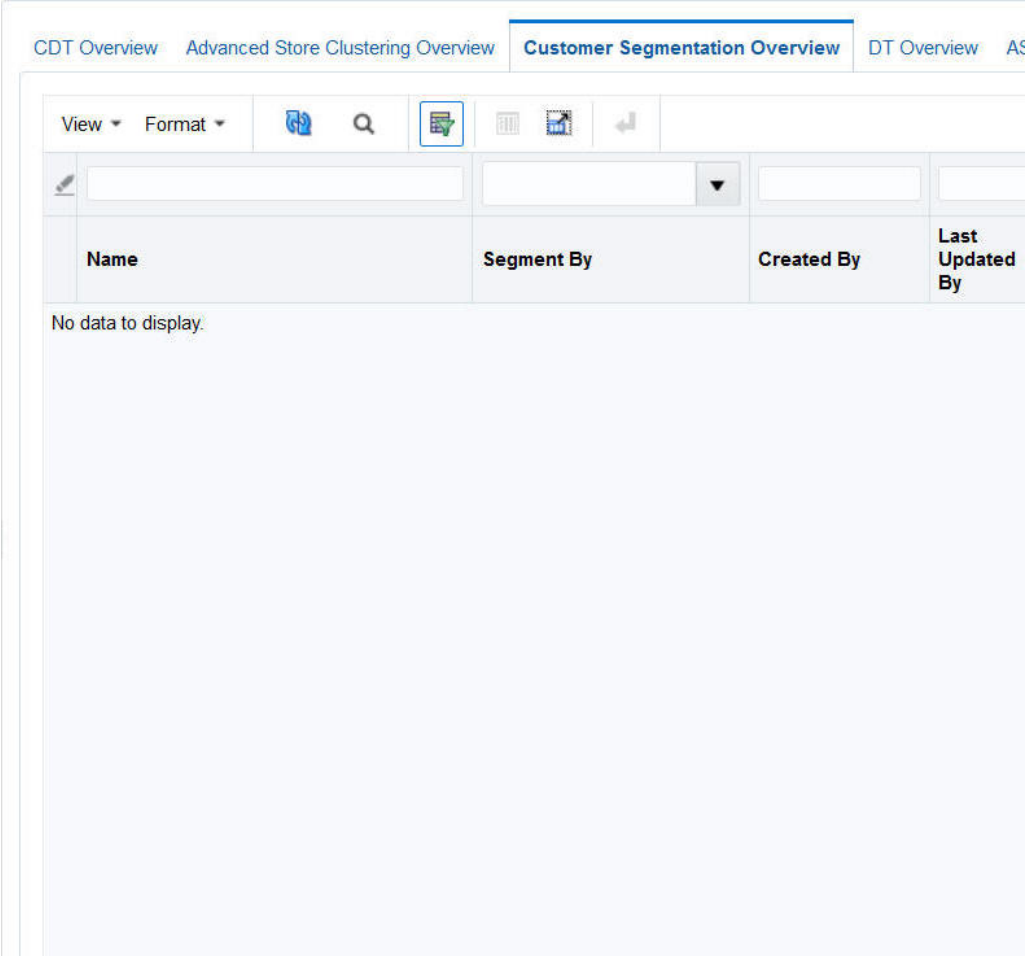
To use Advanced Clustering, follow this general process to create and manage clusters, working in the Generate Store Clusters tab and the Manage Store Clusters tab:

- Cluster Criteria. View all available clusters for specified merchandise, location, and calendar. Review cluster criteria or scenario details for each cluster. Use an existing cluster as the basis for creating a new cluster.
- Explore Data. Examine the input data for the cluster. Review multiple data points and significant attributes using the contextual information.
- Cluster Setup. Define multiple what-if scenarios. Such scenarios can be compared with one another throughout the clustering process.
- Cluster Results. View the scenario results and compare scenarios.
- Cluster Insights. Gain an understanding about cluster results and cluster performance prior to approval by examining the contextual information.
- Manage Store Clusters. Manage existing cluster criteria. Perform manual overrides and approve clusters.

Customer Segmentation

The Customer Segmentation Overview tab is shown in the following figure.

Figure 1-5 Customer Segmentation Overview Tab



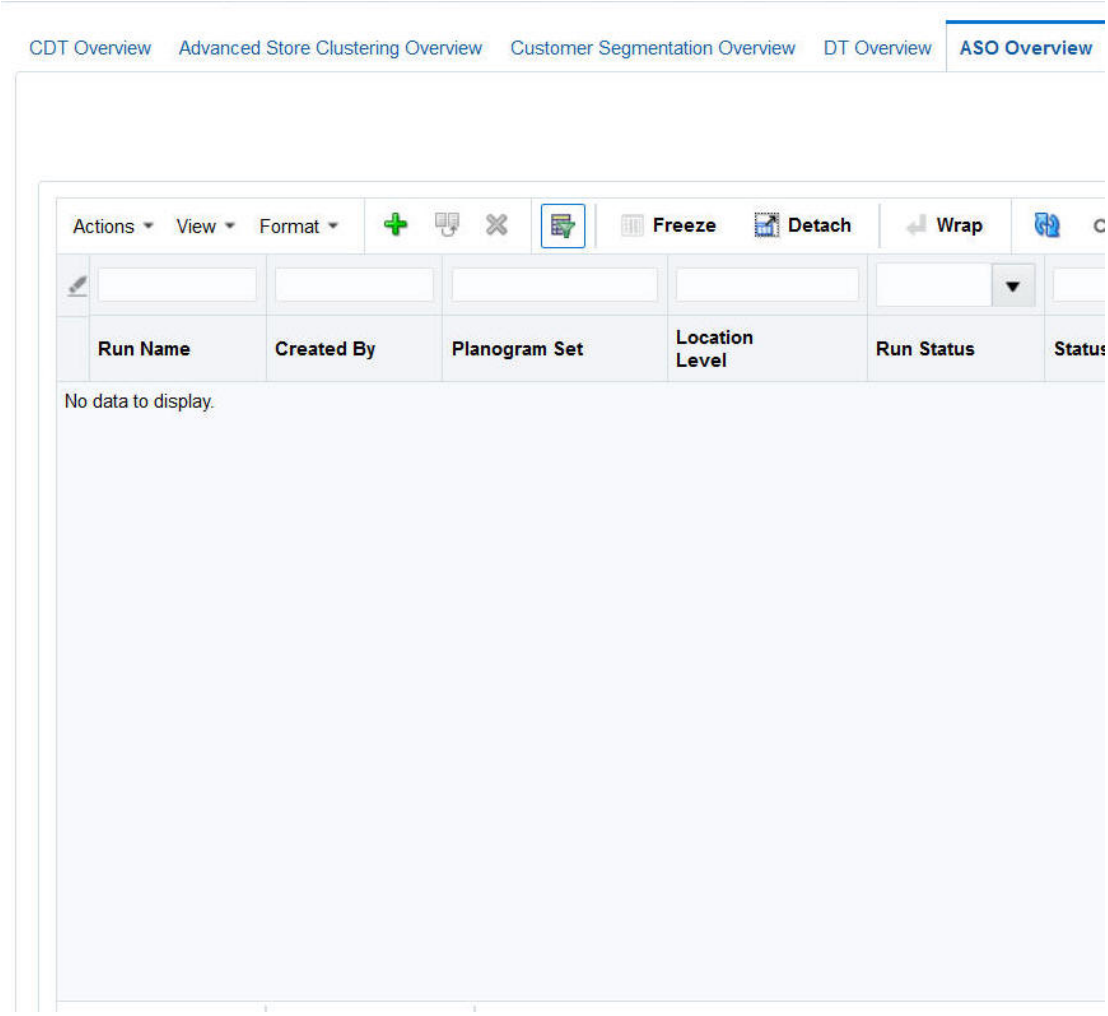
To use Customer Segmentation, follow this general process to create segments, working in the Generate Customer Segment tab:

- Segment Criteria – View all available segments for the specified merchandise, location, and calendar. Review the segment criteria or scenario details for each segment. Use an existing segment as the basis for creating a new segment.
- Explore Data – Examine attributes and their summaries for the segment. Review multiple significant attributes and their correlations using the contextual information.
- Segment Setup – Define multiple what-if scenarios. Such scenarios can be compared with one another throughout the segmentation process.
- Segment Results – View the scenario results and compare scenarios.
- Segment Insights – Gain an understanding of segment results and segment performance prior to approval by examining the contextual information.

Assortment and Space Optimization

The Assortment and Space Optimization Overview tab is shown in the following figure.

Figure 1-6 ASO Overview Tab



ASO is used to optimize a category manager's assortment plan by creating optimal planograms. The category manager creates a preliminary assortment and wants to determine how well that list of products can fit in stores, given the available space, product sizes, and merchandising goals, constraints, and rules. The application creates virtual planograms that organize products onto fixtures in a way that best achieves the optimization objectives. Once the manager finalizes the optimization results, the product level data can be exported for use in planning applications.

Overview of the User Interface

The user interface includes functionality described in this section.

Process Train

You can use the process train to navigate through the stages of each module. You can also use the **Back** button and the **Next** button to move through the train. The color changes for each stage once you visit that stage. Certain stages require you to run that stage before you can go to the next stage.

View Menu

The View Menu provides access to a variety of functionality that you can use to customize the display of the tables in the user interface.

Embedded Help

Embedded help, which you access by clicking the **Question Mark** icon, provides additional information about the type of details required by certain fields.

Process Indicator

At the top of the user interface, in the right-hand corner, is a process indicator that you can use to monitor the status of a user action such as clicking **Next** to go to the next stage.

Search

In certain cases, you can customize your search, using advanced search capabilities to specify the search criteria.

Icons

The following icons are used in the user interface.

Table 1-1 Icons Used in the User Interface

Icon	Icon Name	Icon Description
	Add	Add a category for CDT or DT. Add a cluster in Advanced Clustering. Create new scenarios in Advanced Clustering.

Table 1-1 (Cont.) Icons Used in the User Interface

Icon	Icon Name	Icon Description
	Approve Version	Approve a version of a CDT.
	Calculate	Initiate calculation of substitutable demand percentages in DT.
	Compare Two CDTs	Look at two CDTs side-by-side in the CDT Editor.
	Complete	Indicates the CDT is ready to be activated.
	Delete an entry in a table. Delete a scenario in Clustering.	Delete.
	Detach	Detach the table from the user interface for better viewing.
	Duplicate	Make a copy of a scenario.
	Edit	Edit the category attributes (CDT and DT)
	Embedded Help	Indicates that embedded help is available for the adjacent field.
	Execute	Execute the scenario.
	Export to Excel	Export the selected data to Excel.
	Go To Top	Adjusts table display.
	Go Up	Adjusts table display.
	No	Indicates No in Advanced Clustering.
	Query By Example	Provides access to a text entry field at the top of each column that you can use to search for data by an initial set of characters.
	Refresh	Update the table display or the stage status.
	Revert	Reverts the DT calculation.
	Save	Save the scenario.
	See Similarities	See similarities in DT Similarities Display.
	Set Version As Complete	Set a CDT version as complete.
	Select Date	Access a calendar in order to select a specific date.

Table 1-1 (Cont.) Icons Used in the User Interface

Icon	Icon Name	Icon Description
	Show As Top	Adjusts table display.
	View One CDT	Access a CDT in the CDT Editor.
	Withdraw From Approved Version	Un-approve a version of a CDT.
	Yes	Acts as an indication that something exists in Advanced Clustering.

Buttons

Buttons are used for navigation and to perform certain actions.

Table 1-2 Buttons

Name	Description
Action	Provides access to Save, Approve, and Reject.
Advanced	Provides access to advanced search functionality.
Approve	Used to approve a CDT or a cluster.
Back	Used to navigate the process train.
Cancel	Cancels the action.
Complete	Used to make a CDT or DT model active.
Next	Moves to the next stage.
Reject	Used to navigate the process train.
Reset	Resets the values to the original ones.
Run	Initiates a run.
Search	Provides access to search functionality.
Stop	Stops a process.

Browser Settings

The supported browsers include Mozilla Firefox 68+ ESR, Google Chrome (Desktop) 79+, and Microsoft Edge 44+.

Concurrent Browser Sessions

Users should not log into more than one browser session at the same time using the same user name.

Localization

The default language for the application is English. If you are using a different language on your computer, you should adjust the language settings on your browser as appropriate.

Supported Characters for Text Entry

The following characters are valid for text input: all letters, all numbers, and the following characters: '_', '#', '%', '*', '$', ' ', ',' & '-'

Histograms

Certain stages have associated histograms that can help you analyze the data presented in that stage. You can adjust the way the histogram presents the data in two ways. You can select the number of bins that are used to display the data. In addition, you can select how the bins are defined: Equiwidth or Custom. Each of these options uses a specific algorithm to determine how the bins are defined.

The Equiwidth approach takes the minimum and maximum values in a set of numbers and divides that range into equally sized bins. For example, using the numbers from 1 to 100 with 10 bins, the histogram shows bins for 1-10, 11-20, and so on. If specific bins have no value represented (for example, if all the values are in the range of 1-10 and 91-100), then the histogram will not show that bin in the UI. Additionally, the histogram data series ranges are shown using the actual minimum and maximum values for each of the bins. So rather than showing a range of 1-10, if the only value available was a 5, then the range for the first bin would appear as 5-5. In the Custom approach, each of the bins has an equal number of values represented, while the minimum and maximum number associated with the bin is adjusted. However, the bins are defined using distinct values instead of all the available values. The bins may or may not be of equal height, depending on how diverse the numbers are.

The two approaches differ in that the Custom approach only shows fewer bins than requested if there are fewer distinct values than what was requested for the number of bins.

The two approaches are similar in that both handle the Min/Max value display in a similar manner, using actual data values that are associated with the bin.

To determine which approach to use, you should consider what type of data you are trying to see and the amount of detail you want. For example, if you are trying to set a data filter value, and you want to do so using a common value, you may be able to see where most of the data falls using one of the algorithms, while the other algorithm may help you pinpoint a specific value within the range. The Equiwidth approach is negatively affected by values that are at the extreme ends of a value being binned. This can cause the majority of the data values to appear in a single bin. The Custom approach puts a greater emphasis on a value that is repeatedly found in a dataset. Depending on the values being charted, you may find that one of the approaches presents better data than the other approach.

2
Customer Decision Trees

This chapter describes the use of the Customer Decision Tree Science Cloud Service module.

Introduction

A customer decision tree is a decision support tool that uses a tree-like graph to model a customer hierarchical decision-making process for a specific product. The branches of the tree provide a visual representation of the choices a customer makes and the order of importance of various product characteristics. Transaction data is used in the analysis.

A customer decision tree identifies the decisions a customer makes when choosing a particular product. The decision tree is produced by algorithms that analyze historical customer sales data. It illustrates how customers shop and how they evaluate the importance of different product attributes when making buying decisions. Such information can be useful to a retailer in terms of product selection and display.

The CDT Cloud Service module consists of three tabs: Overview, Generate CDTs, and Manage CDTs. You use the Overview tab to keep track of the status of each stage within the Generate CDTs tab. You use the Manage CDTs tab to assess the set of CDTs you and other users have created using the Generate CDTs tab, setting versions as complete, approving versions, and deleting versions, as necessary.

Overview of CDT Process

When you use the CDT Cloud Service module, you follow this general iterative process to create and manage CDTs:

- Setup. You define the domain of calculation by selecting the categories and time frame for the calculation of the CDTs.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that might lead to inaccurate or unreliable answers.
- Calculation. You calculate a specific version of a CDT. Different versions of the CDT based on different configurations can be calculated and compared.
- Evaluation. You examine the results of the calculation and determine the reliability and accuracy of the answers. You can prune inaccurate or unreliable results.
- Escalation. When you are satisfied with the results, you can set the escalation path to fill in the holes for partitions whose CDTs were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then, you can set a version of a CDT as complete.

Within each stage, you set the required parameters. In some stages, once you do that you click **Run** to initiate the process. To determine when a run is complete, go to the Overview tab and click the **Refresh** icon in order to see the current status of each stage. If you make changes to a stage, you must re-run that stage and all the relevant stages that follow it as the results of those stages are made invalid by the changes you made.

Overview Tab

The Overview tab displays information that you can view and use to monitor the progress of the CDT stages as well as to view some aggregate statistics and the CDT results from the last successful run.

This tab contains the following sections:

- CDT Generation Stage Status
- Aggregate Statistics
- CDT Calculation Report
- CDT Score Distribution

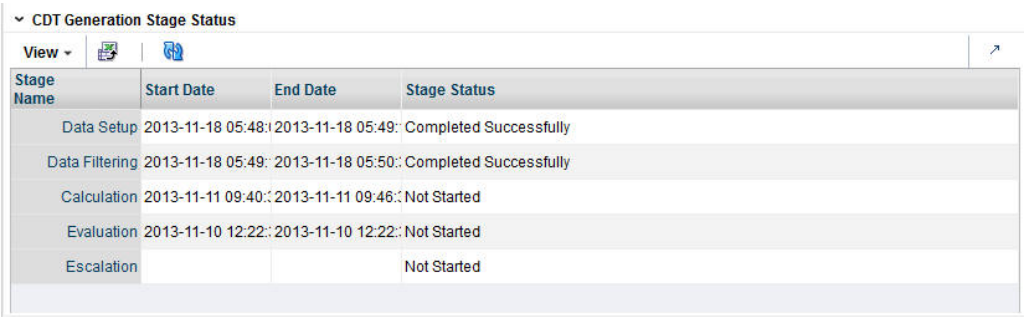
CDT Generation Stage Status

The CDT Generation Stage Status table displays the status of each of the CDT stages.

Click the **Refresh** icon to update the fields and see the latest status for each stage.

In cases where the status is either Failed or Stopped/Paused, or the run never completes, you should consult the database logs in RSE_LOG_MSG as well as the WebLogic console logs in your WebLogic domain in order to troubleshoot the problem.

Figure 2-1 CDT Generation Stage Status



You can use the status information to monitor the progress of each stage. It contains the following fields, which can be arranged and viewed, but not modified.

Table 2-1 CDT Generation Stage Status Fields

Field Name	Description
Stage Name	A row exists in the table for each CDT stage that provides detailed status information. The five stages are Setup, Data Filtering, Calculation, Evaluation, and Escalation.
Start Date	The date and time when a run for the stage most recently started.
End Date	The date and time when a run for the stage most recently ended.

Table 2-1 (Cont.) CDT Generation Stage Status Fields

Field Name	Description
Stage Status	The current status of the stage: Not Started, Not Started (Scheduled for Later), Processing, Completed Successfully, Completed with Errors, Stopped/Paused, Cancelled, or Failed.

The following provides an explanation of the different values for the status of a stage.

Table 2-2 Status Explanations

Stage Status	Description
Not Started	This indicates that the stage has not yet been initiated. This can occur when you first begin to create a CDT or when you update an existing version.
Not Started (Scheduled for Later)	The stage is scheduled to start after the previous stage is complete.
Processing	This indicates that the stage is currently being processed.
Stopped/Paused	This indicates either that the stage has been stopped by the application or that the user has chosen to stop the process (for example, to make a change to an option). In the latter case, once any background processing has stopped, the user can re-run the stopped stage.
Completed Successfully	This indicates that the stage has been successfully processed.
Completed with Errors	This indicates that the stage was able to complete the processing of all requested CDTs, but one or more CDTs encountered errors during the run and were not able to complete successfully. This most commonly occurs when the data used for a CDT is not available or is too sparse to produce a result.
Failed	This indicates that a problem occurred during the processing. If the failure occurred during the Calculation stage, and only a few CDT results did not produce a CDT score, then the user can continue without those results since pruning will remove them. If the failure occurred during any of the other stages, the user must re-run the stage.
Cancelled	This indicates that the database has cancelled the execution of the stage either because of missing data or an exception. The user should review the RSE_LOG_MSG log to determine the problem.

Aggregate Statistics

The Aggregate Statistics table displays statistical details about the existing CDT versions.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 2-2 CDT Aggregate Statistics

Aggregate Statistics

View

Version Name	Creation User	Distinct Category Selections	Distinct Locations	Distinct Customer Segments	Start Date	End Date	Average Score
Coffee CS Brand	cdmUser0	1	1	6	1/4/2010	1/1/2012	68.02
Coffee CS Brand S	cdmUser0	1	1	6	1/4/2010	1/1/2012	73.62
Coffee CS Brand S	cdmUser0	1	1	6	1/4/2010	1/1/2012	76.61
Coffee CS Brand S	cdmUser0	1	1	6	1/4/2010	1/1/2012	76.39
Coffee CS Initial	cdmUser0	1	1	6	1/4/2010	1/1/2012	69.61
Smoke Test	cdtUser0	1	10	1	1/4/2010	1/1/2012	70.85

Each existing version has a row in the table. The table contains the following fields, which can be arranged and viewed, but not modified.

Table 2-3 Aggregate Statistics Fields

Field Name	Description
Version Name	User-created name that identifies this CDT version. The name must be unique to the user.
Creation User	The login name of the person who created this version.
Distinct Category Selections	The number of categories associated with this version.
Distinct Locations	The number of locations associated with this version.
Distinct Customer Segments	The number of customer segments associated with this version.
Start Date	The beginning date associated with the CDTs in this version.
End Date	The end date associated with the CDTs in this version.
Average Score	The average of the CDT scores for a version. A CDT score ranges from 0 to 100. A higher value indicates a better score.

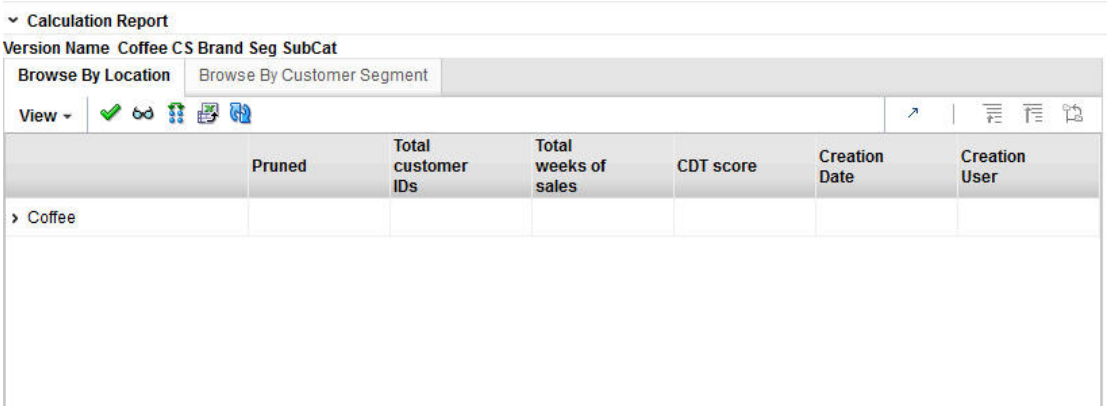
Calculation Report

The Calculation Report displays the CDT results for the indicated version from the last successful run, if one has occurred.

Here you can review information by Location or by Customer Segment. You can also set a version as complete, view and edit a CDT, or compare two CDTs.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 2-3 CDT Calculate Report



The Calculation Report contains the following fields:

Table 2-4 Calculation Report Fields

Field Name	Description
Partition	A partition is the combination of category, segment, and location. The column identifies the names of all nodes in the tree structure in the table. The node type may be either category, location, or customer segment.
Pruned	The number of CDTs removed from the list of usable CDTs. CDTs are removed that do not meet the filtering thresholds. During the Escalation stage, the escalation process makes adjustments for the CDTs that are removed.
Total Customer IDs	The number of customers used in the calculation of the CDT.
Total Weeks of Sales	The total number of weeks of sales used in the calculation of the CDT. This provides an indication of the amount of sales data used in the calculation. All the CDTs in a given version should have a similar value; if this is not the case, the results should be evaluated.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Creation Date	The date when the version whose data is displayed was created.
Creation User	The login name of the user who created the version.

The following icons are available for the Calculation Report:

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version, then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, only the overlapping categories are replaced in the new version.

If a version is overwritten, then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This icon provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

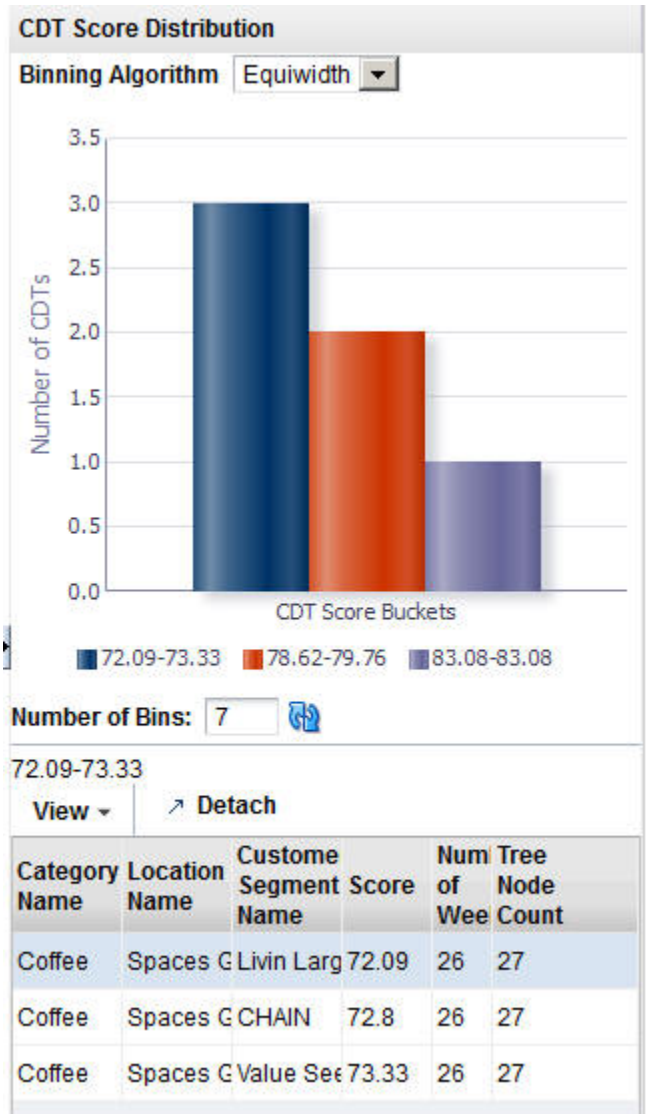
This icon provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs

CDT Score Distribution Histogram

The CDT Score Distribution histogram displays, for the current calculation and the current user, the range of CDT scores and their frequencies. CDT scores range from 0 to 100; a higher number indicates a better score. You can use the information in this histogram to assess the quality of your CDTs.

For detailed information about using the histograms, see Getting Started..

Figure 2-4 CDT Score Distribution



Below the histogram you see a table that provides details about the histogram. This information can help you assess the quality of a CDT. Click on a specific bin in the histogram to populate the table with information about that bin. You see a list of all the specific partitions for the bin, along with the score for each, the number of weeks of data used to calculate the score, and the number of tree nodes for each partition for the selected bin.

Table 2-5 CDT Score Distribution Fields

Field Name	Description
Category Name	The category name of the partition for the specific CDT score.
Location Name	The location name of the partition for the specific CDT score.
Customer Segment Name	The customer segment name of the partition for the specific CDT score.

Table 2-5 (Cont.) CDT Score Distribution Fields

Field Name	Description
Score	The CDT score for the defined partition. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Number of Weeks	The number of weeks of data used to calculate the CDT score.
Tree Node Count	The number of tree nodes present in the CDT tree structure used to calculate the CDT score.

Generate CDT Tab

The Generate CDTs tab is used to configure, run, evaluate, modify, and deploy a CDT. The process is divided into five stages that must be run in order. You can return to a stage you have already completed and make changes, but if you do, you must rerun that stage and all the stages that follow that stage, as the calculations are invalidated by the modifications you just made to the settings in the stage you changed.

To monitor the progress of any stage, go to the Overview tab and click the **Refresh** icon.

The five CDT stages are:

Table 2-6 Generate CDTs Tab: Stages

Stage Name	Description
Data Setup	Select CMPO groups and define the time intervals for the data to be used in the calculation.
Data Filtering	Filter out input data that may result in inaccurate or unreliable answers.
Calculation	Calculate the CDT.
Evaluation	Assess the reliability and accuracy of the results of the calculation. Prune unreliable or inaccurate answers.
Escalation	Set the escalation path for the CDT. Use the Escalation Report to evaluate the results of the escalation.

Data Setup

The Data Setup stage provides two sections to configure: Category Selections and Time Interval Setup. In this stage, you select the category or categories you want to calculate CDTs for and specify the time interval for the calculation.

Process

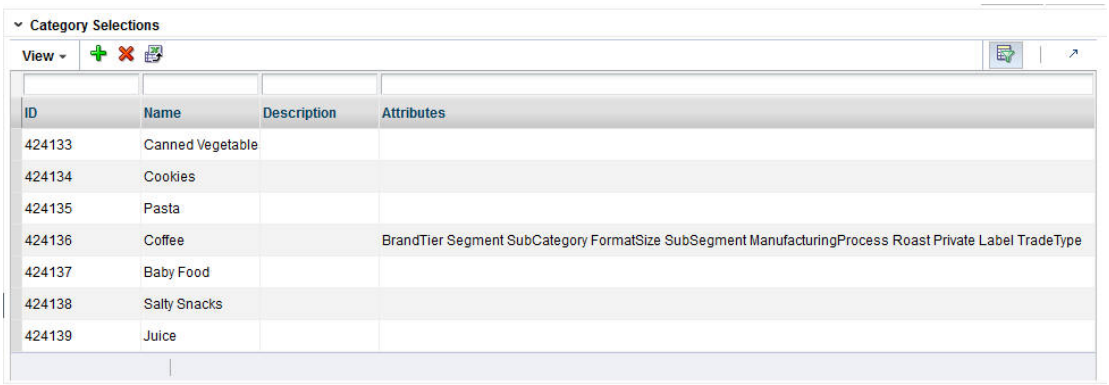
Here is the high-level process for setting up the data for CDT.

1. Select the category or categories you want to calculate CDTs for.
2. Set up the time intervals for the CDT calculation.
3. Click **Next** to go to the Data Filtering stage.

Category Selections

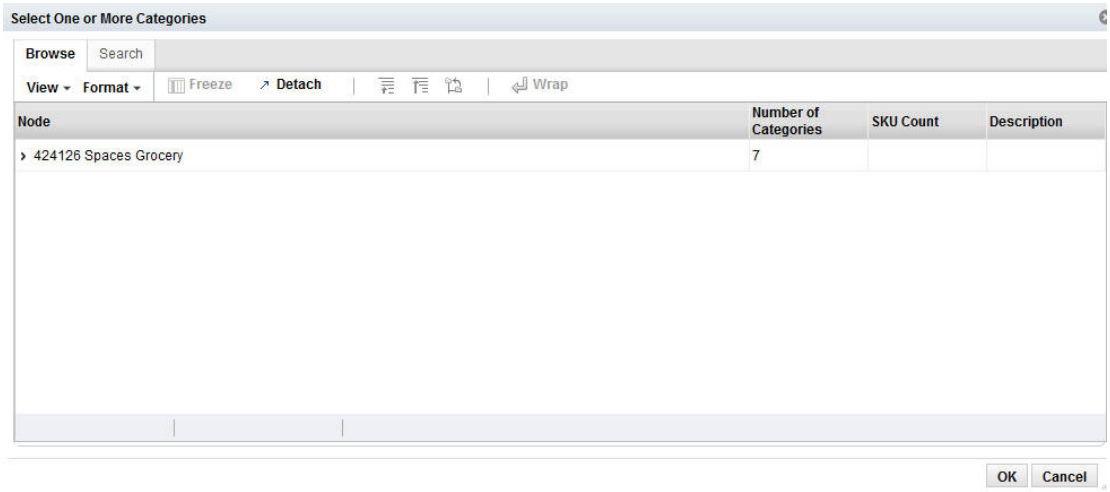
Use this table to add categories you want to include in the CDT calculation or delete categories that you want to remove from the CDT calculation.

Figure 2-5 CDT Category Selections



To display a list of available nodes and the categories included in those nodes, click the **Add** icon. You see the Select One or More Categories dialog box, which contains two tabs, Browse and Search. You can use either tab to find the categories you are looking for.

Figure 2-6 CDT Categories Browse Tab



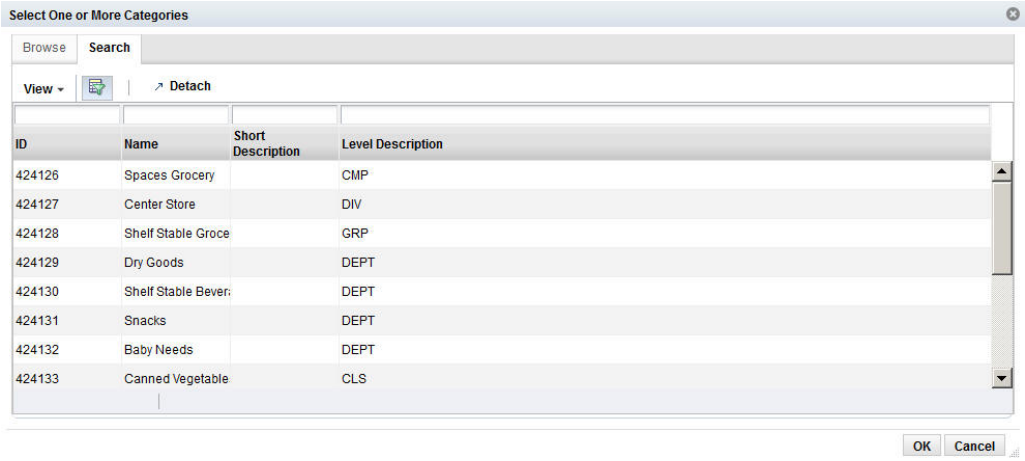
The Browse tab displays a table with the following fields:

Table 2-7 Category Selections: Browse

Field Name	Description
Partition	The node tree structure can be expanded in order to view its categories.
Number of Categories	The number of categories within the node that has been selected. The number can help you understand the amount of processing required for the calculation.
SKU Count	The number of SKUs in a category. A category with too few SKUs may not produce good CDTs.
Description	A description that provides additional information about the category.

Select the category or categories within the node that you want to be part of CDT and click **OK**.

Figure 2-7 CDT Categories Search Tab



The Search tab displays a table with the following fields:

Table 2-8 Category Selections: Search

Field Name	Field Description
ID	An external code used to identify the category in other systems such as CMPO.
Name	The category name.
Short Description	A description that provides additional information about the category.
Level Description	A description of the level of the merchandise hierarchy that the node belongs to.

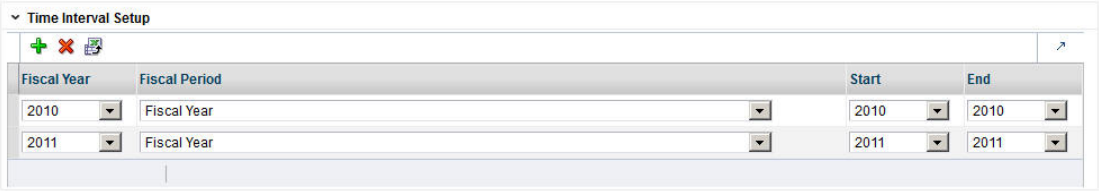
Select the category or categories within the node that you want to be part of the CDT and click **OK**.

Your selections are displayed in the Category Selections table.

Time Interval Setup

The time interval defines the time span for the historical data that is used to calculate the CDT. Only the specified weeks of sales data are used in the calculation. A group of intervals can be defined. Gaps between intervals are permitted; however, intervals cannot overlap. A six-month period is recommended.

Figure 2-8 CDT Time Interval Setup



To define a time interval for the CDT to be generated, click the **Add** icon. A row displays in the dialog box. Select from the following drop-down menus in order to define the time interval:

Table 2-9 Time Interval: Menus

Menu Name	Description
Fiscal Year	The fiscal year for the time interval.
Fiscal Period	The fiscal period within the fiscal year (Fiscal Quarter, Fiscal Period, or Fiscal Week).
Start	This defines when the time interval specified in Fiscal Period begins.
End	This defines when the time interval specified in Fiscal Period ends.

After selecting the category and specifying the time interval, click **Next** to go to the Data Filtering stage.

Data Filtering

The Data Filtering stage applies to all the categories and time intervals that you select in the Data Setup stage.

Process

Here is the high-level process for setting up and running data filtering.

1. Enter the appropriate values into the Filter Setup text entry boxes.
2. Click **Run** in order to filter the data.
3. Review the filtering results in the Data Filtering Summary table and the Data Filtering histograms.
4. After reviewing the results, if necessary, make changes to the values for the filters in Filter Setup and re-run the stage.

5. When you are satisfied with the results, click **Next** to go to the Calculation stage.

Filter Setup

You configure the following filters in order to filter out data you consider unacceptable for the calculation of the CDT. Note that the two attribute filters listed in Table 2-10 are stored and used during the Calculation stage. You also set additional data filtering parameters in the Calculation stage.

Figure 2-9 CDT Filter Setup

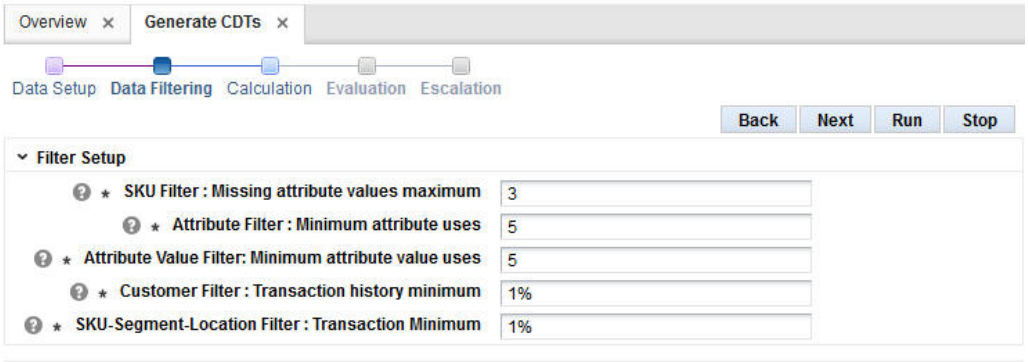


Table 2-10 Data Filters

Data Filter Name	Data Filter Description
SKU Filter: Missing attribute values maximum	Each SKU is defined by its attribute values. If a certain absolute value for the attribute values is not defined, then the product definition is not accurate. A SKU with too many missing attribute values should be filtered out. The default value is 25 for the total attribute values (that is, a SKU with greater than 25 missing attribute values is not included in the calculation of the CDT).
Attribute Filter: Minimum attribute uses	An attribute that is used by only a few SKUs should be filtered out. The default value is an absolute value of 5 for the total SKUs in the category (that is, the data for an attribute that is used by fewer than five of the SKUs is not included in the calculation of the CDT). This filter does not remove t-log level data but instead removes the attribute and attribute values from the CDT creation process.
Attribute Value Filter: Minimum attribute value uses	An attribute value that is used by only a few SKUs should be filtered out. The default value is an absolute value of 5 for the SKUs in a category (that is, the data for an attribute value that is used by fewer than five of the SKUs is not included in the calculation of the CDT). This filter does not remove t-log level data but instead removes the attribute and attribute values from the CDT creation process.

Table 2-10 (Cont.) Data Filters

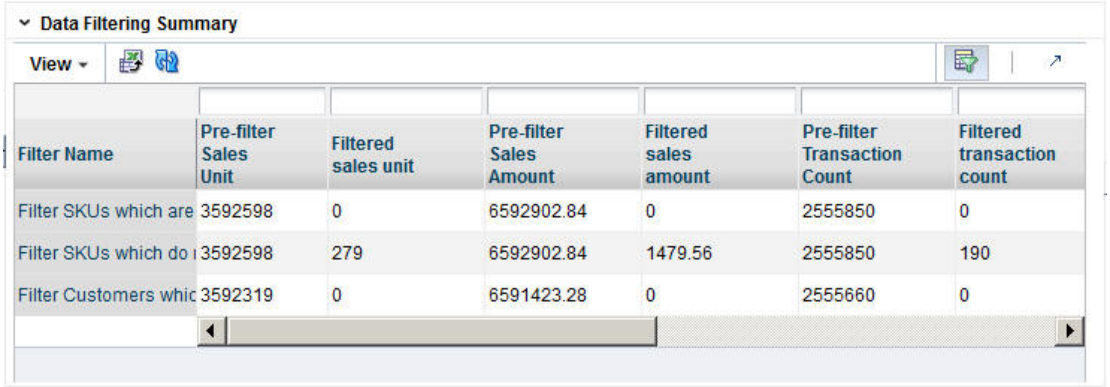
Data Filter Name	Data Filter Description
Customer Filter: Transaction history minimum	Customers with short transaction histories are considered outliers. You assign a percentage value that is applied to the median number of transactions for all customers. Such customers are filtered out. The default value is 10% (that is, a customer who has fewer than 10% of the median number of transactions for all customers is not included in the calculation of the CDT).
SKU-Segment-Location Filter: Transaction minimum	SKUs that have few transactions for a given location-segment partition are considered outliers. You assign a percentage value that is applied to the median number of transactions for the SKUs in a specific partition. Such transactions are filtered out. The default value is 10% (that is, a SKU that is involved in fewer than 10% of the median number of transactions for a specific partition is not included in the calculation of the CDT).

Once you have configured the filters, click **Run** to start the filtering process.

Data Filtering Summary

The following information is provided after the filtering is complete and quantifies the amount of data filtered out for the three indicated filters for sales units, sales amounts, transaction counts, SKU counts, and customer counts. Use this information to assess the effects of the pruning.

Figure 2-10 CDT Data Filtering Summary



Filter Name	Pre-filter Sales Unit	Filtered sales unit	Pre-filter Sales Amount	Filtered sales amount	Pre-filter Transaction Count	Filtered transaction count
Filter SKUs which are	3592598	0	6592902.84	0	2555850	0
Filter SKUs which do	3592598	279	6592902.84	1479.56	2555850	190
Filter Customers whic	3592319	0	6591423.28	0	2555660	0

Table 2-11 Filter Data Summary Fields

Field Name	Field Description
Filter Name	The following filter names are listed: Filter SKUs that are missing too many attribute values. This maps to the SKU Filter: Missing attribute values maximum in the Filter Setup. Filter SKUs that do not have enough sale transaction history. This maps to the Customer Filter: Transaction history minimum in the Filter Setup. Filter Customers that do not have typical transaction history. this maps to the SKU-Segment-Location Filter: Transaction minimum in the Filter Setup.
Pre-filter Sales Unit	Amount prior to the application of the filter.
Filtered Sales Unit	Amount remaining after the application of the filter.
Pre-filter Sales Amount	Amount prior to the application of the filter.
Filtered Sales Amount	Amount remaining after the application of the filter.
Pre-filter Transaction Count	Amount prior to the application of the filter.
Filtered Transaction Count	Amount remaining after the application of the filter.
Pre-filter SKU Count	Amount prior to the application of the filter.
Filtered SKU Count	Amount remaining after the application of the filter.
Pre-filter Customer Count	Amount prior to the application of the filter.
Filtered Customer Count	Amount remaining after the application of the filter.

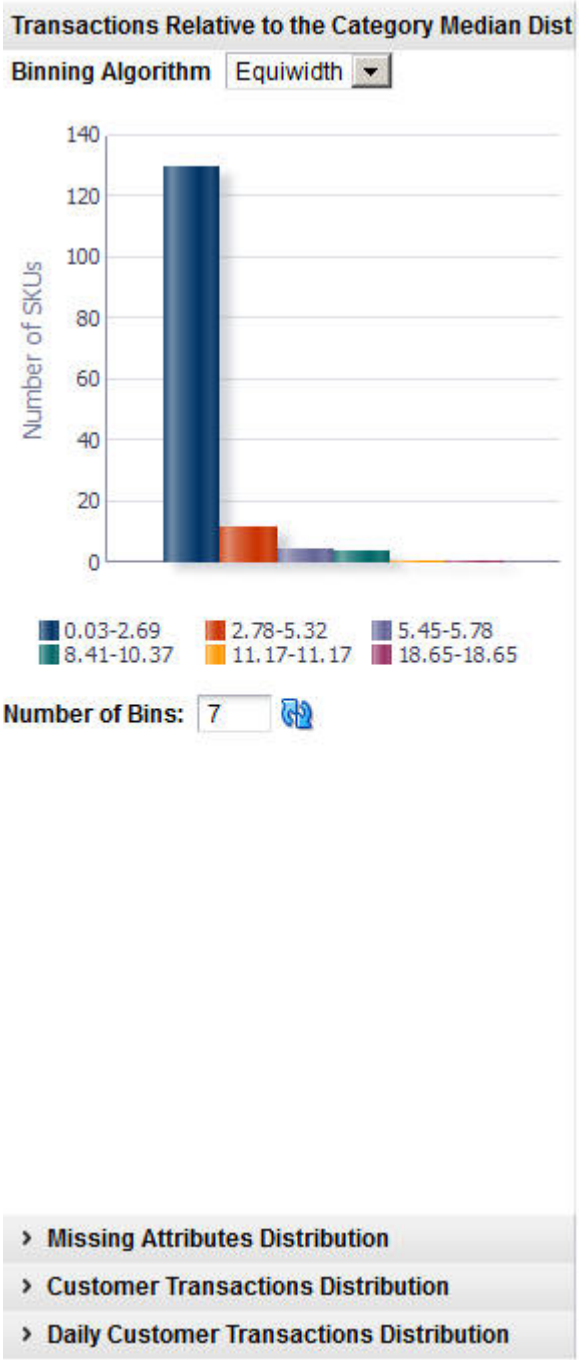
Data Filtering Histograms

The following histograms illustrate the effects of filtering. You can use the information displayed in the histograms to adjust the configuration of the filters in order to eliminate outlier data. If you modify the filters, you must re-run the stage.

Table 2-12 Filter Histograms

Histogram Name	Description
Transactions Relative to the Category Median Distribution	Displays the number of transactions relative to the category median for a SKU's category.
Missing Attribute Distribution	Displays the number of attributes that have missing values per number of SKUs.
Customer Transaction Distribution	Displays the number of transactions in historical data per number of customers.
Daily Customer Transaction Distribution	Displays the relative number of transactions per day per number of customers.

Figure 2-11 CDT Transactions Relative Histogram



Once you are satisfied with the pruning results, click **Next** to go to the Calculation stage.

Calculation

The Calculation stage creates the CDTs for all of the partitions that you selected in the Data Setup stage. A separate CDT calculation occurs for each partition. For example, if you selected two CMPO nodes, and the system also has two customer segments and three

locations, then that is a total of 2 x 2 x 3 = 12 CDTs at the lowest level. In addition, there are CDTs at the higher levels (for example, above location).

Process

Here is the high-level process for calculating CDTs.

1. Enter a unique name for the version of the CDTs to be calculated.
2. Enter values for lowest tree level and minimum percentage of SKUs.
3. Use the check boxes to indicate whether or not only the top-level processing calculation should occur for location or customer segment.
4. Customize the ranking of the category attributes if necessary.
5. Click **Run** to start the calculation.
6. Review the calculation results in the Calculation Report and the CDT Score Distribution histogram.
7. After reviewing the results, if necessary, make changes to the values for the calculation in Version Setup and re-run the stage.
8. When you are satisfied with the results, click **Next** to go to the Evaluation stage.

Version Setup

At the top of this stage you see three text boxes and two check boxes that you use to configure the levels at which the calculation occurs.

Figure 2-12 CDT Version Setup

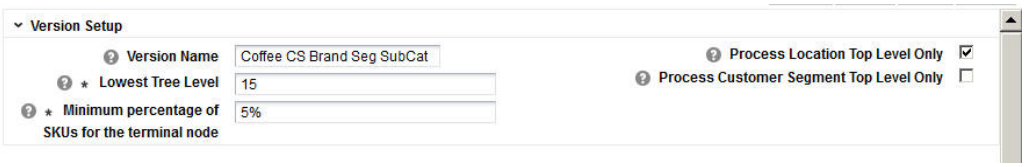


Table 2-13 Calculation Stage: Fields

Field Name	Field Description
Version Name	Assign a name to each version of a CDT calculation. This allows you to create and save more than one version of a CDT. The version name you assign here is used in the Calculation Report, Aggregate Statistics table, and in the Manage CDTs tab. Version names can be re-used; however, if the version name in question has active CDTs, then you will see a warning that the active CDTS will be removed from the version if you do re-use the version name.

Table 2-13 (Cont.) Calculation Stage: Fields

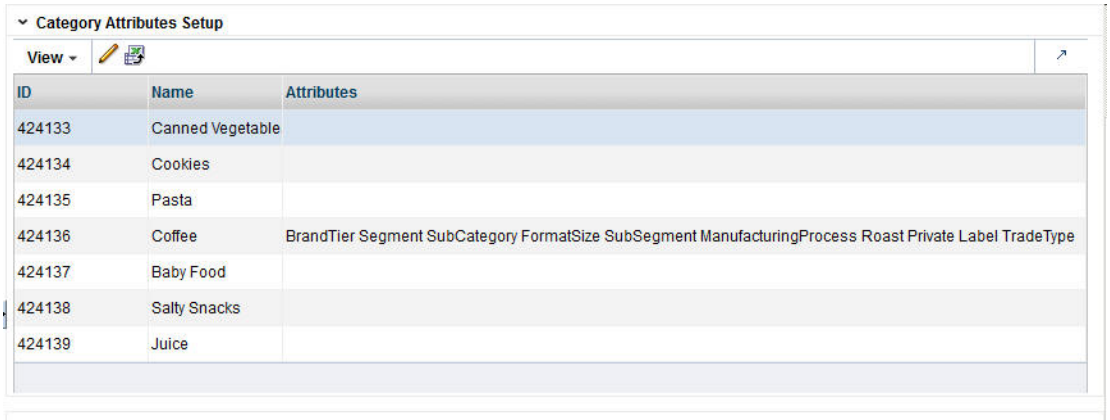
Field Name	Field Description
Lowest Tree Level	Use this option to define the maximum number of levels of CDTs to be calculated. The default value is 15. The number of levels in a CDT can at most be equal to the number of attributes used by the SKUs in the category plus 1. (The number 1 is used for the top level of the tree.) When determining the maximum number of levels, you should consider how many attributes should be represented in the tree. (For example, if you only want to see the top six attributes in the CDT, then set the lowest level to 7.)
Minimum Percentage of SKUs for the Terminal Node	Use this option to make sure the terminal nodes have a sufficient number of SKUs. For example, if the current CDT partition contains 100 SKUs, then any one node in the tree containing fewer than 5 SKUs will be considered a terminal node and the tree will not expand further along that branch of the tree.
Process Location Top Level Only	Check this option if you want CDTs to be calculated for the Location Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.
Process Customer Segment Top Level Only	Check this option if you want CDTs to be calculated for the Customer Segment Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.

Category Attributes Setup

The application determines a specific ranking order for the category attributes. You can optionally change this order and create your own ranking of the attributes for a category by ordering the attributes from most important to least important. The Category Attributes Setup table lists the categories you are calculating CDTs for. Use this list to select a category to edit.

Click the **Edit** icon to access this functionality. You can adjust the ranking for all the attributes or a subset of the attributes.

Figure 2-13 CDT Category Attributes Setup



The Category Attributes Setup dialog box contains the following fields. For each attribute you want to rank, select the Top Attribute check box and enter a value for the Attribute Rank. Click **OK**.

Figure 2-14 CDT Category Attributes Setup

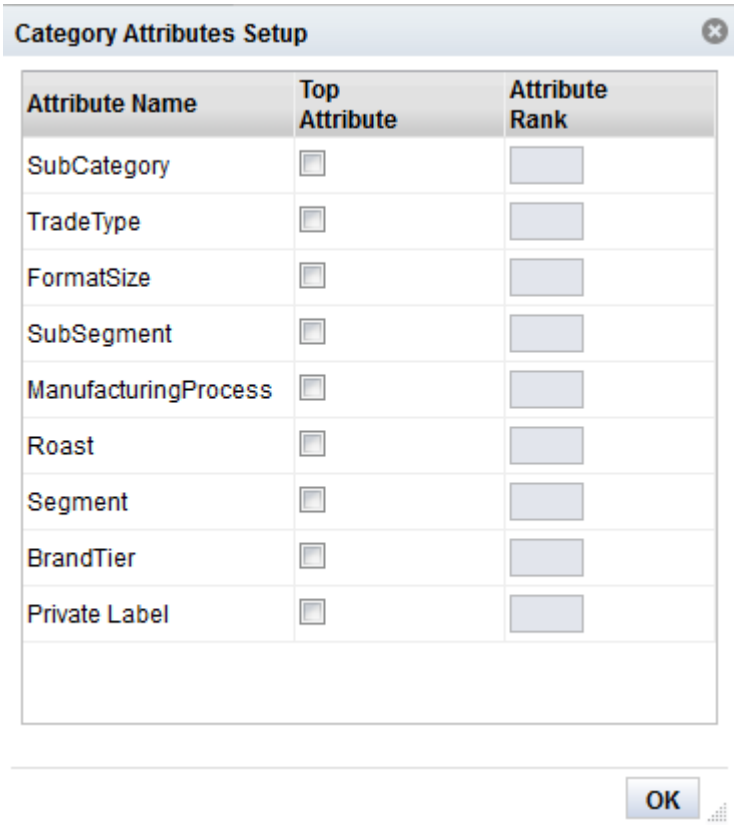


Table 2-14 Category Attributes Setup Dialog Box

Field	Description
Attribute Name	Identifies the attribute.
Top Attribute	Check to enable ranking. This indicates that this attribute must be one of the first attributes in the resulting tree. This box is checked automatically for functional fit attributes and cannot be deselected. A functional fit attribute is one for which a customer cannot choose a different product with a different attribute value. For example, windshield wiper blades must be selected according to the car manufacturer's requirements. The numerical rank is optional. If you do not enter rank values, then the system will determine the order of all the top attributes first and then put any remaining attributes below the top attributes in the tree.

Table 2-14 (Cont.) Category Attributes Setup Dialog Box

Field	Description
Attribute Rank	Enter a value for each attribute, starting with the value 1, to indicate the relative rank of the attribute. A value can be entered only if the attribute has been identified as a top attribute. You do not have to specify the rank if Top Attribute is checked. If multiple attributes are checked, then the system rank orders them.

Once you have configured the parameters for the calculation, click **Run** to start the process.

Calculation Report

The Calculation Report displays the CDT results for the indicated version from the last successful run (if one has occurred). Here you can review information by Location or by Customer Segment. You can also set a version as active, view and edit a CDT, or compare two CDTs.

Figure 2-15 CDT Calculation Report

Table 2-15 Calculation Report Fields

Field Name	Description
Partition	A partition is the combination of category, segment, and location. The column identifies the names of all nodes in the tree structure in the table. The node type may be either category, location, or customer segment.
Pruned	The number of CDTs removed from the list of usable CDTs. CDTs are removed that do not meet the filtering thresholds. During the Escalation stage, the escalation process makes adjustments for the CDTs that are removed.
Total Customer IDs	The number of customers used in the calculation of the CDT.
Total Weeks of Sales	The total number of weeks of sales used in the calculation of the CDT. This provides an indication of the amount of sales data used in the calculation. All the CDTs in a given version should have a similar value; if this is not the case, the results should be evaluated.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Creation Date	The date when the version whose data is displayed was created.

Table 2-15 (Cont.) Calculation Report Fields

Field Name	Description
Creation User	The login name of the user who created the version.

The following icons are available for the Calculation Report:

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version, then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, only the overlapping categories are replaced in the new version.

If a version is overwritten, then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This icon provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

This icon provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs

CDT Score Distribution Histogram

The CDT Score Distribution histogram displays, for the current calculation, the range of CDT scores and their frequencies. CDT scores range from 0 to 100; a higher number indicates a better score. You can use the information in this histogram to assess the quality of your CDTs.

Below the histogram you see a table that provides details about the histogram. This information can help you assess the quality of a CDT. Click on a specific bin in the histogram to populate the table with information about that bin. You see a list of all the specific category name/location name/customer segment name partitions for the bin, along with the score for each, the number of weeks of data used to calculate the score, and the number of tree nodes for each partition.

Table 2-16 CDT Score Distribution Fields

Field Name	Description
Category Name	The category name of the partition for the specific CDT score.
Location Name	The location name of the partition for the specific CDT score.
Customer Segment Name	The customer segment name of the partition for the specific CDT score.
Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.
Number of Weeks	The number of weeks of data used to calculate the CDT score.
Tree Node Count	The number of tree nodes used to calculate the CDT score.

When you have finished with the calculation process, click **Next** to go to the Evaluation stage.

Evaluation

Once the CDT has been generated, you can evaluate it and make adjustments to parameters used in the CDT calculation during the Evaluation stage. Only the current version of the CDT, the one just generated in the Calculation stage, can be assessed in the Evaluation stage.

Process

Here is the high-level process for using the Evaluation stage to prune the results of the CDT calculation.

1. If you have determined that you need to prune the results of the calculation, enter appropriate values into the text boxes in the Pruning Setup area.
2. Click **Run** to start the pruning process.
3. Review the pruning results in the Pruning Results section and in the Pruning histograms.
4. After reviewing the results, if necessary, make changes to the values for the pruning and re-run the stage.
5. When you are satisfied with the results, click **Next** to go to the Escalation stage.

Pruning Setup

You can make changes by pruning the CDT based on the customer count, the SKU count, the tree-level count, and the minimum CDT score. To do this, enter a minimum value for each pruning filter.

Figure 2-16 CDT Pruning Setup

Once you have completed setting up the filters, click **Run** to begin the processing. After you review the pruning results, you can change the values for the filters if you find it necessary. Once you make changes, you must run the stage again in order to see the results of your changes.

Table 2-17 Pruning Setup: Filters

Filter Name	Filter Description
Minimum Customer Count	The minimum number of customers to be used in the CDT calculation. The default is 1000.
Minimum SKU Count	The minimum number of SKUs to be used in the CDT calculation. The default is 10.
Minimum Tree Level Count	The minimum number of levels to be used in the CDT calculation. The default is 2.
Minimum CDT Score	The minimum measure of the quality of a CDT to be used in the CDT calculation. The default is 25.

Pruning Results

The Pruning Results are located below Pruning Setup and display information that can help you assess the effects of the values you provided for the pruning filters.

Figure 2-17 CDT Pruning Results

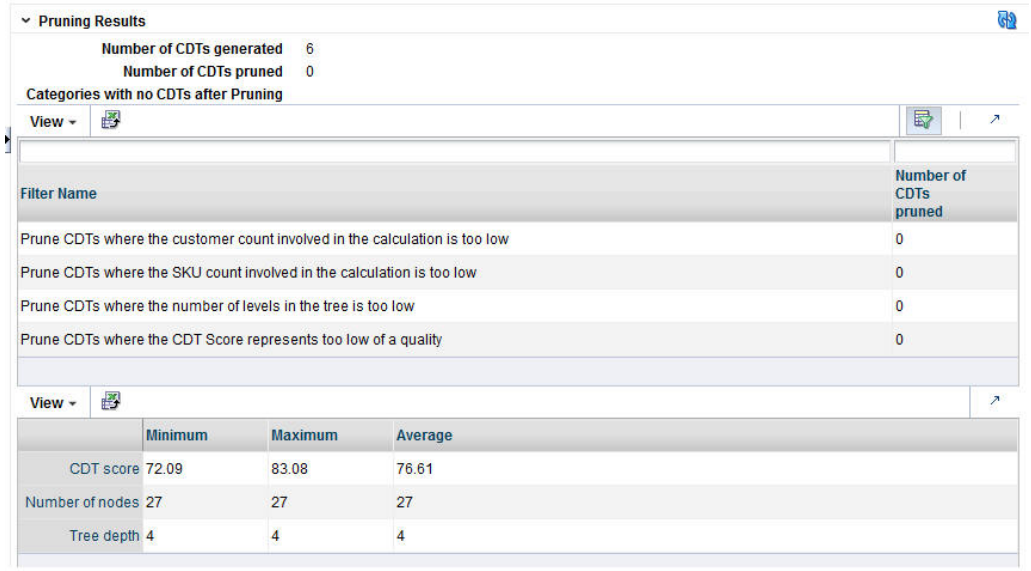


Table 2-18 Pruning Results

Field Name	Field Description
Number of CDTs Generated	The number of CDTs that were generated by the Calculation stage.

Table 2-18 (Cont.) Pruning Results

Field Name	Field Description
Number of CDTs Pruned	The number of CDTs that were pruned after the filters were applied.
Categories with no CDTs After Pruning	The names of the categories from which all CDTs were pruned.

Below this list is a table that identifies the number of CDTs that have been pruned by the following filters.

- Prune CDTs where the customer count involved in the calculation is too low.
- Prune CDTs where the SKU count involved in the calculation is too low.
- Prune CDTs where the number of levels in the tree is too low.
- Prune CDTs where the CDT score represents too low a quality.

A second table provides an overview of the pruning results, including the minimum, maximum, and average values for the CDT score, the number of nodes, and tree depth.

The number of nodes indicates the size of the tree. If a CDT contains few nodes, this can indicate a problem with the data or that too many nodes were excluded during the Calculation stage (because of a parameter setting).

Tree depth also indicates the size of the tree. This value can be used in conjunction with the Lowest Tree Level setting in the Calculation stage to analyze the results in terms of the number of levels in the tree.

Evaluation Report

The Evaluation Report displays the CDT results for the indicated version from the last successful run, if one has occurred. Here you can review information by Location or by Customer Segment. You can also set a version as active, view and edit a CDT, or compare two CDTs.

Table 2-19 Evaluation Report Fields

Field Name	Description
Partition	A partition is the combination of category, segment, and location. The column identifies the names of all nodes in the tree structure in the table. The node type may be either category, location, or customer segment.
Pruned	The number of CDTs removed from the list of usable CDTs. CDTs are removed that do not meet the filtering thresholds. During the Escalation stage, the escalation process makes adjustments for the CDTs that are removed.
Total Customer IDs	The number of customers used in the calculation of the CDT.
Total Weeks of Sales	The total number of weeks of sales used in the calculation of the CDT. This provides an indication of the amount of sales data used in the calculation. All the CDTs in a given version should have a similar value; if this is not the case, the results should be evaluated.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.

Table 2-19 (Cont.) Evaluation Report Fields

Field Name	Description
Creation Date	The date when the version whose data is displayed was created.
Creation User	The login name of the user who created the version.

The following icons are available for the Evaluation Report:

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version, then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, only the overlapping categories are replaced in the new version.

If a version is overwritten, then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This icon provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

This icon provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs.

Three histograms are provided in the Evaluation stage.

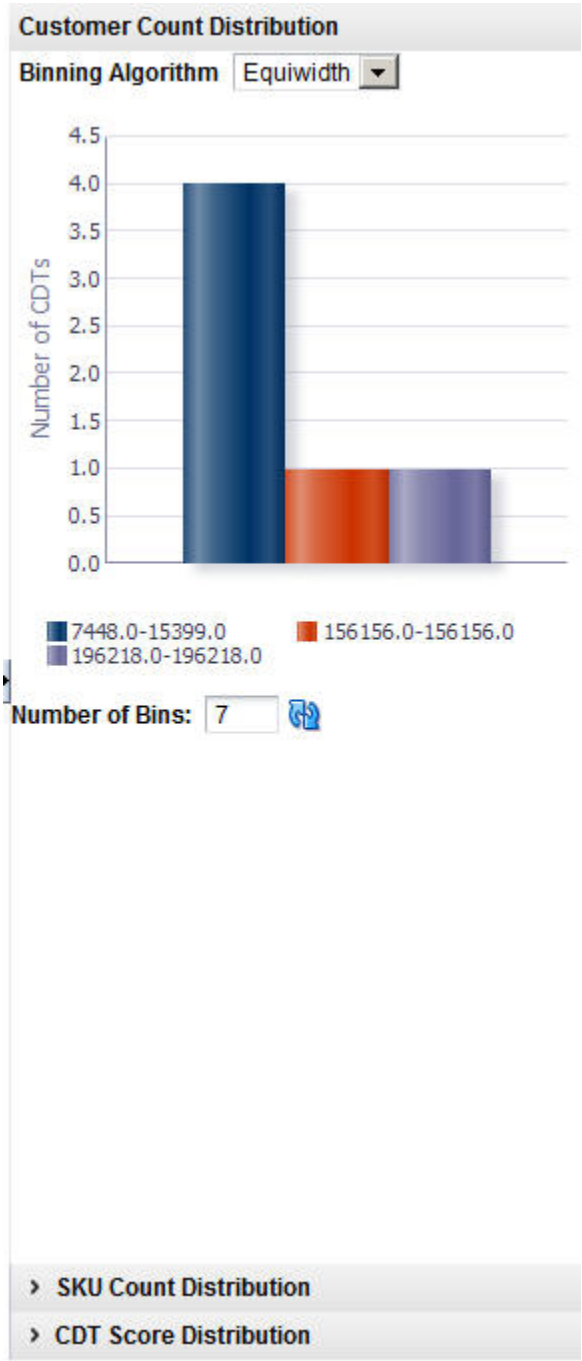
Evaluation Histograms

The following histograms are displayed in the Evaluation stage. You can use the information from the histograms to understand the data that was pruned by the filters.

Table 2-20 Evaluation Histograms

Name	Description
Customer Count Distribution	Displays the number of CDTs for a specific customer count.
SKU Count Distribution	Displays the number of CDTs for a specific SKU count.
CDT Score Distribution	Displays the CDT score distribution.

Figure 2-18 CDT Customer Count Distribution



When you have completed the evaluation of the results, click **Next** to go to the Escalation stage.

Escalation

The Escalation stage is used to fill in the holes for partitions whose CDTs were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy.

Process

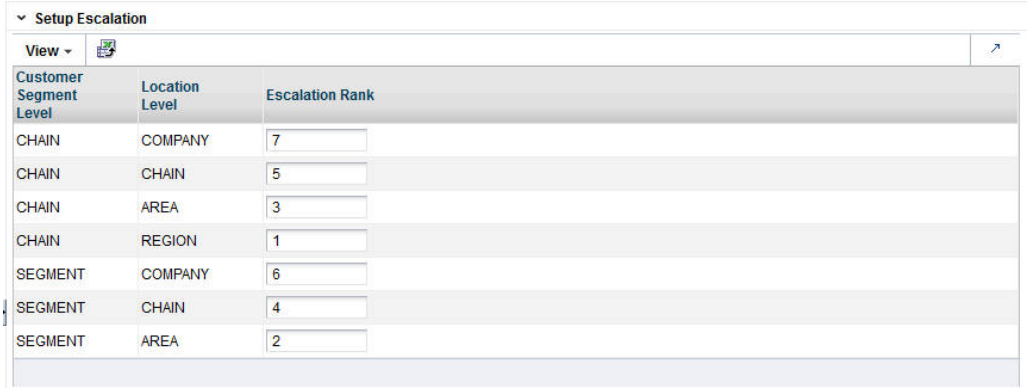
Here is the high-level process for setting up an escalation.

1. Enter a series of numbers to indicate the escalation rank, which determines the order in which the escalation occurs.
2. Click **Run** to start the escalation process.
3. Review the escalation results in the Escalation Report.
4. After reviewing the results, if necessary, make changes to the escalation ranks and rerun the stage.
5. When you are satisfied with the results, you can complete the version and make the version active so that it is available for other applications to use.

Setup Escalation

Escalation occurs along the segment hierarchy and the location hierarchy. Here is an example of an escalation path:

Figure 2-19 CDT Setup Escalation



Customer Segment Level	Location Level	Escalation Rank
CHAIN	COMPANY	7
CHAIN	CHAIN	5
CHAIN	AREA	3
CHAIN	REGION	1
SEGMENT	COMPANY	6
SEGMENT	CHAIN	4
SEGMENT	AREA	2

The following fields are required to set up the escalation.

Table 2-21 Setup Escalation

Field	Description
Customer Segment Level	Identifies the customer segment level in the escalation.
Location Level	Identifies the location level in the escalation.
Escalation Rank	Used to assign the ranks for the escalation, which determines the order in which the escalation occurs.

Here is an example of an escalation path.

Table 2-22 Example of Escalation Path

Segment Level	Location Level	Escalation Rank
Segment chain	Location chain	8
Segment chain	Region	7
Segment chain	Location area	6
Segment chain	Store cluster	5
Segments	Location chain	4
Segments	Region	3
Segments	Location area	2
Segments	Store cluster	1

You fill in the order of numbers. Every row must have an ordering number, and no ordering number can be reused.

The escalation path is specific to the user and the current version that the user is working on.

The default ordering is to go up the location hierarchy first, and then up the segment hierarchy, as shown in the example above. The reason is that the segment hierarchy has only two levels, and so its top level is very general.

Escalation Report

The Escalation Report breaks down the numbers to provide counts for the number of partitions filled with higher-level CDTs and the number of partitions that have not been changed by escalation. In addition, the fraction of CDTs for each partition is displayed.

Figure 2-20 CDT Escalation Report

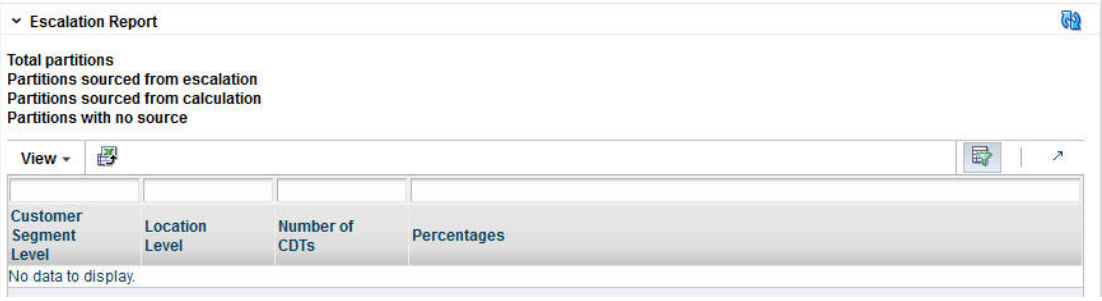


Table 2-23 Escalation Report

Field Name	Description
Total Partitions	The number of partitions in the version.
Partitions Sourced from Escalation	The number of partitions removed during escalation.
Partitions sourced from Calculation	The number of partitions removed during calculation.

Table 2-23 (Cont.) Escalation Report

Field Name	Description
Partitions with No Source	A partition that does not have CDTs assigned to it because all CDTs related to the partition have been pruned.
Customer Segment Level	Identifies the customer segment level.
Location Level	Identifies the location level.
Number of CDTs	The number of partitions that are trying to have a CDT assigned. This is generally the number of customer segments by the number of locations.
Percentages	The percentage of partitions that have been assigned a CDT from a given escalation level.

Completion of Process

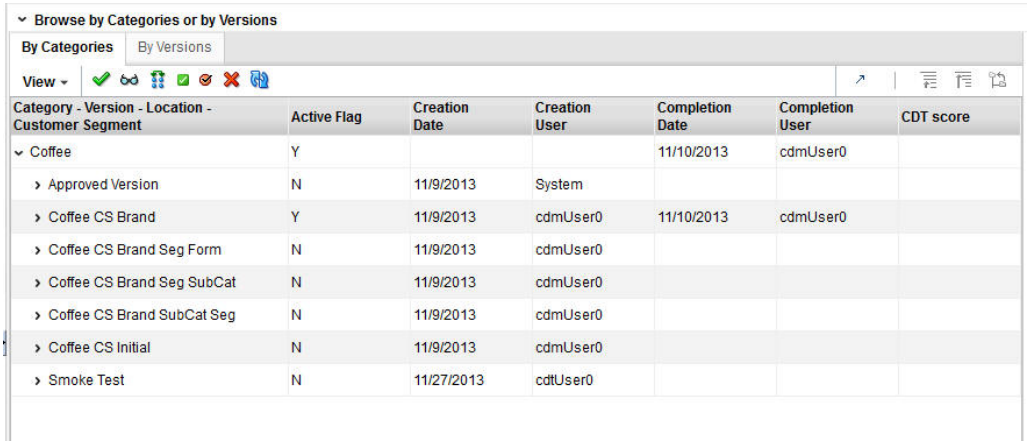
When a version is complete, the results for the version are activated so that other applications can use the information. The similarity data that has been calculated during the generation process is also activated for use.

After the completion of this step, the intermediate results from each stage are removed from the database and can no longer be used.

Manage CDTs Tab

The Manage CDTs tab is used to control which version of a CDT is active or approved. You can also access the functionality to edit a CDT or compare two CDTs.

Figure 2-21 Manage CDTs



Category - Version - Location - Customer Segment	Active Flag	Creation Date	Creation User	Completion Date	Completion User	CDT score
Coffee	Y			11/10/2013	cdmUser0	
Approved Version	N	11/9/2013	System			
Coffee CS Brand	Y	11/9/2013	cdmUser0	11/10/2013	cdmUser0	
Coffee CS Brand Seg Form	N	11/9/2013	cdmUser0			
Coffee CS Brand Seg SubCat	N	11/9/2013	cdmUser0			
Coffee CS Brand SubCat Seg	N	11/9/2013	cdmUser0			
Coffee CS Initial	N	11/9/2013	cdmUser0			
Smoke Test	N	11/27/2013	cdtUser0			

You can find the CDT you are interested in by:

- Browsing by categories or versions
- Searching by version name or user name.

Set Version as Complete

Each partition must have an active CDT. The active CDT is used by other applications that require a CDT.

All CDTs within a version are activated when you click **Set Version as Complete**.

If a CDT exists in more than one version, then the most recently activated version takes precedence.

If a version contains multiple categories and is Complete and then a different version with a partial overlap of categories is later marked as Complete, only the overlapping categories are replaced in the new version.

If a version is overwritten, then the CDTs in that version are no longer active.

When a CDT is at risk of going from active to inactive, you will see a warning message.

To determine whether or not a CDT is active, go to the Manage CDTs tab. The Browse by Categories or by Versions table displays a flag that indicates whether or not a CDT is active.

View One CDT

This button provides access to the CDT Editor. For details about this functionality, see Using the CDT Editor.

Compare Two CDTs

This button provides access to CDT Compare. For details about this functionality, see Comparing Two CDTs.

Approve a Version

Only one approved CDT is permitted for a given CMPO node/segment/location combination. If you approve a second CDT with the same combination, it overwrites the first one.

If an Approved CDT exists, then the Approved CDT is returned to the requesting application.

If no Approved CDT exists, then the Active CDT is returned to the requesting application.

If no Approved CDT exists and no Active CDT exists, an error is returned to the requesting application.

Approval can be done for individual CDTs or for an entire version; activation is for an entire version.

An approved version is used by the CMPO application. An active version is used by Demand Transference.

Unapprove

Use this to change the status of a CDT that was previously approved. Select a single CDT record in the data table and click the **Unapprove** button to remove it from the approved version.

Delete a Version

If you delete a version, it cannot be activated. You cannot delete an approved version, but you can remove a CDT from an approved version.

Browse by Category or Version

You can browse by category or version. The following information is displayed:

Table 2-24 Browsing by Category or Version

Field Name	Field Description
Version...Customer Segment	Partition
Active Flag	Y = Version is active. N = version is not active.
Creation Date	Date on which version was created.
Creation User	User who created the version.
Activation Date	Date on which the version was made active.
Activation User	User who made the version active.
CDT Score	A confidence score assigned to help in assessing the quality of the CDT. A CDT score ranges from 0 to 100. A higher value indicates a better score.

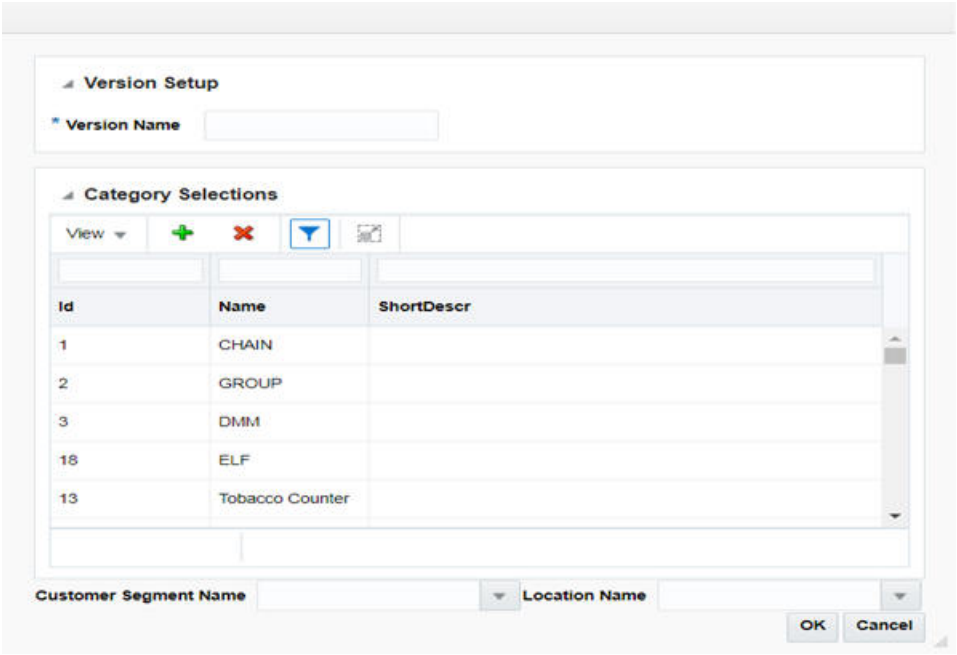
Search

You can search by version name or creation user name. In addition, you can design a search with specific criteria and save that search for future use.

Create CDT

You can create a new CDT in two ways: by modifying an existing CDT or by defining a new CDT.

Figure 2-22 Create CDT



1. Click the **Create CDT** icon, shown in Figure 2-22.
2. If you want to create a new CDT based on an existing one, select the version you want to use as a template from the list of Category Selections. If you want to create a new CDT, skip this step.
3. Enter the Version Name for the new CDT.
4. Select the Product Category, Customer Segment Name, and Location Name.
5. Click **OK**. You see the CDT Editor.
6. You can use the CDT editor, described in "Using the CDT Editor", to edit the new CDT and add a child, make changes to a branch, or change the assigned names.
7. When you have finished creating and editing the new CDT, click **Save CDT**.
8. To approve the new CDT, click **Approve CDT**.

Deleting a CDT

To delete a CDT, click the **Delete CDT** icon.

Using the CDT Editor

You can view or edit a CDT using the CDT Editor, which you access by clicking the **Edit** button. The CDT Editor displays the Customer Segment Name and Location Name for the CDT. It provides tools to navigate through the display using Zoom, Pan, Center, and Layout functionality.

The following functionality is available:

- Add a Child
- Edit
- Delete
- Delete Branch
- Copy Branch
- Paste Branch
- Save and Approve

Add a Child

To add a child node to the CDT, do the following:

1. Click **Add Child**. You see the Add CDT Node window.
2. Select an attribute from the **Attribute** drop-down menu.
3. Select a value from the **Attribute Value** drop-down menu.
4. Select the Create separate branch per value check box, if applicable. When more than one of the attribute values are selected, this field create a separate node for every attribute value selected.

> **Note:**
>
> A node can be created by selecting one or more attribute values. If multiple attribute values are chosen, and the check box to create a separate branch is not selected, one node that represents multiple attribute values is created.
>
> If one branch is created for all values of an attribute, select all attribute values and only one node will appear with the attribute value of ALL.
>
> If all the attribute values for an attribute are used, either by creating one node or separate nodes, a child cannot be added to the node and the **Add Child** is disabled. Creating a node with ALL does not associate the specific attribute values with the node. Instead, it encompasses any attribute value.

5. Click **OK**.

Editing Nodes

Editing is allowed on all nodes as long as there are additional attribute values available. The selections from the node being edited are preselected in the dialog. The attribute cannot be changed, so the attribute drop-down menu is not disabled. Any attribute values that are in use by other siblings are not be displayed in the list.

Deleting Nodes

Nodes can be deleted from the tree. There are two options for deleting: Delete and Delete Branch.

Selecting Delete deletes the selected node. Delete is available for any leaf node (a node without any children). It is also available for a non-leaf node that represents all the attributes for an attribute value: the children under this node move under the parent.

Selecting Delete Branch deletes the entire branch of nodes under the selected node, but not the selected node itself.

Copying/Pasting in Customer Decision Tree Editor

Copy and Paste functionality is provided to copy nodes from one branch to another. **Copy Branch** is enabled when a node is selected that has children. The Copy Branch function copies the full branch of children of the selected node, but not the selected node itself. **Paste Branch** is enabled when a branch has been copied and the selected node is a leaf node.

Viewing a Customer Decision Tree

The Customer Decision Trees have the potential to become large and occupy more space than the screen real estate allows for. The hierarchy viewer component used to display the Customer Decision Tree provides several features to assist in viewing the Customer Decision Tree effectively.

Expanding and Collapsing Branches

One way to limit the amount of space taken up by the Customer Decision Tree is to collapse branches of the tree. For any node that has children, a small triangle appears at the bottom of the box for that node. Hovering over the triangle enlarges it and displays an option to collapse that node if it is expanded and expand the node if it is collapsed.

Moving the Tree

If the entire Customer Decision Tree is not visible on one screen, the Customer Decision Tree can be moved to make other parts of the tree visible. The view can be moved by either clicking and dragging or by using the panning controls in the control bar for the hierarchy viewer.

Zooming

The hierarchy viewer provides some controls for zooming in and out to allow more or less of the tree to be in view at a time. Zooming out shrinks the size of the nodes, which may make them difficult to read.

Comparing Two CDTs

You can also compare two CDTs using the same CDT Editor functionality. You select the two CDTs you want to compare from the list. Both CDTs are displayed side by side.

3
Demand Transference

This chapter described the use of the Demand Transference Science Cloud Service module.

Introduction

Demand Transference (DT) helps you to compare products based on their similarities in order to determine what, if any, products customers might buy if the product they want to buy is for some reason unavailable. In this way, planning and ordering can be optimized. DT calculates similarities by comparing the attributes of the two products. If you are using CDT in conjunction with DT, you also have available the similarities calculated by CDT, which are based on customer-supplied transaction data.

The DT Cloud Service module consists of three tabs: Overview, Generate Models, and Manage Models. You use the Overview tab to keep track of the status of each stage during the main work you do with the application within the Generate Models tab. You use the Manage Models tab to evaluate the demand elasticity results and override the Maximum Substitutable Demand Percentage value, if that is needed.

Overview of DT Process

When you use the DT Cloud Service module, you follow this general iterative process to create and manage DT models:

- Data Setup. You define the categories to be used in the DT calculation.
- Data Filtering. You configure filters that remove input data that might cause errors in the calculation or that can lead to inaccurate or unreliable results.
- Similarity Calculation. You calculate similarities and assess the results of the calculation.
- Elasticity Calculation. You calculate the assortment elasticities and assess the results in terms of substitutable demand, which is the percentage of demand of a SKU that is retained when the SKU is deleted from the stores where it is selling.
- Escalation. When you are satisfied with your results, you can set the escalation path you can set the escalation path to fill in the holes for partitions whose DT models were removed during pruning by setting up a search path through the segment hierarchy and the location hierarchy. Then you can set a version of the DT model as complete.
- Manage Models. Use this tab to set time intervals for evaluating your results and to override the value for the maximum substitutable demand percentage.

Overview Tab

The Overview tab displays information that you can view and use to monitor the progress of the DT stages as well as to view some aggregate statistics and the DT results from the last successful run.

This tab contains the following sections:

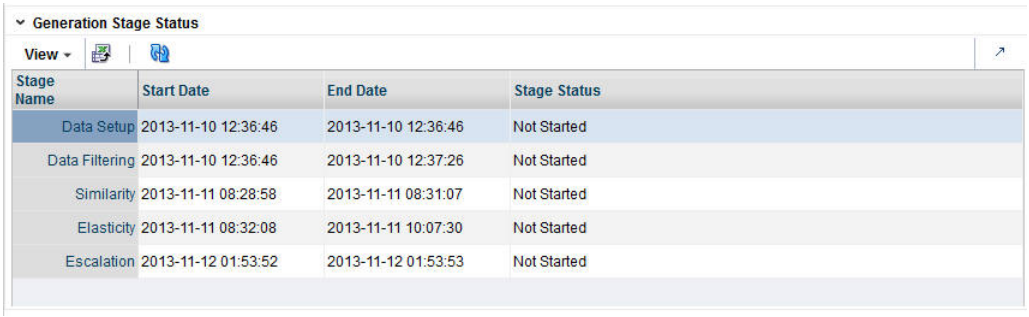
- Generation Stage Status
- Aggregate Statistics
- Calculation Report

Generation Stage Status

The Generation Stage Status table displays the current status of each of the DT stages.

Click the **Refresh** icon to update the fields and see the latest status for each stage.

Figure 3-1 DT Generation Stage Status



Stage Name	Start Date	End Date	Stage Status
Data Setup	2013-11-10 12:36:46	2013-11-10 12:36:46	Not Started
Data Filtering	2013-11-10 12:36:46	2013-11-10 12:37:26	Not Started
Similarity	2013-11-11 08:28:58	2013-11-11 08:31:07	Not Started
Elasticity	2013-11-11 08:32:08	2013-11-11 10:07:30	Not Started
Escalation	2013-11-12 01:53:52	2013-11-12 01:53:53	Not Started

You can use the status information to monitor the progress of each stage. It contains the following fields, which can be arranged and viewed, but not modified.

Table 3-1 Generation Stage Status Fields

Field Name	Description
Stage Name	A row exists in the table for each DT stage that provides detailed status information. The five stages are Data Setup, Data Filtering, Similarity, Elasticity, and Escalation.
Start Date	The date and time when a run for the stage most recently started.
End Date	The date and time when a run for the stage most recently ended.
Stage Status	The current status of the stage: Not Started, Not Started (Scheduled for Later), Processing, Completed Successfully, Stopped/Paused, Cancelled, or Failed.

The following table provides an explanation of the different values for the status of a stage.

Table 3-2 Stage Status Values

Stage Status	Description
Not Started	This indicates that the stage has not yet been initiated. This can occur when you first begin to create a DT model or when you update an existing version.

Table 3-2 (Cont.) Stage Status Values

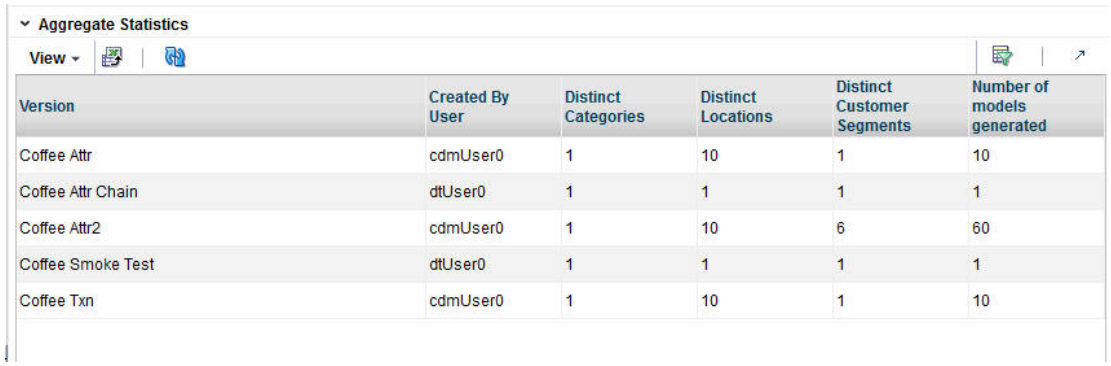
Stage Status	Description
Not Started (Scheduled for Later)	This indicates that the stage is scheduled to start after the previous stage is complete.
Processing	This indicates that the stage is currently being processed.
Stopped/Paused	This indicates either that the stage has been stopped by the application or that the user has chosen to stop the process (for example, to make a change to an option). In the later case, once any background processing has stopped, the user can re-run the stopped stage.
Completed Successfully	This indicates that the stage has been successfully processed.
Completed with Errors	This indicates that the stage was able to complete the processing of all requested CDTs, but one or more CDTs encountered errors during the run and were not able to complete successfully. This most commonly occurs when the data used for a CDT is not available or is too sparse to produce a result.
Cancelled	This indicates that the database has cancelled the execution of the stage either because of missing data or an exception. The user should review the RSE_LOG_MSG log to determine the problem.
Failed	This indicates that a problem occurred during the processing.

Aggregate Statistics

The Aggregate Statistics table displays statistical details about the existing DT model versions.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 3-2 DT Aggregate Statistics



Version	Created By User	Distinct Categories	Distinct Locations	Distinct Customer Segments	Number of models generated
Coffee Attr	cdmUser0	1	10	1	10
Coffee Attr Chain	dtUser0	1	1	1	1
Coffee Attr2	cdmUser0	1	10	6	60
Coffee Smoke Test	dtUser0	1	1	1	1
Coffee Txn	cdmUser0	1	10	1	10

Each existing version has a row in the table. The table contains the following fields, which can be arranged and viewed but not modified.

Table 3-3 Aggregate Statistics Fields

Field Name	Description
Version	User-created name that uniquely identifies this DT model version
Created By User	The user name of the person who created this version.
Distinct Categories	The number of categories associated with this version.
Distinct Locations	The number of locations associated with this version.
Distinct Customer Segments	The number of customer segments associated with this version.
Number of Models Generated	The number of DT models that have been calculated for this version.

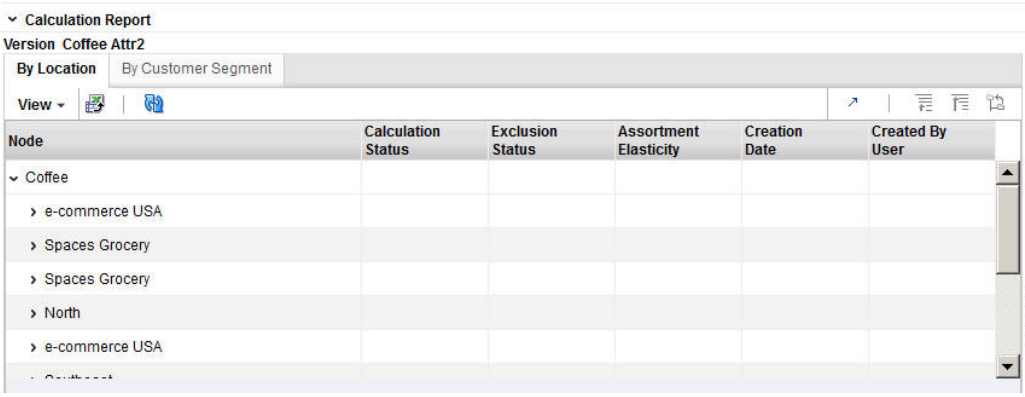
Calculation Report

The Calculation Report displays the DT model results from the last successful run, if one has occurred.

Here you can review information by Location or by Customer Segment.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 3-3 DT Calculation Report



The Calculation Report contains the following fields:

Table 3-4 DT Calculation Report Fields

Field Name	Description
Node	Identifies the node name.
Calculation Status	**Yes** indicates that the calculation is complete. **No** indicates that the calculation is not complete.
Exclusion Status	Yes indicates that data has been pruned. No indicates that data has not been pruned.

Table 3-4 (Cont.) DT Calculation Report Fields

Field Name	Description
Assortment Elasticity	A number calculated by the application that is a parameter in the DT model. A larger magnitude indicates larger overall transference.
Creation Date	The date when the version whose data is displayed was created.
Created By User	The login name of the person who created the version.

Generate Models Tab

The Generate Models tab is used to configure, run, evaluate, modify, and deploy a DT model. The process is divided into five stages that must be run in order. You can return to a stage you have already completed and make changes, but if you do, you must re-run that stage and all the stages that follow that stage, as the calculations are invalidated by the modifications you just made to the settings in that stage.

The five stages are:

Table 3-5 Generate Models Tab: Stages

Stage Name	Description
Data Setup	Select the nodes for the DT model calculations.
Data Filtering	Filter out input data that may result in inaccurate or unreliable answers.
Similarity Calculation	Calculate the similarities in customer demand.
Elasticity Calculation	Calculate the assortment elasticities for customer demand.
Escalation	Set the escalation path for the DT model. Use the Escalation Report to evaluate the results of the escalation.

Data Setup

The Data Setup stage is used to add and delete the categories to be used in the DT model generation process.

Process

Here is a high-level process for setting up the data for DT.

1. Select the category or categories you want to calculate DT models for.
2. Click **Next** to go to the Data Filtering stage.

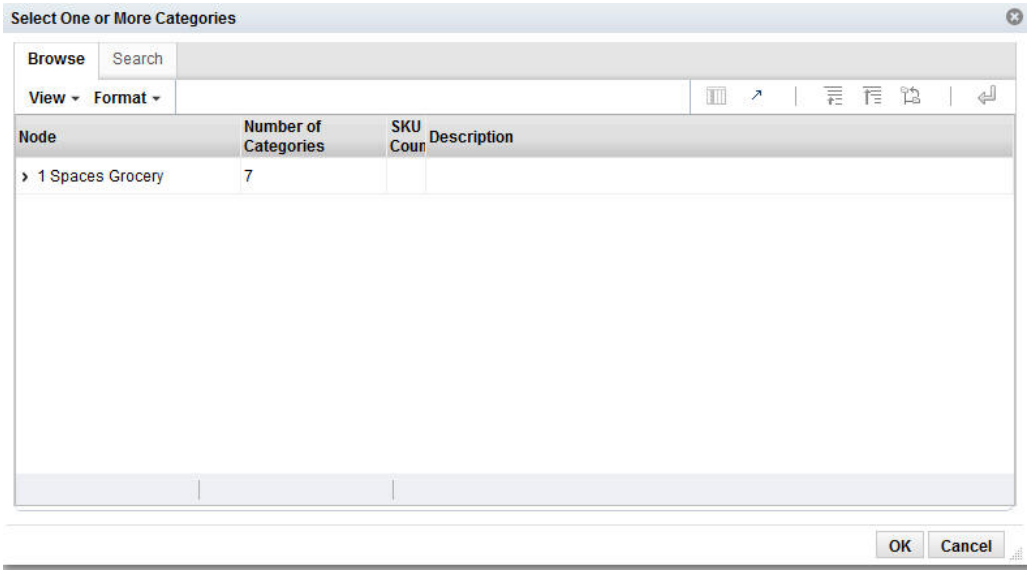
Category Selections

Use this table to add categories you want to include in the DT model calculation or delete categories that you want to remove from the DT model calculation.

Figure 3-4 DT Category Selections

To display a list of available nodes and the categories included in those nodes, click the **Add** icon. You see the Select One or More Categories dialog box, which contains two tabs, Browse and Search. You can use either tab to find the categories you are looking for.

Figure 3-5 DT Browse Categories



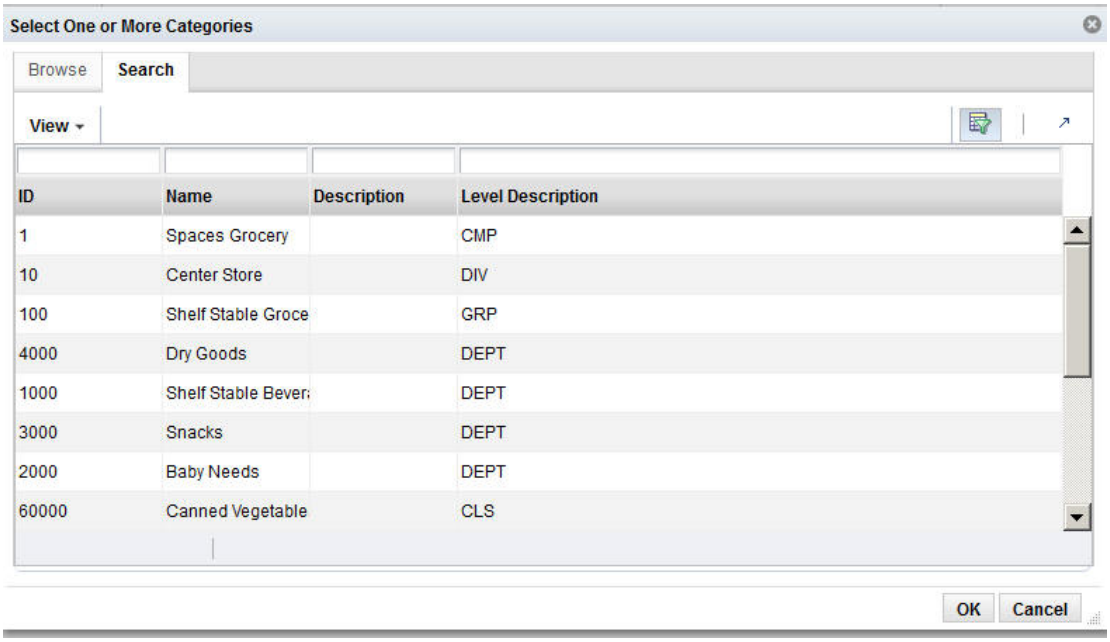
The Browse tab displays a table with the following fields:

Table 3-6 Category Selections: Browse

Field Name	Description
Node	The node tree structure can be expanded in order to view its categories.
Number of Categories	The number of categories within the node that has been selected. The number can help you understand the amount of processing required for the calculation.
SKU Count	The number of SKUs in a category. A category with too few SKUs may not produce good DT models.
Description	A description that provides additional information about the category.

Select the category or categories within the node that you want to be part of the DT model and click **OK**.

Figure 3-6 DT Search Categories



The Search tab displays a table with the following fields:

Table 3-7 Category Selections: Search

Field Name	Field Description
ID	An external code used to identify the category in other systems such as CMPO.
Name	The category name.
Short Description	A description that provides additional information about the category.
Level Description	A description of the level of the merchandise hierarchy that the node belongs to.

Select the category or categories within the node that you want to be part of the DT model and click **OK**.

Your selections are displayed in the Category Selections table.

After selecting the categories, click **Next** to go to the Data Filtering stage.

Data Filtering

The Data Filtering stage applies to all the categories that you select in the Data Setup stage. You should set the filters based on the histograms for each filter. The histograms help identify what data is actual outlier data, as compared to the rest of the data. In most cases, the default settings should be sufficient. However, if a histogram shows a flatter distribution, then you should consider modifying the default settings.

Process

Here is the high-level process for setting up and running data filtering.

1. Enter the appropriate values into the Filter Setup text entry boxes.
2. Click **Run** in order to filter the data.
3. Review the filtering results in the Data Filtering Summary table and the Data Filtering histograms.
4. After reviewing the results, if necessary, make changes to the values for the filters in Filter Setup and re-run the stage.
5. When you are satisfied with the results, click **Next** to go to the Similarity Calculation stage.

Filter Setup

You configure the following filters in order to filter out data you consider unacceptable from the calculation of the DT model.

Figure 3-7 DT Filter Setup

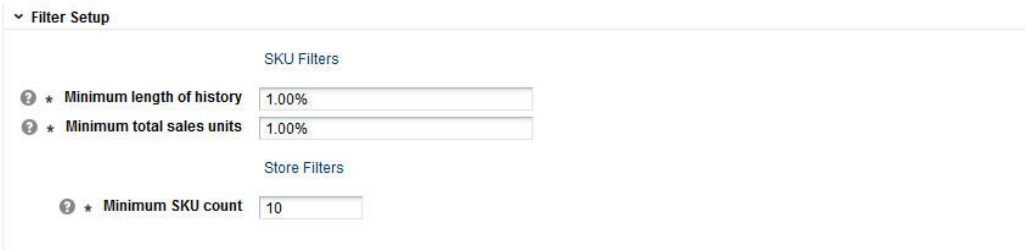


Table 3-8 Data Filters

Filter Name	Description
Minimum Length of History	This filter prunes SKU-segment-store combinations that have a short transaction history. The threshold is defined as a percentage of the median value for the category. The default value is 1%.
Minimum Total Sales Units	This filter prunes SKU-segment-store combinations that have a small number of total sales units during a given sales history for a specified customer segment and store. The threshold is defined as a percentage of the median value for the category. The default value is 1%.
Minimum SKU Count	This filter is applied after the above two filters and looks at the remaining data to determine if a store does not have enough SKUs. The threshold is defined as a set number of SKUs per store. The default value is 10 SKUs.

Data Filter Summary

This section describes the Data Filter Summary, shown in Figure 3-8.

Figure 3-8 DT Data Filter Summary

Data Filtering Summary

View

Filter Name	Pre-filter Sales Unit	Post-filter Sales Unit	Filtered Sales Unit	Filtered Sales Unit Percentage	Pre-filter SKU Count	Post-filter SKU Count	Filtered SKU Count
Length of History	1745156	1745156	0	0%	56	56	0
Total Sales Units	1745156	1745156	0	0%	56	56	0
SKU Count	1745156	1709337	35819	2.05%	56	56	0

The following information is provided after the filtering is complete and quantifies the amount of data filtered out for baseline history, total sales units, sales amounts, and SKU counts. Use this information to assess the effects of filtering.

Click the **Refresh** icon to update the fields and see the latest information for this table.

Table 3-9 Data Filter Summary Fields

Field Name	Field Description
Filter Name	The relevant filter of the three listed above.
Pre-filter Sales Unit	Amount prior to the application of the filter.
Post-filter Sales Unit	Amount remaining after the application of the filter.
Filtered Sales Unit	Amount filtered.
Filtered Sales Unit Percentage	Amount filtered, expressed as a percentage.
Pre-filter SKU Count	Amount prior to application of the filter.
Post-filter SKU Count	Amount remaining after application of the filter.
Filtered SKU Count	Amount filtered.
Filtered SKU Count Percentage	Amount filtered, expressed as a percentage.

Data Filtering Histograms

The following histograms illustrate the effects of filtering. You can use the information displayed in the histograms to adjust the configuration of the filters in order to eliminate outlier data. If you modify the filters, you must re-run the stage.

For information about adjusting the display of the histograms, see the Getting Started.

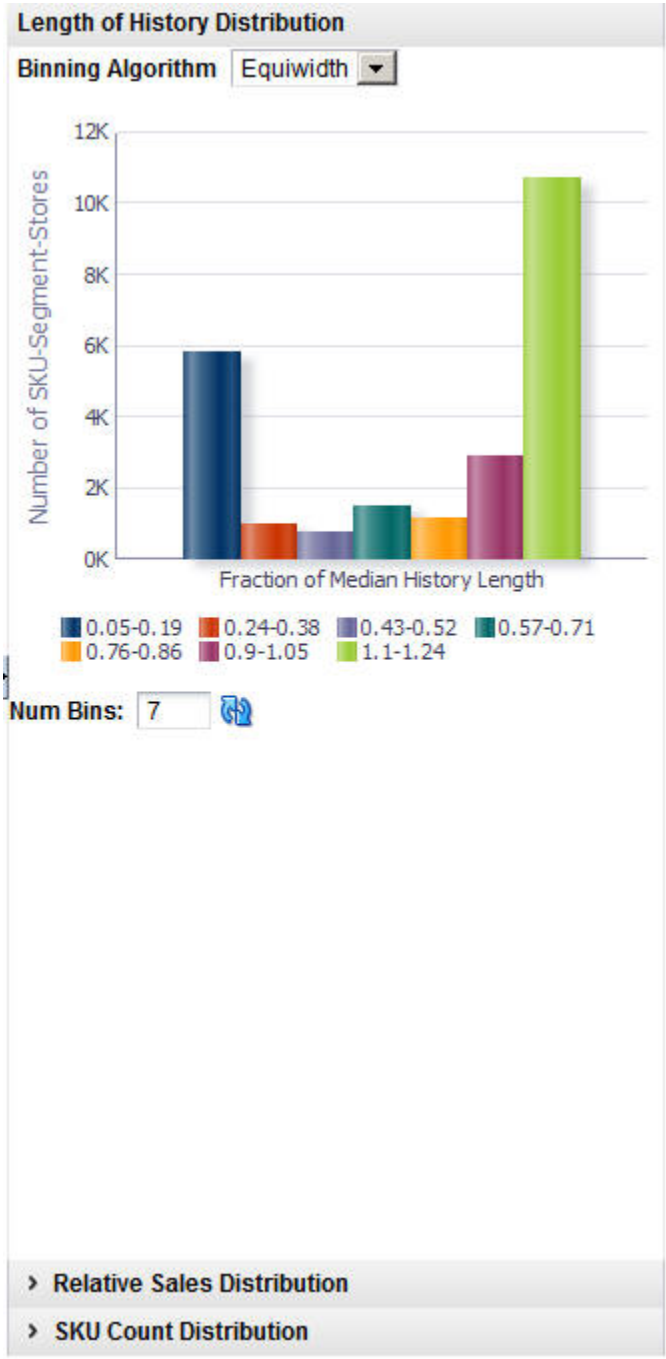
Table 3-10 Data Filtering Histograms

Histogram Name	Description
Length of History Distribution	Displays the percentage of median history length relative to the number of SKU-segment-stores.

Table 3-10 (Cont.) Data Filtering Histograms

Histogram Name	Description
Relative Sales Distribution	Displays the percentage of median category sales relative to the number of SKU-segment-stores.
SKU Count Distribution	Displays the percentage of the median category SKU count relative to the number of segment-stores.

Figure 3-9 DT Length of History Distribution Histogram



Similarity Calculation

Similarity in demand can be determined using either transaction-based data or attribute data. You can calculate the similarity using each type of data, if both types are available. You can only view the most recent run in the UI, so in order to compare runs, you must query the database to obtain the results from earlier runs.

Each category has its own set of similarities, relevant to the SKUs that are in that category. A similarity is calculated for each pair of historical SKUs in a category.

If transaction-based similarities are available, it is recommended that you use them instead of attribute-based similarities. Note that transaction-based similarities are only available through the Customer Decision Tree application.

Process

Here is the high-level process for calculating similarities:

1. Enter a unique name for the version of the DT model to be calculated.
2. Select the source of the data to be used: transaction-based or attributed-based. Transaction-based data is only available from the Customer Decision Tree generation.
3. Use the check boxes to indicate whether or not only top-level processing should occur for location or customer segment.
4. Customize the ranking of the category attributes if necessary.
5. Click **Run** to start the calculation.
6. Review the calculation results in the Similarity Display.
7. After reviewing the results, if necessary, make changes to the values for the calculation in the Version Setup and Category Attribute Setup and then re-run the stage.
8. When you are satisfied with the results, click **Next** to go to the Elasticity Calculation stage.

Version Setup

At the top of this stage you see two text boxes and two check boxes that you use to configure the parameters for the calculation.

Figure 3-10 DT Version Setup

Table 3-11 Version Setup: Fields

Field Name	Field Description
Version Name	Assign a name to each version of a DT model calculation. This allows you to create and save more than one version of a DT model. The version name you assign here is used in the Calculation Report, Aggregate Statistics table, and in the Manage DTs tab. Version names can be re-used; however, if the version name in question has active DT models, then you will see a warning that the active DT models will be removed from the version if you do re-use the version name.
Select the Source of Similarities	Use this option to define the type of data used in the calculation: transaction data or attribute data. Transaction-based data uses similarities calculated by Customer Decision Tree using transaction-based data. Attribute-based data calculates similarities within DT based on the attribute values associated with every SKU in the category.
Process Location Top Level Only	Check this option if you want DT models to be calculated for the Location Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.
Process Customer Segment Top Level Only	Check this option if you want DT models to be calculated for the Customer Segment Chain *only*. You can select this option in order to decrease the amount of time it takes the system to perform the calculation.

Once you select the source for the similarities, you will see either Category Attribute Setup, if you have selected to use attribute-based similarities, or Transaction-based Similarity Availability Per Category, if you have selected to use transaction-based similarities.

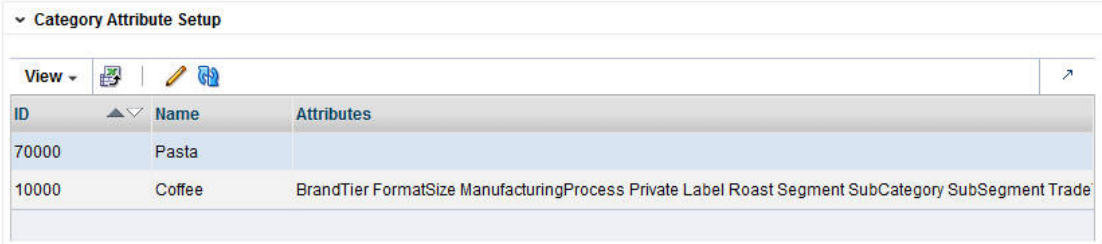
Category Attribute Setup

The application, using historical data, determines a specific weight for the category attributes. You can optionally change this weight and assign your own weight to the attributes for a category.

The weights indicate the importance of the attribute to the customers when they are making purchasing decisions. The attribute with the highest weight is the one the customer considers first when making a purchase. The system-generated weights are determined by the application from historical sales data. However, if a user disagrees with those weights, the user can override them. For example, in the case of coffee, the system may assign a weight of 0.7 to brand and 0.2 to size. This indicates that brand is historically more important to the customer than size when purchasing coffee. If the user disagrees with this analysis and thinks that brand and size are actually much closer together, the user can assign a weight of 0.5 to brand and 0.4 to size.

The Category Attribute Setup table displays the following:

Figure 3-11 DT Category Attribute Setup



Highlight the category you want to adjust the weights for and click the **Edit** icon. You see the Category Attributes Setup dialog box.

Figure 3-12 DT Category Attribute Setup

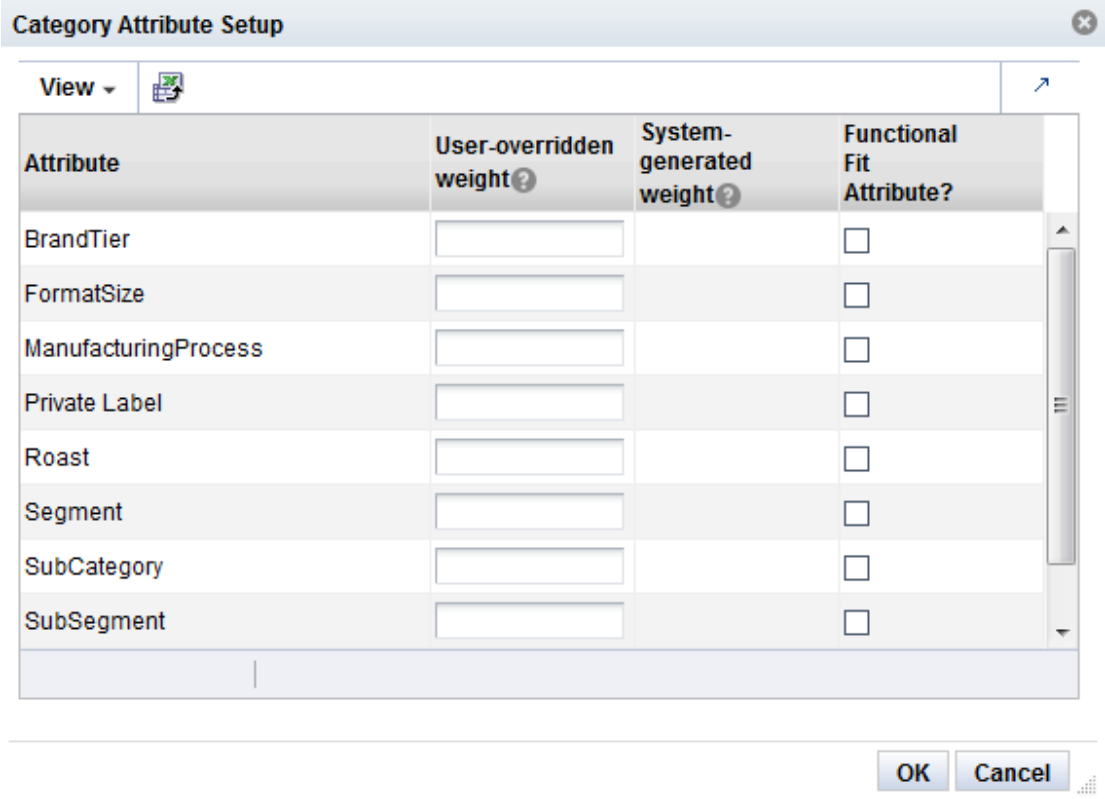


Table 3-12 Category Attributes Setup

Field	Description
ID	An external code used to identify the category in other systems such as CMPO.
Name	The category name.
Attribute	The specific attribute you are configuring.

The Category Attributes Setup pop-up lists the categories you are calculating DT models for. The system-assigned weights are also displayed. You can adjust the weight for all the attributes or a subset of the attributes.

The Category Attributes Setup dialog box contains the following fields. For each attribute you want to assign a custom weight to, enter a number between 0.000 and 1.000. For attributes that have no substitutes (such as windshield wipers of a specific length), the Functional Fit check box is checked by the system, so that similarities are not calculated for these attributes. When you are finished configuring the category attributes, click **OK**.

Table 3-13 Category Attribute Setup Fields

Field	Description
Attribute	The category attribute to assign a weight to.
User-Overridden Weight	The user-defined weight for the attribute.
System-Generated Weight	The system-generated weight for the attribute.
Functional Fit Attribute?	This is checked by the system if the attribute has no substitutes.

After you have finished configuring the similarity parameters, click **Run** to calculate the similarities. You see the results via the Similarity display table.

Transaction-Based Similarity Availability Per Category

The Transaction-Based Similarity table displays the following:

Table 3-14 Transaction-Based Similarity Availability Per Category

Field	Description
ID	An external code used to identify the category in other systems such as CMPO.
Name	The category name.
Description	A description that provides additional information about the category.
Available	A flag that indicates that a CDT version that contains data for this category has been made active.
Available As Of	Indicates the date that the CDT version was activated. This information can help you identify whether the CDT results are recent, or if they are potentially too old to use. For example, if the CDT data became available two years ago, you may consider that data to be out of date.

Similarity Display

The Similarity Display table shows the list of SKUs for which similarities have been calculated so that you can sort and analyze the results. You can search through the list of results by Category Name, Location, or Customer Segment.

Click the **Refresh** icon to update the fields and see the latest information in this table.

Figure 3-13 DT Similarity Display

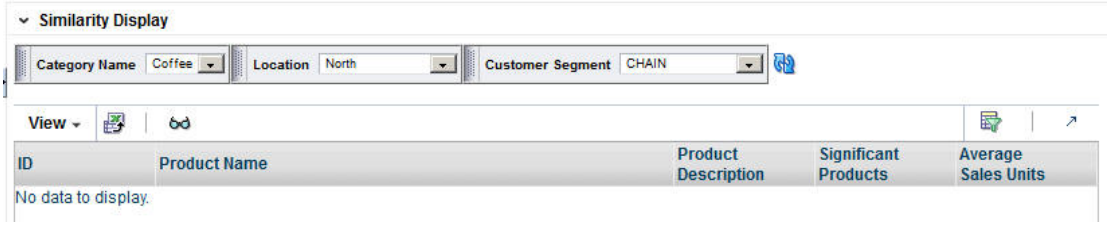


Table 3-15 Similarity Display

Field	Description
ID	An external code used to identify the category in other systems such as CMPO.
Product Name	The name identifying the product.
Product Description	A detailed description of the product.
Significant Products	A list of products that have High or Very High similarity. The threshold for how many products are considered significant can be configured in the database.
Average Sales Units	The average number of units used in the calculation.

Click the **See Similarities** icon to see detailed results for a specific set of SKUs.

The detailed results include the following fields:

Table 3-16 Similarity Display Results

Field Name	Field Description
ID	The product SKU.
Product Name	The name identifying the product.
Product Description	A detailed description of the product.
Similarity Strength	An indication of the similarity for the product: Very High, High, Medium, Low, Very Low.
Similarity Value	The calculated value for the similarity, from 0 to 1. A higher value indicates a higher degree of similarity.
Similarity Code	The numeric value associated with the similarity: 4 = Very High, 3 = High, 2 = Medium, 1 = Low, 0 = Very Low.

When you are satisfied with the Similarity results, click **Next** to go to the Elasticity Calculation stage.

Elasticity Calculation

During the Elasticity Calculation stage, the assortment elasticity is calculated. You do not configure any parameters. Click **Run** to initiate the calculation.

The assortment elasticity should not be a positive value because the transference model does not work properly if the value is positive. In addition, it should not be a null value because a null value indicates that the calculation of assortment elasticity failed and did not

produce an assortment elasticity value. If an assortment elasticity value is positive, it must be replaced with a negative value. The replacement occurs during the escalation process.

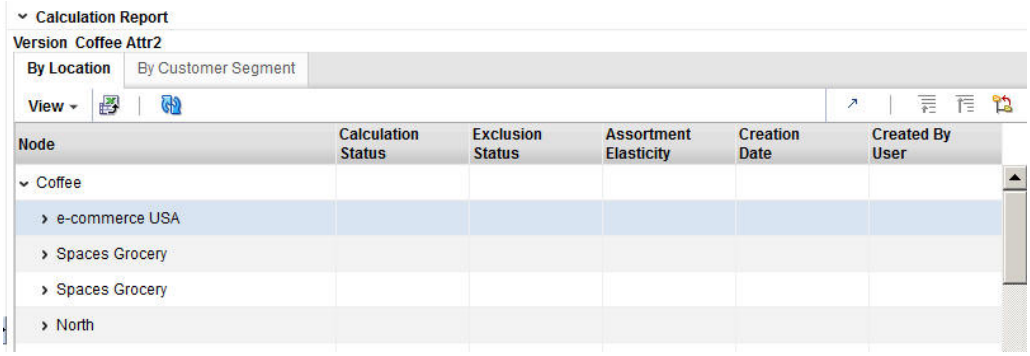
Process

The elasticity calculation is a background process. You use this stage to view the results. Note that the substitutable demand information is displayed and the percentages for the DT models are calculated after you set the time intervals within the Manage Models tab.

Calculation Report

The Calculation Report lists the status of the elasticity calculation and the exclusion, either by Location or by Customer Segment. An assortment elasticity is calculated for each category/location/segment combination selected during the Data Setup stage and the Calculation stage. The numerical result of the calculation, an output of DT generation, is used by Manage Models to calculate substitutable demand percentages.

Figure 3-14 DT Calculation Report



The Calculation Report has two tabs: By Location and By Customer Segment. Each tab has the following fields:

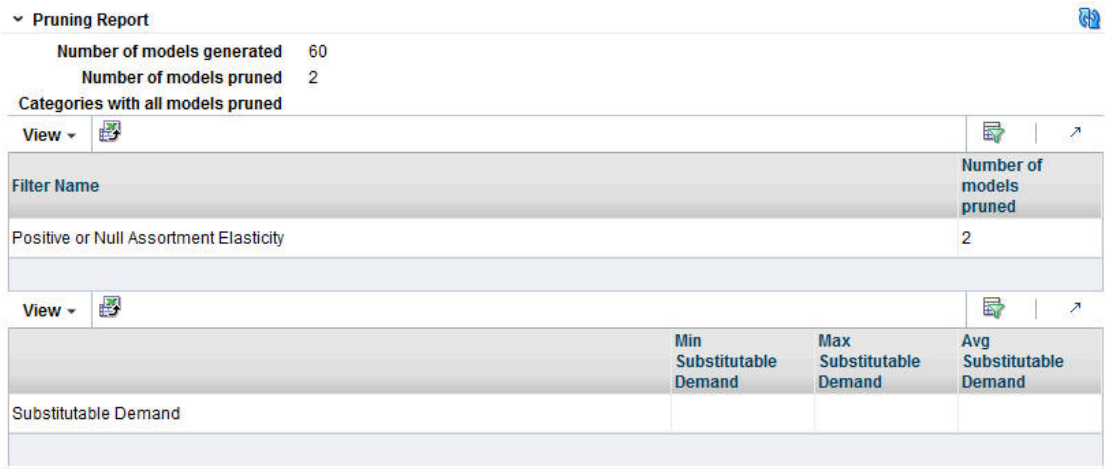
Table 3-17 Calculation Report Fields

Field	Description
Node	The node tree structure can be expanded in order to view its categories
Calculation Status	Indicates whether or not (Yes/No) the elasticity calculation has occurred.
Exclusion Status	Indicates whether or not (Yes/No) pruning has occurred.
Creation Date	The date when the version whose data is displayed was created.
Created By User	The login name of the person who created the version.
Assortment Elasticity	A number calculated by the application that is a parameter in the DT model. A larger magnitude indicates larger overall transference.

Pruning Report

The Pruning Report displays statistics about the results of data pruning.

Figure 3-15 DT Pruning Report



The Pruning Report contains the following fields:

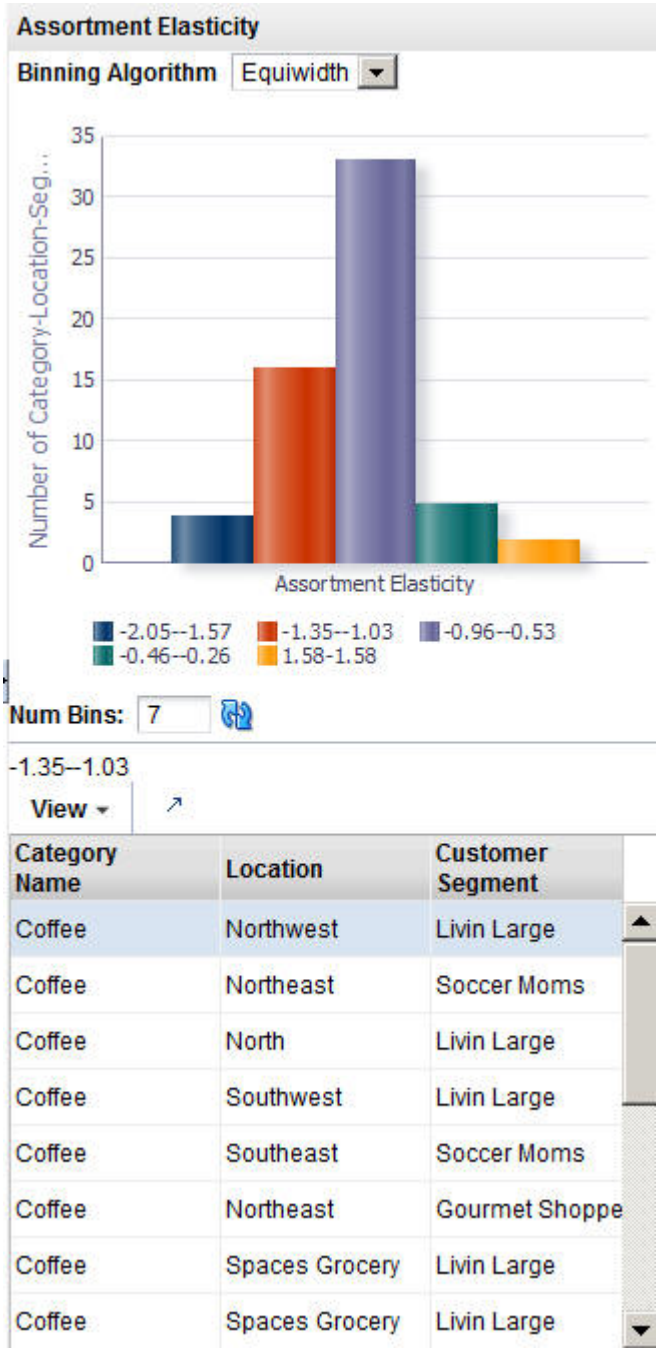
Table 3-18 Pruning Report Fields

Field	Description
Number of Models Generated	The number of DT models produced by the calculation.
Number of Models Pruned	The number of DT models pruned by the calculation.
Categories with All Models Pruned	The names of the categories for which all DT models have been pruned.
Positive or Null Assortment Elasticity	The Assortment Elasticity (AE) value is a negative number used in calculating substitutable demand percentages and demand transference effects. A positive AE value may be produced as a result of missing or unreliable input data for some partitions. Such meaningless transference effects must be pruned. In addition, if a null AE value is generated, it must also be pruned.
Substitutable Demand	Substitutable demand is a measure of how much demand is retained by the rest of the assortment when an item that is removed. When the item is removed, a portion of its demand is transferred to the remainder of items in the assortment. These values are populated when you set up the time intervals in the Manage Models tab and run the calculation there.

Assortment Elasticity Histogram

Figure 3-16 shows the DT Assortment Elasticity Histogram.

Figure 3-16 DT Assortment Elasticity Histogram



Escalation

The Escalation stage is used to fill in the holes for partitions whose DT models were removed during pruning by setting up a search path through the customer segment hierarchy and the location hierarchy. The DT models used to fill in the holes are not used by Manage Models calculations.

Process

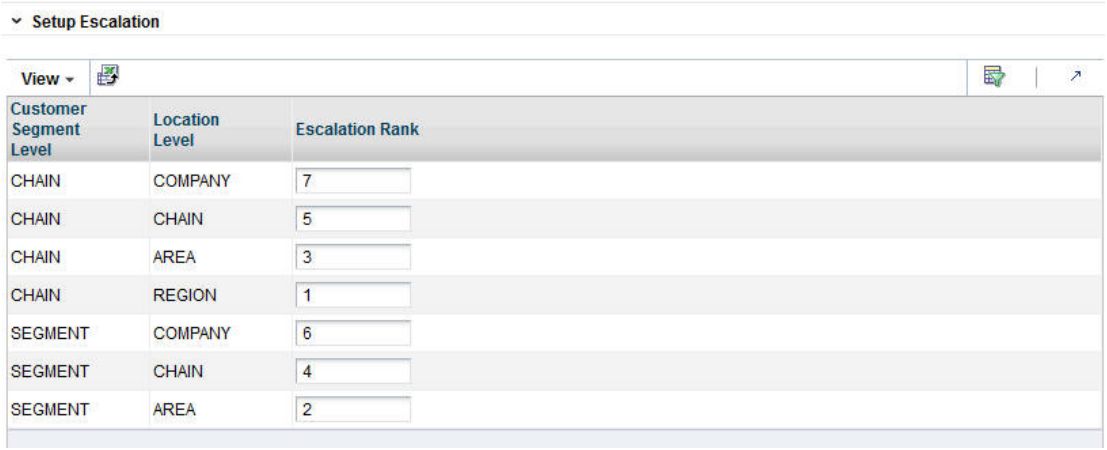
Here is the high-level process for setting up an escalation.

1. Enter a series of numbers to indicate the escalation rank, which determines the order in which the escalation occurs.
2. Click **Run** to start the escalation process.
3. Review the escalation results in the Escalation Report.
4. After reviewing the results, if necessary, make changes to the escalation ranks and re-run the stage.
5. When you are satisfied with the results, you can complete the version and make the version active so that it is available for other applications to use.

Setup Escalation

Escalation occurs along the customer segment hierarchy and the location hierarchy. Here is an example of an escalation path:

Figure 3-17 DT Setup Escalation



Customer Segment Level	Location Level	Escalation Rank
CHAIN	COMPANY	7
CHAIN	CHAIN	5
CHAIN	AREA	3
CHAIN	REGION	1
SEGMENT	COMPANY	6
SEGMENT	CHAIN	4
SEGMENT	AREA	2

The following fields are required to set up the escalation.

Table 3-19 Setup Escalation

Field	Description
Customer Segment Level	Identifies the customer segment level in the escalation.
Location Level	Identifies the location level in the escalation.
Escalation Rank	Used to assign the ranks for the escalation, which determines the order in which the escalation occurs.

Here is an example of an escalation path.

Table 3-20 Example of Escalation Path

Segment Level	Location Level	Escalation Rank
Segment chain	Location chain	8
Segment chain	Region	7
Segment chain	Location area	6
Segment chain	Store cluster	5
Segments	Location chain	4
Segments	Region	3
Segments	Location area	2
Segments	Store cluster	1

You fill in the order of numbers. Every row must have an ordering number, and no ordering number can be re-used.

The escalation path is specific to the user and the current version that the user is working on.

The default ordering is to go up the location hierarchy first, and then up the segment hierarchy, as shown in the example above. The reason is that the segment hierarchy has only two levels, and so its top level is very general.

Escalation Report

The Escalation Report breaks down the numbers to provide counts for the number of positions filled with higher-level DT models and the number of partitions that have not been changed by escalation. In addition, the fraction of DT models for each partition is displayed.

Figure 3-18 DT Escalation Report

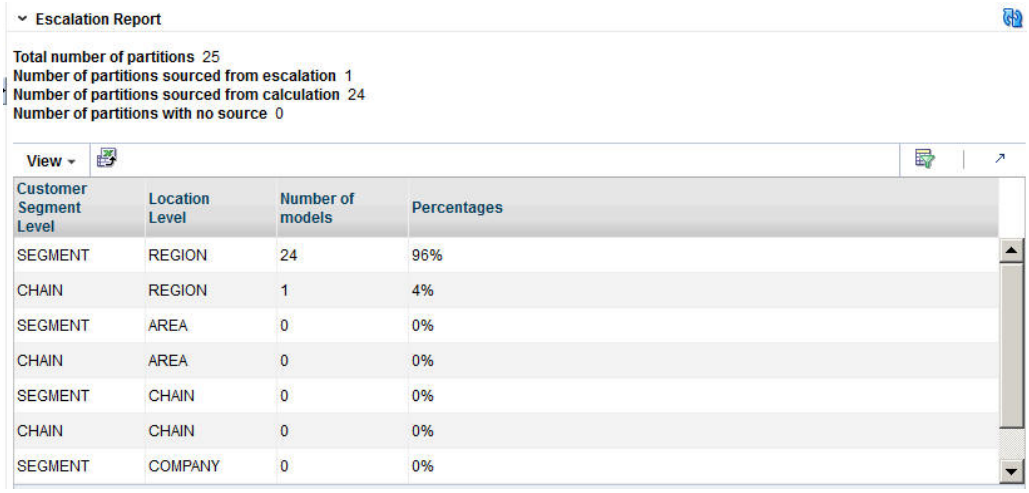


Table 3-21 Escalation Report

Field Name	Description
Total Number of Partitions	The number of partitions in the version.
Number of Partitions Sourced from Escalation	The number of partitions removed during escalation.
Number of Partitions Sourced from Calculation	The number of partitions removed during calculation.
Number of Partitions with No Source	A partition that does not have a model assigned to it because all models related to the partition have been pruned.
Customer Segment Level	Identifies the customer segment level.
Location Level	Identifies the location level.
Number of Models	The number of partitions that are trying to have a model assigned. This is generally the number of customer segments by the number of locations.
Percentages	The percentage of partitions that have been assigned a model from a given escalation level.

Completion of Process

When a version is complete, the results for the version are activated so that other applications can use the information. The similarity data that has been calculated during the generation process is also activated for use.

After the completion of this step, the intermediate results from each stage are removed from the database and can no longer be used.

Be aware that once a version is completed, it cannot be completed again unless a different version is completed first. Changes made to the version's data after completing it will not be copied to the relevant output tables.

Manage Models Tab

You can set up various time intervals to use in the evaluation of a version of a DT and configure the value for maximum substitutable demand and see how different maximum values affect the substitutable demand. This allows you to change the maximum value to one you find more suitable. You can see the percentage of demand of a SKU that is retained when the SKU is deleted from the stores where it is selling. This is the substitutable demand percentage. In this way you can evaluate the accuracy and usability of the elasticity calculation.

Substitutable demand is a measure of how much demand is retained by the rest of the assortment when an item is removed. When the item is removed, a portion of its demand is transferred to the remainder of items in the assortment. This portion is considered the retained demand. If the magnitude of the assortment elasticity is larger, then the amount retained will be higher. By examining the retained demand, you can evaluate the assortment elasticity value to see if its magnitude is too large. The key value to examine is the maximum substitutable demand percentage. For a given category, you may decide that this value is too large.

Process

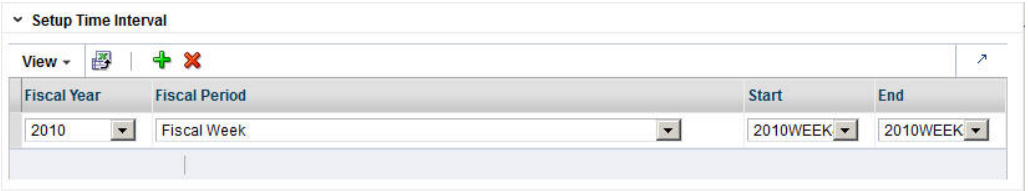
Here is the high-level process for determining suitable substitutable demand values and thus suitable DT models.

- Set up the time intervals you are interested in.
- Select the versions you want to evaluate.
- Click the **Calculate** icon to obtain an initial set of percentages.
- Enter various values for maximum substitutable demand and use the **Edit** icon to enter override values for Maximum Substitutable Demand Percent and the **Calculate** icon to determine the impact.
- You can use the **Revert** icon to restore the original percentage values.
- When you are satisfied with the results, you can choose which version to make active using the **Set Version as Complete** icon.

Time Interval Setup

The time interval defines the span of time for the sales history to be used to determine the amount of history that is retained when SKUs are dropped. A group of intervals can be defined. Gaps between intervals are permitted; however, intervals cannot overlap.

Figure 3-19 DT Setup Time Interval



You should select a time interval for which the historical assortments are reasonably representative of the assortments that will be used in the CMPO application. Because the time interval is used to calculate the substitutable demand information, selecting a representative interval provides substitutable demand information that is highly relevant to the actual application of demand transference in CMPO. Typically, the most representative time period is a recent time interval, since that is generally when assortments are most similar to the current assortments. If you use a time period that is not recent, you run the risk of using assortments that are not as similar to the current ones. You should also make sure not to select an interval that is too large, because a large interval necessarily includes several assortment changes within that interval. An interval size of approximately four weeks is recommended.

The fields that define a time interval are:

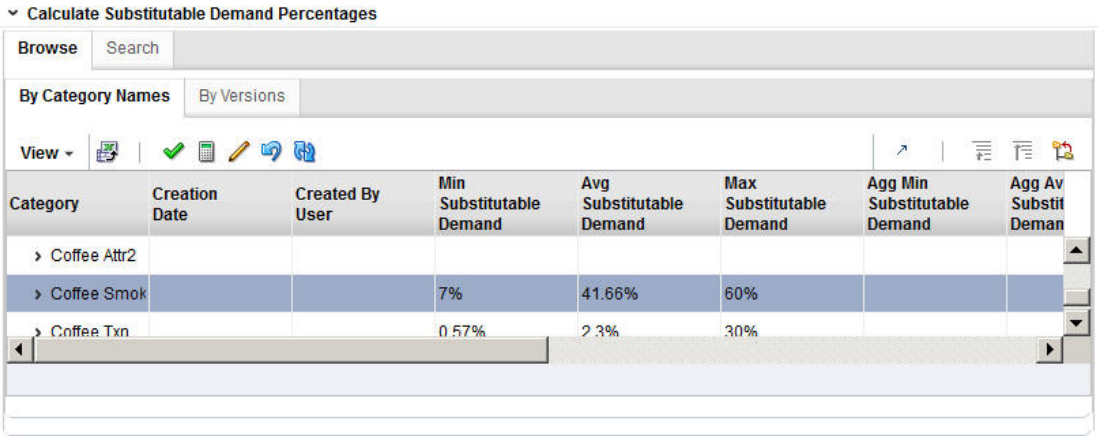
Table 3-22 Setup Time Interval: Fields

Field Name	Field Description
Fiscal Year	The fiscal year for the time interval.
Fiscal Period	The fiscal period within the fiscal year (Fiscal Quarter, Fiscal Period, or fiscal Week).
Start	The time unit when the time interval specified in Fiscal Period begins.
End	The time unit when the time interval specified in Fiscal Period ends.

Calculate Substitutable Demand Percentages

You can vary the value for the maximum substitutable demand percentage and see the impact on selected categories or versions.

Figure 3-20 DT Calculate Substitutable Demand Percentages



Browsing and Searching

You can browse by Category Names or by Versions. You can also search by name or by creation user. You see the following DT model data displayed:

Table 3-23 Browsing

Field Name	Description
Category/Version	The name of the category.
Creation Date	The date when the version was created.
Created By User	The user name of the person who created this version.
Min Substitutable Demand	The percentage value for the minimum substitutable demand.
Avg Substitutable Demand	The percentage value for the average substitutable demand.
Max Substitutable Demand	The percentage value for the maximum substitutable demand.
Agg Min Substitutable Demand	The aggregated value for the minimum substitutable demand.

Table 3-23 (Cont.) Browsing

Field Name	Description
Agg Avg Substitutable Demand	The aggregated value for the average substitutable demand.
Agg Max Substitutable Demand	The aggregated value for the maximum substitutable demand.
Calculation Status	Indication of whether or not (Yes/No) the calculation is complete.
Completion Status	Indication of whether or not (Yes/No) the status is active.
Completion Date	The date when the status became active.
Completion User	The user name of the person who activated the version.

Once you make a selection, click the **Edit** icon and enter an override value between 0% and 100% for the Maximum Substitutable Demand Percent. Click the **Calculate** icon to initiate the calculation. To revert the calculation, click the **Revert** icon.

Once you have determined the substitutable demand value you want, you can click the **Complete** icon to make the version active.

Escalation Report

The Escalation Report breaks down the numbers to provide counts for the number of positions filled with higher-level DT models and the number of partitions that have not been changed by escalation. In addition, the fraction of DT models for each partition is displayed.

Figure 3-21 DT Escalation Report

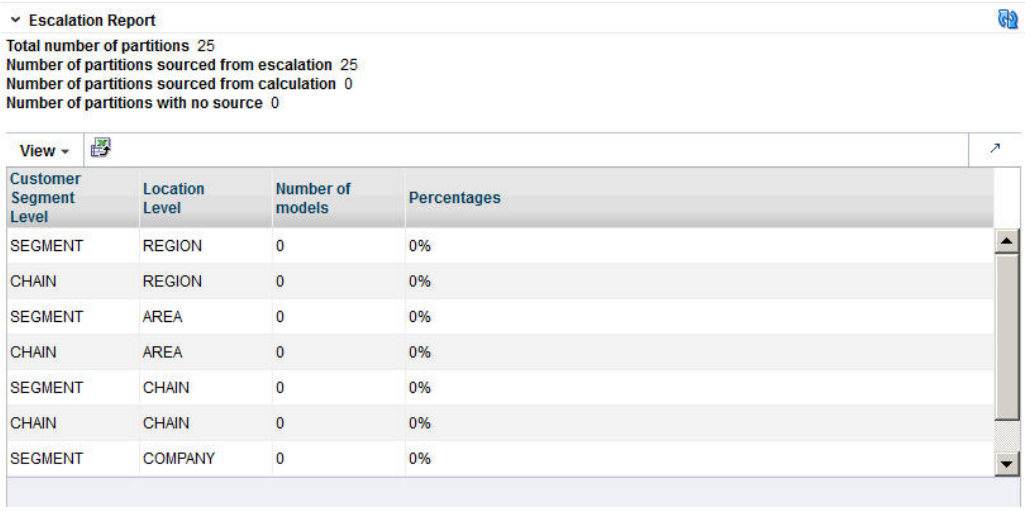


Table 3-24 Escalation Report

Field Name	Description
Total Number of Partitions	The number of partitions in the version.
Number of Partitions Sourced from Escalation	The number of partitions removed during escalation.
Number of Partitions Sourced from Calculation	The number of partitions removed during calculation.
Number of Partitions with No Source	A partition that does not have a model assigned to it because all models related to the partition have been pruned.
Customer Segment Level	Identifies the customer segment level.
Location Level	Identifies the location level.
Number of Models	The number of partitions that are trying to have a model assigned. This is generally the number of customer segments by the number of locations.
Percentages	The percentage of partitions that have been assigned a model from a given escalation level.

Histogram

The Substitutable Demand Distribution histogram displays the distribution of the substitutable demand values for SKU/stores.

4
Advanced Clustering

This chapter describes the Advanced Clustering Cloud Service module.

Introduction

Advanced Clustering is an enterprise-specific clustering solution that uses data mining to create store groupings at different product levels using a variety of inputs. These inputs include performance data (sales dollars, sales units, and gross profit), product attributes (brand, color, and size/fit), store attributes (climate, store format, size, and servicing distribution center), third-party data such as demographics (income, ethnicity, and population density), and customer segments.

The application's embedded science and automation helps you to identify unique patterns within your data that you can use to create the necessary customer-centric and targeted clusters. These can be used by the assortment planning, allocation and replenishment, pricing, and promotion processes.

It optimizes clusters in order to determine the minimum number of clusters that best describes the historical data used in the analysis and that best meets your business objectives, which you define during the design of your clusters.

You can use Advanced Clustering to execute localized or customer-centric assortments and for pricing. In addition, the application can help you when forecasting, for example, if you want to cluster stores based on similar seasonal patterns. You can also use the application for allocation, by clustering stores based on similar selling patterns.

Features

The key features of Advanced Clustering Cloud Service include:

- Scenario-based cluster generation, based on store or product attributes, customer segment profiles, or performance.
- Three-step cluster-generation process.
- What-if capabilities that can be used to create multiple clustering scenarios and then measure them against one another. This can help ensure that the most appropriate clusters are used by the applicable planning and execution processes.
- Automatic ranking of cluster scenarios to support what-if comparisons. Recommendations for the optimal cluster scenario and number of clusters are provided.
- Dynamic nesting of clusters, in which nested or mixed attribute clusters are created based on multiple attributes, performance data, and customer segments.
- Two types of algorithms are used.
  - Proprietary BaNG (Batch Neural Gas) algorithm for convergent cluster parameters
  - K-means approach for creating clusters in a hierarchical manner, which automatically determines the best attributes to split into an additional cluster.

- A variety of distance metrics that are suitable for real-value attributes, categorical attributes, profile-based measurements, and time-based performance.

Overview of Advanced Clustering Process

To use Advanced Clustering, follow this general process to create and manage clusters, working in the Generate Store Clusters tab and the Manage Store Clusters tab:

- Cluster Criteria. View all available clusters for specified merchandise, location, and calendar. Review cluster criteria or scenario details for each cluster. Use existing cluster as the basis for creating a new cluster.
- Explore Data. Examine the input data for the cluster. Review multiple data points and significant attributes using the contextual information.
- Cluster Setup. Define multiple what-if scenarios. Such scenarios can be compared with one another throughout the clustering process.
- Cluster Results. View the scenario results and compare scenarios.
- Cluster Insights. Gain an understanding about cluster results and cluster performance prior to approval by examining the contextual information.
- Manage Store Clusters. Manage existing cluster criteria. Perform manual overrides and approve clusters.

Cluster Criteria Overview Tab

The Cluster Criteria Overview tab displays a list of the most recently defined cluster criteria and provides the status, clustering criteria, applicable merchandise, location, and calendar nodes. You can click on the criteria name in order to access it within the Generate Store Clusters tab.

Figure 4-1 Cluster Criteria Overview Tab

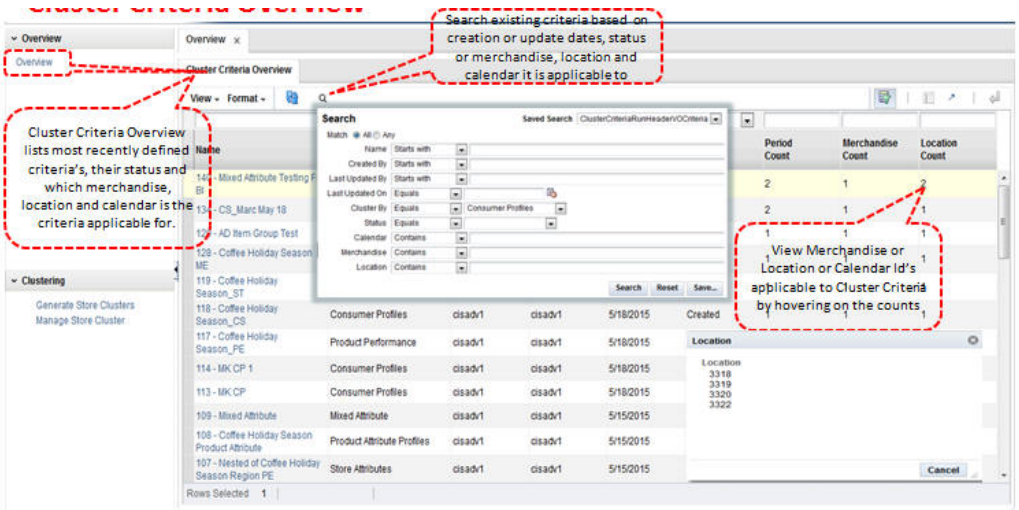


Table 4-1 Cluster Criteria Overview Tab

Field	Description
Name	The criteria ID and user-assigned name of the cluster.
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Created By	The name of the user who created the cluster.
Last Updated By	The name of the user who most recently updated the cluster.
Last Updated On	The date when the cluster was most recently updated.
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
Period Count	The number of calendar nodes defined for the criteria. Hover over the count in order to see a list of the calendar keys associated with the criteria.
Merchandise Count	The number of merchandise nodes defined for the criteria. Hover over the count in order to see a list of the merchandise keys associated with the criteria.
Location Count	The number of location nodes defined for the criteria. Hover over the count in order to see a list of the location keys associated with the criteria.

Clustering Criteria

The following clustering criteria (which are also called "Cluster by") are the defaults:

Consumer Profile

Cluster stores based on the similarities in the customer profile mix whose members shop in the stores or trading areas. These clusters form the basis for additional analysis that can provide an understanding of which customers shop in which stores and how they shop. Information from market research firms such as the Nielsen Corporation can help retailers develop customer profiles. Such information can be provided via a data interface.

Location Attributes

Cluster stores based on how shopping behavior varies by store attribute. In combination with the profile mix, this provides an understanding of demographic details such as income level, ethnicity, education, household size, and family characteristics. Such knowledge can help the retailer to make assortment and pricing decisions. By analyzing cluster composition and studying business intelligence, the retailer can make informed decisions based on shopper demographics.

Product Attributes

Store share is generated based on product attributes. The store clusters produced can be used in an assortment. In this type of cluster, stores with a similar share of sales for one or more attributes are grouped together. For example, for the product coffee, stores can be differentiated by the sales patterns for premium, standard, and niche brands. The percentage of each store contribution is calculated using Sales Retail $ for each product attribute value to

the total sales retail for the category or subcategory in a specified location. Product attributes can only be configured at the category or subcategory level.

Performance Criteria

Cluster stores based on the historical sales metrics by performance at various merchandise levels. Determine how shopping behavior varies by category. This information can be helpful in identifying low, medium, and high-volume stores that all have similar sales patterns.

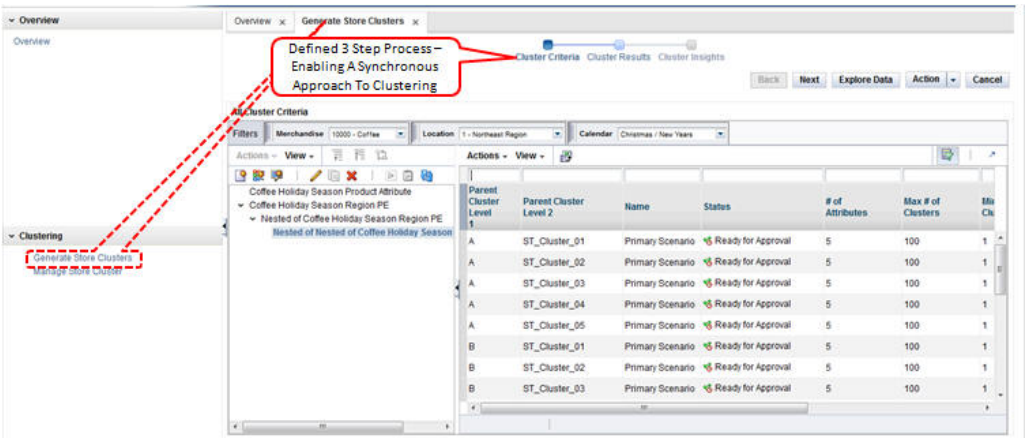
Mixed Criteria

Mixed criteria combine discrete and continuous attributes together. This allows a retailer to cluster stores using attributes from all the first four listed cluster criteria at the same time.

Generate Store Clusters Tab

The Generate Store Clusters tab is used to create clusters and then model the clusters with various scenarios in order to determine the best clusters. It consists of three stages: Cluster Criteria, Cluster Results, and Cluster Insights.

Figure 4-2 Generate Store Clusters Tab



Cluster Criteria Stage

In this stage, you can view summary data about existing clusters and define the characteristics of new clusters.

Process

Here is the high-level process for defining a cluster.

1. Provide a unique name for the cluster.
2. Define the type of data used to characterize the cluster.
3. Select merchandise and location nodes.
4. Define the time period for the cluster.

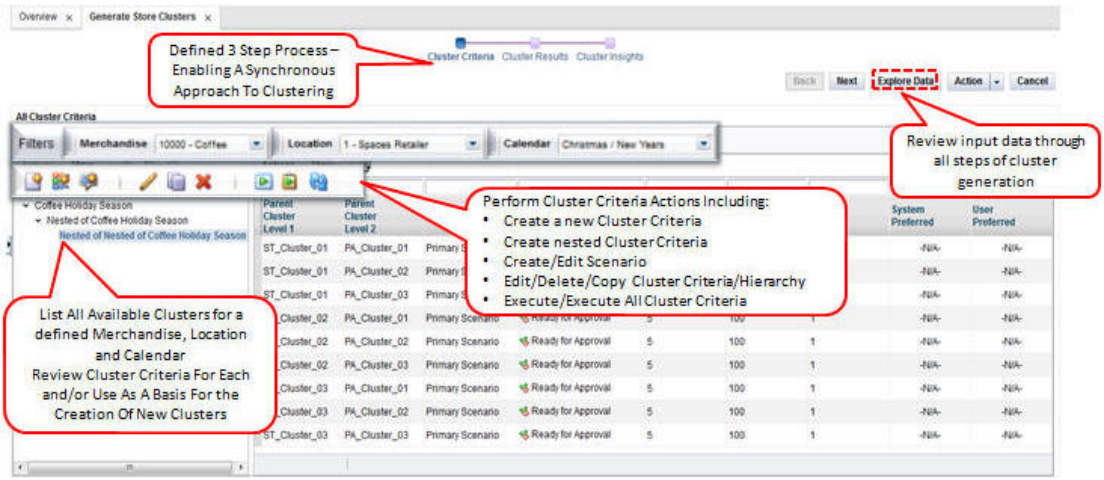
5. Define the historical time period for the data.

All Cluster Criteria

In this area of the page you can view information about existing clusters.

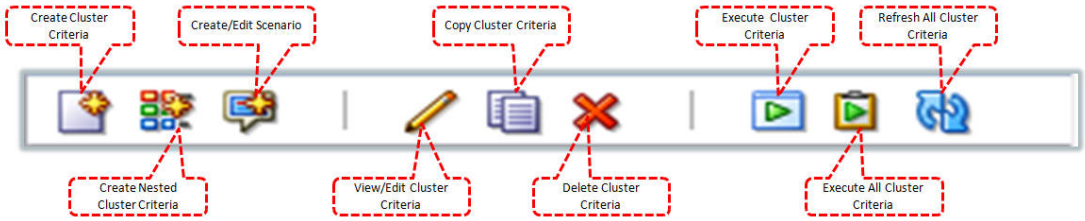
Use the View list to select existing cluster criteria. You can tailor your search for existing clusters by Merchandise, Location, and Calendar. Once you select a cluster, the defining details for that cluster are displayed in the Worksheet area.

Figure 4-3 All Cluster Criteria



In addition, you can use the toolbar buttons to:

Figure 4-4 All Cluster Criteria Toolbar



- Create a cluster criteria. You can define an initial cluster. The criteria include cluster name, the category for the cluster, the effective date for the cluster, and the history to use.
- Created a nested cluster criteria. In this way you can subdivide an existing cluster in order to analyze it further. Once you create a nested cluster, the name "Nested of <name of original cluster>" appears in the Cluster Criteria area. You can then define its characteristics in the same way you define any cluster.
- Create or edit a scenario. Modify the configured primary scenario or create another scenario to perform what-if analysis for a selected cluster criteria.

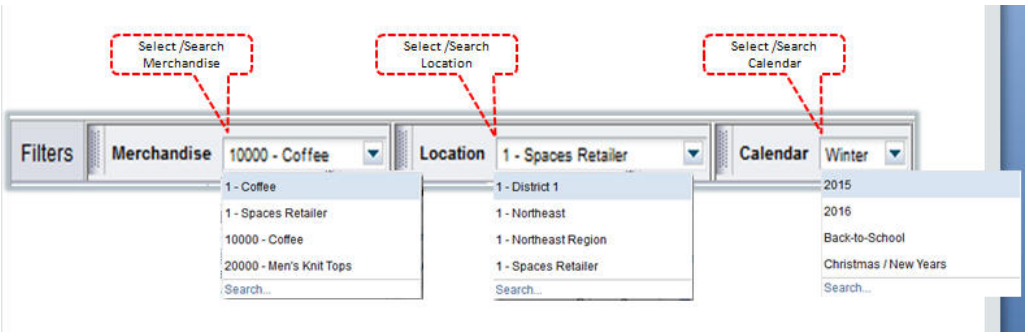
- View or edit a cluster criteria. The characteristics of the cluster are displayed in the Cluster Criteria pop-up.
- Copy a cluster criteria. Once you have copied it, you can modify it.
- Delete a cluster criteria. Delete the selected cluster criteria.
- Execute a cluster criteria. Execute all non-executed scenarios for the selected criteria.
- Execute all cluster criteria. Execute the entire cluster hierarchy criteria for the selected criteria at once.
- Refresh all. Refresh all cluster criteria in order to view any updates to the existing cluster criteria.

All Cluster Criteria Filter

The All Cluster Criteria filter provides the following:

- Merchandise allows the user to filter the existing cluster criteria by searching for or selecting the merchandise for the supported hierarchies.
- Location allows the user to filter the existing cluster criteria by searching for or selecting the location for the supported hierarchies.
- Calendar allows the user to filter the existing cluster criteria by searching for or selecting the calendar for the supported hierarchies.

Figure 4-5 All Cluster Criteria Filter



All Cluster Criteria Summary

Once you have highlighted a cluster criteria to examine, details about that cluster criteria are displayed in a pop-up. The details include information about the cluster criteria and the scenarios created for that cluster criteria.

Table 4-2 Pop-Up Details

Field Name	Description
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Shared Criteria	A check mark indicates that more than one merchandise or location node is used in the cluster criteria.
Merchandise Type	The merchandise type.
Scenario Created	The number of scenarios created for the cluster.
Scenario Executed	The number of scenarios executed for the cluster.
Location Type	The location type.
Parent Cluster Level	The name of the ancestor cluster that has been further clustered.

All Cluster Criteria Scenario List

This displays the scenarios for the selected cluster criteria in the All Cluster Criteria tree.

Table 4-3 Scenario List

Field Name	Description
Name	The name assigned to each scenario that has been created for the cluster.
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
User Preferred	Indicates whether or not the user prefers the cluster.
System Preferred	Indicates whether or not the system prefers the cluster.
# of Attributes	The number of attributes that were used in the cluster.
Max. # of Clusters	A user-provided value for the maximum clusters centers that the clustering process should consider.

Cluster Criteria

In this pop-up, you define the initial clustering parameters for the cluster criteria of a new cluster. Note that multiple hierarchies are supported in order to facilitate comparisons between clusters. For example, you can compare clusters for the market and retail location hierarchy.

Figure 4-6 illustrates how to use a simple approach to clustering by selecting attributes from a Cluster by. For example, you can select the performance Cluster by and generate clusters using store sales units or revenue.

Figure 4-6 Cluster Criteria

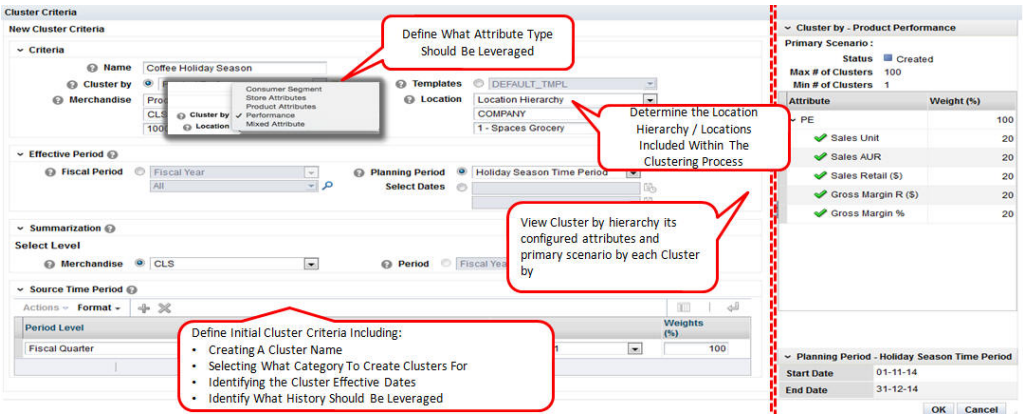
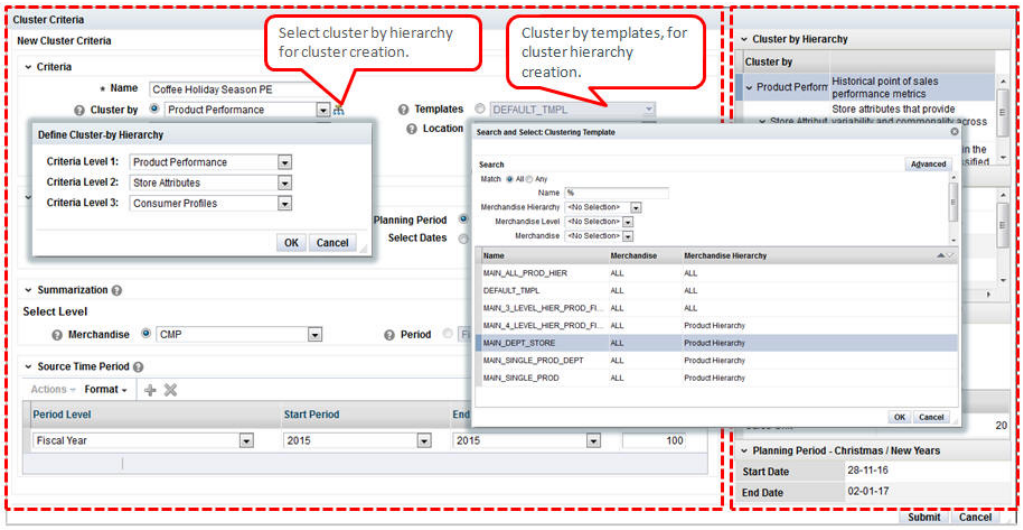


Figure 4-7 illustrates the use of a nested approach to clustering. Select a Cluster by hierarchy using the dynamic hierarchy pop-up or the pre-configured template hierarchies. For example, you can first create a cluster using the performance Cluster by and then further cluster using location attributes.

Figure 4-7 Cluster Criteria - Selection



The following information defines a cluster:

Table 4-4 New Cluster Definition

Field Name	Description
Name	A unique name to identify the cluster.

Table 4-4 (Cont.) New Cluster Definition

Field Name	Description
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Merchandise	Once you choose the merchandise level for the cluster, you must select the hierarchy type, the hierarchy level, and the hierarchy node. These are specific to the merchandise level you select.
Location	Once you choose the location level for the cluster, you must select the hierarchy type, the hierarchy level, and the hierarchy node. These are specific to the location level you select.
Template	Select by name a predefined template that can be used to create a cluster hierarchy.

Effective Period

You can define a time interval for the cluster by either choosing a period from the list provided or by selecting a start date and an end date.

To define the Effective Period, you select either Planning Period, Fiscal Period, or Select Date:

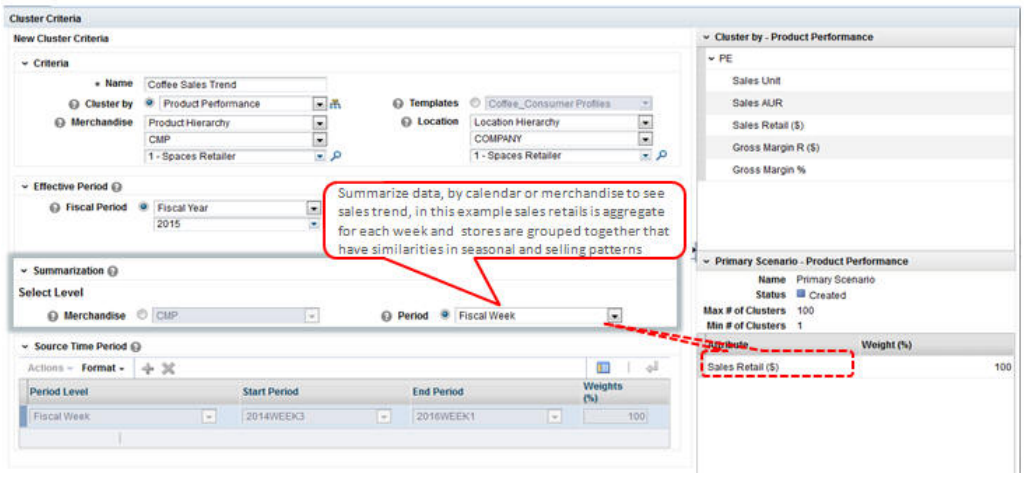
Table 4-5 Effective Period

Option	Description
Fiscal Period	If you select this option, choose the period and the subdivisions of that period from the drop-down lists.
Planning Period	Select from the range of values provided for the period. Planning periods are user-defined buying periods for a season or a season subset.
Select Dates	If you select this option, choose the start and end dates using the calendar pop-up.

Summarization

Data summarization is available when you select the product performance Cluster by. Select the dimensions of the hierarchy (merchandise or calendar) to summarize the data and consider the dimension position in the clustering process. For example, when you use the category/week sales data to generate store clusters, you can select the week dimension in the calendar hierarchy summarization. The clustering process considers all weeks (week1, week2 ,... week52) as attributes and clusters stores based on their weekly sales patterns.

Figure 4-8 Summarization



Source Time Period

Select historical sales data for clustering and view contextual data to analyze cluster performance. You can specify more than one time period and assign different weights to different periods in order to place more or less emphasis on different periods.

Source time periods are available for all cluster criteria. With product and performance or mixed criteria, when performance metrics are used for clustering, these define the historical data used for the calculation. This time period also defines the historical data used to display BI when sales metrics are shown.

Table 4-6 Source Time Period

Field	Description
Period Level	Select from Fiscal Year, Fiscal Quarter, Fiscal Period, or Fiscal Week.
Start Period	Once you select the Period Level, you select the starting subdivision within that period.
End Period	Once you select the Period Level, you select the ending subdivision within that period.
Weight (%)	Used to define the weight given to the historical data from the defined time period.

Contextual Area

When you are creating a new cluster criteria, you can see details about the following parameters that can help you understand the cluster you are creating.

Cluster By Hierarchy

The following information is displayed when you select a template or use the icon to select the Cluster by hierarchy.

Template

Table 4-7 Template Display

Property	Description
Template Name	Name of template configured during deployment.
Description	Description of template.
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.

Hierarchy

A dynamic Cluster by hierarchy is displayed. For example, the template PE-ST-ST has a Cluster by hierarchy of performance/store attribute/store attribute.

Cluster By Primary Scenario

You see this when you select Cluster by in the Criteria panel when you are setting the cluster parameters or when you select Cluster by in the contextual area for the hierarchy.

The attributes configured and the primary scenario properties defined for the selected cluster are displayed. The attributes listed are those that are significant for the clustering defined during deployment.

The primary scenario is the default scenario defined during deployment. The following information is displayed.

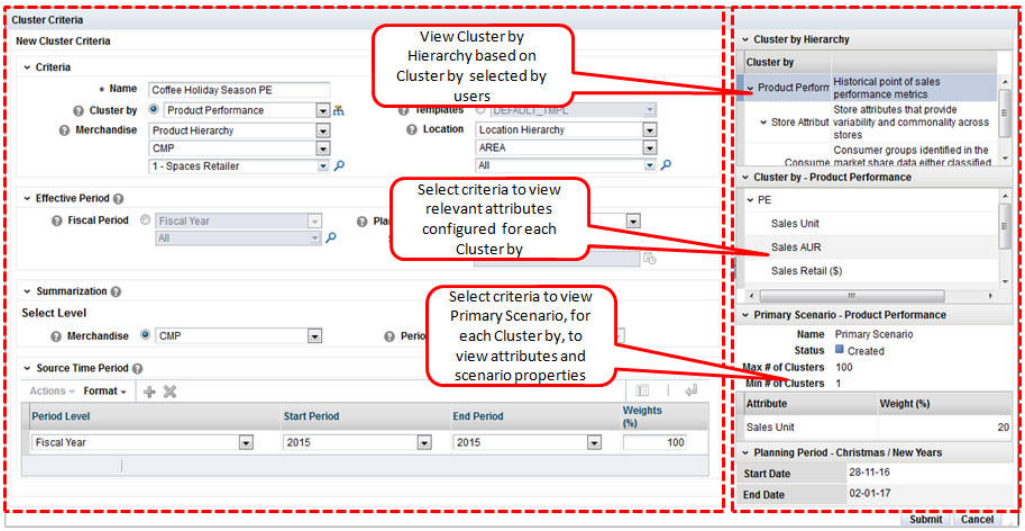
Table 4-8 Primary Scenario

Property	Description
Name	The name of the primary scenario.
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
Maximum # clusters	The maximum number of clusters. The default value is 100. This is used for analyzing the clusters.
Minimum # clusters	The minimum number of clusters. The default value is 1. This is used for analyzing the clusters.
Attribute	A list of the attributes configured during clustering.
Attribute weight	The weights associated with each attribute. This is used to calculate distance.

Planning Period

This list displays the time period you selected for the cluster definition. This information is available only for planning periods, where it provides the start and end dates of the planning period. This content changes whenever planning period is selected in Effective Period when you are setting cluster parameters.

Figure 4-9 Contextual Information



Explore Data

Use the Explore Data pop-up to examine data for the cluster you defined. You can view the store that provides input into the clustering process.

Process

In this pop-up you can only view the data, so the only actions you can perform are drilling down through the data in the table and altering the arrangement of the table.

Summary

This area lists the criteria you initially selected to define the cluster.

Figure 4-10 Cluster Criteria Summary

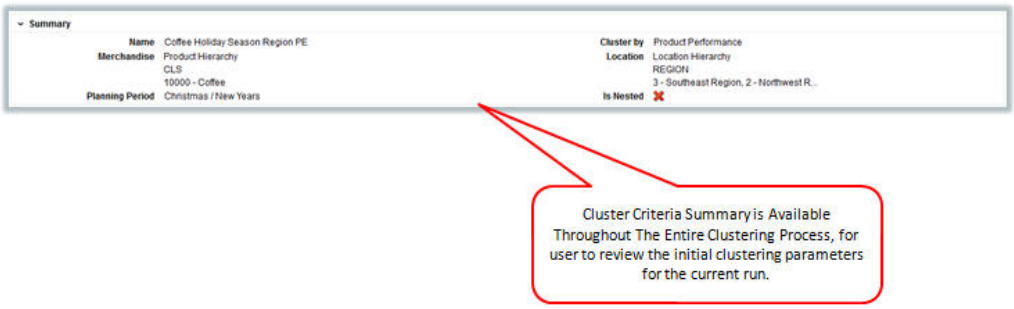


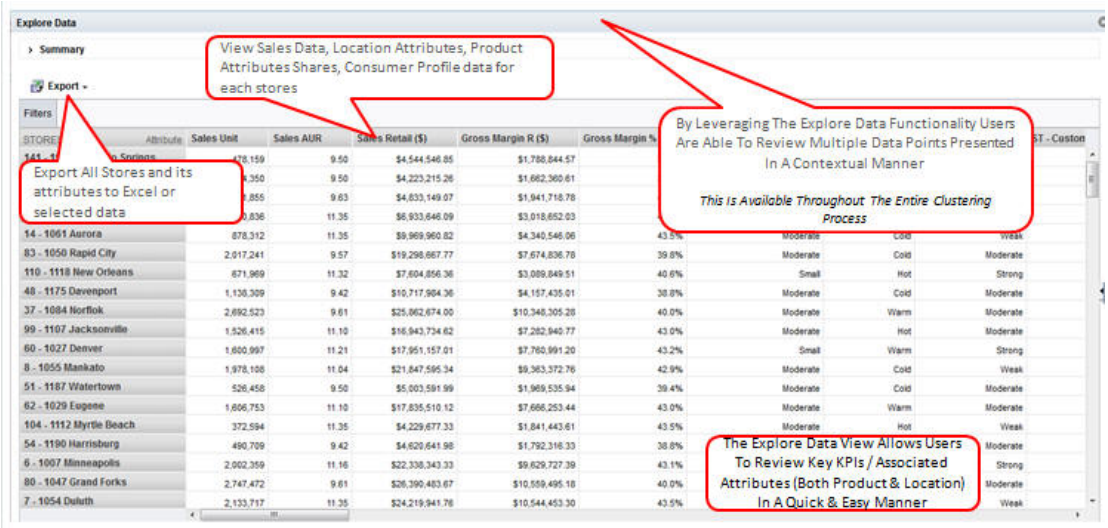
Table 4-9 Explore Data: Summary

Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Merchandise	The merchandise level and nodes for the cluster.
Location	The location level and nodes for the cluster.
Fiscal Period	The time period for the cluster.
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Merchandise Hierarchy Type	Provides details about which type of hierarchy the cluster criteria have been created for.

View Stores

This area displays a nested list of the stores in the cluster you have defined and data for each store for each of the relevant attributes for the Cluster by option you selected to define the cluster. You can see data at the aggregated level as well as at the individual level. Filters are provided so that you can filter the display, for example, by category. You can see aggregated data at a higher level as well as at the store level.

Figure 4-11 Explore Data



Product Performance

This section displays stores, store sales metrics, location attributes, customer profiles, and product attribute profiles.

Table 4-10 Product Performance

Field	Description
Sales Unit	The sales units for the merchandise, location, and source time period selected when setting up clustering parameters.
Sales Average Unit Retail	The average unit retail sales for the merchandise, location, and source time period selected when setting up clustering parameters.
Sales Retail	The sales revenue for the selected merchandise, location, and source time period.
Gross Margin	The retail sales minus the cost of goods sold for the merchandise, location, and source time period selected when setting up clustering parameters.
Gross Margin Percent	The retail sales minus the cost of goods sold divided by the retail sales for the merchandise, location, and source time period selected when setting up clustering parameters.
Location Attribute	The retailer-configured location attributes.
Customer Profile	The retailer-configured customer profiles.
Product Attribute	The retailer-configured product attributes.

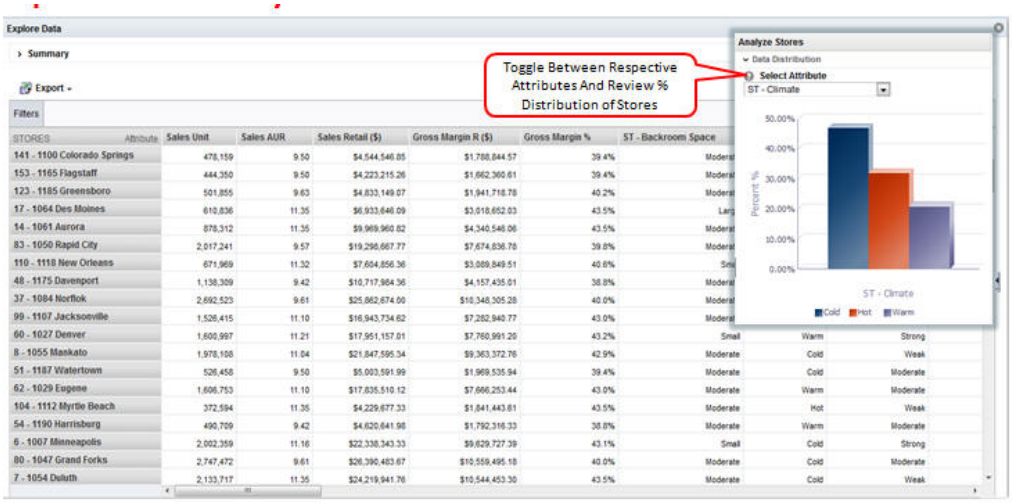
Contextual Area

This area provides a graphical illustration of the detailed data distribution about the cluster.

Analyze Stores

In Explore Data, the BI displays the data distribution of the location by each participating attribute as well as other configured informational attributes. Advanced Clustering identifies the bins based on the underlying data and displays the histograms. It provides the percentage of stores that are present in a location. For example, a company may have 45 percent of stores in cold regions.

Figure 4-12 Clustering Analyze Stores



Category Variability

By analyzing store variability, you can determine if it is worth creating store clusters for the selected categories in the selected location. Three sections are displayed.

A grid is displayed for the selected categories and the sales contribution for a selected location.

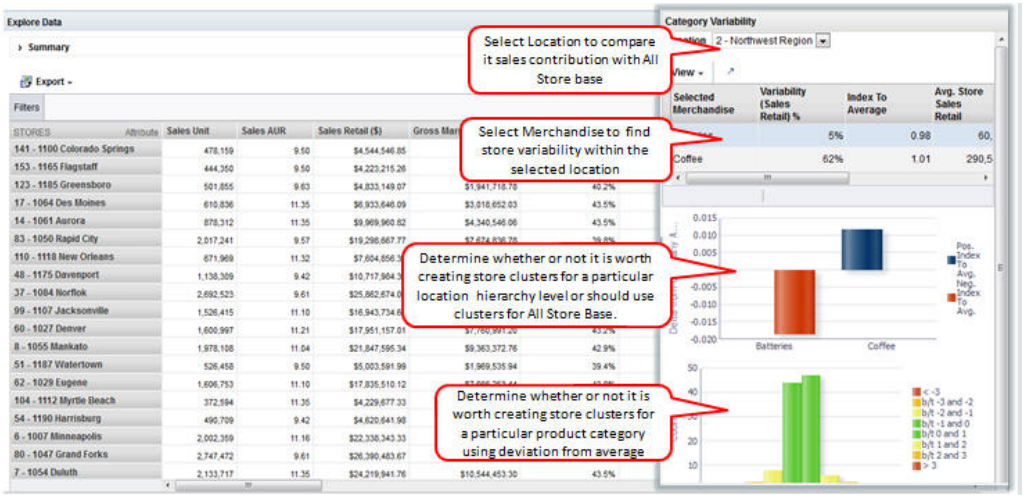
Table 4-11 Category Variability

Property	Description
Categories	A list of the selected categories that are used for store variability analysis.
Variability	The relative standard deviation of the stores in the category. A larger value for the standard deviation indicates greater store variability for the category. Such a category is a possible candidate for store clustering.
Index to average	For a selected location, an indication of how the store performs compared to the all store base. A value close to 1 indicates that the selected location is similar to the all store base. If the value is lower or higher, it indicates that the sales averages for the stores in the selected location are different from the all store base and that you should consider creating store clusters for the selected location.
Average store retail	Average store retail $ for the category for the selected location.
Average store unit	Average store units for the category for the selected location.
Positive/negative index to average	The difference in value for the index to average for the all store base to selected location. For example, a value of 1-index to average < 1 or a value of index to average -1 > 1.

A graph is displayed for the index to average. This shows how the selected location performs compared to the all store base if the average sales metric is below, above, or the same when compared to the all store base average. A red color indicates a value below the all store base average. A blue color indicates a value above the all store base average.

A graph for standard deviation is displayed. This shows the standard deviation for the selected category. If the store value is greater than two standard deviations, then store clustering should be considered for the selected merchandise because the stores sales variability is sufficient.

Figure 4-13 Category Variability



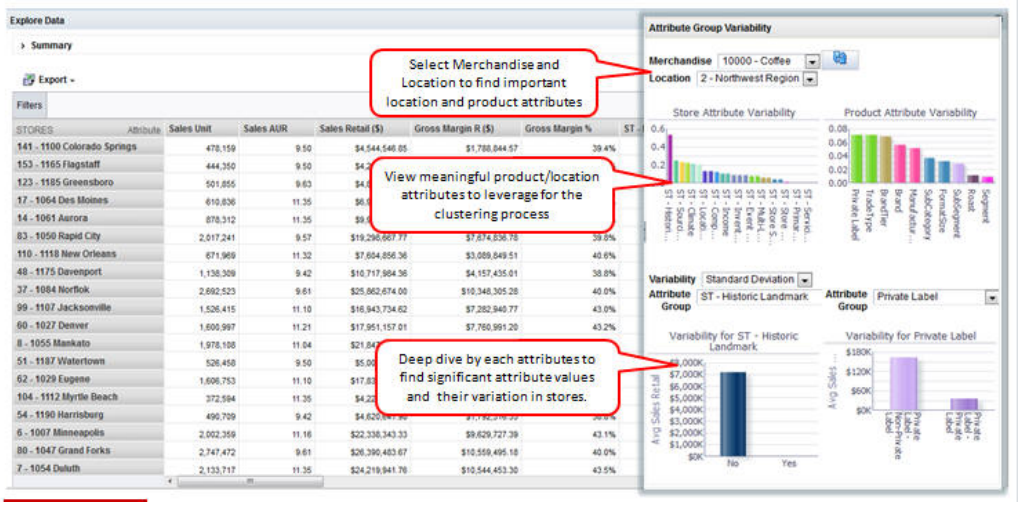
Attribute Group Variability

This is used to analyze the most significant attributes (product or location) for a specific merchandise in a selected location.

The attribute group variability section shows attribute graphs for location and product, indicating the key attributes that are driving sales performance.

The attribute variability section, for each attribute group, shows the sales variability for each attribute value in the selected stores. It calculates the index for the attribute group, indicating attribute significance.

Figure 4-14 Attribute Group Variability



Cluster Setup Stage

You can use this stage to perform what-if analysis by defining one or more scenarios that are based on a specified number of clusters and attributes. You can select one or more attributes and assign different weights to the attributes. The attributes and weights you assign are then fed to the clustering analytics to calculate the weighted distance. Using these scenarios, you can experiment with different numbers of clusters, participating attributes, and their weights. You can either define the maximum number or the minimum number of clusters or alternatively define a specific number of clusters that you want to be generated. Once the scenarios are generated, different scenarios can be compared. You can also use other features in this stage to copy or delete scenarios.

Process

Here is the high-level process for setting up scenarios.

1. Either select the name of a scenario you want to modify or enter a name for the new cluster you want to create.
2. If you want the application to optimize the number of clusters, enter minimum and maximum values for the number of clusters.
3. If you want the application to generate a specific number of clusters, enter that value. In this case, the application generates the exact number of clusters and provides the optimal number of clusters as informational data.
4. Optionally, configure the weights assigned to the attributes. The total must add up to 100 percent. Use a value of 0 percent if you do not want a specific attribute to be part of the clustering process.
5. Click the **Execute** icon to execute the scenario. Once the processing is complete, you see the results in the Cluster Results stage.
6. To see a list of all scenarios and the status for each, go to the Scenario List tab.
7. To compare the defining characteristics of two different scenarios, go to the Scenario Compare tab.

Summary

This lists the criteria you initially selected to define the cluster.

Figure 4-15 Cluster Criteria Summary

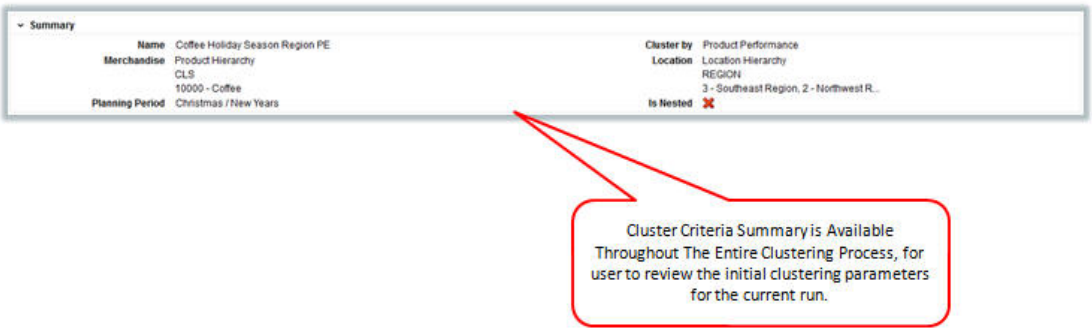


Table 4-12 Cluster Criteria Summary

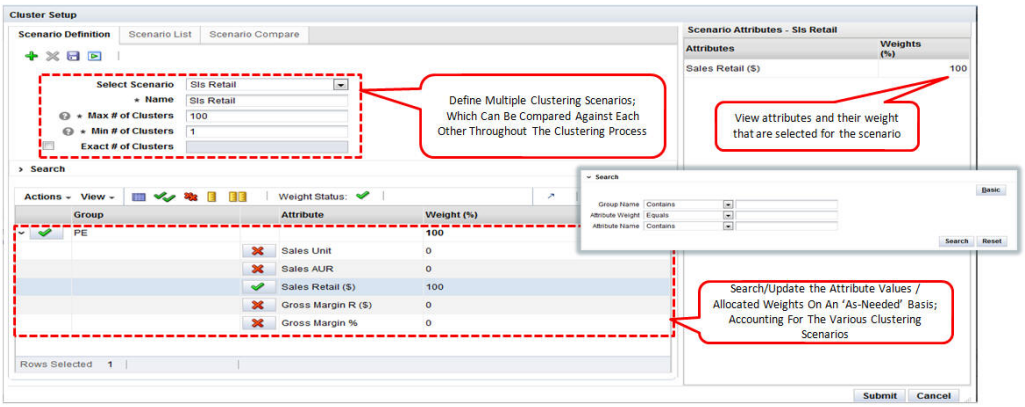
Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Merchandise	The merchandise level and nodes for the cluster.
Location	The location level and nodes for the cluster.
Fiscal Period	The time period for the cluster.
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Merchandise Hierarchy Type	Provides details about which type of hierarchy the cluster criteria have been created for.

Scenario Definition Section

This area has three tabs: Scenario Definition, Scenario List, and Scenario Compare.

Scenario Definition Tab

Figure 4-16 Clustering Scenario Definition



The following information is needed to define a scenario.

Table 4-13 Scenario Definition

Field Name	Description
Select Scenario	Select an existing scenario if you want to modify it.
Name	A unique name that identifies the scenario being defined.

Table 4-13 (Cont.) Scenario Definition

Field Name	Description
Max. # of Clusters	Set the maximum number for the total number of clusters that can be generated. The application determines the optimal number of clusters during the generation process.
Min. # of Clusters	Set the minimum number for the total number of clusters that can be generated. The application determines the optimal number of clusters during the generation process.
Exact # of Clusters	Indicates that the exact number of clusters should be generated. The application does not determine the optimal number of clusters.

Attributes

The Attributes table is used to define which attributes are included in the cluster criteria and the weights that should be assigned to each participating attribute. You can

- search by attribute, attribute value, and attribute weight
- assign equal weight to the selected attribute
- assign weight to the selected attribute or attribute value
- reset weights to default values

The following information defines the attributes that are participating or non-participating.

Table 4-14 Attributes

Field Name	Description
Participating	A check in this column indicates that the attributes participate in the cluster criteria.
Groups	Identifies the group.
Attributes	A description of the attribute.
Weights	The weight assigned to the attribute. All participating attributes can have the same weight or each participating attribute can have a unique weight. The total of all the weights must add up to 100 percent.

The Attributes toolbar, shown in Figure 4-17, includes the following functionality:

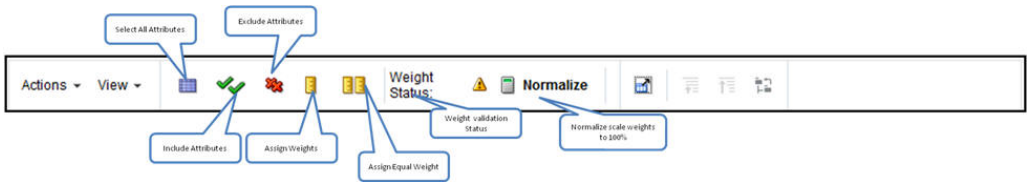
Table 4-15 Attribute Toolbar

Function	Description
Action menu	Resets the weights to the default value overrides that the user provided during configuration.
Weight status	Provides the weight validation status. If the weights do not add up to 100 percent, then a warning is displayed and the scenario cannot be executed.
Include or exclude attributes	Any attribute with a weight equal to zero is not included in the clustering process.

Table 4-15 (Cont.) Attribute Toolbar

Function	Description
Normalize	Scaling attribute weights to ensure that weights are valid. User-provided weights are normalized by applying a weighting average that adds up to 100 percent.

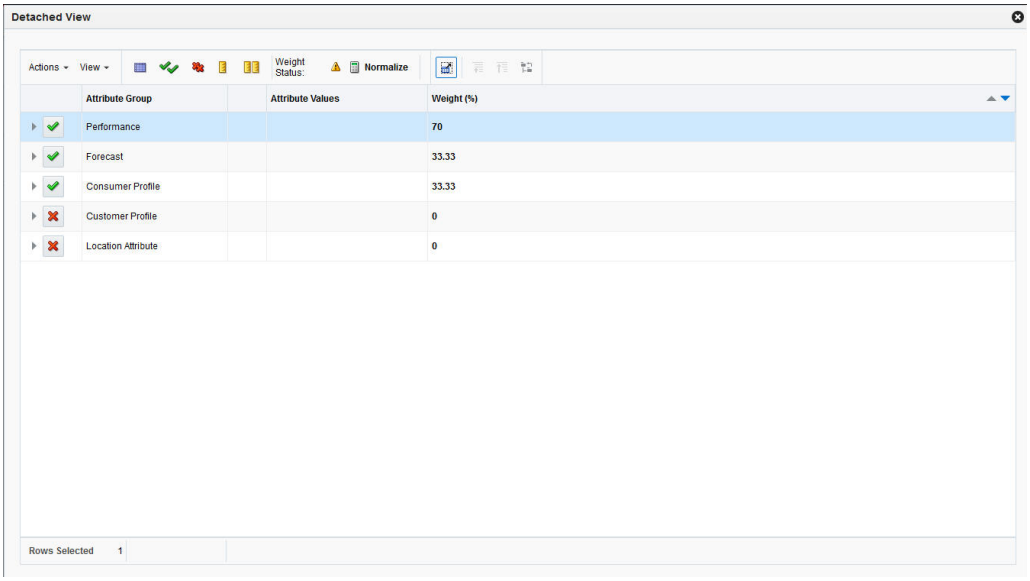
Figure 4-17 Attribute Toolbar



Contextual Area

The contextual business intelligence lists a set of attributes and weights that the current scenario includes as the participating attributes for the clustering process.

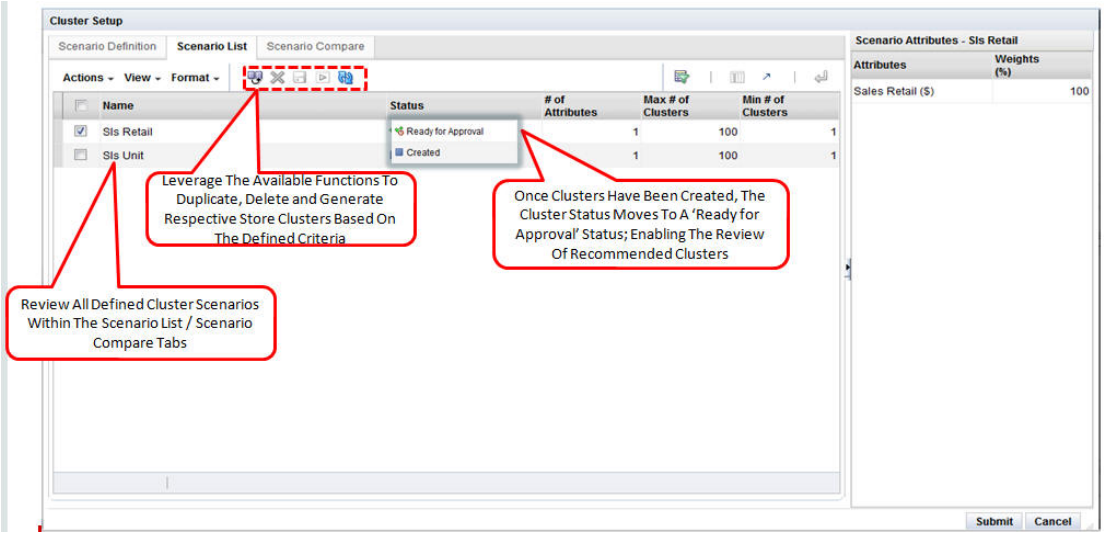
Figure 4-18 Scenario Attributes



Scenario List Tab

The Scenario List summarizes the characteristics for each scenario.

Figure 4-19 Scenario List



You can make a copy of a specific scenario in order to modify it in some way, delete a specific scenario, execute a specific scenario, or save a specific scenario. You can also refresh the scenario list in order to view the updated scenario status.

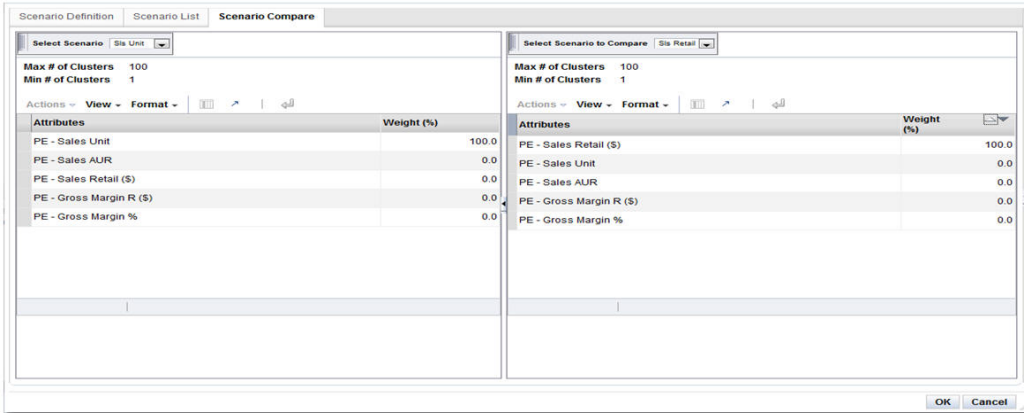
Table 4-16 Scenario List

Field Name	Description
Name	The unique name that identifies the scenario.
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
# of Attributes	The number of attributes is defined by the Cluster by option you select and the weights you optionally assign.
Max. # of Clusters	If you provided a value for this in the scenario definition, that number is displayed here.
Min. # of Clusters	If you provided a value for this in the scenario definition, that number is displayed here.

Scenario Compare Tab

You can select two scenarios from the list to compare. The scenarios you select from the Scenario list are shown side-by-side to facilitate the comparison.

Figure 4-20 Scenario Compare



Cluster Results Stage

After you select a scenario and execute it, you can see the results in this stage. The application uses the data and the parameters you defined in order to group stores together that are most similar according to the characteristics you selected and to separate stores that are most dissimilar. You can also use this stage to rename a cluster.

Process

You use this stage to review clusters, their composition, and the cluster hierarchy, using the grid view and the graph view. This includes

- Reviewing a cluster to see the goodness of fit by using the scores. Determine if any clusters are outliers that warrant further analysis.
- Rank the scenarios (cluster sets) to see how well they are separated and how compact the stores are within each cluster.
- View the optimality of the clusters recommended by the application to determine if increasing the number of clusters beyond the optimal number is significant.
- Rename the cluster after analyzing the centroids and before the cluster is approved.

Summary

This lists the criteria you initially selected to define the cluster.

Figure 4-21 Cluster Criteria Summary

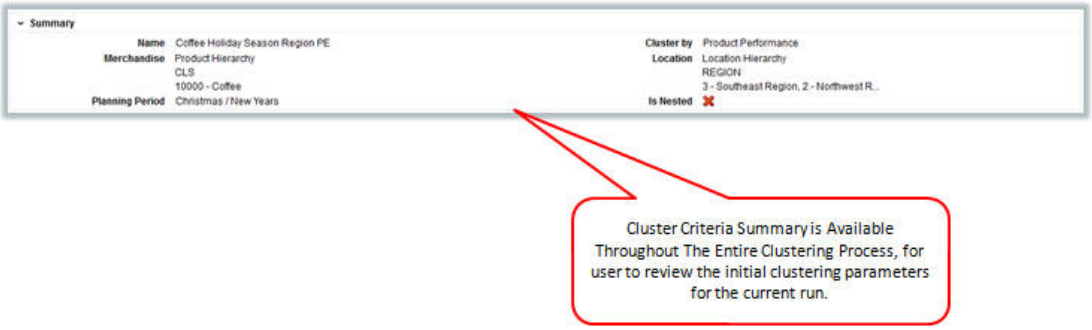


Table 4-17 Cluster Criteria Summary

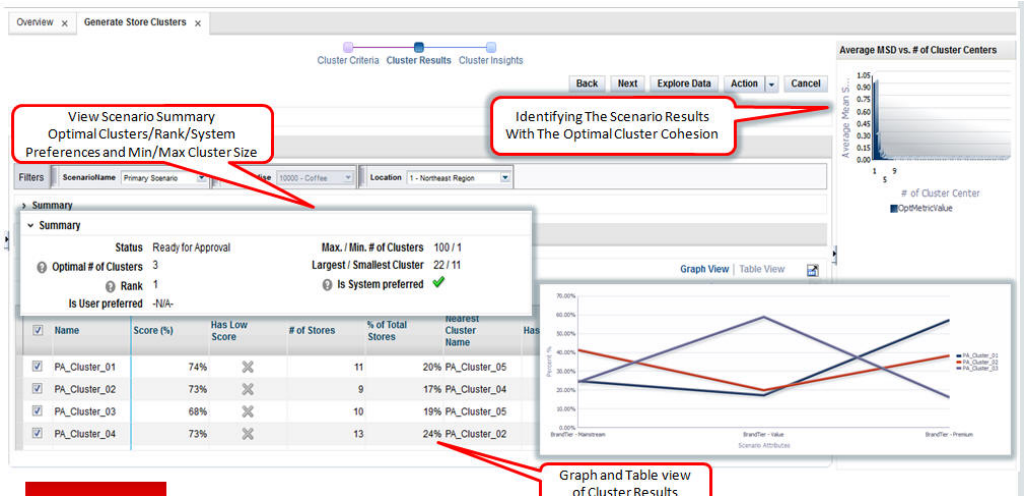
Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Merchandise	The merchandise level and nodes for the cluster.
Location	The location level and nodes for the cluster.
Fiscal Period	The time period for the cluster.
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Merchandise Hierarchy Type	Provides details about which type of hierarchy the cluster criteria have been created for.

Scenario Results Section

The Scenario Results section displays the following:

- The Scenario Summary, which provides key cluster set attributes for the executed scenario as well as its status.
- The Scenario Results, which has three tabs: Clusters, Cluster Composition, and Cluster Hierarchy.

Figure 4-22 Clustering Scenario Results



The Summary section provides an overview of the characteristics of the clusters.

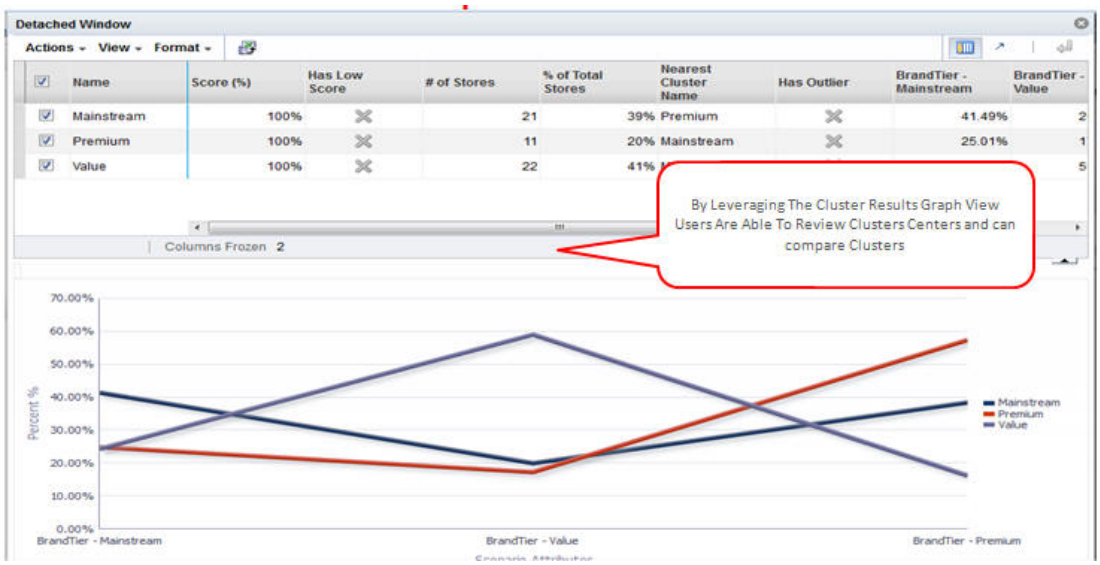
Table 4-18 Scenario Results Summary

Field	Description
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
Optimal # of Clusters	The optimal number of clusters determined by the optimization.
Rank	The application compares executed scenarios and ranks them. A value of 1 indicates the best scenario.
Max./Min # of Clusters	The number you provided for the maximum and minimum number of clusters to calculated.
Largest/Smallest Cluster	Provides the sizes of the largest cluster and the smallest cluster in order to show the range of values.
Is System Preferred	Indicates whether or not the system prefers the scenario.
Is User Preferred	Indicates whether or not the user prefers the scenario.

The Clusters section provides the cluster results for each individual cluster in the scenario in either a Graph View or a Table View. The attributes displayed depend on the Cluster by option chosen in the Cluster Criteria stage.

The Graph View shows the percentage for each attribute in the cluster.

Figure 4-23 Cluster Results Graph View



The Table View provides details that can help you analyze the cluster.

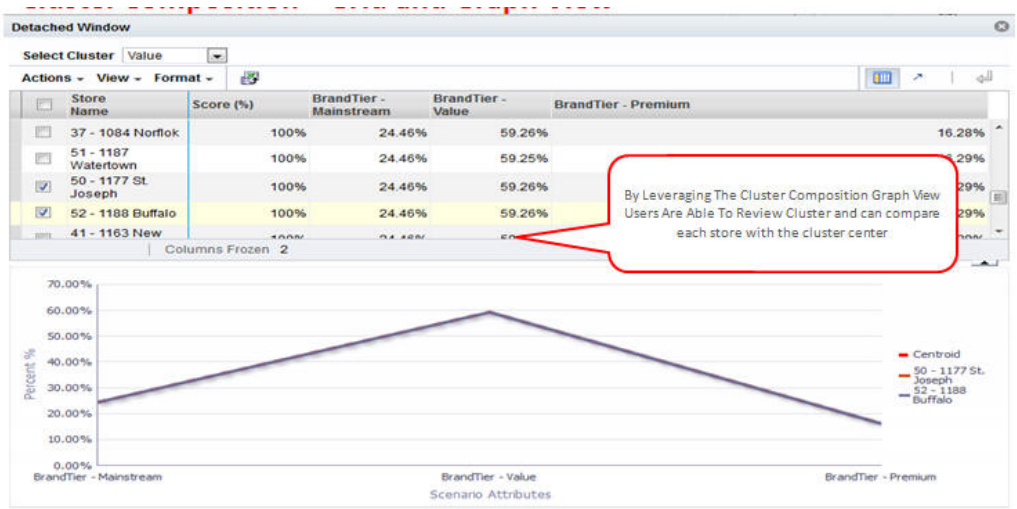
Table 4-19 Scenario Results - Clusters: Table View

Field	Description
Name	The name you assigned to the cluster.
# of Stores	The number of stores in the cluster.
% of Total Stores	The percentage of the total stores that the number of stores represents.
Nearest Cluster Name	The name of the cluster that is most similar to this cluster.
Score %	This value is calculated at the level of store and then averaged to the cluster. The probability, expressed as a percentage, of a store being present in this cluster rather than any of the other clusters.
Has Outlier	Indicates a cluster with the number of stores below a threshold. For example, the number of stores are below certain percentage of the number of stores in a cluster.
Has Low Score	The score threshold can be defined in two ways. The default threshold is calculated as a probability that a store exists in any one of the clusters. The user can further override this threshold at the time of deployment by each Cluster by.
Attributes and their %	For each attribute specific to the Cluster by option, the value indicates the percentage that attribute represents within the total cluster.

The Clusters Composition sub-tab breaks down the cluster into its component parts and shows the percentages for each attribute.

The Table View, shown in Figure 4-24, shows attributes and the score percent.

Figure 4-24 Cluster Composition Table View



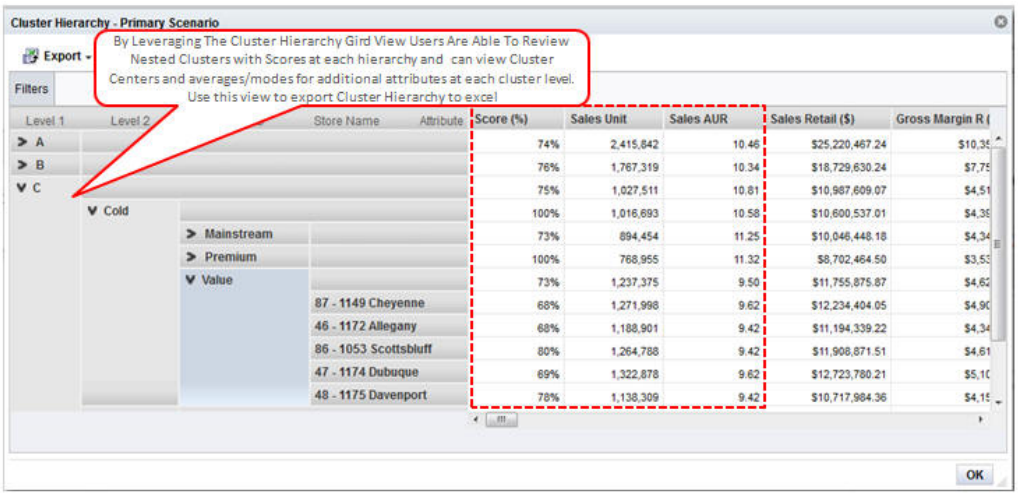
The Graph View shows the centroid of a cluster and allows the user to compare each store with the cluster centroid.

Cluster Hierarchy

The cluster hierarchy shows the parent-to-child cluster relationship with the cluster centers. Any additional store attributes are rolled up as averages or modes for each cluster level. You can view the cluster hierarchy for each scenario.

The cluster hierarchy is displayed for each scenario. In the case of nested hierarchies, the cluster hierarchy is displayed in tree format. You can select the attributes to view and export the selected attributes with a cluster hierarchy tree in Excel format.

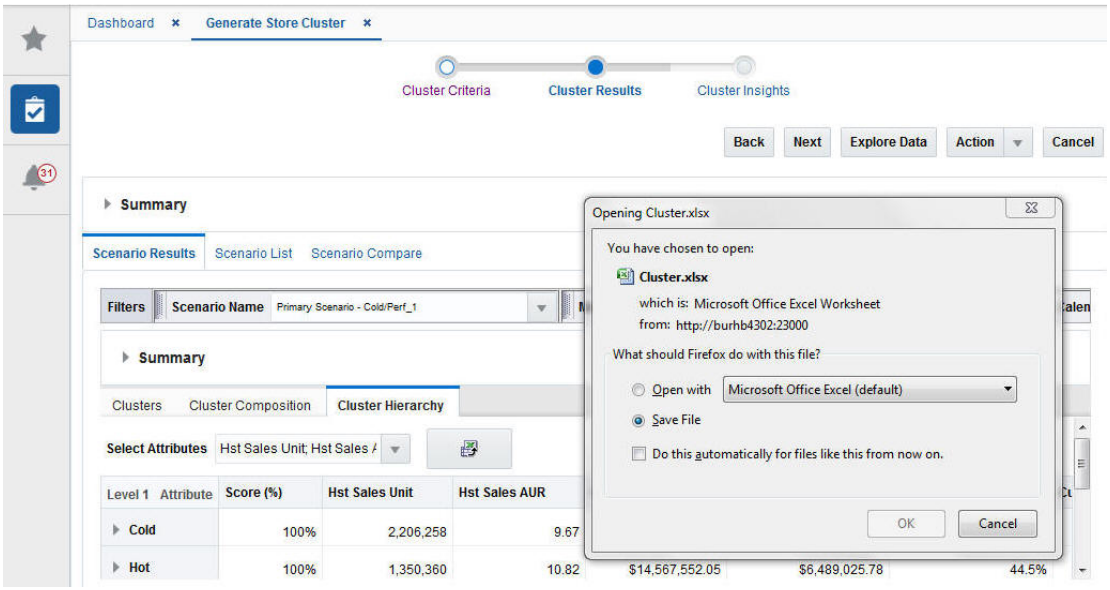
Figure 4-25 Cluster Hierarchy Grid View



Export to Excel

You can select attributes to view data in Pivot and Excel to Export. On selecting icon export to excel user can open or save excel sheet. See excel format below to review cluster hierarchy and their aggregate.

Figure 4-26 Save as Excel File



The Excel-generated file can be then used by external visualization tools to generate graphs not supported in the application.

Figure 4-27 Excel File

	A	B	C	D	E	F	G	H	I	J
1	LEVEL1	LEVEL2	LEVEL3	Store Name	Score (%)	Hst Sales Uni	Hst Sales AUR	Hst Sales Retail	lst Gross Margin	t Gross Margin
2	Cold				100%	2,206,258	9.67	$20,994,088.41	$8,488,147.56	40.6%
3	Cold	Cold/Perf_3			81%	571,643	10.13	$5,781,374.03	$2,490,862.27	42.8%
4	Cold	Cold/Perf_4	CP_Cluster_01		100%	571,643	10.13	$5,781,374.03	$2,490,862.27	42.8%
5	Cold	Cold/Perf_5	CP_Cluster_02	68 - 1035 Spokane	100%	299,051	11.01	$3,293,860.63	$1,555,842.29	47.2%
6	Cold	Cold/Perf_6	CP_Cluster_03	22 - 1069 Detroit	100%	323,853	11.07	$3,583,657.79	$1,692,901.38	47.2%
7	Cold	Cold/Perf_7	CP_Cluster_04	51 - 1187 Watertown	100%	606,559	8.71	$5,285,405.82	$2,002,515.98	37.9%
8	Cold	Cold/Perf_8	CP_Cluster_05	76 - 1043 Casper	100%	342,003	10.19	$3,484,737.64	$1,361,867.45	39.1%
9	Cold	Cold/Perf_9	CP_Cluster_06	89 - 1151 Vail	100%	592,923	8.83	$5,237,635.10	$2,019,730.73	38.6%
10	Cold	Cold/Perf_10	CP_Cluster_07	66 - 1033 Great Falls	100%	813,199	10.89	$8,854,491.61	$4,056,034.53	45.8%
11	Cold	Cold/Perf_11	CP_Cluster_08	18 - 1065 Omaha	100%	589,169	11.31	$6,661,449.26	$3,174,290.53	47.7%
12	Cold	Cold/Perf_12	CP_Cluster_09	21 - 1068 Lansing	100%	305,039	10.98	$3,348,926.53	$1,587,049.30	47.4%
13	Cold	Cold/Perf_13	CP_Cluster_10	20 - 1067 South Bend	100%	275,063	10.96	$3,014,701.44	$1,419,485.37	47.1%
14	Cold	Cold/Perf_14	CP_Cluster_11	16 - 1063 Cedar Rapids	100%	709,508	11.33	$8,035,512.33	$3,837,476.49	47.8%
15	Cold	Cold/Perf_15	CP_Cluster_12	91 - 1197 Pierre	100%	607,093	8.75	$5,313,445.67	$2,025,722.48	38.1%
16	Cold	Cold/Perf_16	CP_Cluster_13	14 - 1061 Aurora	100%	701,834	11.32	$7,943,662.79	$3,796,386.44	47.8%
17	Cold	Cold/Perf_17	CP_Cluster_14	52 - 1188 Buffalo	100%	572,589	8.76	$5,015,910.56	$1,906,445.48	38.0%
18	Cold	Cold/Perf_18	CP_Cluster_15	17 - 1064 Des Moines	100%	496,015	11.22	$5,563,602.56	$2,652,263.81	47.7%

Scenario List

The Scenario List section contains one table with details about each cluster. Here, you can approve or reject a cluster.

After a cluster is approved, it is available for other applications.

Figure 4-28 Clustering Scenario List

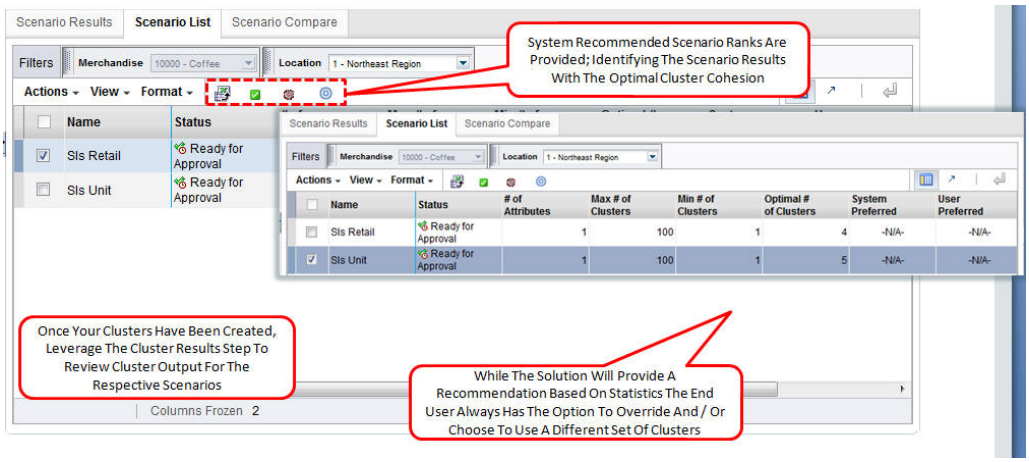


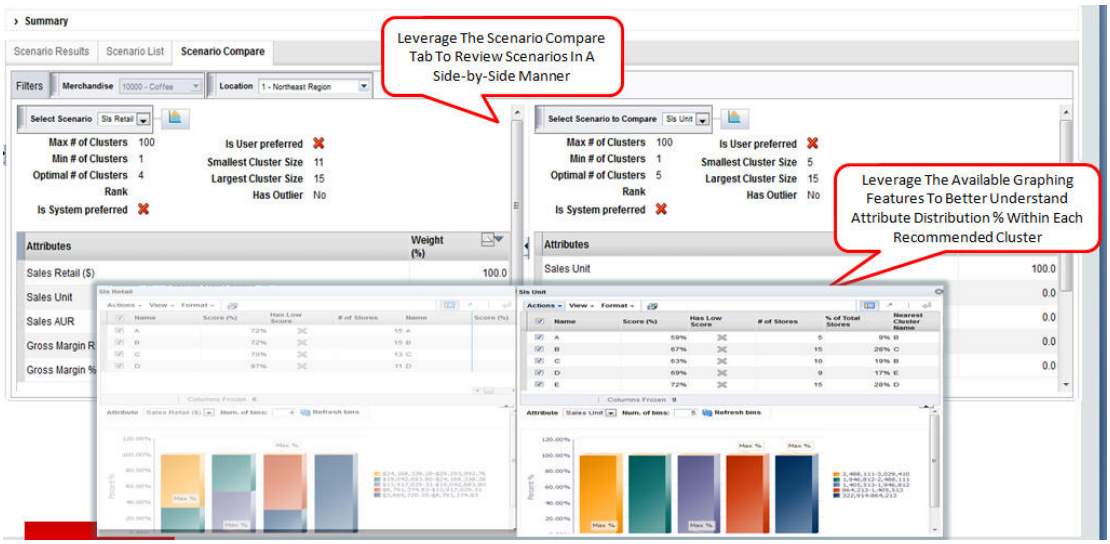
Table 4-20 Scenario List

Field	Description
Name	Name assigned to the scenario.
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
Parent Cluster Level #	The parent cluster name to which the scenario results apply if nested cluster criteria is selected.
# of Attributes	The total number of attributes used, as determined by the weight assigned to each attribute.
Max. # of Clusters	The value used for the maximum in the scenario execution, if this option used.
Min. # of Clusters	The value used for the minimum in the scenario execution, if this option used.
Optimal # of Clusters	The value used for the optimal number of clusters in the scenario execution, if this option used.
System Preferred	Indicates whether the scenario is the one the application prefers.
User Preferred	Indicates whether the scenario is one the user prefers.
Rank Sequence	Indicates the ranking the scenario is given by the application.

Scenario Compare

The Scenario Compare section shows two clusters of your choosing side by side so that you can compare the main results of each, using the same characteristics used in Scenario Results and Scenario List.

Figure 4-29 Compare Scenarios



The information displayed includes:

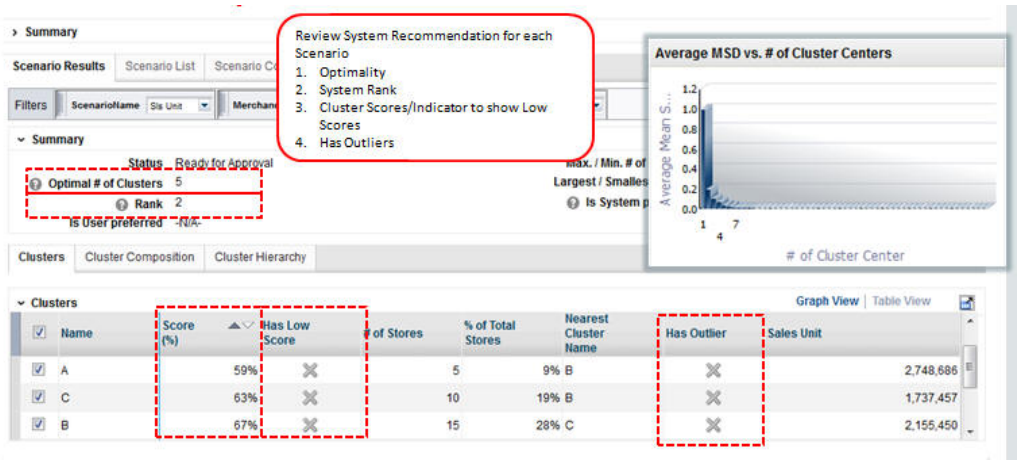
Table 4-21 Scenario Compare

Field Name	Description
Max. # of Clusters	The value used for the maximum in the scenario execution, if this option used.
Min. # of Clusters	The value used for the minimum in the scenario execution, if this option used.
Optimal # of Clusters	The value used for the optimal number of clusters in the scenario execution, if this option used.
Rank	The value for the rank.
Is System Preferred	Indicates whether the scenario is the one the application prefers.
Is User Preferred	Indicates whether the scenario is the one the user prefers.
Smallest Cluster Size	The size of the smallest cluster.
Largest Cluster Size	The size of the largest cluster.
Has Outlier	Indicates a cluster with the number of stores below a threshold. For example, the number of stores are below certain percentage of the number of stores in a cluster.
Attributes	A list of relevant attributes.

Scenario System Recommendations

The application provides the following recommendations at the scenario (cluster set), cluster, and store levels.

Figure 4-30 Scenario Recommendations



Scenario Optimality

This graph indicates how the system identifies the best number of clusters for a given data set. It starts with a small number of cluster centers and searches for the number beyond which there is little improvement in the mean squared distance (MSD). At this point, increasing the number of cluster centers any more only decreases the MSD by a small amount, and the marginal improvement is small.

Scenario Rank

You can see the ranking of all scenarios in the Cluster Results step. The scenario with the highest rank is designated as System Preferred. The ranking is based on the following:

- How many similar stores are contained in the cluster.
- How well separated the clusters are from each other.

Outlier Indicator

This provides an outlier indicator in the cluster list if the cluster has an outlier store. Two outlier rules are supported. The distance from the centroid indicates that if a store is beyond a certain limit for the configuration threshold from the centroid, then the cluster to which the store belongs is marked as an outlier. When the size of a cluster is compared with the total stores, if the number of stores in the cluster is below a certain configured percentage of the total stores, then the cluster is marked as an outlier.

Cluster Scores

The application provides scores for clusters, based on the calculated threshold score. The score is based on the assumption that each store has an equal chance of being a member of the cluster. A high score indicates that the store is close to the centroid. A low score indicates that the store is an outlier.

New Stores and Stores with a Poor History

Advanced Clustering supports post-processing rules in order to allocate stores that are new or that have a poor history. These rules can be configured for each criterion and can be changed during deployment.

- Like Stores. This rule allocates new stores or stores with a poor history to the same clusters that the like location belongs to. It requires data to be provided to Advanced Clustering that defines the mapping between the location and like locations. This mapping can be configured by merchandise, and one location can be mapped to multiple locations with different weights. For example, a like location can be used to correct a store with poor history or to allocate a new store to a valid performance cluster.
- Largest Clusters. This rule allocates new stores or stores with a poor history to the largest cluster identified by Advanced Clustering. Stores can be allocated to a bigger group of stores. For example, a store that has not yet formed a customer base can be allocated to the largest cluster.
- Cohesive Clusters. This rule allocates new stores or stores with a poor history to the most compact cluster identified by Advanced Clustering. Stores can be allocated to a compact group of stores. For example, stores can be assigned to a cluster that has not been affected because of outliers.

Insights Stage

Use the Insights stage to analyze a scenario, its clusters, and its hierarchy, based on performance and attribute contributions, prior to the approval of the scenario. This stage includes the following tasks:

- Approve a cluster scenario
- Create a new cluster within a scenario
- Rename clusters within a scenario
- Rank scenarios, if not completed earlier
- Flag a cluster scenario as "system preferred"
- Review a cluster hierarchy in a nested cluster

Select from the following views in this stage:

- Criteria view. Displays the parent cluster, if it exists, or the scenario for root-level clusters.
- Parent cluster. Displays any child clusters.
- Cluster. Displays the stores under the selected cluster.

Each cluster is identified by the following information:

Table 4-22 Insights Stage

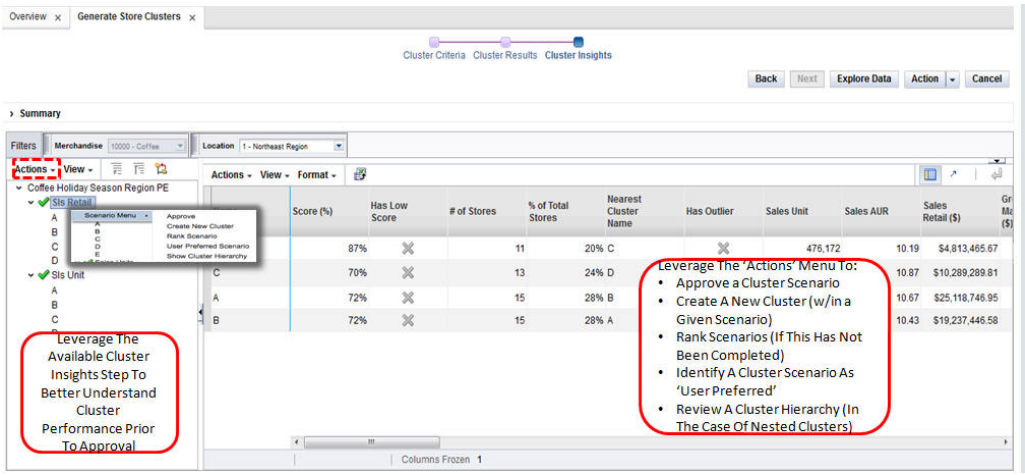
Field Name	Description
Name	The name assigned to the cluster.
Nearest Cluster Name	The name of the cluster that is most similar to the named cluster.
# of Stores	The number of stores in the cluster.

Table 4-22 (Cont.) Insights Stage

Field Name	Description
Is Outlier	Whether or not the cluster is considered an outlier. If it is an outlier, you may want to review that store.
All Sales Metric	Financial information about the store.
Score %	This value is calculated at the level of store and then averaged to the cluster. The probability, expressed as a percent, of a store being present in this cluster rather than any of the other clusters.
Has Low Store	Indicates a cluster that falls below a defined threshold.
% Total Stores	The percentage of the total stores that the number of stores represents.
Sales Retail	The sales revenue for the cluster or store. The merchandise and the source time period are selected when the cluster parameters are set up.
Sales Unit	The sales units for the cluster or store. The merchandise and the source time period are selected when the cluster parameters are set up.
Sales Average Unit Retail	The average unit retail sales for the cluster or store. The merchandise and the source time period are selected when the cluster parameters are set up.
Gross Margin Retail	The retail sales minus the cost of goods sold for the cluster or store. The merchandise and the source time period are selected when the cluster parameters are set up.
Gross Margin Percent	The retail sales minus the cost of goods sold divided by the retail sales for the cluster or store. The merchandise and the source time period are selected when the cluster parameters are set up.
Cluster Centers	Additional attributes for each cluster.

You use this stage to view all scenarios, examine and compare sales metrics for the various clusters, and manage clusters by approving, rejecting, merging, or deleting clusters.

Figure 4-31 Clustering Insights



Summary

This lists the criteria you initially selected to define the cluster.

Figure 4-32 Cluster Criteria Summary

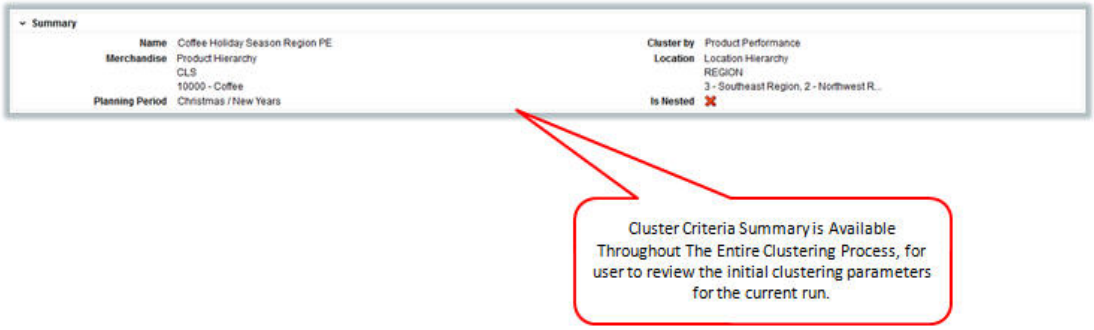


Table 4-23 Cluster Criteria Summary

Field	Description
Name	The name you provided for the cluster in the Cluster Criteria stage.
Cluster By	A predefined group of attributes that include Consumer Profile, Product Performance, Store Attribute, Product Attribute, and Mixed Attribute. These criteria types are sets of attributes. For example, store attributes are the properties of a store. These properties can include ethnicity, store format, and store size.
Merchandise	The merchandise level and nodes for the cluster.
Location	The location level and nodes for the cluster.
Fiscal Period	The time period for the cluster.
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Merchandise Hierarchy Type	Provides details about which type of hierarchy the cluster criteria have been created for.

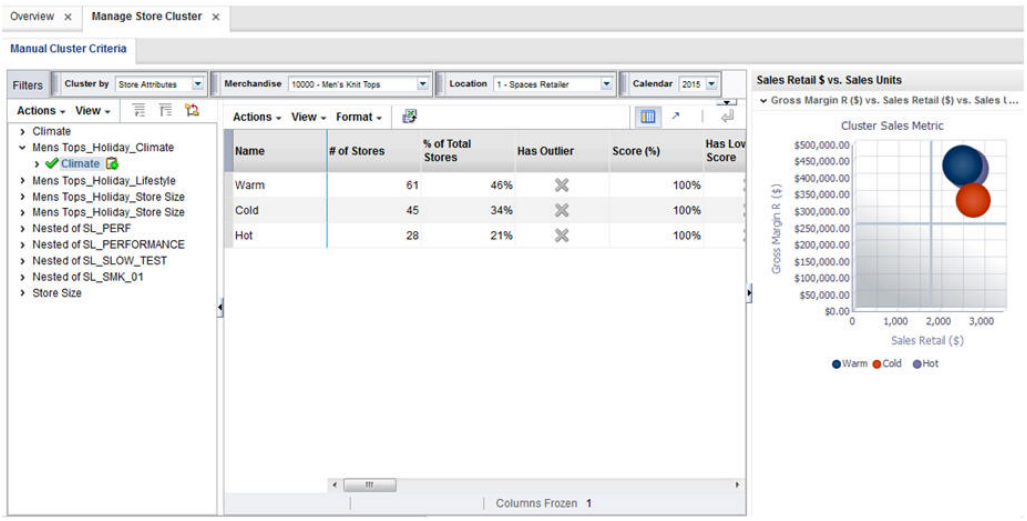
Contextual Information

The graphs displayed in the Clustering Insights stage are also displayed in the Manage Store Clusters tab. For a discussion, see Contextual Information.

Manage Store Clusters Tab

You can use the Manage Store Clusters tab to view a list of the already executed cluster criteria and associated summary details. You can create manual clusters in a scenario, rank/approve/reject scenario, mark a scenario as user preferred, rename clusters, and override any cluster composition within a scenario.

Figure 4-33 Clustering Manage Cluster Criteria



The following information is displayed.

Table 4-24 Manage Store Clusters

Field	Description
Is Nested	Indicates whether or not the cluster is nested within another cluster.
Is Deployed	Indicates that the cluster has been deployed.
Is Shared	A check mark indicates that more than one merchandise or location node is used in the cluster.
Scenario Created	The number of scenarios that were created.
Scenario Executed	The number of scenarios that were executed.
Name	The name of the scenario.
Status	Created, Ready for Approval, Completed with Errors, Approved, Rejected.
# Attributes	The total number of attributes used, as determined by the weight assigned to each attribute.
Max # Clusters	The value used for the maximum number of cluster centers in the scenario execution, if this option used.
Min # Clusters	The value used for the minimum number of cluster centers in the scenario execution, if this option used.
Optimal # Clusters	The value used for the optimal number of clusters in the scenario execution, if this option used.
System Preferred	Indicates whether the scenario is the one the application prefers.
Rank Sequence	Indicates the ranking the scenario is given by the application.

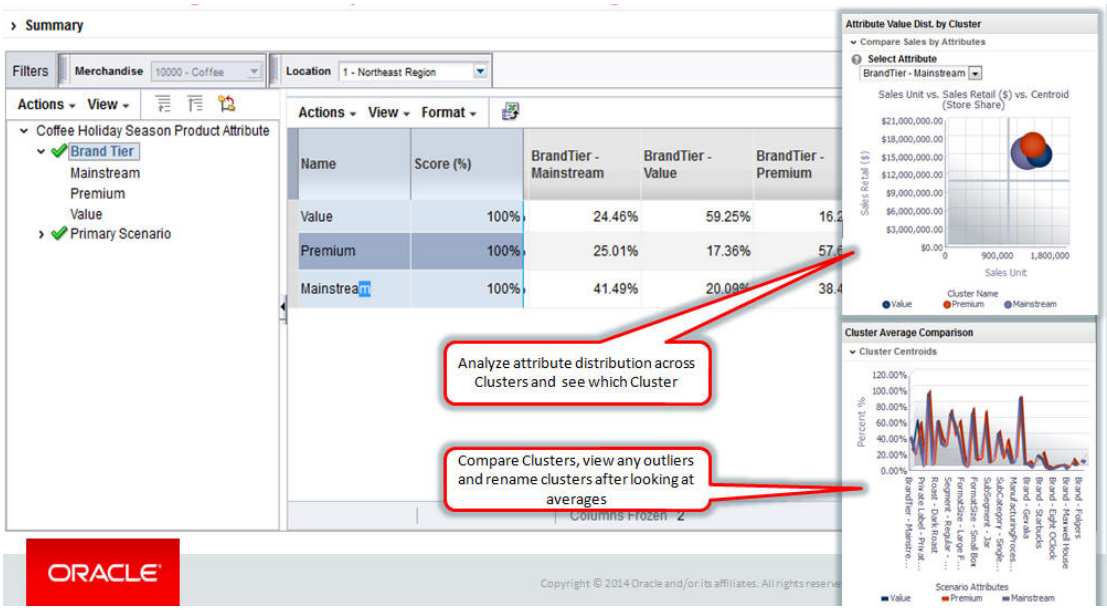
Contextual Information

The following charts are available.

Attribute Value Dist By Cluster

The chart shows a comparison of sales with the average value distribution by clusters. The *y* axis is sales retail and the *x* axis is sales units. The *z* bubble size shows the average value for the selected attribute for the cluster.

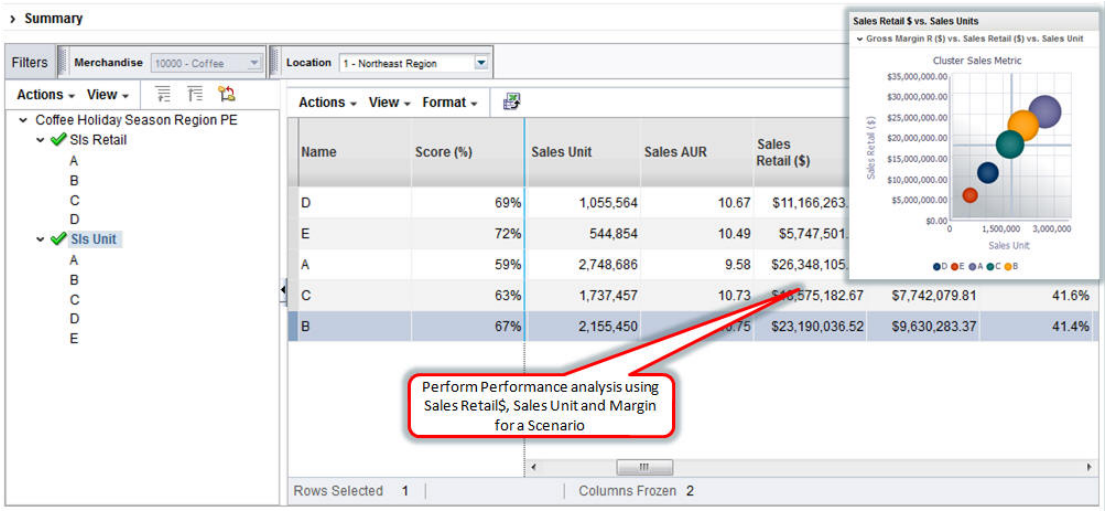
Figure 4-34 Attribute Value Dist. by Cluster



Sales Retail vs. Sales Units

The chart shows a comparison of sales retail and sales unit. The *y* axis is sales retail and the *x* axis is sales unit. The *z* bubble size shows the gross margin retail for a cluster.

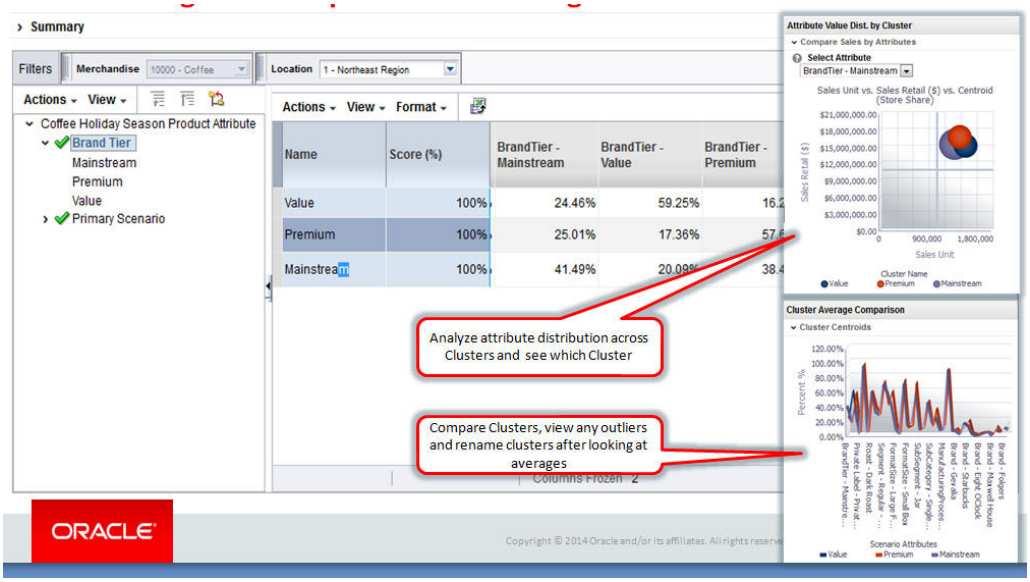
Figure 4-35 Sales Retail vs. Sales Units



Cluster Average Comparison

The chart overlays the centroid of the cluster and provides a comparison of cluster averages so that you can rename the clusters based on the information in the graph.

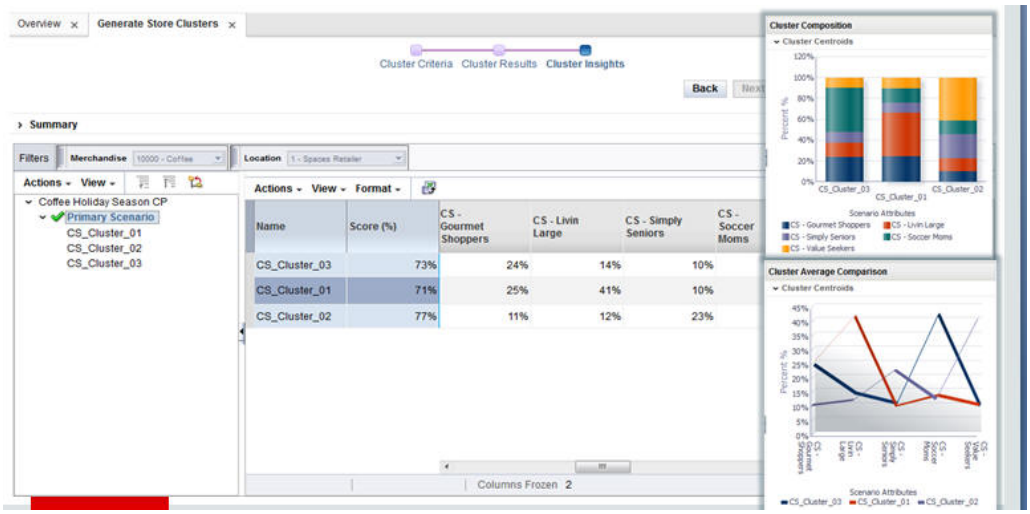
Figure 4-36 Cluster Average Comparison



Cluster Comparison

The chart shows the stacked contribution of each attribute by percentage for each cluster. You can use this chart to determine which cluster contributes the most for an attribute by viewing the clusters side by side.

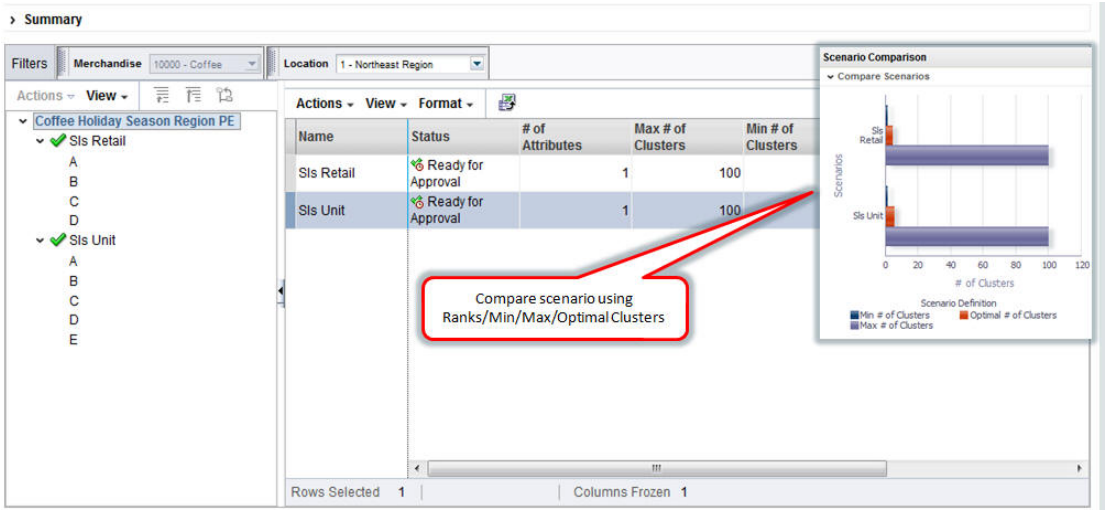
Figure 4-37 Cluster Comparison



Scenario Compare

Multiple scenarios are compared using optimality and scenario rank. This chart is displayed when multiple scenarios for the criteria are available. The following properties are used by the chart: minimum number of clusters, maximum number of clusters, optimal number of clusters, and the rank of the scenarios, from top to bottom.

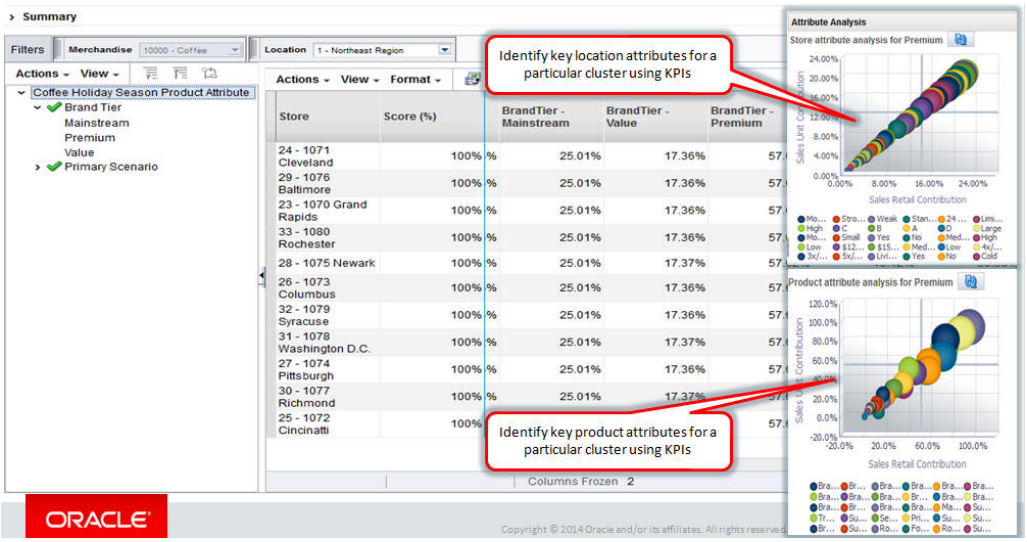
Figure 4-38 Scenario Compare



Attribute Analysis

Stores and product attributes are analyzed and compared to identify the most prominent selling attributes within a cluster. This indicates how the store and product attributes are correlated and what each attribute in the specified cluster contributes to sales. You can make inferences about which attributes in the cluster contribute to significant sales and the potential attributes that should be considered for assortment planning in order to improve sales even more. With both store and product attribute graphs, you can see which location attributes drive the product attribute sales.

Figure 4-39 Attribute Analysis



The following properties are displayed by the graph. Note that the store and product attributes are only displayed when they are configured as part of the Cluster by process.

Table 4-25 Properties

Axis	Description
x-axis	Sales Retail $ contribution, calculated using the sales revenue share of each attribute in the cluster with respect to the total cluster sales revenue.
y-axis	Sales Unit $ contribution, calculated using the sales unit share of each attribute in the cluster with respect to the total cluster sales units.
z-axis	Total Sales Retail $ of each attribute in the selected cluster, indicating, via the bubble, the magnitude of the sales contribution.

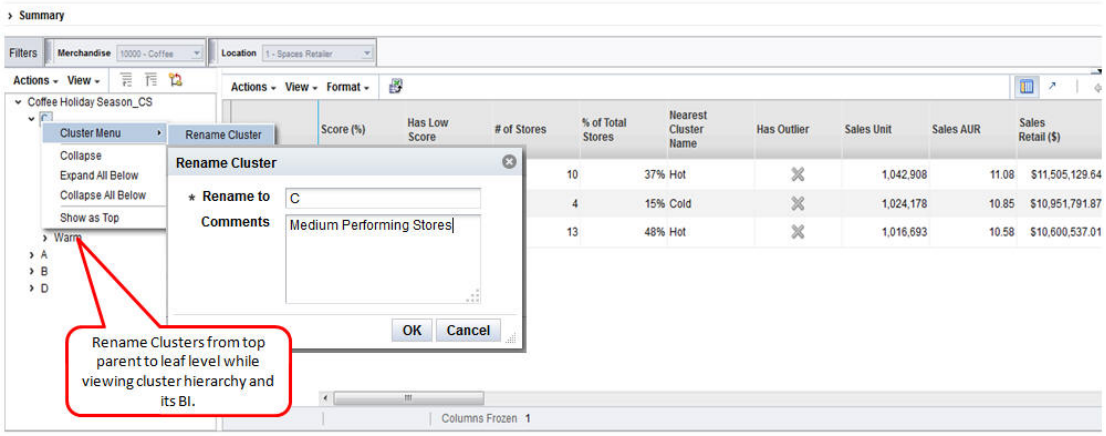
Cluster Overrides

Both Cluster Insights and Manage Store Cluster can be used to override clusters manually. The following tasks, to be completed before the approval process, are available.

Rename Cluster

To complete this on a nested cluster, select the parent or child clusters in the tree. With the contextual information provided, you have the necessary details to understand the cluster across the hierarchy. This task can be performed using the Cluster Contextual menu.

Figure 4-40 Rename Cluster



Create New Manual Cluster

You can create a new manual cluster, describe it, and tag the cluster as Inactive, Flagship, or Manual. This task can be performed using the Scenario Contextual menu.

Inactive

Inactive stores are allocated to these types of clusters. Such stores are either closed, tagged as invalid, or in construction for a specific effective period.

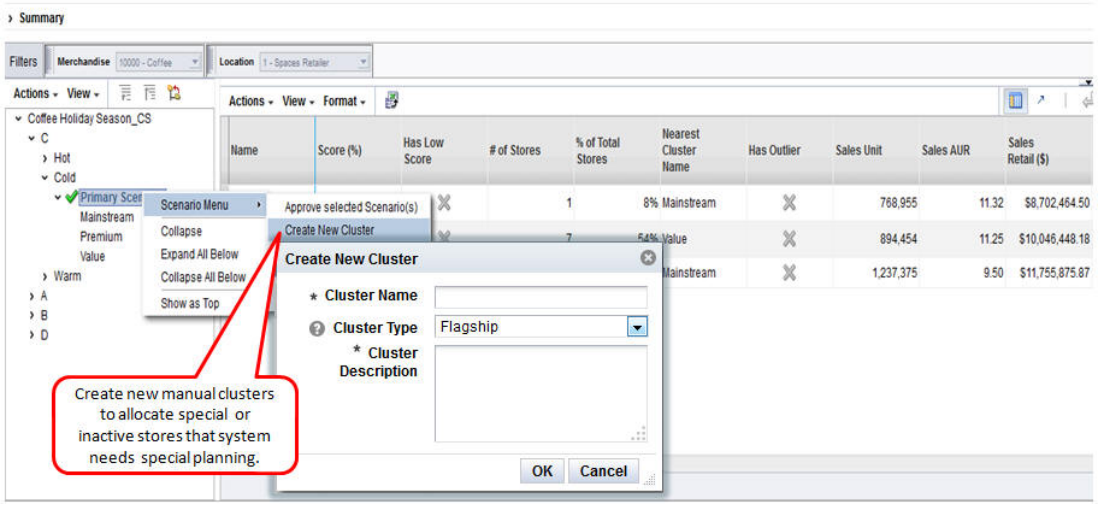
Flagship

Special clusters in which only certain stores reside.

Manually Created

A user manually creates an empty cluster and allocates stores to the cluster using the drag and drop feature. The application automatically re-calculates the cluster centers after the cluster composition changes.

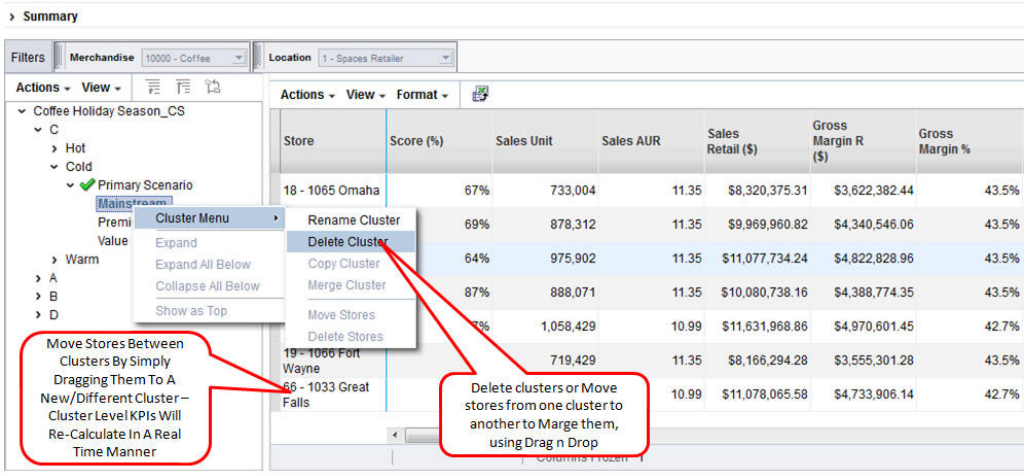
Figure 4-41 Create Manual Cluster



Delete Cluster

You can delete an empty cluster. Prior to deleting the cluster you must move the already allocated stores to another cluster. This can help you to merge clusters at the same level. The application automatically re-calculates the cluster centers after the cluster composition changes. This task can be performed using the Cluster Contextual menu.

Figure 4-42 Delete Cluster



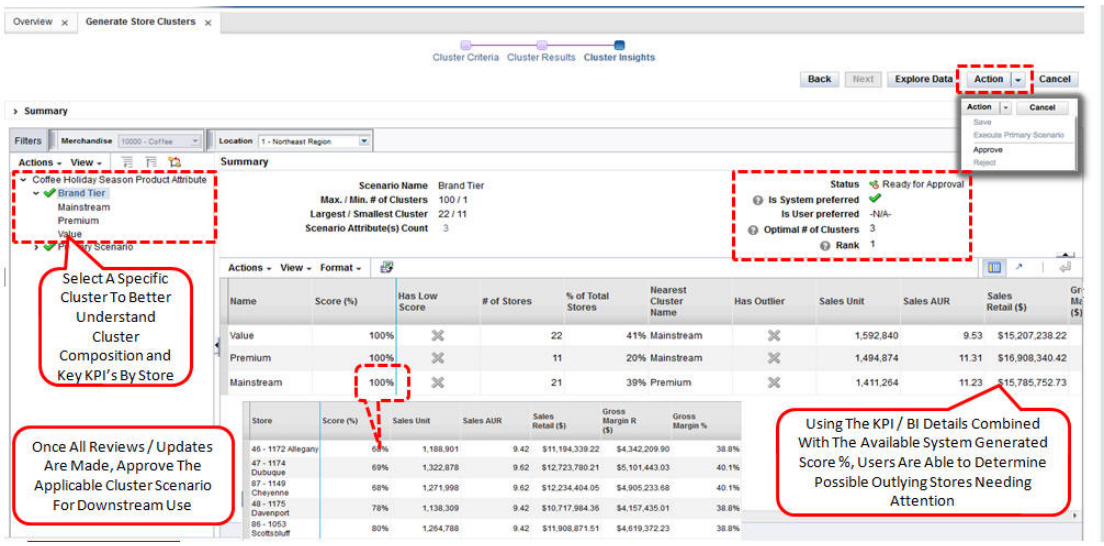
Move Stores to Clusters

Stores can be moved from one cluster to another using drag and drop. The application automatically re-calculates the cluster centers after a cluster composition changes.

Cluster Review, Approve, and Adjust

The approval step is the last step after the review and any manual overrides for the scenario are completed. The scenario results are a set of clusters that are effective for a merchandise, location, and calendar combination. The batch export process selects the last updated approved cluster to deploy to the subscribing applications using an interface file. You can reject an already approved cluster and deploy it if the selected scenario results are better. You are notified of any manual overrides in which approved clusters are modified. They are automatically redeployed.

Figure 4-43 Cluster Review, Approve, and Adjust



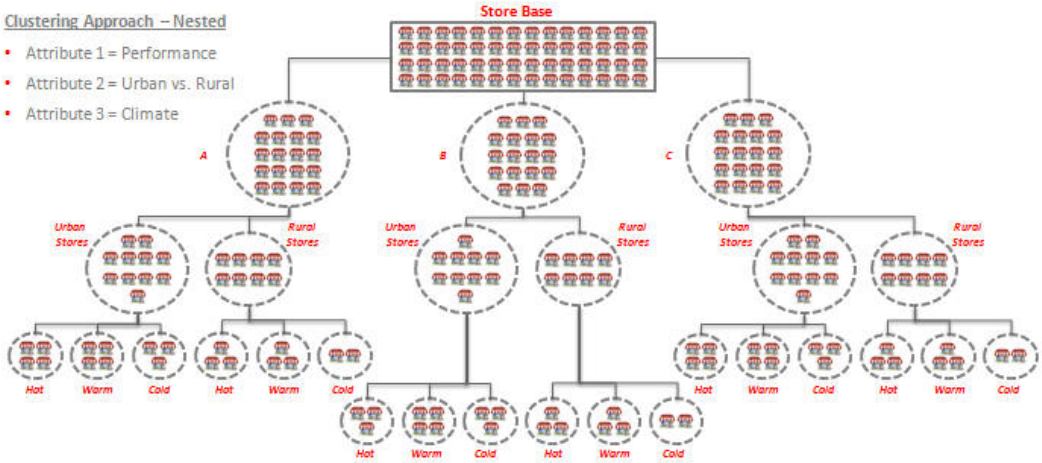
Nested vs. Mixed Attribute

This section describes nested attributes as compared to mixed attributes.

Nested

By default, all Cluster by except mixed attributes can have nested hierarchies. Performance attributes can be further clustered by location attributes, which can be further clustered by location attributes. This approach facilitates dynamic hierarchies in clusters. Nesting can be configured to be enabled or disabled.

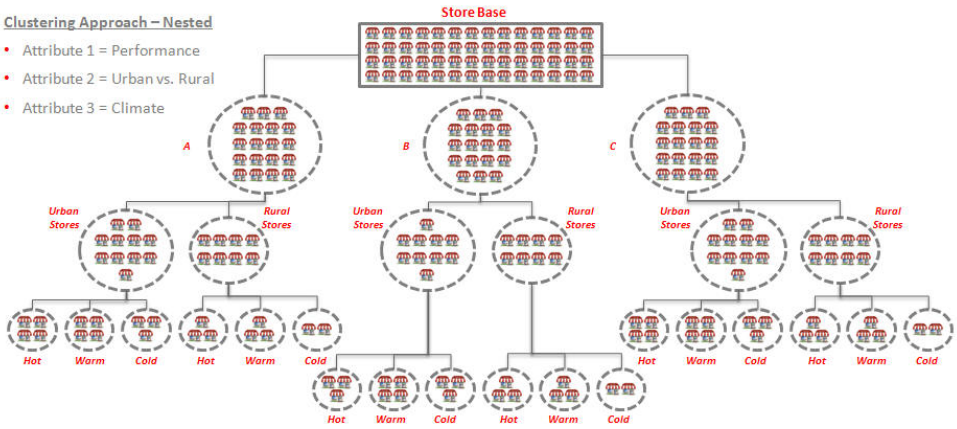
Figure 4-44 Nested Hierarchies



Nested clusters can be created by

- Dynamic hierarchy. Select the Cluster by for each level of a hierarchy while creating a cluster criteria.
- Templates. Select a predefined Cluster by hierarchy while creating a cluster criteria.
- Manual nesting. Create a single cluster criteria, review the results, and then determine whether or not to further cluster. The number of clusters using this approach is granular. The cluster results are hierarchical.

Figure 4-45 Nested Clusters



Mixed Attributes

The following mixed attributes are supported by default: performance, customer segment, location attributes, and product attributes. You can combine attributes from different Cluster by. For example, you can combine attributes from customer segment and performance Cluster by and generate a cluster using sales revenue and customer segment distributions. The number of clusters that are generated using mixed attributes are usually limited, as compared to nested clusters. This approach generates flat clusters with no hierarchy, as attributes are scaled based on the weights you provide.

Figure 4-46 Mixed Attribute Clusters

Figure 4-47 Create Cluster Using Mixed Attributes

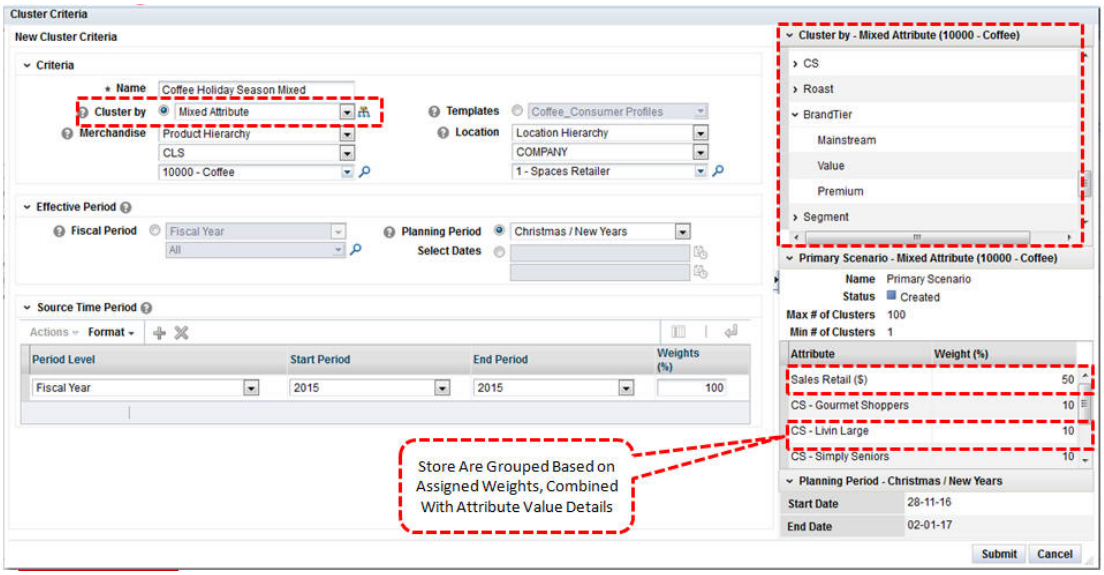
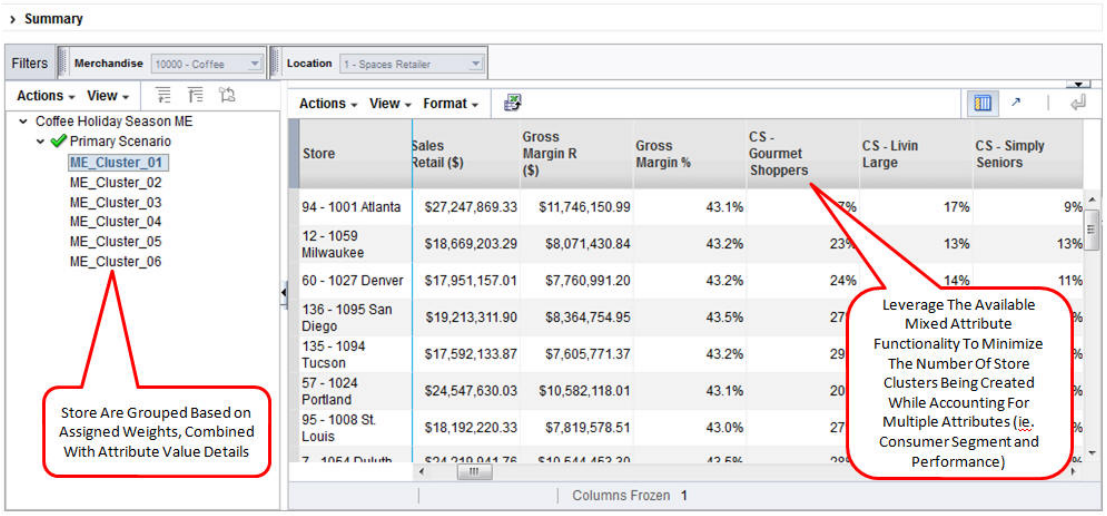


Figure 4-48 Renaming Mixed Attribute Cluster Results



5
Customer Segmentation

This chapter describes the Customer Segmentation module.

Introduction

Customer segmentation is an enterprise-specific solution that uses data mining to group customers based on customer attributes and customer transactions. The retailer can use this information to describe and predict customer behavior. It provides the retailer with a vehicle to target customers with offers, pricing, assortment, and experience.

Retailers understand that shoppers are heterogeneous in nature, that they possess different wants and needs, and that it is impossible to satisfy them all. Retailers can differentiate themselves from their competitors by specializing and offering goods and services that are tailored to one or more market segments.

Customer Segmentation can be used to group customers and to discover hidden customer segments based on the contents of customer shopping baskets and the number of shopping trips they make. Loyalty card data is used to determine if these customer segments differ in terms of socio-demographic or lifestyle characteristics and whether these characteristics can be used to target different customer segments with more relevant product offers.

Retailers can create localized assortments and use customer insights to determine which products to offer by location or channel. This provides insights into the importance of a product to key customer segments and helps when making drop or keep decisions about products within an assortment.

This science-driven tool helps to automate segmentation in a repeatable process and bridges the gap between targeted marketing to targeted assortments.

Features

The key features of Customer Segmentation include:

- Scenario-based segment generation, based on customer attributes, customer behavior, and transactions.
- Attribute importance and correlation mining to identify significant attributes and their associations.
- The ability to generate granular customer segments via the UI. These customer segments include departments such as men, women, and children, health and beauty, or groceries.
- A three-step segment-generation process.
- What-if capabilities that can be used to create multiple segmentation scenarios and then measure them against one another. This can help ensure that the most appropriate segments are used by the applicable planning and execution processes.
- Automatic ranking of segment scenarios to support what-if comparisons.
- Recommendations for the optimal segment scenario and number of segments.

- Descriptions of the main characteristics of the data assigned to each segment using segment rules.
- The use of Oracle data mining hierarchical segmentation and an enhanced version of the k-means algorithm.

Overview of the Customer Segmentation Process

To use Customer Segmentation, follow this general process to create segments, working in the Generate Customer Segment tab:

- Versions – Perform preprocessing actions such as filter and sample a customer's shopping-specific categories or departments. Also identify important attributes that are specific to merchandise and location for customer segmentation.

  Segment Criteria – View all available segments for the specified merchandise, location, and calendar. Review the segment criteria or scenario details for each segment. Use an existing segment as the basis for creating a new segment.
- Explore Data – Examine attributes and their summaries for the segment. Review multiple significant attributes and their correlations using the contextual information.
- Segment Setup – Define multiple what-if scenarios. Such scenarios can be compared with one another throughout the segmentation process.
- Segment Results – View the scenario results and compare scenarios.
- Segment Insights – Gain an understanding about segment results and segment performance prior to approval by examining the contextual information.

Customer Segmentation Overview Tab

The Customer Segmentation Overview tab displays a list of the most recently defined segment criteria and provides the status, segmentation criteria, and applicable merchandise, location, and calendar nodes. You can click on the criteria names in order to access them within the Generate Customer Segments tab.

Figure 5-1 Customer Segmentation Overview Tab

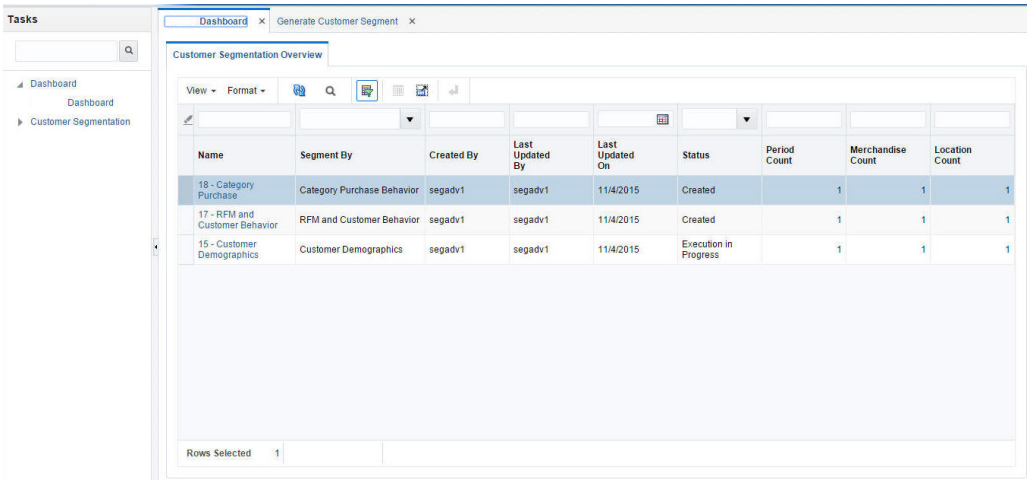


Table 5-1 Cluster Criteria Overview Tab

Field	Description
Name	The criteria ID and user-assigned name of the segment.
Segment By	The Segment By option used for the segment.
Created By	The name of the user who created the segment.
Last Updated By	The name of the user who most recently updated the segment.
Last Updated On	The date when the segment criteria were most recently updated.
Status	The most recent, up-to-date status across the scenarios for the segment criteria. Value include Created, Ready for Preview, Ready for Approval, Approved, and Rejected.
Period Count	The number of calendar nodes defined for the criteria. Hover over the count in order to see a list of the calendar keys associated with the criteria.
Merchandise Count	The number of merchandise nodes defined for the criteria. Hover over the count in order to see a list of the merchandise keys associated with the criteria.
Location Count	The number of location nodes defined for the criteria. Hover over the count in order to see a list of the location keys associated with the criteria.

Segmentation Criteria

The following segmentation criteria are supported by default:

Customer Demographics

This descriptive segmentation technique leverages customer loyalty programs and demographic information (such as residence, profession, age, gender, ethnicity, marital status, and education) about customers to generate demographics based on customer segments.

RFM and Customer Behavior

Segmentation based on purchase behavior aims at discovering groups of customers who exhibit similar purchasing behavior. However, the definition of behavior in this context includes many factors. For example, retailers may want to distinguish between light and heavy users, regular stock-up shoppers versus emergency top-up shoppers, lunchtime shoppers versus evening shoppers, home and daytime shoppers versus work and weekend shoppers, or fast-checkout customers versus regular checkout customers. The two important behavioral dimensions for understanding customer motivations are visit behavior (identified by the time of day and the day of week that the visits take place) and shopping behavior (identified by the customer's spend dispersion across categories purchased during the trip).

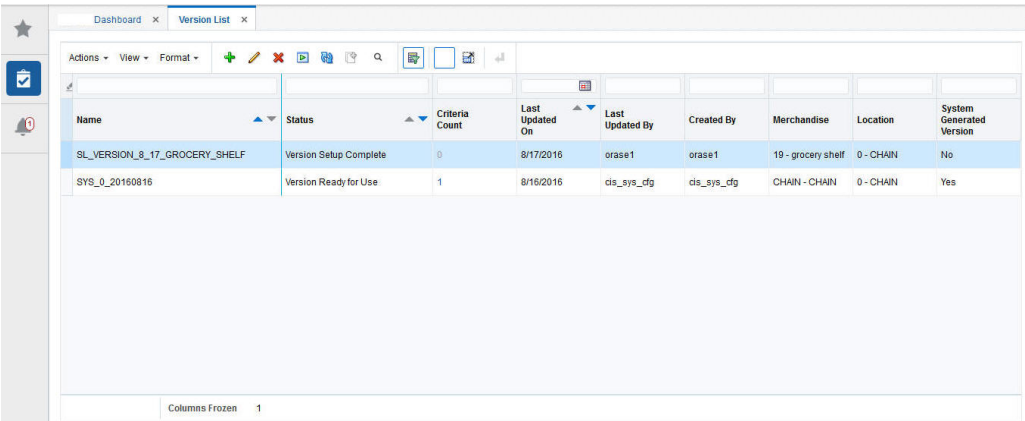
Category Purchase Behavior

Another type of behavior segmentation aims at segmenting the customers who seek similar benefits when evaluating and choosing or purchasing products. These benefits can be measures such as economical price, bulk products, durability, or free shipping. Here, the segment process considers factors that capture customer sensitivity to price and promotions for each category. This can help retailers to segment customers by distinguishing predicted customer responses to the targeted or general promotion of products.

Version List Tab

The Version List tab is used to review, define, and execute preprocessing criteria for customer segments. Version criteria allow users to define versions for merchandise and location combinations. These preprocessed versions can be used to generate customer segments at global or granular levels for merchandise and location. The user can then perform version actions such as creating, updating, and deleting versions and can use selected versions to define new segment criteria.

Figure 5-2 Version List Tab



The Version List, which, by default, lists the global segment version, is defined as part of the batch process. This version is defined for all merchandise and all locations. The version has three key components: filtering, sampling, and attribute mining. It helps when performing the preprocessing of granular customer segments for departments such as men/women/children or segments such as health and beauty, because customers who are shopping for such categories can differ from customers who are shopping for groceries. These segments can be used by applications such as CMPO for planning at the global level or for promotion planning for granular level segments. The version list is sorted by last update, in descending order. The version status is sorted in descending order. The name is sorted in ascending order. Searching, using QBE and a search panel option, is provided for both basic and advanced searches.

Version Actions

The following actions are available.

Figure 5-3 Actions

- Create, edit, and delete a version. If a version has a status of Ready for Use and at least one segment is associated with the version, then edit and delete are not available.
- Execute version. This is available when the setup is complete.
- Create a customer segment using an existing version. This opens the Generate Customer Segment tab. The New Segment Criteria pop-up is displayed, and the default values are filled in. (See the defaults that are selected on selection of the version in segment criteria.) This option is only available if the version has a status of Ready for Use.

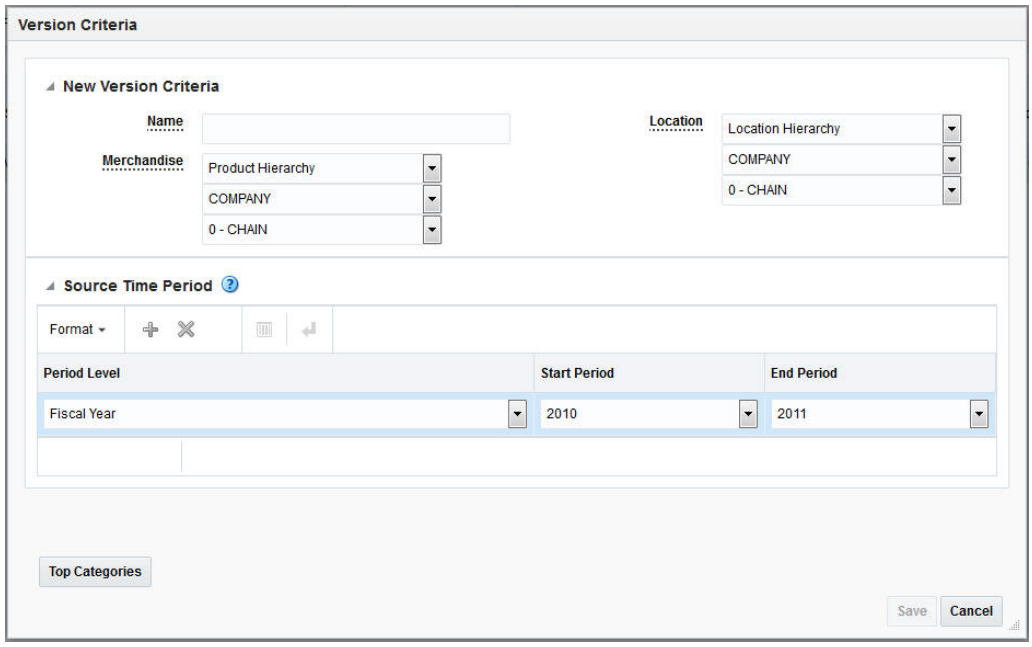
Table 5-2 Version Details

Name	Description
Name	The name of the version.
Criteria Count	The number of criteria associated with the version. When you click the link, a pop-up is displayed that lists the segment criteria details that are associated with the version.
Last Updated By	The name of the user who updated the version most recently.
Last Updated On	The date when the version was last updated.
Created By	The name of the user who created the version.
Status	The current status of the version. Values include Version Setup Complete, Filtering Completed, Sampling Completed Successfully, Attribute Mining Completed Successfully, Version Ready for Use.
Merchandise	The merchandise node defined for the version.
Location	The location node defined for the version.
System Generated Version	This column provides a flag indicating whether or not the version has been created using a batch process and is set up with a default configuration.

Version Criteria Pop-up

This section describes the Version Criteria pop-up, shown in Figure 5-4.

Figure 5-4 Version Criteria Pop-Up



Merchandise

Once you choose the merchandise level for the version, you must select the hierarchy node. The merchandise defined as part of the version is then used for creating segments.

Location

Once you choose the location level for the version, you must select the hierarchy node. These are specific to the location level you select. The location defined as part of the version is then used for creating segments.

Source Time Period

Select historical sales data for the version. You can specify more than one time period. Source time periods selected as part of the version are then applicable for the segments.

Table 5-3 Source Time Period

Time Period	Description
Period Level	Select from Fiscal Year, Fiscal Quarter, Fiscal Period, or Fiscal Week.
Start Period	Once you select the Period Level, select the starting subdivision within that period.
End Period	Once you select the Start Period, select the End Period.

Top Categories

Use this pop-up to select top categories by reviewing category KPIs, such as Sales Share, Revenue, and Margin. The system selects the top categories by each department. You can override these top categories further while defining versions for merchandise. By default, the top categories in the selected merchandise are driven based on the sales share for each category. You can remove categories to reduce processing time and storage, using the user interface and adding emerging to the top category list.

Version Status

The version status has one of the following values.

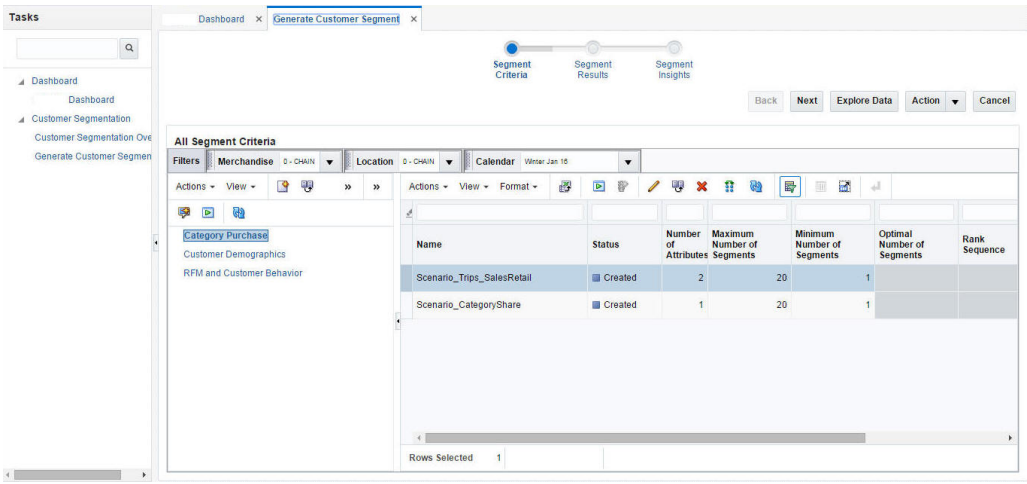
Table 5-4 Version Status

Status	Description
Version Setup Complete	Indicates that version setup is complete.
Filtering Completed with Errors	Indicates that version filtering has completed with errors.
Filtering Execution in Progress	Indicates that version filtering is in progress.
Filtering Completed Successfully	Indicates that version filtering has completed successfully.
Sampling Completed with Errors	Indicates that sampling has completed with errors.
Sampling Execution in Progress	Indicates that sampling execution is in progress.
Sampling Completed Successfully	Indicates that sampling has completed successfully.
Attribute Mining Completed with Errors	Indicates that attribute mining has completed with errors.
Attribute Mining Execution in Progress	Indicates that attribute mining execution is in progress.
Attribute Mining Completed Successfully	Indicates that attribute mining has completed successfully.
Version Ready for Use	Indicates that all three execution phases (filtering, sampling, and attribute mining) have completed successfully. Once the version is ready for use, it is ready for the creation of segment criteria.

Generate Customer Segments Tab

The Generate Customer Segments tab is used to create segments and then model the segments with various scenarios in order to determine the best set of segments. It consists of three stages: Segment Criteria, Segment Results, and Segment Insights.

Figure 5-5 Generate Customer Segments Tab



Segment Criteria Stage

Use the Segment Criteria stage to view summary data about existing segments and define the characteristics of new segments.

Process

Here is the high-level process for defining a segment.

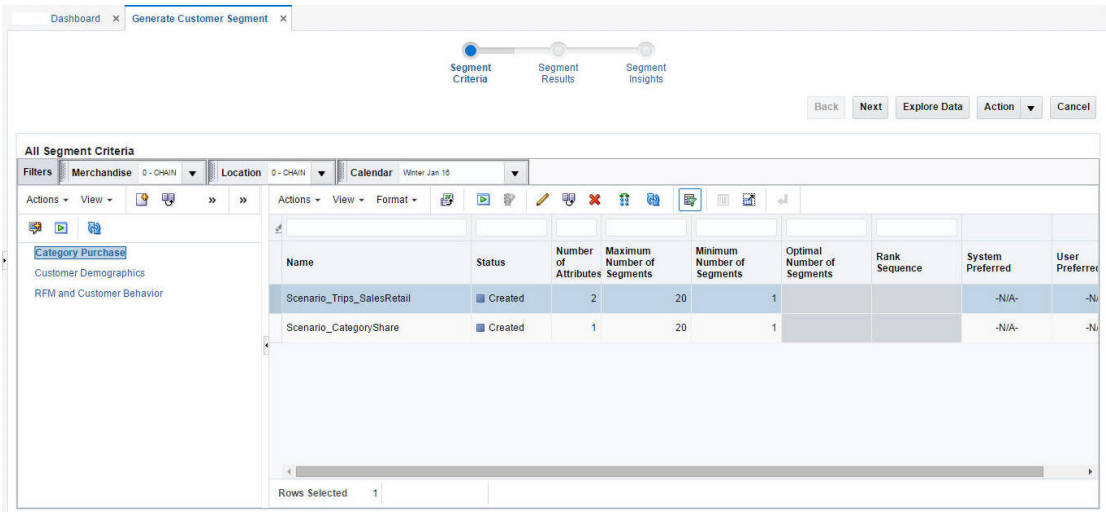
1. Provide a unique name for the segment.
2. Define the type of data used to characterize the segment.
3. Select the merchandise and location nodes.
4. Define the time period for the segment.
5. Define the historical time period for the data.

All Segment Criteria

In this area of the page you can view information about existing segments.

Use the View list to select existing segment criteria. You can tailor your search for existing segment criteria by Merchandise, Location, and Calendar. Once you select the segment criteria, the defining details for the segment criteria are displayed in the Worksheet area.

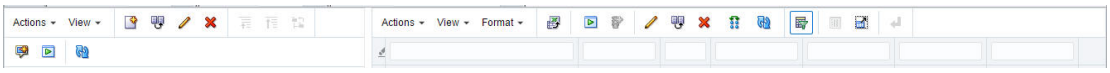
Figure 5-6 All Segment Criteria



Toolbar Buttons

The Toolbar buttons are shown in Figure 5-7.

Figure 5-7 Toolbar



You can use the Toolbar buttons to:

- Create segment criteria. You can define an initial segment. The criteria include segment name, the merchandise for the segment, the effective date for the segment, and the history to use.
- View or edit segment criteria. The characteristics of the segment are displayed in the Segment Criteria pop-up.
- Copy segment criteria. Once you have copied the criteria, you can modify them.
- Delete segment criteria. Delete the selected segment criteria.
- Execute segment criteria. Execute all non-executed scenarios for the selected criteria.
- Evaluate KPIs after the execution of the scenario is complete.
- Refresh all. Refresh all segment criteria in order to view any updates to the existing segment criteria.

All Segment Criteria Filter

Once you have highlighted the segment criteria to examine, details about the segment criteria are displayed in a pop-up.

Figure 5-8 All Segment Criteria Filter

The details include information about the segment criteria and the scenarios created for segment criteria.

- Merchandise allows the user to filter the existing segment criteria by searching for or selecting the merchandise for the supported hierarchies.
- Location allows the user to filter the existing segment criteria by searching for or selecting the location for the supported hierarchies.
- Calendar allows the user to filter the existing segment criteria by searching for or selecting the calendar for the supported hierarchies.

All Segment Criteria Summary

Once you have highlighted the segment criteria to examine, details about the segment criteria are displayed in a pop-up.

Figure 5-9 Segment Criteria Details

The details include information about the segment criteria and the scenarios created for the segment criteria.

Table 5-5 Pop-Up Details

Field Name	Description
Name	The name of the segment criteria.
Shared Criteria	A check mark indicates that more than one merchandise or location node is used in the segment criteria.
Merchandise Type	The merchandise type.
Location Type	The location type.
Segment By	A set of attributes to be used in the creation of the segment criteria: Customer Demographics, RFM and Customer Behavior, Category Purchase Driven.
Scenario Created	The number of scenarios created for the segment.

Table 5-5 (Cont.) Pop-Up Details

Field Name	Description
Scenario Executed	The number of scenarios executed for the segment.

All Segment Criteria Scenario List

This displays the scenarios for the selected segment criteria in the All Segment Criteria tree.

Table 5-6 Scenario List

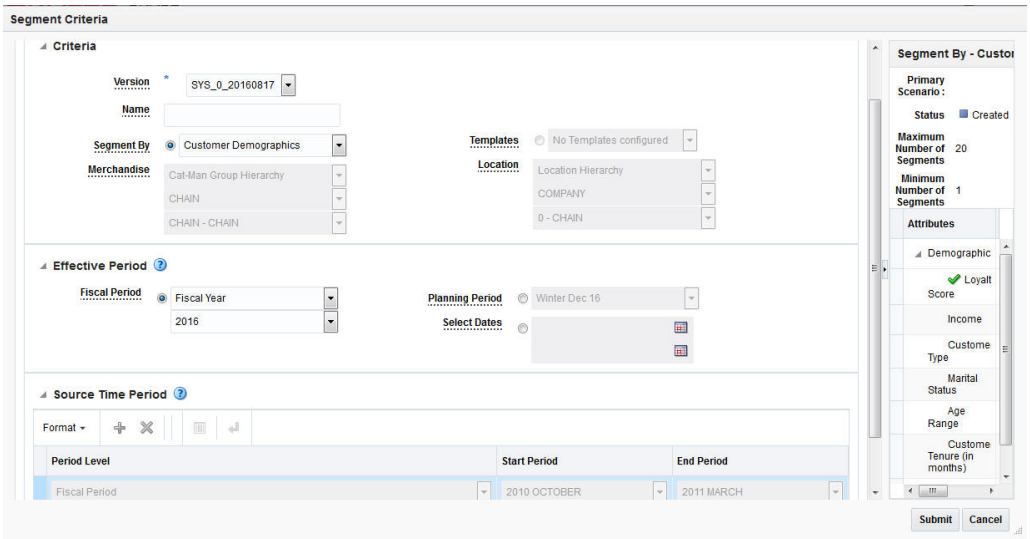
Field Name	Description
Name	The name assigned to each scenario that has been created for the segment.
Status	Created, Ready for Preview, Ready for Approval, Completed with Errors, Approved, Rejected.
Rank	The system-calculated rank for the segment.
Optimal # of Segments	The system-calculated optimal number of segment centers.
User Preferred	Indicates whether or not the user prefers the segment.
System Preferred	Indicates whether or not the system prefers the segment.
# of Attributes	The number of attributes that were used in the segment.
Max. # of Segments	A user-provided value for the maximum segment centers that the segmenting process should consider.
Min. # of Segments	A user-provided value for the minimum segments centers that the segmenting process should consider.

Segment Criteria

In this pop-up, you define the initial segmenting parameters for the segment criteria of a new segment. Note that the hierarchy type supported can be configured at the time of deployment.

Figure 5-10 illustrates how to use a simple approach to segmenting by selecting attributes from a Segment by. For example, you can select the RFM and Customer Behavior Segment by and generate segments using the total number of trips and the amount spent by customer.

Figure 5-10 New Segment Criteria Pop-Up



The following information defines a segment:

Table 5-7 New Segment Definition

Field Name	Description
Name	A unique name to identify the segment.
Version	Select a predefined version to use for creating segment criteria. When the version is selected, the merchandise and location are selected, based on the version.
Segment By	A predefined group of attributes that include Customer Demographics, RFM and Customer Behavior, Category Purchase Driven. These criteria types are sets of attributes. For example, customer demographics are the properties of a customer. These properties can include ethnicity, income, and age.
Merchandise	Merchandise is selected based on the version selected for the segment criteria.
Location	Location is selected based on the version selected for the segment criteria.
Template	Template is selected based on the version selected for the segment criteria.

Effective Period

You can define a time interval for the segment by either choosing a period from the list provided or by selecting a start date and an end date.

To define the Effective Period, you select either Planning Period, Fiscal Period, or Select Date.

Table 5-8 Effective Period

Option	Description
Fiscal Period	If you select this option, choose the period and the subdivisions of that period from the drop-down lists.
Planning Period	Select from the range of values provided for the period. Planning periods are user-defined buying periods for a season or a season subset.
Select Dates	If you select this option, choose the start and end dates using the calendar pop-up.

Summarization

Data summarization is available by default and set to either Category or Sub Category. It is applied to Category Purchase Driven Segment by. The segmentation process considers the top selected categories and their attributes and groups customers based on their sales patterns.

Source Time Period

Source time periods are selected based on the version selected for the segment criteria.

Table 5-9 Source Time Period

Field	Description
Period Level	Select from Fiscal Year, Fiscal Quarter, Fiscal Period, or Fiscal Week.
Start Period	Once you select the Period Level, you select the starting subdivision within that period.
End Period	Once you select the Period Level, you select the ending subdivision within that period.

Contextual Area

When you are creating new segment criteria, you can see details about the following parameters that can help you understand the segment you are creating.

Segment By Hierarchy

The following information is displayed when you select a template or use the icon to select the Segment by.

Table 5-10 Template Display

Property	Description
Template Name	Name of template configured during deployment.
Description	Description of template.

Table 5-10 (Cont.) Template Display

Property	Description
Segment By	A predefined group of attributes that include Customer Demographics, RFM and Customer Behavior, Category Purchase Driven. These criteria types are sets of attributes. For example, customer demographics are the properties of a customer. These properties can include ethnicity, income, and age.

Segment By Primary Scenario

You see this when you select Segment by in the Criteria panel when you are setting the segment criteria parameters or when you select Segment by in the contextual area for the hierarchy.

The system displays the primary scenario, its preconfigured properties, and the significant attributes identified during the attribute importance process for each segment by.

The following information is displayed.

Table 5-11 Primary Scenario

Property	Description
Name	The name of the primary scenario.
Status	Created, Ready for Preview, Ready for Approval, Completed with Errors, Approved, Rejected.
Maximum # segments	The maximum number of segments. The default value is 20. This is used for analyzing the segments.
Minimum # segments	The minimum number of segments. The default value is 1. This is used for analyzing the segments.
Attribute	A list of the attributes configured during segmentation.
Attribute importance	The weighted average across attributes and importance index for each attribute.

Planning Period

This list displays the time period you selected for the segment criteria. This information is available only for planning periods, where it provides the start and end dates of the planning period. This content changes whenever the planning period is selected in Effective Period when you are setting segment criteria parameters.

Explore Data

Use the Explore Data pop-up to examine data for the segment criteria you defined. You can view the customer and attribute summary that provides input into the segmentation process.

Process

In this pop-up you can view criteria and attribute summaries as well as their related contextual BIs.

Summary

The Summary lists the criteria you initially selected to define the segment.

Table 5-12 Explore Data: Summary

Field	Description
Name	The name you provided for the segment in the Segment Criteria stage.
Segment By	A predefined group of attributes that include Customer Demographics, RFM and Customer Behavior, Category Purchase Driven. These criteria types are sets of attributes. For example, customer demographics are the properties of a customer. These properties can include ethnicity, income, and age.
Merchandise	The merchandise level and nodes for the segment.
Location	The location level and nodes for the segment.
Fiscal Period	The time period for the segment.
Merchandise Hierarchy Type	Details about which type of hierarchy the segment criteria have been created for.

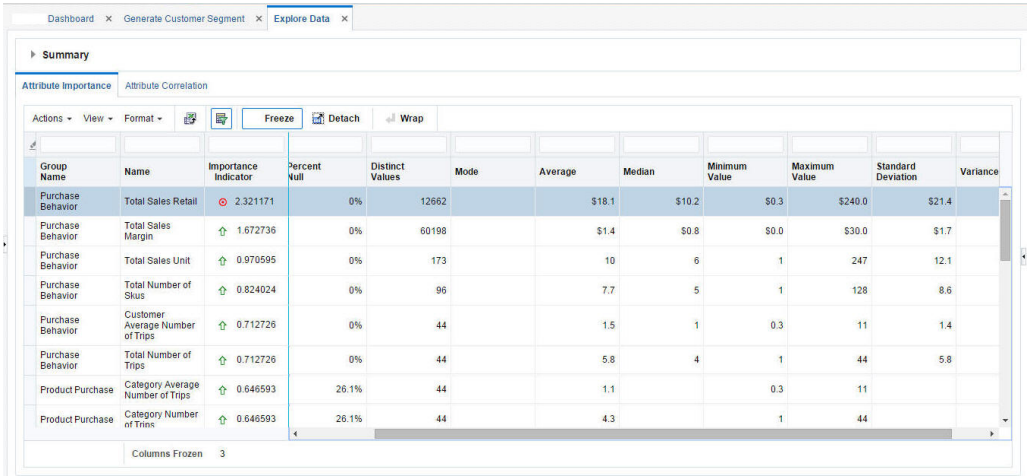
Attribute Mining

This screen provides you with insights about the attribute mining process, which lists the attributes' significance and their correlations. The attributes required for the customer segmentation process come from different dimensions such as customers, their households, demographics, and purchasing behaviors. This process helps to eliminate redundant attributes and to identify the attributes that may have the most influence on generating customer segment. The attribute summary displays information about data availability and data quality by providing the attributes' distinct values, percentage of nulls, and statistics summaries, such as mean, median, and standard deviation.

Attribute Importance

Along with attribute summaries, the system generates an attribute importance index that sums up data quality, data distribution, and its representation of each attribute in the data. See Figure 5-11.

Figure 5-11 Attribute Importance



Group Name	Name	Importance Indicator	Percent Null	Distinct Values	Mode	Average	Median	Minimum Value	Maximum Value	Standard Deviation	Variance
Purchase Behavior	Total Sales Retail	2.321171	0%	12662		$18.1	$10.2	$0.3	$240.0	$21.4	
Purchase Behavior	Total Sales Margin	1.672736	0%	60198		$1.4	$0.8	$0.0	$30.0	$1.7	
Purchase Behavior	Total Sales Unit	0.970595	0%	173		10	6	1	247	12.1	
Purchase Behavior	Total Number of Skus	0.824024	0%	96		7.7	5	1	128	8.6	
Purchase Behavior	Customer Average Number of Trips	0.712726	0%	44		1.5	1	0.3	11	1.4	
Purchase Behavior	Total Number of Trips	0.712726	0%	44		5.8	4	1	44	5.8	
Product Purchase	Category Average Number of Trips	0.646593	26.1%	44		1.1		0.3	11		
Product Purchase	Category Number of Trips	0.646593	26.1%	44		4.3		1	44		

The system indicates the relative importance of attributes, based on upper and lower threshold values.

Table 5-13 Attribute Importance Values

Category	Range
Low Importance	Minimum value for the lower threshold, indicated in red.
Average Importance	Values between the lower threshold and the upper threshold, indicated in yellow.
High Importance	Values above the upper threshold are indicated in green and are the best candidates for mining attributes.

The following attributes are displayed in Attribute Importance.

Table 5-14 Attributes

Attribute Name	Description
Group Name	The attribute group name (attributes such as demographics, purchase behavior, or product profile).
Name	The name of the attribute (for example, income, ethnicity, or total sales retail)
Histogram	The spark chart for displaying the attribute distribution. This corresponds to the current data distribution graph.
Importance Indicator	The attribute importance index with an image to indicate if attribute has high, average, or low importance.
Percent Null	The percentage of data that is null for the attribute.
Distinct Value	The distinct value applicable to discrete attributes. If no value is available, this is empty.
Mode	The most common value of the discrete attributes. If no value is available, this is empty.

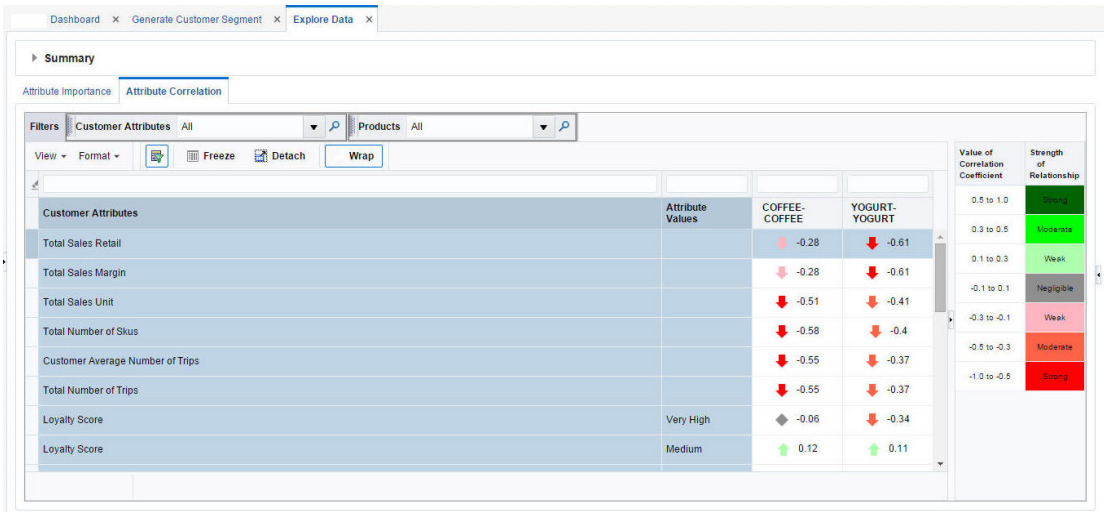
Table 5-14 (Cont.) Attributes

Attribute Name	Description
Average	The mean value of the numeric attributes. If no value is available, this is empty.
Median	The median value of the numeric attributes. If no value is available, this is empty.
Minimum Value	The minimum value of the numeric attributes. If no value is available, this is empty.
Maximum Value	The maximum value of the numeric attributes. If no value is available, this is empty.
Standard Deviation	The standard deviation indicates the deviation from the average for the numeric attributes. If no value is available, this is empty.
Variance	The variance indicates the dispersion from the average for the numeric attributes. If no value is available, this is empty.

Attribute Correlation

The attribute mining process generates a correlation between customer attributes and products. This information can help you understand which combination of customer attributes may drive product purchase behavior.

Figure 5-12 Attribute Correlation



Filter

You can use customer and product filters to reduce the number of customer and product attributes that are displayed in the correlation matrix.

Figure 5-13 Customer and Product Filter

Correlation Matrix

The correlation matrix displays attributes, attributes values (in the case of discrete attributes), and correlations between customer and products. Each cell in the matrix provides a visual indication of how attributes are correlated, along with the correlation value. Various colors indicate the strength of the correlation among attributes, with -1 and 1 indicating strong association.

Table 5-15 shows the strength of the correlation for the different ranges of values for the correlation coefficient.

Table 5-15 Correlation Coefficient Values

Correlation Coefficient Value	Strength of Relationship
1.0 to 0.5	Strong
0.3 to 0.5	Moderate
0.1 to 0.3	Weak
-0.1 to 0.1	None or very weak
-0.3 to -0.1	Weak
-0.5 to -0.3	Moderate
-1.0 to -0.5	Strong

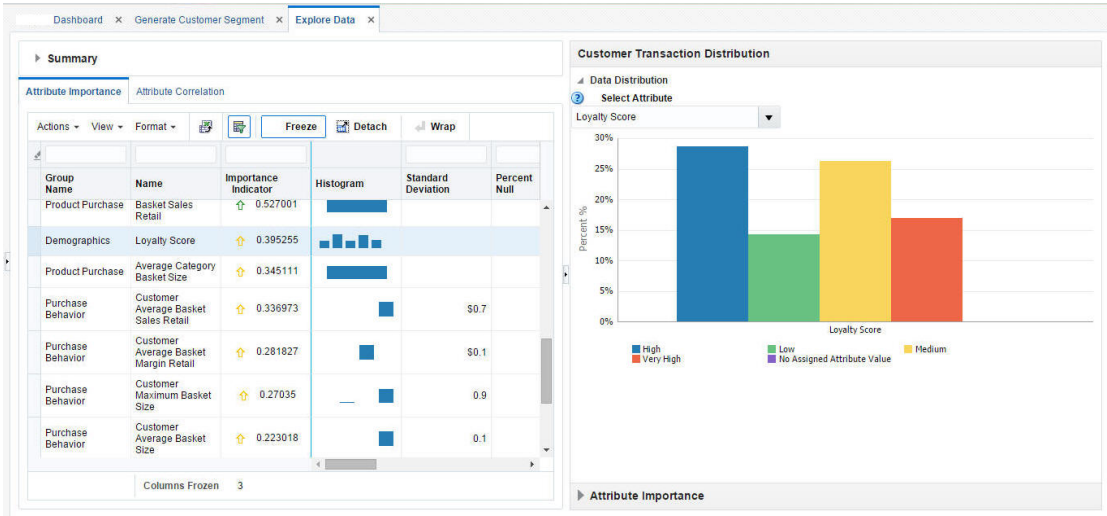
Contextual Area

This area provides graphical illustrations of the detailed data distribution about the customers and their attribute importance.

Analyze Customers

In Explore Data, the BI displays the data distribution of the customers by each participating attribute as well as other configured informational attributes. Customer Segmentation identifies the bins based on the underlying sample data and displays the histograms. It provides the percentage of customers that are present in a selected location. For example, a company may have ten percent of premium customers who are high spenders and who shop frequently.

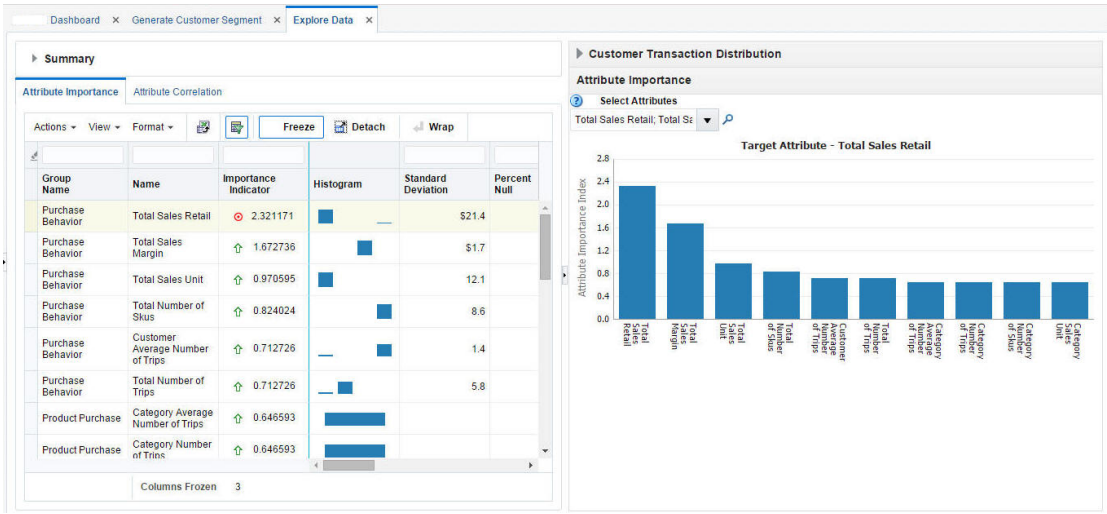
Figure 5-14 Analyze Customers



Attribute Importance

The attributes are ranked in order of their significance in predicting a target (the default is set to Customer Total Spent) and by analyzing the quality and variability of the data. The results in Figure 5-15 shows that loyalty score and income have the most effect on whether or not the customer spends more shopping at the retailer.

Figure 5-15 Attribute Importance

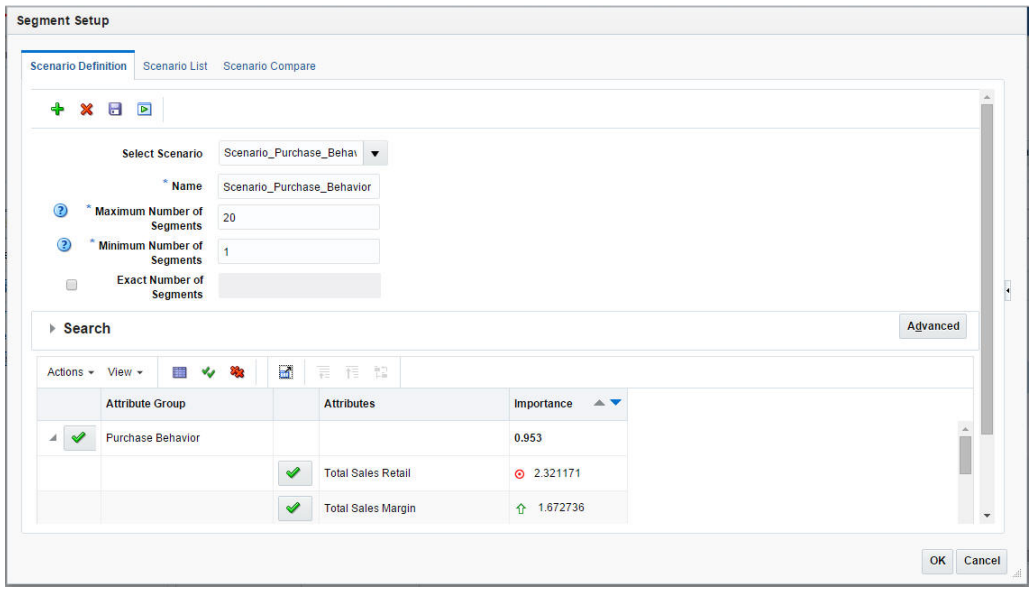


Segment Setup Stage

You can use this stage to perform what-if analysis by defining one or more scenarios that are based on a specified number of segments and attributes. You can select one or more attributes based on the attribute importance index. The attribute importance index guides you when you are defining the scenario setup. Using these scenarios, you can experiment with different numbers of segments and participating attributes. You can either define the

maximum number or the minimum number of segments or alternatively define a specific number of segments that you want to be generated. Once the scenarios are generated, different scenarios can be compared. You can also use other features in this stage to copy or delete scenarios.

Figure 5-16 Segment Setup



Process

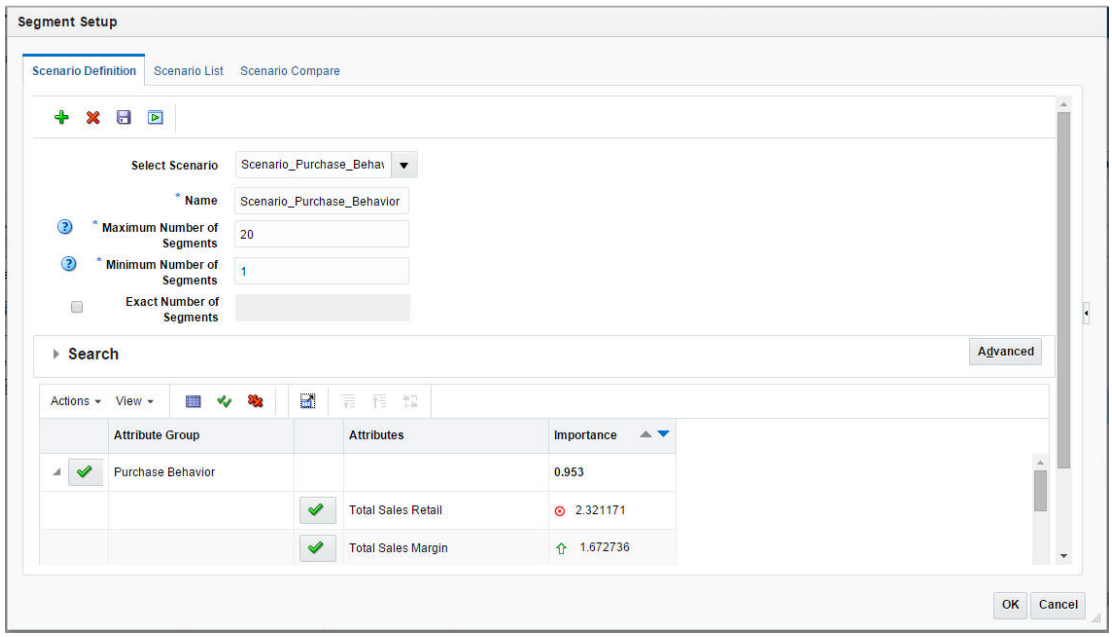
Here is the high-level process for setting up scenarios.

1. Either select the name of a scenario you want to modify or enter the name for the new segment criteria you want to create.
2. If you want the application to optimize the number of segments, enter the minimum and maximum values for the number of segments.
3. If you want the application to generate a specific number of segments, enter that value. In this case, the application generates the exact number of segments and provides the optimal number of segments as informational data.
4. Optionally, review the attribute importance assigned to the attributes and select an attribute if you want a specific attribute to be part of the segmentation process.
5. Click the **Execute** icon to execute the scenario. Once the processing is complete, you see the results in the Segment Results stage.
6. To see a list of all scenarios and the status for each, go to the Scenario List tab.

Scenario Definition Tab

This section describes the Scenario Definition tab.

Figure 5-17 Scenario Definition Tab



The following information defines a scenario.

Table 5-16 Scenario Definition

Field Name	Description
Select Scenario	Select an existing scenario if you want to modify it.
Name	A unique name that identifies the scenario being defined.
Max. # of Segments	Set the maximum number for the total number of segments that can be generated. The application determines the optimal number of segments during the generation process.
Min. # of Segments	Set the minimum number for the total number of segments that can be generated. The application determines the optimal number of segments during the generation process.
Exact # of Segments	Indicates that the exact number of segments should be generated. The application does not determine the optimal number of segments.

Attributes

The Attributes table is used to define which attributes are included in the segment criteria. You can

- search by attribute, attribute value, or attribute importance
- select and deselect attributes

The following information defines the attributes that are participating or non-participating.

Table 5-17 Attributes

Field Name	Description
Participating	A check in this column indicates that the attribute participates in the segment criteria.
Attribute Group	A logical grouping of attributes such as demographics or purchase behavior.
Attributes	Attributes that are potential candidates for generating segments.
Importance	System-generated attribute importance index that indicates the significance of each attribute.

The Attributes toolbar includes the following functionality:

Figure 5-18 Attribute Toolbar

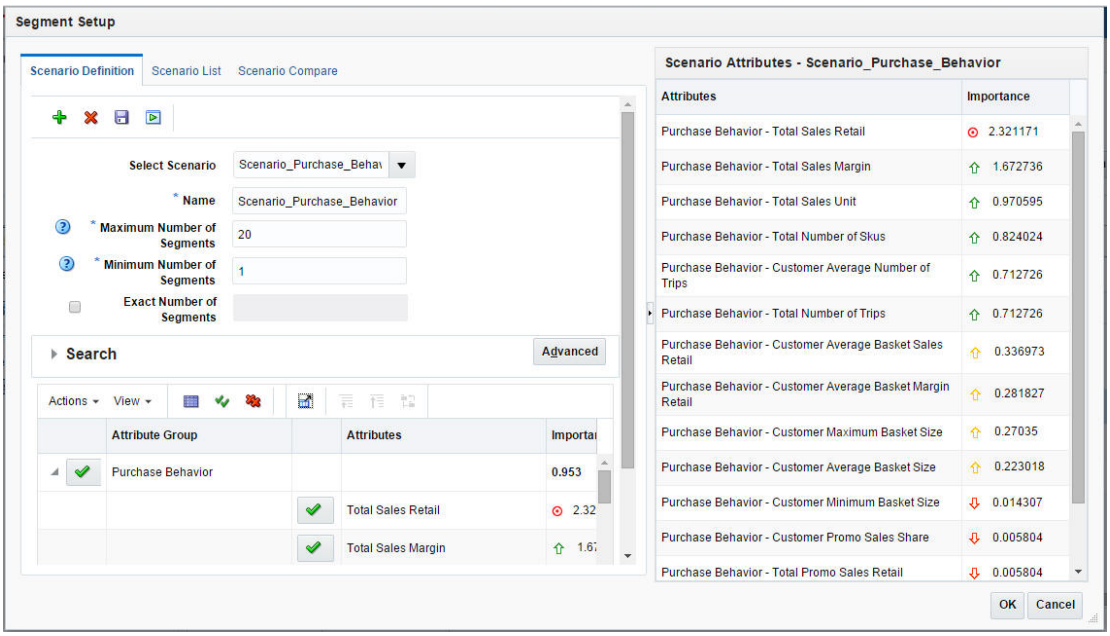
Table 5-18 Attribute Toolbar

Function	Description
Action menu	Resets the attribute selection to the default selection that system identified using attribute importance thresholds.
Include or exclude attributes	Any attribute beyond a certain threshold is not included in the segmentation process.

Contextual Area

The contextual business intelligence lists a set of attributes that the current scenario includes as the participating attributes for the segmentation process.

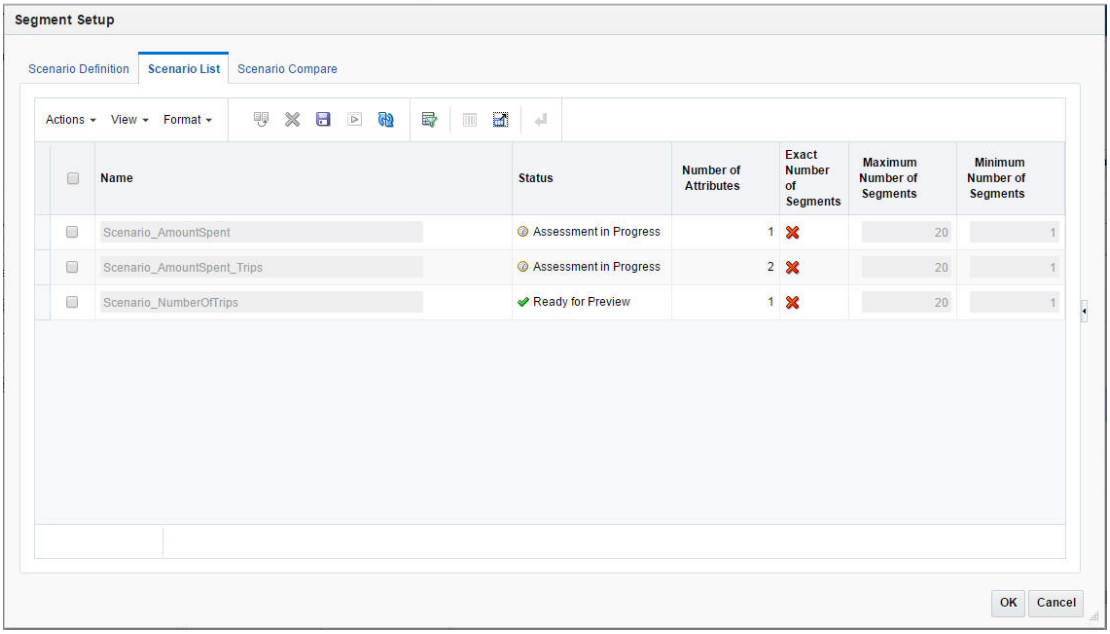
Figure 5-19 Scenario Definition: Contextual Area



Scenario List Tab

The Scenario List summarizes the characteristics for each scenario.

Figure 5-20 Scenario List Tab



You can make a copy of a specific scenario in order to modify it in some way, delete a specific scenario, execute a specific scenario, or save a specific scenario. You can also refresh the scenario list in order to view the updated scenario status.

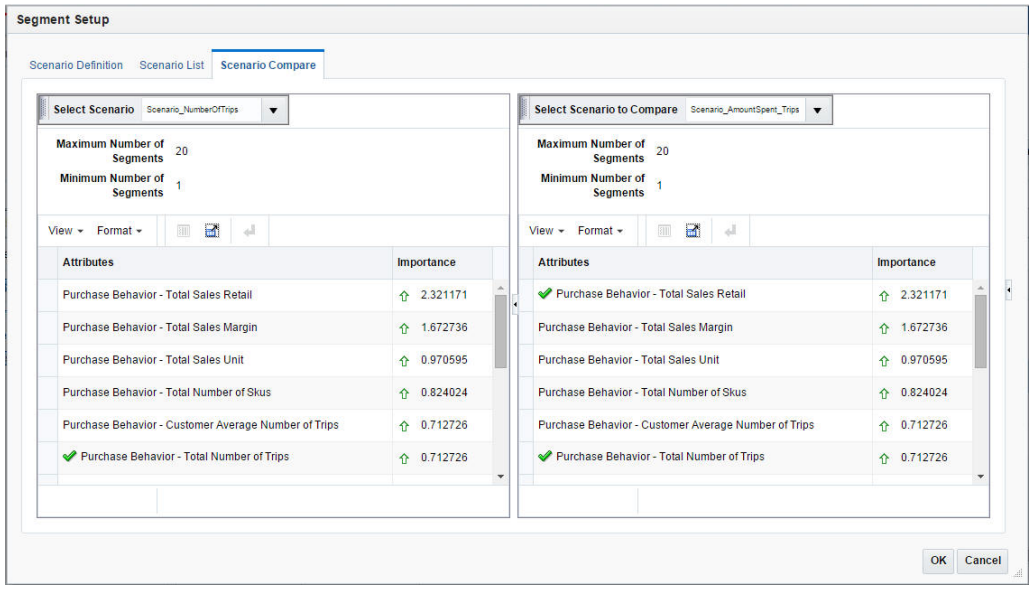
Table 5-19 Scenario List

Field Name	Description
Name	The unique name that identifies the scenario.
Status	These include Created, Ready for Preview, Ready for Approval, Completed with Errors, Approved, Rejected.
# of Attributes	The number of attributes is defined by the Segment by option you select.
Max. # of Segments	If you provided a value for this in the scenario definition, that number is displayed here.
Min. # of Segments	If you provided a value for this in the scenario definition, that number is displayed here.

Scenario Compare Tab

You can select two scenarios from the list to compare. The scenarios you select from the Scenario list are shown side-by-side to facilitate this.

Figure 5-21 Scenario Compare Tab



Segment Results Stage

After you select a scenario and execute it, you can see the results in this stage. The application uses the data and the parameters you defined in order to group customers together that are most similar according to the characteristics you selected and to separate customers that are most dissimilar. You can also use this stage to rename a segment.

Process

You use this stage to review segments and their composition, using the grid view and the graph view. This includes

- Review a segment to see the goodness of fit by using the scores. Determine if any segments are outliers that warrant further analysis.
- Rank the scenarios (segment sets) to see how well they are separated and how compact the customers are within each segment.
- View the optimality of the segments recommended by the application to determine if increasing the number of segments beyond the optimal number is significant.
- Rename the segment after analyzing the centroids and before the segment is approved.

Summary

This lists the criteria you initially selected to define the segment.

Figure 5-22 Results Summary

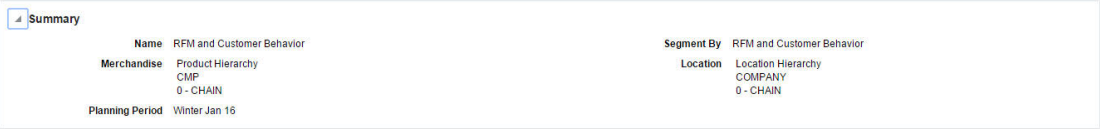


Table 5-20 Segment Criteria Summary

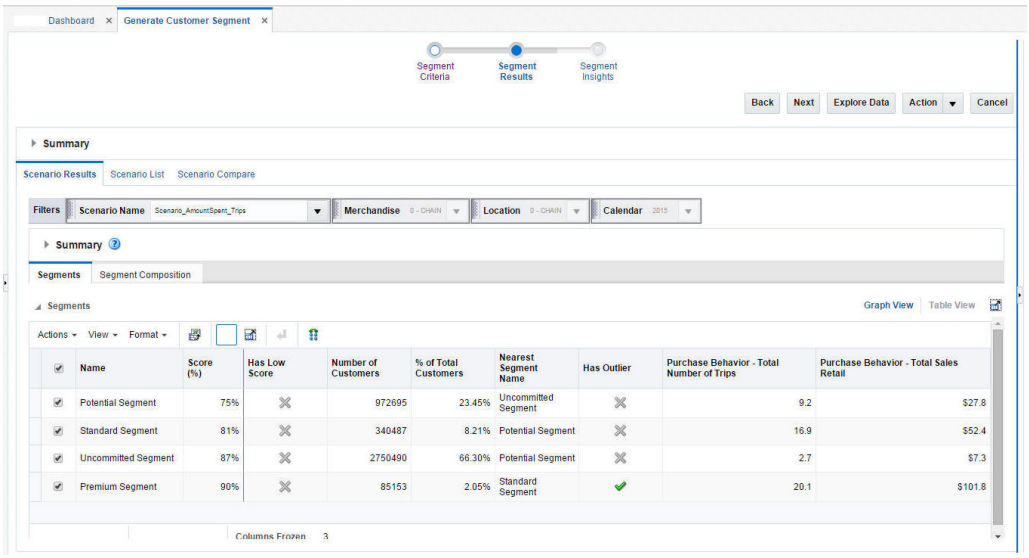
Field	Description
Name	The name you provided for the segment in the Segment Criteria stage.
Segment By	A predefined group of attributes that include Customer Demographics, RFM and Customer Behavior, and Category Purchase Driven. These criteria types are sets of attributes. For example, customer demographics are the properties of a customer. These properties can include ethnicity, income, and age.
Merchandise	The merchandise level and nodes for the segment.
Location	The location level and nodes for the segment.
Fiscal Period	The time period for the segment.
Merchandise Hierarchy Type	Provides details about which type of hierarchy the segment criteria have been created for.

Scenario Results Section

The Scenario Results section displays the following:

- The Scenario Summary, which provides key segment set attributes for the executed scenario as well as its status.
- The Scenario Results, which has two tabs: Segments and Segment Composition.

Figure 5-23 Customer Segmentation Scenario Results



The Summary section provides an overview of the characteristics of the scenario.

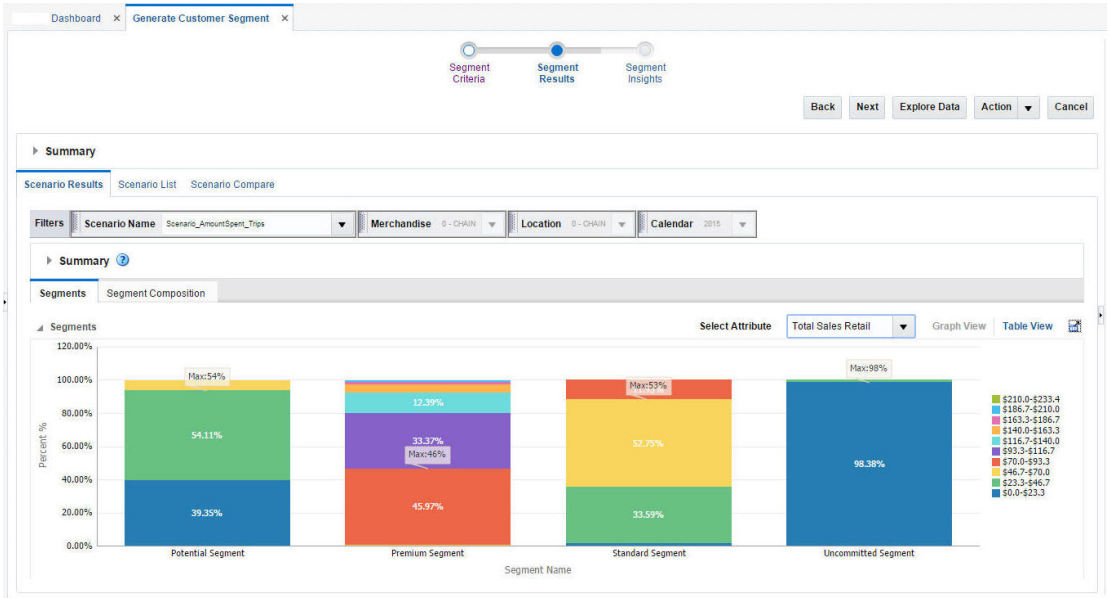
Table 5-21 Scenario Results Summary

Field	Description
Status	Ready for Preview, Ready for Approval, Approved, Rejected.
Optimal # of Segments	The optimal number of segments determined by the optimization.
Rank	The application compares executed scenarios and ranks them. A value of 1 indicates the best scenario.
Max./Min # of Segments	The number you provided for the maximum and minimum number of segments to calculated.
Largest/Smallest Segment	Provides the sizes of the largest segment and the smallest segment in order to show the range of values.
Is System Preferred	Indicates whether or not the system prefers the scenario.
Is User Preferred	Indicates whether or not the user prefers the scenario.

The Segments section provides the segment results for each individual segment in the scenario in either a Graph View or a Table View. The attributes displayed depend on the Segment by option chosen in the Segment Criteria stage.

The Graph View shows the percentage for each attribute in the segment.

Figure 5-24 Segments Graph View



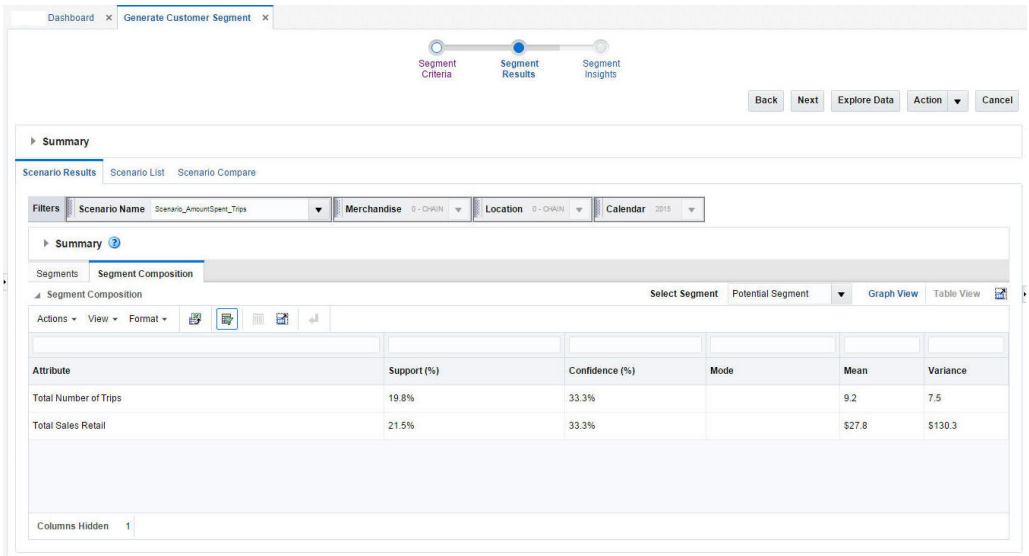
The Table View provides details that can help you analyze the segment.

Table 5-22 Scenario Results - Segments: Table View

Field	Description
Name	The name you assigned to the segment.
# of Customers	The number of customers in the segment.
% of Total Customers	The percentage of the total customers that the number of customers represents.
Nearest Segment Name	The name of the segment that is most similar to this segment.
Score %	This value is calculated at the level of customer and then averaged to the segment. The probability, expressed as a percent, of a customer being present in this segment rather than any of the other segments. See Segment Scores.
Has Outlier	Indicates a segment with the number of customers below a threshold. For example, the number of customers are below certain percentage of the number of customers in a segment. See Outlier Indicator.

The Segment Composition sub-tab breaks down the segment into its component parts and shows the percentages for each attribute.

Figure 5-25 Segment Composition



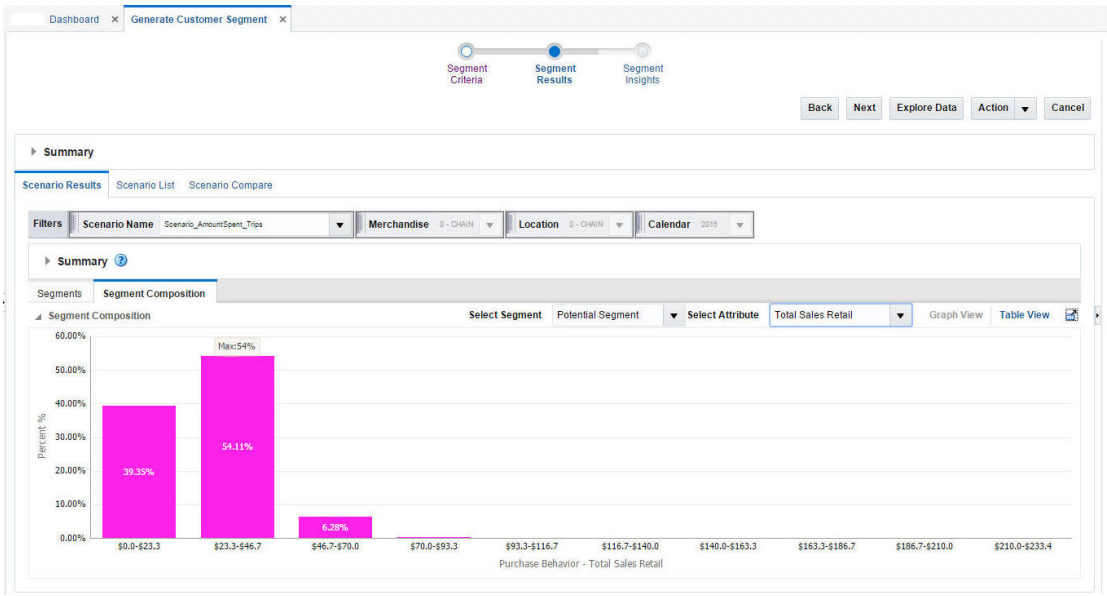
The Table View shows attributes, their confidence, and summary statistics within each segment.

Table 5-23 Segments Composition Table View

Field	Description
Attribute	The name of the attribute that was used in generating the segment.
Category	The category name, which is only available for Category Purchase Driven segment by.
Support	The percentage of the customers that are concentrated within the segment according to the attribute.
Confidence	Measures the goodness of match between the distribution of the data points for the attribute, indicating the power in recognizing a new pattern in data. An attribute with a higher confidence number plays a more important role in that segment.
Mean	The average value of the numeric attribute within the selected segment. If no value is available, this is empty.
Mode	The most common value of the discrete attributes within the selected segment. If no value is available, this is empty.
Variance	This explains the dispersion from the average for the numeric attributes within the selected segment. If no value is available, this is empty.

The Graph View shows the segment composition by attribute.

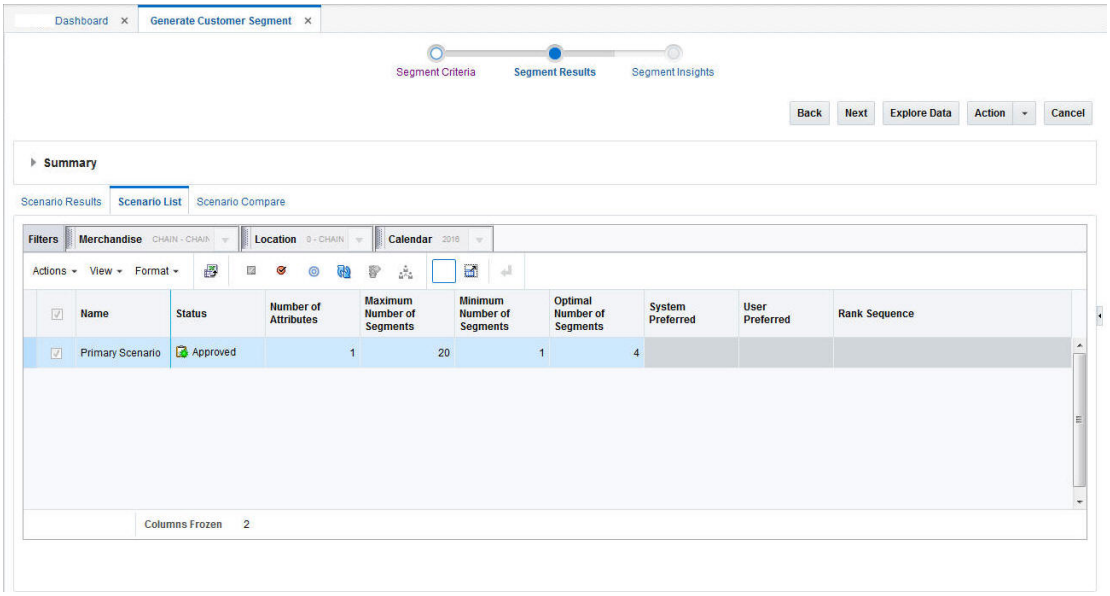
Figure 5-26 Segment Composition Graph View



Scenario List

The Scenario List section contains one table with details about each segment. Here, you can evaluate KPIs and approve or reject a segment.

Figure 5-27 Segment Results Scenario List



After a segment is approved, it is available for other applications.

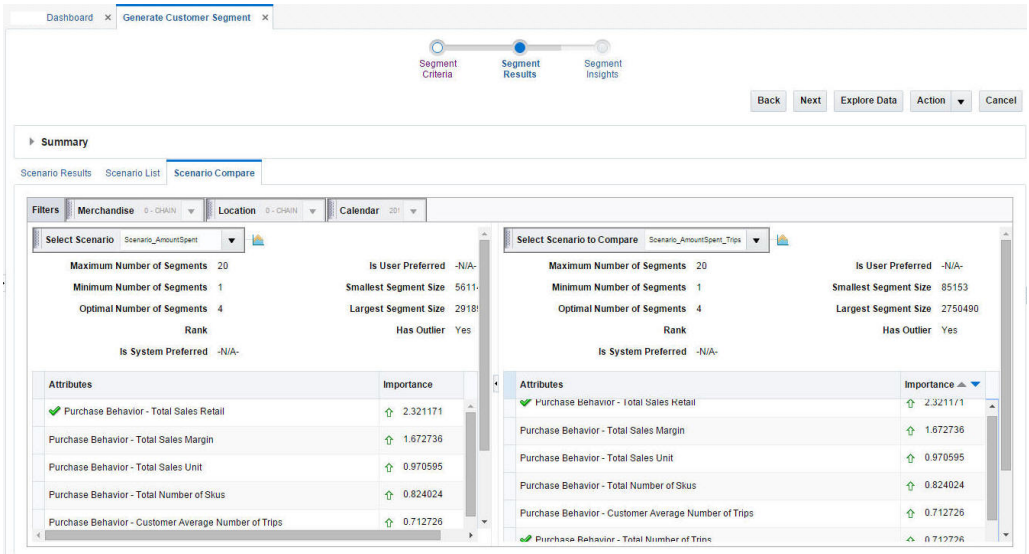
Table 5-24 Scenario List

Field	Description
Name	The name assigned to the scenario.
Status	Ready for Preview, Ready for Approval, Approved, Rejected.
# of Attributes	The total number of attributes used in the segmentation process.
Max. # of Segments	The value used for the maximum in the scenario execution, if this option used.
Min. # of Segments	The value used for the minimum in the scenario execution, if this option used.
Optimal # of Segments	The value used for the optimal number of segments in the scenario execution, if this option used.
System Preferred	Indicates whether the scenario is the one the application prefers.
User Preferred	Indicates whether the scenario is one the user prefers.
Rank Sequence	Indicates the ranking the scenario is given by the application.

Scenario Compare

The Scenario Compare section shows two clusters of your choosing side by side so that you can compare the segment results of each, using the same characteristics used in Scenario Results and Scenario List.

Figure 5-28 Segment Results Scenario Compare



The information displayed includes:

Table 5-25 Scenario Compare

Field Name	Description
Max. # of Segments	The value used for the maximum in the scenario execution, if this option used.
Min. # of Segments	The value used for the minimum in the scenario execution, if this option used.
Optimal # of Segments	The value used for the optimal number of segments in the scenario execution, if this option used.
Rank	The value for the rank.
Is System Preferred	Indicates whether the scenario is the one the application prefers.
Is User Preferred	Indicates whether the scenario is the one the user prefers.
Smallest Segment Size	The size of the smallest segment.
Largest Segment Size	The size of the largest segment.
Has Outlier	Indicates a segment with the number of customers below a threshold. For example, the number of customers is below a certain percentage of the number of customers in a segment.
Attributes	A list of relevant attributes.

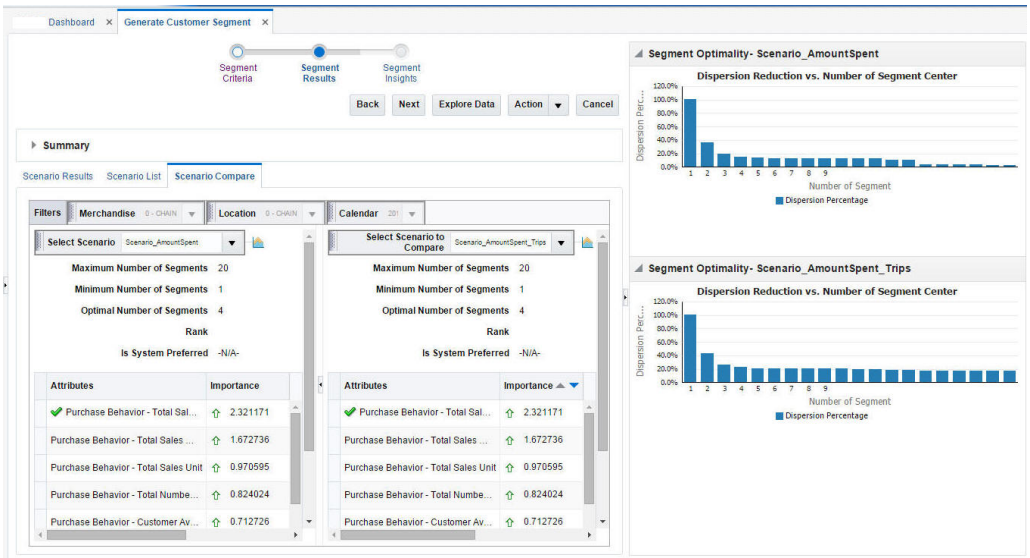
Scenario System Recommendations

The application provides the following recommendations at the scenario (segment set), segment, and customer levels.

Scenario Optimality

This graph indicates how the system identifies the best number of segments for a given data set. It starts with a small number of segment centers and searches for the number beyond which there is minimal dispersion. At this point, increasing the number of segment centers any more only reduces dispersion by a small amount, and the marginal improvement is small.

Figure 5-29 Scenario Optimality



Scenario Rank

You can see the ranking of all scenarios in the Segment Results step. The scenario with the highest rank is designated as System Preferred. The ranking is based on the following:

- How many similar customers are contained in the segment.
- How well separated the segments are from each other.

The rank for a scenario is calculated by taking the average of the probability of all the data points in that scenario. A higher probability value for a customer is an indication of the goodness of fit of the segment.

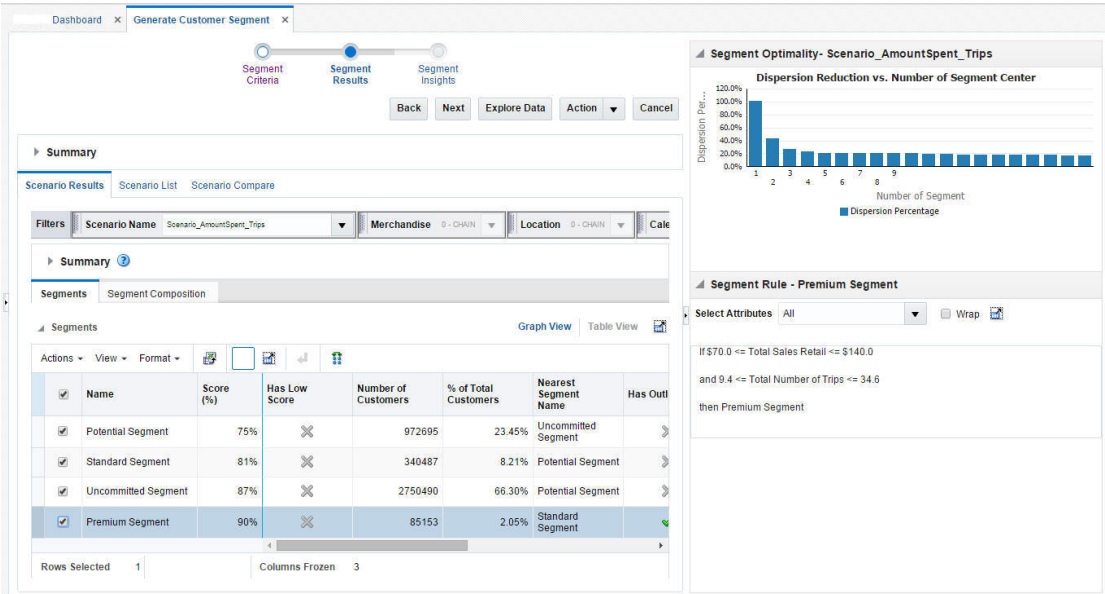
Outlier Indicator

This provides an outlier indicator in the segment list if the segment has an outlier customer. Two outlier rules are supported. The distance from the centroid indicates that if a customer is beyond a certain limit for the configuration threshold from the centroid, then the segment to which the customer belongs is marked as an outlier. When the size of a segment is compared with the total customers, if the number of customers in the segment is below a certain configured percentage of the total customers, then the segment is marked as an outlier. The distance from the centroid is the outlier rule that is configured by default.

Segment Rules

Segment rules describe the data in each segment where most of the data is concentrated. A rule is a conditional statement that captures the logic of how the system generates the hierarchy of segments and what conditions it uses for further splitting a segment. Each rule describes the conditions for a customer to be assigned with some probability to a segment. Figure 5-30 provides an example of rules.

Figure 5-30 Segment Rules



Segment Scores

The application provides scores for segments, based on the threshold score. The scoring measures the goodness of a segment. A score for a particular segment is calculated using the average of the probability scores of the customers in a segment. The score is based on the assumption that each customer has an equal chance of being a member of the segment. A high score indicates that the customer is close to the centroid. A low score indicates that the customer is an outlier. The low score indicator is turned on and off.

New Customers

Customer Segmentation supports periodic processing rules in order to allocate customers who are new. If a customer over time has built enough sales history, the system begins to consider customers and allocates them to segments based on the probability scores.

Generate Store Profile

This allows the user to calculate the sales share of customer segments for each store. These store profiles can be generated by the user from the user interface for the approved customer segments. The store profiles can then be consumed by RI or a subscribing application to be used in generating business reports. The profiles are also consumed in Advanced Store Clustering and used to generate customer-centric store clusters.

Segment Insights Stage

Use the Insights stage to analyze a scenario, its segments, and its hierarchy, based on performance and attribute contributions, prior to the approval of the scenario. This stage includes the following tabs:

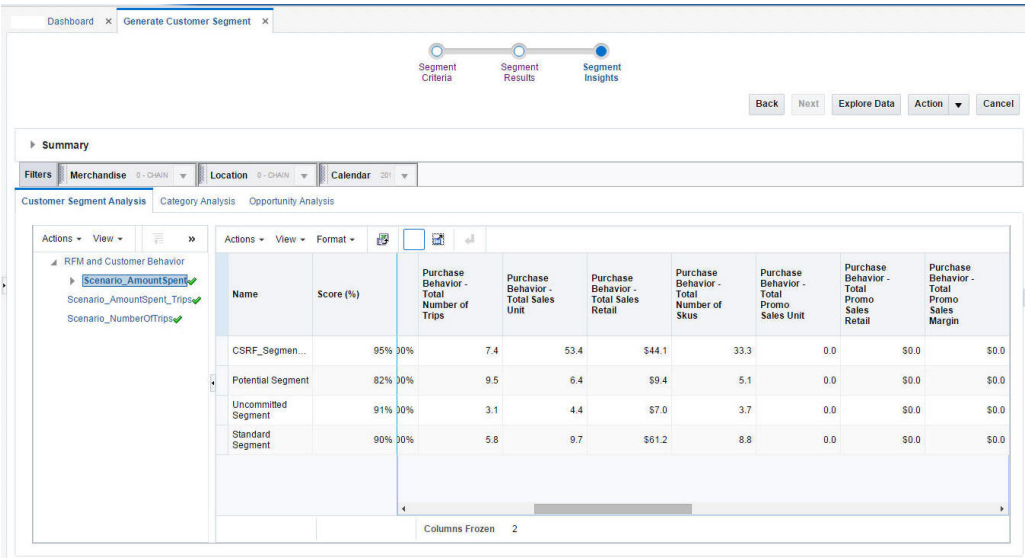
- Customer Segment Analysis
- Category Analysis

- Opportunity Analysis

Customer Segment Analysis

The Customer Segment Analysis tab lists all the scenarios, the customer segments for each scenario, and the related KPIs. Aggregates (mean and mode) of the attributes are displayed for each segment. The attributes by default are included in the grids customer purchase behavior and customer demographics.

Figure 5-31 Customer Segment Analysis



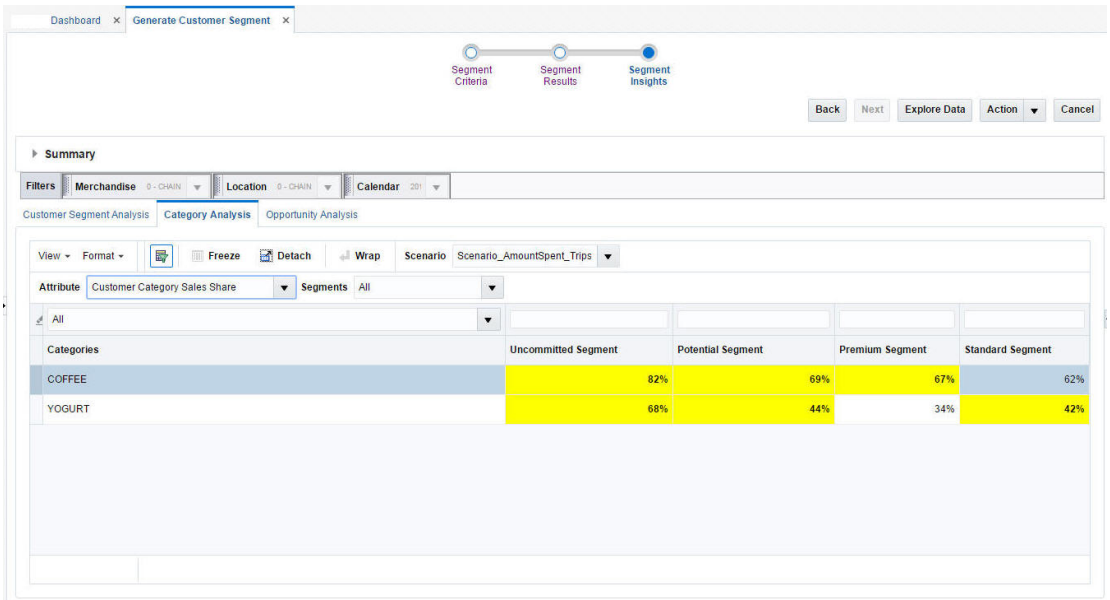
This tab includes the following tasks:

- Evaluate the KPIs for a scenario, it this has not been completed earlier
- Approve a scenario
- Rename segments within a scenario
- Rank scenarios, if not completed earlier
- Flag a segment scenario as system preferred
- Generate store profile for an approved scenario

Category Analysis

The Category Analysis tab lists the top categories vs. the customer segment matrix. You can analyze the segment performance in each category using different sets of attributes such as sales share or number of trips.

Figure 5-32 Category Analysis



The system highlights the top three performing segments for each category, and you can compare category performance across segments. You can filter categories or select a set of segments while comparing the results.

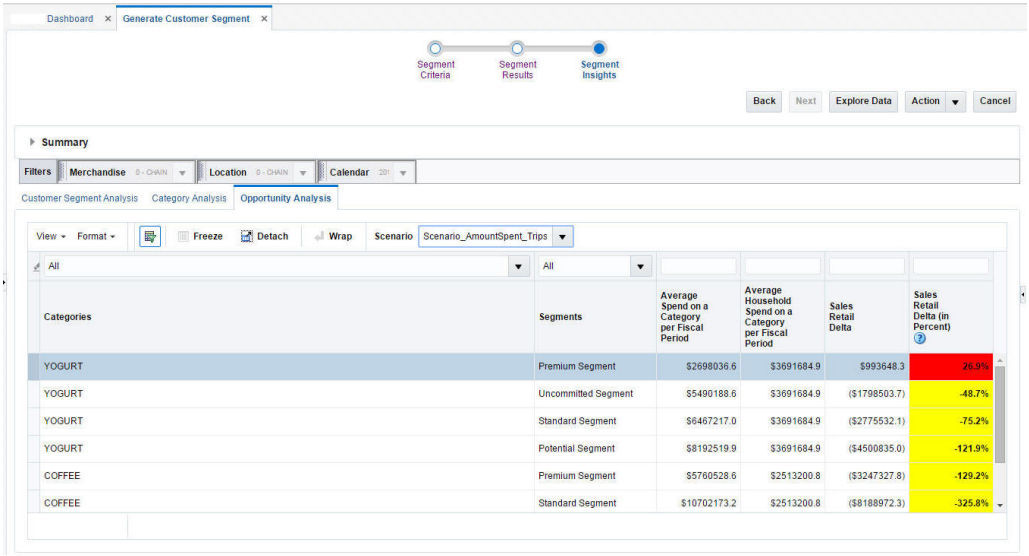
Table 5-26 Axis

Field	Description
Columns	The segments for the selected scenario.
Rows	The top identified categories in the preprocessing.
Cell	The KPI values based on the attribute selected.

Opportunity Analysis Tab

The Opportunity Analysis tab provides a comparison of the retailer to market data analysis for the category and provides an analysis about where there is an opportunity for the retailer to grow for each segment.

Figure 5-33 Opportunity Analysis



You can compare where the retailer is performing better by comparing the average sales retail spend per <configurable time period>.

Table 5-27 Opportunity Analysis

Field Name	Description
Segment	The segments for the selected scenario.
Category	The top identified categories in the preprocessing.
Average spend on a category per <configurable time period>	The average amount spent for a segment.
Average household spend on a category per <configurable time period>	The average household amount spent for a segment.
Sales retail delta	The difference in sales retail between retailer and market data.
Sales retail delta as a percent	The percentage difference in sales retail between retailer and market data.

The color indicators listed in the following table help you see if the difference in sales retail is above or below a certain threshold or within an acceptable range. This grid is sorted with the sales retail difference as negative or highlighted as red, so that you can review the values below threshold first.

Color	Value
Red	X percent below
Green	Y percent above
Yellow	In between X below and Y above

This grid provides a filter for customer segment and category, so that you can review the category for each segment or vice versa.

Summary

This lists the criteria you initially selected to define the segment.

Figure 5-34 Summary

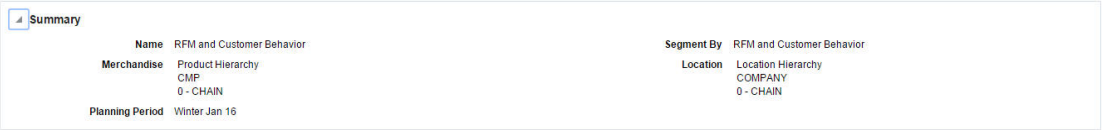


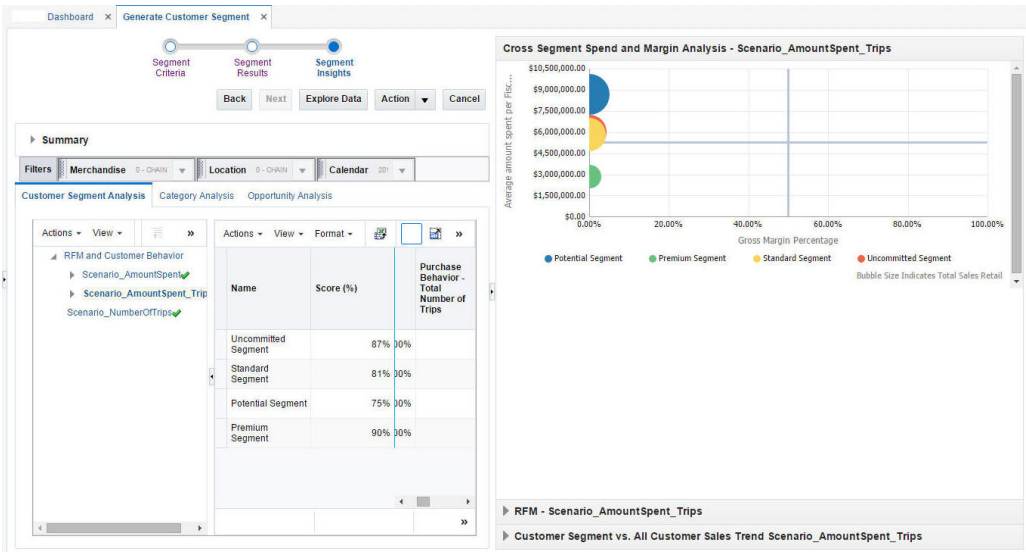
Table 5-28 Summary

Field Name	Description
Name	The name you provided for the segment in the Segment Criteria stage.
Segment by	A predefined group of attributes that include Customer Demographics, RFM and Customer Behavior, Category Purchase Driven. These criteria types are sets of attributes. For example, customer demographics are the properties of a customer. These properties can include ethnicity, income, and age.
Merchandise	The merchandise level and nodes for the segment.
Location	The location level and nodes for the segment.
Fiscal period	The time period for the segment.
Merchandise hierarchy type	This provides details about which type of hierarchy the segment criteria have been created for.

Cross Segment Spend and Margin Analysis BI

This BI compares how segments vary by Gross Margin% (*x* axis), Average Amount Spent <per configuration period> (*y* axis), and Total Sales Retail (*z* axis) for the customer segment. This can help you understand which segments are underperforming and have the potential for growth. The segments in the middle of the chart are the segments that have growth potential, as compared to the premium segments to the top right.

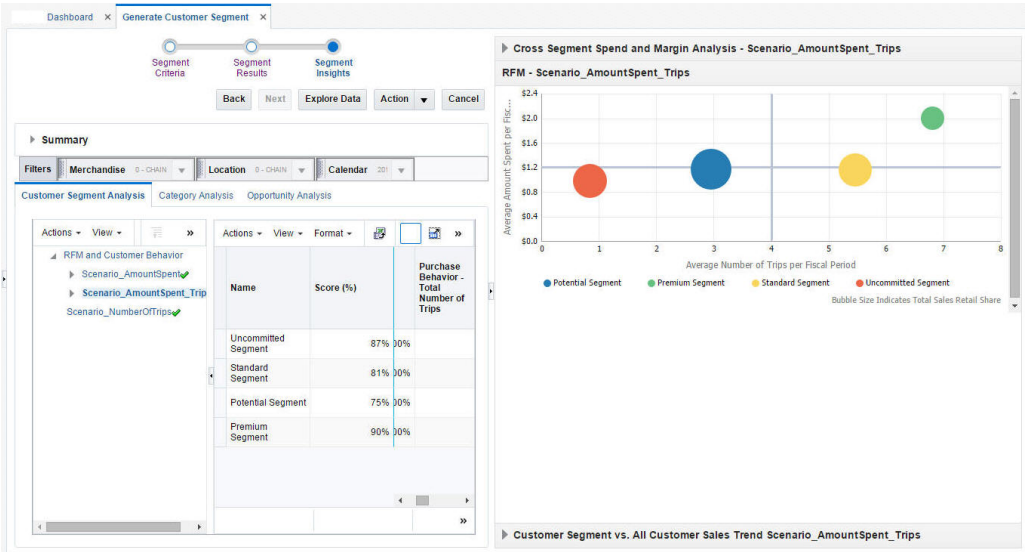
Figure 5-35 Cross Segment Spend Analysis



RFM Analysis BI

This BI provides analysis of segments based on frequency and amount spent. The more frequent customers tend to have a significantly better response and average spend than those who have shopped less recently or less frequently. This indicates customers that are high spender and frequent shoppers are premium segments, vs. low spender and infrequent shopper are uncommitted customers. Figure 5-36 shows how the premium customer segment is in the top right corner, the potential customer segments are close to the center, and the uncommitted shoppers are in the bottom left quadrant. This helps in planning how potential and uncommitted customers can be attracted by certain offers and how existing committed customers can be satisfied.

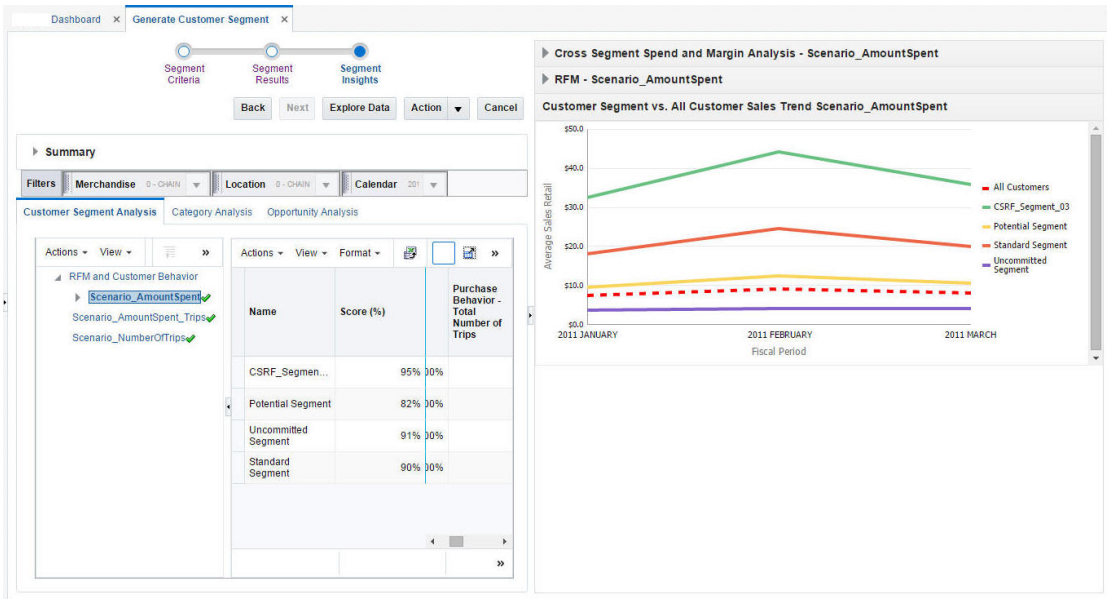
Figure 5-36 RFM Analysis



Customer Segment vs. All Customer Sales Trends BI

This BI shows a customer segment's sales trends by comparing Average Sales Retail per Segment vs. Average Sales Retail for all customers for the selected criteria time period. You can then see if sales retail has decreased for a specific customer segment and for which customer segment it has remained consistent for the defined quarters. This can help you determine if certain customers are migrating from one customer segment to another over the time period. Figure 5-37 shows how, for each segment, sales increased in 2011 February and the average sales of entire customer base, compared to the segments created.

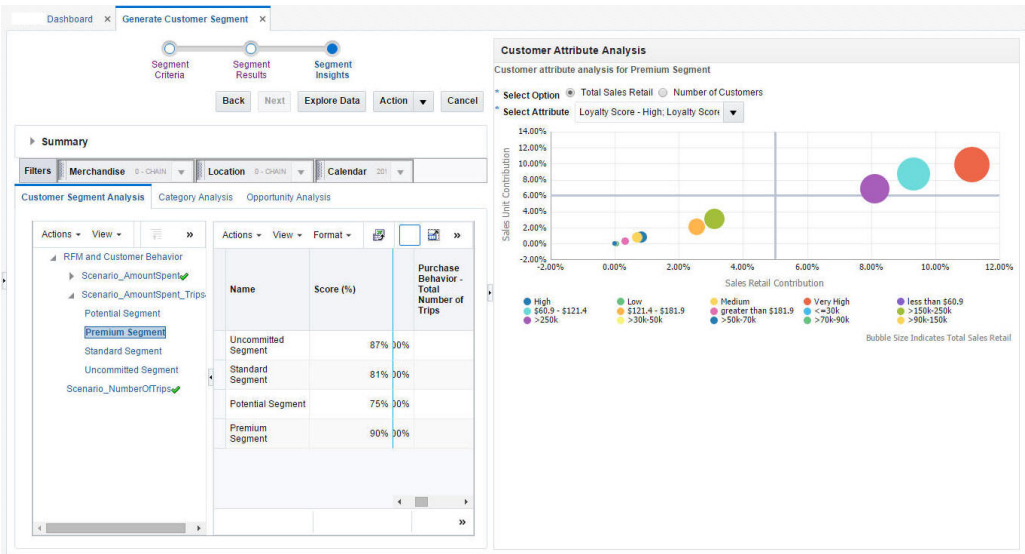
Figure 5-37 Customer Segment vs. All Customer Sales Trends



Customer Attribute Analysis BI

Customer attributes are analyzed and compared to identify the most prominent selling attributes within a segment. This indicates how the customer attributes are correlated and what each attribute in the specified segment contributes to sales. You can make inferences about which attributes in the segment contribute to significant sales and the potential attributes that should be considered for assortment planning in order to improve sales even more. This chart also provides you with the option to see the number of customers in each segment instead of total sales retail, thus providing insights about how many customers with certain attributes and attribute values are present in certain segments. Figure 5-38 shows how customers with high income and high loyalty score contribute to most of the sales in the segment.

Figure 5-38 Customer Attribute Analysis



The following properties are displayed by the graph. Note that the customer attributes are only displayed when they are configured as part of the Cluster by process.

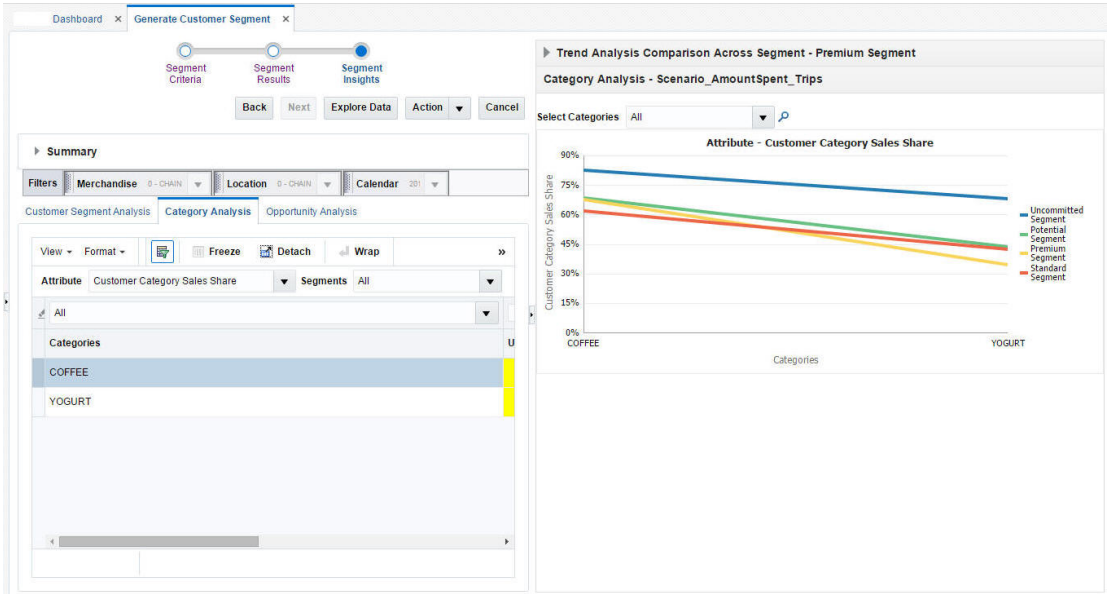
Table 5-29 Properties

Axis	Description
x-axis	Sales Retail $ contribution, calculated using the sales revenue share of each attribute in the segment with respect to the total segment sales revenue.
y-axis	Sales Unit $ contribution, calculated using the sales unit share of each attribute in the segment with respect to the total segment sales units
z-axis	Total Sales Retail $ or number of customers of each attribute in the selected segment, indicating, via the bubble, the magnitude of the sales contribution.

Category Analysis BI

You can perform category/segment analysis by comparing categories using a selected attribute (such as category sales share, category promotion sales share, average amount spent per month, or number of trips). This can indicate the segments and categories that customers are shopping frequently.

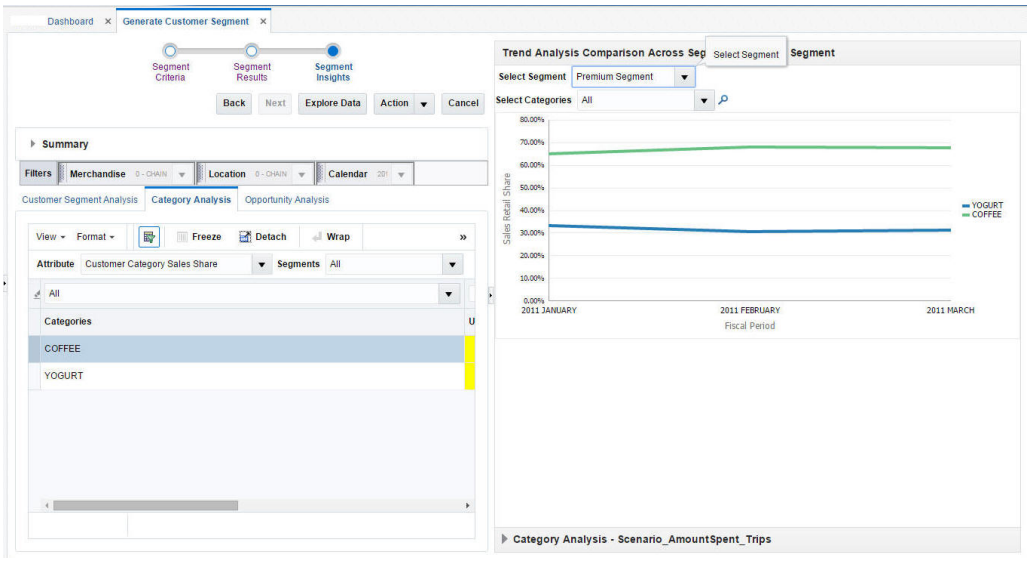
Figure 5-39 Category Analysis



Trend Category Analysis BI

This BI compares sales for selected categories in a segment. You can compare sales by <Configurable Period (Period or Quarter)> to view trends for each category. You can limit the number of categories by selecting a subset of the categories in the drop-down list. You can use this information to plan certain categories for certain seasons by comparing sales with previous periods or by modifying a category plan that is underperforming for a segment in a certain season. The time period displayed is driven by the selection you make in the Segment Criteria. If you select a source time period to include two years' worth of data then you will also see the data for the previous year.

Figure 5-40 Trend Category Analysis



6
Attribute Extraction

This chapter describes the Attribute Extraction Cloud Service module.

Introduction

Attribute Extraction (AE) is an enterprise-specific solution that uses machine learning to extract product attributes from free-form product description strings.

The application's embedded science and automation helps you to extract the attributes (such as brand, color, flavor, and so on) of each product in a particular category and to normalize the attribute values by correcting short forms, misspellings, and other inconsistencies. The product attributes can be used by Demand Transference, Customer Decision Trees, Advanced Clustering, and other retail applications that require product attributes in a structured format.

The AE Cloud Service module consists of the following tabs: Overview, Edit Labels, Annotation, Errors, Normalization, and Results. You use the Overview tab to select one of the previously added product categories or to add a new category. You use the Edit Labels to define category-specific attributes that you want to extract. In the Annotation and Errors tabs, you follow an iterative process to extract attributes and correct any mislabeled attributes. In the Normalization tab, you can use the embedded List of Values (LOV) or create your own LOV to standardize the attribute values. You use the Results tab to review and export the table of attributes.

Overview Tab

The Overview tab displays general information about the product categories previously added by you and other users. You can view this information to determine which user has created each category, when it was last updated, and whether or not the attribute extraction process is complete.

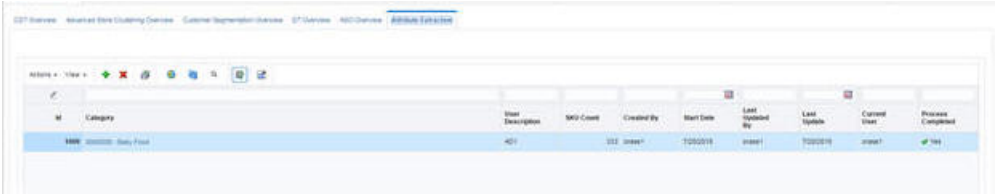
To use the attribute extraction process, you can either select one of the existing categories or you can add a new category. Note that only one copy of each category can exist at any time.

To add a new category, click the **Add** icon. In the pop-up screen, select a category from the list and enter a description. Then click **OK** to return to the Overview tab.

To select a category and start the attribute extraction process, simply click on its name in the overview table.

The AE overview tab is shown in Figure 6-1.

Figure 6-1 Attribute Extraction Overview Tab

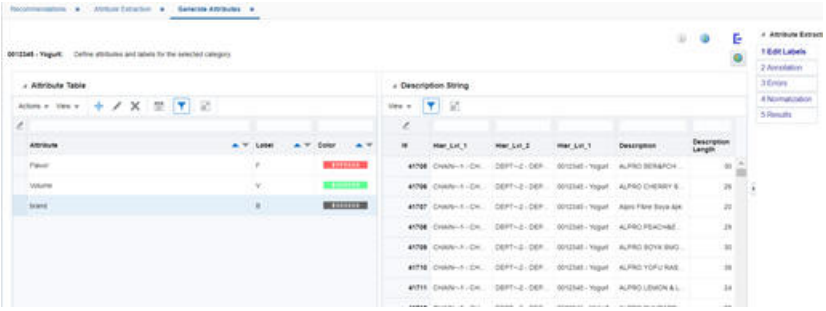


Edit Labels Tab

The Edit Labels tab is used to define category-specific attributes and their labels for the selected category. Defining attributes and labels is the first step in the attribute extraction process. You can return to this tab at any time to add, remove, or modify the attributes.

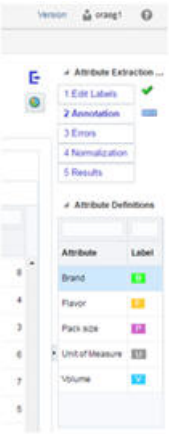
On the right-hand side of the tab, you can view the description strings of the products in the selected category as well as three levels of hierarchy for the products. The hierarchy that is shown in the Attribute Extraction tables is based on the product hierarchy data that was loaded for the AE application. (See the Implementation Guide for further information about the hierarchy.) To add an attribute, click the **Add** icon. In the pop-up screen, enter the attribute name and attribute label and select a color. The attribute label must be a single character (the characters "o" or "O" are not permitted) and must be unique for each attribute.

Figure 6-2 Edit Labels Tab



The attributes, labels, and colors that you define in the Edit Labels tab are displayed on the right-hand side of all other screens for your reference.

Figure 6-3 Attribute Definitions



Annotation Tab

The Annotation tab is used to assign labels to a single token (referred to as a unigraph) or combination of two adjacent tokens (referred to as a digraph). A token is defined as an alphabetic or numeric sequence of characters in a description string. For example, the string "ABC vanilla yogurt 500gr" is a description that consists of the following five tokens (or unigraphs): "ABC", "vanilla", "yogurt", "500"and "gr". At the same time, it consists of the following digraphs: "ABC vanilla", "vanilla yogurt", "yogurt 500" and "500 gr".

The Annotation tab provides two screens to annotate unigraphs and digraphs.

Unigraphs

In the Unigraph screen in Annotation tab, you can assign attribute labels to tokens, run the machine learning algorithm to find new attribute labels, and review and approve machine-recommended labels.

Assigning User Labels and Reviewing Machine Labels

On the top left of Unigraph screen, you see the Annotation section that contains a table listing all tokens along with the following fields:

Table 6-1 Annotation Fields

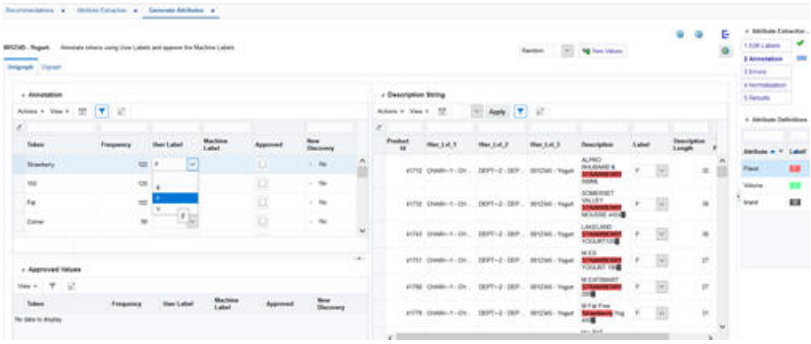
Field	Description
Frequency	The number of times the token appears across all product description strings
User Label	The label you have assigned to the token
Machine Label	The label recommended by the machine learning algorithm
Approved	A check box used to approve and apply the machine label to the token
New Discovery	A Yes/No flag that indicates whether or not the machine-recommended label is a new discovery from the most recent run

In the Annotation section, shown in Figure 6-4, you can perform one of the following tasks for each token:

Table 6-2 Annotation Section Tasks

Task	Description
Change/assign the user label	Select a value from the drop-down menu to change the label or assign a label to a token. (Values in the drop-down menu are the attribute labels that you defined in the Edit Labels tab.) The user label is applied to all instances of the token across all product description strings.
Approve a machine label	Check the Approve check box to assign the machine-recommended value to a token. Once you check the box, the token with the approved machine label is moved from the top tables to the bottom table (the Approved Values section).

Figure 6-4 Annotation Tab with Label Tokens in Unigraph Screen



The Approved Values section on the bottom left displays the tokens that you have approved so far. You can uncheck the Approve check box in this section to remove the approved machine label and move the token back to the top table.

The Description String section on the right side of the Unigraph screen displays all description strings that contain a token selected on the left, along with the length of the strings and position of the selected token in the string, as well as the product hierarchy. The Label column is populated based on the user label or approved machine label for the token. All labeled tokens are colored based on the colors you assigned in the Edit Labels tab. The token selected on the left (in the Annotation section) is shown using a bold font.

You can perform one of the following tasks in the Description Strings section:

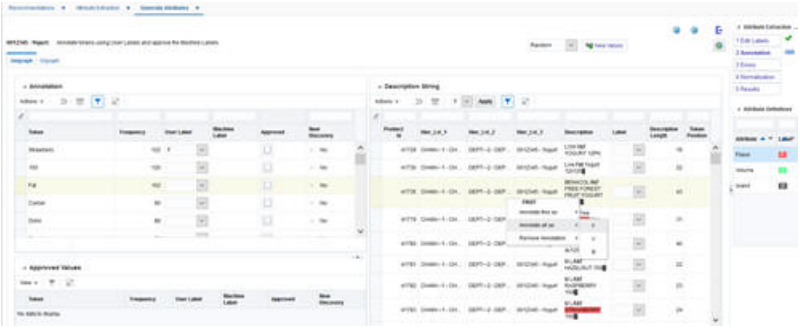
Table 6-3 Description Strings Section Tasks

Task	Description
Change or remove the label for one instance of the selected token	Use the inline drop-down menu in the label column to change or remove the label of the token in a particular description string. Alternatively, right click on any token and use the first option in the menu (Annotate this as) to assign a label to a particular instance. Use the third option in the menu (Remove Annotation -> Remove this label) to remove the label for a particular instance

Table 6-3 (Cont.) Description Strings Section Tasks

Task	Description
Change or remove the label for multiple instances of any of the tokens	Select multiple rows using the Ctrl or Shift keys. Then, change the label using the drop-down list at the top of the table and click **Apply**. To remove the label for multiple rows, change the label to NULL by selecting the first option in the drop-down list at the top of the table. Alternatively, to change or remove the label for ALL instances of a token, right click on any token and use the second option in the menu (Annotate all as) to assign a label to all instances, or use the third option in the menu (Remove Annotation -> Remove all labels) to remove the label for all instances. Example: if you wish to change the label of a token for all products that belong to a certain department, you can first filter the hierarchy columns based on department. Then, select all of the rows using the Shift key and change the label for all selected rows using the drop-down list at the top of table and click **Apply**.

Figure 6-5 Description Strings Tasks



Finding New Values

Once you assigned labels to some of the tokens in Unigraph screen, you may use the New Values button to run the machine learning algorithm that finds new attributes and recommends labels based on the labels that you have assigned and approved so far. Alternatively, you may go to the Digraph screen and label some digraphs before running the algorithm. Note that assigning and approving unigraph and digraph labels and running the algorithm to find new values is an iterative process; you can repeat these two steps as frequently as you want until you are satisfied with the quality of the results and number of attributes extracted. If the machine learning algorithm does not find enough new values, it is an indication that it requires more information to discover the patterns and be able to recommend new values, and you might want to label some more unigraphs and digraphs before re-running the algorithm.

To run the machine learning algorithm, first select a mode from the drop-down menu next to the New Values button on the top right. You can select from the following three modes:

Table 6-4 Machine Learning Algorithm Modes

Mode	Description
Random	Randomly partitions the data into a training set and a test set to be used by machine learning algorithm.
By attribute	If you select this mode, you will be prompted to select one attribute. The data is then partitioned in such way that all description strings that have one or more tokens labeled as the selected attribute are used as the training set. The remaining descriptions strings are used as test set.
By annotation	The description strings that are 80% labeled are used as training set. The remaining descriptions strings are used as test set.

The By Attribute mode is most effective in the early stages, where very few attribute values have been identified.

Note that after you click the New Values button, it may take few seconds for the machine learning algorithm to run and find new attributes. Then a pop-up message shows how many new values were found. The Machine Label and New Discovery columns in the Annotation section on the top left will be populated to show the new labels and indicate which tokens were labeled by the machine in the most recent run.

Digraphs

The Digraph is used for labeling two adjacent tokens (as opposed to a single token) with the same attribute. The Digraph screen consists of an Annotation section on the top left for labeling digraphs, an Approved Values section on the bottom left, and a Description String section on the right for labeling individual instances of a digraph.

Assigning User Labels to Digraphs

On the top left of Digraph screen, you see the Annotation section that contains a table listing all combinations of adjacent tokens (labeled as "token1" and "token2") where exactly one of the two tokens has been labeled and the other token has no label assigned or approved. The token that already has a label is shown using a bold font.

Other fields in this section are as follows:

Table 6-5 Diagraph Fields

Field	Description
Frequency	The number of times that the two tokens are adjacent to one another across all product description strings.
Label	The label you have assigned to or approved for one of the tokens. (The token that is shown in bold font already has this label.)
Approved	A check box used to approve and apply the label of labeled token to the unlabeled token.

If you think the two tokens are of the same attribute type indicated by the label, check the Approved check box to apply the label to both tokens. Once you check the box, the

approved row is moved from the top table to the bottom table (the Approved Values section).

The Approved Values section on the bottom left displays the adjacent tokens that you have approved so far. You can uncheck the approve check box in this section to remove the approved row and return it to the top table.

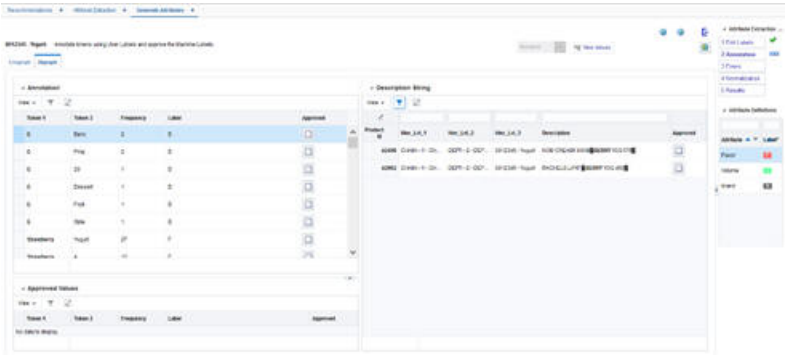
The Description String section on the right side of the Digraph screen displays all description strings that contain the adjacent tokens selected on the left, as well as the product hierarchy. All labeled tokens are colored based on the colors you assigned in the Edit Labels tab. The two tokens selected on the left (in the Annotation section) are shown using a bold font.

You can do one of the following in the Description Strings section:

Table 6-6 Description Strings Section Tasks

Task	Description
Remove the label of the originally unlabeled token for one instance	If you believe the combination of two tokens must not be labeled the same in one or few description strings, un-check the Approved check box to remove the label for the token that was not originally labeled. Note that the token that was originally labeled (i.e., the token displayed in bold font in the Annotation section on the left) remains as labeled and is not affected.
Change or remove the label for one or all instances of any of the tokens	As with the Unigraph screen, you can use the right click menu on any token and use the Annotate this as option to assign a label to a particular instance or use the Annotate all as option to assign a label to all instances of the token. To remove the label for a particular instance or all instances of the token, use Remove annotation -> Remove this label and Remove annotation -> Remove all labels.

Figure 6-6 Annotation Tab Diagraph Screen



Errors Tab

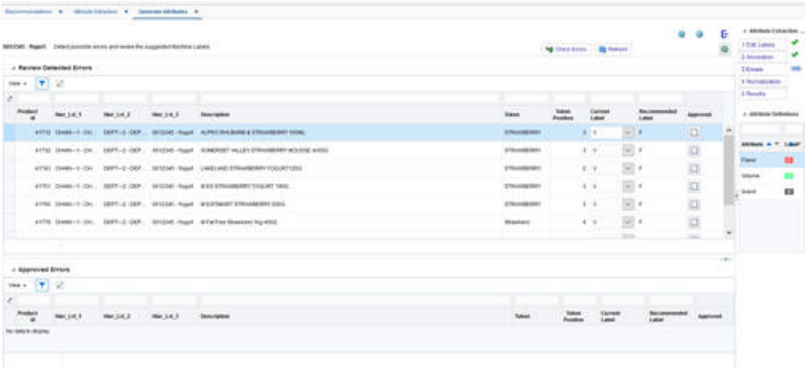
The Errors tab is used to review the instances where the current attribute type assigned to a token is different from the attribute type recommended by the machine. You can review these potential user errors and fix them if needed. For each instance, you can take one of the following actions:

Table 6-7 Error Tab Actions

Action	Description
Approve the Recommended Label	Check the Approved check box to approve the Recommended Label. The approved instance will be moved from the top table to the bottom table. You can use the check box in the bottom table to undo the approval.
Change the Current Label	If neither the Current Label nor the Recommended Label is correct, use the drop-down list to change the Current Label. Do not check the Approved check box in this case. If the Current Label is correct, no action is needed.

It is recommended that you review the errors after you complete the annotation step and extract the attributes. However, you can navigate to the Error tab at any time to review the errors and change or approve the labels for some or all instances.

Figure 6-7 Errors Tab



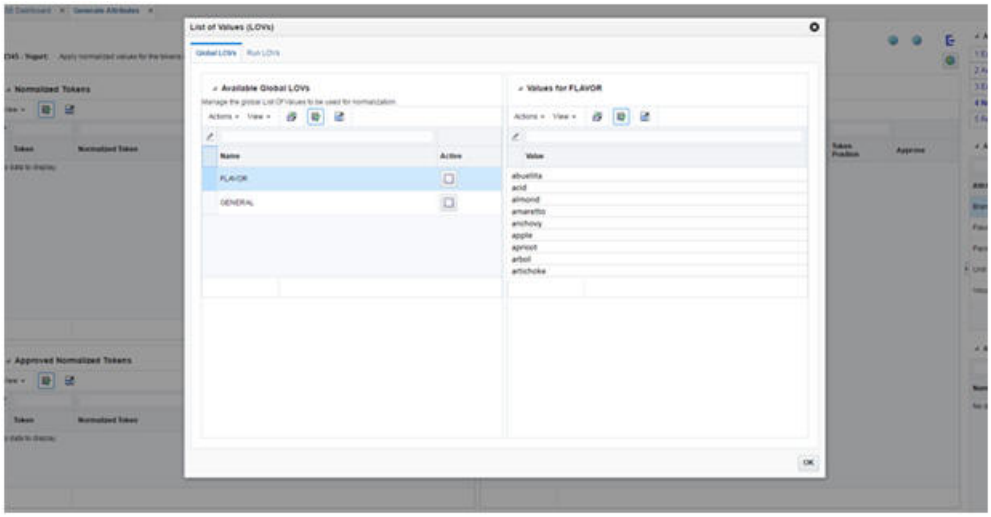
Normalization Tab

After you extract the attributes, you can use the Normalization tab to correct mis-spellings, sort forms, and other inconsistencies in the attribute values to makes sure that each attribute value is displayed in a consistent form across all product descriptions. For example, strawberry is a flavor that may appear in many different forms such as "sberry", "strawb", and ""strberry" across description strings. The purpose of using the Normalization screen is to convert all these different forms to the correct form (i.e., strawberry).

To run the spell-correcting algorithm, you can use one of the pre-defined List of Values (Global LOVs) or you can create your own LOV (Run LOVs) specific to one attribute. To use either of the LOVS, click the List of Values button on the top right.

To select a Global LOV, check the Active check box next to it. There are two Global LOVs: flavor (list of all different flavors) and general (a comprehensive list of English words). You can click the row for each LOV to see all the values in the list. If you select the flavor LOV, it will only be used to correct the values of the flavor attribute. If you select the general LOV, it will be used for correcting the values of all attribute types.

Figure 6-8 Normalization Tab with Global List of Values



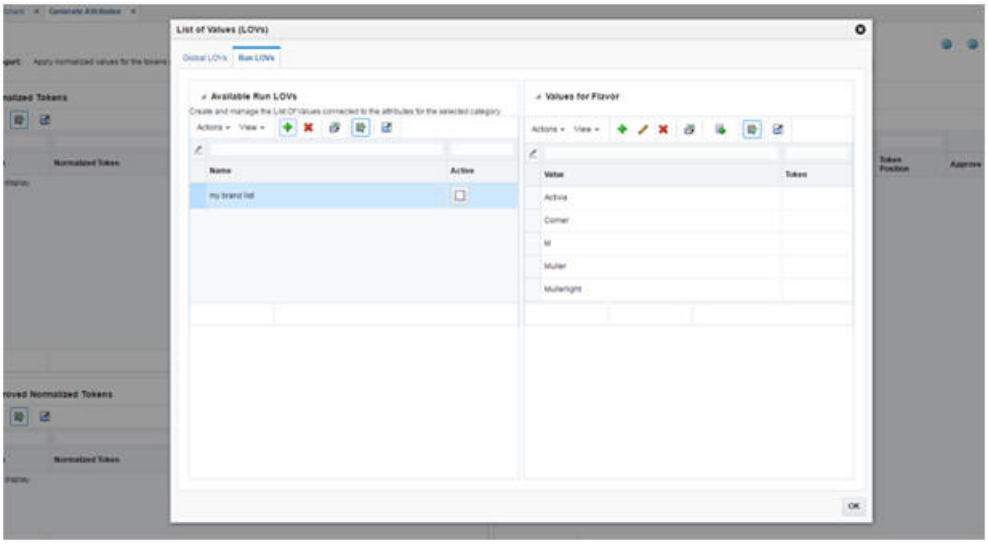
To create a new Run LOV or select an existing Run LOV, navigate to the RUN LOVs tab in the pop-up screen. To create a new list, click the Add button on the left table (Available Run LOVs). Then select the attribute that you want to create the list for and assign a name to the LOV. You can also pre-populate the list with tokens that are labeled as the selected attribute and appear more than a certain number of times across all description strings. The reason for this option is that usually the tokens with high frequency have the correct form and spelling. For example, there may be few instances of "sberry" and "strawb", but most certainly there are many instances of "strawberry". So the high frequency tokens are likely to have the correct form and can be used by the spell correcting algorithm.

After you create a Run LOV, you can edit or remove the values in the list or add new values. To edit or remove values, select a row in the right table and use the Edit or Delete button. To add a new value, click the Add button on the top of table on the right.

In addition to adding a new value to the list (to be used by the spelling correction algorithm), you can define a value/token pair so that all instances of the token across all description strings are replaced by the defined value. This is a useful option when the data has lots of abbreviations and short forms that may be difficult to correct using a spelling correction algorithm. For example if the name "Hello Kitty" is a brand that appears as "HK" in many of the description strings, then you can define a pair as "Hello Kitty|HK" to have all instances of "HK" replace by "Hello Kitty". (Note that the correct value and the token must be separated by a pipe delimiter (|)).

Once you are done adding the Run LOV and editing the values, make sure to check the Active check box on the left table to select the list.

Figure 6-9 Normalization Tab with Run List of Values



After you select a Global or Run LOV, click the OK button to close the pop-up screen and return to the Normalization tab. You can see the active LOVs on the bottom right of the screen.

To run the spell correcting algorithm, click the Normalize button. It may take few seconds for the algorithm to run. When the run is complete, the recommended corrections are displayed in the Normalized Tokens table on the top left. This table has the following fields:

Table 6-8 Normalized Tokens

Field	Description
Token	The token identified by the algorithm as misspelled.
Normalized token	The recommended correct value for the token or the value that you defined for replacement (i.e., the value/token pair).
Frequency	The number of times the token appears across all product description strings.
Approved	A check box used to approve and apply the recommended correct value.

You can perform one of the following in Normalization tab:

Table 6-9 Normalization Tab Tasks

Task	Description
Edit the normalized token	If you do not agree with the recommended correction and want to edit the normalized token, you can edit the text in the Normalized Token column before approving it.

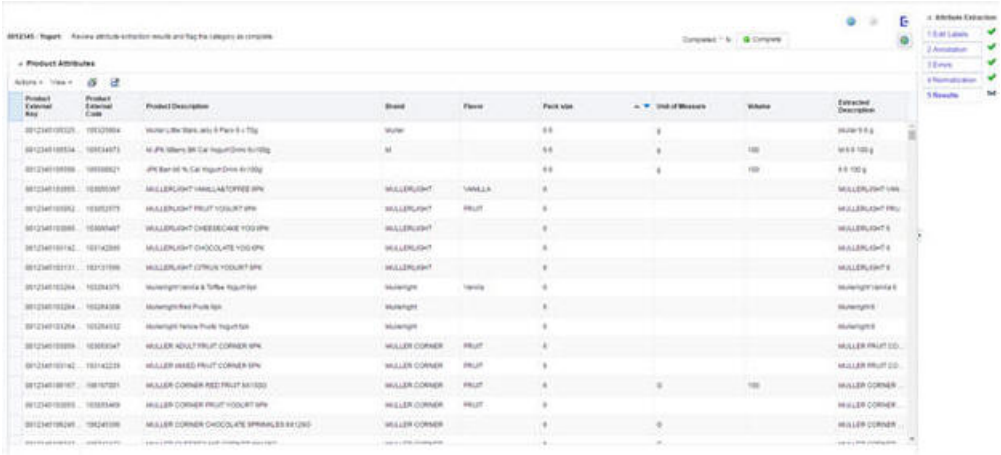
Table 6-9 (Cont.) Normalization Tab Tasks

Task	Description
Approve/reject the normalized token for all instances	Check the Approved check box to replace all instances of the token with a normalized token. The approved rows will be moved from top table to the bottom table (Approved Normalized Tokens). To undo the approval, uncheck the check box in the bottom table.
Approve/reject the normalized token for one instance	When you click on a row in the Normalized Token table, all descriptions that contain the selected token will be displayed in the description strings table on the right. You can approve or reject individual instances of the normalization by checking or un-checking the Approved check box in the right table.
Clear all recommended corrections	To clear all recommended normalized tokens, click the Reset button on top of the Normalized Tokens table on the top left.
Refresh	It is recommended that you click the Refresh button on the top right after you make changes to a LOV or select/deselect a LOV.

Results Tab

The Results tab is used to view the table of attributes. You can export the results into a spreadsheet. You can click the Complete button on the top right to change the status of the category in the overview tab to a value of "complete". This indicates that the attribute extraction process for this category is complete. Other users who log in to the application may decide to work on incomplete categories.

Figure 6-10 Results Tab



Status

The Status screen is accessible from every tab. To open, click the pie chart button on the top right.

In the Status screen, you can see the progress of the attribute extraction process. It displays different statistics, including total number of tokens, number of distinct tokens, and number of labeled tokens.

You can also see the breakdown of labeled tokens by attribute type. To see the list of all distinct values for a certain attribute, click the respective section of the attribute in any of the pie charts.

Figure 6-11 Status Screen

7
Affinity Analysis

This chapter describes the use of the Affinity Analysis (AA) Cloud Service module.

Introduction

Market basket analysis involves the use of data mining techniques to search for sales patterns between products within a given group of transactions. The output of that analysis provides a rule that defines the association found between products at the subclass or class level of the merchandise hierarchy.

A rule consists of one to three antecedents (IF attributes) and a single consequent (THEN attribute). For example:

IF (milk) and (juice), THEN (cereal)

In other words, if a customer purchases an item from the subclasses milk and juice, the customer will also purchase an item from subclass cereal. After a rule is defined, a user can use the AA interface to understand how strong the affinity is, using rule confidence and support. The probability that a customer will buy milk, juice, and cereal is known as the *support percentage*, while the conditional probability that they will buy cereal when they buy milk and juice is known as *confidence*. Rules with a very high support value occur frequently in your transaction history, while rules with a high confidence value represent a strong affinity between products.

After users have identified selling patterns, they can begin to take action based on those patterns, as well as the needs and goals of their product category. Suppose that a merchant is tasked with bringing in more margin dollars to the cereal category. Using the affinity rule in the preceding example, the merchant might work with the dairy category on a milk promotion to increase sales of milk. This in turn increases the sales of cereal, without sacrificing margin dollars on a cereal promotion. Note that this can require cross-category planning in some cases, as product affinities can sometimes occur between seemingly unrelated products (such as pet food and beer).

Another component of market basket analysis relates to the product assortments being sold in stores. Using product affinities and sales history, AA provides assortment recommendations that improve the revenue or margin of a category by suggesting product additions or removals. Products may be recommended for removal if they are found to be too similar to other products in the assortment (and thus cannibalize the sales of those products). Conversely, products that are not similar to any other items in the assortment may be candidates for inclusion, as they will not divert sales from the existing assortment.

Market basket rules are used to improve the assortment recommendation by showing the potential lift (or halo effect) on your overall sales due to any known affinities on recommended item additions. For example, if AA is analyzing an assortment for Coffee, and a particular item is part of a market basket rule that drive sales for Milk, then that item has a greater potential value for the lift it brings to the Milk category. AA may then recommend that item over other items in the category, because including it will bring in additional revenue to other assortments without changing those assortments directly.

User Interface

The AA user interface consists of several screens that help the user analyze and take action on the results of the market basket analysis and assortment recommendation processes. The following list summarizes the main purpose of each screen:

- Assortment Recommender - Review the product assortment recommendations made by AA, including the expected halo effects from items in the optimized assortment.
- Top N Class Affinities - Review the market basket affinity rules identified by AA between different product classes, including insights around the frequency and profitability of the top rules.
- Top N Subclass Affinities - Review the market basket affinity rules identified by AA between different product subclasses, including insights around the frequency and profitability of the top rules.
- Top N Promotion Affinities - Review the market basket affinity rules identified by AA between different product subclasses under the effects of a promotion, which can be used to identify the effects promoting a category has on other non-promoted products.
- Top N Customer Segment Affinities - Review the market basket affinity rules identified by AA between different product subclasses under the effects of a promotion and separated by customer segment, which provides insights into the top promoted product affinities for a targeted group of customers.

The way that you interact with the AA user interface depends on your business role and which insights you want to take action on. For example, a category planner looking to make changes to their assortment to increase margin dollars might start from the Assortment Rec screen. The planner might select the category and one of the top-selling locations, and then review the added or dropped items AA has recommended. From there, the planner can choose to move forward with the recommendations and make the necessary adjustments to the assortment plan or dive deeper into the Product Affinities screens to better understand why certain recommendations are being made.

Regardless of which screen you are examining, a set of global prompts determines the data displayed throughout the application.

On first accessing the application, you make selections from the prompts and click on the **Submit** button, which will load the remaining application screens based on your selections. Modifying and submitting the prompt values after that time will refresh the screens with new data.

Table 7-1 AA Global Prompts

Prompt Name	Description
Top 10 By	Select the primary metric used to rank the market basket results in all places where affinity rules are displayed.
Fiscal Week From	Select the starting week from which results must be displayed. The values displayed are the week of the Fiscal Calendar.
To	Select the ending week that results must be displayed for.

Table 7-1 (Cont.) AA Global Prompts

Prompt Name	Description
Department	Select the department from the merchandise hierarchy that results must be displayed for. All of the classes or subclasses on the THEN side of an affinity rule will be from the selected Department.
Location	Select the location from the location hierarchy that results should be displayed for.

The Top 10 By prompt allows you to select from the following metrics, which are used to determine which affinity rules are displayed and how to sort the results that are shown.

Table 7-2 Top 10 By Options

Metric Name	Description
Net Sales	Chooses the top N rules in each screen, based on the net sales retail amount (e.g., Avg Sales from If Items)
Net Quantity	Chooses the top N rules in each screen, based on the net sales quantity (e.g., Avg Sales Quantity from If Items)
Net Profit	Chooses the top N rules in each screen, based on the net sales profit (e.g., Avg Profit from If Items)

Although this prompt indicates "Top 10", you have the ability to select more than 10 results by changing the Top selection in the Market Basket Filters area of the screen, as discussed later in this chapter.

The Fiscal Week From and To prompts allow you to select a range of dates for which the results are returned. Market basket affinity rules may change over time as new sales transactions are processed, so selecting the weeks to review ensures that only the rules that are applicable for that period of time are displayed.

The Department prompt displays a list of departments from the merchandise hierarchy. Selecting a department will limit the data in all screens such that the consequent (or THEN components) of an affinity rule will be from that department.

After all selections are made and the Submit button is clicked, the interface will refresh and display results. Each section of the application is organized under an informational tile that describes the purpose of the screen, along with summarization metrics.

Figure 7-1 Info Tiles

Each tile has the following general properties.

Table 7-3 Info Tile Components

Component	Description
Title	A descriptive name for the data contained below the selected tile.
Primary Metric	A key measure that highlights important data, such as the number of affinity rules available for review for a given set of prompt selections.
Secondary Metrics	Additional measures that describe the underlying data available when clicking on the info tile.

Clicking on any info tile's title or primary metric will display the associated data in the lower section of the screen. The data shown in each tile in the application is described in the following Table.

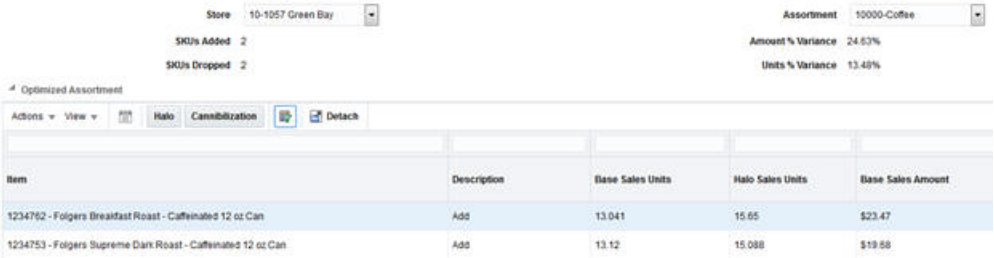
Table 7-4 Info Tile Metrics

Tile Name	Primary Metric Description	Secondary Metric Description
Assortment Rec	The number of assortment recommendations available for the selected global prompt values.	For the assortments in the department, the percentage the sales amount changed and the amount the sales changed.
Class	The number of class level market basket affinity rules available for the selected global prompt values.	The share of sales profit and sales amount that is represented by all discovered association rules out of the total sales profit and sales amount processed.
Sub Class	The number of subclass level market basket affinity rules available for the selected global prompt values.	The share of sales profit and sales amount that is represented by all discovered association rules out of the total sales profit and sales amount processed.
Promotion	The number of promoted subclass level market basket affinity rules available for the selected global prompt values.	The share of sales profit and sales amount that is represented by all discovered association rules out of the total sales profit and sales amount processed.
Customer Segment	The number of customer segment promoted subclass level market basket affinity rules available for the selected global prompt values.	The share of sales profit and sales amount that is represented by all discovered association rules out of the total sales profit and sales amount processed.
All	The total number of market basket affinity rules available for the selected global prompt values.	The last date that market basket insights were executed on.

Assortment Recommender

The assortment recommender screen provides insights on ways to improve the mix of products in a store assortment, using a combination of data from Demand Transference and Affinity Analysis. This section describes the data shown in the screen, as well as how to use the results to take action within your business. For more information on Demand Transference, refer to Demand Transference.

Figure 7-2 Assortment Recommender



The first step in using the assortment recommender screen is to select a store and assortment to review. The stores listed in the drop-down menu are limited to those stores that have completed an assortment recommendation calculation for the selected department in the selected period. The assortment drop-down menu is limited to those assortments with a completed assortment recommendation calculation for the date range, department, and location previously selected, and align with the level of the product hierarchy that assortment-planning operations occur at. For example, if we process Coffee on Week 1, Yogurt on Week 2, and Milk on Week 3, and we pick a date range of Weeks 1-3 in the global prompts, we will show results for those three categories.

After selecting a store and an assortment, the screen displays data showing the system-optimized product list and the original product list, along with several summary metrics.

Table 7-5 Assortment Optimization Summary Metrics

Field Name	Description
SKUs Added	The number of SKUs added to the assortment by the optimization process.
SKUs Dropped	The number of SKUs removed from the assortment by the optimization process.
Amount % Variance	The change in sales retail amount for the assortment after SKUs are added and dropped by the optimization, based on average weekly sales.
Units % Variance	The change in sales units for the assortment after SKUs are added and dropped by the optimization, based on average weekly sales.

The following table lists the summary metrics that the optimized assortment table displays insights into expected product performance if the recommendations are applied to your assortment.

Table 7-6 Optimized Assortment Table

Field Name	Description
Item	The item number and description of the SKU in the assortment.
Description	The recommended change for a SKU in the assortment. This may either be to keep an existing SKU or add a SKU that was not previously in the assortment.
Base Sales Units	The average weekly sales units of the SKU at the selected location.
Halo Sales Units	The average weekly sales units of the SKU at the selected location, after adjusting for demand as a result of Halo sales due to market basket affinity rules involving the SKU.
Base Sales Amount	The average weekly sales amount of the SKU at the selected location.
Halo Sales Amount	The average weekly sales amount of the SKU at the selected location, after adjusting for demand as a result of Halo sales due to market basket affinity rules involving the SKU.
Base Sales Profit Amount	The average weekly sales profit of the SKU at the selected location.
Halo Sales Profit Amount	The average weekly sales profit of the SKU at the selected location, after adjusting for demand as a result of Halo sales due to market basket affinity rules involving the SKU.
Incremental Demand	The number of sales units of the SKU that do not transfer to any other SKU in the assortment if it were to be deleted from the assortment.
Substitutable Demand	The number of sales units of the SKU that can transfer to other SKUs in the assortment if it were to be deleted from the assortment.

The primary purpose of the recommendations is to highlight the optimal mix of products, based on a pre-defined optimization target (which can be sales units, amount, or profit). The optimization process takes into account the product similarities and assortment elasticity (as generated by Demand Transference) to determine which products are candidates for addition or removal from the current assortment. It also uses the market basket affinity rules to identify additional sales generated by the SKUs in the assortment in other product categories. For example, a SKU that belongs to Subclass A, which has a strong affinity with sales of Subclass B, is effectively generating a certain number of additional sales outside of the assortment itself just by being included.

This combination of inputs into the optimization process may result in SKUs being added or dropped from the assortment. A SKU that has been dropped from the assortment may have been found to be too similar to other existing SKUs (thus its removal will transfer the majority of the demand for that SKU to other similar items). A dropped SKU may also have had a low contribution to sales in that assortment, either due to a lack of affinity rules generating additional halo effects, or because other SKUs available for addition provided better results and were swapped in. The system may recommend product additions when it has identified SKUs that will improve the optimization target, such as sales profit. It may also recommend adding a SKU when there are no similar SKUs currently in the assortment, so adding the new SKU will not cannibalize much demand from other items.

If you want to analyze the recommendations further, the system provides pop-up windows displaying Halo and Cannibalization effects relating to the selected item from the Optimized Assortment table. Select a row in the table, and then click the Halo or Cannibalization buttons on the task bar to review the data.

Figure 7-3 Halo Effects Pop-Up

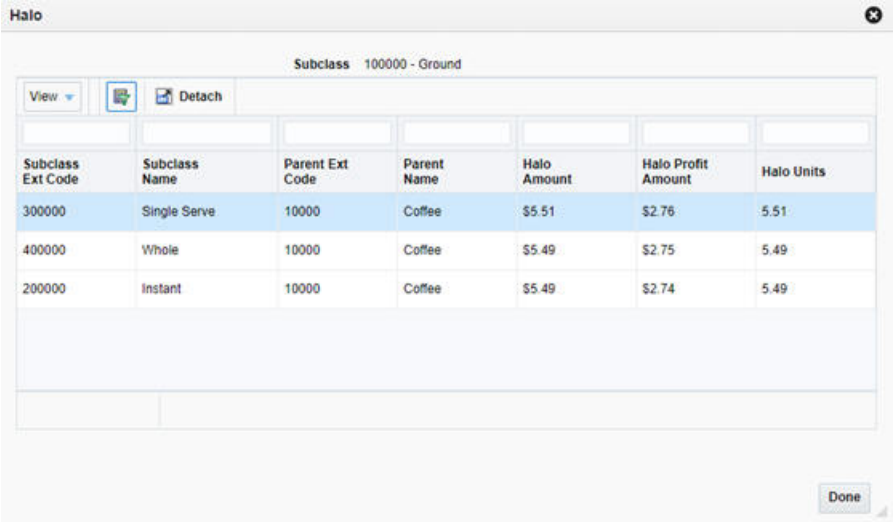


Table 7-7 Halo Effects

Field Name	Description
Subclass Ext Code	The numeric identifier for a subclass with a halo effect from the SKU.
Subclass Name	The descriptive name for a subclass with a halo effect from the SKU.
Parent Ext Code	The numeric identifier of the parent class for a subclass.
Parent Name	The descriptive name for the parent class for a subclass.
Halo Amount	The calculated halo effect on the subclass's sales amount.
Halo Profit Amount	The calculated halo effect on the subclass's sales profit.
Halo Units	The calculated halo effect on the subclass's sales units.

The Halo pop-up window provides insights into the secondary effect of a SKU on other products. Products with strong market basket affinities may generate large halo effects and thus be good candidates for your assortment.

When evaluating the additions and removals on your assortment, it can also be helpful to understand cannibalization effects. These effects are based on the product similarities and assortment elasticities obtained from Demand Transference. Products that are very similar to each other tend to split up the customer demand for this kind of product, since the customer is not likely to buy two similar products. Conversely, products that are not very similar will not have as much overlapping demand. The Cannibalization pop-up window provides some of these insights.

Figure 7-4 Cannibalization Effect Pop-Up

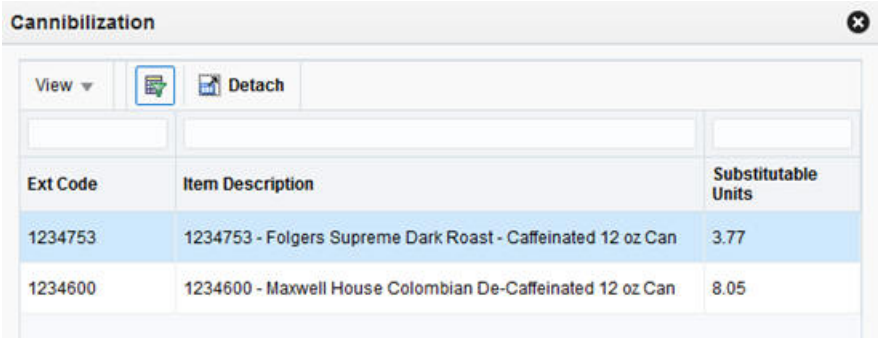


Table 7-8 Cannibalization Effects

Field Name	Description
Ext Code	The numeric identifier for a SKU that is cannibalized by the selected assortment SKU.
Item Description	The descriptive value for the SKU.
Substitutable Units	The calculated number of sales units that can be cannibalized from other SKUs when the assortment SKU is added, or the number of units transferred to these SKUs when the assortment SKU is dropped.

At the bottom of the screen, you can also see the original assortment, along with any SKUs from the assortment that have been dropped in the recommendation.

Figure 7-5 Original Assortment

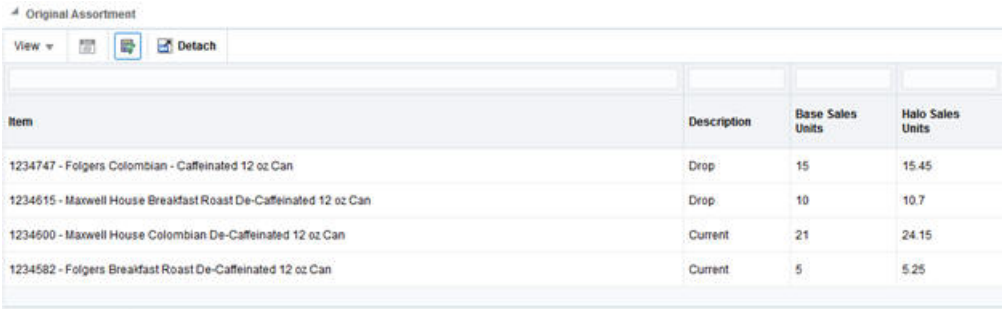


Table 7-9 Original Assortment

Field Name	Description
Item	The item number and description of the SKU in the assortment.
Description	The recommended change for a SKU in the assortment. This may either be to keep an existing SKU or drop a SKU that was in the assortment.

Table 7-9 (Cont.) Original Assortment

Field Name	Description
Base Sales Units	The average weekly sales units of the SKU at the selected location.
Halo Sales Units	The average weekly sales units of the SKU at the selected location, after adjusting for demand as a result of Halo sales due to market basket affinity rules involving the SKU.
Base Sales Amount	The average weekly sales amount of the SKU at the selected location.
Halo Sales Amount	The average weekly sales amount of the SKU at the selected location, after adjusting for demand as a result of Halo sales due to market basket affinity rules involving the SKU.
Base Sales Profit Amount	The average weekly sales profit of the SKU at the selected location.
Halo Sales Profit Amount	The average weekly sales profit of the SKU at the selected location, after adjusting for demand as a result of Halo sales due to market basket affinity rules involving the SKU.

Using Assortment Recommender Results

The assortment recommendations provided within AA aim to achieve the maximum possible value for the chosen target sales measure, such as profit or revenue. Assortment planners may use the results of this process while preparing for the next selling season or fiscal period in order to adjust their current assortments based on their business strategy. Planners may choose to accept the recommendations as-is, or choose a subset of changes in order to see how that will affect their financial targets and assortment plan. This modified assortment can then be provided to AA and reprocessed in the future to refine the recommendations.

The following example describes the workflow of an assortment planner (Anne) working to improve sales in the Coffee category.

1. Anne starts with the existing assortment of coffee items available in stores, along with current performance measures describing how each item is selling. She knows that certain items are performing poorly and wants to mark them for removal from the assortment. She also has a list of new items that can be added to the assortment. All of this information is provided as inputs to AA prior to performing the optimization process.
2. After AA has been provided with the assortments and lists of possible changes that Anne wants to make, the system will schedule and execute the optimization process using pre-defined business rules (such as the maximum number of items that can be added or dropped by the system).
3. Knowing the process executed over the weekend, Anne logs into AA on Monday morning to review the results. She selects a location and her Coffee assortment in the UI prompts, and then begins to analyze the recommendations.
4. AA has selected three items in the current assortment to be dropped and another four items that should be added. Anne first looks at the dropped items and compares them to the other items she was thinking about removing. She notes that the dropped items are not the worst-selling ones she had chosen, but AA shows them as contributing very little to halo effects in other areas, so they may have very weak market basket affinities that contribute to their removal.

5. Anne next reviews the recommendation for items to add to the assortment. Of the four items that AA has chosen, three of them have significant halo effects driven by the market basket affinities. Anne is not as sure about the last item, so she selects it and clicks on the Cannibalization button. She notes that the item has relatively low substitutable demand from other items, suggesting it is not very similar to anything in her current assortment. This makes it a good candidate for addition, as it can bring in new demand that the assortment may not have today.
6. Anne decides to accept all of the recommendations made by AA for this assortment, exports the results to Excel for later reference, and then exits the system to make the necessary changes to her assortment plan.

Market Basket Analysis

The market basket analysis screens of the application all relate to the market basket affinity rules generated by AA. The tiles and screens provide a way to view different sets of data, depending on the configurations used for your business.

Market basket rules and their associated metrics are a key insight into product demand and customer buying behaviors. Understanding which products are more likely to sell together allow your business to coordinate targeted offers and promotions for products that have strong affinities elsewhere, without having to promote those other categories. It is also possible to understand market basket differences between customer segments, allowing you to further refine your sales and promotion strategies to target specific groups of customers that will bring in the most value to your business.

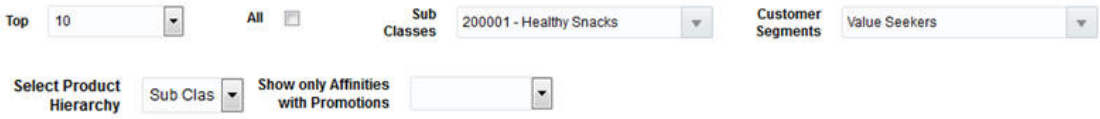
There are five screens available for viewing the AA results, each containing a different subset of affinity rules with certain characteristics as described below.

Table 7-10 Market Basket

Field Name	Description
Class	Displays market basket affinities between different classes, without any consideration for whether the items were on promotion or not.
Subclass	Displays market basket affinities between different subclasses, without any consideration for whether the items were on promotion or not.
Promotion	Displays market basket affinities between different subclasses when the antecedent (if) components are promoted and the consequent (then) components are not promoted.
Customer Segment	Displays market basket affinities between different subclasses when the antecedent (if) components are promoted and the consequent (then) components are not promoted, further separated by customer segments.
All	Displays a searchable list of all available affinity rules for all types.

After you select one of the five info tiles at the top of the screen, you will be presented with additional filter prompts and a data table for displaying results.

Figure 7-6 Market Basket Filters



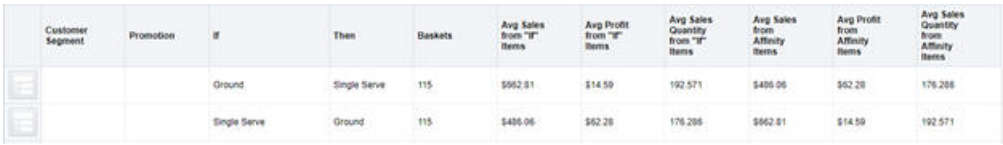
The filters above the table allow you to limit the data displayed in the table. AA can generate a large number of rules, and usually only the most common rules supported by the largest number of sales transactions are used in analyses. Which filters are displayed will depend on the current screen being examined. The following table provides an explanation of how all the filters impact the displayed results.

Table 7-11 Market Basket Filters

Filter Name	Description
Top	Selecting a different number will change how many market basket rules are displayed in the results table.
All	Selecting the check box overrides the Top filter to instead display all available market basket rules.
Classes/Subclasses	Select specific classes or subclasses for which market basket rules should be displayed. This filter will be applied as an OR condition for either the IF components or the THEN components. If there are multiple IF components, then a match is only required for one of the components.
Customer Segments	Select specific customer segments for which market basket rules should be displayed.
Select Product Hierarchy	(All Screen Only) This prompt is used to choose which level of the hierarchy you are entering search criteria for.
Show only Affinities with Promotions	(All Screen Only) This prompt allows you to choose whether the results should be limited to promoted subclass affinities.

Once you have selected your desired filter settings, you may begin to review the market basket affinities. The main table on each screen describes the product affinities that were found when analyzing the market basket data, sorted in descending order by your chosen performance metric (the Top 10 By global prompt described earlier in this chapter). In addition to the product affinities, numerous metrics are provided to describe the market basket data that contains these products.

Figure 7-7 Market Basket Affinities



Customer Segment	Promotion	If	Then	Baskets	Avg Sales from "If" Items	Avg Profit from "If" Items	Avg Sales Quantity from "If" Items	Avg Sales from Affinity Items	Avg Profit from Affinity Items	Avg Sales Quantity from Affinity Items
		Ground	Single Serve	115	$662.81	$14.59	192.571	$486.06	$62.28	176.286
		Single Serve	Ground	115	$486.06	$62.28	176.286	$662.81	$14.59	192.571

Table 7-12 Market Basket Affinities

Field Name	Description
Customer Segment	(Customer Segment screen only) Displays the customer segment for which the product affinity was found.
Promotion	(All affinities screen only) Displays whether the rule represents IF components that were on promotion.
If	Displays the classes or subclasses containing items that were purchased in the market baskets for a given product affinity. All market baskets that contributed to a rule will contain items from these product categories.
Then	Displays the class or subclass that was found to have an affinity with the IF components of the market basket rule. Market baskets containing the IF categories were found to contain items from the THEN category enough times to establish a product affinity between the IF and THEN components.
Baskets	The number of market baskets that contain items from both the IF and THEN components of the rule.
Avg Sales from If Items	The average sales retail amount from items belonging to the IF classes or subclasses.
Avg Profit from If Items	The average sales profit from items belonging to the IF classes or subclasses.
Avg Sales Quantity from If Items	The average sales units from items belonging to the IF classes or subclasses.
Avg Sales from Affinity Items	The average sales retail amount from items belonging to the THEN class or subclass.
Avg Profit from Affinity Items	The average sales profit from items belonging to the THEN class or subclass.
Avg Sales Quantity from Affinity Items	The average sales units from items belonging to the THEN class or subclass.

The purpose of these summary metrics is to provide a quick view of the sales and profitability of the market baskets described by each rule. Even though a market basket affinity is present in a large number of baskets, it may not generate much profit or be of particular interest to the business user at this time. The summary metrics allow the user to quickly identify market basket affinities that are worth analyzing in more detail. Once the user has selected a rule to investigate, clicking that row in the table generates additional information about it in a panel to the right side of the screen.

Figure 7-8 Market Basket Affinity Details

Table 7-13 Market Basket Affinity Details

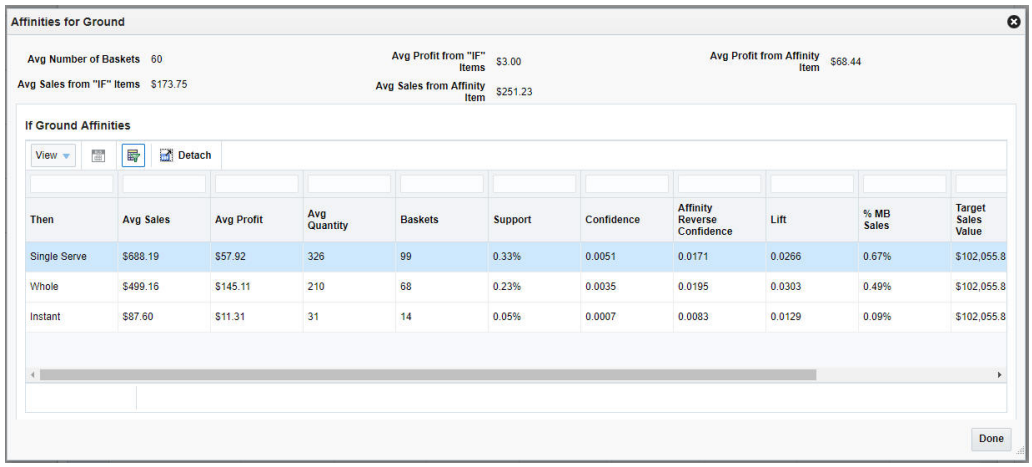
Field Name	Description
Support	The percent of market baskets that contain items from the IF and THEN components of the rule. Written as a formula: Rule Transaction Count/Total Transaction Count
Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing IF items. Written as a formula: Rule Transaction Count/Total IF Transaction Count
Reverse Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing THEN items. If a rule has a higher Reverse Confidence than its own Confidence value, then it indicates that the rule is better represented by the reversal of the IF and THEN components. Written as a formula: Rule Transaction Count / Total THEN Transaction Count
Lift	A statistical measure of how strong the product affinity is, as compared to random chance for the rule (higher is better). Written as a formula: Rule Transaction Count * Total Transaction Count / Total IF Transaction Count/Total THEN Transaction Count
Sales Amt Pct Of Total	The percent of total sales retail amount that the rule represents. Written as a formula: Rule Sales Amount/Average Weekly Total Sales
Avg Rule Sales Amt	The average weekly sales retail amount across all transactions containing the rule.
Avg Rule Sales Profit Amt	The average weekly sales profit across all transactions containing the rule.
Avg Rule Sales Qty	The average weekly sales units across all transactions containing the rule.
Avg Total Sales Amt	The average weekly sales retail amount across all transactions.
Avg Tot Sales Profit Amt	The average weekly sales profit across all transactions.
Avg Tot Sales Qty	The average weekly sales units across all transactions.
Avg Tot Txn Cnt	The average weekly transaction count across all transactions.

The detailed metrics for a rule provide more ways to evaluate if the rule is going to be useful for your business purpose. For example, a rule that has both high support and high

confidence represents a strong product affinity that is purchased by many of your customers. Any action taken on the categories in such a rule, such as promoting a subclass on the IF side of the rule, can have significant benefits for the THEN components of the rule. Another way to use these metrics is to consider the average sales or profit that a rule brings in relative to total sales over the same period. Rules that represent a significant profit amount may be a more valuable target for your efforts than rules with lower profitability.

Another way to view the market basket affinities is to click the button found to the left of each row in the data table. This button displays a pop-up window with all of the product affinities that have been found for the IF components of the rule. This can be especially useful when your goal is to find any relationships with a specific product category, rather than seeing all of the relationships between different products.

Figure 7-9 Targeted Product Affinities Pop-Up



This screen is split into two main sections: the top section displays summary metrics for all rules associated with the IF components, and the bottom section displays all of the product affinities found for the IF components and a selection of detailed metrics for them.

Table 7-14 Targeted Product Affinities Summary Metrics

Field Name	Description
Avg Number of Baskets	The average number of baskets found on a rule containing the IF items.
Avg Sales from IF Items	The average sales retail amount for the IF items in the rules.
Avg Profit from IF Items	The average sales profit for the IF items in the rules.
Avg Sales from Affinity Item	The average sales retail amount for the THEN items in the rules.
Avg Profit from Affinity Item	The average sales profit for the THEN items in the rules.

Table 7-15 Targeted Product Affinities Detail Metrics

Field Name	Description
Then	Displays the class or subclass that was found to have an affinity with the IF components of the market basket rule.
Avg Sales	The average weekly sales retail amount for the IF and THEN items in the rules.
Avg Profit	The average weekly sales profit for the IF and THEN items in the rules.
Avg Quantity	The average weekly sales units for the IF and THEN items in the rules.
Baskets	The number of market baskets that contain items from the IF and THEN components of the rule.
Support	The percent of market baskets that contain items from the IF and THEN components of the rule.
Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing IF items.
Affinity Reverse Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing THEN items.
Lift	A statistical measure of how strong the product affinity is, as compared to random chance for the rule (higher is better).
% MB Sales	The percent of total sales retail amount that the rule represents.
Target Sales Value	The average weekly total sales retail amount across all transactions in the selected periods.
Target Profit	The average weekly total sales profit across all transactions in the selected periods.
Target Sales Quantity	The average weekly total sales units across all transactions in the selected periods.

This view of market basket affinities provides powerful insight into the relationships between different product categories, as it may expose affinities even between seemingly unrelated categories that your customers tend to purchase together. This screen allows you to see how actions taken on the target (IF) products will impact other areas within your business. For example, any action taken to increase the sales of an IF category can lift the sales of all other products with a strong affinity to those items. Conversely, if you have a category that is performing poorly on the THEN side of a rule, you may be able to take action on the IF products to boost that category's sales indirectly.

Using Market Basket Rules

There are a variety of ways to take action within your business based on the insights that AA provides. The most common use is to develop a promotion strategy that takes advantage of product affinities to maximize halo effects across other categories. By reducing the amount spent on promotions while increasing the effect of those promotions, you can realize additional revenue and boost profit margins. For example, by analyzing the AA results, it is possible to identify multiple product categories with affinities on a target category, where some categories are more profitable than others. This can provide alternatives for promoting items that will yield more in sales or profit than a more obvious affinity like Hot Dogs and Hot Dog Buns.

Another common practice is for store planners and planogrammers to take advantage of product affinities when deciding how to arrange the products on the shelf or for adding aisle end-caps for strongly associated product categories. AA can show obvious relationships like Bread and Peanut Butter, but it may also reveal previously unknown associations like Pet Food and Beer. Using market basket rules to inform store layouts allows you to place commonly purchased groups of products close together in the store, increasing the chance of customers purchasing more from those categories.

Combining market basket analysis with customer segments further enhances the potential benefits of AA. Customer behavior information is obtained from mining transaction history, and it is correlated with customer segment attributes to inform your promotion strategies. The ability to understand market basket affinities allows marketers to calculate, monitor, and build promotion strategies based on critical metrics such as customer profitability and preferred categories.

As an example, consider a marketer that is planning promotions for the Soda and Chips categories. While analyzing the data in Affinity Analysis, the marketer finds that a particular customer segment, College Singles, has a very strong affinity for buying Soda any time they purchase Chips.Armed with this knowledge, the marketer may decide to create a targeted offer for this customer segment only on the Chips category, knowing that they are likely to also buy Soda in the same basket. This kind of offer is more cost effective than a company-wide promotion on both the chips and soda categories, but may yield almost the same results, because the offer has focused in on the customers that will generate the most revenue in these categories.

Affinity Analysis

This section addresses MBI analysis. The functionality described here includes creating custom runs and comparing the results of multiple runs.

System runs are executed automatically once a week as part of the weekly batch cycle using settings that have been configured for a specific level of the location hierarchy. Since system runs occur weekly, they can be used to examine results over a long period of time.

User runs can be run at any time by an individual user who specifies the settings for the run.

Two types of comparison are available: Compare Runs and Compare Results.

Use Compare Runs to select specific runs to compare. This comparison is useful for straightforward runs such as comparing the results of two different months.

Use Compare Results to compare system runs only, with specified dates and other filters. In this case, take care in selecting runs that overlap, as some runs may be represented in the results more than once.

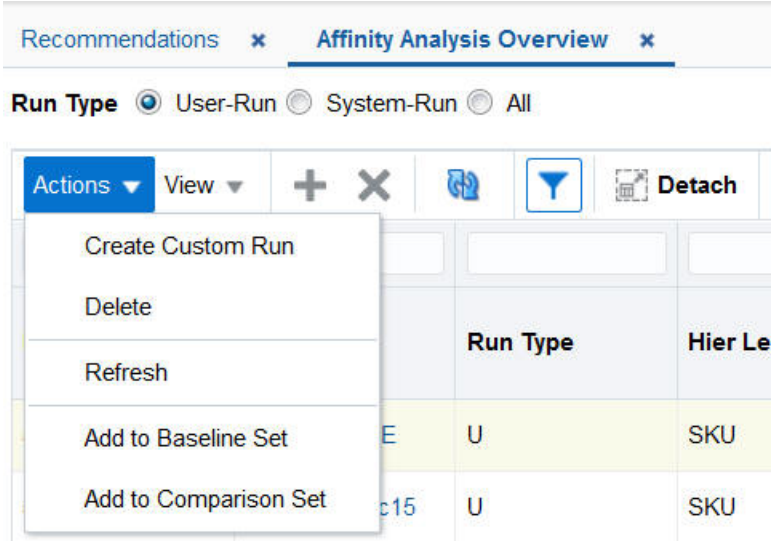
Affinity Analysis Overview

The Affinity Analysis Overview screen provides the tools in the following images. In addition to the standard application functionality, you can also

- Select from the following functionality using the Action drop-down menu:
  - Create Custom run

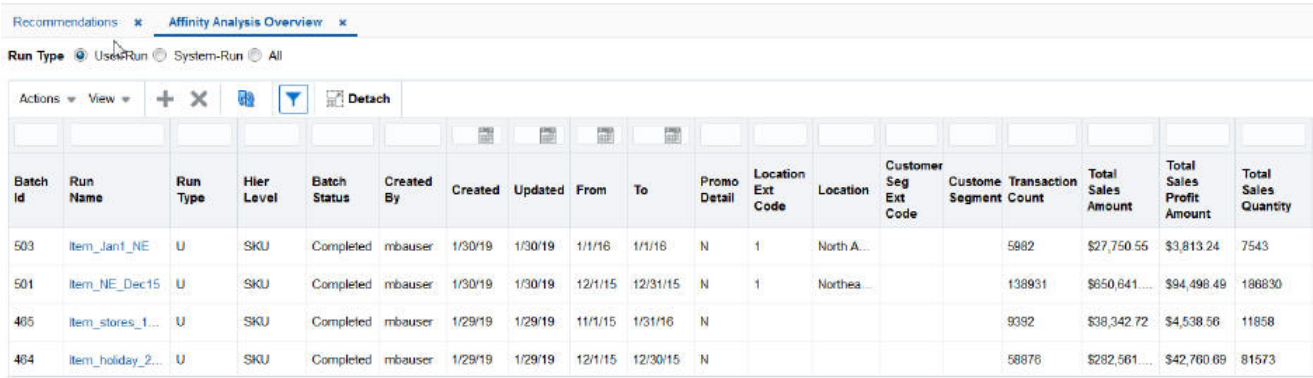
  - Delete
  - Refresh
  - Add to Baseline Set
  - Add to Comparison Set
- Select the type of run to view in this screen: User-Run, System-Run, or all runs.

Figure 7-10 Affinity Analysis Overview Functionality



The Affinity Analysis Overview table, shown in the following figure, displays the list of completed runs and contains the following fields.

Figure 7-11 Affinity Analysis Overview Table



Batch Id	Run Name	Run Type	Hier Level	Batch Status	Created By	Created	Updated	From	To	Promo Detail	Location Ext Code	Location	Customer Seg Ext Code	Custome Segment	Transaction Count	Total Sales Amount	Total Sales Profit Amount	Total Sales Quantity
503	Item_Jan1_NE	U	SKU	Completed	mbauser	1/30/19	1/30/19	1/1/16	1/1/16	N	1	North A...			5982	$27,750.55	$3,813.24	7543
501	Item_NE_Dec15	U	SKU	Completed	mbauser	1/30/19	1/30/19	12/1/15	12/31/15	N	1	Northea...			138931	$650,641....	$94,498.49	186830
465	Item_stores_1...	U	SKU	Completed	mbauser	1/29/19	1/29/19	11/1/15	1/31/16	N					9392	$38,342.72	$4,538.56	11858
464	Item_holiday_2...	U	SKU	Completed	mbauser	1/29/19	1/29/19	12/1/15	12/30/15	N					58876	$282,561....	$42,760.69	81573

Table 7-16 Affinity Analysis

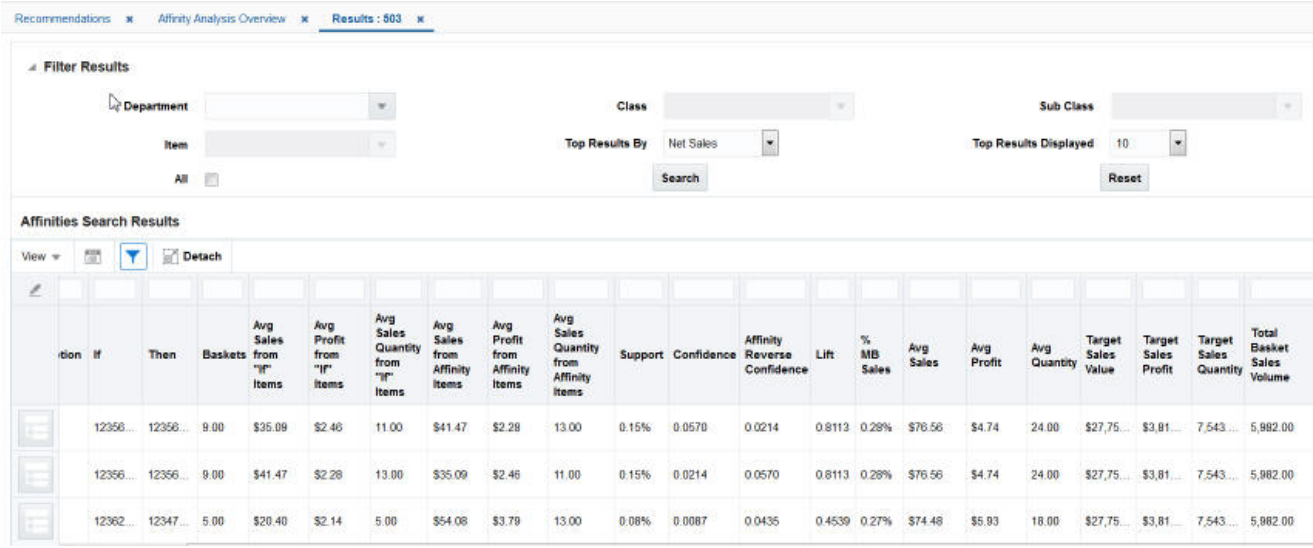
Field Name	Description
Batch ID	Identifies the run.
Run Name	User-provided name for the run.
Run Type	U = User -Run; S = System-Run.
Hier Level	Hierarchy level.
Batch Status	Indicates the status of the run. Helps to determine if the run has completed, encountered any errors, or has not yet completed.
Created By	User name of person who created the run.
Created	Date when run created.
Updated	Date when run last updated.
From	Beginning date for run data collection.
To	Ending date for run data collection.
Promo Detail	Y indicates that the market basket affinities between different subclasses are displayed when the antecedent (if) components are promoted and the consequent (then) components are not promoted. N indicates that no consideration is given to whether or not the items are promoted.
Location Ext Code	The numeric identifier for the location.
Location	The external identifier for the location.
Customer Seg Ext Code	The external identifier for the customer segment.
Customer Segment	The name of the customer segment which the results are for. Values include Soccer Mom, Simply Seniors, Value Seekers, Gourmet Shoppers, Livin Large.
Transaction Count	Total number of transactions processed for this run.
Total Sales Amount	Total sales amount processed for this run.
Total Sales Profit Amount	Total sales profit amount processed for this run.
Total Sales Quantity	Total sales quantity processed for this run.

When you click on a specific **Run Name**, you see a pop-up that displays Filter Results and Affinities Search Results.

Filter Results

The following figure shows the Filter Results and Affinities Search Results, which are displayed when you click on a specific Run Name. The fields are described in following tables. **Search** button to initiate the filtering process and use the **Reset** button to clear the filter fields.

Figure 7-12 Filter Results and Affinities Search Results Table



The following table describes the options you can use to filter the results that will be displayed.

Table 7-17 Filter Results

Field	Description
Department	Select one or more departments for which the results should be filtered.
Item	Select the item within the department from the drop-down list.
All	Select to prevent any other filters from being used.
Class	Select one or more classes for which the results should be filtered.
Top Results By	Select a metric to see the top results by: Net Sales, Net Quantity, of Net Profit.
Sub Class	Select one or more subclasses for which the results should be filtered.
Top Results Displayed	Select a value (10, 20, 30...970, 980) to limit the number of results displayed.

Affinities Search Results

The Affinities Search Results section of the pop-up provides details about the observed associations (affinities) between sales transactions. It is expressed as "If a customer buys X, then that customer may also buy Y." The table describes the relevant values associated with the If/Then transactions and contains the following fields:

Table 7-18 Affinities Search Results

Field	Description
Customer Segment	(Customer Segment screen only) Displays the customer segment for which the product affinity was found.

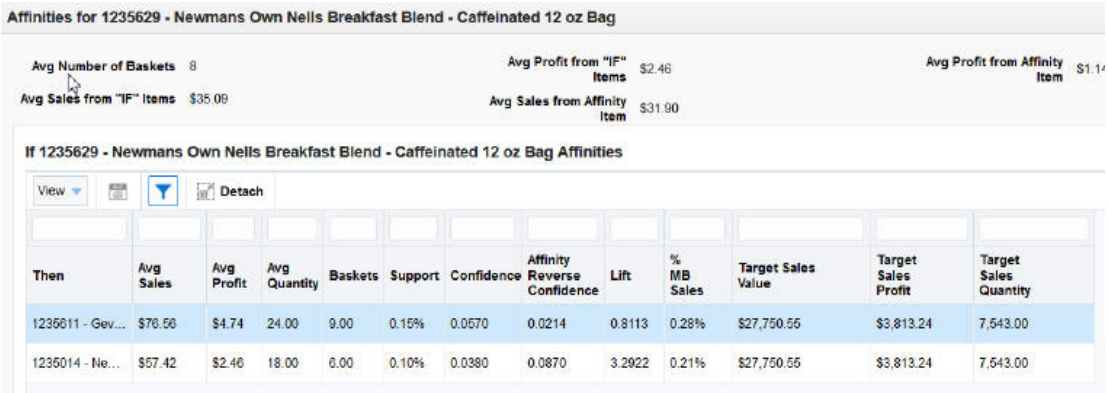
Table 7-18 (Cont.) Affinities Search Results

Field	Description
Promotion	(All affinities screen only) Displays whether the rule represents IF components that were on promotion.
If	Displays the classes or subclasses containing items that were purchased in the market baskets for a given product affinity. All market baskets that contributed to a rule will contain items from these product categories.
Then	Displays the class or subclass that was found to have an affinity with the IF components of the market basket rule.
Baskets	The number of market baskets that contain items from both the IF and THEN components of the rule.
Avg Sales from "If" Items	The average sales retail amount from items belonging to the IF classes or subclasses.
Avg Profit from "If" Items	The average sales profit from items belonging to the IF classes or subclasses.
Avg Sales Quantity from "If" Items	The average sales units from items belonging to the IF classes or subclasses.
Avg Sales from Affinity Items	The average sales retail amount from items belonging to the THEN class or subclass.
Avg Profit from Affinity Items	The average sales profit from items belonging to the THEN class or subclass.
Avg Sales Quantity from Affinity Items	The average sales units from items belonging to the THEN class or subclass.
Support	The percent of market baskets that contain items from the IF and THEN components of the rule.
Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing IF items.
Affinity Reverse Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing THEN items.
Lift	A statistical measure of how strong the product affinity is, as compared to random chance for the rule (higher is better).
% MB Sales	The percent of total sales retail amount that the rule represents.
Avg Sales	The average weekly sales retail amount for the IF and THEN items in the rules.
Avg Profit	The average weekly sales profit for the IF and THEN items in the rules.
Avg Quantity	The average weekly sales units for the IF and THEN items in the rules.
Target Sales Value	The average weekly total sales retail amount across all transactions in the selected periods.
Target Sales Profit	The average weekly total sales profit across all transactions in the selected periods.
Target Sales Quantity	The average weekly total sales units across all transactions in the selected periods.
Total Basket Sales Volume	The average weekly transaction count across all transactions.

Results Pop-up

Click the right-hand icon for a specific row in the Results table to see If/Then Affinities for...pop-up. The pop-up shows all the other rules that have been discovered for the same If item. You can use this to compare the rules. In this way, you may discover another product with a stronger affinity for the item.

Figure 7-13 Affinities For



Then	Avg Sales	Avg Profit	Avg Quantity	Baskets	Support	Confidence	Affinity Reverse Confidence	Lift	% MB Sales	Target Sales Value	Target Sales Profit	Target Sales Quantity
1235611 - Gev...	$76.56	$4.74	24.00	9.00	0.15%	0.0570	0.0214	0.8113	0.28%	$27,750.55	$3,813.24	7,543.00
1235014 - Ne...	$57.42	$2.46	18.00	6.00	0.10%	0.0380	0.0870	3.2922	0.21%	$27,750.55	$3,813.24	7,543.00

The pop-up has the following fields.

Table 7-19 Affinities For

Field	Description
Avg Number of Baskets	The average number of baskets across the runs for the combination of the If and Then items.
Avg Sales from "If" Items	The average sales retail amount from items belonging to the IF classes or subclasses.
Avg Profit from "If" Items	The average sales profit from items belonging to the IF classes or subclasses.
Avg Sales from Affinity Item	The average sales retail amount from item belonging to the THEN class or subclass.
Avg Profit from Affinity Item	The average sales retail amount from item belonging to the THEN class or subclass.
Then	Displays the class or subclass that was found to have an affinity with the IF components of the market basket rule.
Avg Sales	The average weekly sales retail amount for the IF and THEN items in the rules.
Avg Profit	The average weekly sales profit for the IF and THEN items in the rules.
Avg Quantity	The average weekly sales units for the IF and THEN items in the rules.
Baskets	The number of market baskets that contain items from both the IF and THEN components of the rule.

Table 7-19 (Cont.) Affinities For

Field	Description
Support	The percent of market baskets that contain items from the IF and THEN components of the rule.
Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing IF items.
Affinity Reverse Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing THEN items.
Lift	A statistical measure of how strong the product affinity is, as compared to random chance for the rule (higher is better).
% MB Sales	The percent of total sales retail amount that the rule represents.
Target Sales Value	The average weekly total sales retail amount across all transactions in the selected periods.
Target Sales Profit	The average weekly total sales profit across all transactions in the selected periods.
Target Sales Quantity	The average weekly total sales units across all transactions in the selected periods.

Panel

The side panel has two sections. To display the side panel, you must first select a row in the table. The top section reports some of the most useful values that are also displayed in the table. The bottom section explains the details regarding the run you are reviewing.

Figure 7-14 Panel Top

If 1234816 - Tully's French Roast Dar

Support	0.13%
Confidence	0.0741
Reverse Confidence	0.0945
Lift	5.4780
Sales Amt Pct Of Total	0.45%
Avg Rule Sales Amt	$174.42
Avg Rule Sales Profit Amt	$0.00
Avg Rule Sales Qty	70.00
Avg Total Sales Amt	$38,342.72
Avg Tot Sales Profit Amt	$4,538.56
Avg Tot Sales Qty	11,858.00
Avg Tot Txn Cnt	9,392.00

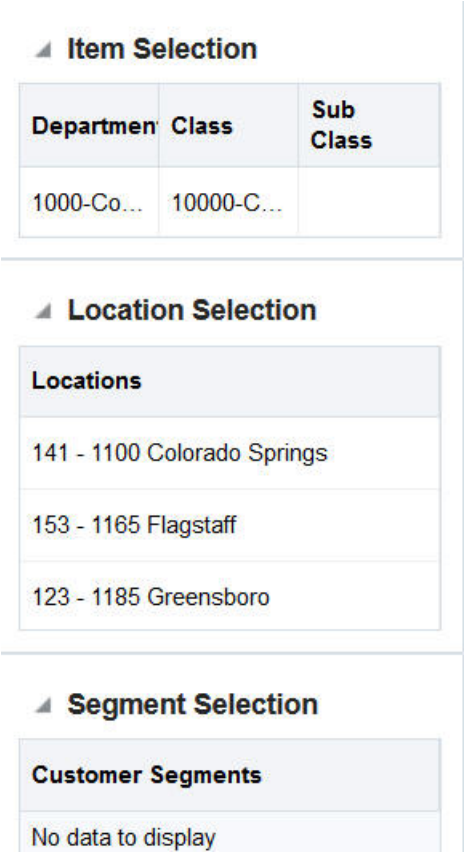
Figure 7-15 Panel Middle

If 1234816 - Tully's French Roas

Results Parameters

Minimum Support	0.02%
Minimum Confidence	0.0000
Minimum Reverse Confidence	0.0000
Minimum Lift	0.0001
Minimum Transactions	2.00
Maximum Number of Rules	9,999.00
Maximum Rule Size	2.00
Date From	11/1/2015
Date To	1/31/2016

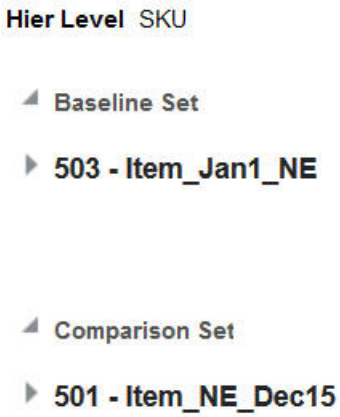
Figure 7-16 Panel Bottom



Compare Results

To compare results, select the two runs you want to compare. Highlight one run and click **Add to Baseline Set**. Highlight the second run and click **Add to Comparison Set**. You see these selections in the Compare Results right panel.

Figure 7-17 Panel Showing Run Selection



Click **View** to see the comparison. the View Results pop-up displays, which can be filtered by a combination of Department Item, Class, Top Results By, Subclass, and Top Results Displayed.

The results are displayed as a pair of rows for each item, a baseline row followed by a comparison row.

Table 7-20 Compare Results

Field	Description
If	Displays the classes or subclasses containing items that were purchased in the market baskets for a given product affinity. All market baskets that contributed to a rule will contain items from these product categories.
Then	Displays the class or subclass that was found to have an affinity with the IF components of the market basket rule.
Set	Indicates whether the row of data is for the baseline runs or the comparison runs.
Set Size	The number of Item/Then components in the affinity rule. For example, "If Coffee and Creamer Then Sugar" has a Set Size of 3.
Baskets	The number of market baskets that contain items from both the IF and THEN components of the rule.
Avg Sales from Affinity Items	The average sales retail amount from items belonging to the THEN class or subclass.
Avg Profit from Affinity Items	The average sales profit from items belonging to the THEN class or subclass.
Avg Sales Quantity from Affinity Items	The average sales units from items belonging to the THEN class or subclass.
Avg Sales from "If" Items	The average sales retail amount from items belonging to the IF classes or subclasses.
Avg Profit from "If" Items	The average sales profit from items belonging to the IF classes or subclasses.
Avg Sales Quantity from "If" Items	The average sales units from items belonging to the IF classes or subclasses.
Support	The percent of market baskets that contain items from the IF and THEN components of the rule.
Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing IF items.
Affinity Reverse Confidence	The ratio of the number of transactions where the entire rule is present, compared to all transactions containing THEN items.
Lift	A statistical measure of how strong the product affinity is, as compared to random chance for the rule (higher is better).
% MB Sales	The percent of total sales retail amount that the rule represents.
Avg Sales	The average weekly sales retail amount for the IF and THEN items in the rules.
Avg Profit	The average weekly sales profit for the IF and THEN items in the rules.
Avg Quantity	The average weekly sales units for the IF and THEN items in the rules.

Table 7-20 (Cont.) Compare Results

Field	Description
Target Sales Value	The average weekly total sales retail amount across all transactions in the selected periods.
Target Sales Profit	The average weekly total sales profit across all transactions in the selected periods.
Target Sales Quantity	The average weekly total sales units across all transactions in the selected periods.
Total Basket Sales Volume	The average weekly transaction count across all transactions.

Create Custom Run

To create a custom run, select **Create Custom Run** from the Action drop-down. You see the Create Custom Run pop-up with two pages - Item Selection and Results Parameters. Need more info about recommendations to create a run.

Use the Item Selection page to select the items to include in the results, including the level of the results and the hierarchy nodes to include.

Provide a meaningful name for the **Run Name** that can help identify the run.

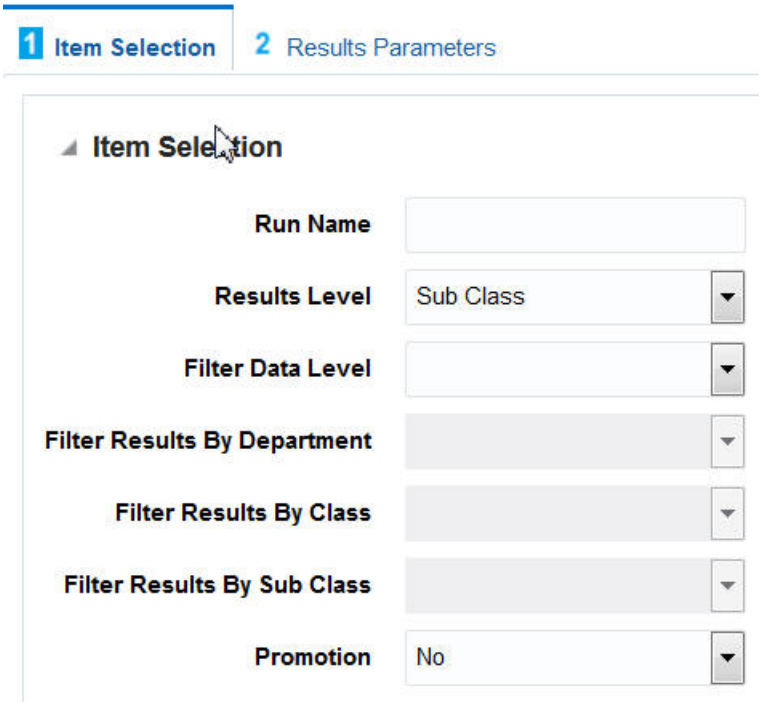
The **Results Level** defines the level of the hierarchy that the data is created for. Lower levels can require lower thresholds on the Results Parameters page (in order to find results), while higher levels can include too many results if the settings are not increased on the Results Parameters page.

The **Filter Data Level** is used to define rules regarding how to filter data. For example, if the results are only required for specific departments, then select Departments as the level and then select the departments.

A Promotion value of **Yes** only includes results in which the If component was on a promotion. A value of **No** is used when there is no concern about whether a promotion was present or not.

After entering item criteria, click **Add Item Selection**. The bottom section of the screen is populated with the results of the selection. You can also see this information in the right panel.

Figure 7-18 Item Selection Tab



The Results Parameters page, which is shown in Figure 7-19, is displayed in the right panel. The minimum value for transactions is used to prevent a situation in which there are too many results because so many transactions have been used in the calculations.

Figure 7-19 Results Parameters Tab

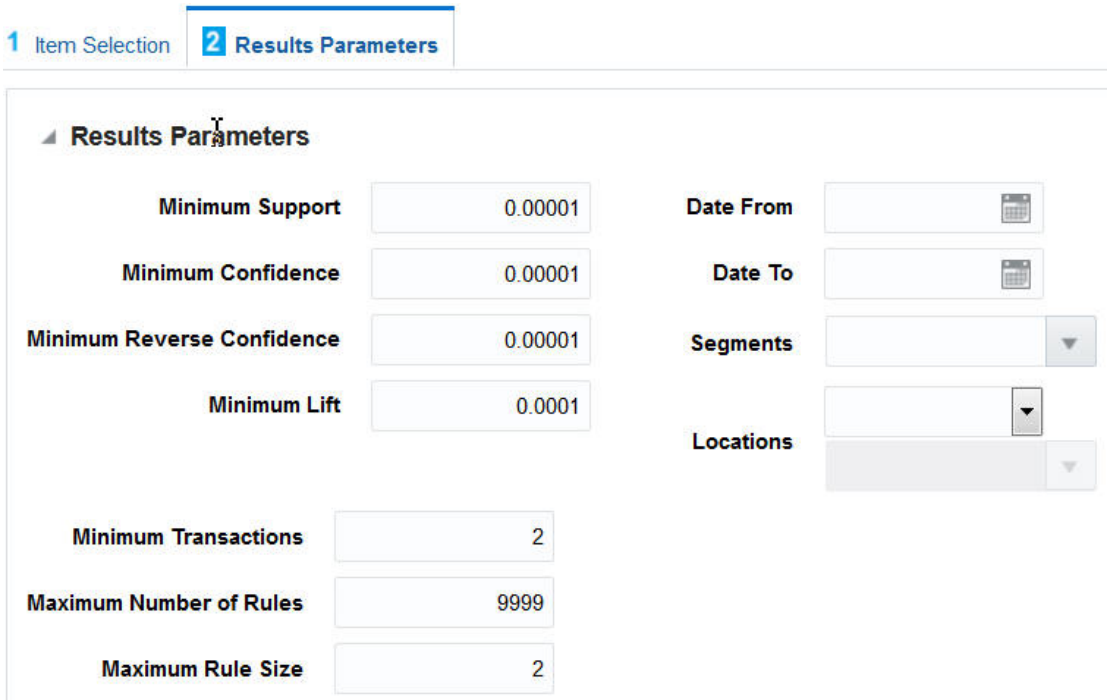


Table 7-21 explains how to interpret these values.

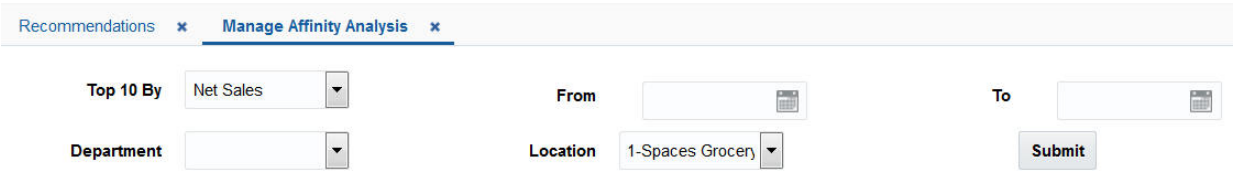
Table 7-21 Results Parameters

Field	Description
Minimum Support	The minimum percent of total transactions required to have this rule expressed as a decimal value. (0 - 1 range)
Minimum Confidence	The minimum confidence ratio required for a rule expressed as a decimal value. (0 - 1 range)
Minimum Reverse Confidence	The minimum confidence ratio required for the reverse rule definition, expressed as a decimal value. (0 - 1 range)
Minimum Lift	The minimum lift value required for a rule.
Minimum Transactions	The minimum value for transactions are used to prevent a situation in which there are too many results because too many transactions have been used in the calculations. This value also helps ensure some consistency if the transaction volume varies enough for Minimum Support to be used.
Maximum Number of Rules	The limit on the top number of rules to be saved from the results. Used to reduce data volume and remove the less significant rules.
Maximum Rule Size	The number of components to be included in a rule. The value of 2 is only supported for rules that include promotion or item results; otherwise, the value of 3 is supported. The additional components are always on the If side of the rule.
Dates From and To	The date range for processing sales data. This value must not span an excessive length of time, as the amount of data processed can become too large to manage reasonably.
Segments	Select any customer segments that should be included when determining the sales transactions to process.
Location	Limit the locations that should be included when determining the sales transactions to process. This can be done by selecting any hierarchy level and then multi-selecting any locations at that level.

Manage Affinity Analysis

Use the Manage Affinity Analysis screen to view the results from all system created runs. Set the filters, select a range of calendar dates, and click **Submit**. The Info Tiles are displayed.

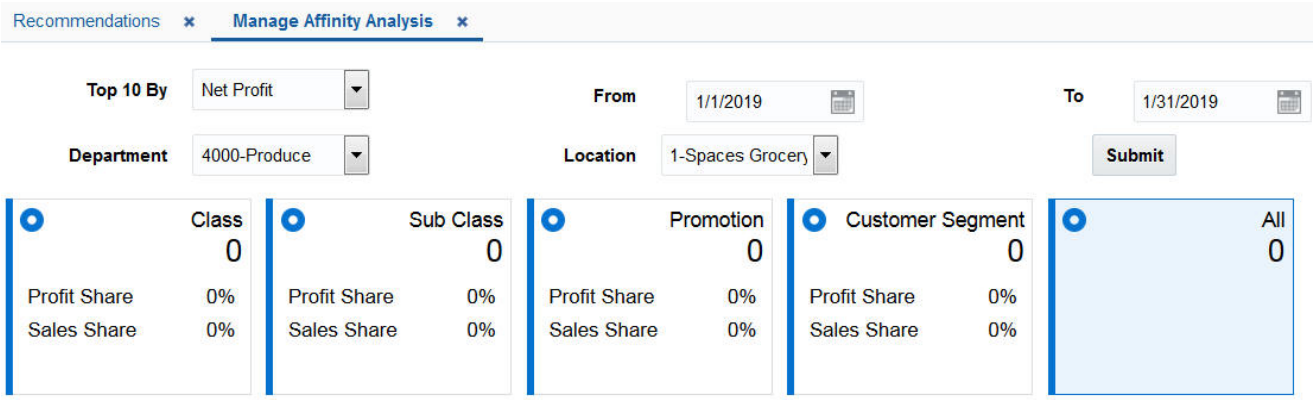
Figure 7-20 Manage Affinity Analysis Filters



Info Tiles

The Info Tiles display the results from all system runs. Once you specify the filters to indicate what runs you want to use, the Info tiles display an aggregate view of the results for different hierarchy levels and the types of processing that were executed during the run. Select an Info Tile and use the filters below the tiles to view the affinities for your selection.

Figure 7-21 Info Tiles



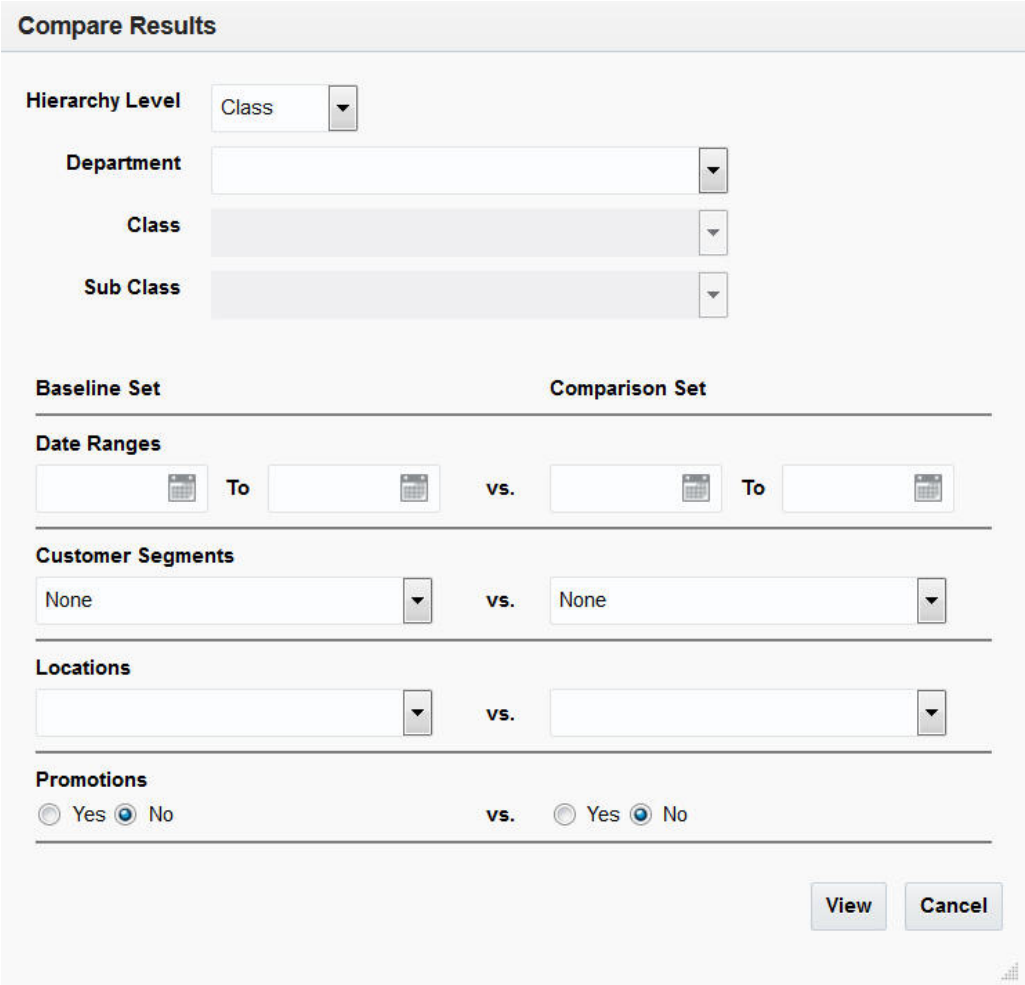
Each tile has specific filters. Here is an example of the filters for the All tile.

Figure 7-22 All Info Tile Filters

Compare Results

You can use Compare Results to compare system runs only, using the dates and filters you specify, as shown in Figure 7-23.

Figure 7-23 Compare Results



Click **View**. Table 7-20 describes the results of the Compare Results functionality.

8
Size Profiles

This chapter describes the Size Profiles Cloud Service module.

Overview

The Size Profiles (SP) is a module under Retail Profile Sciences and is used to estimate the distribution of demand across different sizes (size profile) for different merchandise and location levels.

Size profile is estimated at different levels of merchandise and location. The lowest level of estimation is style-color (for merchandise) and store (for location). Size profiles may also be estimated at higher levels of aggregation on both the merchandise and location dimensions (for example, at subclass-store or style/color-region), depending on the retailer's requirements.

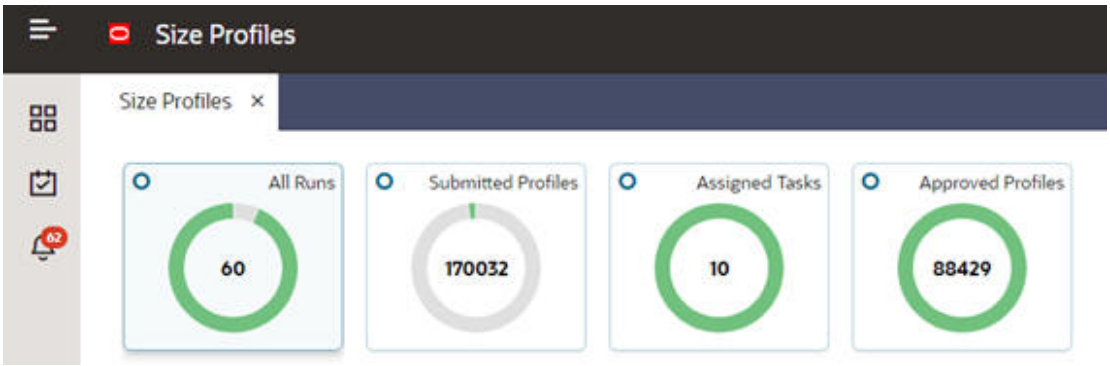
At a high level, the workflow for SP consists of creating a run, reviewing the size profiles, and submitting and approving them. The approved profiles are exported once the export process runs.

Navigation

You can use the Info tiles at the top of the screen, shown in the following image, to navigate to:

- All Runs, which displays all previously created runs and is where you can create new runs.
- Submitted Profiles, which displays the submitted profiles across all runs. It is possible to have multiple submitted profiles for the same merchandise/location/attribute/size range/season from different runs.
- Assigned Tasks, which displays all profiles assigned to the user for review.
- Approved Profiles, which displays the approved profiles that are already exported. Each merchandise/location/attribute/size range/season can have at most one approved profile.

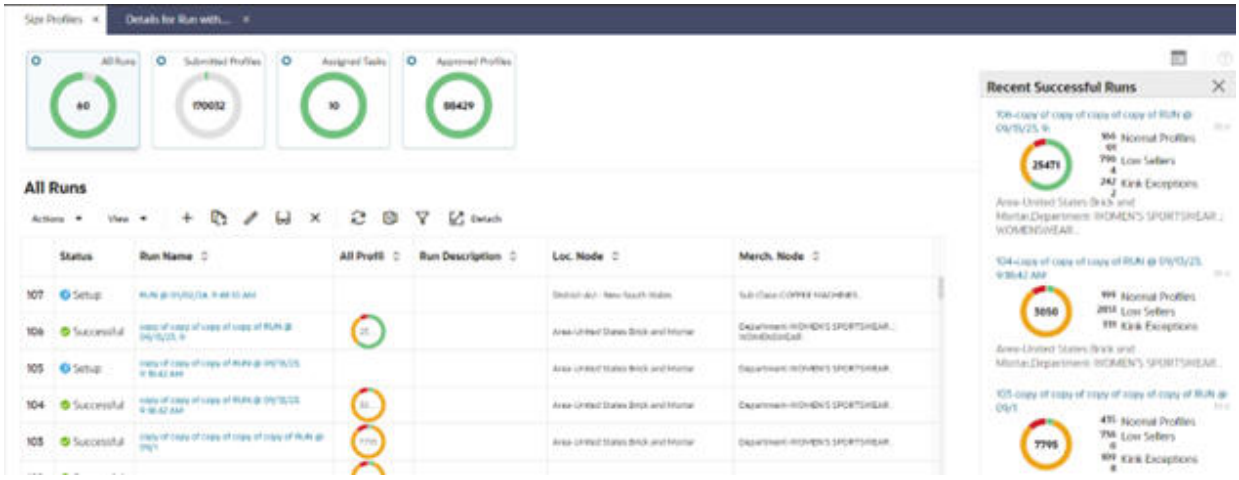
Figure 8-1 Info Tiles



All Runs

The All Runs screen lists all previously created runs, regardless of their status.

Figure 8-2 All Runs



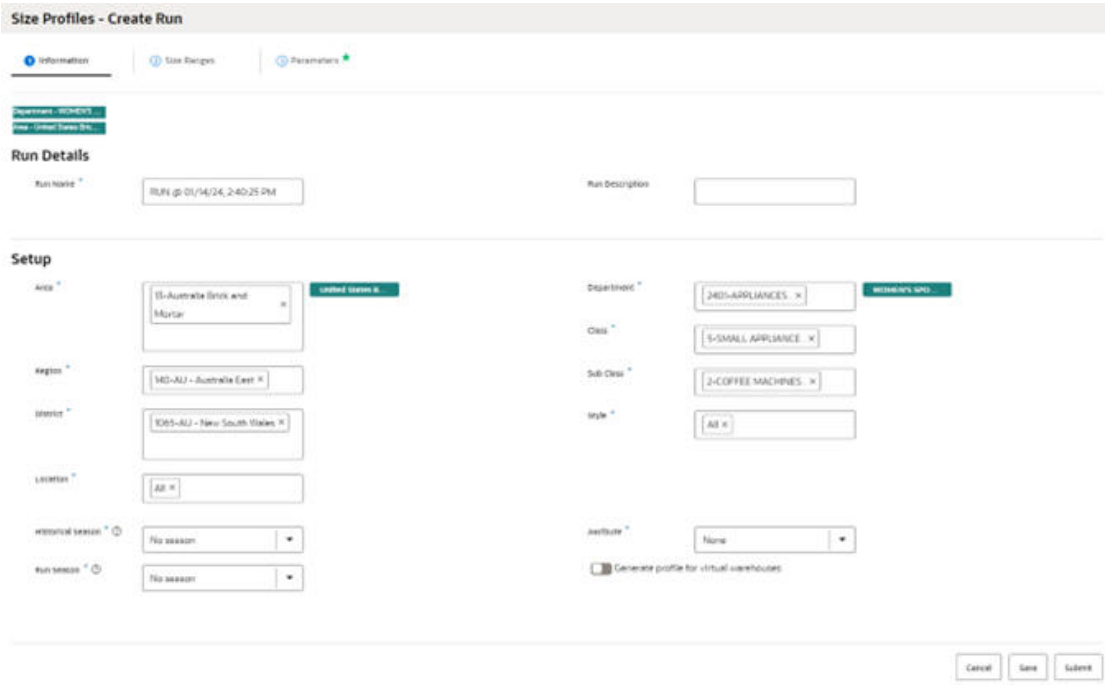
Click the **Run Name** with a status of Successful to see the size profiles as well as the inputs of that run. If the run is in Failed status, only the run inputs will be shown. If the run is in Setup status, the edit run pop-up opens.

The Recent Successful Run list is displayed in the contextual area on the right. You can also click a run in this list in order to open it.

Creating a Run

To estimate size profiles, you create a run and specify the settings to be used for the estimation. To begin the process, either click the **Create** icon at the top of the overview table or select a row in the overview table and click the **Duplicate** icon to duplicate a run. Then click the link of the duplicated run in Run Name column. In either case, the Create Run screen is displayed.

Figure 8-3 Create Run Showing Information Tab



The Create Run screen has the following features:

- Three tabs: the "Information Tab", the "Size Ranges Tab", and the "Parameters Tab".
- The **Save** button, which saves the changes made in any of the tabs.
- The **Cancel** button, which discards the changes in the current tab and closes the Create Run screen.
- The **Submit** button, which saves all changes, begins the execution of the run, and closes the screen. At this point, the status of the run changes from Setup to In Progress.
- The contextual area on the right, which displays a summary of user selections in each of the three tabs.

Information Tab

Use the Information tab, to provide the run name and description in the Run Details section. Use the selectors to select merchandise and location nodes that the size profile should be generated for. The highest/lowest level of selection and the lowest level that multi-select is allowed are controlled by global configurations. These configurations are explained in the *Oracle Retail AI Foundation Cloud Services AI Foundation Implementation Guide*.

Select the historical season to filter the historical data based on the start/end date of that season as well as limit the items to the ones that are mapped to that season. If item-season mapping is not available for a season, no data will be found for processing and no profiles will be generated. You can also select no season. In that case, the start/end of the historical period will be defaulted to the values indicated by SPO_NO_SEASON_START_DATE/ SPO_NO_SEASON_END_DATE in RSE_CONFIG. If those values are null, then the start/end of the historical period will be defaulted to the min/max weeks in the calendar

hierarchy data (RSE_CAL_HIER table). In either case, you can override the start/end of historical period in the third tab of the Create Run screen.

In addition to the historical season, you can also select a run season. This is simply linking the run to the selected season. This is a feature which is used by the RMS Allocation system to link the profiles to the upcoming seasons.

If you want to use a product attribute in the escalation, select an attribute value. In the Parameters tab, you can set the escalation path as desired. Lastly, select the option for warehouse profiles, if desired.

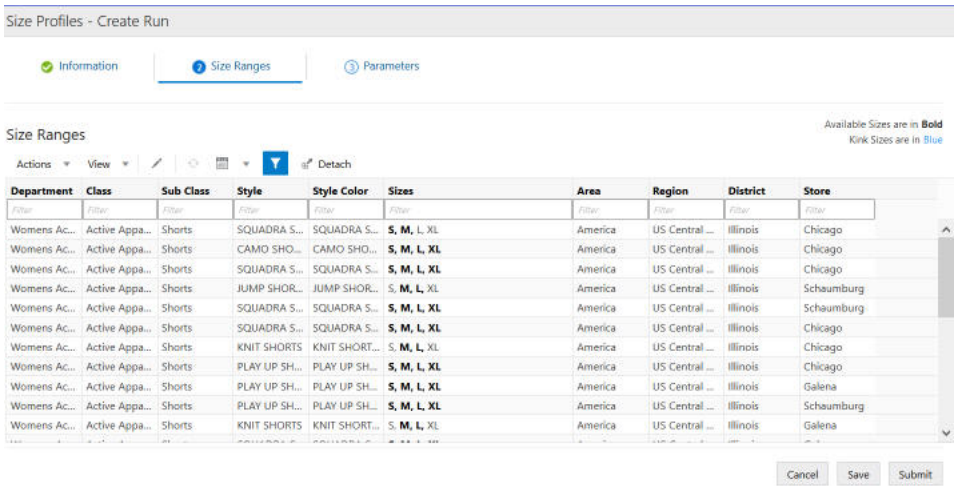
Size Ranges Tab

A size range consists of a set of all available sizes for each style-color. For example, the size range for the style-color of Men's Pullover Black and Store of Boston Downtown is S, M, L, XL. The size range data is provided by the retailer.

Size range data is required for estimating the size profiles. Once you navigate from the Information tab to any of the other tabs, a process runs at the back end to extract size range data for the selected merchandise/location nodes and to populate relevant tables. This process can take some time to complete, depending on the size of data. You will see a pop-up message in the UI to indicate that the process is in progress. While the process is in progress, the merchandise/location filters will be disabled and cannot be modified. You can see the status of this process in the notifications on the left menu in the screen.

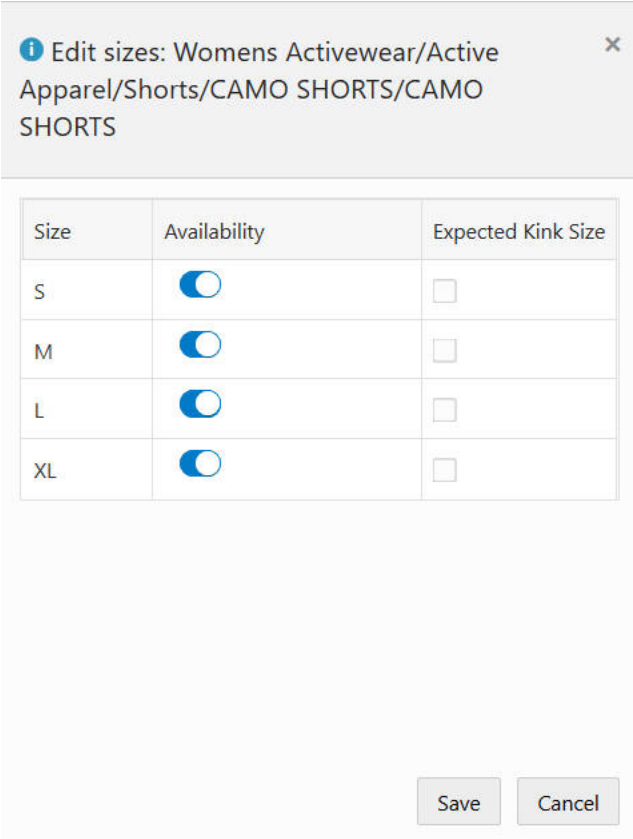
In the Size Ranges tab, shown in Figure 8-4, size range data is shown for each style-color and each store. If a certain size is not available in a size range, it will be shown as greyed out. If a size has been specified as an expected Kink Size (which means it is an odd size that may have significantly fewer or more sales than other sizes), it will be shown in blue.

Figure 8-4 Create Run Showing Size Ranges Tab



To edit the available sizes and kink size flags, select a row and click the **Edit** icon. This opens a pop-up, shown in Figure 8-5

Figure 8-5 Edit Available Sizes



Parameters Tab

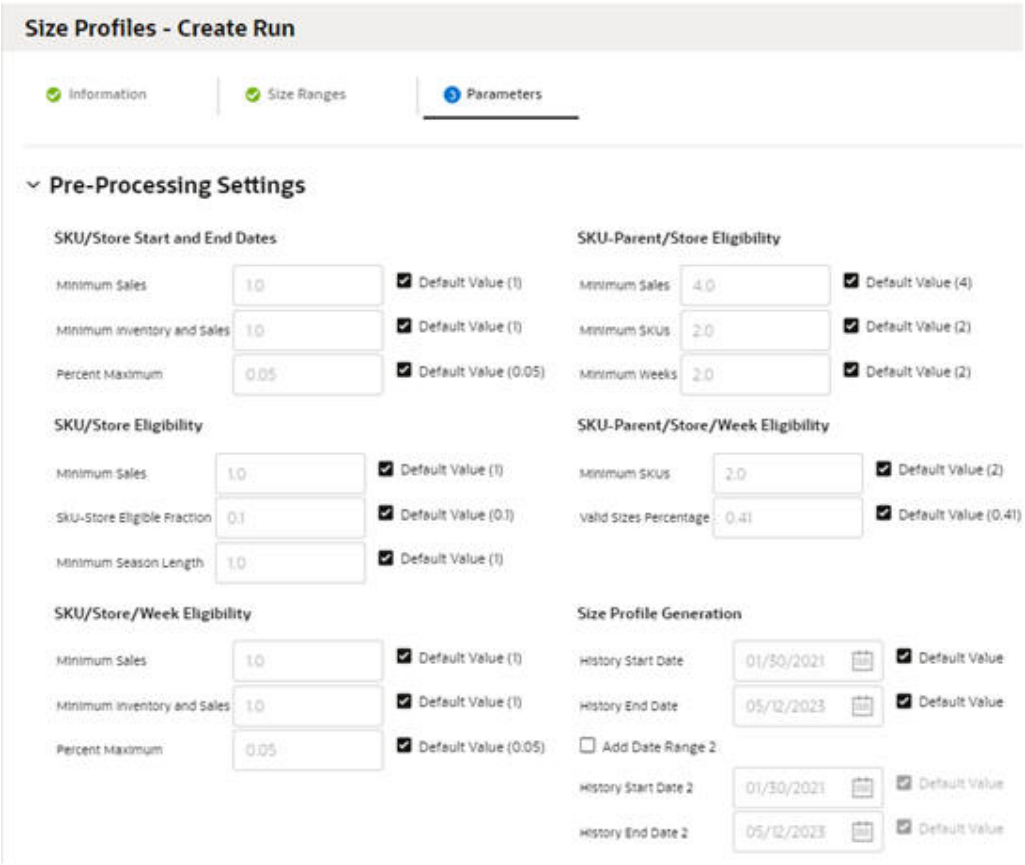
The Parameters tab displays the parameters that are used by the size profile generation algorithm. The escalation path is used to cluster data and estimate profiles at different levels (that is, merchandise/location or merchandise/location/attribute) and find the optimal size profiles that best represent the expected demand pattern and is most robust to stock out periods, imbalanced stock across sizes and other factors. When a size profile at a given level of the escalation path does not meet the post-processing criteria, it automatically gets rejected and the size profile associated with a higher level is examined. This process continues until a valid size profile is fond at one of the levels along the escalation path.

The lowest level of the escalation path in a run defines the level of the profiles generated in that run. For example, if the escalation path consists of department-region, department-location, subclass-region, and subclass-location, the size profiles will be generated at subclass-location level, that is one profile for each subclass/location/size range. In addition, for all runs at any level, size profiles will also be generated at the lowest level (style-color/location).

The parameters of the size profiles estimation algorithm fall into two categories: pre-processing and post-processing.

You can edit the values for these parameters, as shown in Figure 8-6. The default values of the parameters are defined in SPO_DFLT_PARAMETER. You can modify the default values using the Manage System Configuration screen under the Control & Tactical Center. The table below shows the name, description, and default value of the parameters.

Figure 8-6 Create Run Showing Parameters Tab



PARAMETER_NAME	PARAMETER_TYPE	UI_PARAMETER_LABEL	PARAMETER_VALUE	DESCRIPTION
SPO_MIN_SALES	SKU/Store Start and End Dates	Minimum Sales	1	Minimum Sales for determining start and end dates.
SPO_MIN_INV_SALES	SKU/Store Start and End Dates	Minimum Inventory and Sales	1	Minimum Inventory and Sales for determining start and end dates.
SPO_PCT_MAX	SKU/Store Start and End Dates	Percent Maximum	0.05	Percent Maximum for determining start and end dates.
SPO_MIN_SALES_SKU_STR	SKU/Store Eligibility	Minimum Sales	1	Minimum Sales SKU-Store for determining SKU-Store Eligibility.
SPO_ELIG_FRACTION_SKU_STR	SKU/Store Eligibility	SKU-Store Eligible Fraction	0.1	Eligible Fraction SKU-Store for determining SKU-Store Eligibility.

PARAMETER_NAME	PARAMETER_TYPE	UI_PARAMETER_LABEL	PARAMETER_VALUE	DESCRIPTION
SPO_MIN_SEASON_LEN_SKU_STR	SKU/Store Eligibility	Minimum Season Length	1	Minimum Season Length SKU-Store for determining SKU-Store Eligibility.
SPO_MIN_SALES_SKU_STR_WK	SKU/Store/Week Eligibility	Minimum Sales	1	Minimum Sales for determining SKU/Store/Week Eligibilit
SPO_MIN_INV_SALES_SKU_STR_WK	SKU/Store/Week	Minimum Inventory and Sales	1	Minimum Inventory and Sales for determining SKU/Store/Week Eligibility
SPO_PCT_MAX_SKU_STR_WK	SKU/Store/Week Eligibility	Percent Maximum	0.05	Percent Max SKU-Store-Week for determining SKU/Store/Week Eligibility
MIN_SLS_PRNT_STR_WK	SKU-Parent/Store Eligibility	Minimum Sales	4	Minimum sales units for determining SKU-Parent/Store Eligibility.
MIN_SKUS_PRNT_STR_WK	SKU-Parent/Store Eligibility	Minimum SKUs	2	Minimum SKUs for determining SKU-Parent/Store Eligibility
MIN_WKS_PRNT_STR_WK	SKU-Parent/Store Eligibility	Minimum Weeks	2	Minimum Weeks for determining SKU-Parent/Store Eligibility
SPO_MIN_SKU_STR_WK	SKU-Parent/Store/Week Eligibility	Minimum SKUs	2	Minimum SKUs for determining SKU-Parent/Store Eligibility
SPO_PCT_SZ_RNG_PRNT_STR_WK	SKU-Parent/Store/Week Eligibility	Valid Sizes Percentage	0.41	Percent Size Range for determining SKU-Parent/Store Eligibility.
SPO_SZ_PRO_GEN_METHOD	Size Profile Generation	Size Profile Generation Method	Optimization	The size profile generation method (Optimization or Normalized).
SPO_SKU_PRNT_STR_WK	Postprocessing	Valid Sample Threshold	2	Post Processing threshold: Min number of SKU-parent/store/weeks with at least one valid sale.

PARAMETER_NAME	PARAMETER_TYPE	UI_PARAMETER_LABEL	PARAMETER_VALUE	DESCRIPTION
SPO_TOT_UNIT_SLS	Postprocessing	Valid Sales units Threshold	2	Post Processing threshold: Min number of valid sales units that each size in the partition should have.
SPO_TOT_UNIT_SLS_PCT_SIZES	Postprocessing	Valid Sizes Percentage	0.4	Post Processing threshold: Min Percent of sizes.
SPO_SALE_PROF_CORR_THRESHOLD	Postprocessing	Threshold for correlation with Sales Profile	0.9	This parameter provides the correlation threshold between size profiles and sales profile.
SZ_PROF_OVER_CORR_THRESHOLD	Postprocessing	Threshold for size profiles compared with historical sales profile	2	Indicates size profiles are compared with historical sales profile and if the ratio between size profile and sales profile exceeds this value for at least one size, profile is flagged and is shown in UI under over-correction tile in run output
SPO_KINK_INNER_RATIO	Kink Size Analysis	Kink Inner Ratio	1.1	The default value for kink inner ratio.
SPO_KINK_OUTER_RATIO	Kink Size Analysis	Kink Outer Ratio	1.03	The default value for kink inner ratio.
SPO_CORR_LOW_SELL_FLG	Low Seller Correction	Correct Size Profiles for Low Sellers	Y	The flag (Yes or No) to indicate whether to apply post processing to correct size profiles of low sellers
SPO_LOW_SELL_AVG	Low Seller Correction	Average Weekly Sales of SKU	1	Average weekly sale units across a season, below which deems a SKU (Size) to be a low seller.

PARAMETER_NAME	PARAMETER_TYPE	UI_PARAMETER_LABEL	PARAMETER_VALUE	DESCRIPTION
ENABLE_OVER_CORR_FILTER_FLG	Postprocessing	Enable size profiles over-correction filter flag	Y	Y/N flag which indicates whether the profiles that have over-correction should be removed during post-procesing (Y) or should be kept in the final set of profiles and just be flagged as over-corrected (N).

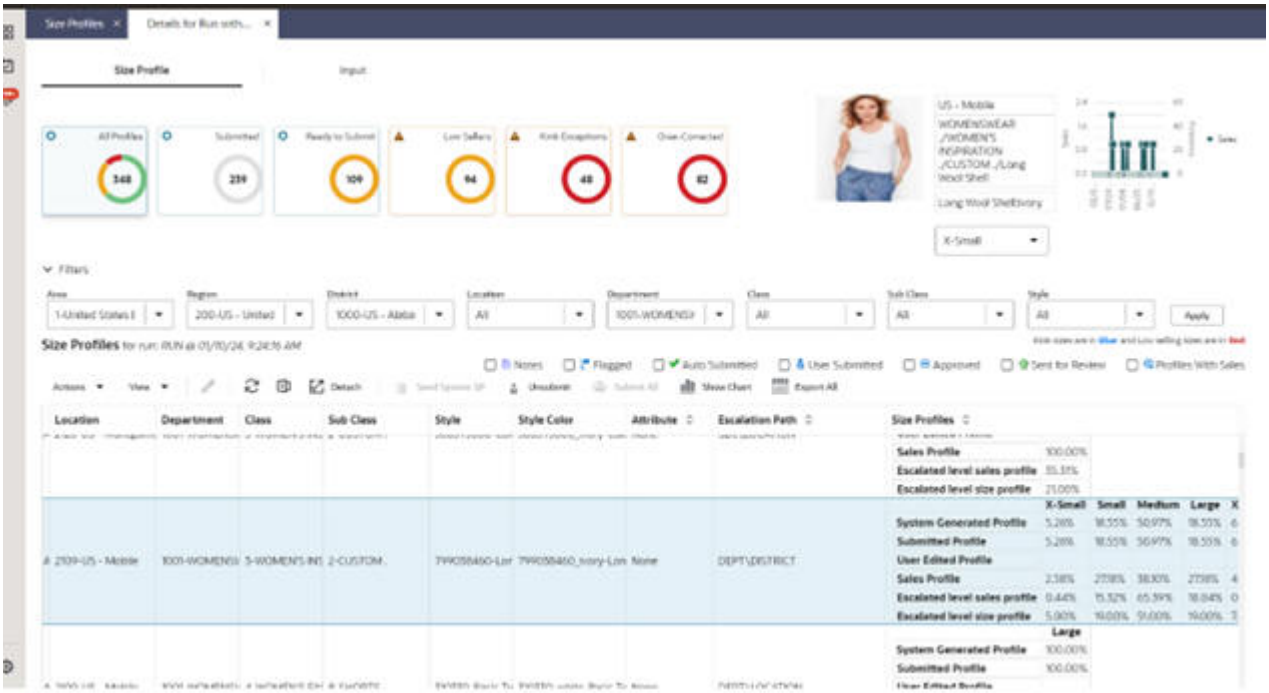
Reviewing a Run

Once the run has completed successfully, you can use various options to filter the profiles and review the size profile output. You can also override and submit/un-submit size profiles.

Run Output

To open a run with a status of Complete, click the **Run Name** in either the All Runs table or from the list of recent runs. The Run Output screen opens in a new tab. This screen has two tabs, Size Profiles and Inputs.

Figure 8-7 Run Output



Size Profiles Tab

The Size Profiles tab displays the generated size profiles for each merchandise/location combination at the run level. You can filter the profiles based on their submitted and alert status using the five tiles at the top of the screen for All Profiles, Submitted, Ready to Submit, Low Sellers, Kink Exceptions, and Over-corrected.

Below the tiles, you can use the collapsible filter section to filter the size profiles based on the merchandise and location hierarchies. You can also use the check-box filters above the table to further filter profiles by status (auto/user, submitted, approved, sent for review) as well as profiles with notes/flag. Lastly, you can filter profiles to show/hide the profiles that had any sales during the historical period.

The size profiles table is multi-row selectable, using the Ctrl and Shift buttons. This table has the following columns:

- The first column (with no header) shows the status of the profile. Values are user submitted ( ), auto submitted ( ), and approved.
- The action column has the note and flag icons. Values are some notes have been added ( ), no notes added yet ( ), flagged ( ), not flagged ( ).

  You can click on the note icon to view or edit the notes in a pop-up. You can also click the flag icon to flag or un-flag a size profile.
- The columns that show the merchandise and location hierarchy (for example, Region, District, Location, Department, Class, Subclass, and so on) and another column to show attributes. If an attribute is not selected when setting up the run, this column will have value of None.
- The escalation path that indicates the level at which the size profile has been generated.

The size profiles column that shows the following values:

- System: the size profile generated by the system (using the estimation algorithm). This profile will be identical to the “escalated level size profile” unless the merchandise/location at the run level has fewer sizes than the merchandise/location at the escalation level. In that case the escalation level profile will be normalized based on the sizes available for the merchandise/location at the run level such that the system generated profile sums to 1.
- User: the values edited by the user. If the user has not edited the values or has not seeded system values to populate user values, this row will be empty.
- Approved: the approved values, if the size profile was approved. If the size profile has not been approved, this row will be empty.
- Sales Profile: shows the profile of the observed sales (without correcting for the lost sales and any other post-processing) at the profile level.
- Escalated level sales profile: shows the profile of the observed sales (without correcting for the lost sales and any other post-processing) at the escalation level.
- Escalated level size profile: shows the estimated profile at the escalation level.

The following actions can be done in the Size Profiles table:

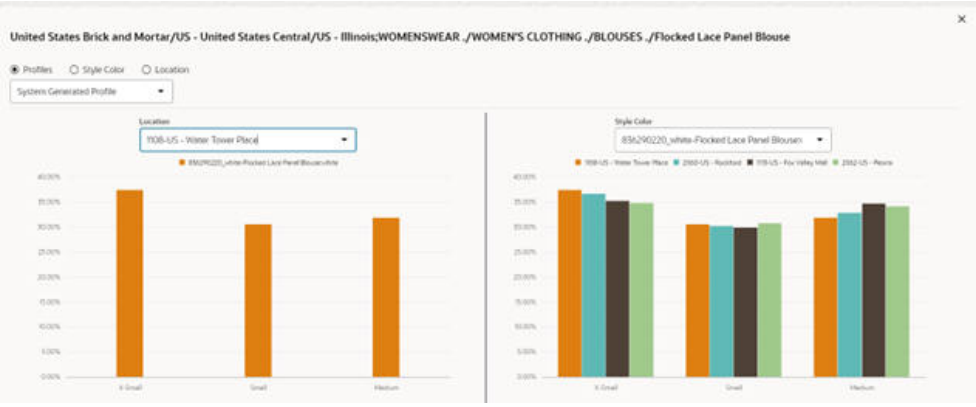
- Seed the user values with the System size profile by selecting one, or multiple, rows and clicking on the Seed System SP button at the top of the table.

- Edit a profile in the edit pop-up by selecting a single row and clicking on the Edit icon at the top of the table. If a size profile is submitted, edit is not allowed. To edit, the user must first un-submit the profile. The edit pop-up is shown in Figure 8-8. The edited values must sum up to 100 percent. If they do not, you will not be able to save. To normalize the values, you can use one of the two options (Normalize All or Normalize Edited). The first option adjusts all values so that they sum up to 100 percent. The second option adjusts only the edited values so that the sum of all the values becomes 100 percent. In the edit profile pop-up, you can also override the profile with the sales profile.

Figure 8-8 Edit Size Profiles

- Submit or un-submit a profile by selecting one or multiple rows and clicking on the Submit/Un-submit button at the top of the table. Alternatively, use the Submit All button to submit all the profiles that match the selected filters.
- Export the output to Excel using the export button above the table. This will export the profiles that match the selected filters. The maximum number of profiles that gets exported is controlled by the configuration SPO_UI_FETCH_SIZE in RSE_CONFIG. To export all profiles that match the selected filters, use the Export All button above the table. This option will export all the profiles even if number of them exceeds the configuration SPO_UI_FETCH_SIZE. In that case the export process may take longer and user will be notified once the process is complete and the Excel file is downloaded.
- View and edit the notes on each profile using the Note icon in the Action column, which opens a pop-up.
- Flag or un-flag a profile by clicking on the Flag icon in the Action column.
- View the graphs for each profile by selecting a row and clicking the “Show Chart” button above the table. This will open the pop-up that is shown in Figure 8–9. There are three different views:
  - Profiles view: In this view on the left side you can see the profiles of multiple merchandise nodes (at the run level) for a specific location node (at the run level). Similarly, on the right side you can see the profiles of multiple location nodes (at the run level) for a specific merchandise node (at the run level).
  - Merchandise view (for example style-color): In this view you can see the system profile, sales profile and user profile for the selected profile merchandise and the location that is selected from the drop-down list.
  - Location view (for example region): In this view you can see the system profile, sales profile and user profile for the selected profile location and the merchandise that is selected from the drop-down list.

Figure 8-9 Profiles View



- View the weekly sales and inventory and product image at the top of the screen for a selected row in the table. This view is available only for the runs that are at lowest level (style-color/location).

> **Note:**
>
> If a profile (for a given merchandise/location/attribute/size range) has already been approved or sent for review, the seed, submit/un-submit, and edit actions are not available. None of these actions can be performed for the profile. Note and flag can still be modified.

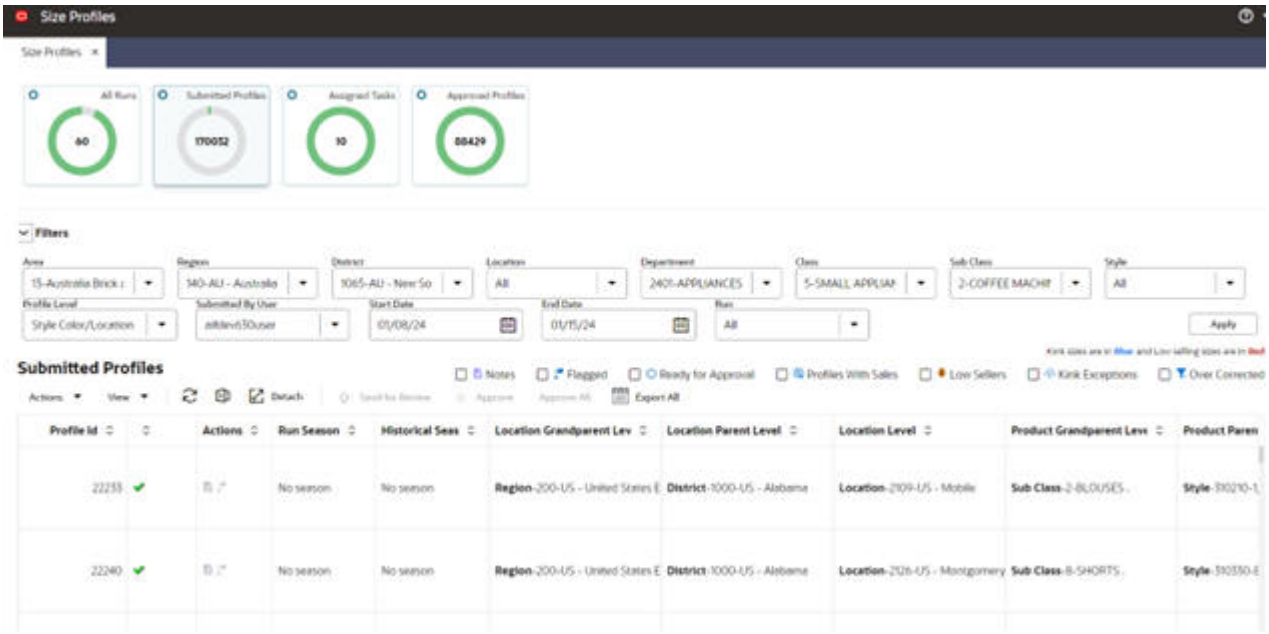
Input Tab

The input tab shows a summary of all inputs for the run in read-only mode.

Submitted Profiles

You can see the submitted profiles across all runs in the Submitted Profiles screen, shown in Figure 8–10. You can send profiles for review to your own user or another user. You can approve selected profiles or bulk approve all profiles. The approved profiles will get exported out once the nightly export process runs.

Figure 8-10 Submitted Profiles

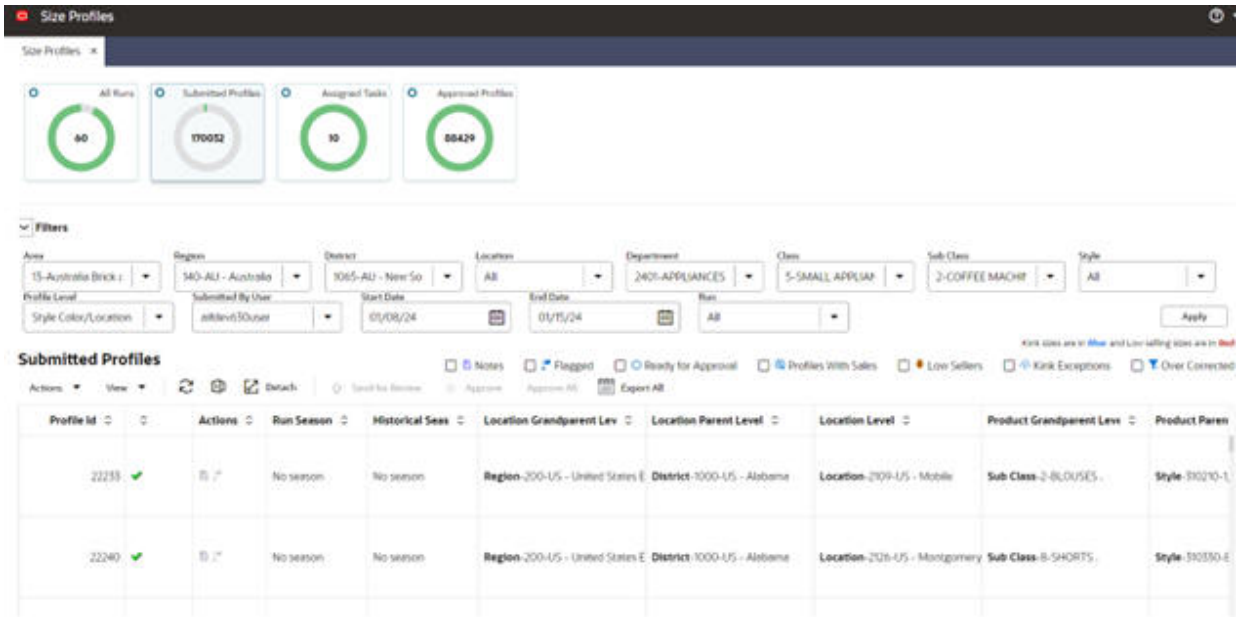


The Filters section is similar to the filters in the Run Output screen. You can use it to filter the size profiles based on product and location hierarchies. Using the profile level filter, you can view either the profiles at lowest level (style-color/location) or the profiles at higher levels. Using the user filter, you can filter the profiles based on the user who submitted the. Using the date filters, you can filter the profiles based on the date that they were submitted. Using the run filter, you can limit the profiles to a certain run. You can also use the check-box filters above the table to further filter profiles by ready for approval status, as well as having notes or flags. Lastly, you can filter profiles to show/hide the profiles that had any sales during the historical period, are low sellers, have kink exception, or have over-correction alert.

Approved Profiles

In the approved profiles screen, shown in Figure 8-11, you can view approved profiles across all runs.

Figure 8-11 Approved Profiles



The Filters section is the same as the one in the Run Output screen. You can use it to filter the size profiles based on product and location hierarchies. You can also use the check-box filters above the table to filter profiles by status (auto/user approved, submitted, and sent for review) as well as a notes/flag.

Submitted Profiles Table

The submitted profiles table is multi-row selectable, using the Ctrl and Shift buttons. This table has the following columns:

- The first column (with no header) shows the status of the profile: Value are user submitted ( ), auto submitted ( ), approved ( ), sent for review ( ).
- The action column has the note and flag icons. Values are some notes have been added ( ), no notes added yet ( ), flagged ( ), not flagged ( ).
- You can click on the Note icon to view or edit the notes in a pop-up. You can also click the Flag icon to flag or un-flag a size profile.
- The run season and historical seasons are the values that are selected when the run was created.
- Columns that show three levels of hierarchy for the merchandise and location of the profile. For example, if the profile is at style-color/location level, the three location hierarchy columns will show the region, district and location. The three merchandise hierarchy columns will show the subclass, style and style-color. Note that the submitted profiles can belong to different runs with different levels. Therefore, the profiles in submitted table are not necessarily at the same level.
- Attribute columns that show the attribute that was selected when creating the run. If attribute was not selected when setting up the run, this column will have value of None.

- The escalation path that indicates the level at which the size profile has been generated.
- Submitted by, Submitted on, and Run (on which the profile was submitted).
- Size profiles shows the submitted profile.

The following actions can be performed in this screen:

- View and edit the notes on each profile, using the Note icon in the Action column that opens a pop-up.
- Flag or un-flag a profile by clicking on the Flag icon in the Action column.
- Send a profile for review to own user or another user, using the **Send for Review** button above the table. It will open a pop-up in which you can select the reviewer role.
- Approve the profile by selecting one or multiple profiles and using the Approve button above the table. To approve all profiles that match the selected filters, use the Approve All button. The approved profiles will be exported out once the nightly or ad-hoc export process runs. .

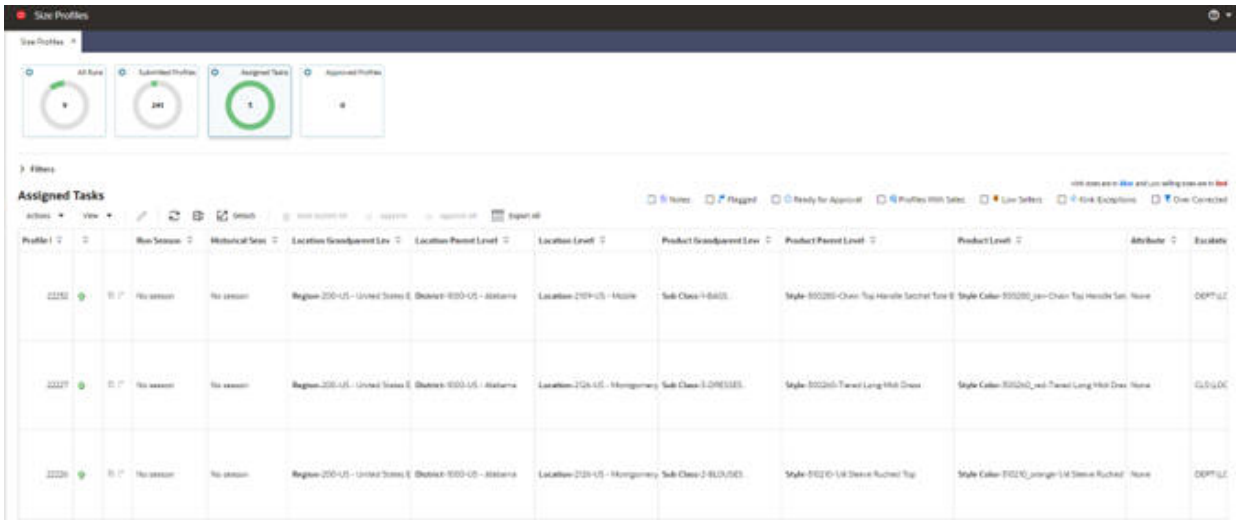
> **Note:**
>
> Once a profile is approved, it will no longer be shown in the submitted profiles screen and table. It will be only available to see in the Approved Profiles screen after the export process runs.

- Export the output to Excel using the export button above the table. This will export the profiles that match the selected filters. The maximum number of profiles that gets exported is controlled by the configuration SPO_UI_FETCH_SIZE in RSE_CONFIG. To export all profiles that match the selected filters, use the Export All button above the table. This option will export all the profiles even if number of them exceeds the configuration SPO_UI_FETCH_SIZE. In that case the export process may take longer and user will be notified once the process is complete and the Excel file is downloaded.

Assigned Tasks

In the Assigned Tasks screen, shown in Figure 8-12, you can see any profile that has been assigned to your user for review.

Figure 8-12 Assigned Tasks



The filters section is similar to the filters in the Run Output screen. You can use it to filter the size profiles based on product and location hierarchies. Using the profile level filter, you can view either the profiles at lowest level (style-color/location) or the profiles at higher levels. You can also use the check-box filters above the table to further filter profiles by ready for approval status, as well as having notes or flags. Lastly, you can filter profiles to show/hide the profiles that had any sales during the historical period, are low sellers, have kink exception, or have over-correction alert.

Assigned Tasks Table

The assigned tasks table is multi-row selectable using the Ctrl and Shift buttons. This table has the following columns:

- The first column (with no header) shows the status of the profile: Values are user submitted ( ) and auto submitted ( ).
- The action column has a note icon and a flag icon. Values are some notes have been added (
- You can click on the Note icon to view or edit the notes in a pop-up. You can also click the Flag icon to flag or un-flag a size profile.
- Columns that show three levels of hierarchy for the merchandise and location of the profile. For example, if the profile is at style-color/location level, the three location hierarchy columns will show the region, district and location. The three merchandise hierarchy columns will show the subclass, style and style-color. Note that the submitted profiles can belong to different runs with different levels. Therefore, the profiles in submitted table are not necessarily at the same level.
- Attribute column to show the attribute that was selected when creating the run. If attribute was not selected when setting up the run, this column will have value of None.
- The escalation path which indicates the level at which the size profile has been generated.

Size profiles chart that shows the following values:

- System: the size profile generated by the system (using estimation algorithm)
- User: the values edited by the user. Initially this row shows the values edited by the sender user (if any edits has been made).
- Submitted: the values that were submitted by the sender user.
- Sales Profile: shows the profile of the observed sales (without correcting for the lost sales and any other post-processing).

The following actions can be performed in this screen:

- Seed the user values with System size profile, by selecting one or multiple rows and clicking on the **Seed System SP** button at the top of the table.
- View and edit the notes on each profile, using the Note icon in the Action column that opens a pop-up.
- Flag or un-flag a profile by clicking on the Flag icon in the Action column.
- Edit a profile in the edit pop-up, by selecting a single row and clicking on the Edit icon at the top of the table.
- Approve the profiles by selecting one or multiple profiles and using the Approve button above the table. To approve all profiles that match the selected filters, use the Approve All button. The approved profiles will be exported out once the nightly or ad-hoc export process runs.

> **Note:**
>
> Once a profile is approved, it will no longer be shown in the assigned tasks screen and table. It will be only available to see in the Approved Profiles screen after the export process runs.

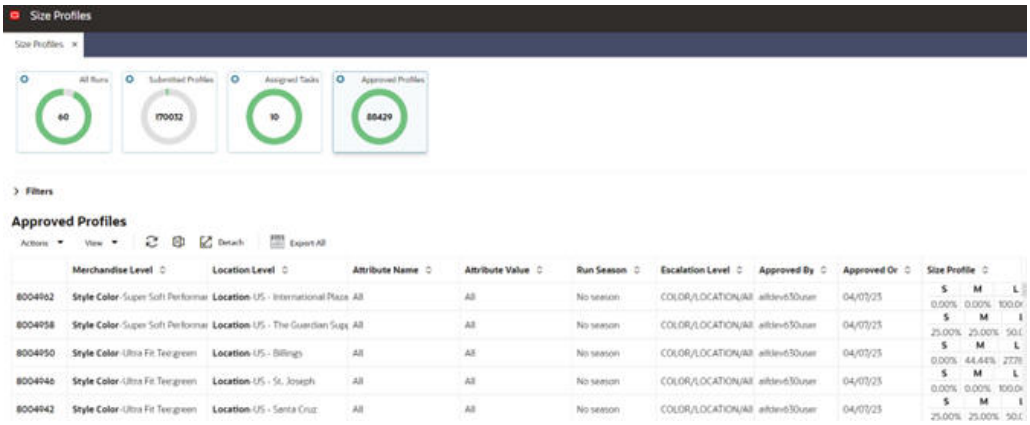
- Export the output to Excel using the export button above the table. This will export the profiles that match the selected filters. The maximum number of profiles that gets exported is controlled by the configuration SPO_UI_FETCH_SIZE in RSE_CONFIG. To export all profiles that match the selected filters, use the Export All button above the table. This option will export all the profiles even if number of them exceeds the configuration SPO_UI_FETCH_SIZE. In that case the export process may take longer and user will be notified once the process is complete and the Excel file is downloaded.

Approved Profiles

In the Approved Profiles screen, you can see the profiles that have been approved. the approved profiles are exported to different flat files during the nightly batch or whenever the ad hoc process for export is executed. The profiles are shown in the Approved Profiles screen only after the export is completed. For details of different export files refer to the *Oracle Retail AI Foundation Cloud Services: AI Foundation Implementation Guide*.

Using the drop-down lists in the filter section you can filter profiles based on the profile level and the merchandise/location combination at that level. For example, to see all approved and exported profiles at the subclass/location level for US-San Jose, set the location level to location. Set the location selection to US-San Jose. Set the merchandise level to subclass and set the merchandise selection to All.

Figure 8-13 Approved Profiles



Useful Tables and Views

The following table/views can be accessed and queried in Innovation Workbench and provide useful information about the input data or run outputs.

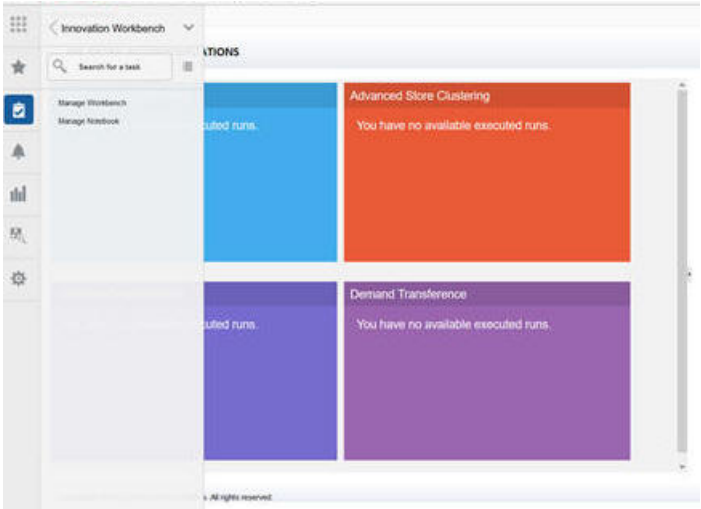
- SPO_RUN_METRIC: provides statistics of retained data after each stage of pre-processing. For example if too much of sales volume is being filtered after a certain stage of pre-processing, it can be an indication of the pre-processing parameters being too restrictive for that stage. It will be useful to do additional runs with more relaxed values for the parameters and compare the profiles and run metrics between the two runs.
- SPO_STYLE_DUP_DIFF_GRP_VW: provides the style-colors having more than one size range. This is considered invalid data and will require review of data in the source system.

9
Innovation Workbench

Innovation Workbench (IW) is a service that enables you to access read-only data using Oracle APEX and implement algorithms and analytics as an extension. This extension is a workbench for advanced analytics users that can be used to add new implementations via Oracle Advanced Analytic (Oracle R/ODM) algorithms that are implemented as SQL/PLSQL functions.

Innovation Workbench allocates a workspace to a retailer. This is a logical work area that is associated with a pre-allocated retailer database schema.

Figure 9-1 Accessing Innovation Workbench



Process

To access the workbench, in the Task menu shown in Figure 9-1, click **Innovation Workbench.** You see a list of workspaces in a new browser tab that are pre-allocated to the retailers you have access to.

Workspace List

The Workspace List displays all the workspaces that you were initially assigned to when workspaces were created at the time of installation and configuration. You can select a workspace from the list to review database objects using Oracle Application Express, shown in Figure 9-2.

Figure 9-2 Oracle Application Express

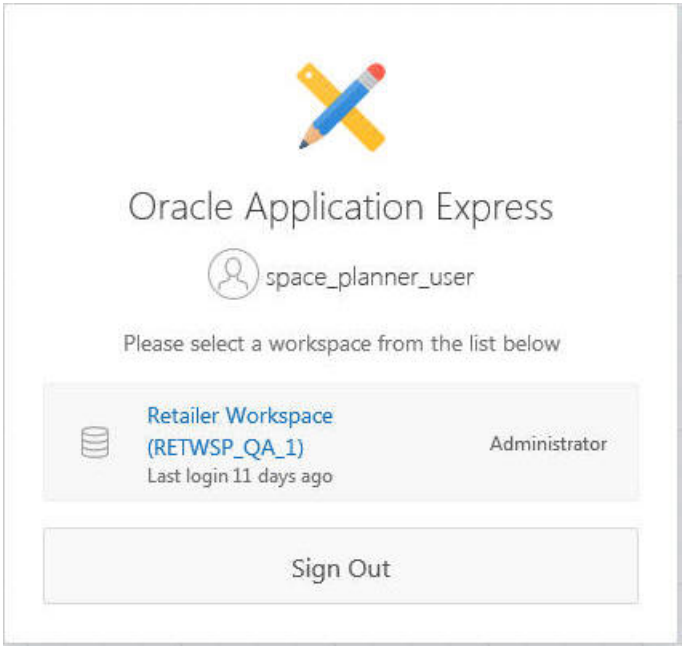
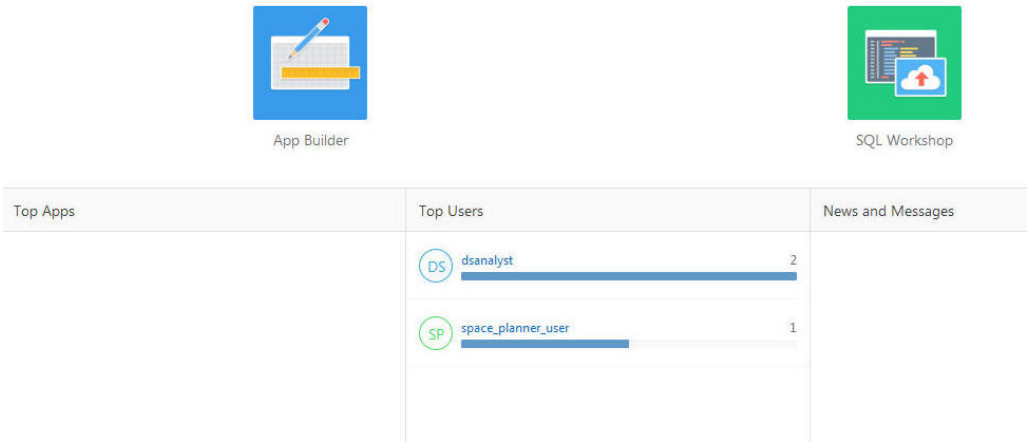


Figure 9-3 shows the Oracle APEX workspace, which displays SQL Workshop and App Builder.

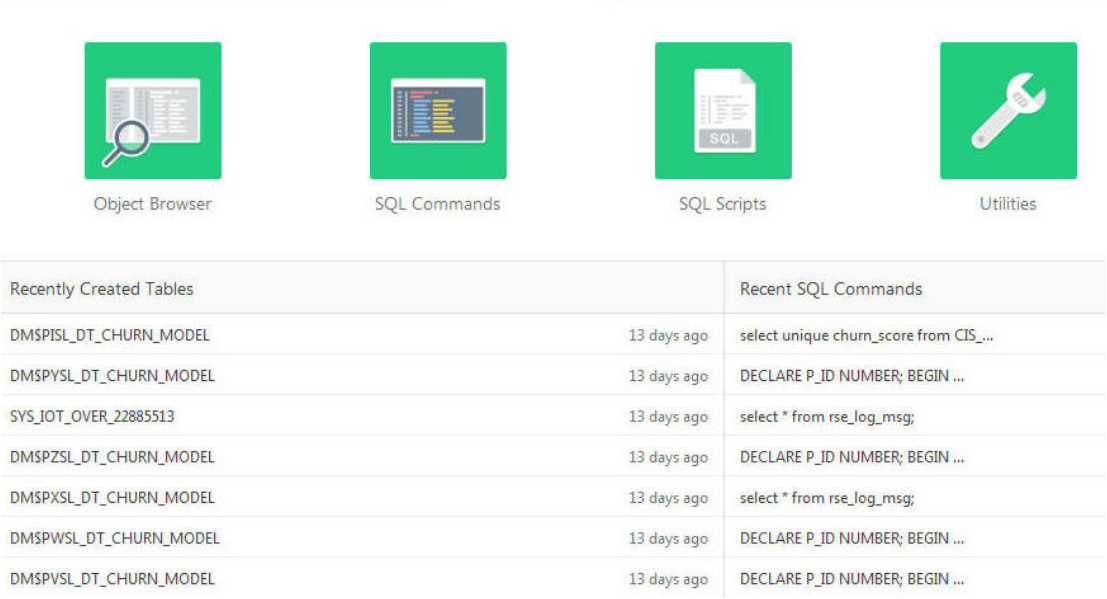
Figure 9-3 Oracle APEX Workspace



SQL Workshop

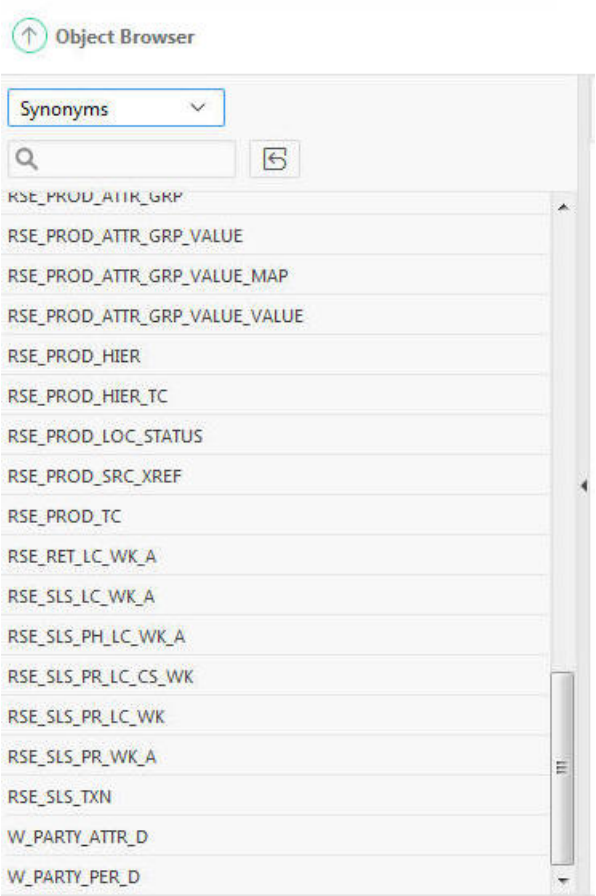
The SQL Workshop shown in Figure 9-4 provides tools to view and manage database objects.

Figure 9-4 SQL Workshop



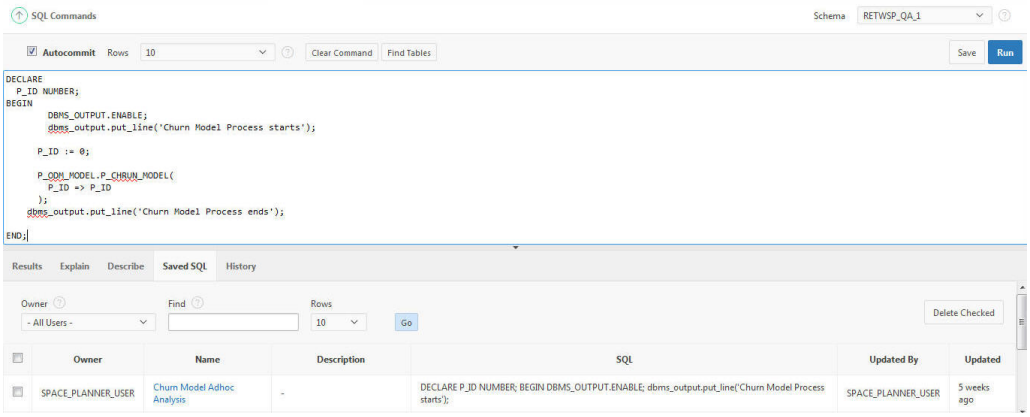
The Object Browser shown in Figure 9-5 lets you to use a tree control to view object properties and create new objects. The read-only database objects that are aggregated and cleansed are available for you to review and organize for data mining. Figure 9-4 shows the schema associated with the Retailer Workspace schema.

Figure 9-5 Contextual Area



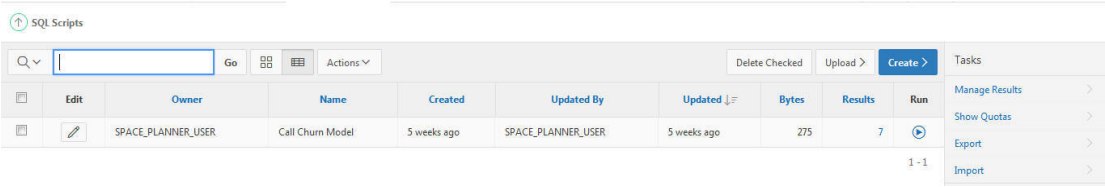
You can use the SQL Commands tool to enter ad-hoc SQL. The Query Builder can be used to create join queries using drag and drop. Figure 9-6 shows how to execute the ad hoc PL/SQL block.

Figure 9-6 SQL Commands



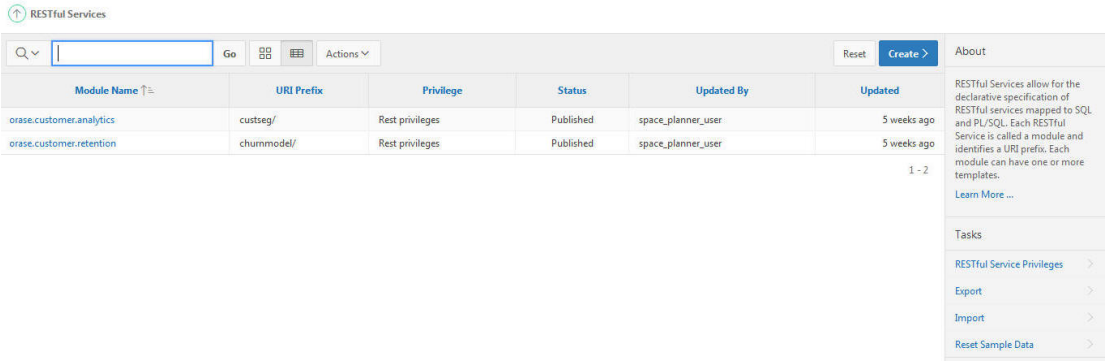
The SQL Scripts shown in Figure 9-7 can be used to store and run scripts. Use the Data Workshop to load and unload text, DDL, and spreadsheet data.

Figure 9-7 SQL Scripts



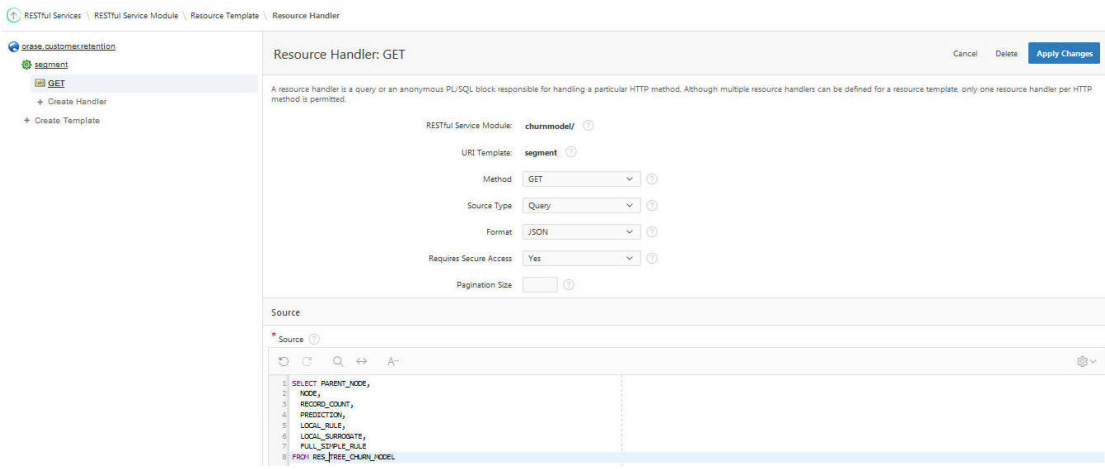
Use RESTful Services to manage web service definitions declaratively using SQL Query and to test the service modules from the web browser. Figure 9-8 shows a list of RESTful service modules.

Figure 9-8 RESTful Services



You can create a new RESTful Service module by clicking **Create** and setting the resource handler with a query or an anonymous PL/SQL block responsible for handling a particular HTTP method.

Figure 9-9 Resource Handler



APEX SQL Workshop Guide is available at http://docs.oracle.com/database/apex-5.1/AEUTL/toc.htm

Application Builder

Application Builder is a development environment that you can use to build and deploy database-centric applications. You can create interactive grids, reports, and charts.

APEX App Builder User's Guide is available at http://docs.oracle.com/database/apex-5.1/HTMDB/toc.htm

You can use wizards to declaratively assemble database applications organized by page to visualize and explore data, as a pre-process or post-process of data mining.

The following functionality is available.

Create

Launches the Create Application Wizard, which you can use to build a complete application containing multiple pages.

Import

Launches the Import Wizard, which you can use to import export files from the export repository.

Dashboard

Links to the App Builder Dashboard, which displays metrics about applications in the current workspace.

Workspace Utilities

Use the Workspace Utilities page to manage App Builder Defaults, news items displayed on the home page, workspace themes, and so on.

Manage Notebook

Data scientists can use the Innovation Workbench Notebook to create notebooks, which are collections of documentation, snippets of code, and visualizations. These notebooks are bundled with key python modules for machine learning, data mining, natural language processing, network analysis, and optimization solvers. Interactive graph visualization is available that supports highlighting, expansion, and applying filters. Custom paragraphs for advanced graph visualization concepts such as visual graph creation and large graph visualization concepts are supported. This powerful widget concept allows the analyst to curate data into a form that is easier to understand, highlights trends and outliers, and shares reports with business.

Features

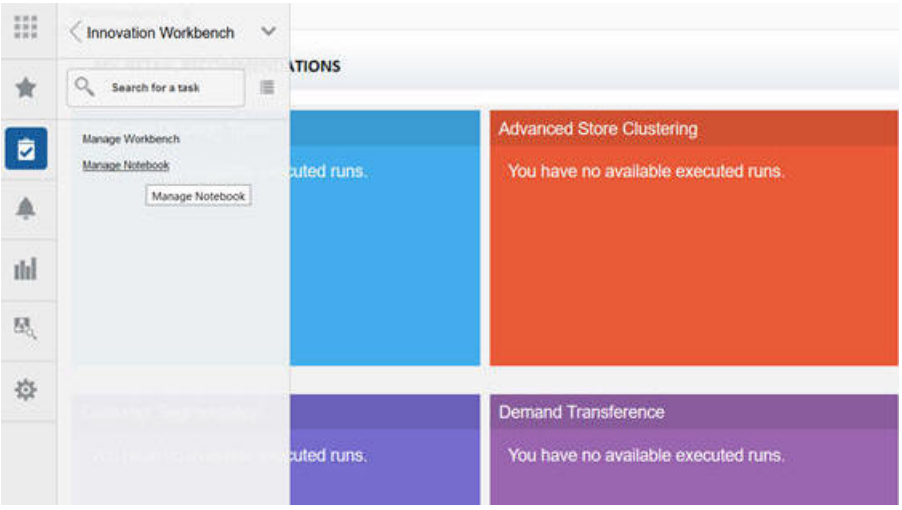
The key features of Manage Notebook include:

- Create, import, export, and clone python notebooks
- Browse, filter, and search for Notebooks.
- Add comments from the user interface (UI)

- Visualize data in the form of a Graph Visualization.
- Use collection of Python Libraries bundled with the product
  - Machine Learning packages such as (Tensorflow, Mxnet and Keras) Analyze
  - Optimization Solvers such as Gurobi

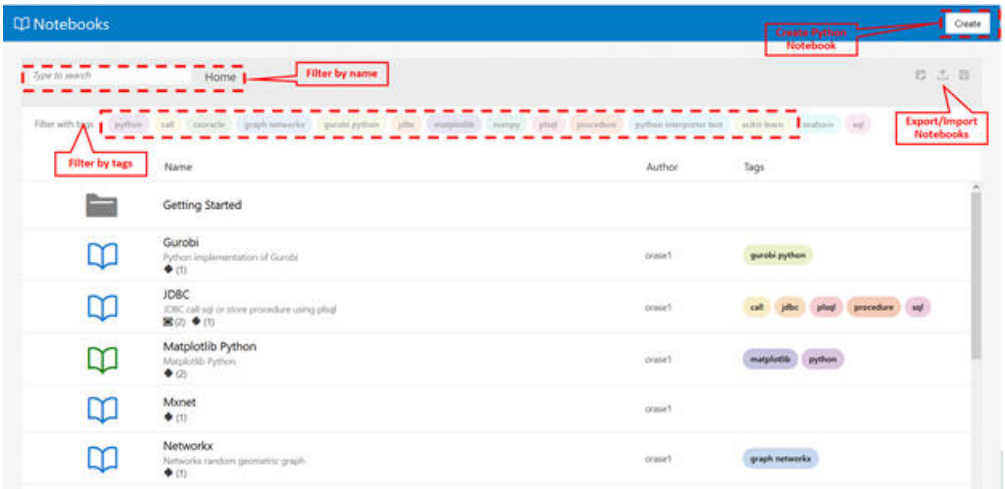
To access Python Notebook, click Innovation Workbench **Manage Notebook**. This opens a new tab in the browser and all the notebooks in the workspace are displayed.

Figure 9-10 Manage Notebook



From the workspace, you can import or export notebooks, create new workbooks, or copy notebooks. You can also browse, filter, and search for Notebooks using either a name or tags.

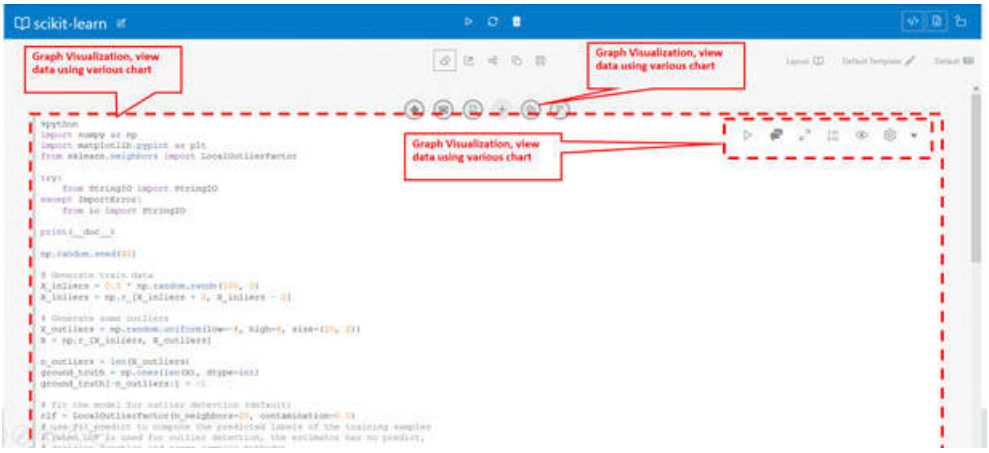
Figure 9-11 Notebook Functionality



You can create a notebook and add paragraphs using the toolbar. These paragraphs can be Python, JDBC, Markdown, or graph analytics paragraphs. You can execute paragraphs

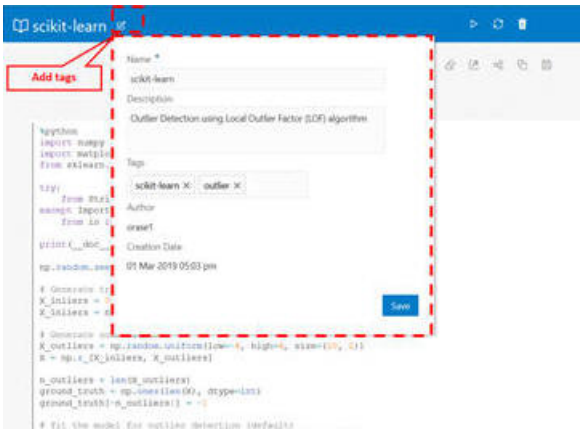
individually and see the results at the bottom of each paragraph. You can also execute all the paragraphs from the toolbar provided at the top.

Figure 9-12 Notebook - Adding Paragraphs



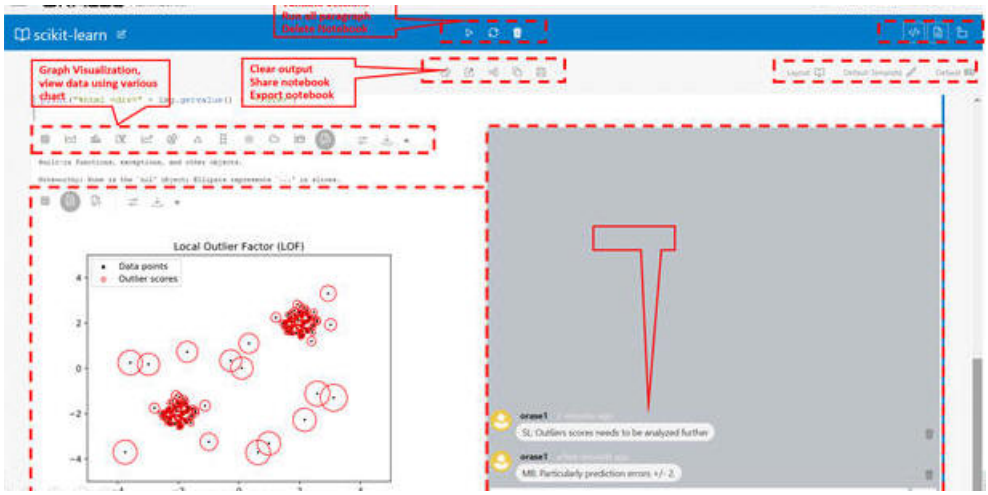
You can associate tags with the Notebook; these tags can be used for filtering and browsing the notebooks.

Figure 9-13 Notebook Tags



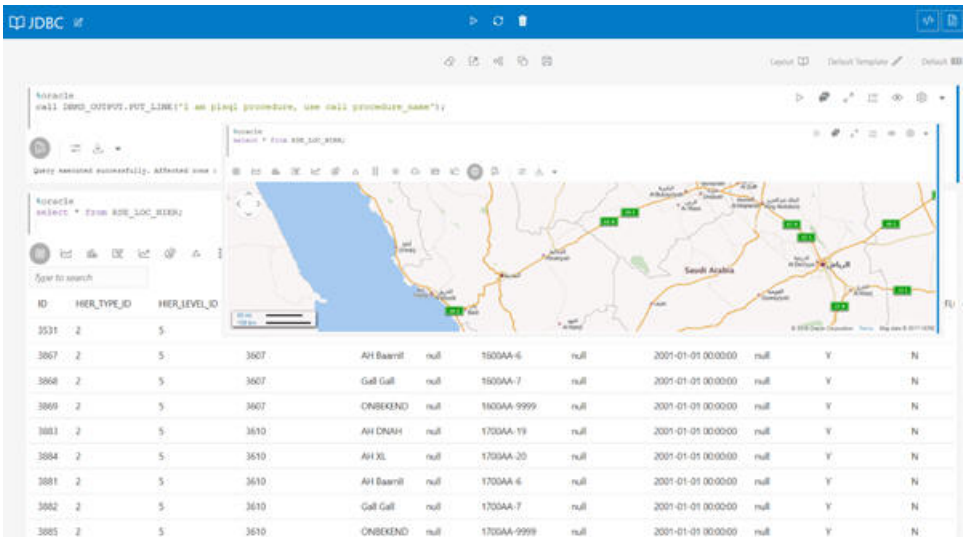
You can also add comments to the paragraphs and manage them from the user interface (UI). The comments can then be shared with others as a PDF. The Comments view shows all paragraph comments. You can visualize data in the form of a Graph Visualization and Exploration. You can explore graphs visually from within the notebook by viewing a summary of graphs or by using interactive exploration such as dragging, hovering, filtering, highlighting, and setting visualization properties. Different types of charts and graphs can be used for different types of datasets such as Line, Tag-Cloud, Tree Map, Scatter Plot, and so on.

Figure 9-14 Notebook Comments



You can visualize data in the form of a map and configure a column from a dataset to represent the latitude, longitude, title, and description.

Figure 9-15 Notebook Data as a Map



10
Oracle Digital Assistance

Oracle Digital Assistance (ODA) is a NLP and Machine Learning-based cloud service that enables application-specific interactions. It is a virtual user that assists end users with complex engagements using transactional data to review key business insights and exceptions. It also provides out-of-the-box knowledge sharing by responding to typical Q&A-style questions for customer service and support. ODA features are supported using voice along with conversational interface to distill end user intents, invoke actions, and provide reasonable responses. This enhances consulting services with minimal training and support.

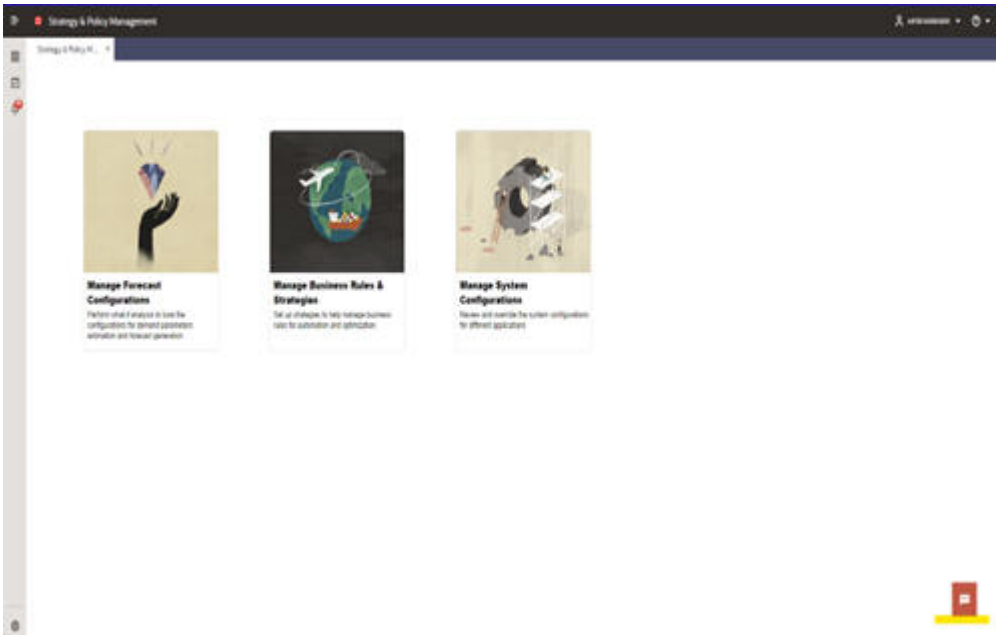
Digital assistance is not a replacement for a web application but a channel that allows the user to converse and query areas related to AIF.

Digital assistance supports Answer Intents, Regular Intents, and SQL Dialogs. (SQL Dialogs are skills that can translate a user's natural language utterances into SQL queries, send the queries to a backend data source, and display the response.)

This chapter describes how to use digital assistance.

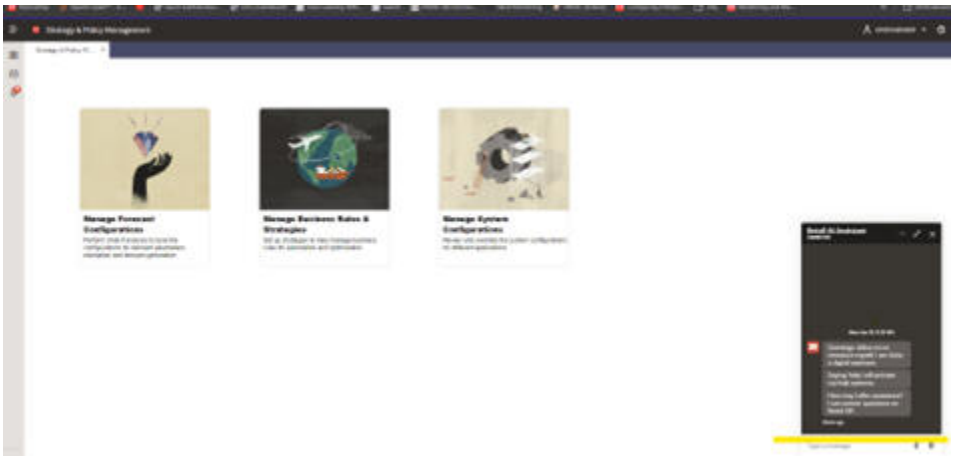
Access: Access to Digital Assistance in the AIF Platform is available with the Chat Icon on the AIF UI.

Figure 10-1 Chat Icon — AIF UI



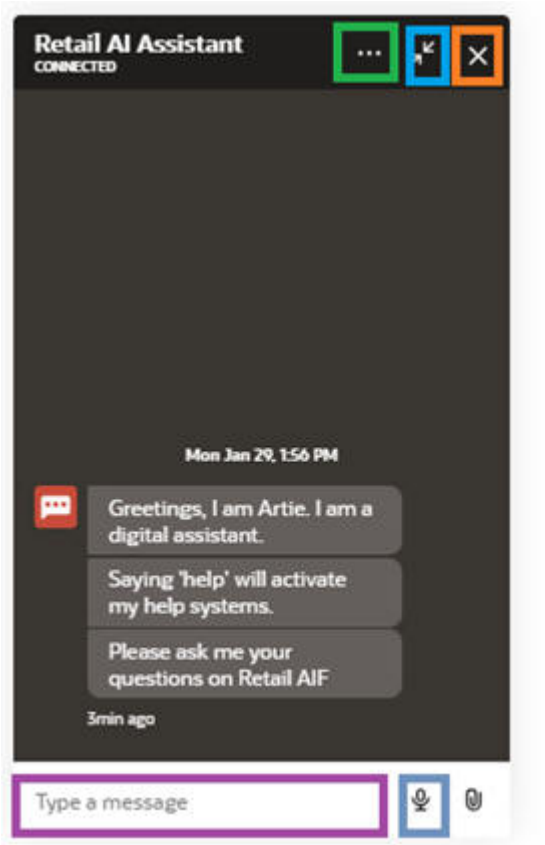
Click the icon to open the Chat widget.

Figure 10-2 Chat Window



Here is the Chat widget.

Figure 10-3 Chat Widget



The widget has the following components.

Figure 10-4 Widget Components

ICON	DESCRIPTION
	Show the Options available
	Closes the chat window without clearing the conversation
	Closes the Chat Window and clears the conversations
Type a message	User can type a Message and click enter
	Access voice component by clicking the microphone icon. If voice is enabled, the bar color next to the microphone icon will flip to a keyboard icon and When user speaks, the content is typed in the message window.

Digital assistance helps answer questions by returning relevant answers for the questions asked. It helps the user find FAQs or other knowledge-based documents. The data source of these bots is pre-loaded CSV and the AIF documentation and database entities that are pre-mapped on the Oracle Digital Assistant Instance.

Here are some examples of requests that you can try:

- How to configure DAT_FILE_VALIDATE_JOB.
- View DV reports.
- How to load product data.
- Why can't I export forecast data from AIF to RDX?

The following DB entities are mapped as part of the SQL Skills RASE01.RSE_LOG_MSG:

- RASE01.RSE_LOG_MSG
- RASE01.RSE_CONFIG
- RADM01.RI_LOG_MSG
- RADM01.C_ODI_PARAM
- RADM01.C_LOAD_DATES
- RADM01.C_HIST_LOAD_STATUS
- RADM01.C_HIST_FILES_LOAD_STATUS

These DB tables can be queried using natural language from the Chat window. Possible sample user utterances are:.

- Show rse log msg with log level error.

- Show the param value of the rse config where the param name is "DEFAULT_LEAD_TIME".
- Show c load dates for target table name W_PDS_FLEXFACT1_F.

11
Control and Tactical Center

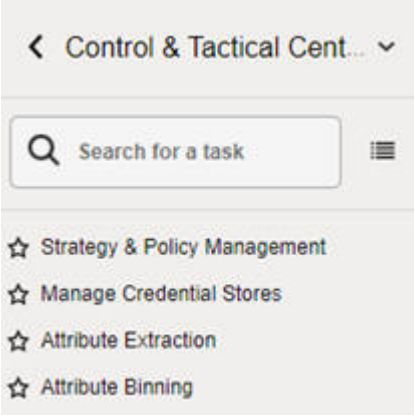
Using the Control and Tactical Center, a user can access Strategy & Policy Management. This is the central place for managing the configurations of different applications and for setting up and configuring the demand forecasting runs.

In addition, a user can access the modules for managing product attributes (Attribute Extraction and Attribute Binning) as well as the link for managing web service-related credentials under the Control & Tactical Center.

Depending on the role, a user will see one or multiple of the following links, as shown below.The administrator user will have access to all of the links.

- Strategy and Policy Management
- Manage Credential Stores
- Attribute Extraction
- Attribute Binning

Figure 11-1 Control and Tactical Center



Strategy & Policy Management

In Strategy & Policy Management, a user can edit the configurations of different applications via the Manage System Configurations functionality. In this screen, the tables can be filtered by applications.

To edit a row in any table, click the row and click the **Edit** icon. You can override the values for the columns that are editable.

Use the Manage Forecast Configurations to set up, manage, and configure the demand forecasting runs for different applications such as Lifecycle Pricing Optimization (LPO), Inventory Planning Optimization-Demand Forecasting (IPO-DF), Inventory Planning Optimization-Inventory Optimization (IPO-IO), Inventory Planning Optimization-Lifecycle

Allocation & Replenishment (IPO-LAR), Merchandise Financial Planning (MFP), and Assortment Planning (AP).

Figure 11-2 Strategy & Policy Management Dashboard

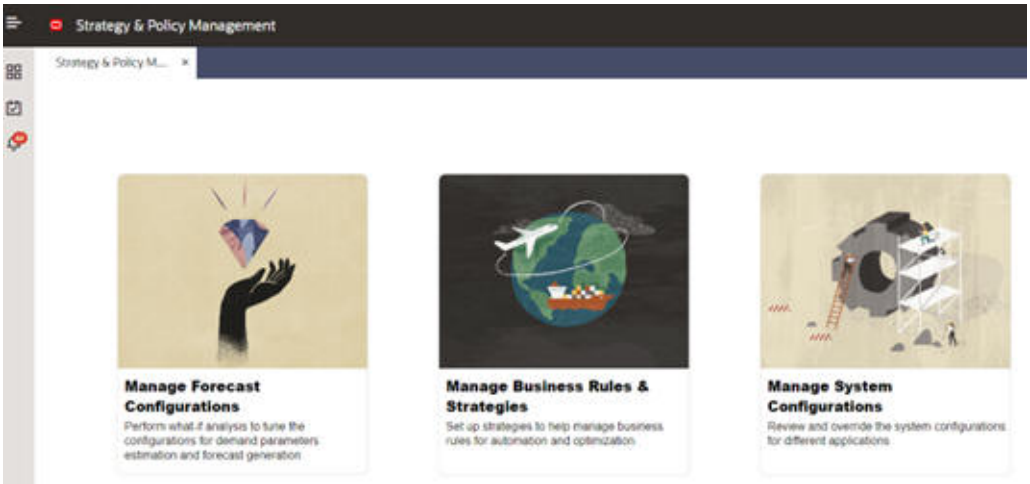
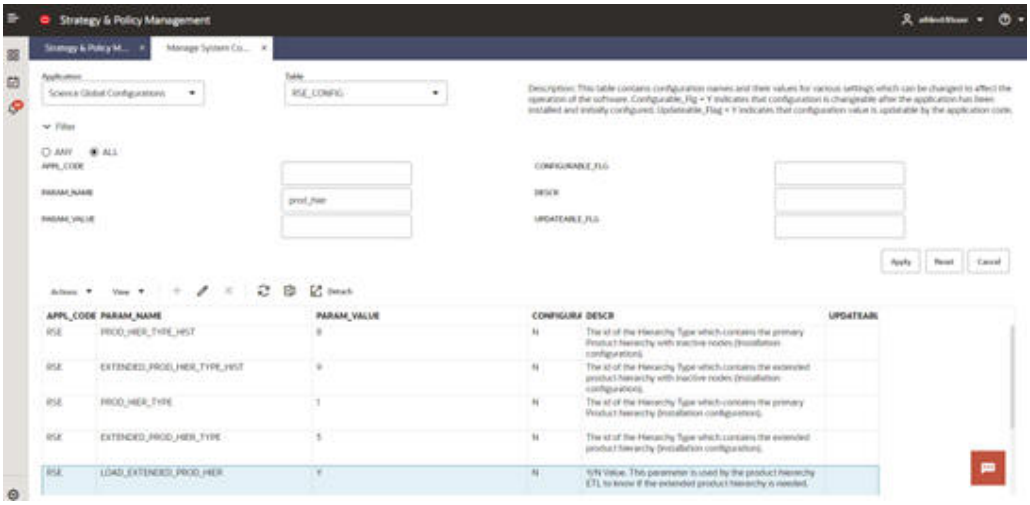


Figure 11-3 Manage System Configurations



Manage Business Rules

Business rules are policy inputs that inform the optimizations about business intent, boundaries, and so on. Each type of rule is created for specific purposes (for example, promotion rules, inventory rules) and are grouped into categories of related concepts.

However, the expected business outcomes can't easily be understood by looking at the pieces (individual rules). Rather, the set of rules that function together must be viewed together to build a complete understanding of the business drivers. This is achieved by grouping rules into Strategies. Strategies are the mechanism to organize rules around business intent.

To help reduce maintenance, rules are re-usable across strategies. Rules are individually maintained in the Rules Library so that all strategies making use of the rule automatically uptake modification.

Alert rules are also defined for specific business purposes. Alert rules are not assigned to strategies as they do not frequently change due to business events.

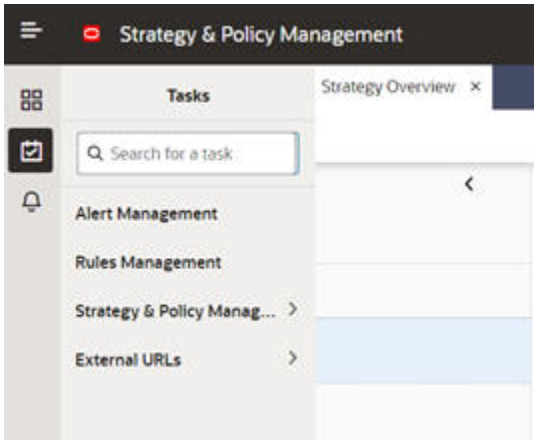
Alert Management

Alerts are user workflow indicators. They are used to indicate situations where the optimization results fall outside of the business defined thresholds. Alert Management is where you will define which alerts are relevant to your business and the associated thresholds for each alert type.

To navigate:

1. Open the Task menu.

Figure 11-4 Task Menu



2. Click Alert Management.

Inventory Alert Rules

Create or edit rule thresholds in any of the Rule Categories.

Overstock	Alerts the inventory planner when a review day's beginning of day inventory meets or exceeds the OUTL for that review day and/or Inventory units exceeds the specified threshold. This means that the inventory in the preceding period is more than enough to meet the demand of the prior, and the next review period. The inventory planner will review preceding stock levels, receipts, and forecast or actual sales.
Out of Stock	Out of stock alerts are based on lost sales. This alert requires a demand forecast in order to calculate lost sales. Lost sales occur any day inventory plus receipts exceeds the forecasted demand. Note that retail value is always calculated as full retail value.

Low Stock	Alerts the inventory planner to the days where inventory falls below acceptable thresholds of Days of Supply (DOS) and/or units. The DOS thresholds requires a demand forecast. The demand forecast is used to calculate average demand of the week. Then each day's beginning inventory is divided by the average demand to determine if it falls below DOS threshold. A day that meets out of stock rules will not be alerted for Low Stock. Tip: to ignore DOS threshold set it to 0 and choose '**OR**' Inventory is Less Than…

Rules Management

In the Promotion Effects module used by the Sales & Promo forecast method, promotion data flows from RI into AIF. This enables the modeling of promotions based on the promotions the retailer has defined at the lowest level.

The following three levels in the hierarchy are used to define a promotion.

- Campaign: The highest level in the promotion hierarchy. A campaign can be used to group one or more promotions specific to an event or a holiday (for example, Fall campaign, Christmas campaign, and Back to School campaign).
- Promotion: A campaign can be further divided into smaller time frames. For example, a Christmas campaign can be divided into a Week 1 promotion, a Week 2 promotion, and a Week 3 promotion. During a promotion, multiple offers can be active.
- Offer: An offer defines the mechanism used during a promotion (for example, BOGO, 20% Off, 50% Off, TV Ad, and Gift with purchase).

The user can manage promotion data in the Promotion Management module to clean up the campaign and offer data.

The raw data provided by the retailer may not be suitable for the modeling of promotions. Consider, for example, information about two fall campaigns, one in 2022 and one in 2023. If the year information is present in the campaign description, it must be removed for modeling purposes as well as to estimate the lifts associated with fall campaign and to apply the observed lift values for future fall campaigns in following years. Similarly, an offer description may contain unwanted information such as month or product descriptions. After the data is cleaned up, the campaign and offer names are referred to as *Campaign alias* and *Offer alias*. These values are passed to the Promotion Effects module so that the lift values can be estimated.

The following section describes the management of promotion data for forecasting. The process for estimating the promotion lifts is described under Promotion Effects.

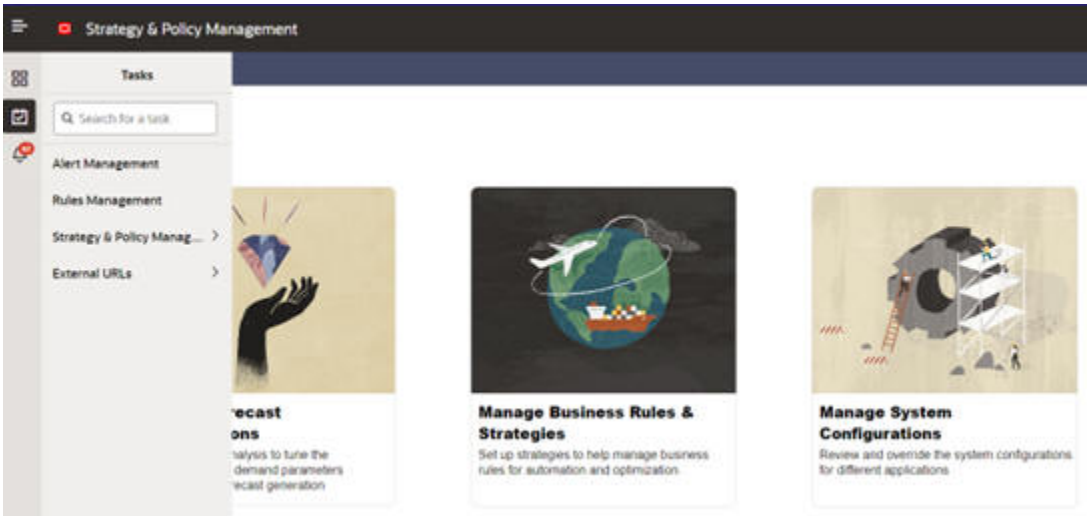
Promotion Management

Access the Promotion Management module from the AI Foundation Cloud Services page as follows:

Control and Tactical Center -> Strategy & Policy Management -> Tasks -> Rules Management -> Promotion Management

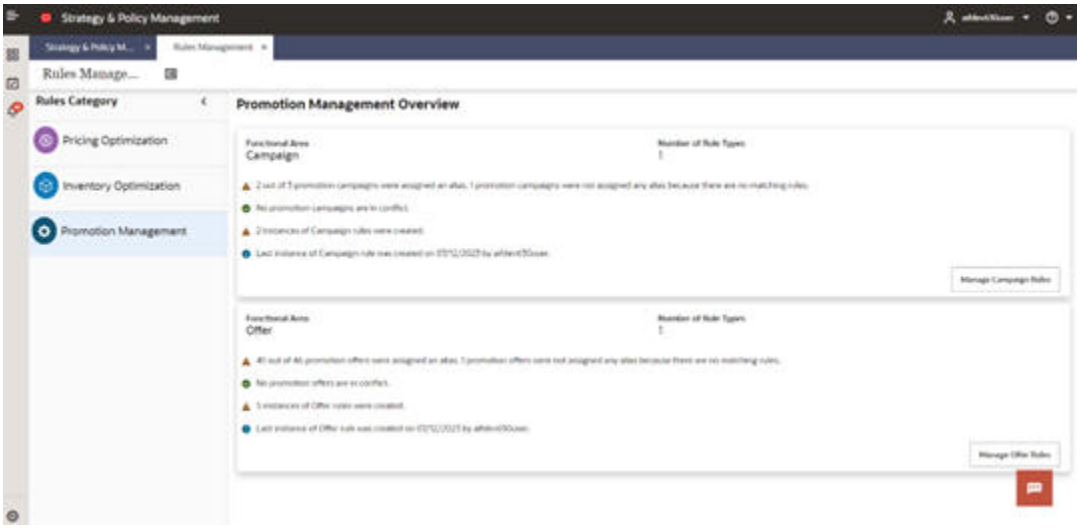
The Strategy & Policy Management window opens.

Figure 11-5 Strategy and Policy Management with the Tasks Pane



Select Rules Management. You see the Rules Category screen. Under Rules Category, select Promotion Management to view the screen for managing the campaign and offer data in AIF.

Figure 11-6 Promotion Management



Both Campaign data and Offer data are managed in a similar way; the following process describes managing the Offer data.

Select Manage Offer Rules

When you select Manage Offer Rules, you see the following three tabs.

- Offers: This tab contains the offer data received from the retailer. Review the data to understand the various offers used and pay attention to prefixes or suffixes to the offers that do not add value or that may have an impact on the lift estimation. Similar offers can be grouped together if necessary. For example, BOGO 50% and BOGO 100% can be combined to generate BOGO.

- Rules: Rules can be created to remove unwanted descriptions from the offer data.
- Results: Review the results after the rule is applied and make changes to the rule if necessary.

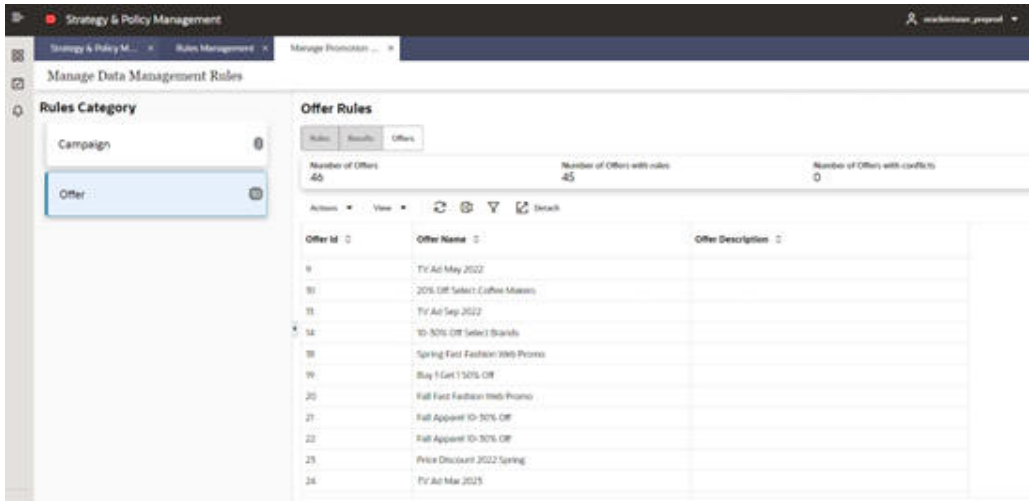
The Summary panel, located below the tab, displays the following metrics.

- Number of offers: Offers received from the retailer through the data file.
- Number of offers with rules: After each rule is created, this metric is updated by the total number of offers for which an alias value is generated by a rule.
- Number of offers with conflicts: Occasionally, multiple rules can impact an offer with conflicting results. This alerts the user of the need to review the rules and make changes.

For example, you may see multiple offers for the following.

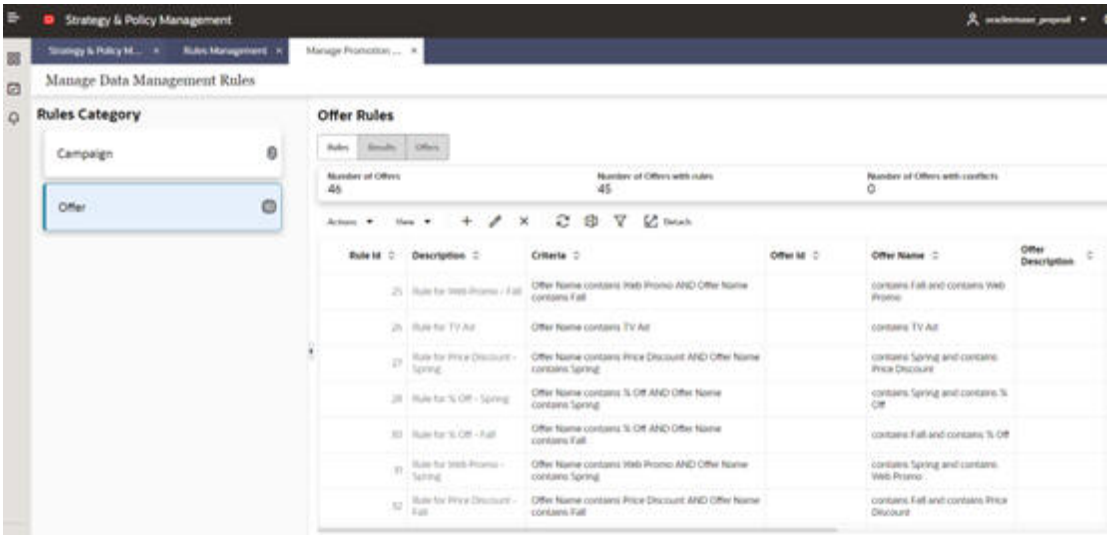
- TV Ad with Month and Year. The month and year can be removed using rules to generate an offer alias of TV AD corresponding to these offers.
- Price Discount with Year and Season: The year and season can be removed using rules to generate an offer alias of Price Discount corresponding to these offers.
- Web Promo: Multiple offers associated with the web promotion can be grouped by creating an appropriate rule to generate an offer alias Web Promo.
- % Off: All the offers with % Off can be grouped to generate an alias % Off.

Figure 11-7 Offers



Select the Rules tab to see the following screen. This is used to create the new rules and review the existing rules. In this example, rules have already been created for Price Discount, Web Promo, TV Ad, and % Off. Select the + symbol to view the Create Rule screen. To edit a rule, select an already defined rule and click **Edit**.

Figure 11-8 Rules



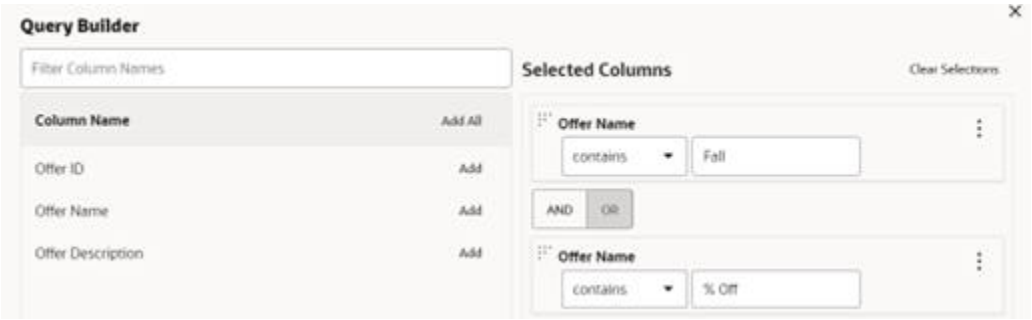
Create or Edit Offer Rules

The following must be provided for each rule.

- Description: Add a description to help manage the rules. This is used only in the Promotion Management screens.
- Rule criteria: Define the rule using the Query Builder. Use the Offer Name, Offer ID, and Offer Description columns to define the rules. Select one or more columns to define the rule. For each selected column, select the criteria from the drop-down menu.
- Rule value: Enter the alias value to associate with offers satisfying the rule criteria.

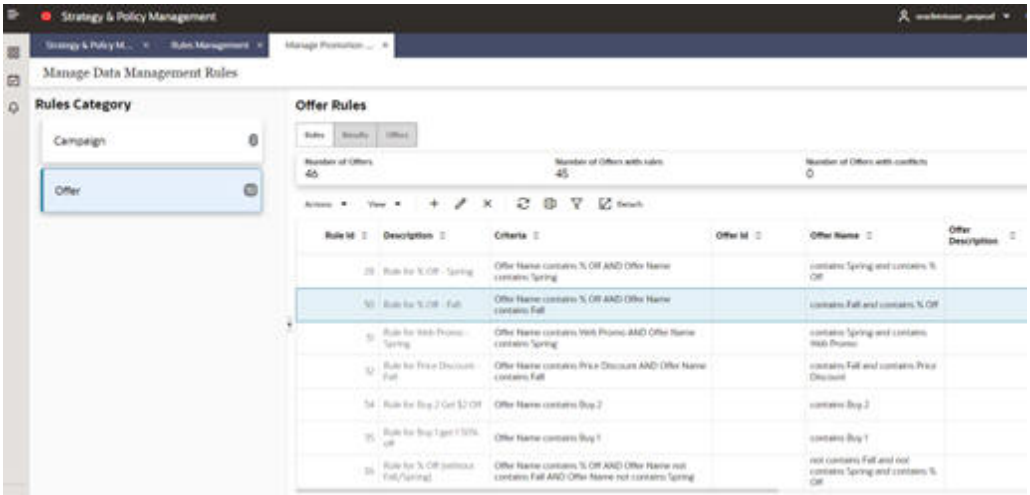
In the following example, all offer name that contain %-Off and Fall are identified and assigned a value of % Off-Fall. Click **OK** and save the rule. The Summary metrics are updated. You then see the Rule tab where you can review all the rules that have been created.

Figure 11-9 Query Builder



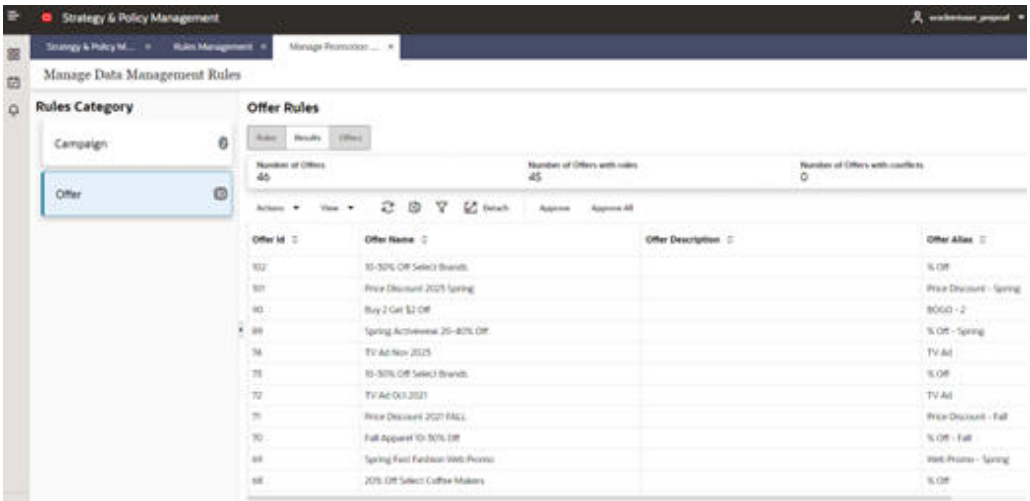
From the Rules tab, review the data for any conflicts in the Summary tab. You can select the Results tab to review the results and any conflicts.

Figure 11-10 Example Rules Tab



In the following Results tab, the Offer alias assigned to each Offer ID is shown. If there are conflicts for an Offer ID, a drop-down menu with multiple offer alias values associated with the Offer ID is displayed. To override the conflict, select one of the alias values in the UI or update the associated rules to remove the conflicts.

Figure 11-11 Example Results Tab



Manage Forecast Configurations

This section describes the workflow and screens in Manage Forecast Configurations. With this functionality, you can set up and manage the forecast run types and forecast runs for different applications.

You can access Manage Forecast Configurations from the Strategy & Policy Management dashboard under the Control & Tactical Center.

Forecast Run Type

The Forecast Run Type is the high-level template for forecast runs. Each run type has certain attributes that must be specified when the run type is being created. All configuration parameters default to the system default value when the run type is created. Use the Manage screen to modify the value of the configuration parameters for the run types.

Forecast Run

Forecast runs are created for a specific run type and are considered instances of the run type. The forecast run shares all the attributes of the run type and by default will have system default values for all configuration parameters. These parameters can be modified when a new run is being created in the Test screen.

Overview

The typical workflow for setting up run types, runs, and configuring the parameters is outlined in this section. Each step is described in detail in the sections that follow.

1. Create a run type in the Setup screen (train stop 1).

   If the aggregation status of the run type is Not Started, start the aggregation.
2. Create a run in the Test screen (train stop 2) and override the configuration parameters as desired. Then submit the run.
3. Once the run is successfully completed, review the summary of outputs from the estimation process by clicking the Summary link next to the run. This feature is available only if the run uses the Life Cycle forecast method.

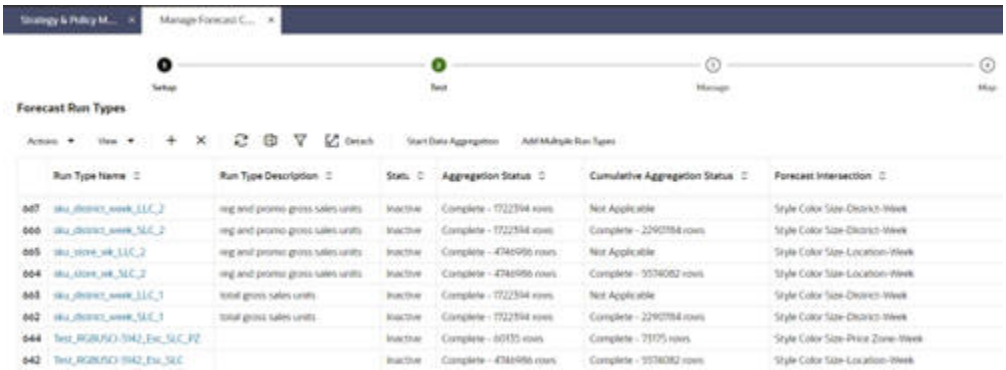
   > **Note:**
   >
   > You can create as many runs as desired. These are considered what-if runs and allow you to experiment with different values for configuration parameters and examine the impact on the estimated demand parameters.
4. Set the configuration parameters for the run type in the Manage screen (train stop 3). This will determine the values that should be used in all subsequent batch runs for the given run type.

   At the end of the implementation process, activate the run type to indicate that the run type should be part of the weekly batch process. During the weekly batch, a new run is created for each active run type.
5. Create a mapping between the run type and the application using the Map screen (train stop 4).

Setup

In the Setup screen in train stop 1, shown below, you can see the overview table of the existing run types and create new run types. In addition to Run Type Name, Run Type Description, Created On, and Created By, the table shows the following information about the run type.

Figure 11-12 Run Types



Run Type Status

The Status column in the run type overview table shows whether the run type is Active or Inactive. Active run types become part of the weekly batch process, and a new batch run is created and executed for each active run type. Run types can be activated in the Manage screen in train stop 3.

Aggregation Status

The Aggregation Status column in the run type overview table shows whether the data aggregation for the run type is Not Started, In Progress, or Complete. Instances of runs can be generated for a run type (in the Test screen in train stop 2) only if the aggregation is complete.

Cumulative Aggregation Status

The Cumulative Aggregation Status column in the run type overview table shows whether the cumulative data aggregation for the run type is Not Started, In Progress, or Complete. Cumulative data aggregation is only applicable for run types with forecast method Life Cycle.

Forecast Intersection

The Forecast Intersection column in the run type overview table shows the merchandise-location-calendar intersection for which the forecast is generated. The intersection includes price zone and/or customer segment if the run type has those dimensions.

Applications

The Applications column in the run type overview table shows the application(s) that the run type is mapped to. You can map run types and applications in the Map screen in train stop 4.

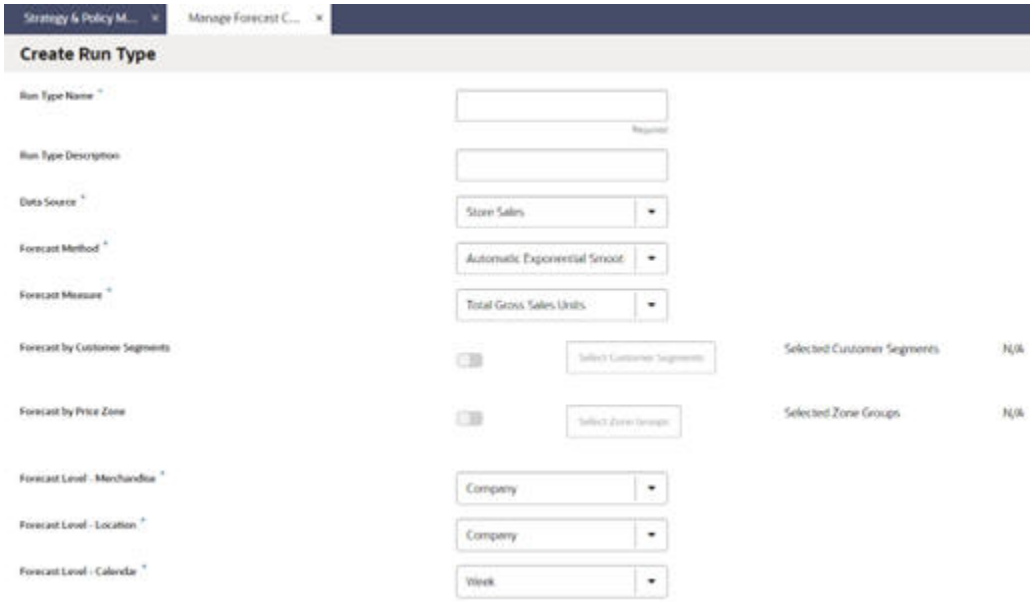
External Application Key

The External Application Key column in the Run Type Overview table shows whether the external application key is not yet assigned to the run type. You can assign the external application key for a run type (in the Map screen in train stop 4) only if the run type is being mapped to the application IPO-DF.

Create a Run Type

To create a new run type, click the **+** (plus) icon above the run type overview table. In the Create Run Type pop-up, as shown below, you must specify the value for the following fields.

Figure 11-13 Create Run Type



Run Type Name

This is a unique name and must be less than 30 characters. The run type name is used as an identifier when the forecast output is exported to other applications.

Forecast Method

A forecast method must be specified for each run type. All the runs created for the run type will use the selected forecast method. The methods that are currently supported are Sales & Promo, Life Cycle, Hierarchical Bayesian, and Automatic Exponential Smoothing. The forecast method must be chosen based on the specific use case and application.

Forecast Measure

This determines the measure that will be used as input by the forecasting algorithm. The final forecast will be generated for this measure. For the Sales & Promo and Life Cycle methods, the supported measures are total gross sales units, total net sales units, and regular & promotion gross sales units. For the Automatic Exponential Smoothing method, the forecast can be generated for other measures such as clearance sales, other sales, and returns.

Data Source

This determines which data will be used as input to the selected forecast method. The sources that are currently supported are store sales, warehouse transfers, warehouse shipments, and fact measures.

Customer Segment

Select Forecast by Customer Segments to generate the forecast by the merchandise/location/segment. If you turn on the switch, you must also click the button next to it in order to select one or multiple customer segments. It is recommended that you only select the segments that you want the forecast to be generated for (that is, un-select inactive segments) as this will make the data aggregation and forecast generation processes more efficient. Forecast by Customer Segment is applicable only for the Life Cycle method.

Price Zone Group

Select the Forecast by Price Zone to generate the forecast by merchandise/price zone. If you turn on the switch you must also click the button next to it in order to select one or multiple price zone groups. It is recommended that you only select the price zone groups that you want the forecast to be generated for (that is, un-select inactive price zone groups) as this will make the data aggregation and forecast generation processes more efficient. Forecast by price zone is applicable for the Life Cycle and Hierarchical Bayesian methods.

Forecast Level- Merchandise/Location/Calendar

Select the intersection that you want the forecast to be generated at. This will also determine the level at which sales data will be aggregated and flow through the various stages of forecasting. If you turn on the Forecast by Price Zone, the location level defaults to a level at the top of the location hierarchy. (This level is configurable and is determined by PMO_MIN_LOC_HIER_PROCESSING_LVL in the RSE_CONFIG table in the Manage System Configuration screen.)

Spread Profile Level — Merchandise/Location/Calendar

You can turn on the switch for the Spread Forecast to Day level only if the Forecast Level – Calendar is Week. Select the Spread Forecast Level for Merchandise/Location/Calendar only after turning on the switch for the Spread Forecast to Day level.

Data Aggregation

After you create a run type, you can see the aggregation status in the run type overview table. If the new run type has the same merchandise-location-calendar intersection as an existing run type, the aggregation status will be shown as complete as soon as you create the new run type because the data aggregation at that intersection has already been completed. This is applicable only for the run types that do not have a price zone dimension. For the run types that do have a price zone dimension, data aggregation must be done for each run type separately.

If any run type has an aggregation status of Not Started, click the **Start Data Aggregation** button above the run type overview table. This opens a pop-up where you can see a list of run types for which the aggregation process has not yet started, as shown in the next section..

You can select one or multiple run types from the top and/or bottom section in the pop-up and click **Submit** to start data aggregation. The Cumulative Aggregation process will be started once the aggregation process is submitted. It is only applicable for run types with forecast method Life Cycle.

Add Multiple Run Types

Click on the **Add Multiple Run Types** button. It will open a tab with a table consisting of available templates to create multiple run types at a time, as shown below. If the Enable column button is turned on and Process Flag column is null, then the corresponding row will be considered for creating run type. Click on the **Create Run Types** button to create multiple run types.

Click on + icon to add new process IDs to the table, as shown below. Select rows from the table and click on X icon to delete process IDs. User can also edit parameters within each row if Process Flag is null. Click the **Save** button to save the changes.

Figure 11-14 Add Multiple Run Types

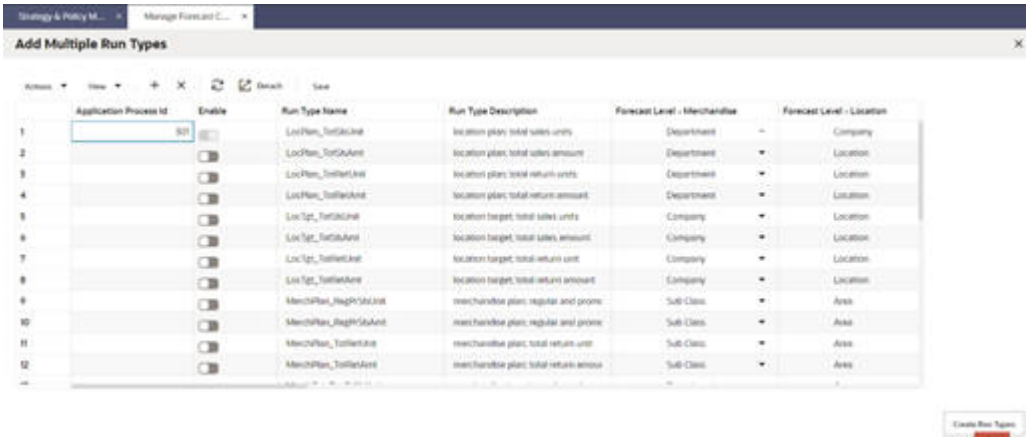
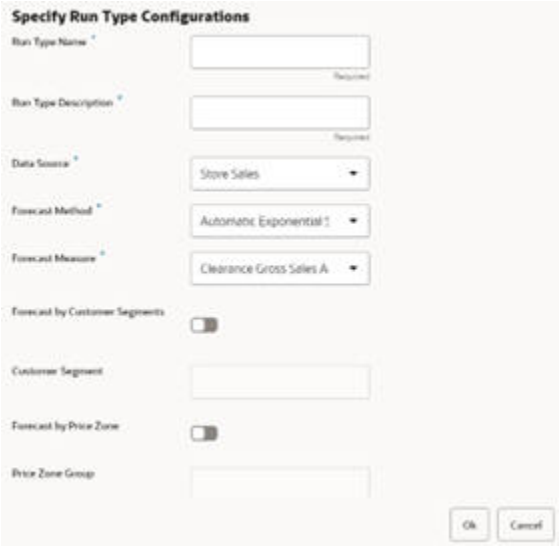


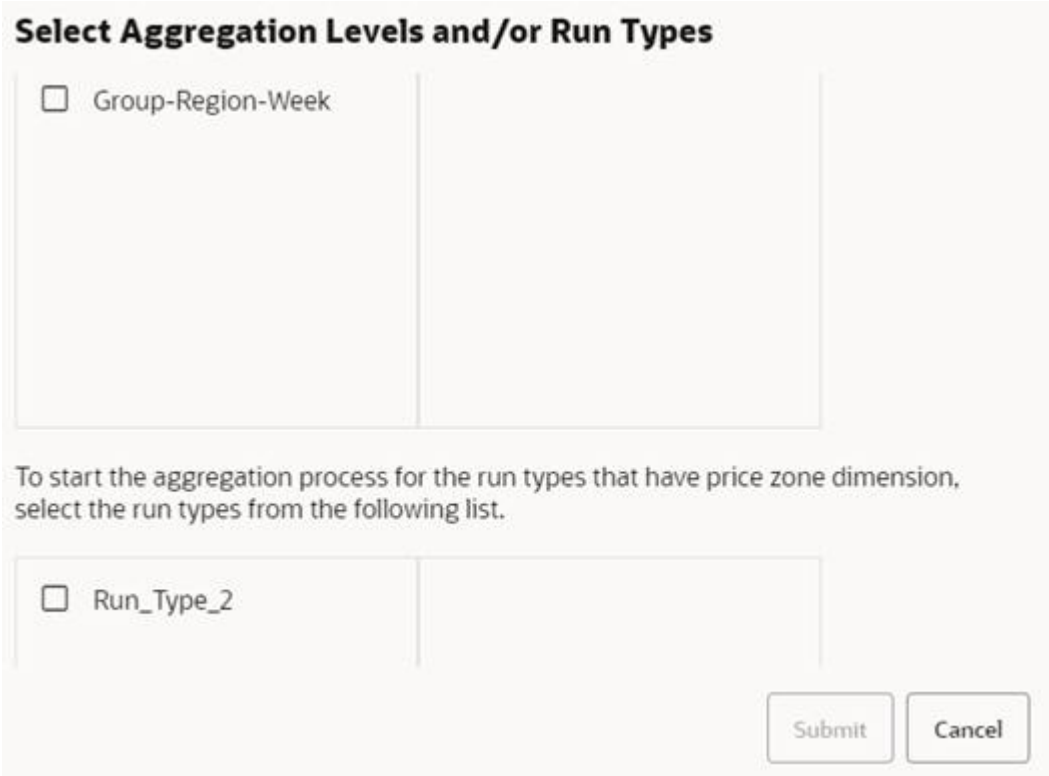
Figure 11-15 Specify Run Type Configurations



Delete Run Type

Select one or multiple rows in the run type overview table, as shown below, and click the **X** icon above the table to delete the run types. If a run type is activated or if a run type has a run that is in progress, it cannot be deleted.

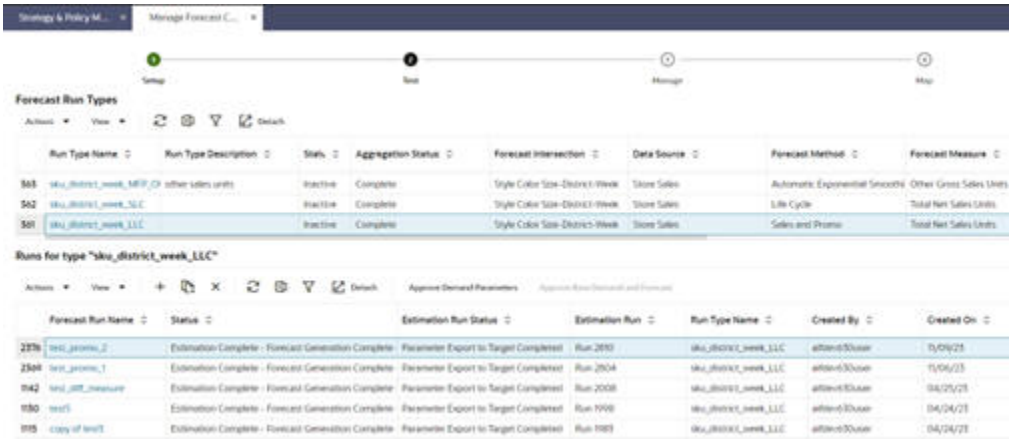
Figure 11-16 Select Aggregation Levels and Run Types



Test

After the run types are created and aggregation is complete, you can create and submit what-if runs and review the summary of the runs in the Test screen in train stop 2, as shown below. Performing what-if analysis is an optional step that can help you to tune the various configuration parameters. In addition, you can see the batch runs associated with each run type in this screen.

Figure 11-17 Test Screen



In the Test screen, the top table shows all the run types that have complete aggregation. This table shows the same information as the run type table in the Setup screen. The bottom table shows all the what-if runs and batch runs associated with the selected run type in the top table. In addition to Forecast Run Name, Created On, Created By, and Completed On, the table shows the following information about the run.

Status

The Status column in the run overview table shows the progress of the run by showing the status of the different stages of estimation and forecasting.

- Setup
- Preprocessing in Progress, Preprocessing Failed
- Preprocessing Complete - Estimation in Progress, Preprocessing Complete - Estimation Failed
- Estimation Complete - Base Demand in Progress, Estimation Complete - Base Demand Failed, Estimation Complete - Base Demand Complete, Estimation Complete - Base Demand Complete with Zero Rows
- Estimation Complete - Forecast Generation in Progress, Estimation Complete - Forecast Generation Failed, Estimation Complete - Forecast Generation Complete, Estimation Complete - Forecast Generation Complete with Zero Rows

Estimation Run

In all forecast methods, the estimation process runs initially to generate the demand parameters such as seasonality and elasticity. Subsequently, the forecast process runs to generate the base demand and the forecast using the outputs from the estimation process. The Estimation Run column shows the name of the estimation run that is behind the forecast run:

Run Type Name

Each run is associated with a run type. This column shows the name of the run type.

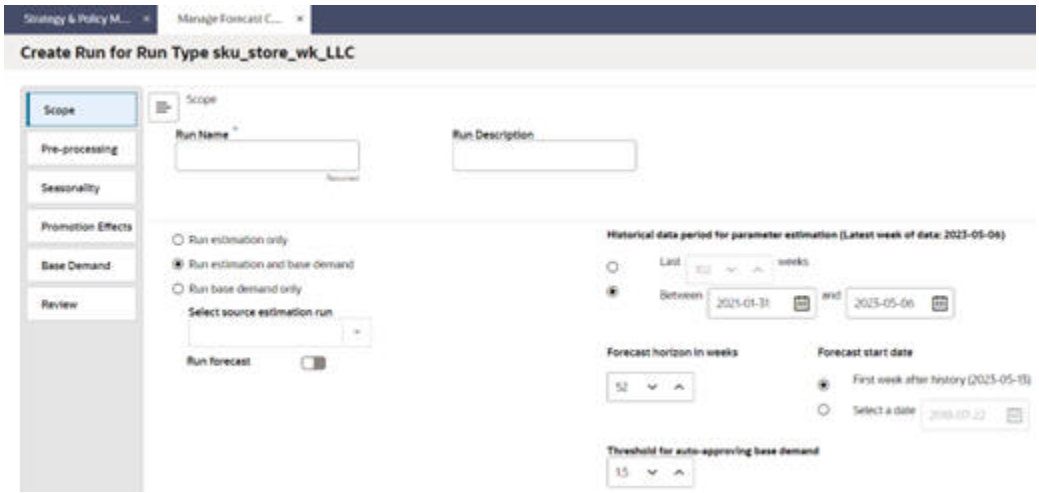
Results

This column shows the link to different output screens. The Summary link opens a tab that provides information about the different stages of demand parameter estimation. This link is enabled for all Life Cycle runs for which the estimation process has completed successfully.

Create What-If Run

To create a what-if run, first select a run type by selecting a row in the top table. Then, click the **+** (plus) icon in the bottom table to open the Create Run pop-up, as shown below.

Figure 11-18 Create a Run



The menu on the left shows the different stages of estimation and forecasting. These stages can be different, depending on the forecast method that was selected for the run type.

You can set the scope of the run, the run name, and the description in the first tab. In the left side of the Scope tab, select Run Estimation Only to execute the end-to-end process for estimating demand parameters. Select Run Estimation and Base Demand to execute the end-to-end process for estimating demand parameters and generating base demand. Alternatively, you can select Run Base Demand Only and then select an estimation run from the list of previously completed estimation runs. If no such estimation run exists, this option will be disabled. If a completed estimation run is selected, then the historical data period section will be populated based on the estimation run and will be disabled. To also generate the forecast, turn on Run Forecast. Forecast generation for a price zone run type will always be disabled.

In the right side of the Scope screen, you can select the period of historical data that should be included in the estimation process. This part will not be enabled if the Run Base Demand Only option is selected. You can either select a period by specifying the start and end date or by specifying the number of weeks to be used from the latest historical data. In the Scope tab, you can also specify the time period for the forecast generation by selecting a start date for forecast period and a forecast horizon. You can also set the threshold used for auto-approving base demand. The base demand values that are less than this threshold times the average rate of historical sales are approved.

In the other tabs on the left menu, you can review and override the configuration parameters for each stage of the estimation and the base demand/forecast. To override a parameter, select the row in the table and click the **Edit** icon above the table. For the Elasticity, Seasonality, and Base Demand tabs, you can override the parameters at different levels of merchandise and location. This feature is applicable only for the Sales & Promo and the Life Cycle run types. The levels that are valid for the override depends on the stage (particular tab), that is, for elasticity and seasonality this is closely tied to escalation levels.

You can click the Add a parameter override button to override parameters at different levels of merchandise and location. The Active flag allows the user to not use an override temporarily without having to delete it, as shown below. Once the parameter is overridden at the selected intersection, you can add the override at multiple intersections by clicking Save and Add Another button. This override parameters table is shown below.

The Base Demand tab override can be performed at three levels: at the global level, at a single merchandise-location parent level, and at the base demand level, which is essentially the aggregation level of the run, as shown in the following figures.

Figure 11-19 Override Parameters at the Intersection for Elasticity/Seasonality

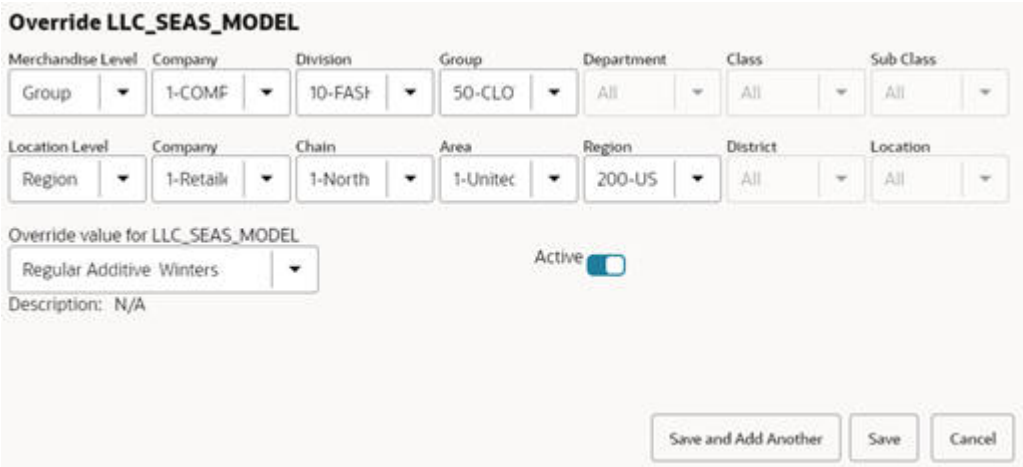


Figure 11-20 Override Parameters Table

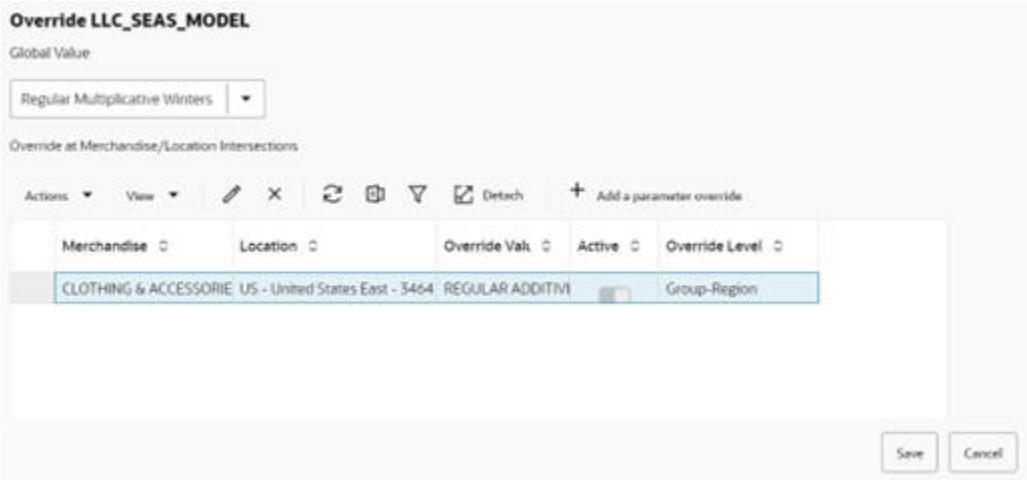


Figure 11-21 Override Parameters at the Intersection for Base Demand

In the Seasonality tab, in addition to the configuration parameters, you can also modify the escalation path by selecting or unselecting the merchandise-location intersections that will be used during escalation process.

In the Review tab, you can see all the parameters that were overridden and submit the run for the execution.

Figure 11-22 Add a Parameter Override Button

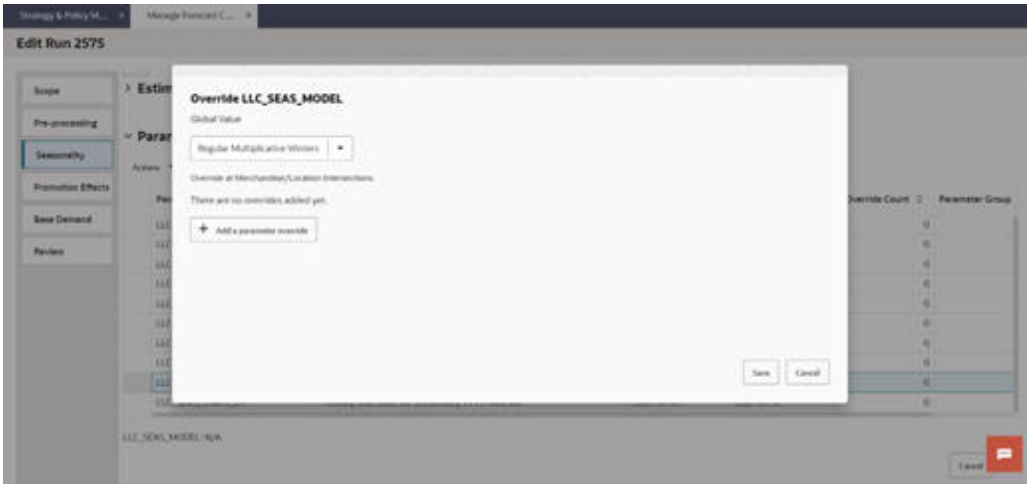
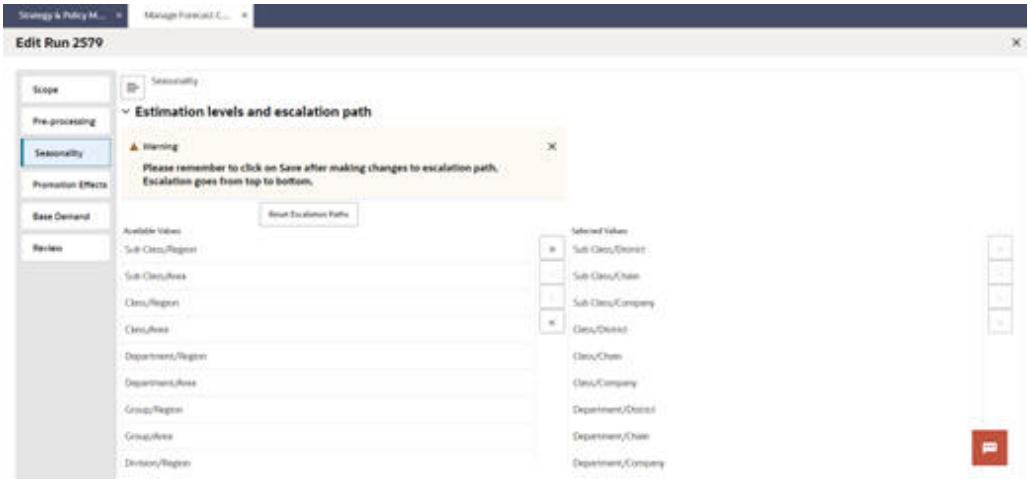


Figure 11-23 Modify Escalation Path for Seasonality/Elasticity



Duplicate Run

Select a row and click the **Duplicate** icon above the table to create a copy of an existing run. The run will be created in Setup status.

Delete Run

Select one or multiple rows and click the **X** icon above the table to delete a run or runs. If a run is in progress, it cannot be deleted. For each run selected to be deleted, you see three options: Cancel, Delete Forecast Run and Delete Forecast and Estimation Run. The Delete Forecast Run button deletes the forecast run. The Delete Forecast and Estimation Run button deletes the forecast run and in addition deletes the estimation run that is behind the forecast run. You can perform the Delete Forecast and Estimation Run only when there is no other forecast run tied with same estimation run. For the Automatic Exponential Smoothing run type, only the Delete Forecast and

Estimation Run can be performed, as each forecast run is mapped with a different estimation run.

Approve Demand Parameters

Select a row and click the Approve Demand Parameters button above the table to approve the estimation runs. You can approve demand parameters only for the runs where the estimation is complete. You can approve another run to replace the previously approved run. Once a run is approved, the Estimation Run column in the below table displays Approved next to the estimation run name. Only one estimation run per run type can remain approved at any point in time. The demand parameters generated by the approved run will be used in weekly forecast batch runs.

Approve Base Demand and Forecast

Select a row and click the Approve Demand and Forecast button above the table to approve base demand and forecast runs. Approve Base Demand and Forecast operation can be performed for the runs that has an approved estimation run and are in base demand complete or forecast generation complete status. Once a run is approved, the Forecast Run Name column in the below table displays Approved besides the forecast run name. Multiple forecast runs per run type can remain approved.

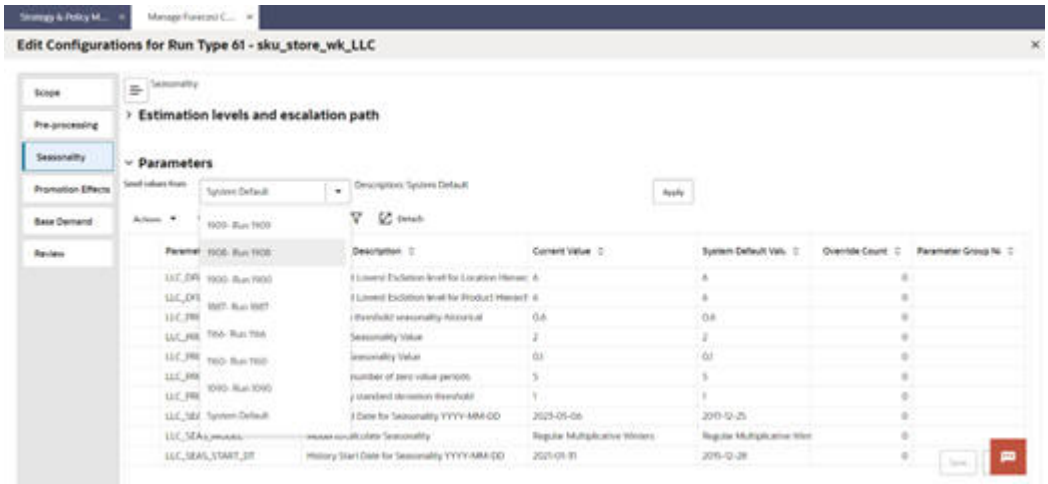
Manage

In the Manage screen in train stop 3 you can set the value of configuration parameters for the run type. These values will be used when creating and executing batch runs for each run type. In this screen, you can also activate or inactivate run types by selecting a row and clicking the Activate button above the table. Active run types become part of the weekly batch process, and a new batch run is created and executed for each active run type. For each active run type, once the batch run is approved, the output of the batch run will be exported to and consumed by the applications that are mapped to that run type. In order to auto-approve the batch runs, select a row and click on **Manage Auto-approve**. For the Life Cycle run type, the start date calculation can be disabled or enabled by clicking the Disable Life Cycle Start Date Calculation or Enable Life Cycle Start Date Calculation button, respectively.

To review and override the value of configuration parameters for a run type, select a row and click **Edit Configuration Parameters** above the table. This button will display as **View Configuration Parameters** if the run type is in active status. The Edit configuration pop-up is similar to the Create Run pop-up that is used to create forecast runs and has a tab for each stage of estimation and forecast. In each tab, you can do one of the following, as shown below. .

- Set the values of all configuration parameters in that stage, based on the system default for the run type. To do this, select System Default from the drop-down menu and click **Apply**.
- Set the values of all configuration parameters in that stage based on the values that were previously set in one of the what-if runs. To do this, select the desired run from the drop-down menu and click **Apply**.
- Override the value of a certain configuration parameter. To do this, select a row and click the **Edit** icon above the table.

Figure 11-24 Edit Run Type Configuration



Promotion Effects

This section describes the process for estimating the promotion lifts and is a continuation from Promotion Management. As part of the Promotion Effects module within the Sales & Promo forecast method (this module is activated when the PROMOTIONAL_SALES_AVAIL parameter is set to Y in the RSE_CONFIG table in the Manage System Configurations screen), select the following.

- Select features for the promotion lift estimation model.
- Model settings.

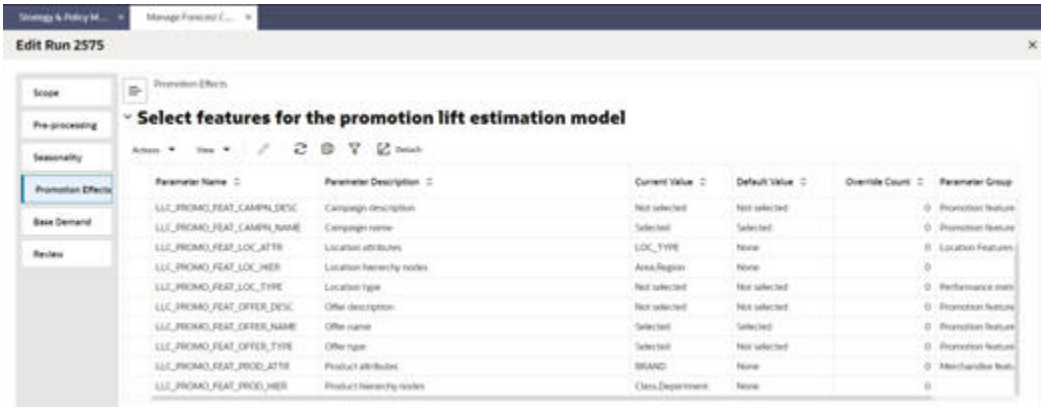
Forecast Configuration — Promotion Features

The Promotion Feature Selection screen within the Promotion Effects tab shows the various features that are available to the user. These various features can be classified according to the following categories.

- Promotion related: Campaign Name, Offer Name, Campaign Description, Offer Type, and Offer Description.
  - Campaign Name and Offer Name contain the campaign alias and offer alias values that are assigned through the Promotion Management module; these are always selected.
  - In addition to the alias values, one or more of the following features can be enabled if they are useful: Campaign Description, Offer Description, and Offer Type. Review the data to decide which of these features will be useful.
- Product Hierarchy: Promotion feature product hierarchy
  - Multiple levels of the product hierarchy can be selected as features to the model (for example, Class and Department). This allows the model to search for variations at each of the selected levels.
- Location Hierarchy: Promotion feature location hierarchy

  - Multiple levels of location hierarchy can be selected as features to the model (for example, Area and Region). This allows the model to search for variations at each of the selected levels.
- Product Attributes
  - Two attributes are supported. Select the relevant attributes that impact the promotion lift.
- Location Attributes
  - Two attributes are supported. Select the relevant attributes that impact the promotion lift.
- Performance Metrics: % sales on promotion, % weeks on promotion
  - In addition to the hierarchy and attributes, products can behave differently. Some products might be highly promotional compared to others. These features can be used to capture the variation lifts observed for different types of products, as determined by each performance metric.

Figure 11-25 Feature Selection for Promotion Effects



For a given retailer, start with a small selection of features that you expect to have an impact on promotion lifts and add new features to see if the addition of the feature improves the model performance. The details for reviewing the model output are provided after the model settings.

Forecast Configuration — Promotion Model Settings

Algorithm: Available options are Decision Tree and Generalized Linear Models (GLM).

Minimum samples: Minimum number of data points for each leaf node of the decision tree. This is used to ensure there are enough data points for each node.

Maximum depth: Number of levels between the leaf node and the top of the decision tree. Start with a value between 4 and 6, depending on the number of features being used and increase the number gradually to see if the model performance improves.

Partition: The model can be partitioned, based on one or more of the features used for building the model. Selecting a partition builds a separate model for each value corresponding to the chosen partition. Multiple features can be selected to determine a partition. Partitioning by higher levels in the product hierarchy and the location hierarchy or based on relevant product/location attributes can help to determine the promotions lifts more

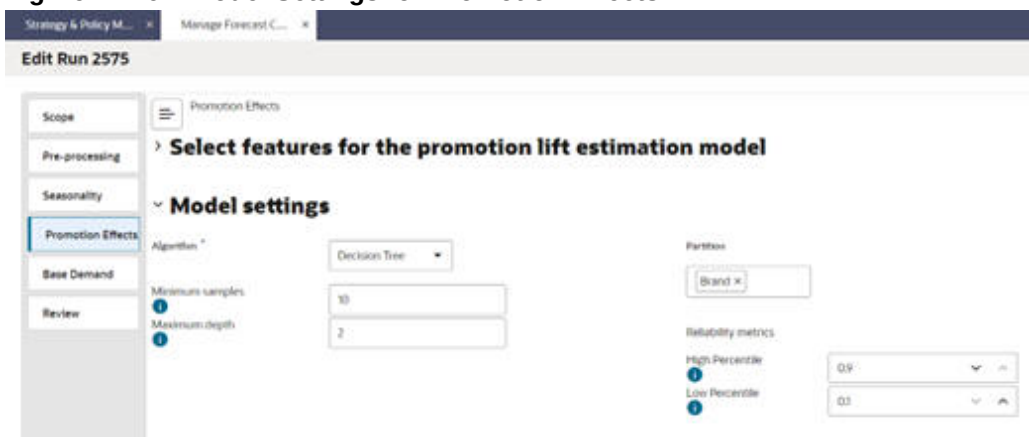
accurately. Lower levels of product/location hierarchy may not have sufficient data to create reliable parameters.

Reliability metrics:

High percentile: A value of 0.9 removes the top 10 percent of the observed lift values for building the model. Adjust the setting to remove outlier values with very high lift.

Low percentile: A value of 0.1 remove the bottom 10 percent of the observed lift values for building the model. Adjust the setting to remove outlier values with very low lift.

Figure 11-26 Model Settings for Promotion Effects



Review the output of the promotion effects in Innovation Workbench (IW):

- PMO_LLC_PL_MODEL table: Contains the details about the promotion model.
- RUN_HDR_ID: Estimation run id.
- CASE_TABLE, CASE_COLUMNS: Contains all the features available for the user to select.
- VIEW_NAME, VIEW_COLUMN_NAMES: Based on the user selection in the UI.
- Using IW, MODEL_NAME can be used to apply the model for a chosen time frame on the CASE_TABLE.
- Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE): Error metrics for the promotion lift values.
- Average Error (AVG_ERROR), Median Error (MEDIAN_ERROR), and Weighted Error (ERROR_WGT): % Error metrics for the promotion lift values.
- ERROR_WGT: % Error weighted by sales units.

Figure 11-27 Promotion Effects Output Review in IW

RUN_ HDR_ID	FCST_RUN _TYPE_ID	FINALIZED _MODEL	MODEL _NAME	SETTINGS _FILE	RMSE	MAE	AVG_ERROR	MEDIAN_ERROR	ERROR_WGT
1558	241	N	RSE_LLC_PL_	TMP$RSE_LLC_PL_SETT$0000001558	0.74	0.55	0.47	0.22	0.30
1440	241	Y	RSE_LLC_PL_	TMP$RSE_LLC_PL_SETT$0000001440	0.79	0.55	0.65	0.18	0.34
1303	201	N	RSE_LLC_PL_	TMP$RSE_LLC_PL_SETT$0000001303	0.00	0.00	0.00	0.00	0.00
1762	341	N	RSE_LLC_PL_	TMP$RSE_LLC_PL_SETT$0000001762					
1600	241	N	RSE_LLC_PL_	TMP$RSE_LLC_PL_SETT$0000001600	0.74	0.55	0.47	0.22	0.30
1743	341	N	RSE_LLC_PL_	TMP$RSE_LLC_PL_SETT$0000001743					

Map

In the Map screen in train stop 4, you can map run types to applications. There can be many-to-many mappings between the run types and applications. For example, if IPO-DF requires a forecast for two different measures (for example, total gross sales units and total net sales units), which may or may not be at the same forecast intersection, there will be two run types and both must be mapped to IPO-DF. Also, if two different applications (for example, IPO-DF and IPO-IO) require a forecast for the same measure and at the same forecast intersection, then both the applications will be mapped to the same run type.

You can select more than one run type at a time to be mapped to an application except when the applications are IPO-DF and RI/RMS, as shown below.

You can assign an external application key to the run types in this screen. An external application run type key value can only be assigned to a run type when the selected application is IPO-DF. Two run types that have the exact same forecast intersection, forecast measure, and data source, but different forecast methods (Sales & Promo vs. Life Cycle) can have the same external application run type key.

Basically, for each external application run type key, there can be at most one Sales & Promo run type and at most one Life Cycle run type, and those two run types must have the same forecast intersection, forecast measure, and data source.

To create a new mapping, click on the **+** (plus) icon above the table in Map screen. Select an application and one or more run types (based on what the application allows) and click **Save**. To delete a mapping, select a row and click the **X** icon above the table.

Figure 11-28 Map Run Types to IPO-DF as a Single-Select Option

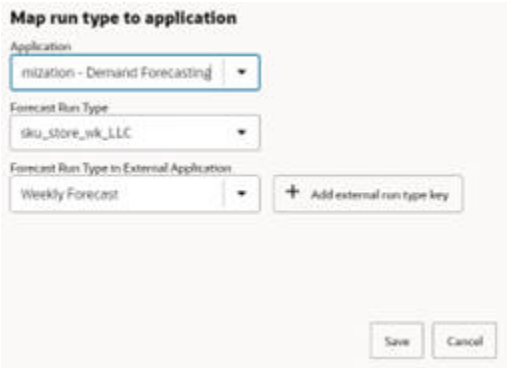
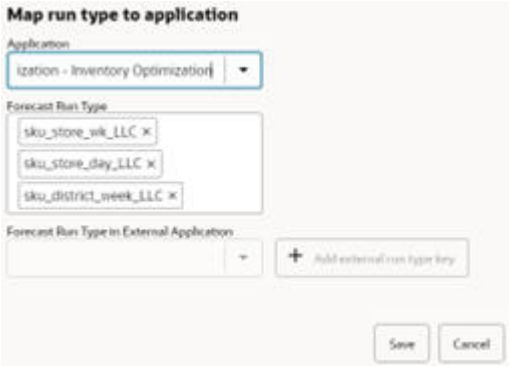


Figure 11-29 Map Run Types to IPO-IO as a Multi-Select Option



Manage Master Data

You can access Manage Master Data from left side Tasks list on the Strategy & Policy Management screen as shown below. This takes you to the Sales Plans and Flexible Groups tabs.

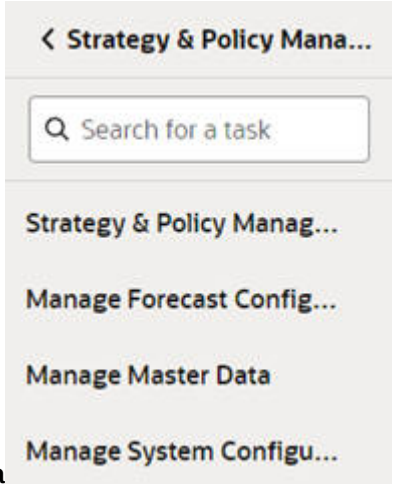


Figure 11-30 Manage Master Data

In Sales Plans tab, as shown below, you can select a row and edit the Merchandise Level/Location Level or reset the value to the default levels. Sales Plans are used for in-season (Bayesian) forecasting within the Automatic Exponential Smoothing run types.

Figure 11-31 Sales Plans Tab

The Flexible Groups provide the grouping of merchandise/locations based on different criteria. These are used in the escalation path to estimate seasonality/elasticity values for Sales & Promo and Life Cycle run types. Flexible Groups ensure that the seasonality/elasticity estimation results are separately calculated for each group of merchandise/location. Group Set is the parent (that is, the grouping definition). Partitions are the children (that is, the groups within each group set). In the Flexible Groups tab, you can create a new flexible group set and edit and delete an existing flexible group set, as shown below. While creating a new flexible group set, you can add a new group set name (which must be unique), group set description, and select an existing run type from the drop-down list, as shown below . Once this step is done, the partitions (groups) inside the group set must be created. The partitions can be provided through IW using custom calculations or loaded through the staging interface (RSE_FLEX_GROUP_DTL_STG). By clicking the icon near group set name, you can see the partitions under it, as shown below. .

Figure 11-32 Flexible Groups Tab

Figure 11-33 Create Flexible Group Set

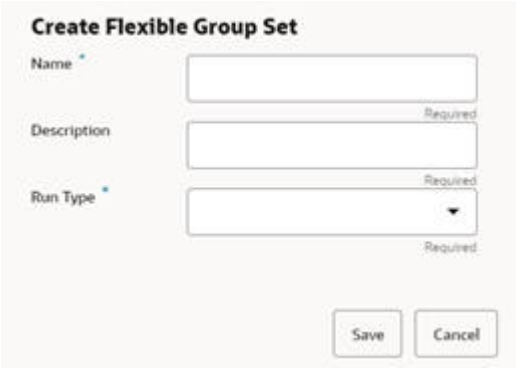
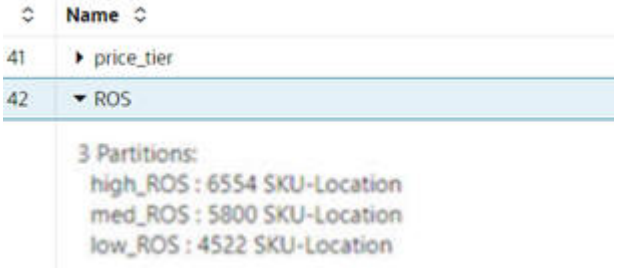
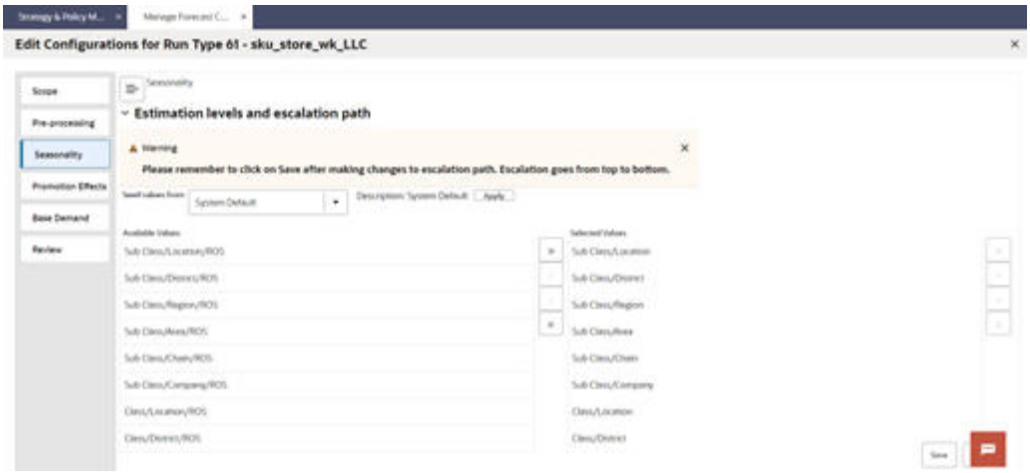


Figure 11-34 Partitions (Groups) under a Group Set



Once a new flexible group set is created and mapped to a run type, new escalation levels corresponding to the flexible group set will be automatically added under the Available Values column in the Estimation Levels and Escalation Path section of the Seasonality tab for the corresponding run type in Edit Configuration Parameters within the Manage screen, as shown below. .

Figure 11-35 Escalation Levels Based on Group Set are Available

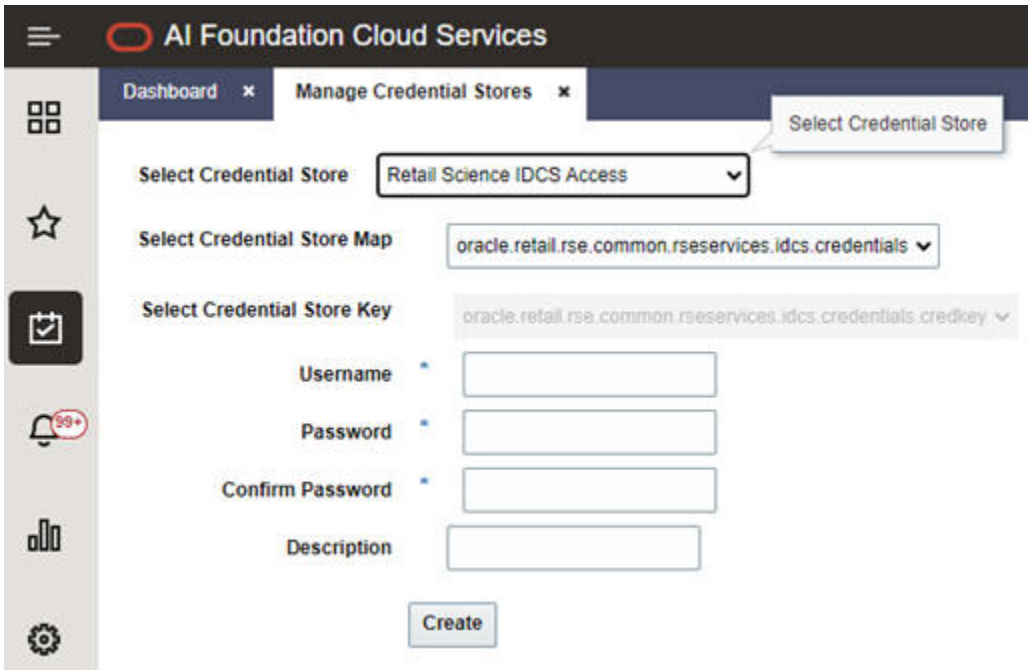


When a new What-If run is created from the Test screen under the corresponding run type, the run will also contain the additional escalation levels corresponding to the flex group set and will be displayed under the Available Values column in the Estimation Levels and Escalation Path section of the Seasonality tab.

Manage Credential Stores

Oracle Customer Engagement web service-related credentials are managed in the credential stores using the interface shown below. The administrator can use this screen to configure the name and password. This information is used to generate the authentication key that is sent as part of the message to Oracle Customer Engagement.

Figure 11-36 Manage Credential Stores



The Credential Stores dialog box requires the following information:

- Label - describes the credential store map and key
- Username - user name provided by Oracle Customer Engagement
- Password - password provided by Oracle Customer Engagement
- Confirm Password - prompt to confirm user password
- Description - used by the administrator to describe the credential store used for Customer Segment integration with Oracle Customer Engagement.

Attribute Extraction and Attribute Binning

For details about the screens of Attribute Extraction, refer to Attribute Extraction.

12
Forecasting Methods in AIF

This chapter discusses the forecast methodologies available in Oracle's Next Generation AI Foundation (AIF). The intent of this chapter is to highlight forecast algorithms that are available with Oracle's Retail & Analytics Planning (RAP) cloud services within AIF. RAP solutions consist of Assortment Planning (AP), Assortment Space Optimization (ASO), Inventory Planning Optimization (IPO), LifeCycle Pricing Optimization (LPO), Merchandise Financial Planning (MFP), and Retail Insights (RI). To facilitate the retailer's planning and analytics needs, RAP solutions may be deployed with the forecasting capabilities of AIF. The utilization of forecasting methodologies discussed in the chapter will differ based on the cloud services implemented and the unique requirements of each retailer's business.

Some Forecasting Concepts

In the Next Generation architecture, the AI Foundation supports all the RAP cloud services by providing a single data point of entry to multiple cloud services. From this data, the Forecasting Engine fuels different forecasts at the different levels required by the different processes of the different cloud services. MFP and AP utilize the base functionality of the Forecast Engine. IPO utilizes the Forecasting Engine and the Forecasting Workflow. IPO offers enhanced forecasting review and workflow management that maintain and adjust forecasts at the item level. The other solutions of ASO, LPO, and RI may be configured to utilize the

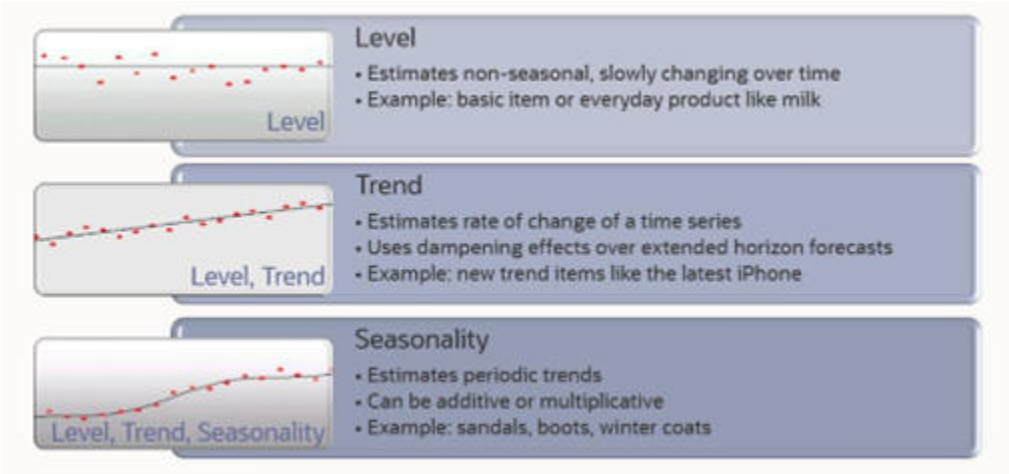
AIF forecasts as best serves the retailers business objectives. AIF forecasts enhance retailer supply chain's planning, allocation, and replenishment processes for a more profitable and customer-oriented approach. This unconstrained demand may be sent to execution and planning solutions to be transformed into revenue-based metrics and strategies. Demand forecast provides a single unified forecast with minimal human intervention. The Forecasting Engine generates accurate forecasts of demand as automatically as possible by using a variety of predictive techniques.

The AIF Forecasting Engine utilizes several concepts when forecasting for planning solutions. These concepts are integral parts of the various forecasting methods available in RAP cloud solutions.

Exponential Smoothing

Exponential Smoothing (ES) evaluates the basic demand patterns of Level, Trend, and Seasonality and projects these into the future. They are labelled smoothing since they use weighted averages on historic data, and the weighting decays exponentially as we go back in time: more recent data is weighted more heavily than past data.

Figure 12-1 Demand Patterns of Level, Trend, Seasonality



Automatic Method Selection

Multiple forecasting methods are valuable if the appropriate model can be selected in an accurate and efficient manner. Oracle Retail has developed different meta-methods that can automatically select the best method among competing models. These models include level information, level and trend information, and level, trend, and seasonality information. The optimal smoothing parameters for each model form are determined automatically (that is, greater smoothing is applied to noisier data). The final selection between the resulting models is made according to a performance criterion that involves a trade-off between the model's fit over the historic data and its complexity. The amount of available historic information can affect the complexity of the model. Automatic Exponential Smoothing (Auto ES) and Seasonal Exponential Smoothing (Seasonal ES) chooses between competing models according to performance criterion, looking at the tradeoff between history and its complexity.

Seasonal Regression Analysis

Regression analysis is another standard technique used in prediction. Regression uses a least-squares estimator to fit a model of predictor variables to another set of target variables. Seasonal regression is Oracle Retail’s extension to the standard seasonal forecast method for application between one and two years of history. This is included in the Automatic Exponential Smoothing (Auto ES) and Seasonal Exponential Smoothing (Seasonal ES) method.

Bayesian Analysis

Bayesian Analysis is effective where no previous sales history exists, such as with new items or items with short lifecycles. It requires a custom plan to serve as the initial sales plan. It may be configured to lend more weight to recent history as it accrues. Once enough history exists, standard forecasting methods must be used, such as Exponential Smoothing. This enables a plan with current data to create a new forecast.

Forecasting Methologies

From the various forecasting concepts detailed above, the following forecasting methodologies are created. These forecast methods are available to the user when creating a forecast run type within the AIF Control and Tactical Center -> Strategy and Policy Management -> Manage Forecast Configurations screen. Different methods have different data needs. Refer to the Data Requirements for Various Forecast Methods section from the AIFCS Implementation guide for more information.

- Automatic Exponential Smoothing
- Sales & Promo
- Life Cycle
- Hierarchical Bayesian

Automatic Exponential Smoothing Method

The primary goal for a demand forecasting solution is automation. The Automatic Exponential Smoothing method (Auto ES) automatically selects an exponential smoothing model to a time series. Auto ES is a proprietary forecasting algorithm that compares the varying ES methods against each other for the best fit while being the least complex.

Auto ES considers a collection of six candidate models:

1. Simple Exponential Smoothing (Simple ES):
   - Simple ES ignores seasonality and trend features in the demand data.
   - Simple ES is effective when there are short time horizons in which less than one year of historic demand data is available or data is untrended and unseasonal.
   - Simple ES produces a flat line (Level only) forecast.
2. Holt Exponential Smoothing (Holt ES):
   - Holt ES captures trends but is non-seasonal.
   - Holt ES treats data as a one-dimensional trend (up or down).
   - Holt ES provides forecast estimates by combining an estimated trend for the forecast horizon and the smoothed level at the end of the series.
   - Holt ES is effective for short or long horizons and provides a damping effect for long horizons.
3. Additive Winters Exponential Smoothing (Additive Winters ES):
   - Additive Winters ES extracts seasonal indexes assumed to have additive effects on the de-seasonalized series. Effects are applied additively to levels and trends.
   - Additive Winters ES requires sufficient data of at least two years of history.
   - Additive Winters ES controls the smoothing of the components of level, trend, and seasonality.
   - Additive Winters ES is effective when there are zeros in the history.
4. Multiplicative Winters Exponential Smoothing (Multiplicative Winters ES):

- Multiplicative Winters ES extracts seasonal indexes assumed to have multiplicative effects on the de-seasonalized series. Effects are applied multiplicatively to levels and trends.
- Multiplicative Winters ES requires sufficient data of at least two years of history.
- Multiplicative Winters ES controls the smoothing of the components of level, trend, and seasonality.
- Multiplicative Winters ES does not work when there are zeros in the history.

5. Seasonal Regression:
   - Seasonal Regression is for sales forecasts based entirely on sales from the same time period last year. Forecasting using only last year sales involves simple calculations and often outperforms more sophisticated seasonal models. –
   - Seasonal Regression performs best when dealing with highly seasonal sales data with a short sales history.
   - Seasonal regression is designed to address retailer needs when sales history is between 56 to 104 weeks.
   - Seasonal Regression uses simple linear regression with last year's sales as the predictor variable and this year's sales as the target variable.
   - If there are significant shifts from year to year, Seasonal Regression will learn the shift and appropriately weigh last year's data, keeping the same shape but adjusting the scale.
6. Bayesian:
   - The Bayesian method combines historic sales data with sales plan data and is effective for new products with little or no historic data. The sales plan provides expert knowledge. The sales plan defines the shape of the selling profile or lifecycle and the scale of the total quantity expected to be sold over the plan's duration. The Bayesian method merges a customer's sales plans with any available historical sales.
   - The Bayesian method starts with a seasonal sales plan and evaluates the following conditions to calculate the forecast:
     - When no sales history exists, the forecast equals the sales plan.
     - As sales data becomes available, the forecast adjusts, and the scale becomes a weighted average reflecting known history.
     - Once sufficient sales data is collected, it can be switched to one of the different standard forecasting methods.

During configuration, an implementer may determine if only Seasonal Exponential Smoothing (Seasonal ES) methodologies are appropriate and remove access to Simple ES and Holt ES methods.

Sales & Promo Method

This method is useful for retailers wanting to account for outliers, out-of-stock, and promotion events. This is a causal approach and goes through multiple stages to generate the final forecast. Typically, the data requirement for this method is at least two years of sales history. This is among the best methods in AIF that can produce

reliable forecasts at the lowest levels of merchandise and location hierarchies with only the minimum requirement of sales data. The stages for Sales & Promo method:

1. Preprocessing corrects bad data points that represent unusual sales behavior, so that the demand forecast does not replicate undesirable patterns. Possible causes of history adjustments include out of stocks, outliers, or events. During preprocessing, the quality of data is analyzed, and faulty data points are corrected before further calculations. This provides pre-processed data to produce reliable demand parameters for the next process steps.
2. Promo Estimation consists of heavy data mining of the relationships between historical sales (sales lift) and promotions (different promotion-relevant features). The two available models to quantify the impact of promotions on demand are:
   - Decision Trees:
     - Tree-based Machine Learning classifier model that predicts an outcome based on several input features.
   - Generalized Linear Model:
     - Regression-based Statistical model that predicts an outcome based on the values of another set of input variables.
3. Seasonality Estimation utilizes various machine learning techniques to understand the time series in which data experiences regular and predictable changes that recur every calendar year. This reflects a sales pattern that repeats over a one-year period based on the time of year (season). The estimation of seasonality occurs at different Escalation levels which are aggregated sales at different merchandise and location hierarchy levels. The four options for estimating seasonality are:
   - Additive Winters Exponential Smoothing (Additive Winters ES):
     - Additive Winters ES extracts seasonal indexes assumed to have additive effects on the de-seasonalized series. Effects are applied additively to levels and trends.
     - Additive Winters ES requires sufficient data of at least two years of history.
     - Additive Winters ES is effective when there are zeros in the history.
   - Multiplicative Winters Exponential Smoothing (Multiplicative Winters ES):
     - Multiplicative Winters ES extracts seasonal indexes assumed to have multiplicative effects on the de-seasonalized series. Effects are applied multiplicatively to levels and trends.
     - Multiplicative Winters ES requires sufficient data of at least two years of history.
     - Multiplicative Winters ES does not work when there are zeros in the history.
   - Auto Seasonal:
     - Selects the best method based on the inverse of Root Mean Square Error of Additive Winters ES or Multiplicative Winters ES.
   - Blend Seasonal:
     - Utilizes the inverse of Root Mean Square Error as the weights for the two methods of Additive Winters ES and Multiplicative Winters ES and combines the weighted seasonal indexes.
4. Base Demand Estimation calculates the average demand signal after the cleansing of data and removal of promotional and seasonality effects. This is generated at the same merchandise and location hierarchy level as the required forecast. The six options for estimating base demand are:

- Simple Exponential Smoothing (Simple ES):
  - Simple ES ignores seasonality and trend features in the demand data.
  - Simple ES produces a flat line (Level only) estimate.
- Holt Exponential Smoothing (Holt ES):
  - Holt ES captures trends but is non-seasonal.
  - Holt ES treats data as a one-dimensional trend (up or down).
  - Holt ES provides estimates by combining an estimated trend for the forecast horizon and the smoothed level at the end of the series.
  - Holt ES provides a damping effect for long horizons.
- Moving Average:
  - Moving Average is calculated as a simple moving average estimate from past time frames.
  - Moving Average gets priority over Simple ES and Holt ES when the data is too noisy.
- Poisson:
  - This helps in estimating for low sellers assuming a Poisson distribution on the sales history.
- Pick Best:
  - System determines which method (Simple ES, Holt ES, Moving Average, or Poisson Base Demand) provides the best fit based on the inverse of Root Mean Square Error.
- Blend:
  - System generates estimates from various methods (Simple ES, Holt ES, Moving Average, or Poisson Base Demand) and uses the inverse of Root Mean Square Error as weights to combine them.

5. Forecast Generation calculates the demand forecast combining the output of all previous steps. This forecast multiplies base demand, seasonality, and promotional lifts. In this manner, changes in seasonality and promotional effects can be evaluated and combined with the de-seasonalized base demand.

Life Cycle Method

This method is useful for retailers wanting to account for price elasticities and markdowns. This is a causal approach and goes through multiple stages to generate the final forecast. Typically, the data requirement for this method is more than one year of sales, inventory, receipts, and prices. This is among the best methods in AIF that can produce reliable forecasts at the lowest levels of merchandise and location hierarchies accounting for prices and markdowns but with a larger breadth of data needs.

The stages for Life Cycle method:

1. Preprocessing filters bad data points that represent unusual sales behavior, low inventory, or irregular prices so that the demand forecast does not replicate undesirable patterns. The filtering of historical data defines a subset of data to produce reliable demand parameters for the next process steps.

2. Event Lift Estimation involves heavy data mining of the relationships between historical sales and events. This estimates the traffic lift during promotional or holiday events. The estimation of traffic lifts occurs at different Escalation levels which are aggregated sales at different merchandise and location hierarchy levels.
3. Elasticity Estimation is generated from calculations of historical demand sensitivity based on markdown or promotional prices changes. The estimation of elasticity occurs at different Escalation levels which are aggregated sales at different merchandise and location hierarchy levels.
4. Seasonality Estimation utilizes various machine learning techniques to understand the time series in which data experiences regular and predictable changes that recur every calendar year. This reflects a sales pattern that repeats based on the time of year (season). The estimation of seasonality occurs at different Escalation levels which are aggregated sales at different merchandise and location hierarchy levels.
5. Base Demand Estimation calculates the average demand signal after the filtering of data, accounting for traffic lifts, and removal of elasticity and seasonality effects. This is generated at the same merchandise and location hierarchy level as the required forecast.
6. Forecast Generation calculates the demand forecast combining the output of all previous steps. This forecast multiplies base demand, seasonality, elasticity, and traffic lifts. In this manner, changes in seasonality, elasticity, and events can be evaluated and combined with the de-seasonalized base demand.

Hierarchical Bayesian Method

Hierarchical Bayesian method is useful for retailers wanting to optimize regular price of items. This is a Bayesian-based approach that uses regular price as one of the features to generate forecast. Typically, the data requirement for this method is more than one year of sales and prices.

The stages for Hierarchical Bayesian method:

1. Data Preparation defines the levels at which data is aggregated on merchandise, location/price zone, customer segment, and time dimensions and the duration of the data used for parameter estimation.
2. Preprocessing filters the historical data and makes the first determination of item eligibility.
3. Training enables the user to select the configurations and features that can be used for model training. Note that some features are mandatory.
4. Forecast Generation generates the ML models that can be later used by applications consuming demand forecasts.

13
Glossary

Assortment Cluster

Store cluster definitions that are used in Category Manager, which are defined at the product category level.

Bay

A physical unit of fixtures in which the configuration is the same from left to right. A fixture configuration can change across bays but not within a bay. Retailers often lay out POGs on a discrete number of bays. Bays can be considered a unit of macro space like linear ft/cm.

Capping

Products are stacked with units of different orientation. The bottom unit is in the main orientation, and the units above use a capping orientation. This applies only to shelves. Only simple capping is supported.

Current/Historical POG Assortment

The list of products from a specific POG that is imported from a third-party POG tool. This assortment is POG-specific; the stores that received this POG have this assortment. A master list for a POG set can be created by finding the union of all products that are on the POGs in the set.

Display Style, Display Style Type

This provides the discrete unit of display for a given product. Display style types include single/unit, tray, case, and pallet. Display style types can also describe peg vs. shelf. These are set up in MSM.

For a given product, one or more display styles is established in MSM. Each one has its own set of size and spacing data, as well as the number of total units. For example, a specific SKU may have two display styles: single/unit and tray. Unit then corresponds to a single unit of the product. Tray is then 12 units. The number of units per tray depends on the product.

Note that MSM uses the term unit for a single unit display. In order to describe the number of units, ASO uses the term single/unit instead.

Empty Planogram

An empty POG is a POG skeleton. The bays and fixtures are defined but no products have been placed on the fixtures.

Facing

A facing is the smallest unit of micro space that is allocated to a product. On a shelf, facings are counted horizontally. In other words, it is the number of units across. On a pegboard or a freezer, facings are the number of total units that the customer can see when looking at the fixture.

Finalized Assortment

This is the final approved assortment.

Finger Space Above

The product-specific space that is required between the top of a product's facings and the bottom of the shelf above. This accounts for the space required to remove the product from the fixture. When available, the specified space is added to the top of the top unit in the product's facings in the *z* direction. This applies to pegboards, freezer chests, and shelves. In MSM, this is also called finger space above.

Finger Space Behind

The product-specific space that is required on the front and back of units of a product. When available, the specified space is added on either side of the unit in the *y* direction; half of the quantity goes on the front and half of the quantity goes on the back. This applies to shelves only. In MSM, this is called a gap in the *y* direction.

Finger Space Beside

The product-specific space that is required on the left side and right side of units of a product. When available, the specified space is added on either side of the product's facings; half of the quantity goes on either side in the *x* direction. For example, if there are two facings of product A with finger space beside equal to 1 inch, then horizontally it looks as follows: one-half inch, product A, one-half inch, one-half inch, product A, one-half inch. This applies to pegboards, freezer chests, and shelves. In MSM, this quantity is called a gap in the *x* direction.

Fixture

A fixture is defined as equipment such as a set of shelves, a pegboard, or a freezer chest. A single shelf is not a fixture, nor is a single peg a fixture.

Fixture Configuration

This provides details about the dimensions of shelving (that is, the number of shelves and the height, width, and depth of each one), freezer chests (that is, the height, width,

and depth) or pegboard (in this case, the height, width, and depth equal the peg length, which may also have peg hole spacing details).

Fixture Type

This is the attribute that defines what kind of fixture is in a POG: shelf, pegboard, or freezer chest.

Inter Product Gap

The product-specific space that is required to separate adjacent products on a shelf. This accounts for space that holds product dividers. When available, the specified space is added to the left and right of the entire block of a product; half of the quantity goes on either side of the run of that product in the *x* direction. For example, if there are two facings of product A with inter product gap equal to 1 inch, then horizontally it looks as follows: one-half inch, product A, product A, one-half inch. This applies to pegboards, freezer chests, and shelves. In MSM, this is called finger space beside.

Item Performance Index (IPI)

This is a value generated by Category Manager that provides indication of the relative importance of products within an assortment. A value of 1 means the product is average, a value of 1.3 means it is 130% of average, and a value of 0.5 means it is 50% of average. ASO uses these values optionally as attributes or weights in the optimization objective function.

Mandatory Items

These are items from the preliminary assortment that cannot be dropped.

Optimization Level

This is the location level for micro-optimization runs and is set for a specific optimization run. The options are cluster or store.

Optimization Location

This is a specific location node defined for optimization. It can be a store or a store cluster.

Order Point (OP) and Order Up to Level (OUTL)

These are proxies for the replenishment triggers and are used in the inventory and sales model. These values are not directly input from the client feeds, but are calculated from theinput data.

Orientation, Legal Orientation, Default Orientation

This describes how a product's package faces the front of the fixture. The orientation has two pieces of information, the part of the package that is to the front (front, back, top, bottom, left,

right) and how it is rotated in degrees (0, 90, 180, 270), for a total of 24 possible orientations. How a product is oriented determines how much space it takes in the *x,y,z* directions on a fixture.

Legal orientations are all orientations that can be used for the product.

The default orientation is the legal orientation that is normally used for the product.

Planogram (POG)

The layout of product on shelves or pegboards, which has an underlying fixture assumption (for example, number of shelves or overall size).

A third-party POG typically includes a visual depiction of the product layout with images. POGs are the underlying data and not the pictures.

Planogram (POG) set

Current and historical POGs in the same POG category or subcategory and with the same seasonal attribute.

Preliminary Category Manager Assortment

A product list created in Category Manager by product category, this varies by assortment cluster. The preliminary assortment is a list of products that are eligible. A preliminary assortment has a master list that is the union of all the cluster-specific preliminary assortments.

Preliminary assortments from Category Manager are forward looking and can include new products or new product assignments and store assignments.

POG Sub-Category, POG Category, POG Department

These are levels on the POG hierarchy and are used to organize POGs. For example, the leaf to root path in the POG hierarchy is Grocery -> Crisps and Snacks -> Crackers.

Product Stacking Height Limit

This defines a product-specific limit on how high a given product can be stacked or nested in the *y* direction on any shelf. This applies only to shelves.

Receive Up-To Level (RUTL)

This denotes the target level of inventory when an order is generated. It is calculated differently based on the replenishment method.

Re-Order Point (RP)

This represents the inventory level below which an order must be triggered. It is calculated differently based on the replenishment method.

Replenishment Policy

The replenishment policy for each sku-location is characterized by the two parameters: RP and RUTL, calculated for that sku-location.

Safety Stock

The inventory held at a location to mitigate the risks of stockouts caused by variability in forecasted demand. It is calculated differently based on the replenishment method.

Season/Seasonal Attribute

This refers to a specific year-independent time period for a Category Manager assortment and a POG set. Examples include spring, holiday, back to school, and year-round.

Sell-Through

This is calculated as the ratio between the historical sales (of the past weeks) and the sum of the historical sales (of the past weeks) and forecast sales (of the future weeks). Future weeks are the weeks until the end of the current season or the end of the life cycle of the sku-store, whichever is earlier.

Service Level

This is defined as the percentage of time that the stock must be able to meet demand. The service level is used in the safety stock calculations for the Dynamic and Poisson replenishment methods.

Smart Start Process

The Smart Start Process (SSP) is a back-end process that assigns shelves to a Partial Shelf Fixture Planogram (PSFP), based on a combination of default values, user selections, and algorithms. SSP supports one or more algorithms that add shelves to PSFPs only. SSP does not perform fixture optimization.

Space Optimized Assortment

This is the product list after products have been dropped because of limited space. This varies by optimization location.

Stacking/Nesting, Nesting Height

Multiple units of a product can be stacked on top of one another in the *y* direction on a specific facing on a shelf. When the units are stacked, the total height equals the number of

units x the height of a single unit. Nesting is defined as stacked units that fit inside each other so that the total height is less than the number of units x the height of a single unit. The nesting height is the incremental space that the nested units require. The total height of a facing with nested units is determined as follows: total height equals the height of a single unit + the number of nested units x the nesting height. This applies only to shelves.

Top Shelf Stacking Height Limit

This defines the limit on how high products can be stacked or nested on the top shelf in the *y* direction. This is not required for intermediate shelves in a fixture because of the physical limit created by the shelf above. This applies only to shelves.

Units of Measure and Currency

This document will use "ft/cm" when talking about units of space, and will use "value" when talking about currency

Virtual Planogram (VPOG)

This refers to a POG that was created by ASO.