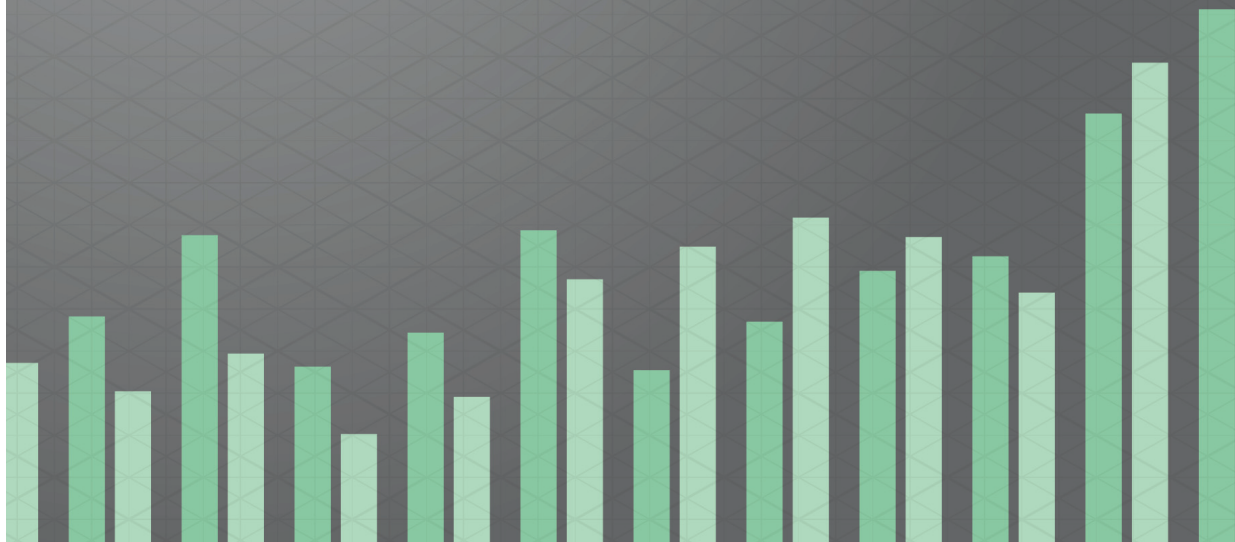




ACTUATE.
The BIRT Company™



BIRT Analytics



Using BIRT Analytics

Information in this document is subject to change without notice. Examples provided are fictitious. No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, for any purpose, in whole or in part, without the express written permission of Actuate Corporation.

© 2003 - 2014 by Actuate Corporation. All rights reserved. Printed in the United States of America.

Contains information proprietary to:
Actuate Corporation, 951 Mariners Island Boulevard, San Mateo, CA 94404

www.actuate.com

The software described in this manual is provided by Actuate Corporation under an Actuate License agreement. The software may be used only in accordance with the terms of the agreement. Actuate software products are protected by U.S. and International patents and patents pending. For a current list of patents, please see <http://www.actuate.com/patents>.

Actuate Corporation trademarks and registered trademarks include:

Actuate, ActuateOne, the Actuate logo, Archived Data Analytics, BIRT, BIRT 360, BIRT Analytics, BIRT Data Analyzer, BIRT Performance Analytics, Collaborative Reporting Architecture, Dynamic Data Web, e.Analysis, e.Report, e.Reporting, e.Spreadsheet, Encyclopedia, Interactive Viewing, OnPerformance, Performancesoft, Performancesoft Track, Performancesoft Views, Quite4Me, Quiterian, Report Encyclopedia, Reportlet, The people behind BIRT, X2BIRT, and XML reports.

Actuate products may contain third-party products or technologies. Third-party trademarks or registered trademarks of their respective owners, companies, or organizations include:

Mark Adler and Jean-loup Gailly (www.zlib.net): zlib. Apache Software Foundation (www.apache.org): Axis2, log4, Tomcat. Boost.org: Boost libraries, licensed under the Boost Software License. CURL (curl.haxx.se): Curl, licensed under a MIT/X derivate license. International Components for Unicode (ICU): ICU library. Marcin Kalicinski (rapidxml.sourceforge.net): RapidXML, licensed under the Boost Software License. Bruno Lowagie and Paulo Soares: iTextSharp, licensed under the Mozilla Public License (MPL). Math.NET: Math.NET, licensed under the MIT/X11 License. Microsoft Corporation: Access Database Engine, SQL Server Express. opensv team (sourceforg.net): opensv. openssl.org: OpenSSL, licensed under the OpenSSL license. gooxdoo.org: gooxdoo, licensed under the Eclipse Public License (EPL). Dave Scriven (svg.codeplex.com): SVG Rendering Engine, licensed under the Microsoft Public License. SQLAPI: SQLAPI++. sqlite.org: SQLite, public domain. stlsoft.org: STLSoft libraries, licensed under the BSD license. Matthew Wilson and Garth Lancaster (www.pantheios.org): Pantheios, licensed under a modified BSD license.

All other brand or product names are trademarks or registered trademarks of their respective owners, companies, or organizations.

Document No. 131215-2-580331 February 19, 2014

Contents

About <i>Using BIRT Analytics</i>	v
Chapter 1	
Understanding BIRT Analytics	1
About BIRT Analytics main interface	2
Understanding the home page	2
Laying out the feature tabs	3
Setting preferences	3
Changing a password	3
Changing the language	3
Changing the theme	4
Understanding the sample data model	4
Chapter 2	
Understanding BIRT Analytics work areas	5
About BIRT Analytics work areas	6
Understanding Data Tree	6
Using My Data	6
Making a calculated field permanent	7
Using Discrete Values	7
Using My Folders	8
Understanding Scratchpad	9
Understanding Data Explorer	11
About Record View	11
About Summary	11
About Discrete Values	12
About Chart	12
About Statistics	12
About Frequency	13
Exploring views of a database	13
Filtering views of a database	14
Understanding table resolution	15
Viewing results of simple queries	15
Changing table resolution	15
Chapter 3	
Working with your data	17
Using BIRT Analytics basic tools	18
Understanding the basic tools	18
Calculate	18
Export	18
Clear	18
Convert	18
Save or Save As	18
Applying a filter	18
About advanced filters	19
More about filters and resolution changes	19
Creating a parametric filter	20
Defining a selection	20
Adding a block	22

Returning all rows from a table	22
Changing resolution	22
Returning all rows and changing resolution level	23
Inverting a selection	24
Selecting discrete values	24
Sorting values	24
Specifying a sample in a selection	24
Creating an inner selection	24
Using import and export tools	24
Using the import tool	25
Understanding links	25
Using the export tools	25
Using downloads	25
Using BIRT Analytics engineering tools	26
Aggregating values	26
Decoding a field name	28
Working with expressions	30
Creating numeric ranges	31
Working with quantile ranges	33
Understanding parametric columns	33
Understanding ranking	35
Chapter 4	
Analyzing your data	37
About analyzing your data	38
Using crosstabs	38
Creating a crosstab	39
Understanding dimensions	40
Understanding measures	41
About crosstab filters	42
Selecting crosstab options	42
Other crosstab functions	42
Using Venn diagrams	43
Using bubble analyses	45
Using evolution	51
More about viewing an evolution	54
Recommendations	55
Using profile analyses	55
Using map analyses	58
Using Pareto analyses	61
Chapter 5	
Visualizing your data	65
Visualizing your data	66
Using the Gallery	66
Using a dial	66
Converting data measures to another indicator type	72
Using a meter	74
Using a label	78
Using a sphere	80
Using a cylinder	85
Using a funnel	87
Working with a Canvas	90

Chapter 6	
Identifying and predicting data trends	93
Understanding data mining and predictive analytics	94
Preparing data for mining	95
Standardizing data in a column	96
Understanding normalization	96
Understanding linear scaling	97
Understanding logistic scaling	98
Understanding Softmax scaling	98
Remapping a column	100
Understanding Clustering	101
Understanding Forecasting	102
More about outliers	102
Understanding decision trees	104
Training and testing a predictive model	104
Understanding the confusion matrix	104
Understanding sensitivity and specificity	105
Understanding association rules	107
Chapter 7	
Managing campaigns	113
Understanding campaigns	114
Configuring campaign elements	114
Creating a campaign workflow	114
Creating a stage	115
About assigning permissions	116
Defining a campaign resolution level	117
Defining a media condition	118
Defining an action goal	119
Planning a campaign	119
About campaign properties	120
Creating a strategy	120
Creating a campaign	120
About campaign cells	125
Running a campaign	128
Starting a campaign	128
Managing campaign stages	129
Viewing campaign summaries	129
Executing a campaign	130
Chapter 8	
Scheduling tasks	131
Automating a task	132
About event types	132
About action types	132
Creating a scheduled task	134
Managing scheduled tasks	138
Duplicating a scheduled task	138
Modifying a scheduled task	139
Using a conditional query to automate actions	139
Glossary	143
Index	157

About Using BIRT Analytics

BIRT Analytics is a web application, data repository, and set of web services that support analysis and forecasting of big data. In today's business climate, business users must be able to address the problems presented by big data, drawn from both structured and unstructured sources. BIRT Analytics supports selecting, grouping, analyzing, and presenting big data in a way that makes it actionable. BIRT Analytics enables a business user to process massive amounts of data, predict business outcomes, and make informed decisions. By making better decisions faster, business strategists can deliver vibrant and informative visual analysis of inherent trends in big data.

BIRT Analytics consists of three key components:

- The main user interface described in this manual, *Using BIRT Analytics*.
- BIRT Analytics Loader, a tool described in *Using BIRT Analytics Loader*. BIRT Analytics Loader extracts, transforms, and loads records from multiple external data sources to FastDB, the BIRT Analytics data repository.
- Admin, a toolset described in *Administering BIRT Analytics*. Admin technology supports administering access and privileges for each module and component of BIRT Analytics.

Using BIRT Analytics describes how to use Actuate BIRT Analytics technology to carry out dynamic analyses. *Using BIRT Analytics* includes the following chapters:

- *About Using BIRT Analytics*. This chapter provides an overview of this guide.
- *Chapter 1. Understanding BIRT Analytics*. This chapter introduces Actuate BIRT Analytics and provides information about the application's home page.
- *Chapter 2. Understanding BIRT Analytics work areas*. This chapter describes the BIRT Analytics work areas: Data Explorer, Data Tree, and Scratchpad.
- *Chapter 3. Working with your data*. This chapter describes how to select your data for analysis using BIRT Analytics fundamental tools.
- *Chapter 4. Analyzing your data*. This chapter describes how to analyze data.
- *Chapter 5. Visualizing your data*. This chapter describes how to create appealing data analysis visualizations.
- *Chapter 6. Identifying and predicting data trends*. This chapter describes how to use BIRT Analytics to mine data.
- *Chapter 7. Managing campaigns*. This chapter describes how set up and run a business campaign using BIRT Analytics.
- *Chapter 8. Scheduling tasks*. This chapter describes how to automate tasks and events using BIRT Analytics.
- *Glossary*. This chapter provides definitions of terms used in the BIRT Analytics product and documentation.

1

Understanding BIRT Analytics

This chapter contains:

- About BIRT Analytics main interface
- Understanding the home page
- Understanding the sample data model

About BIRT Analytics main interface












The BIRT Analytics main interface provides the following tools:

- Data exploration
- Purely analytical tools: crosstabs, Venn diagrams, bubble diagrams, evolutions, profiles, maps, and Pareto analyses
- Advanced analytical tools for data mining: clustering and forecasting
- Segmentation tool: selections
- Data engineering and enrichment tools: aggregates, decodes, expressions, numeric ranges, quantile ranges, parameters, and rankings
- Data export and import tools
- Configuration tools for user preferences: password, language, and theme

Understanding the home page

The home page appears in the browser when you open BIRT Analytics. This page provides access to all of the features of the product. To access the features, use the icons shown in Table 1-1.

Table 1-1 Icons for BIRT Analytics features

Icon	Label	Purpose of the BIRT Analytics feature
	Start	The initial view
	Explore	Viewing and interacting with data in a segment or table
	Enrichment	Creating more informative data by defining new fields in existing tables or using the data outside BIRT Analytics
	Enrichment— Engineering	Enriching analyses by creating new fields in existing tables
	Enrichment— Import-Export	Inputting and outputting data using various formats
	Analytics	Performing in-depth examinations of data and drawing conclusions from the patterns
	Analytics— Analysis	Examining data in depth using statistical tools
	Analytics— Advanced	Mining data to produce information from operations such as grouping and prediction
	Analytics— Selections	Segmenting the data by identifying groups of items that meet certain criteria
	Gallery	Displaying data in visual form, providing a quick way to see trends and patterns
	Workflow	Scheduling tasks and events and managing campaign workflow.

Laying out the feature tabs

A tools menu provides customization options for each set of tabs in the user interface, as shown in Figure 1-1. These options support placing the tabs in different positions and displaying text, an icon, or both on each tab. To access these options, right-click a tool's tab.

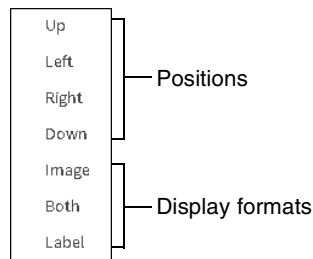


Figure 1-1 Options for tool tab placement and text

Setting preferences

Use Preferences to change the password, language, and theme for BIRT Analytics, as shown in Figure 1-2. To save one change and prepare to make further changes, choose Change. To discard unsaved changes, choose Cancel. To save changes and return to BIRT Analytics, choose Accept. Changes take effect after exiting BIRT Analytics and reentering.

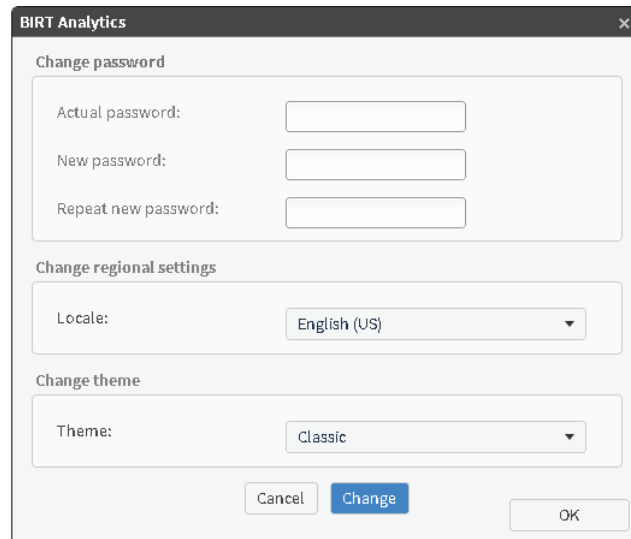


Figure 1-2 Available user preferences

Changing a password

The new password must comply with the password policy associated with the user account and must contain at least one of each of the following character types:

- Lowercase letters
- Uppercase letters
- The following special characters `_`, `-`, `!`, `.`, `$`, `%`, `(`, `)`, `=`, `|`, `@`, `#`, `€`, `*`, `~`, `"`, `'`, `.`
- Numeric characters

Changing the language

You can change the language used in the user interface to an available language.

Changing the theme

You can change the appearance of BIRT Analytics using predefined themes. BIRT Analytics provides a single initial theme.

Understanding the sample data model

Figure 1-3 shows an outline of the demo database analytical model. Not all fields from each table appear in this example.

- Household is the parent table.
- Customer is child table 1.
- Order is child table 2.
- Order Detail is child table 3.

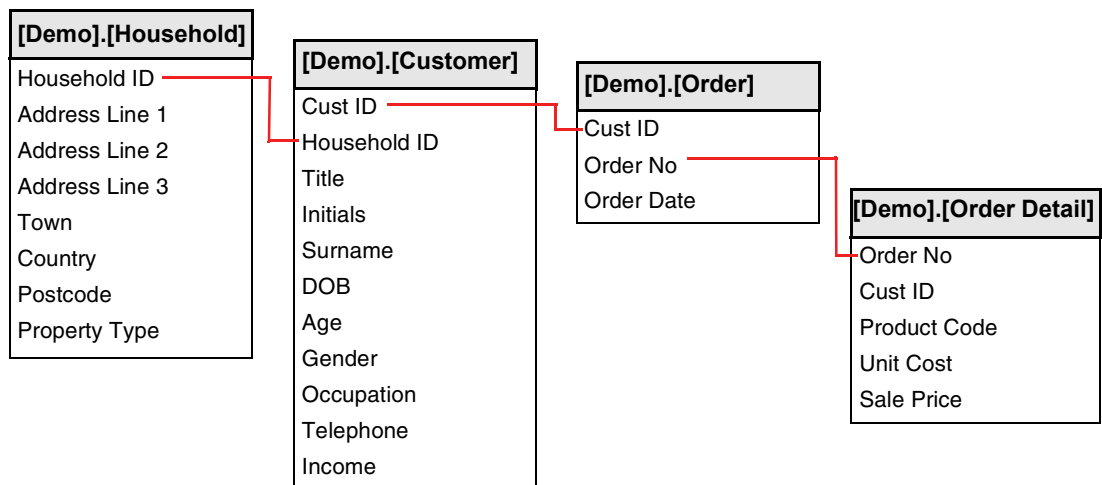


Figure 1-3 Demo database tables, fields, and associations

When the administrator loads the demo database into BIRT Analytics, the loader sets up the associations among the tables. Use these implied associations in BIRT Analytics to change the resolution on many types of analyses:

- A household can have from 1 to N associated customers.
- A customer can have from 1 to N associated orders. A customer must have an associated household.
- An order can have from 1 to N associated order detail lines. An order must have an associated customer.
- An order detail line must have an associated order.

2

Understanding BIRT Analytics work areas

This chapter contains:

- About BIRT Analytics work areas
- Understanding Data Tree
- Understanding Scratchpad
- Understanding Data Explorer
- Understanding table resolution

About BIRT Analytics work areas

The BIRT Analytics user interface consists of three main areas:

- Data Tree
- Scratchpad
- Data Explorer

Understanding Data Tree

Data Tree provides the following three views to navigate through your data and saved analysis files: My Data, Discrete Values, and My Folders, as shown in Figure 2-1.

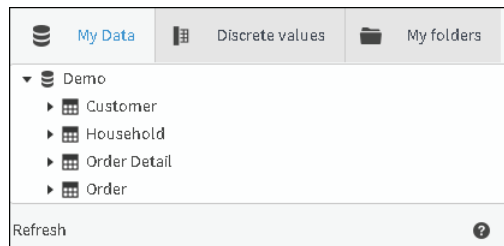


Figure 2-1 Data Tree showing the tables in the Demo database

My Data works with the Discrete Values viewer, which shows the content of explored fields, their values, and the number of stored entries with each value. Values are sometimes referred to as categories and entries as records.

Using My Data

My Data is available from Data Tree. Use My Data to display and navigate through databases, tables, and fields. To view the tables in a database, select the triangle icon beside the database name. To view the fields in a table, select the triangle next to the table, as shown in Figure 2-2. In this figure, the Demo database and the Customer table in the database are expanded and the fields in the Customer table are visible. Selecting the triangle next to an expanded database or table collapses the view of the items.

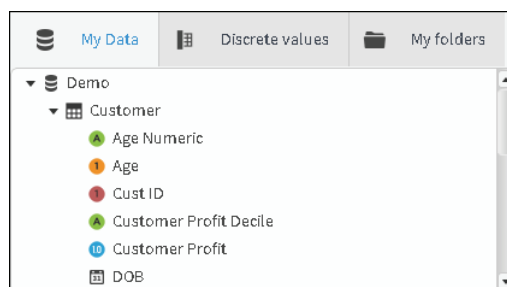












Figure 2-2 My Data showing fields in the Demo Customer table

Data items appear with a different identifying icon for each item type, as shown in Table 2-1. Unindexed fields appear without color but are otherwise the same as the associated physical field type.

Table 2-1 Icons for database, table, and field data types

Icon	Data structure or type
	Database
	Table
	Full numeric field
	Real numeric field
	Text field
	Date field
	Time field
	Date/time field
	Calculated field
	Unindexed field

In My Data, right-click an item to enable the following field operation options:

- Delete: Permanently remove the field or table. Delete is available only for users who have the Delete privilege granted by an administrator.
- Edit: Change a calculated field.
- Selection: Select a range of values stored in a column.
- Make permanent: Write a calculated field as an entry in the database.

Making a calculated field permanent

A calculated field can be stored as a permanent field in the associated table for use by others. Identify a calculated field by the small, gray cog in the lower left corner of the calculated field icon. Unlike a calculated field, a permanent field can be used to sort columns in a report. Making a calculated field permanent replaces the calculated field. The database does not retain the original calculated field definition.

How to make a calculated field permanent

- 1 Right-click a calculated field appearing in My Data.
- 2 Choose Make permanent.

Using Discrete Values

Discrete Values is a view that shows the contents of individual database fields. Each unique value, or category, is represented by a total of all its occurrences or records.

Discrete Values appears when you double-click a database field in My Data. For example, double clicking the Age Numeric field in the Customers table displays a list of age ranges and the number of customer records matching each age range, as shown in Figure 2-3.

Value	Count
25 under	5,533
25-30	12,876
30-35	26,259
35-40	40,857
40-45	50,685
45-55	86,312
55-65	32,351
65 plus	5,001

Figure 2-3 Discrete values showing the values in the Age Numeric field

Discrete Values also supports dragging fields and dropping them in places where selections or segments are used. For example, drag one or more specific categories to Scratchpad to examine the values in detail or store them for future use.

When a field has many categories, the viewer displays them on multiple pages. Typically, the list of discrete values is paginated, because most fields have a large number of categories. By default, the viewer shows 100 categories per page. For example, [Household].[Town] is a field that contains many discrete values. [Order].[Order No] contains more discrete values as there is a unique category for every order.

Categories can be sorted by name and the number of records they contain. The default sort order is by number of records. To sort categories in alphabetical order, click Value. A search engine supports finding particular categories by Name. Select multiple discrete values by holding CTRL while clicking the desired categories.

Using My Folders

My Folders displays the files and folders available to a user, as shown in Figure 2-4.

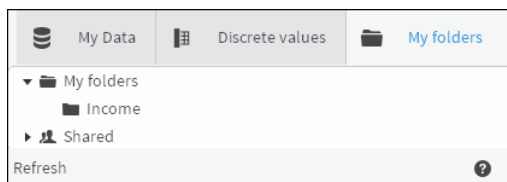


Figure 2-4 My Folders showing a personal folder, Income

Files and folders in the My Folders section in My Folders are available only to you. Files in the Shared section are visible to all users. Files can contain analyses, selections, and exports. There is no fixed limit to the number of folders you can create. To share a file with other users, move the file under the Shared heading.

Create new folders directly or while saving a file:

- Right-click the red folder labeled My folders. Choose New. Type the name and description of the new folder, as shown in Figure 2-5.

New...

Name:

Description:

Figure 2-5 Creating a new folder named Income

- When saving an analysis, choose Save As to create a folder in which to save the analysis.

To save a complete data structure into a file with a .ddw extension, right-click a file and choose Export. The file can be downloaded to your local system. To load previously stored structures, right-click a folder and choose Import.

Understanding Scratchpad

Scratchpad is a work area located below Data Tree, as shown in Figure 2-6. Segments can be placed in Scratchpad for use in an analysis or for examining them in detail. Items in Scratchpad are not saved between sessions.

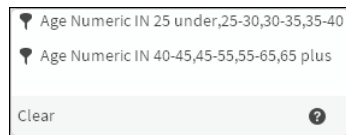


Figure 2-6 Scratchpad showing two segments

Use Scratchpad in the following ways:



- Drag segments directly from Data Tree to use as filters in analyses.
- Drag results of different analyses to Scratchpad for detailed analysis.
- Manipulate segments from Data Tree or from an analysis to create intersections, invert content, and so on.

All changes applied to a segment in Scratchpad affect only the current session. To retain these segments for later use, save them in My Folders, as described in “Using My Folders.”

Scratchpad consists of four elements:




- Segments window. Drag segments from other windows in BIRT Analytics and drop them in this main area of Scratchpad. Available segments are tables from a database, discrete values, and the results of some analyses.
- Status bar. This information section appears below the Scratchpad segments window and displays the number of records in the selected segment.
- Segment handling menu. When you drag one segment over another in Scratchpad, the join functions in Table 2-2 appear in a menu.

Table 2-2 Functions used to join segment contents

Icon	Function	Description
	AND	Joins the two categories with the resulting segment containing those records that meet both conditions simultaneously. For example, you could select Customers who are Sales Assistants and under the age of 25.
	OR	Joins the two categories with the resulting segment containing those records that meet one or both of the conditions. For example, you could select customers who are either Sales Assistants or under the age of 25, which would include Sales Assistants under the age of 25.

(continues)

Table 2-2 Functions used to join segment contents (continued)

Icon	Function	Description
	Exclude	Joins the two categories, with the resulting segment containing those records that meet the conditions in the first category but not the second one. For example, you could select customers who are Sales Assistants but not under the age of 25, which would exclude Sales Assistants under the age of 25.
	NAND	Inversion of AND.
	NOR	Inversion of OR.

- Secondary menu. When you right-click a segment, the options in Table 2-3 are available.

Table 2-3 Segment options

Option	Description
Rename	Prompts for a new name for the segment. Use this option if you have changed the results of the segment by joining it with another one and the default name no longer describes the segment accurately.
Delete	Removes the segment from the scratchpad.
Invert	Change the segment to return results that exclude the selected segment. For example, to select all records from households that are not in London, select London and invert the selection. [Household].[Town] contains 1335 different towns. Selecting all towns, except London, one by one would be very time-consuming. Invert takes the category of London and returns all of the records that do not belong to this category.
Make permanent	Save the segment as a new field. The field appears in the table currently being used for resolution. Save a segment to make it available in a subsequent BIRT Analytics session. After refreshing the screen view, the new field appears in the table in which it belongs.
Change resolution level	You can resolve results at different levels within the database. These levels correspond to the tables that make up the segment. For example, resolve a category from the Customer table at the Household level to display the households of the customers. To interpret the results of changing the resolution, keep in mind the direction of the resolution change, either many (N) to one or one to N.
First discrete	Choose a record for each of the selection's attributes based on a field in the database.
View definition	Examine the definition of the segment. The definition contains all of the operations applied to the segment, indicating the type of operation, the segment's total records, the number of records after performing the operation, and the query that obtains this segment. After you make a change in resolution, the operations performed are displayed in groups.

Table 2-3 Segment options

Option	Description
Sort	Uses a field from the database to sort the segment. The result of the sort appears in Scratchpad and not in the data explorer.
Select sample	Returns a data sample with the size determined by the selection criteria: Top, Bottom, Middle, 1 in N, Random.
Refresh	Updates the count of records in the segment.

Understanding Data Explorer

Data Explorer is a tool that displays detailed and summary information about databases, tables, and fields. To access Data Explorer, choose Explore. Then, drag a segment from Data Tree or Scratchpad, and drop it in Record View.

Depending on the selection you make, different views are available, as shown in Table 2-4.

Table 2-4 Data Explorer views for data selection types

Selection type	Available Data Explorer view
Database	Summary
Table	Record, Summary
Unindexed field	Record, Summary
Indexed textual field	Record, Summary, Discrete Values, Chart
Indexed numeric field	Record, Summary, Discrete Values, Chart, Statistics, Frequency

Data Explorer supports exporting data to other applications, such as Microsoft Excel.

To remove data from Data Explorer, choose Clear.

About Record View

Record View displays the records in the table you are exploring or the table that contains the field you are exploring. For example, when you explore the Customers table you see 50 of the 259,874 records. You can browse the pages into which the records are divided by using the arrows on the pagination bar. You can navigate to a specific page by typing the page number and pressing Enter. You can select the fields you want to view using the column selector icon in the upper-right corner.

About Summary

Summary displays information about the database, table, or field you are exploring.

For a database, Summary displays the name of the database; the tables in the database; and the number of rows, columns, and cells in each table.

For a table, Summary displays the name of the table; the number of rows, columns, and cells it contains; the other database tables to which it is joined; and the information about each field in the table, such as data type, number of discrete values, and whether or not the column is indexed.

For a field, Summary displays the name of the field; the table that contains it; and the other information, such as data type, number of discrete values, and whether or not the column is indexed.



To export the contents of the Summary tab to an RTF file, choose the export icon in the upper-right corner. RTF format is compatible with Microsoft Word.

About Discrete Values

Discrete Values shows the categories contained in a column, the number of occurrences or records for each category, and the percentage for each category. To export the contents of Discrete Values to a CSV file, choose Export. CSV format is compatible with Microsoft Excel and text editors, such as Ultraedit and Notepad++.

About Chart

Chart displays a graphical representation of the records of categories in a database column. The values are ordered from most frequently occurring to least frequently occurring. The following chart types are supported:

- Columns 3D (default)
- Areas
- Bars
- Columns
- Doughnut
- Doughnut 3D
- Lines
- Pie
- Pie 3D

To change the chart type, use the chart icon in the upper right corner. To export a chart to a JPG file, choose the export icon.

About Statistics

Statistics displays the following information in a tabular format for numeric fields.

Table 2-5 Information provided for a numeric field

Column name	Value displayed
Count	Total number of non-discrete values in the field.
Kurtosis	A measure of the distribution of values, or peakedness. The value varies between -0.5 and 0.5. The Kurtosis coefficient indicates how sharp a distribution is, relative to a standard normal distribution.
Maximum	Largest value in the field.
Mean	Arithmetic average of all non-discrete values in the field.
Median	Value for which there are an equal number of larger values and smaller values in the field.
Minimum	Lowest value in the field.
Mode	The most frequently occurring value in the field.
Skewness	A measure of symmetry that varies between -0.5 and 0.5. If skewness is negative, the distribution is skewed to the left. If skewness is positive, the distribution is skewed to the right.

Table 2-5 Information provided for a numeric field

Column name	Value displayed
Standard deviation	A measure of how much variation or dispersion there is from the mean. A low standard deviation indicates that the data points tend to be very close to the mean. A high standard deviation indicates that the data points are spread over a large range of values.
Sum	The sum of all the values in the field.
Sum of squares	The sum of all the squared values.



To export the contents of the Statistics tab to an RTF file, choose the export icon in the upper-right corner. RTF format is compatible with Microsoft Word.

About Frequency

Frequency displays a graphical representation of the number of occurrences of values in a database column. The values are in ascending order. The following chart types are supported:

- Areas (default)
- Bars
- Columns
- Columns 3D
- Lines

To change the chart type, use the chart icon in the upper-right corner. To export a chart to a JPG file, choose the export icon.

Exploring views of a database

Data Explorer supports viewing different levels of a database, depending on the level of data you choose to view. For example, you can choose a database to view summary information for that database. You can choose a column in a database table to view records, discrete values, a chart of values, or summary information about the column.

How to view a database summary

- 1 Drag a database from My Data and drop it in Data Explorer.
- 2 Examine the information appearing on Summary, as shown in Figure 2-7.

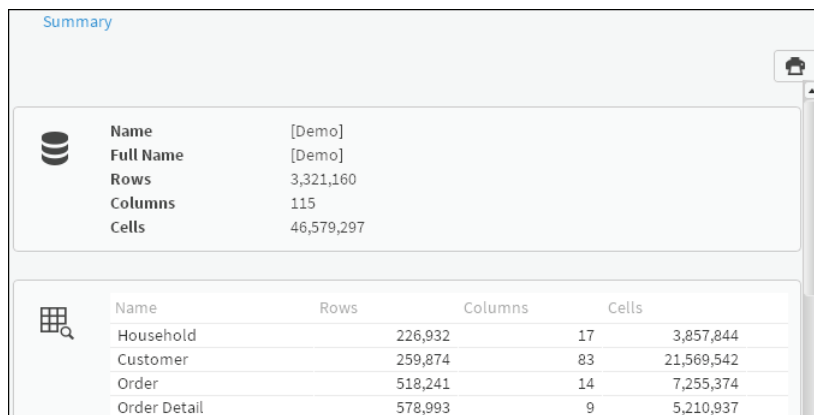
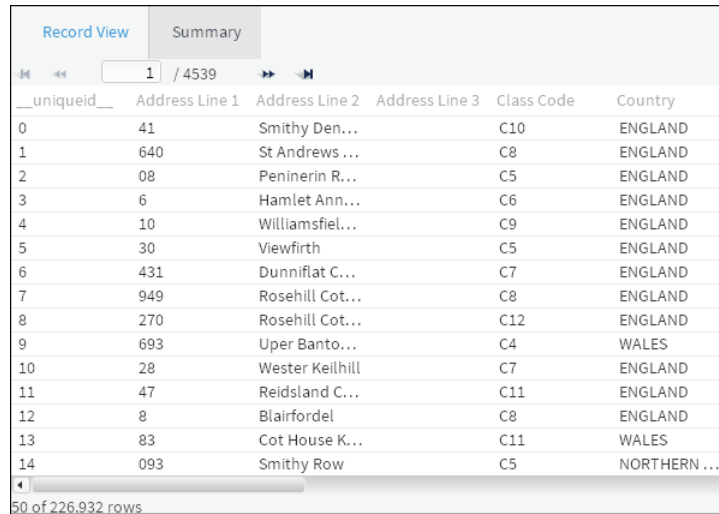


Figure 2-7 Viewing a database summary in Data Explorer

How to view a database table

- 1 From a database appearing in My Data, drag a table and drop it in Data Explorer.
- 2 Examine the information on Record View and Summary, as shown in Figure 2-8.



__uniqueid__	Address Line 1	Address Line 2	Address Line 3	Class Code	Country
0	41	Smithy Den...		C10	ENGLAND
1	640	St Andrews ...		C8	ENGLAND
2	08	Peninerin R...		C5	ENGLAND
3	6	Hamlet Ann...		C6	ENGLAND
4	10	Williamsfiel...		C9	ENGLAND
5	30	Viewfirth		C5	ENGLAND
6	431	Dunniflat C...		C7	ENGLAND
7	949	Rosehill Cot...		C8	ENGLAND
8	270	Rosehill Cot...		C12	ENGLAND
9	693	Uper Banto...		C4	WALES
10	28	Wester Keilhill		C7	ENGLAND
11	47	Reidsland C...		C11	ENGLAND
12	8	Blairfordel		C8	ENGLAND
13	83	Cot House K...		C11	WALES
14	093	Smithy Row		C5	NORTHERN ...

Figure 2-8 Viewing records from a table in Data Explorer

How to view a database column

- 1 From a database table appearing in My Data, drag a column and drop it in Data Explorer.
- 2 Examine the information in Record View, Summary, Discrete Values, and Chart, as shown in Figure 2-9.

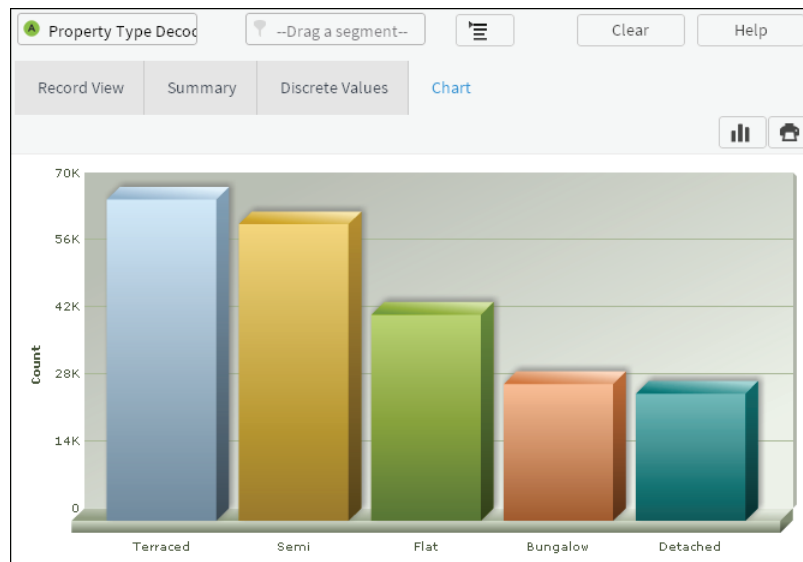


Figure 2-9 Viewing a database column using a chart in Data Explorer

Filtering views of a database

You can filter the view of a table or column appearing in Data Explorer using segments. Adding a segment to the view of a table or column limits the number of records shown to only those records having a value that matches the segment.

How to filter a database view using a segment

- 1 Drag a column from My Data and drop it in Data Explorer.
- 2 Select Discrete Values.
- 3 Drag a row in Discrete Values and drop it in Scratchpad. For example, from the Age Numeric column, drag Age Numeric EQ 25 under. A segment appears in Scratchpad.
- 4 In Data Explorer, select Record View.
- 5 Drag one or multiple segments from Scratchpad and drop the segments in Drag a segment. A view of data, filtered by the selected segment, appears in Data Explorer.

Understanding table resolution

BIRT Analytics supports saving your selections, queries and analyses for reuse. To retrieve records from a different part of your database, replace the table from which you retrieved records in a previous selection, query, or analysis, then recalculate the results. To demonstrate how the concept of table resolution works in practice, this section presents examples that demonstrate viewing different tables in the BIRT Analytics demo database using Data Explorer.

Viewing results of simple queries

Examining results returned by a simple query from one database table shows the discrete values in that table. For example, using My Data, expand the Household table, then double-click Property Types Decode. In Discrete Values, you see 28,514 records for households having the type Bungalow.

For a similar example, expand Customers, then double-click Gender. In Discrete Values, you see 102,042 F and 157,832 M records, which represent 102,042 female and 157,832 male customers in the database. To resolve questions about customers in each household, you can change the resolution or perspective of your query.

Changing table resolution

To demonstrate changing table resolution, modify the second example from the preceding section in the following way. Drag the Female value from Discrete Values and drop it in Record View of Data Explorer. You see complete records for 102,042 female customers. Choose My Data, then replace Customer with Household. To do this task, drag Household from My Data and drop it on Customer in Data Explorer. 90,765 records that represent households having one or more female customers appear in Record View, as shown in Figure 2-10.

__uniqueid__	Address Line 1	Address Line 2	Address Line 3	Class Code	Country	County
1	640	St Andrews ...		C8	ENGLAND	
2	08	Peninerin R...		C5	ENGLAND	HAMPSH
3	6	Hamlet Ann...		C6	ENGLAND	DERBYSH
7	949	Rosehill Cot...		C8	ENGLAND	
8	270	Rosehill Cot...		C12	ENGLAND	WEST YO
12	8	Blairfordel		C8	ENGLAND	
15	70	Town House		C7	SCOTLAND	WEST LO
16	979	West Cairnb...		C1	ENGLAND	WILTSHI
17	50	Nethergate ...		C7	WALES	CLWYD
19	896	Easter Muck...		C2	ENGLAND	
22	68	Henhouse		C3	ENGLAND	
24	861	Candy Mill		C9	ENGLAND	HAMPSH
25	7	Bal Road Or ...		C11	WALES	MID GLA
26	914	Fowler Terr...		C12	ENGLAND	
27	4	Moncur St		C10	ENGLAND	KENT

Figure 2-10 Changing table resolution in Data Explorer

Changing the table on which a query resolves returns results different than those returned by a simple query. To return expected results, the tables you replace must relate or join on a common field.

Working with your data

This chapter contains:

- Using BIRT Analytics basic tools
- Using BIRT Analytics engineering tools

Using BIRT Analytics basic tools

Basic tools appear throughout BIRT Analytics to support common data analysis operations such as calculating and saving results, clearing work spaces, selecting data, and importing and exporting files. Engineering tools support creating new data fields that support your analysis of existing data values. This chapter explains the fundamental tools provided in BIRT Analytics.

Understanding the basic tools

A toolbar appears in the window for each analysis type. Some or all of the following basic tools are available in each window.

Calculate

Calculate runs and displays the analysis using the parameter values.

Export

There are two export modes.

- Exporting data
Press the Export button in the analysis toolbar. In this case the results table is exported from the crosstab, Venn, bubble or profile analysis. Values are exported to a comma-separated values (CSV) file, a standard format supported by tools such as Excel and text editors such as Ultraedit and Notepad ++.



- Full export
The export icon displayed in the graphic exports the full analysis (data and graphic) in RTF format, which is compatible with word processors such as MS Word.

Clear

Clear removes all entries from the window, without saving.

Convert

This option changes one analysis or indicator into another one with the same features.

Save or Save As

You can save all BIRT Analytics analyses you create using analysis tools or selections. The definition of the analysis is saved, but not the results themselves. Any modification to the database which affects its configuration is automatically applied when the saved analysis is run.



These analyses must always be saved in a folder. If a folder has not been created previously, it can be created at the time of saving the first analysis. You can access saved analyses from Data Tree using My Folders.

Any folder or subfolder is personal, unless you indicate otherwise and give viewing permission to other users. These permissions can be given for both folders and analyses.

If you run a saved analysis, make a change to its configuration, and want to keep both the initial and modified versions, use Save As to save a new version of the analysis.

Applying a filter

Filters are used throughout BIRT Analytics and are based on data segments.

You usually can drag a discrete value directly to a filter area. For example, in an analysis of recent orders, you could drag the Customer Gender discrete value “female” to the analysis filter to see only orders placed by female customers.

Some tools offer more advanced filters.

About advanced filters

Crosstab, bubble, and map analyses support the following three types of filters: universal, target, and baseline.

A universal filter is applied before any change in resolution occurs. A target filter is applied after a change in resolution occurs. For example, to view only records for female customers, add as a universal filter: Gender equals female. If you add Gender equals female as a target filter and change the resolution from Customer to Household, only records that include households with females appear. Some of those households may also include males.

Target and baseline filters are used together to create comparative analyses. Be sure to use segments that can be compared. For example, compare one year with another or one population group with another. When calculating a comparative analysis, you can choose to display a measure as:

- Result

The default for measures is to produce the count of records in both filtered segments. This is not directly useful for most comparisons but can be used as a total when creating calculated fields.

- Index

Shows the degree to which compared groups differ using an indicator. An index value greater than 0 means that the baseline is as many times greater than the value shown by the index with respect to the target. An index value less than 0 means the reverse is true. The formula for Index is:

$$(\text{Target/Total}) / (\text{Baseline/Total})$$

- Difference

Displays size differences between the baseline and the target as measured in units. A negative result means that the baseline has as many more values than the displayed number. A positive result indicates the opposite. The formula for Difference is:

$$\text{Target} - \text{Baseline}$$

More about filters and resolution changes

If you are using filters, note that specific situations require certain filter types. You must use a target filter for pivoted analyses when there is a change in resolution between the axes and the measures in the direction N-to-1. You must use a universal filter with a non-pivoted table when there is a change in resolution between the axes and the measures in the direction N-to-1. When no such size disparity exists between axes and measures, the type of filter used for each analysis does not matter.

Consider creating a crosstab using axes from one table and resolving the results in another. Use, as a filter, a segment from the source table for the axes.

For example:

- Universal filter. Apply the filter before carrying out the change in resolution. For example, the field low salary [axis 1: customer table, salary field] is used with the filter. Then, change the resolution to a different table, such as Households. The filter conditions are met by the same household and person. The resolution unit of the filter is the unit indicated by the axes of the crosstab, in this case Customers.
- Target filter. Apply the filter after carrying out the change in resolution. In other words, select the segments from the table to which the selected axes belong, and carry out the

change in resolution for a table chosen as the resolution level for the results. Then, apply the filter. For example, low salary [axis 1: customer table, salary field] and the resolution is changed to Household. In this example, you see all households with at least one customer whose salary is low. The filter is applied. For example, [customer tables, gender field = female] gives a result qualitatively higher than the result from the universal filter. All households meeting the condition of low salary and female appear. The condition is not necessarily met by the same person. The filter's resolution unit is Households, the resolution table for the crosstab.

- **Baseline filter.** Selecting a target filter activates a baseline filter. Use a baseline filter for the purpose of comparison. For example, compare two periods of time using the following two filters: 2008 target and 2007 baseline.

To configure the table, you must first select the axis or axes by dragging to the appropriate space, then dragging the measures. By default, when you drag the axes, the value count for the table to which they belong appears.

Creating a parametric filter

Parameters used in filters should be included if a table is calculated or in a situation where you introduce a new data table in order to calculate the final output. The filter is determined by prompting you for the value when the analysis is calculated. You can use either a pre-set filter or a prompted filter, but not both.


Defining a selection

A selection is a segment of data, a set of values chosen for a specific purpose. A database is made up of values, and these values internally form groups that have similar features, with these segments being homogenous. For example, an organization's customers can be gender=F or gender=M, but not both. You can combine these segments (gender=F, aged between 25 and 35, city=Barcelona, and average purchasing power, for example) and thus specify the target audience for a marketing campaign or promotional offer.

The BIRT Analytics Selections tool supports drag-and-drop configuration of elements in a selection. After configuring a selection, save it in Scratchpad or in a shared or private location for reuse. Calculate the selection to return all records meeting the criteria defined in the selection. As with full analyses, BIRT Analytics saves any defined parameters for each selection and runs them in real time.

For example, a simple selection may include only one query returning all rows from a single database table. To create a more complex selection, define multiple blocks. Each block may return records from a specific table or column, use a specific operator to compare values, or define ordering and grouping conditions. By combining elements in logical blocks, you refine the set of records the selection returns.

How to define a selection

- 1 Choose Analytics—Selections.
-  2 Choose Selections to open a new Selections window.
- 3 In the top-level block, type a name for the selection.
- 4 From the Selections toolbar, drag a query element and drop it in a block.
- 5 From My Data, drag a table, column, segment, or report, and drop it in an element.
- 6 Choose Calculate to execute a selection. Results indicating a number of records selected appear for each element.
- 7 Choose Save. Then, choose a private or shared location in which to store the selection.
- 8 Optionally, drag the selection name and drop it in Scratchpad.

For example, Figure 3-1 shows a simple selection returning all rows from the Customer table in the BIRT Analytics Demo database in which the customer is over 65 years old.

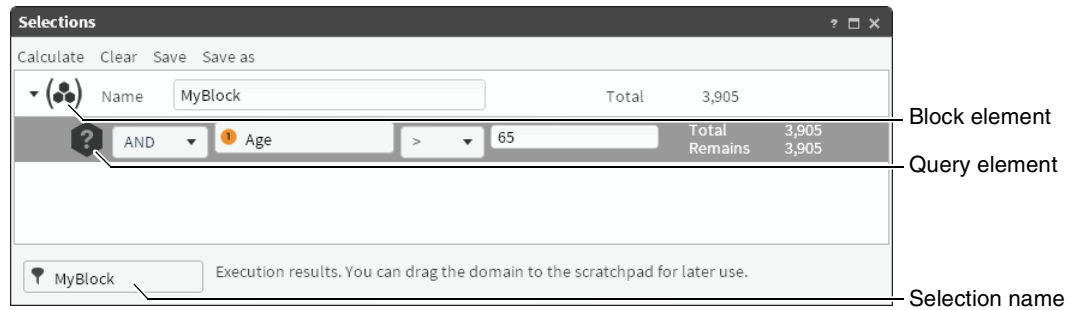


Figure 3-1 Examining basic elements in a simple selection



How to create a new query

- 1 Drag New query from Selections toolbar and drop it in a block.
- 2 Drag a field from the database, for example Gender, and drop it in the query element.
- 3 Select an operator, for example =, from the drop-down list.
- 4 Drag a value, for example Female, from Discrete values and drop it after the operator.
- 5 Choose Calculate.

Follow the same procedure for any other value you want to enter. Segments interact with each other by adding or excluding those records that do not match both values. For example, you have the value Female from the Gender field; if you drag the value 25 from the Age field, you can return:

- All women and everyone aged 25, regardless of gender
- All women aged 25

To accomplish this task, you select an operator from the drop-down list to the left of the dragged column. In the first example, the appropriate operator is Or. In the second example, it is And. By default, the And operator is used for values from different fields in the database. When the values are from the same field in the database, the default operator is Or. For example, if you try to combine the values Age=25 with Age=26, and the operator is And, the number of values that meet both of these conditions is zero.

You can also create a new query by dragging a segment from the data tree or the scratchpad to:

- A query in the selection. The new query is entered above.
- A block. The new query is entered at the end of the block.

To parameterize a query, right-click the query and choose Parameter.



How to create a new parametric query

A parametric query is a query based on a value determined at calculation time, which may be the default value of the parameter. You create a parametric query the same way you create a non-parametric query.

For example, create a selection to calculate the number of customers under 25 years of age with occupation janitor. You could have a query with age, occupation and another third query sentence. In this particular example, our third query sentence is parametric gender. You could include a default value for the parameter, such as Female. When you run the selection, you are prompted to type a parameter value or accept the default. The result is a selection parameterized by gender.

To modify the parameter prompt, right-click the query.



Adding a block

Use a block to group elements in a selection. For example, consider the following basic mathematical operation. If you want to determine the result of the sum of two values multiplied by the sum of two other values, group the sum operations as follows:

$$(34 + 89) \times (23 + 65)$$

Now consider the following example as it applies to selections:

(Woman+salesperson) or (Woman+director)

If you do not include parentheses in this selection, you cannot be certain that the operators you use are invoked in the intended order. Hence one block is (Woman + salesperson) and another block is (Woman + director). A new selection has one block, in which you name the selection, by default. Every block must include a query element.

How to add a block

- 1 Drag a block element from Selections toolbar and drop it in an existing block.
- 2 Drag and drop additional elements in the block, as necessary.



Returning all rows from a table

To return all rows from a database table, add a simple query, using the All element.

How to return all rows from a table

- 1 Drag All from Selections toolbar and drop it in a block.
- 2 Drag a database table from My Data and drop it in the All element.
- 3 Optionally, select an operator from the list.
- 4 Choose Calculate.



Changing resolution

Adding a change resolution element to a selection or block changes the perspective of the selection. For example, to select customers who are both women and directors, you create a selection including query elements that return records for all customers who are female and directors. To see the households where the people who meet these conditions live, add a change resolution element to the selection.

A change resolution element cannot hold the top-level position in a selection. A change resolution element holding the last position in a selection causes the selection to return all records from the new table linked to all previous conditions. For example, the selection defined in the example shown in Figure 3-2 returns 868 records for households in which customers who are women directors live.

Name	Total
Selection having one resolution change	868
Gender Decode = Female	Total: 102,042 Remains: 102,042
Occupation Decode = Director	Total: 8,034 Remains: 924
Household	Total: 0 Remains: 868

Total records

Change resolution element

Figure 3-2 Examining a selection with one resolution change

For a contrasting example, add a change resolution element for each query element. This approach changes the perspective of each query in the selection. The selection in the example

shown in Figure 3-3 returns all houses where customers who are female live, intersected with all houses where customers who are directors live; a total of 1009 records.

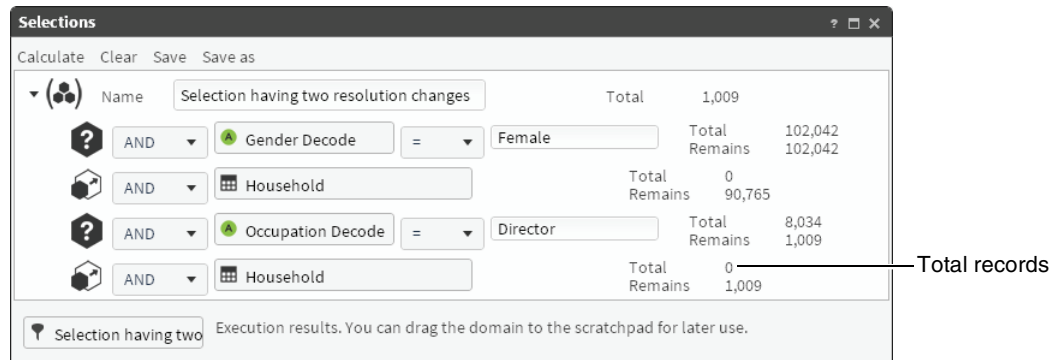


Figure 3-3 Examining a selection with two resolution changes

Returning all rows and changing resolution level

You may choose to create a selection using query elements that return records from different tables. Placing an All query element as the first, or the last element in a selection returns different numbers of records. Placing the All element first in a selection affects all single queries added after the All element by changing the resolution level to the table defined in the All element.

An example showing a selection created in this way appears in Figure 3-4. The selection returns all houses, intersected with houses where female customers live, intersected with houses where director customers live.

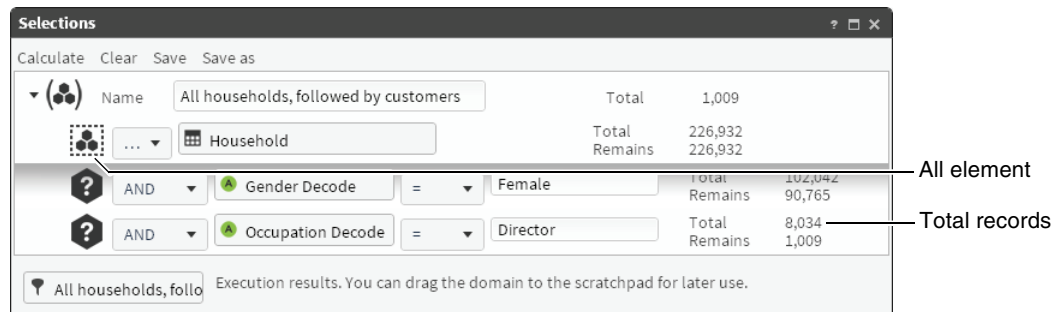


Figure 3-4 Selecting all records from a table first

Placing an All element last in a selection affects the selection by changing the resolution level to the table defined in the All element. In this case, the selection returns all female customers, intersected with director customers, intersected with all customers that have a house. In other words, this selection returns all female customers who are directors and have a house. In the BIRT Analytics Demo database, 868 records match this criteria. An example showing a selection created in this way appears in Figure 3-5.

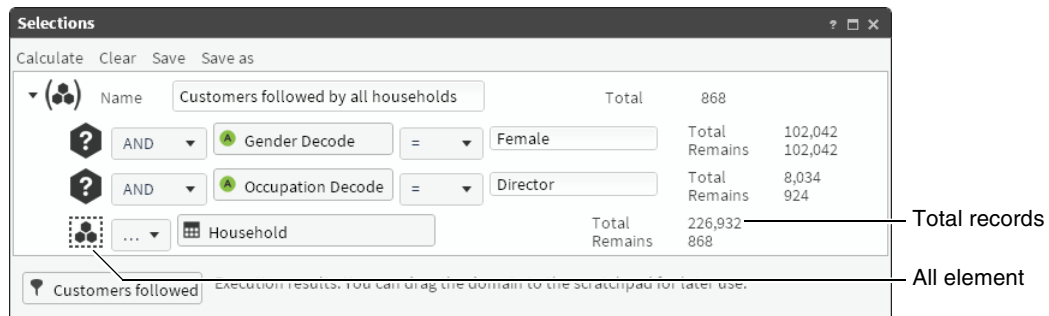


Figure 3-5 Selecting all records from a table last

Inverting a selection



To return the inverse set of values in a selection, drag Invert and drop it in a selection. For example, consider a selection that returns customer records containing Woman and Director; in other words, the set of women who are directors. Adding Invert to this selection causes this section to return all records in the database for customers who are not women directors.

Selecting discrete values



Add a Discrete Values element to a selection to return specific values. Drag Discrete Values from the Selections toolbar and drop it in a block. Then, drag a database field name from My Data and drop it in the element. Choose Calculate to return the field values. For example, a discrete values element using the Age field returns seventy two values from the Age field.

Sorting values



Add a Sort element to a selection to sort the values in the selection. Drag a column and drop it in the Sort element, then select Ascending or Descending order.

Specifying a sample in a selection



Add a Sample element to a selection to return a defined range of N records for a block or a selection. Sample requires that you select a range parameter and number of values that define the sample. For example, select Top and type 10 to sample the top ten records of a selection.

Creating an inner selection



Use an Inner selection element to add an existing filter to a selection. For example, consider how to create a selection of female directors who are 50 years old, using an existing selection of female directors.

How to create an inner selection

- 1 Create a new selection and provide a name for the block.
- 2 Drag Inner selection from the toolbar and drop it in the block element.
- 3 From a shared location in My Folders, drag the Female directors selection and drop it in the Drag a report box in the inner selection element.
- 4 To the current selection, add a query element, Age=50.
- 5 Choose Calculate to return the number of female directors who are 50 years old.

Using import and export tools

Use these tools to import and export data and to create and delete links.

Using the import tool

Use this tool to import data to a new table created in the database currently loaded in the Engine. For example, this tool is useful when you want to create a Master table. Microsoft Access files are only supported in 32-bit installations.

How to import a table

- 1 Select the source of the data to be imported. Options range from MS Access to text files.
- 2 For a text file, select file features and whether the first row contains the file header.
- 3 In the columns section, create the definition of the columns by selecting Get columns or Add columns. The definition of a column can be changed by double-clicking it. You should define as many columns as there are in the file.
- 4 Use the Column up and Column down buttons to change the position of a column.

Understanding links

Use links to delete existing relations between columns in different tables and to create new relations between columns. Creating links is important when working with the Engine. If the links have not been created or have been created incorrectly, it is not possible to work with various levels of resolution in the analysis. You cannot obtain reliable results when you mix columns from different tables in the database in a crosstab or a Venn diagram.

Using the export tools

BIRT Analytics provides two export tools. Export File exports a segment to a plain text file. Export Database exports a segment to a table in the database. Supported export file formats are: CSV (comma-separated values) file, a standard format supported by programs such as Excel and text editors including Ultraedit and Notepad++, and RTF file, a format compatible with word processors such as MS Word. Both tools provide Save and Save As options. When you save an export definition, both the segment and the export configuration are saved in your personal folders.

How to export to a text file

- 1 In Enrichment—Import-Export, choose Export►File.
- 2 Drag the segment you want to export from Scratchpad and drop it in Segment to export.
- 3 In Delimiter, select a column character separator: Tab, Pipe, or Flat.
- 4 In End of line, select an end-of-line character: CR, LF, CRLF, or LFCR.
- 5 To save the file in the application server, choose Deferred. You can download this file from the application server by choosing Downloads.
- 6 To add a field containing the Uniform Resource Name (unique identifier) to the generated file, choose URN.
- 7 In Available columns, to select the fields to export, drag each field from the left pane to the right pane. To change the order of the fields, use the up and down arrow buttons.

How to export to a database

The export-to-database tool exports a segment to a new table in the analytical database engine. You must select the database in which you want to create the new table, the name of the new table, and the fields to create in the destination table.

Using downloads

This tool shows a list of deferred export files. This list shows the date, type, name, and file size. To download or delete an export file from the application server, you must open it.

Using BIRT Analytics engineering tools

BIRT Analytics engineering tools support creating new fields that you can include in your data analysis. To better quantify data, you can create fields that summarize, rename, and define expressions for existing fields within the database. You also can create ranges, groups, parameters, and ranks based on your existing data values. This chapter describes how to use the following engineering tools: Aggregates, Decodes, Expressions, Numeric Ranges, Quantile, Parametric, and Ranking.

Aggregating values

Aggregates supports aggregating, or grouping, values in a data field. You can aggregate values from one field in any table. For example, create an aggregate using the Count function in the Order field of the Customer table to see how many orders each customer placed.

How to create an aggregate

- 1 In Enrichment—Engineering, choose Aggregates.
- 2 From My Data, drag the name of the table in which you will create an aggregate, and drop it in Create a new Aggregate based on, as shown in Figure 3-6.

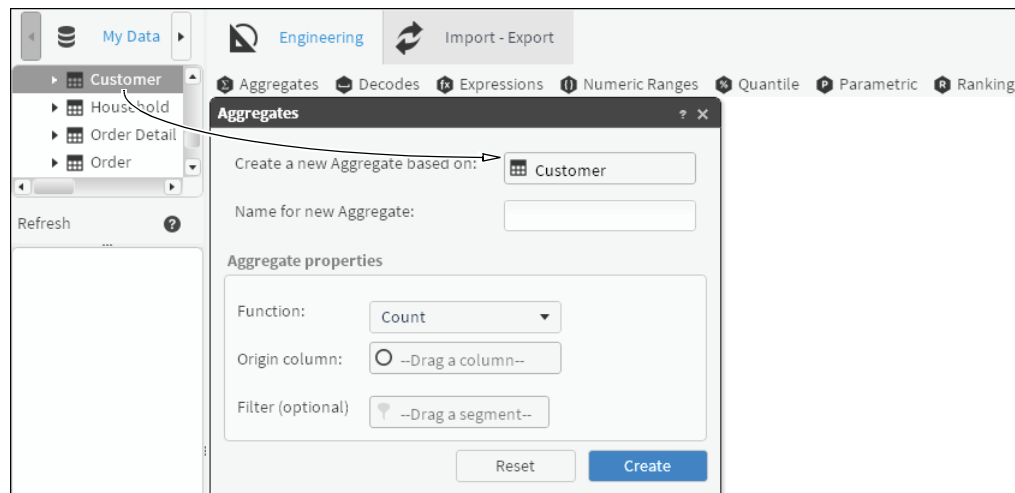


Figure 3-6 Assigning a table as a basis for an aggregate

- 3 In Name for new Aggregate, type the name of the new, aggregate field.
- 4 From My Data, drag the column that contains values to be aggregated, and drop it in Origin column.
- 5 In Function, select one of the following aggregate functions:
 - Average
 - Count
 - First
 - Last
 - Maximum value
 - Mean (integer)
 - Mean (real)
 - Minimum value

- Quadratic sum
- Standard deviation
- Sum

6 Choose Create.

7 In My Data, review the new aggregate to verify your selections.

For example, to aggregate orders placed by Customers in London, select Count in Function, drag Order No. from Customers and drop Order No. in the Origin column.

How to aggregate totals for a specific group

- 1 In Enrichment—Engineering, choose Aggregates.
- 2 From My Data, drag a table on which to base the new aggregate, and drop it in Aggregates, as shown in Figure 3-6.
- 3 In Name for new Aggregate, type the name of the new, aggregate field.
- 4 In Aggregate properties—Function, select a function appropriate to aggregate values.
- 5 From My Data, drag a column that contains values to be aggregated, and drop it in Origin column, as shown in Figure 3-7.

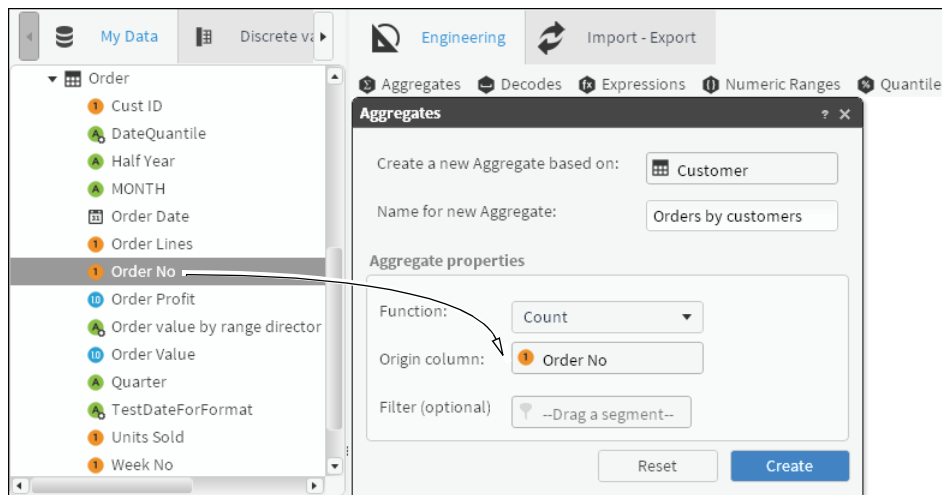


Figure 3-7 Defining properties for an aggregate

- 6 To create a filter that defines a specific group to be aggregated, double-click a column in My Data. From Discrete Values, drag a segment, and drop it in Filter (optional), as shown in Figure 3-8.

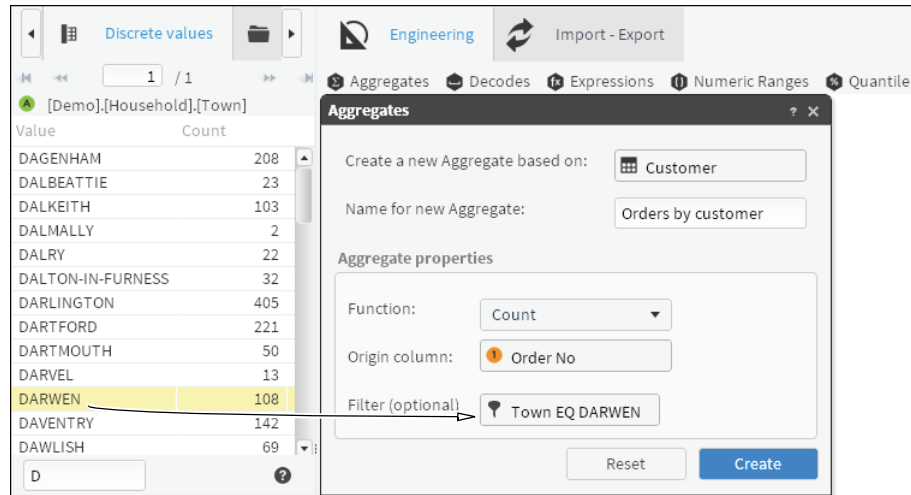


Figure 3-8 Adding a filter to an aggregate

- 7 Choose Create.
- 8 In My Data, double-click the newly created aggregate to view the results. Calculated values appear in Discrete Values, as shown in Figure 3-9.

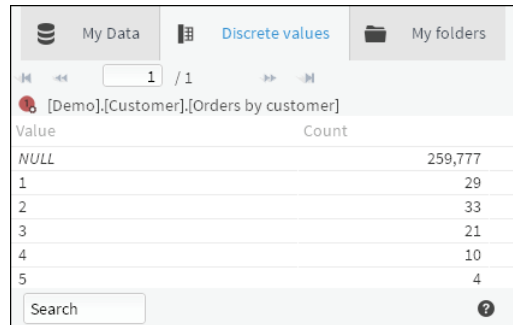


Figure 3-9 Viewing values in an aggregate

Decoding a field name

You can decode any field that contains up to 100 discrete values. For example, gender is usually encoded as M and F, or even as 0 and 1. Decodes supports changing the name to something more recognizable, such as Male and Female.

How to create a decode

- 1 In Enrichment—Engineering, choose Decodes.
- 2 From My Data, drag a column to decode, and drop it in Decodes, as shown in Figure 3-10.

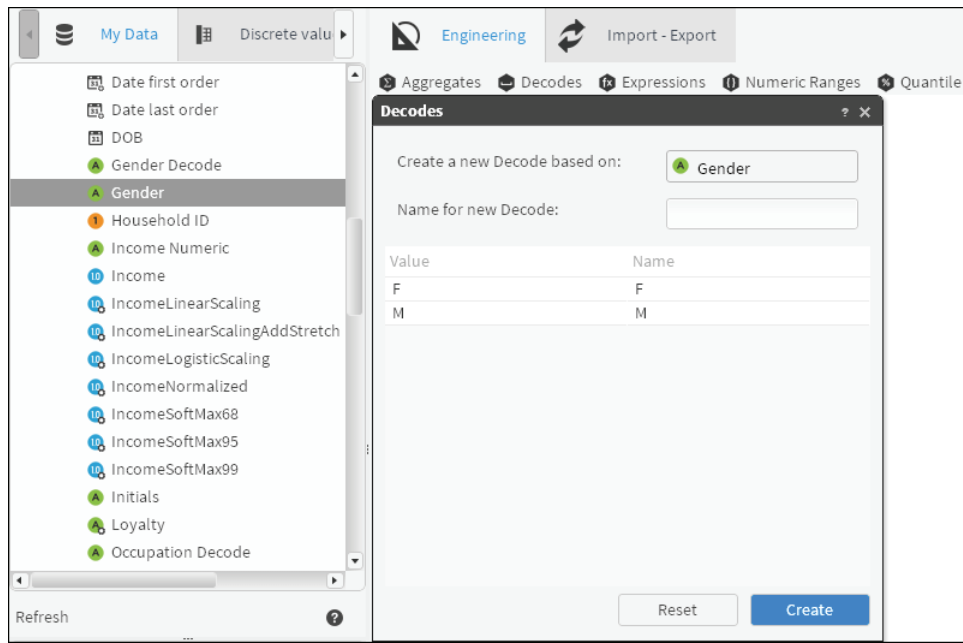


Figure 3-10 Defining a column to be decoded

- 3 In Name for new Decode, type a name for the decoded column.
- 4 To edit each field name, double-click a value in Name. Then, type any character, except the following characters, in the highlighted name field, as shown in Figure 3-11:

/ \ ° ^ - accents, ?, !, *, @, #, &, \$, o, ñ

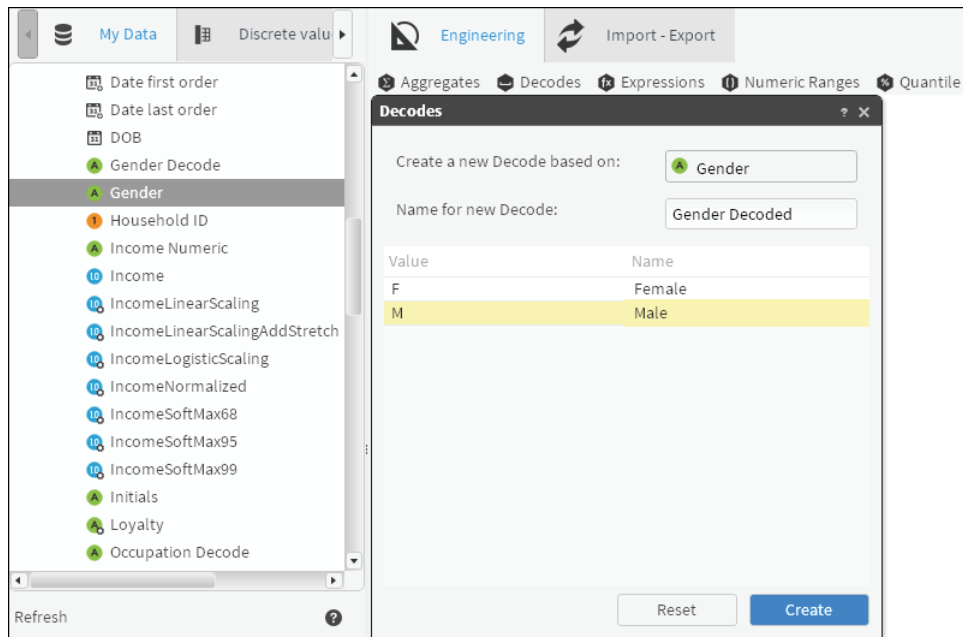


Figure 3-11 Providing a new name for a decoded column

- 5 Choose Create.
- 6 Double-click the newly created decoded field in My Data. Renamed values appear in Discrete Values, as shown in Figure 3-12.

Value	Count
Female	102,042
Male	157,832

Figure 3-12 Viewing values in a decoded field

Working with expressions

Use an expression to create a calculated field from one, or multiple fields in a database. You can create a simple expression that concatenates strings by joining several fields or a complex expression that uses mathematical operations, such as logarithms. For expressions generating a decimal field, you can specify precision up to six decimal places. For example, a typical calculated field can display customer orders based on the year in which each order was placed.

To further demonstrate, you can use the following equations to accomplish the listed tasks:

- Concatenate Text (+): join two text fields.

```
[Demo] . [Customer] . [Title] + [Demo] . [Customer] . [Surname]
```

- Concatenate (+): join two numeric fields or strings.

```
STRING ( [CustomerID] ) +5+ [DaysSinceLastOrder] -1
```

- Mathematical functions (+, -, *, /): numeric fields can be added, subtracted, multiplied, or divided.

```
STRING ( [Demo] . [Customer] . [Cust_ID] + [DEMO] . [Customer] . [Household_ID] )
```

- Boolean values: (logic 1, logic 0) assignment of Boolean values to fields, based on the conditions that are established.

```
[Demo] . [Customer] . [Edats] EQ "Adults" > 0 or  
[Demo] . [Customer] . [Edats] EQ "Joves" > 0
```

- Constants: a constant is a fixed value; in this case a constant function is one that gives a single value for a set of variables.

```
[Demo] . [Customer] . [Household_ID] = 1
```

This expression returns the value to position 1 in the Household_id field.

- Switch statement: checks the content of a field and, depending on its value, replaces it with the specified value.

```
SWITCH ( [Demo] . [Customer] . [Gender] ) CASE "M" : "  
MALE" , CASE "F" : "FEMALE" SWITCHEND
```

- Conditional Boolean value (Boolean if statements): complex conditions can be tested, such as whether the customer's age has a specific characteristic. Returns either true or false. Instead of returning true or false, the expression can return other values, such as yes or no.

```
If ( [DEMO] . [Order] . [OrderDate] > [Demo] . [Customer] . [DOB] , 1, 0)
```

To display help for a function, double-click the function. Figure 3-13 shows the help for the AGE function.

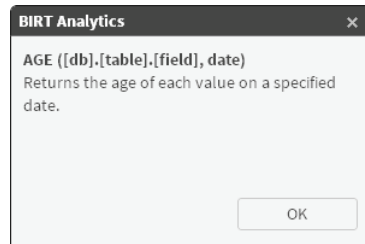


Figure 3-13 Help for AGE function

How to create an expression

- 1 In Enrichment—Engineering, choose Expressions.
- 2 From My Data, drag a destination table for the field you want to create.
- 3 Type a name for the field, using any characters except:
/ \ ° * - accents, dieresis, ?, !, *, @, #, &, \$, o, ñ
- 4 Drag a function from the left pane and drop it in the function editor. The function appears in the editor, followed by parentheses.
- 5 Drag a field name from the destination table and drop it in the parentheses.
- 6 If necessary, edit the expression syntax.
- 7 Choose Create. A new calculated field appears in the destination table.

Creating numeric ranges

To create a unique field in which the values from a numeric field appear as a set of numeric ranges that you define, use Numeric Ranges. Two typical examples follow:

- From a numeric field that contains profit per order, you can create a group of numeric profit ranges. Label those ranges, for example: Very High, High, Medium, Low, and Very Low.
- From a numeric field that contains age values, you can create a group of age ranges, such as 10–20, 20–30, 30–40 and so on, that fit one or more types of data analysis.

How to create a field containing numeric ranges

- 1 In Enrichment—Engineering, choose Numeric Ranges.
- 2 Drag the appropriate field to Numeric Ranges, as shown in Figure 3-14.

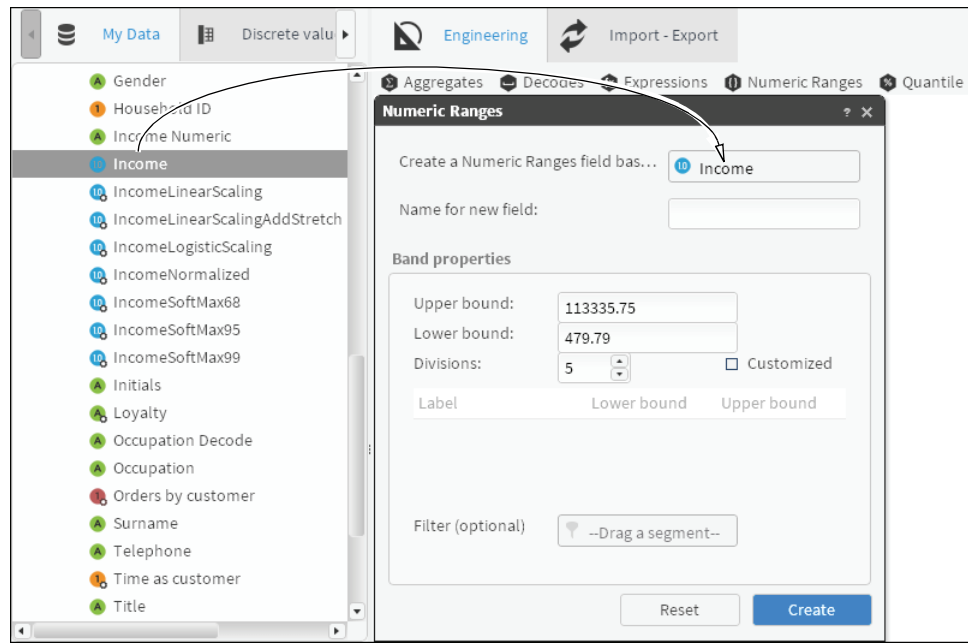


Figure 3-14 Dragging a field to use as a basis for a new numeric range

- 3 Type a name for the new field.
- 4 In Band properties, set the following properties:
 - To create default ranges, type the highest field value in Upper bound, the lowest field value in Lower bond, and a number of ranges to create, in Divisions.
 - To create customized ranges, select Customized, as shown in Figure 3-15. Then, type a name for each range, and the upper and lower limit for each range. The lower limit must be equal to the lowest value in the range. The upper limit must be greater than the values in the range. For example, the range 15 to 20 includes values 15 through 19. The range 20 to 25 includes values 20 through 24.

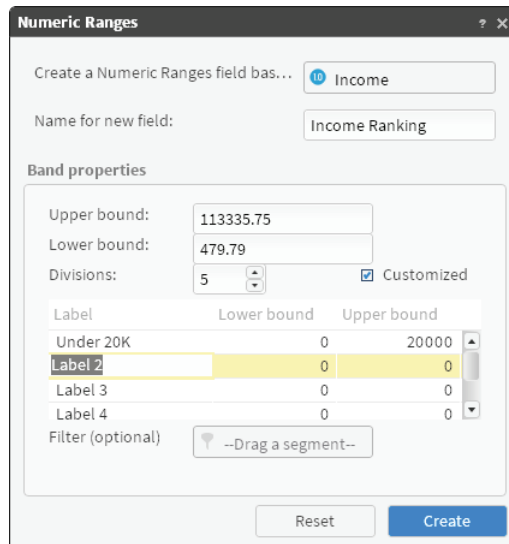


Figure 3-15 Creating customized numeric ranges

- 5 Optionally, to limit the number of values in the new field, drag and drop a segment in Filter (optional).
- 6 Choose Create.

Working with quantile ranges

A quantile range field is created from a numeric field. It creates groups that contain the same number of values. These groups are created by sorting the values in ascending order so that the first range contains the smallest values and the last range contains the largest values. For example, you can identify customers that generate the greatest profit.

How to create a quantile range field

- 1 In Enrichment—Engineering, choose Quantile.
- 2 Drag a numeric field from My Data, and drop it in New quantile based on.
- 3 In Name for new field, type a name for the quantile range field, as shown in Figure 3-16.

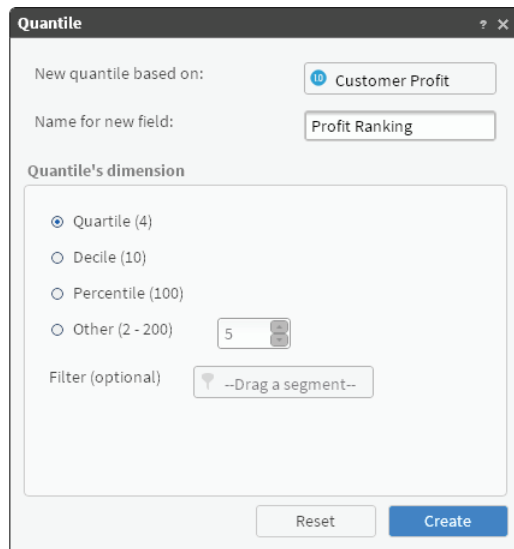


Figure 3-16 Creating a new quantile field

- 4 Select one of the following quantile types:
 - Quartile. Divides the sample into four equal parts.
 - Decile. Divides the sample into ten equal parts.
 - Percentile. Divides the sample into one hundred equal parts.
 - Other. Divides the sample into a selected number of two to two hundred equal parts.
- 5 Choose Create.

Understanding parametric columns

Parametric columns consist of query-based values. These queries are formulated with values or fields belonging to any table in the database as long as the tables are properly joined. The resulting field is created in the table indicated during the process, so the results are very different depending on the resolution marked by the selected table.

For example, you can determine sales of a particular product. For the customers table, the query returns the number of customers who bought one or more quantities of the product (customers account), while for the products table the query returns how many X products have been sold (products account). The number of products sold is probably greater than the number of customers who have bought them, because one customer may buy more than one unit of a product, but one unit of a product cannot be sold to more than one customer.

How to create a parametric query

- 1 In Enrichment—Engineering, choose Parametric.

- 2 Drag a table from My Data, and drop it in Parametric, as shown in Figure 3-17.

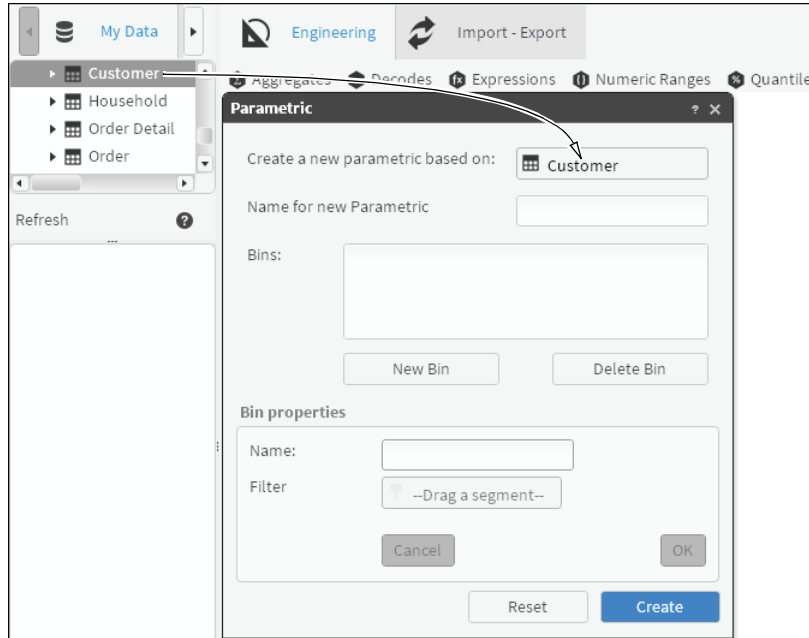


Figure 3-17 Setting the resolution table

- 3 In Name for new Parametric, type a name for the new parametric.
- 4 Choose New Bin.
- 5 Type a name for the bin.
- 6 Drag a segment from Discrete Values or Scratchpad and drop it in Filter, as shown in Figure 3-18.

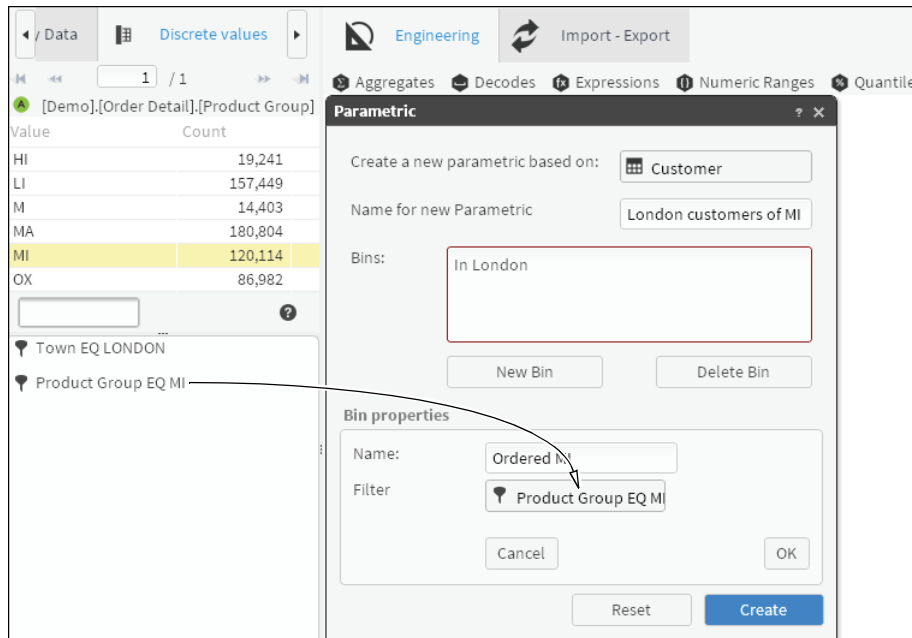


Figure 3-18 Setting a bin filter

- 7 Choose OK.
- 8 Repeat steps 4 through 7 to add more criteria to the query, if desired.

9 Choose Create.

To use the newly created parametric field, drag and drop it into Data Explorer. The parametric field itself does not appear in Data Explorer, because the same record may meet the condition of several values in the parametric field.

Understanding ranking

Use ranking to create a new column with the rank for each subset of records relating to the same key. Ranking allows you to rank a table based on a specific column. For each subset of records associated with the same key in a table, a sequence number is assigned according to the relative value of a given field. Ranking is used, for example, when trying to identify the first products purchased by each customer, or the most recent product purchased.

How to create a ranking column

- 1 In Enrichment—Engineering, choose Ranking.
- 2 In New column name, type a name for the new column.
- 3 Select a ranking order, Ascending or Descending.
- 4 To select the key to rank, drag a column from My Data.
- 5 To select the field to use for ranking, drag another column from the same table, as shown in Figure 3-19. One table must include both key column and rank field.

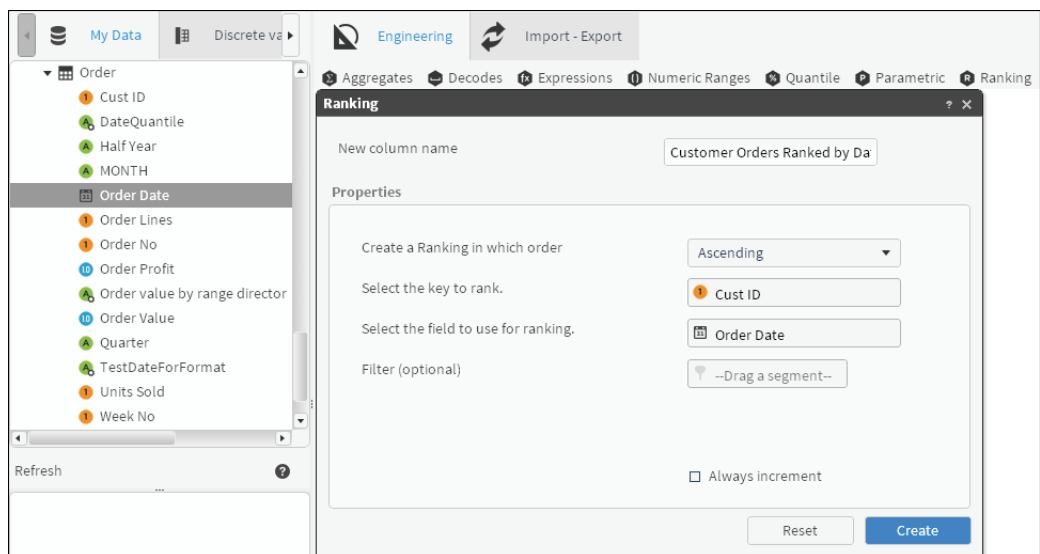


Figure 3-19 Selecting properties for a new ranking field

- 6 Optionally, to filter the ranked column, drag a segment from either Discrete Values or Scratchpad and drop it in Filter (optional).
- 7 In cases where two or more records have exactly the same value, select Always increment to ensure that each record is allocated a different rank. If you do not select Always increment, records with the same value have the same rank.
- 8 Choose Create. The new field appears in the table of rank elements. For example, you obtain 230,466 orders ranked with value 0, corresponding to the first purchases for each customer.

4

Analyzing your data

This chapter contains:

- About analyzing your data
- Using crosstabs
- Using Venn diagrams
- Using bubble analyses
- Using evolution
- Using profile analyses
- Using map analyses
- Using Pareto analyses

About analyzing your data

BIRT Analytics provides many ways to analyze data. The analyses make it possible to get new perspectives on your organization's information by crossing various fields, searching for characteristics that are or are not shared by groups, identifying suitable targets for each purpose, showing the evolution over time of different variables, and grouping values in different clusters.

BIRT Analytics supports the following types of data analysis: crosstabs, Venn diagrams, bubbles, evolutions, profiles, maps, and Pareto. You can save the definition of the analyses (the means of generating them) for reuse. When you run a saved analysis, any changes in the data appear in the current result of your analysis.

Using crosstabs

A crosstab is an analysis that supports cross-tabulation, or pivoting, of different fields from a data source. The discrete values in a field are used as labels for either rows or columns. The default value for the cells in the table is the count of records that match the intersection of the column and row value. You can, however, add a data value for the cells separately.

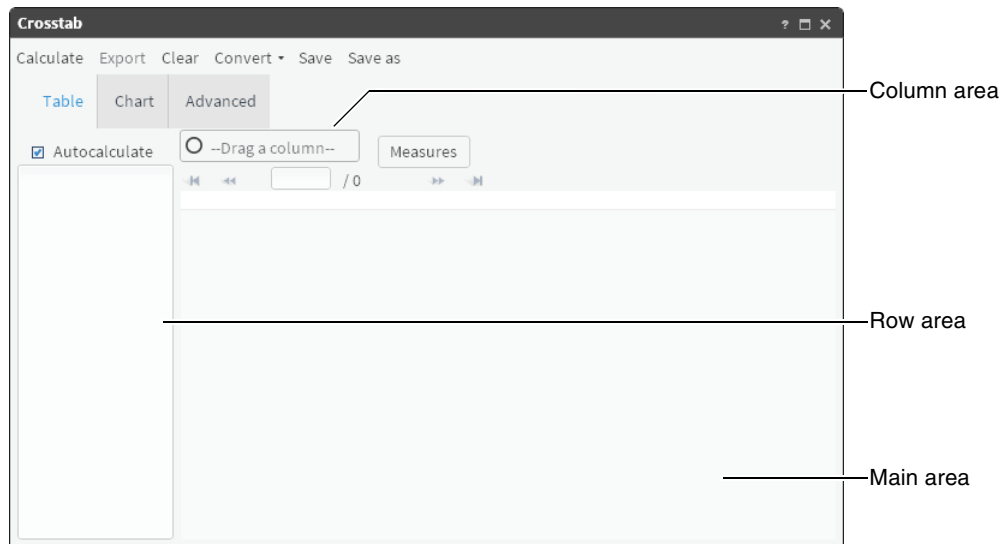


Figure 4-1 The crosstab window

For example, you can set customer gender for columns, resulting in one column for male and one column for female. Then, you can set product line for rows, resulting in a row for each product line. The cells of the table by default show total product line purchases by male and female customers. Instead, you can show average product line profit by male and female customers or maximum product line unit sales per order for male and female customers.

A crosstab that defines only rows is called non-pivoted. In a non-pivoted crosstab, a column appears for each field and rows appear labelled with the discrete values from those fields. The single, generated column on a non-pivoted table shows counts of records for each row combination. Figure 4-2 shows an example of a non-pivoted crosstab.

Occupation Decode	Income Numeric	Count(Customer)
Director		19
Director		252
Director	20 - 30K	1
Director	30 - 40K	86
Director	40 - 50K	1,408
Director	50 - 60K	3,909
Director	60 - 70K	2,130
Director	70 - 80K	246
Director	80 - 90K	2
House Person		1,002
House Person	10 - 20K	50
House Person	10K under	11,652
Manager		1,054
Manager	10 - 20K	2,289

Figure 4-2 Creating a non-pivoted crosstab

When you have specified a resolution table that is different from the axis table and is at a higher resolution, use a field from the resolution table as a column. In other words, if the table results, or cell values, come from a different table than the row labels, you must use a pivoted crosstab. For example, to analyze Customer Occupation and Income Bracket at a higher level, by Household, define Occupation and Income Bracket fields as rows on a pivoted crosstab. Then, change the resolution table for the existing measure from Customer to Household, and add Household Country as a column.

When considering two fields to use in a crosstab, you typically use the field with more discrete values to label rows. Fields with a very high number of discrete values are less useful as a basis for pivoting a data set.

Creating a crosstab

To create a crosstab, in Analytics—Analysis, choose Crosstab. Then, define properties of the crosstab, using the following features:

- **Table**
 Table provides an area for defining multiple rows and a column. To define a row or column, drag a field from Data Tree and drop it in the row area or the column area. As you drop the field, the crosstab appears in Table.

 You can change the configuration of the table by changing the order of fields in the rows definition or by eliminating measures. You must define at least one measure to produce valid output.

 Double-clicking a field in the row area provides the option to deactivate some of its values. If you select Autocalculate, data is updated immediately when you make a change to the analysis. If Autocalculate is not selected, you must choose Calculate to see your changes.
- **Chart**
 You can display data in charts and dynamic tables. For example, you can examine the number of orders by type of article to get average profits or a sum of purchase amounts.

A pivoted crosstab supports the following chart types:

- Areas
- Areas with scroll
- Bars
- Bars 3D
- Columns
- Columns 3D
- Columns with scroll
- Lines
- Lines with scroll
- Stacked areas
- Stacked bars
- Stacked columns
- Stacked columns 3D
- Stacked columns with scroll

The Columns 3D chart is shown by default.

A non-pivoted crosstab supports the following chart types:

- Areas
- Bars
- Columns
- Columns 3D
- Doughnut
- Doughnut 3D
- Lines
- Pie
- Pie 3D
- Advanced
 - Use Advanced to specify data for a table, including table axes or dimensions, simple or calculated measures, and filters. The following sections provide details about Advanced.

Understanding dimensions

A dimension is an axis of the crosstab, one of the fields which is crossed with another field to provide an analysis of the data.

You usually populate the dimensions of your crosstab by dragging a field from Data Tree and dropping it in the row area, column area, or main area of Table.

Alternatively, use Advanced—Dimensions to define rows and columns for a crosstab. Choose Calculate to update the table results.

Using a field as an axis on a crosstab automatically creates a measure for the count of the number of records with that discrete value. You can change the measures to something other than a simple count and set filters to limit the data included on the table.

Understanding measures

A measure in a crosstab is a value that appears in an individual cell. By default, a measure value is a count of the values that match the cross of the fields in the rows and columns.

To create a new measure, drag a field and drop it in the main area of Table. Select an operator for this field, for example sum, mean, minimum, or maximum. To render the crosstab, calculate the cells using the chosen dimensions and operator.

Use Advanced—Measures to change or add measures used in the analysis. In contrast to Table, Advanced supports adding complex measures that interact with other measures, changing or adding axes, and changing the order of the measures.

How to edit a measure

- 1 In Advanced, on Measures, select a measure, as shown in Figure 4-3. Double-click a measure to open it for editing.

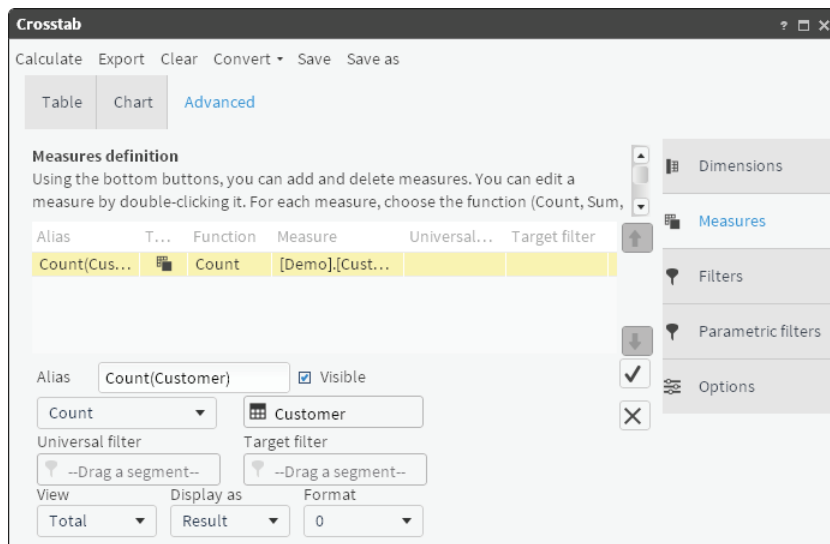


Figure 4-3 Editing a crosstab measure

- 2 Select an operation. For example, select sum, average, maximum, minimum, or count.
- 3 Select the field or table to which to apply the operation. The count operation requires that you select a table, not a field.
- 4 Select filters for each measure, if necessary.
- 5 Select a display format for the results. For example, select total, % column, % row, or % total.
- 6 Select a display type for the result. For example, select values, index, or difference.

To reuse a measure, drag the measure from one analysis window and drop it in another.

After defining simple or standard measures, you can define calculated measures. For example, to estimate cost per operation, define a calculated measure that divides the sum of operation costs by the number of operations.

How to define a calculated measure

- 1 Create simple or standard measures. For example, create Cost Sum and Cost Count.
- 2 In Advanced, on Measures, choose New formula.

- 3 In Alias, type a name for the calculated measure. For example, type:

Approximate cost of operations

- 4 Right-click and choose the simple measures and operators from the context menu.
- 5 Choose the check mark to confirm that the syntax is correct.

How to create a new measure based on an existing measure

- 1 In Advanced, on Measures, select a measure.
- 2 Choose Duplicate.
- 3 Replace fields, or edit filters, formulas, values or properties of the duplicate measure.
- 4 Save the new measure.

How to delete a measure

- 1 In Advanced, on Measures, select a measure.
- 2 Choose Remove selected.

About crosstab filters

You can set universal, target, and baseline filters. For more information about setting up filters for comparing values, see “Applying a filter” in Chapter 3, “Working with your data.”

Selecting crosstab options

After creating a crosstab, use Advanced—Options to modify the crosstab appearance. You can choose whether or not to display blank rows and columns. For a graphical display of the table, you can set the maximum characters for axis labels, which determines whether the labels are complete, truncated, or hidden. Only pivoted and pie-style charts support showing a legend.

Optionally, you can set *y*-axis lower and upper limits for the crosstab chart according to crosstab maximum and minimum values. Setting these limits improves chart readability for a chart on which values appear too closely spaced. You can sort the results in ascending or descending order by measures.

Other crosstab functions

Crosstab provides the following tools:

- Calculate
When Autocalculate is not selected, choosing Calculate manually re-draws the table after you make configuration changes. Also, use this tool to redraw a chart.
- Export
Saves your crosstab as a comma-separated values (CSV) file for use in common spreadsheet programs or in other databases.
- Clear
Removes the contents of your crosstab without saving.
- Convert
Changes a crosstab analysis into a bubble, an evolution, or a map analysis. You may need to modify some settings to make the new chart type render in a correct and logical way.

- Save
Keeps a crosstab configuration in a previously named file. The file name appears in My Folders in Data Tree, overwriting the previous file. A prompt to establish the file name appears when you choose Save for a crosstab that has not been saved previously.
- Save As
Keep a crosstab configuration in a new file. The new file appears in My Folders.

Using Venn diagrams

A Venn diagram provides an analysis of data by crossing different fields. A Venn analysis identifies coincidences and differences between N segments. Up to five different segments appear on the graphic display in Chart. Results for segments beyond five appear in Table only.

How to create a Venn diagram analysis

- 1 Drag a segment from Discrete Values or from Scratchpad and drop it in Chart in Venn, as shown in Figure 4-4.
- 2 Drag additional segments to Chart, up to a maximum of five. The Venn diagram is drawn as you drop each item into Chart.

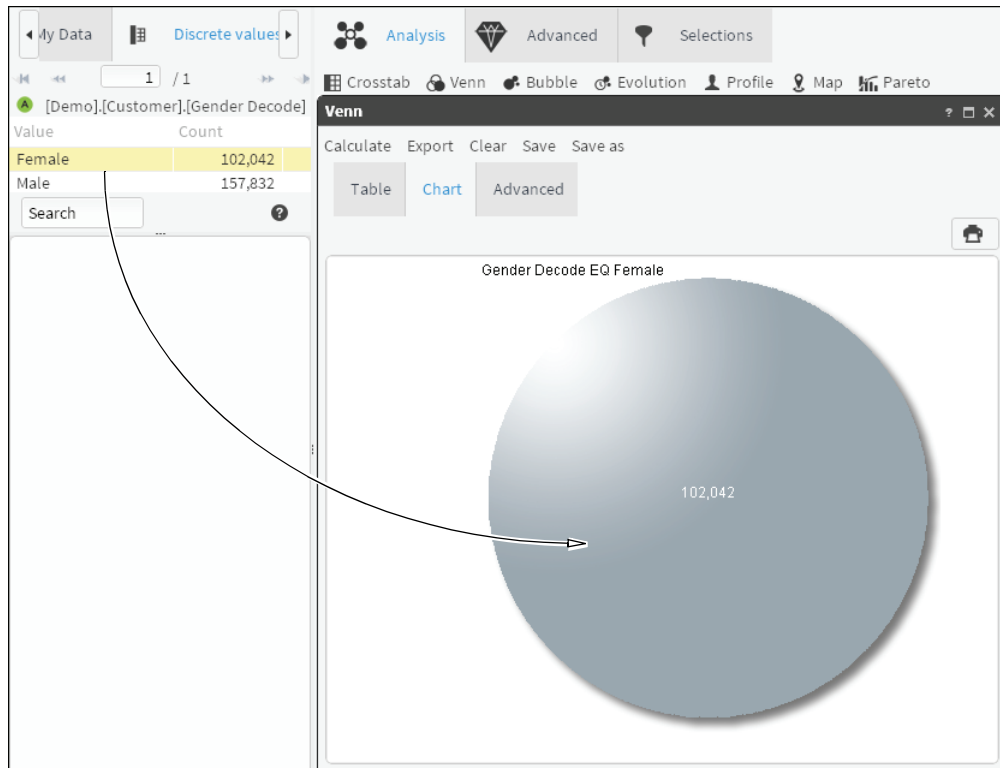


Figure 4-4 Dragging a discrete value to a Venn analysis in Chart

For example, to identify cross-selling opportunities, examine customers who have purchased different categories of products. Using the Venn analysis, cross the data segments representing three different product categories. Zero coincidences occur, because a product is of one type or another, but not both, as shown in Figure 4-5.

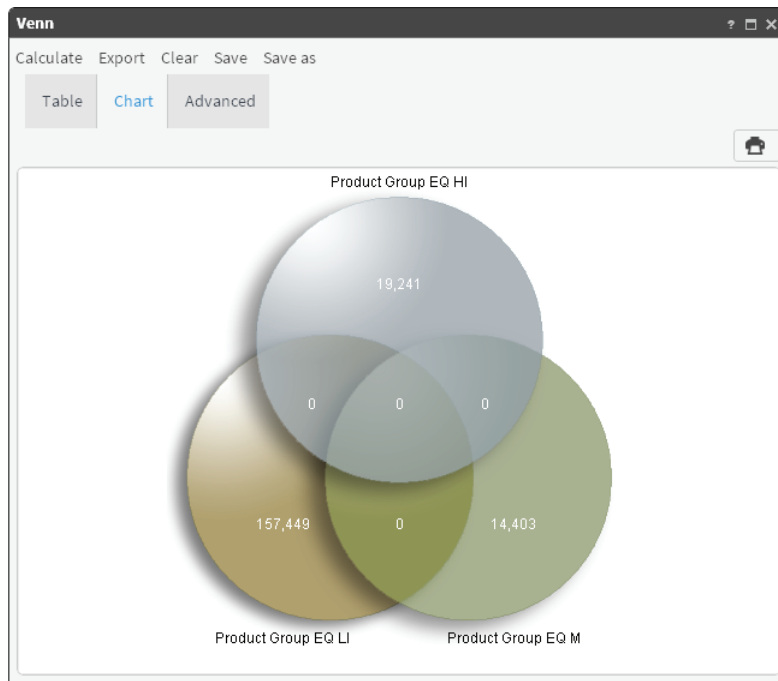


Figure 4-5 Showing three discrete values with no coincidences

A Venn diagram can show coincidences or differences at other levels of database resolution. For example, resolving the Venn diagram at the customer level shows coincidences between the two or three segments, including customers purchasing one, two, or three products.

How to change the resolution

- 1 Change the resolution by dragging a different table to Resolution Table, as shown in Figure 4-6.

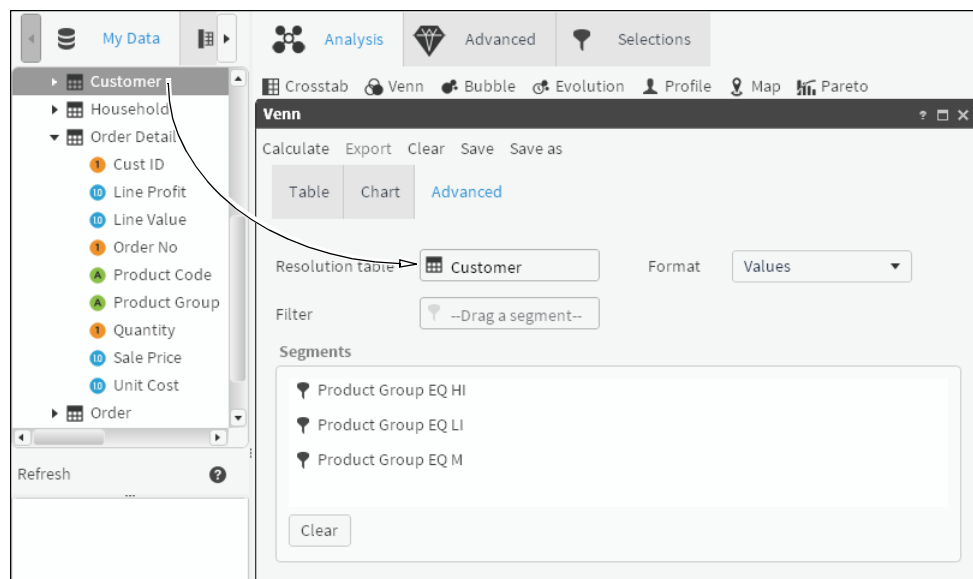


Figure 4-6 Changing the Venn diagram's resolution

- 2 Choose Calculate to redraw the chart after changes. The resulting chart is shown in Figure 4-7.

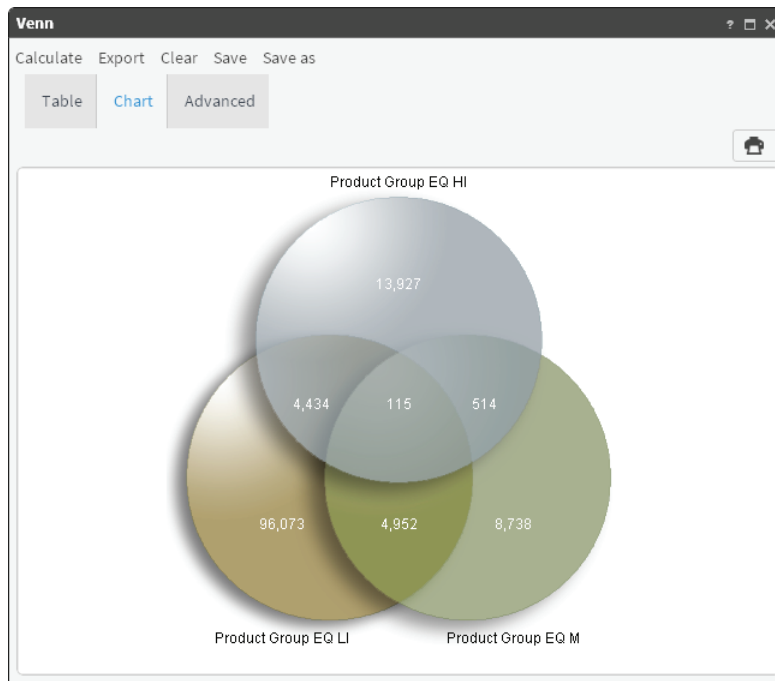


Figure 4-7 Showing a Venn diagram with a changed resolution table

On Advanced, in addition to the resolution table, you can set Filter and Format.

- **Filter**
You can include a filter at the analysis level, which means that a Venn diagram can be constrained to show results for a specific and smaller universe. For example, include the customers who have purchased one, two, or three product types, but only in Madrid.
- **Format**
You can format your results to show a count of values that match the criteria you selected, a percentage of the total number of records included the analysis, or a global percentage, which means that each percentage is based on the total possible records from the resolution table being analyzed.

Using bubble analyses

A bubble diagram shows the distribution of categorical data across two axes of numeric variables. You can also view results on a data table.

Use a third variable to set the size of each bubble. To configure a bubble analysis, set two or three variables: two on the axes and an optional third to size the bubbles.

Instead of showing measures by attributes or categorical variables, as in the case of crosstabs, a bubble diagram shows groups or categories according to numeric variables, providing a different way of showing numeric data versus categorical data. For example, you can examine customers by age, order amounts, and average salary.

Drag a field to use as a categorical variable, such as profession, gender, age ranges, or number of children. Next, select continuous or numeric variables, such as age, salary, or purchase amount. The continuous variables are the axes of the analysis. For each variable, choose a function, such as count, sum, average, maximum, minimum, and a display type, such as percentages or totals. You may optionally choose a universal or target filter that applies only to this variable or measure.

You can set filters on a bubble analysis at either the measure level or at the complete analysis level. For example, you can set a partial filter when you are analyzing customers by age, and you need to view the order average for only the last month. You also, however, want to check the average salary for these age ranges, so you do not want the filter to be applied at the global-analysis level, only at the order-measure level.

You can use parametric filters as analysis filters.

Results can be sorted in ascending or descending order by measures.

The check box Show legend is only available for pivoted and pie-style charts, but there is also an option to set the maximum number of characters to display in chart labels. Hide chart labels by setting the maximum number of characters to 0.

The legend configuration is bound to the chart configuration, so if the chart configuration (type, rows, columns, or filters) changes, the legend changes. When you select a pie-style chart, clicking a value label separates the corresponding piece of the pie, but when the bubble is multidimensional, clicking a value label hides its graphical representation.

You can convert this type of analysis into a crosstab analysis.

How to add data to a new bubble analysis

- 1 In Analytics—Analysis, choose Bubble.
- 2 Drag a column from Data Tree to use as the categorical variable and drop it in the left column. A message that you must specify two continuous variables to calculate the analysis appears, as shown in Figure 4-8. Choose OK.

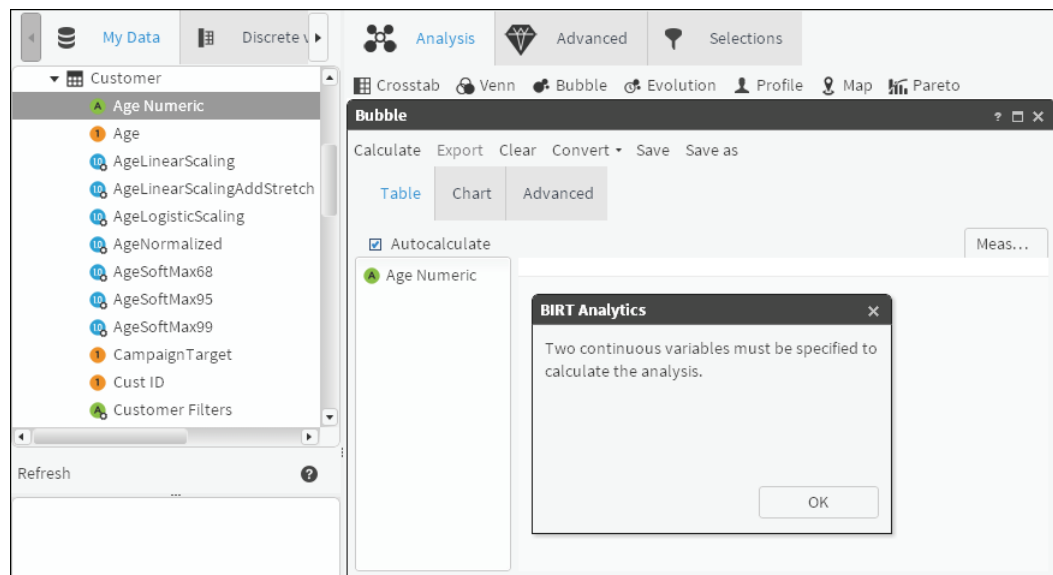


Figure 4-8 Defining a continuous variable for a bubble analysis

- 3 Drag a column from Data Tree to use as the continuous variable and drop it in the right column of the table, as shown in Figure 4-9. Choose Sum.

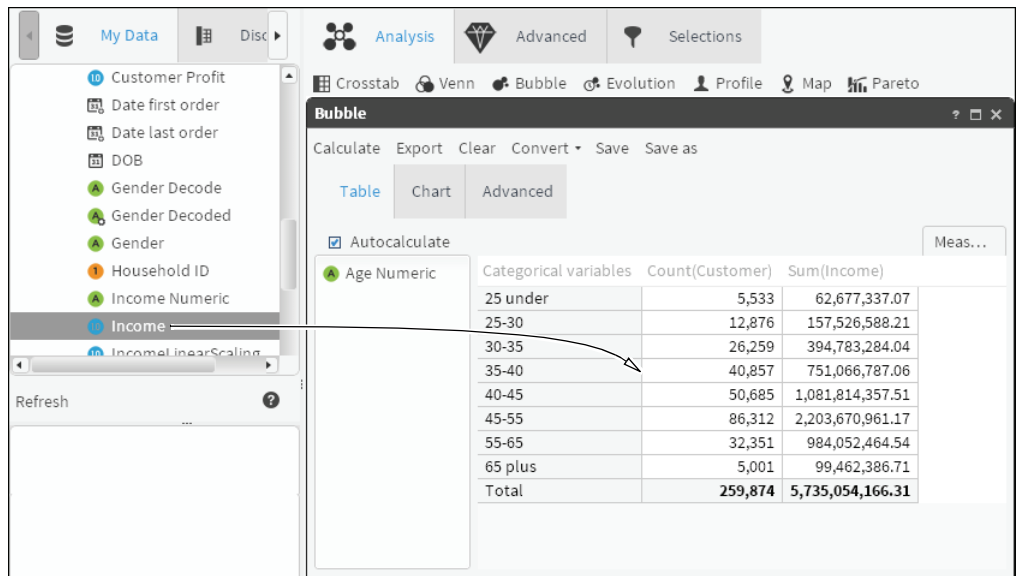


Figure 4-9 Defining a continuous variable for a bubble analysis

- 4 Choose Advanced—Measures.
- 5 Choose New measure.
- 6 Type a name for the new measure.
- 7 Select a function for the new measure, and drag a column the measure uses, as shown in Figure 4-10.

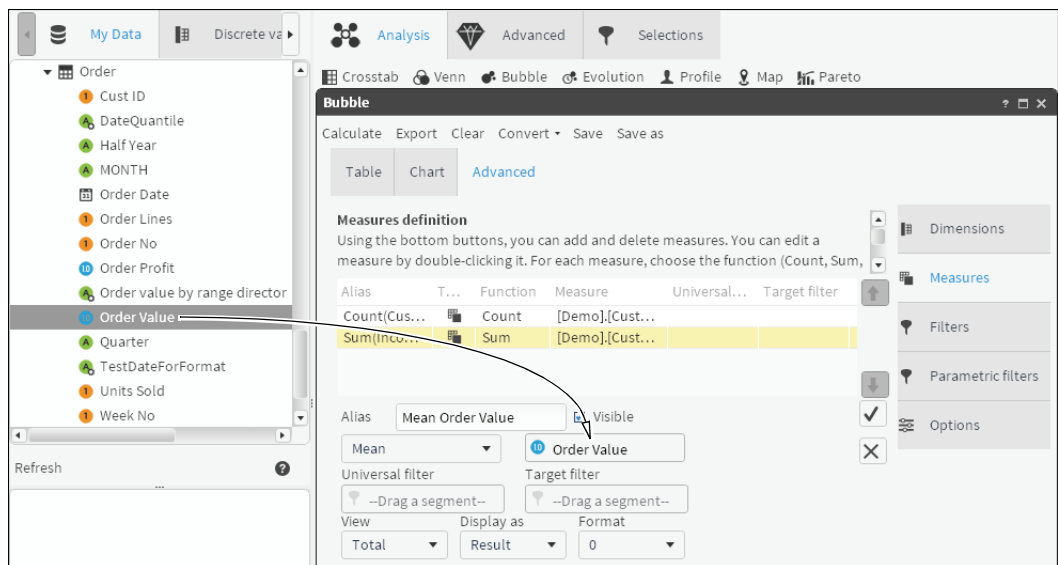


Figure 4-10 Defining a new measure for a bubble analysis

- 8 In Format, select the number of decimal places to display.
- 9 Choose the check mark to create the measure.

How to add a filter to a new bubble analysis

- 1 From Bubble—Advanced, select Filters.
- 2 From Data Tree, drag a filter to use.

- 3 To see the new bubble analysis, choose Calculate. The bubble analysis appears on Chart, as shown in Figure 4-11.

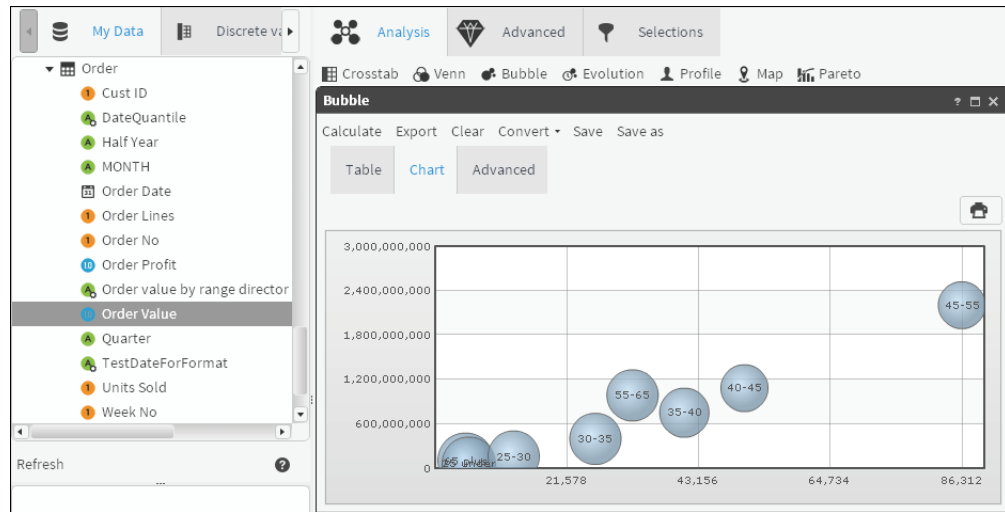


Figure 4-11 Visualizing a data measure using a bubble analysis chart

How to display multiple segments using a bubble analysis

- 1 In Analytics—Analysis, choose Bubble—Advanced.
- 2 From Data Tree, drag columns to use as the categorical variable, as shown in Figure 4-12.

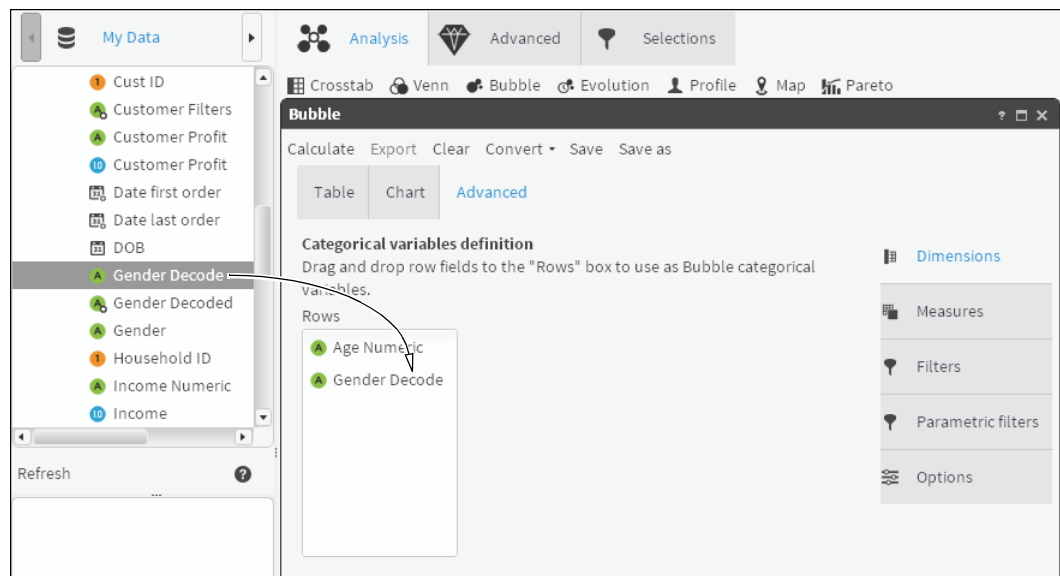


Figure 4-12 Adding multiple categorical variables to a bubble analysis

- 3 Choose Measures.
- 4 To rename an existing measure, double-click the measure and type a new name, as shown in Figure 4-13. Choose the check mark to finalize the edit.
- 5 Choose New measure.
- 6 Type a name for the new measure.

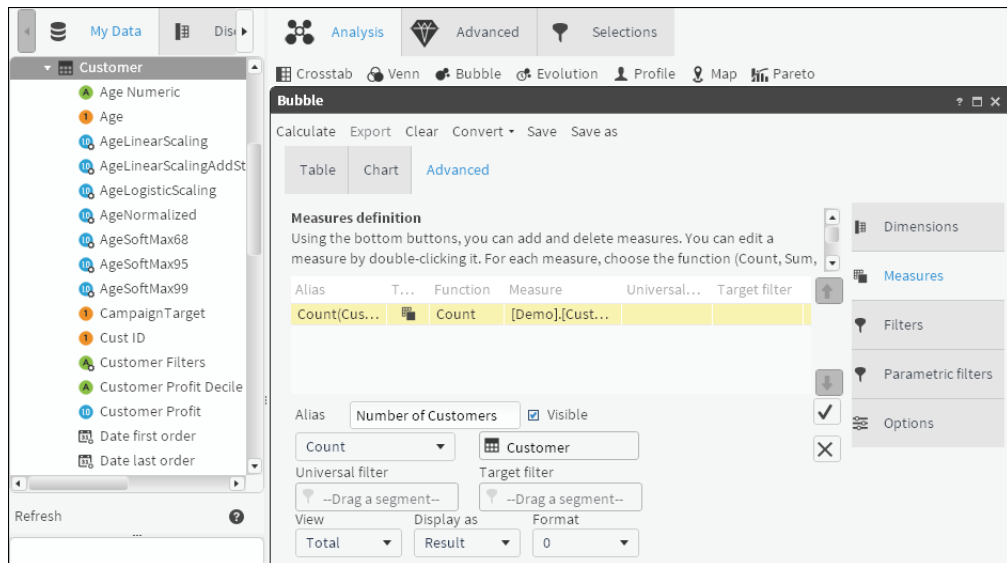


Figure 4-13 Renaming a measure

- 7 Choose a function for the new measure and drag a column the measure will use, as shown in Figure 4-14.
- 8 In Format, select the number of decimal places to display.
- 9 Choose the check mark to create the measure.

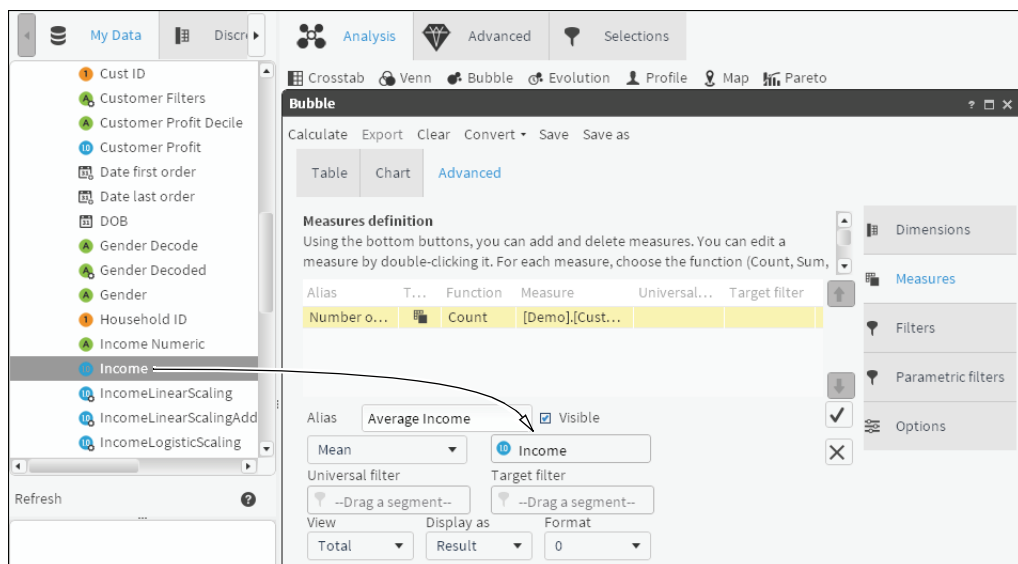


Figure 4-14 Adding a new measure to a bubble analysis

How to duplicate measures on a bubble analysis

- 1 Choose Duplicate to create a measure similar to the one that is highlighted.
- 2 Type a name for the new measure, as shown in Figure 4-15.

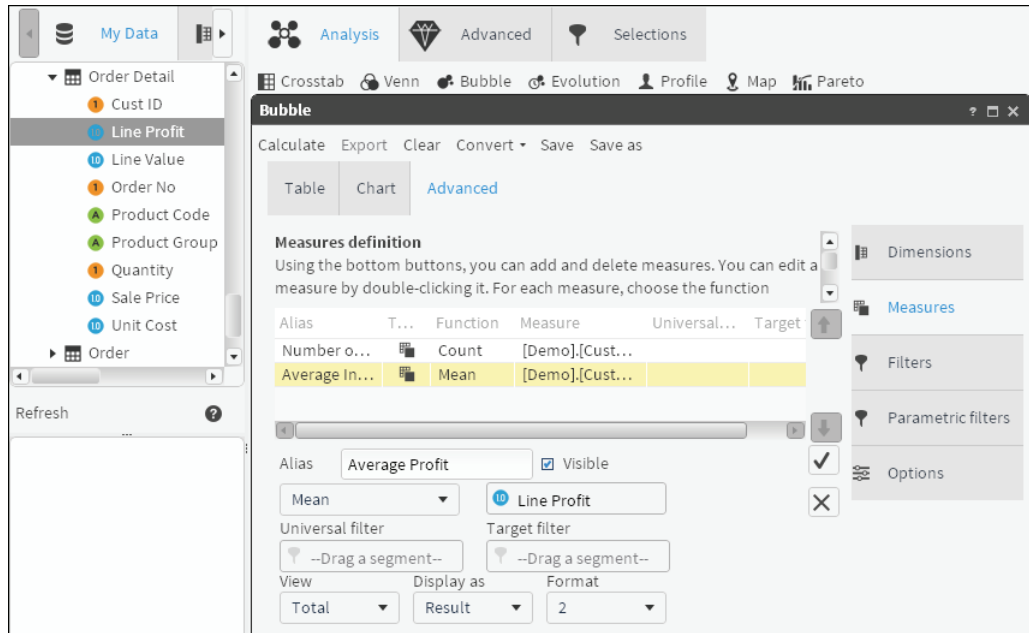


Figure 4-15 Duplicating a measure

- 3 Drag a column from Data Tree that is different from the one already displayed, as shown in Figure 4-16.

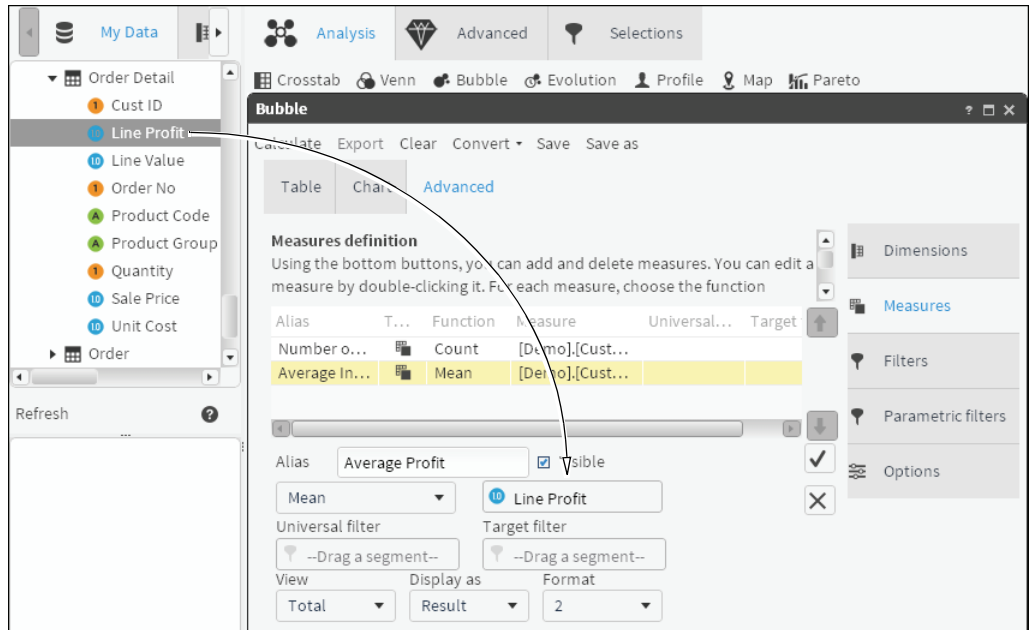


Figure 4-16 Replacing a column in a duplicated measure

- 4 Choose the check mark to create the measure.
- 5 Choose Calculate. The bubble analysis appears on Chart, as shown in Figure 4-17.

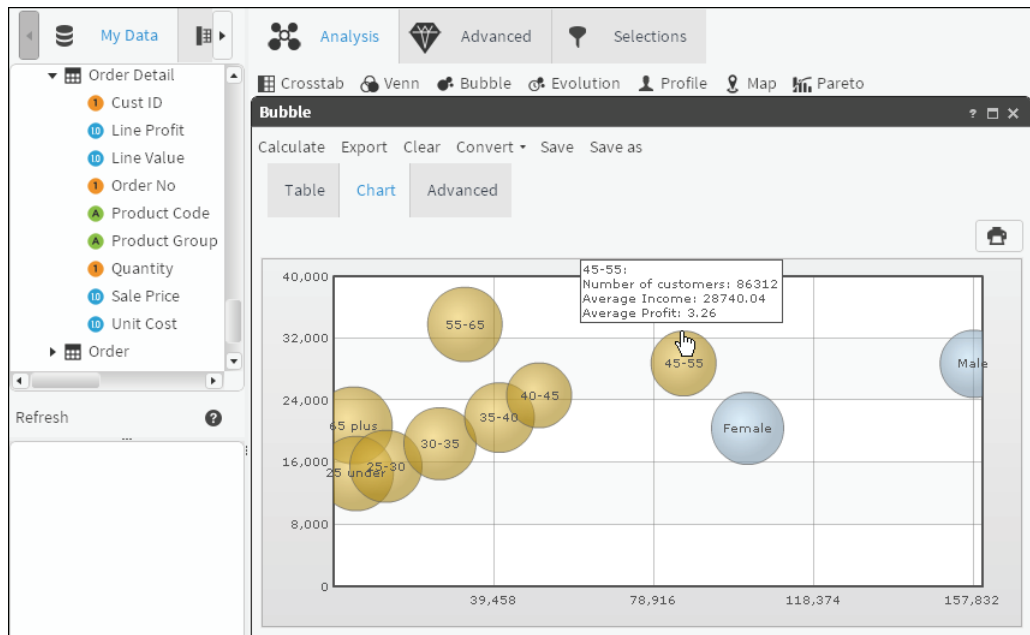


Figure 4-17 Visualizing a bubble analysis that shows multiple variables

Using evolution

An evolution analysis shows a progression of data over time. You can examine the behavior of certain measures in different periodic scenarios. For example, examine how the sales of some product families evolve over a period of months. The product family field is the categorical variable under study, while the different scenarios are determined by the different values in the month of purchase variable.

Define the following components to create an evolution analysis:

- **Categorical variable:** Field that contains the categories whose behavior is to be analyzed, or in other words, the variable under study.

To indicate the variable you want to work with, drag the field from Data Tree to the control that is in the top part of the configuration form, as shown in Figure 4-18.



Figure 4-18 Control without variables

After dragging the variable to the control, it checks the number of discrete values and shows all of the divisions as well as the single values that it contains, as shown in Figure 4-19. Each green strip corresponds to a product type.



Figure 4-19 Control with variables

You can drag to this control only fields with a maximum of 100 discrete values. In the analysis, a sphere with a specific color selected at random represents each of these categories.

If you want to delete one or more of the categories from the analysis variable, click the box that represents the category. If you position the cursor over each of the green boxes, a tag appears, displaying the value of the option.

You can also select or deselect all of the boxes by using the buttons located just below the discrete values control (all, none).

- Measures: Evolution can display up to a maximum of three measures, two of which position the spheres on the axes of the x - and y - coordinates, while the third (optional) defines the relative measure of the sphere inside the group.

To define the measures of the axes, drag the numeric fields from Data Tree to the vertical bar (y -axis) or to the horizontal bar (x -axis).

When you drag a column to the x -axis, the latter changes color, indicating that it is ready to accept a column. After you drop the field, the control shows the applicable functions available in accordance with the type of column.



To use the third measure (sphere measure), drag the column you want to use to the icon that is located in the top-right part of the form.

If the cursor is positioned over the definition controls of the measures (axes, measure control), the current operation as well as the field involved in it are displayed.

- Transition variable: It is necessary to define what data tree variable to use as a base for creating transitions. To indicate the field, you drag it to the discrete values control at the bottom of the form.

This control operates in the same way as the categorical variable control and enables you to indicate which elements to use in the analysis.

How to create an evolution

- 1 In Analytics—Analysis, choose Evolution.
- 2 From Data Tree, drag a column to use as the categorical variable, as shown in Figure 4-20.

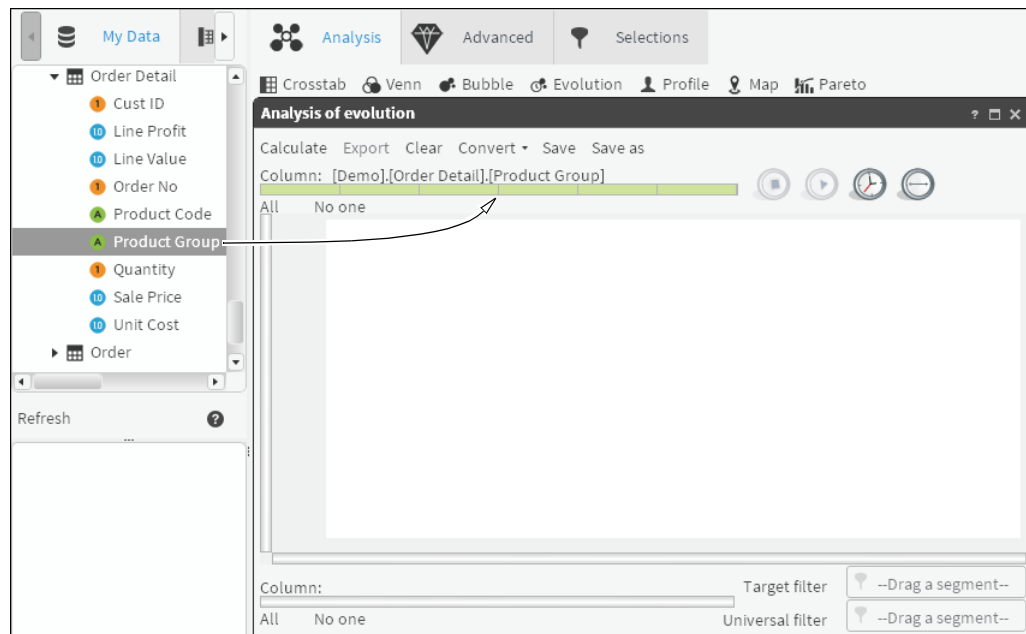


Figure 4-20 Defining a categorical variable for an evolution analysis

- 3 From Data Tree, drag a column to use as the transition variable, as shown in Figure 4-21.

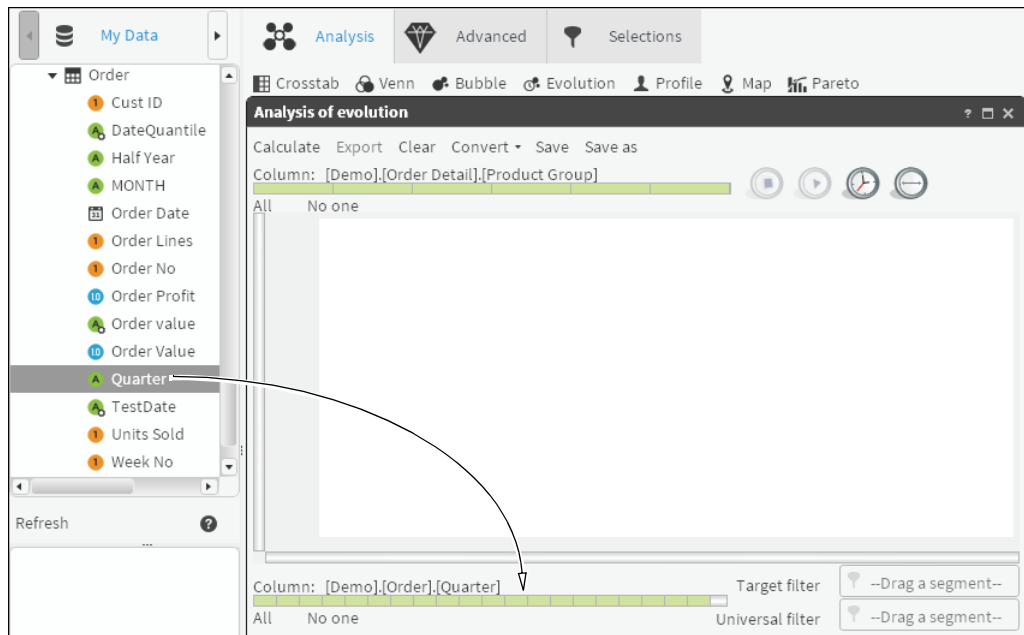


Figure 4-21 Defining a transition variable for an evolution analysis

- 4 From Data Tree, drag columns onto the x-axis and y-axis. Choose appropriate functions for each axis. Figure 4-22 shows the y-axis being added as mean order profit with the x-axis in blue, to indicate that a column has already been applied.

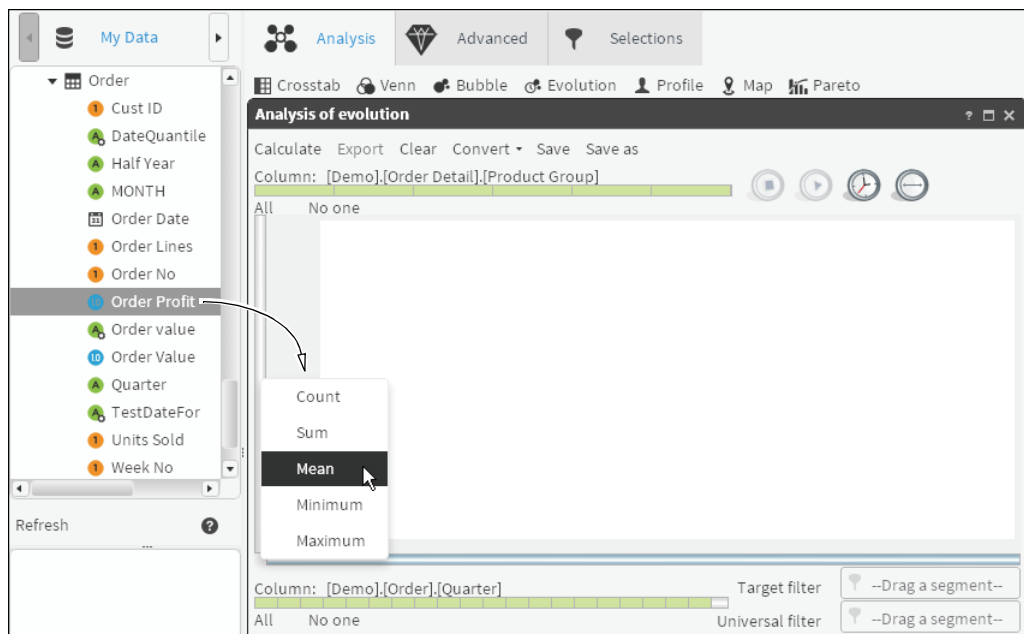


Figure 4-22 Defining y-axis properties for an evolution analysis



- 5 Choose Calculate. Choose Play to show the evolution over time. Figure 4-23 shows the evolution analysis.

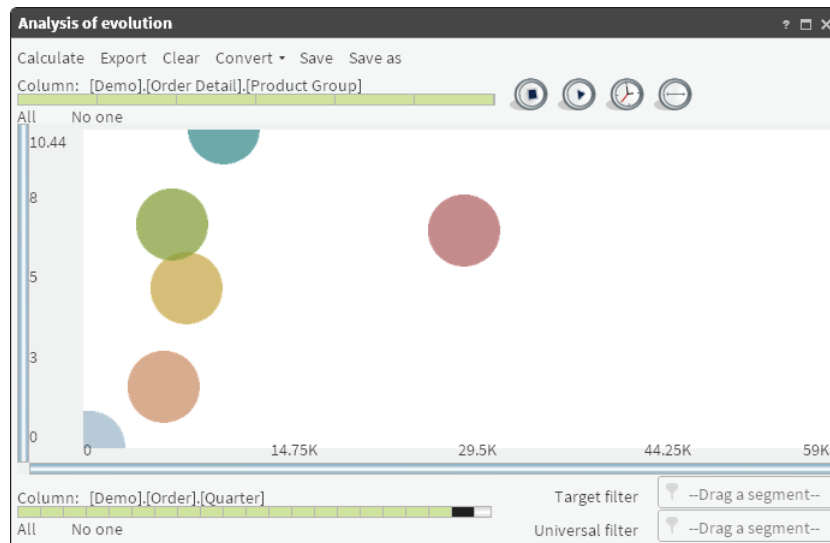


Figure 4-23 Viewing an evolution analysis



- 6** To indicate a third measure, from Data Tree, drag a column to the far right icon.
- 7** To modify the transition of time, select the clock icon and choose an interval. For example, an interval of 0.25 is faster than an interval of 3.

More about viewing an evolution

When evolution is defined, BIRT Analytics internally creates a crosstab in which the categorical variable that you want to study is the row dimension (in the case of the example that accompanies this explanation, the product category) and the transition variable (the temporary variable) is the column dimension. The crosstab also includes the two or three defined measures (x, y, and measure).

The internal result of the crosstab is similar to Table 4-1.

Table 4-1 Internal result of crosstab

	Scenario 1			Scenario 2			Scenario 3		
Category 1	x1	y1	measure1	x2	y2	measure2	x3	y3	measure3
Category 2	x1	y1	measure1	x2	y2	measure2	x3	y3	measure3

The graphic representation of the evolution analysis consists of data extracted from each of the scenarios, displayed as spheres. The animation extracts data from each scenario and displays each category as a sphere in a particular color, positioning it in a way that takes into account the maximum and minimum values of the measures for other categories.

During execution, you can hover the cursor over one of the spheres to display the name of the category and the values for each of its measures for the current transition. A transition is the category of the transition variable that is displayed by an evolution.

As the transitions occur, the active scenario of the transition variable appears.



It is possible to modify the time between transitions by choosing the time icon and selecting the speed.

The analysis definition form also allows you to set the filters for the calculation.

Recommendations

- To make this analysis useful, the number of discrete values of the categorical variable must be as low as possible.
- The analysis is deemed to exceed the maximum if it occupies more than one page.
- To learn more about internal calculations, consult the crosstab help utility.

For example, consider the question, “How do the sales of a product family change over the months?” To answer this question, create an evolution indicating the group of products as a categorical variable and the month when the order is placed as a transitional variable. Possible measures are the orders count (x -axis) and the average profit (y -axis), which enable you to see quickly the group of products that sells most over time, and the group of products that produces the most profit. You can convert this type of analysis into a crosstab analysis.

Using profile analyses

A profile analysis groups values and determines their relatedness to another group, called the profile segment. This analysis helps you draw a profile of a group of values from attributes selected in Data Tree.

In advertising, a target group refers to the people to whom the advertising is directed. The target group is equivalent to a profile segment. If a profile segment is chosen, the profile carries out a comparison with each of the attributes that are selected based on all the database’s values, including the analysis segment itself. In other words, a request is made for an analysis of the significance of the attributes to define the segment, versus all other values in the database.

Alternatively, if a base segment is selected in addition to the profile segment, the analysis is based on the values that are part of that chosen segment. In the first case the segment is defined. In the second case, whether the attributes are suitable for showing differences between both groups (profile and base segment) is defined. Three different scenarios are possible:

- Customers who provide higher profits vs. customers who provide lower profits (no overlapping of values).
- Customers who have responded to a campaign via e-mail vs. all who have responded (the segment to be profiled is part of the base).
- Customers who have bought product X this year vs. customers who bought product X last year. There are some values that are part of both segments, profile and base, but not all.

There are several indicators that measure the significance of the attributes to define the analysis segment, including Z-score. The Z-score determines whether the difference between two proportions is statistically significant. In the case of profile, this determination is carried out between the group to be analyzed and the group considered Rest, values that belong to the attribute whose significance you want to measure but not to the analysis segment.

The Z-score uses the binomial for unequal variances, whose formula is:

$$Z = (p1-p2) / \text{sqrt}((p1*(1-p1)/n1 + (p2*(1-p2)/n2))$$

where

- p1 is the sampled proportion and n1 is the number of sampled values.
- p2 is the proportion in Rest and n2 is the number of Rest values. Table 4-2 shows calculated Z-scores.

Table 4-2 Calculating Z-scores

	Men	Rest	Z-score
M	34.4%	19.0%	171.31
X	27.9%	16.6%	132.66
W	17.0%	9.8%	102.52
S	20.7%	54.6%	-336.01
	682,493	317,507	

In other words:

- You would first choose the characteristics or attributes whose significance you want to measure to determine the analysis segment.
- Significance means those attributes or characteristics that meet the highest possible number of cases within the segment analyzed.
- You set the characteristics or attributes of those whose significance you want to measure to define or characterize the analyzed segment:
 - How is this significance measured?
 - Using a Z-score indicator.
 - This indicator is obtained by applying a statistical formula.
 - The higher the Z-score of the attribute, the higher the defining capacity, or relatedness, of the segment analyzed.

A Z-score higher than 0 can be considered significant to describe a segment, but values below 1.96 cannot be considered statistically significant. Only values above this threshold can be considered significant. You set attributes of which you want to measure the significance, which means that other values that are not chosen for the study may be the ones that are truly statistically significant to define the analysis segment.

How to create a Profile analysis

- 1 In Analytics—Analysis, choose Profile.
- 2 In Profile, choose Advanced.
- 3 To create a segment, drag one or more values from Discrete Values to Scratchpad, as shown in Figure 4-24. Alternatively, use an existing segment. The segment of data represents a profile group against which you test the significance of various attributes.

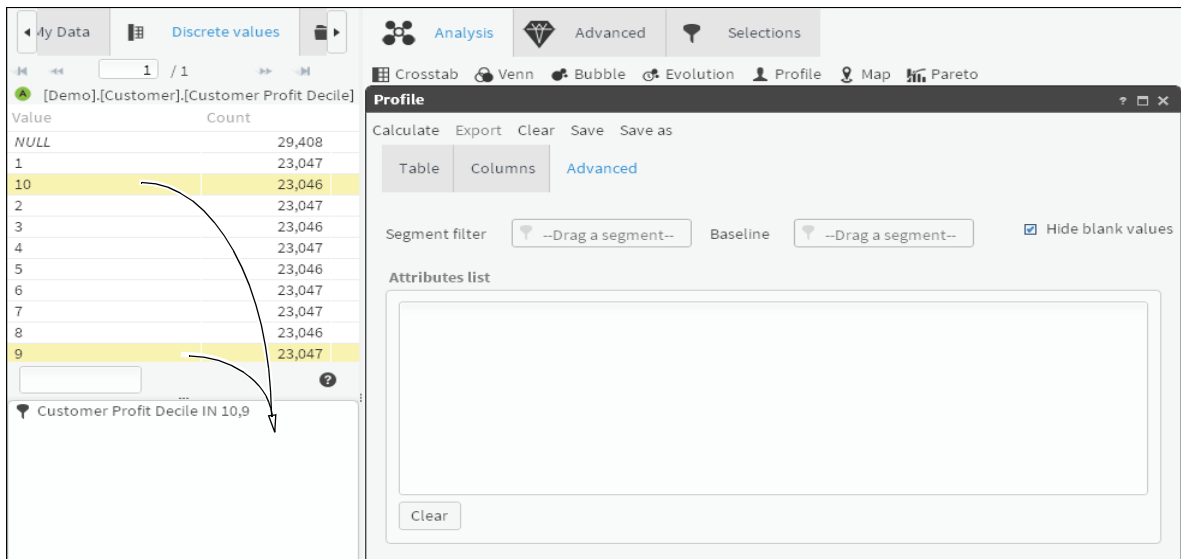


Figure 4-24 Selecting a segment using two discrete values

- 4 If necessary, rename the segment appropriately.
- 5 Drag the segment from Discrete Values and drop it in Segment filter, as shown in Figure 4-25.

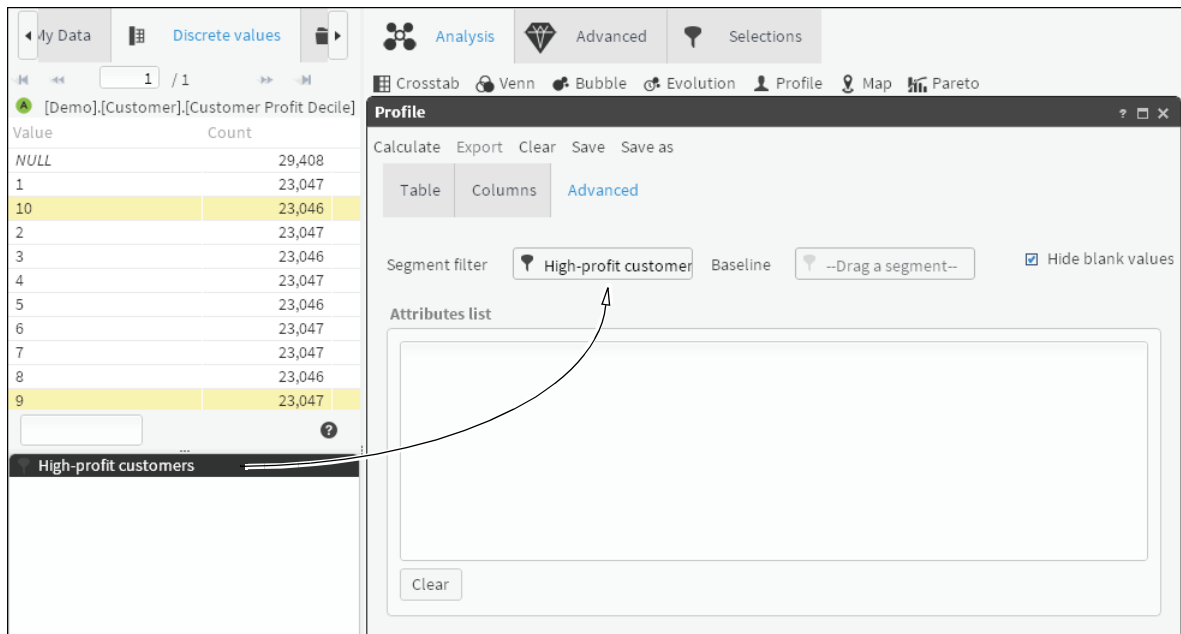


Figure 4-25 Adding a segment filter

- 6 Drag fields from Data Tree into Attributes list, as shown in Figure 4-26. Using the Profile analysis, you test these attributes for how significantly they contribute to the profile.

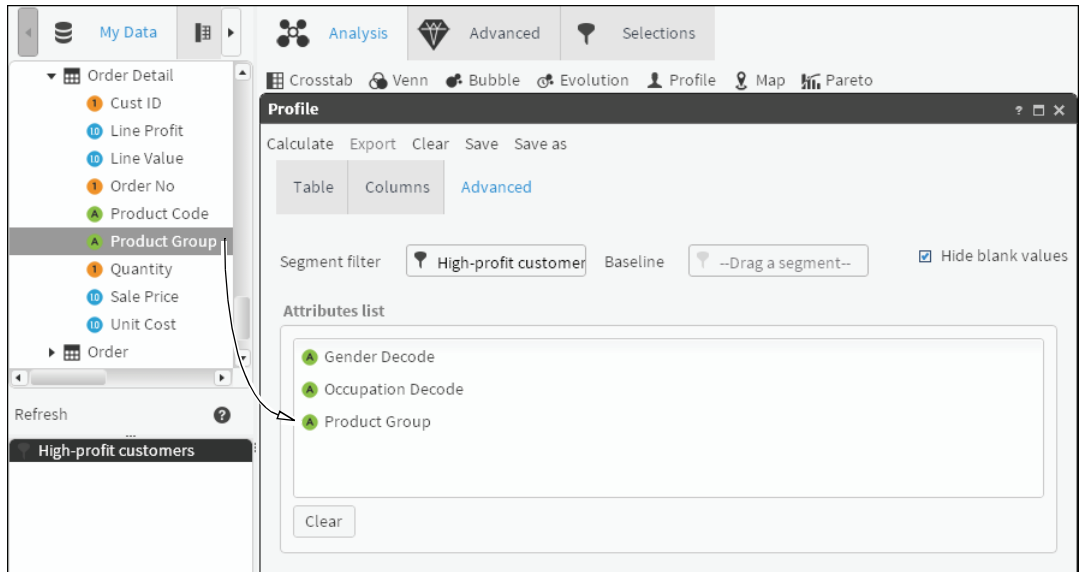


Figure 4-26 Adding profile attributes

7 Choose Calculate.

8 Examine the results on Table. The gray bars in the Graphic column show which discrete values from the list of attributes contribute most significantly to the profile, as shown in Figure 4-27. Note that the gray bars represent the Z-score, and the column is sorted by default from most significant to least significant.

Value	Column	In group	Remainder	Index	ZScore	Graphic
OX	[Demo],[Order...	28,854	67,493	213.76	136.25	★
MI	[Demo],[Order...	28,177	94,614	148.91	80.60	★
M	[Demo],[Order...	7,896	14,319	275.72	59.80	★
Retired	[Demo],[Cust...	6,954	19,742	198.60	42.89	★
LI	[Demo],[Order...	25,302	105,574	119.83	35.77	★
HI	[Demo],[Order...	5,655	18,990	148.89	24.69	★
Female	[Demo],[Cust...	20,492	102,042	113.22	20.73	★
Unemployed	[Demo],[Cust...	4,508	19,303	131.67	15.94	★
MA	[Demo],[Order...	25,713	125,773	102.22	4.78	★
Office Worker	[Demo],[Cust...	4,303	24,379	99.51	-0.31	
Self Employed	[Demo],[Cust...	3,357	19,553	96.80	-1.83	
Director	[Demo],[Cust...	1,350	8,034	94.74	-1.90	
Manual Worker	[Demo],[Cust...	5,922	36,827	90.66	-7.77	★
Manager	[Demo],[Cust...	4,326	28,330	86.09	-10.18	★
Shop Worker	[Demo],[Cust...	7,126	45,414	88.47	-10.95	★
Senior Manager	[Demo],[Cust...	2,094	15,192	77.71	-12.14	★
Professional	[Demo],[Cust...	4,490	30,377	83.34	-12.83	★
House Person	[Demo],[Cust...	1,662	12,704	73.76	-13.28	★
Male	[Demo],[Cust...	25,601	157,832	91.45	-20.73	★

Figure 4-27 Examining profile results

Using map analyses

A map analysis supports visualizing data on a geographic map. Analysis results appear as a choropleth map on which each predefined region is assigned a color or shade. Each shade corresponds to the magnitude of data values for that region. For example, a map analysis assigns each region in a country map a different shade of red that corresponds to the number

of unemployed customers in each region. In this example, a region that appears dark red has a high number of unemployed customers.

A map analysis assigns, by default, increasingly darker shades of red to regions having values of correspondingly higher magnitude. Optionally, create a custom set of color shades. When selecting custom colors, consider the following points:

- Darker colors typically represent data items with a higher magnitude.
- To make clear, distinct data ranges, select a distinct color shade for each range. Test map analysis results to verify that the map conveys meaning when viewed in color or in grayscale shades.
- To provide effective contrast between data ranges, limit the number of colors to seven, if possible.

BIRT Analytics provides multiple maps with region names encoded. You can also add new maps. If necessary, you can use a decode field to match the names in the map to the region names in your data.

How to add a measure to a map analysis

- 1 In Analytics—Analysis, choose Map.
- 2 In Map, choose Definition, as shown in Figure 4-28.

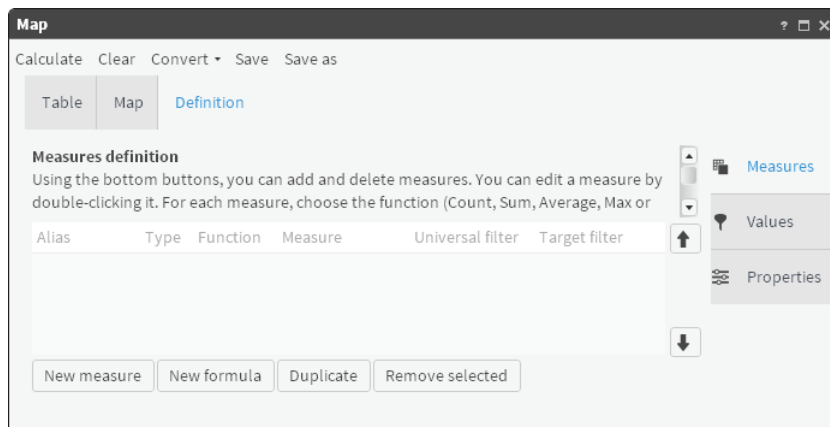


Figure 4-28 Map definition view

- 3 In Measures, choose New measure.
- 4 Type a descriptive alias.
- 5 To define the new measure, drag a table from My Data and drop it as shown in Figure 4-29.

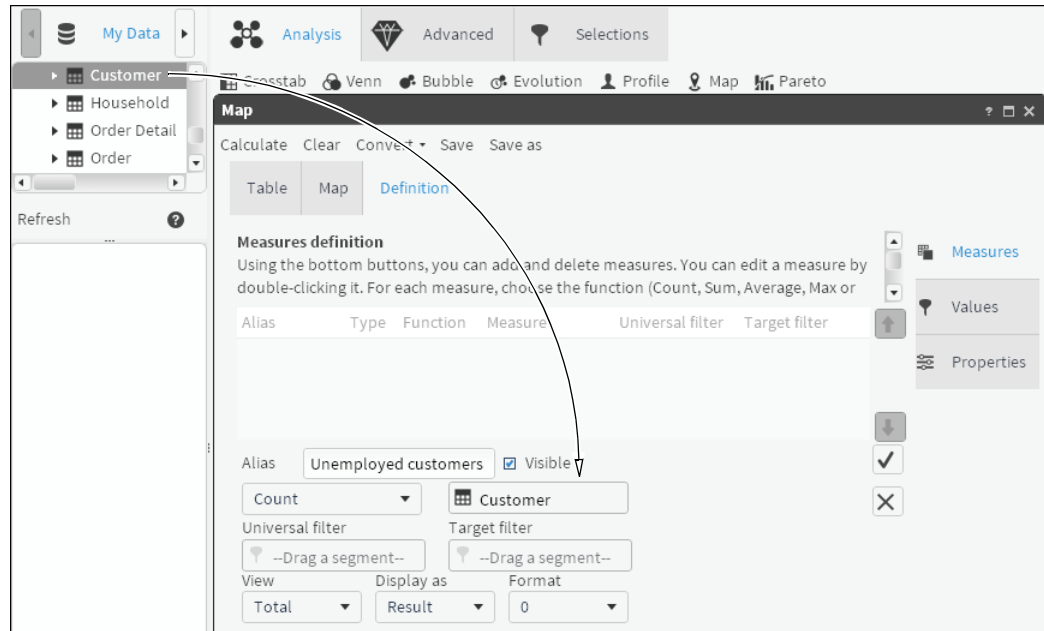


Figure 4-29 Adding a table to the measure definition

- 6 To filter the measure, drag a segment from Discrete Values and drop it in Universal filter, as shown in Figure 4-30.

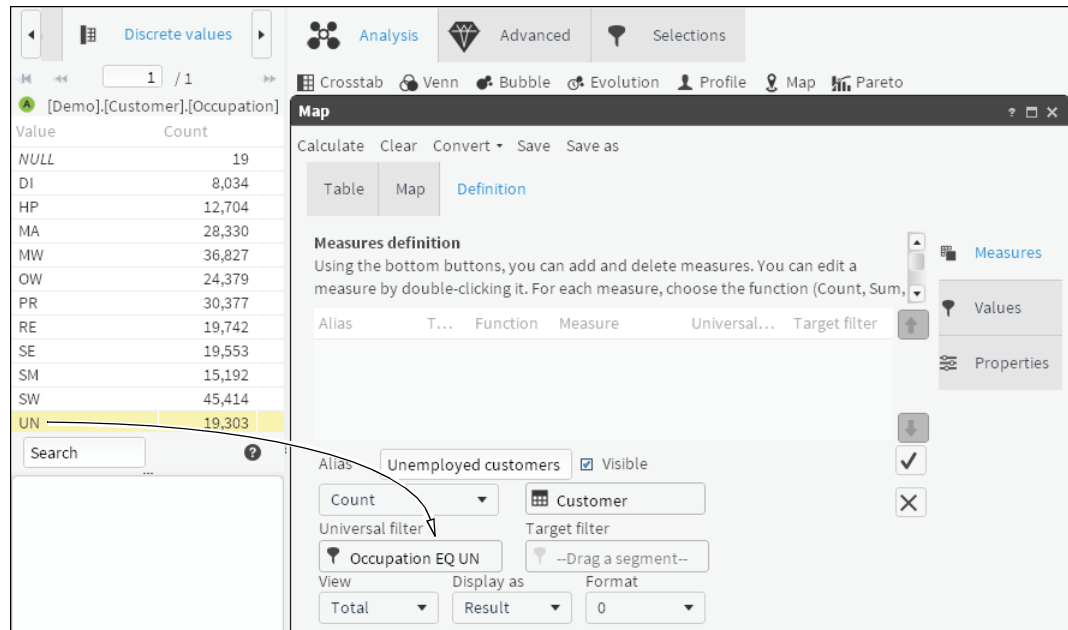


Figure 4-30 Adding a filter to the measure definition

- 7 Choose the check mark. A defined measure appears in Measures definition.

How to generate map analysis results

- 1 Choose Values.
- 2 To link geographic data to a measure, drag a column that contains geographic data from My Data and drop it in Geographical Column, as shown in Figure 4-31.

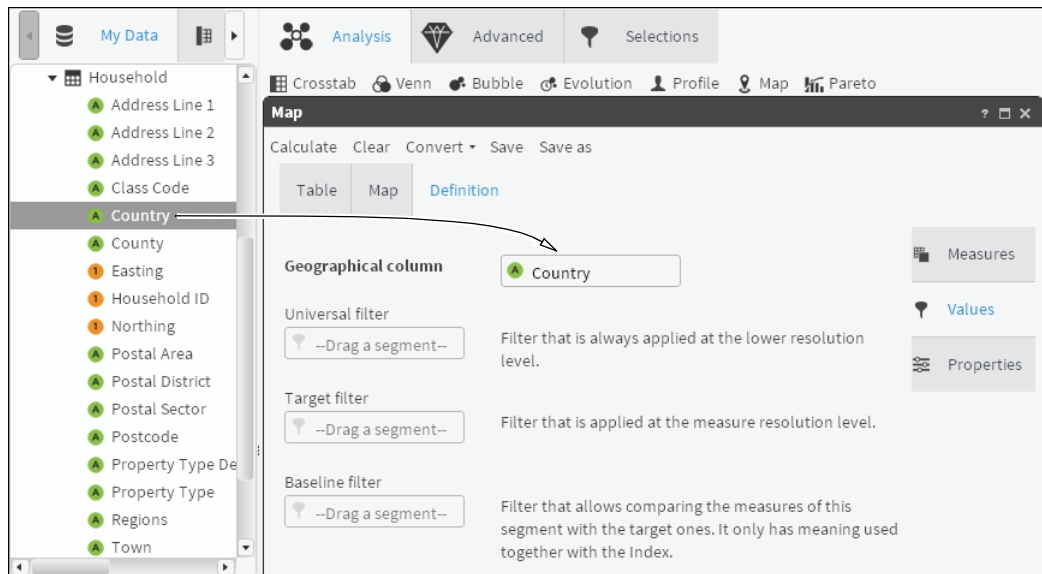


Figure 4-31 Adding geographic data values to a measure definition

- 3 Choose Properties.
- 4 In Select map, select a map name that matches the geographic region defined for the measure. If no available map appears to match the geographic region for your analysis, ask a BIRT Analytics administrator to load the appropriate map.
- 5 Choose Calculate.

How to view map analysis results



- To view graphical results, choose Map, if necessary.
- To view data values that correspond to each color appearing in Map, choose Legend.
- To see numerical results of the map analysis, choose Table.
- To see analysis results appearing in Map on a crosstab, select Convert. Then, choose Crosstab.

Using Pareto analyses

A Pareto analysis represents Pareto’s 80-20 theory with available data. Pareto’s theory states that:

- A minority of the population (approximately 20%) bears 80% of something.
- The remaining majority group (approximately 80%) bears 20% of something.

For example:

- 20% of clients are responsible for 80% of turnover.
- 80% of turnover comes from 20% of the product catalog.

You can use this theory to explore the relationship between a numerical and a categorical variable. For example, you can analyze the accrued benefit of customers (continuous numerical variable) through the grouping of customer's accrued benefit into deciles (categorical variable).

How to create a Pareto analysis

- 1 In Analytics—Analysis, choose Pareto.
- 2 In Pareto, choose Advanced, as shown in Figure 4-32.

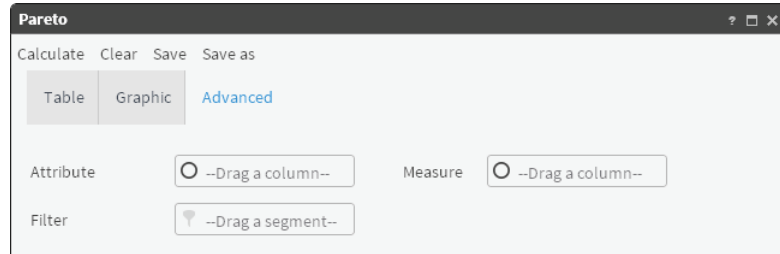


Figure 4-32 Pareto—Advanced

- 3 Drag a categorical field from Data Tree to Attribute.
- 4 Drag a related numeric field from Data Tree to Measure. The resulting entries are shown in Figure 4-33.

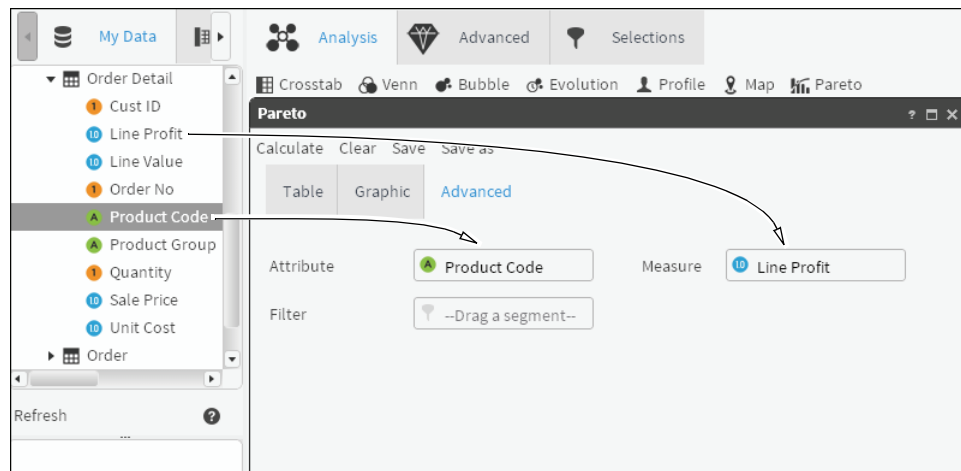


Figure 4-33 Dragging an attribute and a measure to Pareto

- 5 Add a filter to limit the results, if necessary. Drag a segment from Discrete Values or Scratchpad to Filter, as shown in Figure 4-34.

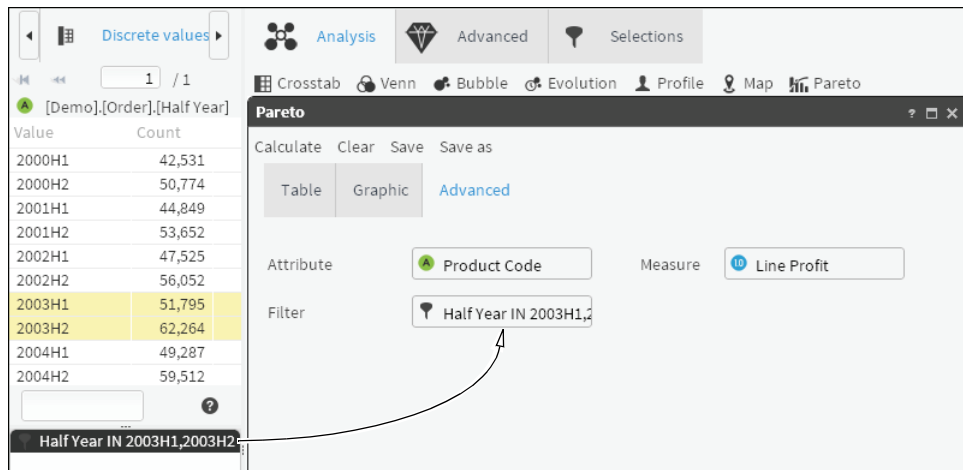


Figure 4-34 Dragging a filter to Pareto

- 6 Choose Calculate.
- 7 Examine the results on Graphic.

For example, select a numeric field as a measure, and a quantile rank of the same field as an attribute. This quantile groups its values into n equal groups. The analysis displays a Pareto curve. You can see if the Pareto analysis satisfies Pareto’s theory by looking at the growth curve. The graph shows cumulative percentages.

Table shows, sorted by the amount, the number of records, both amount and percentages, cumulative and cumulative percentages.

Double-clicking Count (record number), the records appear in Scratchpad. By choosing the Cumulative Count value, you can display the record and all previous records.

5

Visualizing your data

This chapter contains:

- Using the Gallery
- Working with a Canvas

Visualizing your data

BIRT Analytics provides many ways to visualize data. By defining data measures, refining their properties, and assigning a visual indicator to display the data measures, you use BIRT Analytics to visually compare and contrast data values.

Using the Gallery

Gallery supports creating visual indicators that display data values as graphic visualizations. For each indicator, you define one or multiple data measures. All BIRT Analytics indicators provide the following two features:

- Graphic displays a visual preview of the indicator.
- Definition supports defining data measures, values, and properties that determine how data, shapes, and colors appear on the indicator.

To view an indicator on Graphic, you must create valid measures using Definition. This section describes how to visualize data in the following ways:

- Using a dial, meter, label, sphere, cylinder, and funnel indicator
- Converting data measures for one indicator to another

Using a dial

A dial indicator displays data values using a needle that points to data values and range boundaries appearing on a semi-circular dial. To compare values with defined bounds, define data measures, values, and properties for a dial using the following features on Definition:

- Measure: Add, edit, duplicate, or delete measures and formulas.
- Values: Assign maximum, mean, and minimum values to a dial by selecting an appropriate measure. To indicate multiple values, use Add dial to create a new needle, then set alias, color, and size properties for each new needle.
- Properties: Choose values to appear as boundaries for defined ranges, and assign a unique color or shade to each range.

You can convert a dial into a meter, sphere, or cylinder indicator.

How to create measures to display on a dial indicator

- 1 In Gallery—Dial, choose Definition.
- 2 To create a measure for the dial, choose New measure.
- 3 In Alias, type a name for the new measure.
- 4 Select Maximum as the function for the new measure. From My Data, drag a column on which to base the measure and drop it next to Maximum, as shown in Figure 5-1.

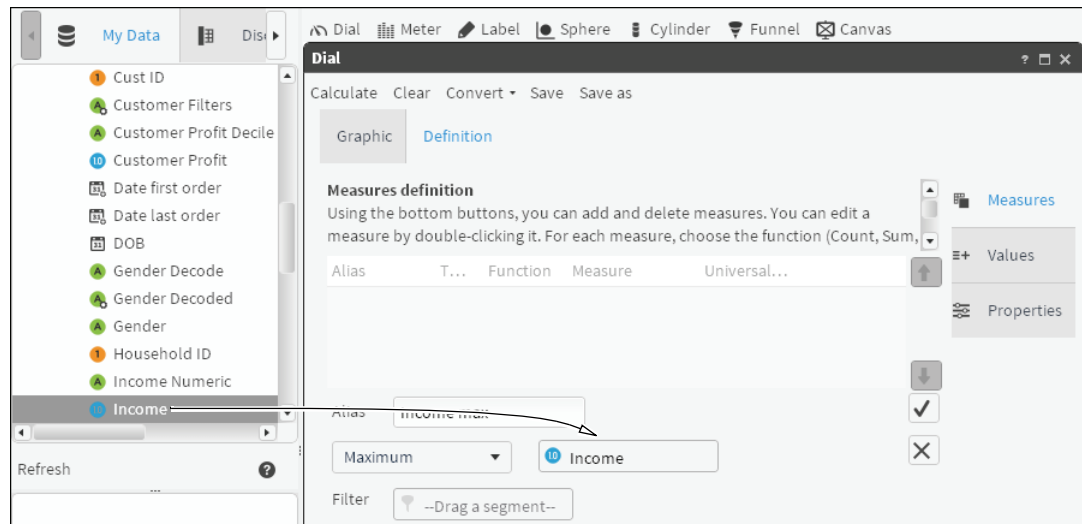


Figure 5-1 Defining a new measure for a dial

5 To create a filter, double-click a column in My Data. Drag a field from Discrete Values and drop it in Filter, as shown in Figure 5-2.

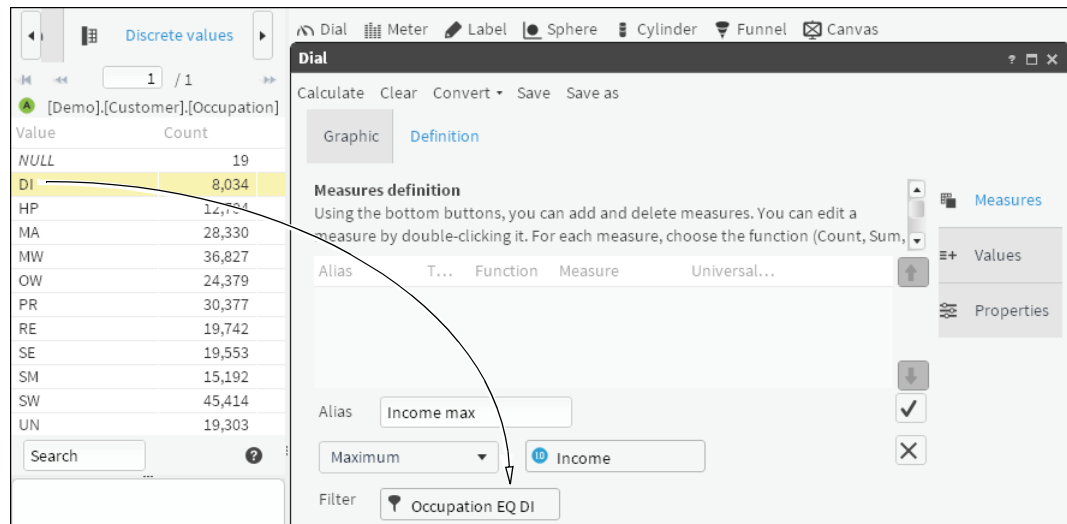


Figure 5-2 Adding a filter to a measure

6 Choose the check mark to create the new measure.

7 To create another measure with properties similar to the first, select a measure and choose Duplicate.

8 Type a name and change the function to Minimum, as shown in Figure 5-3.

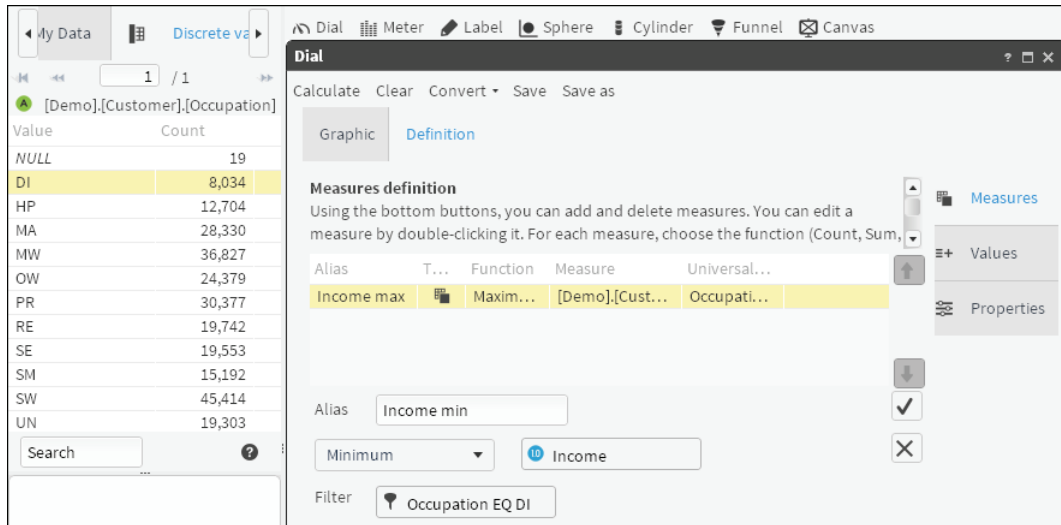


Figure 5-3 Creating a new measure based on a measure having similar properties

- 9 Choose the check mark to create the new measure.
- 10 Repeat steps 7–9, changing the function to Mean.

How to add formulas and thresholds to a dial indicator

- 1 Choose New formula to create a formula.
- 2 In Alias, type a name for the formula, and type a formula for a lower threshold in the dialog box below Alias. In this example, the formula is a value.
- 3 Choose Validate. At the message that the formula definition is valid, shown in Figure 5-4, choose OK.

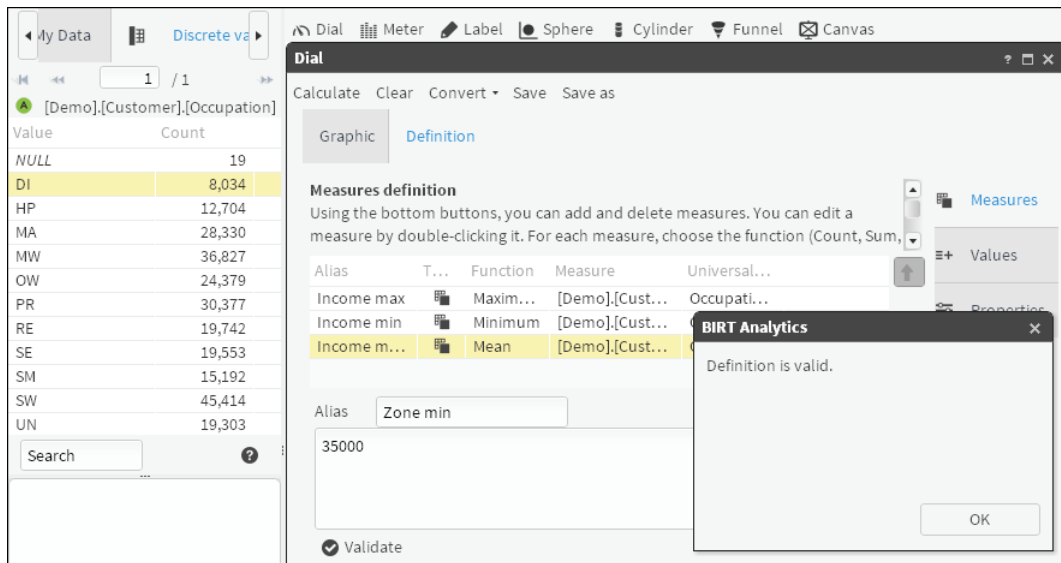


Figure 5-4 Validating the definition for a formula

- 4 To create the new formula, choose the check mark.
- 5 Repeat steps 1–3, entering a formula for a higher threshold.
- 6 Choose Values.

- For each value appearing on the dial, choose an appropriate measure from the list. For example, choose measures that correspond to minimum, mean, and maximum values, as shown in Figure 5-5.

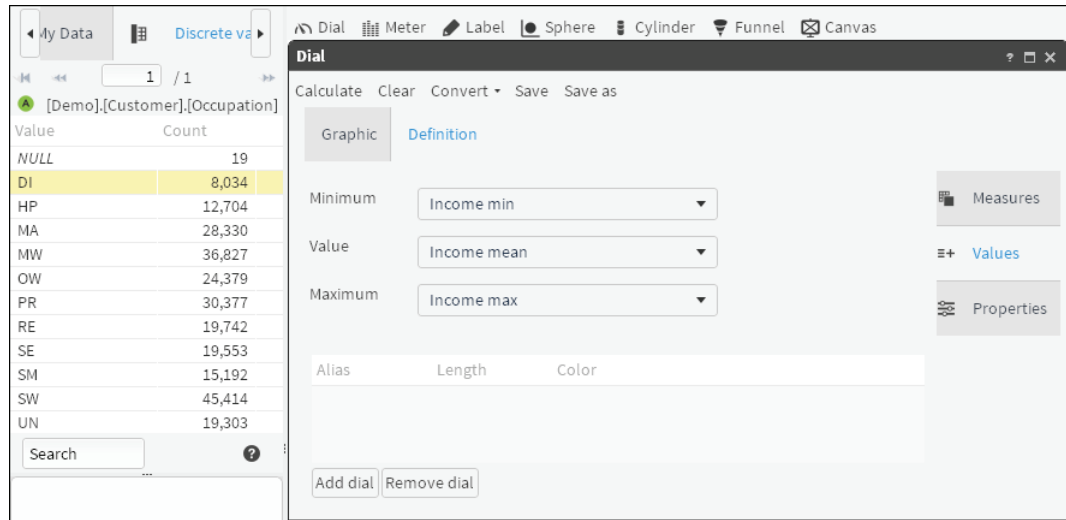


Figure 5-5 Assigning measures to values that appear on a dial

How to add properties to the dial indicator

- Choose Properties.
- Choose Add range.
- Choose a measure or formula to indicate the minimum and maximum thresholds for the range, as shown in Figure 5-6.
- In Color, open the palette and select a color for the range. Choose OK.

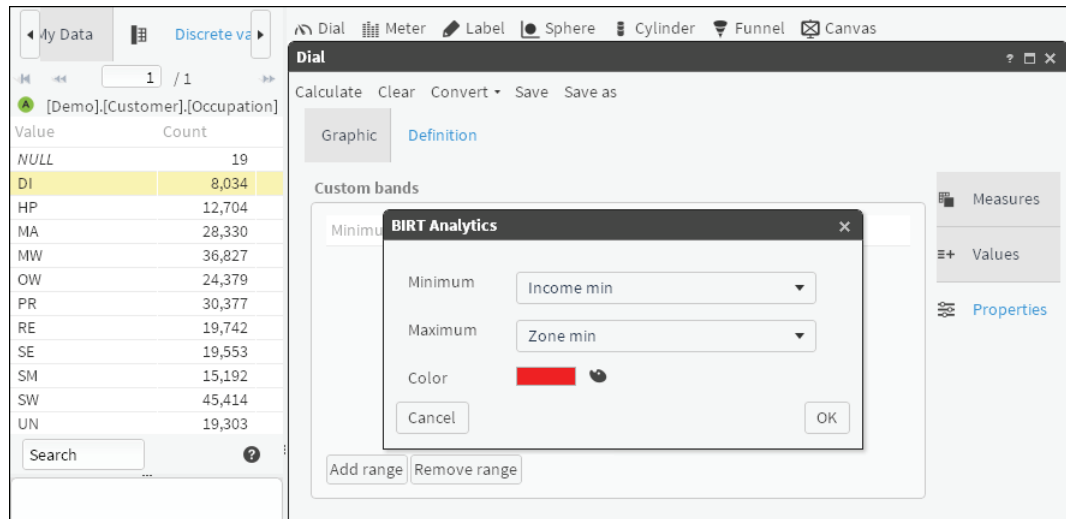


Figure 5-6 Configuring properties for a dial

- Repeat steps 2–4 until you have created three ranges.
- Choose OK.
- Choose Calculate to view the dial. To see the average value, hover the cursor over the needle, as shown in Figure 5-7.

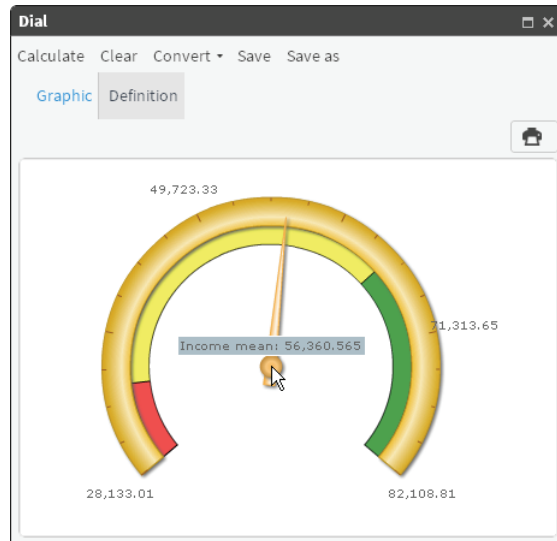


Figure 5-7 Viewing the value of an average on a dial

How to configure a dial so that min, max, and average values for two similar groups appear

- 1 In Data Tree, choose My folders.
- 2 Locate and open a previously saved dial. Choose Calculate. The dial appears on Graphic, as shown in Figure 5-8.

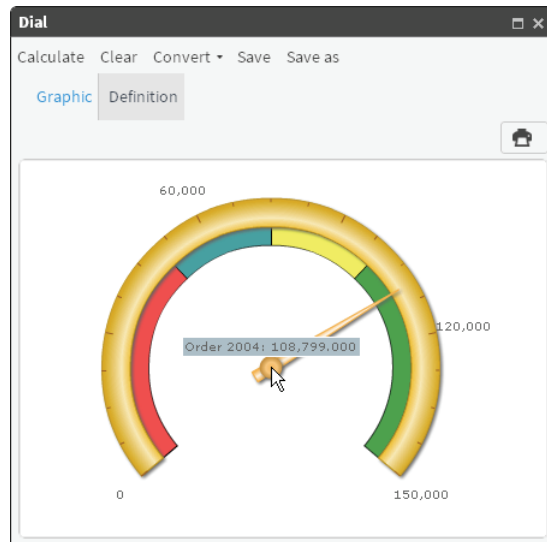


Figure 5-8 Calculating results for a dial

- 3 Choose Definition.
- 4 Select a measure and choose Duplicate.
- 5 In Alias, type a name, then replace the filter. To replace a filter, drag a segment, drop it in Filter, as shown in Figure 5-9, then choose Replace.

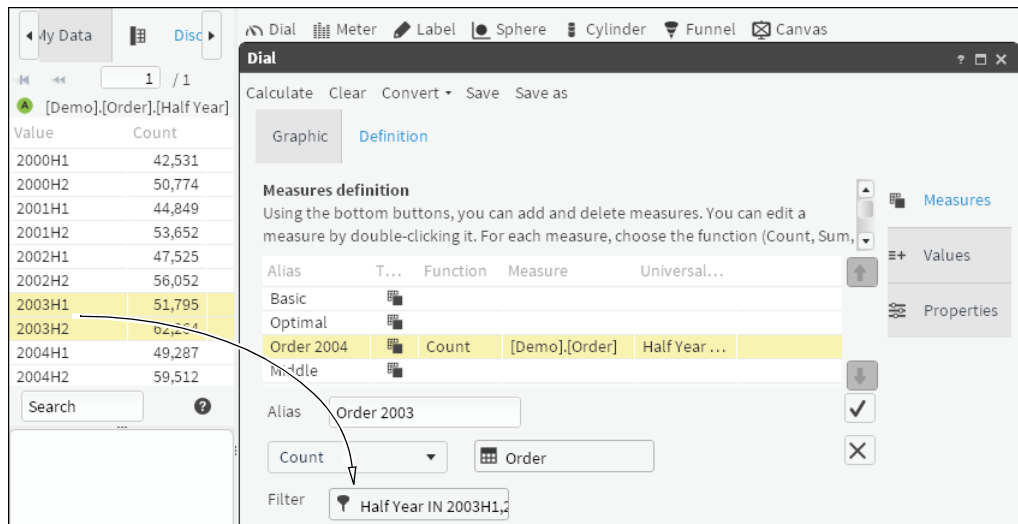


Figure 5-9 Replacing the filter in a duplicated measure

- 6 Choose the check mark to create the new measure.
- 7 Choose Values.
- 8 Choose Add dial.
- 9 In Alias, select the measure shown to open a pull-down menu and choose the newly created measure, as shown in Figure 5-10.

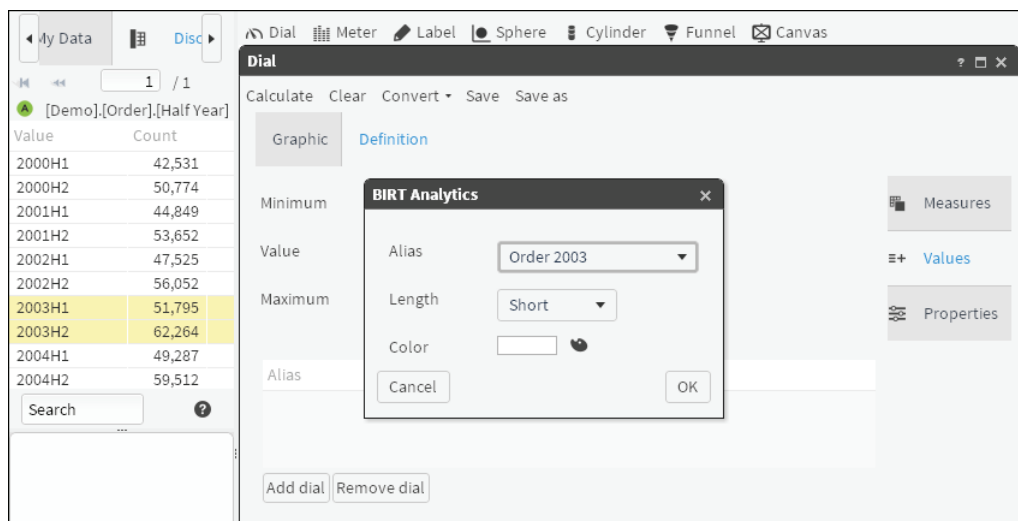


Figure 5-10 Adding a new measure to a dial

- 10 In Color, open the palette and select a color for the range. Choose OK.
- 11 Choose Calculate to view the dial. The dial now represents two similar groups. Hover the cursor over each needle to see and compare values represented by each needle, as shown in Figure 5-11.

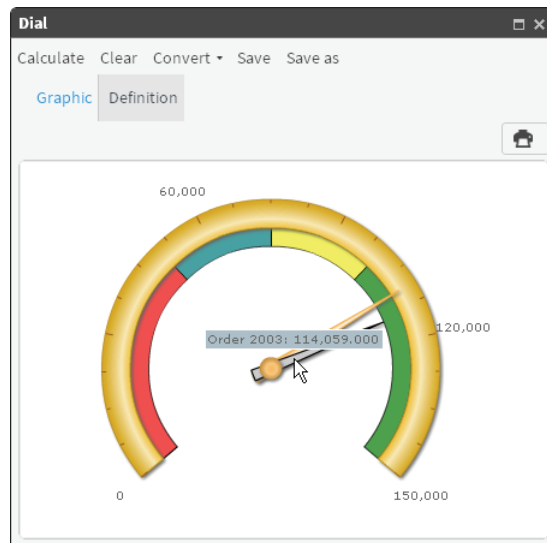


Figure 5-11 Viewing a dial representing values from two similar groups

Converting data measures to another indicator type

Data measures that you define for one BIRT Analytics indicator may appear on another indicator. Convert lists indicators for which a calculated indicator supports conversion.

How to convert a dial to a meter

- 1 In Data Tree, select My Folders.
- 2 Locate and open a previously saved dial. Choose Calculate. The dial appears on Graphic, as shown in Figure 5-12.

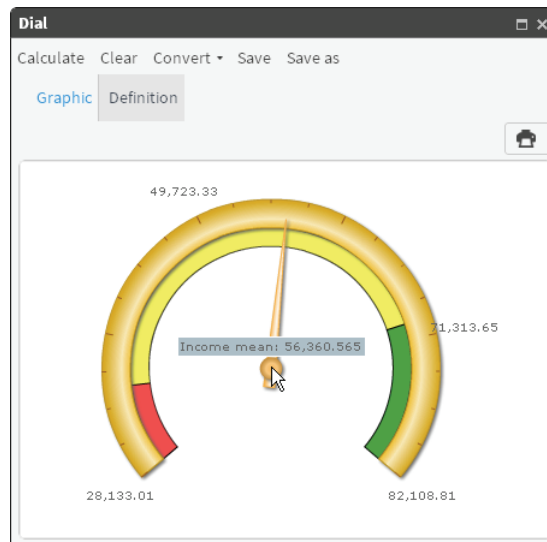


Figure 5-12 Visualizing a data measure using a dial

- 3 Select Convert, then choose Meter.
- 4 Choose Definition, then view the Measures, Values, and Properties tabs.
- 5 Choose Calculate. The new meter appears on Graphic, as shown in Figure 5-13.

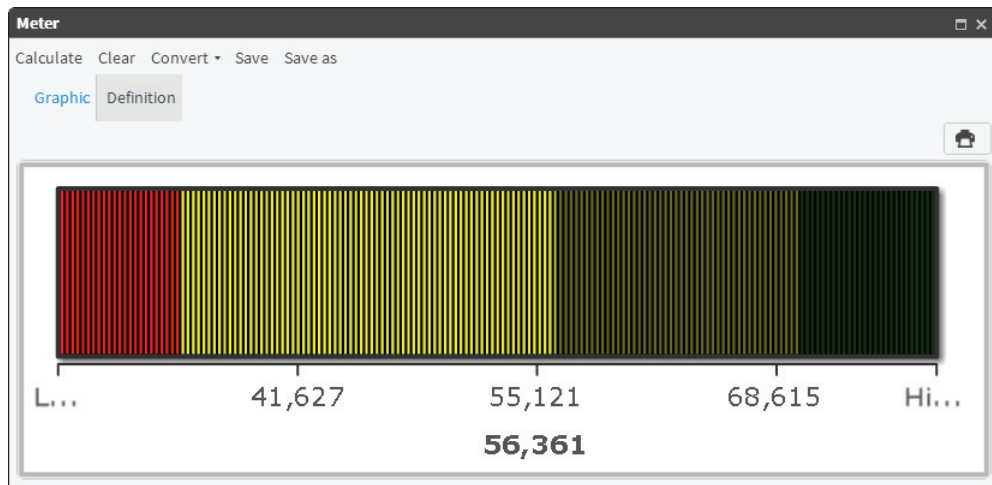


Figure 5-13 Visualizing a data measure using a meter

How to convert a dial graph to a sphere

- 1 For a calculated dial, select Convert, then choose Sphere.
- 2 To view Measures, Values, and Properties, choose Definition.
- 3 Choose Calculate. The new sphere appears on Graphic, as shown in Figure 5-14.

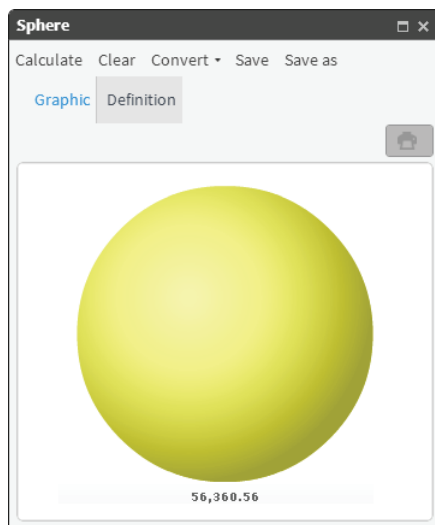


Figure 5-14 Visualizing a data measure using a sphere

How to convert a dial graph to a cylinder

- 1 For a calculated dial, select Convert, and choose Cylinder.
- 2 To view Measures, Values, and Properties, choose Definition.
- 3 Choose Properties.
- 4 Using the palette, choose a color, as shown in Figure 5-15.

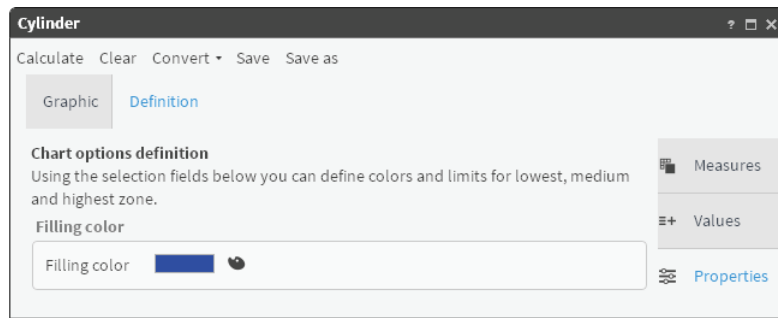


Figure 5-15 Formatting properties of a data visualization gadget

- 5 Choose Calculate. The cylinder appears on Graphic, as shown in Figure 5-16.

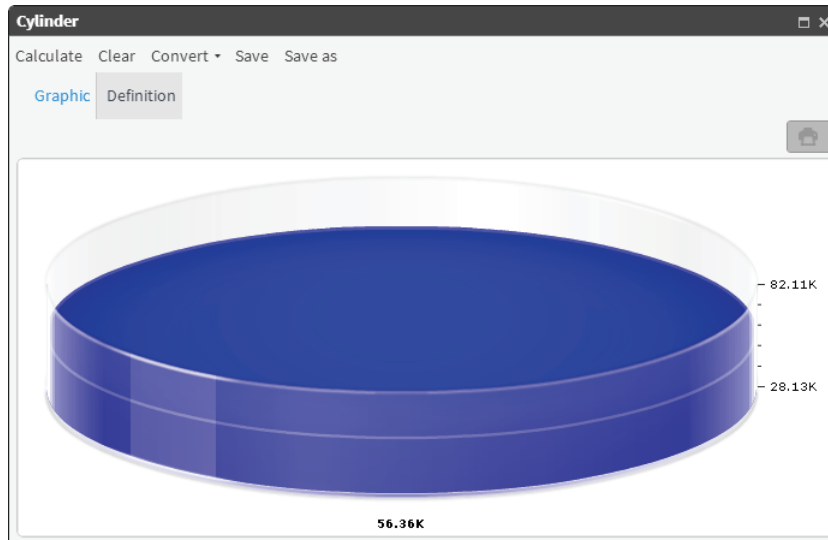


Figure 5-16 Visualizing a data measure using a cylinder

Using a meter

A meter is an indicator that uses a color bar display to show if the data are within established parameters. The following tabs are in the Definition tab:

- Measure tab: Use the buttons to add or delete measures and formulas. To edit an existing measure, double-click it.
- Values tab: Type maximum and minimum values and the value to be represented by measures.
- Properties tab: Type the values to be used as the limits of the areas to be represented. Choose the color.

You can convert a meter into a dial, sphere, or cylinder indicator.

How to create measures to display on a meter

- 1 In Gallery—Meter, choose Definition.
- 2 To create a measure for the meter, choose New measure.
- 3 In Alias, type a name for the new measure.
- 4 Select Maximum as the function for the new measure. From My Data, drag a column on which to base the measure and drop it next to Maximum, as shown in Figure 5-17.

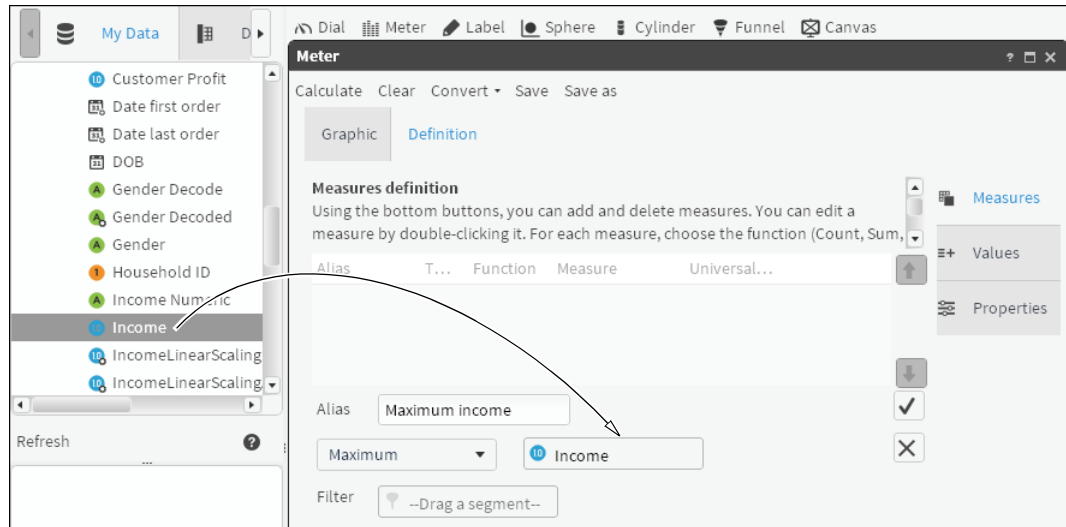


Figure 5-17 Defining a new measure for a meter

5 To create a filter, double-click a column in My Data. Drag a field from Discrete Values, and drop it in Filter, as shown in Figure 5-18.

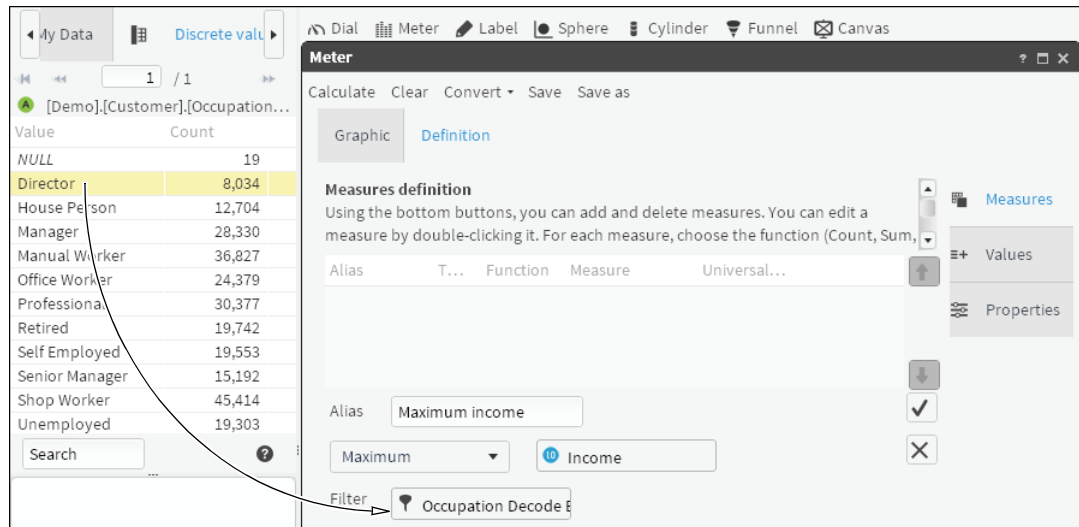


Figure 5-18 Adding a filter to a measure

6 To create the new measure, choose the check mark.

7 To create another measure with properties similar to the first, select a measure, and choose Duplicate.

8 In Alias, type a name and change the function to Minimum, as shown in Figure 5-19.

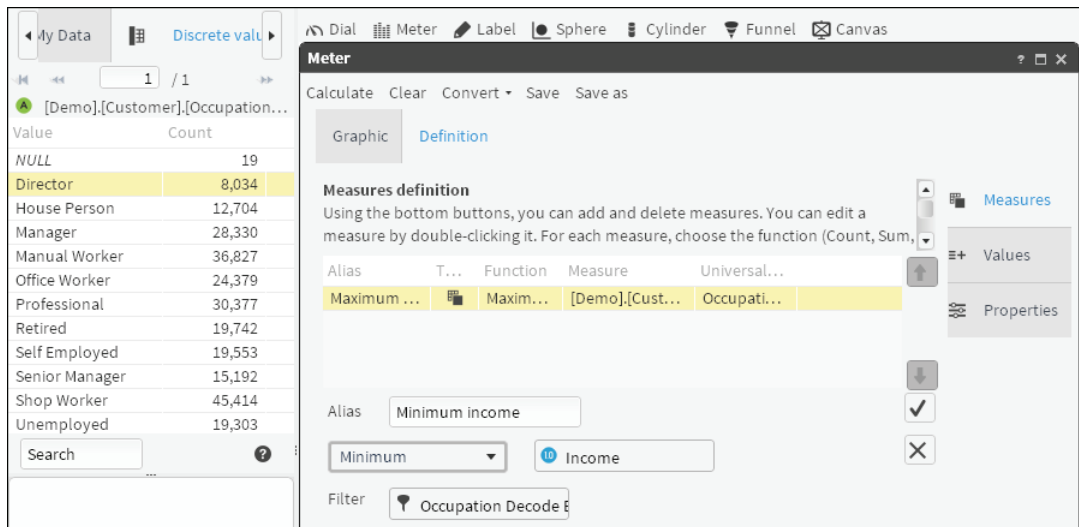


Figure 5-19 Creating a new measure based on a measure having similar properties

- 9 To create the new measure, choose the check mark.
- 10 Repeat steps 7–9, changing the function to Mean.

How to add formulas and thresholds to a meter

- 1 To create a formula, choose New formula.
- 2 In Alias, type a name for the formula, then type a formula for a lower threshold in the dialog box below Alias. In this example, the formula is a value.
- 3 Choose Validate. At the message that the formula definition is valid, shown in Figure 5-20, choose OK.

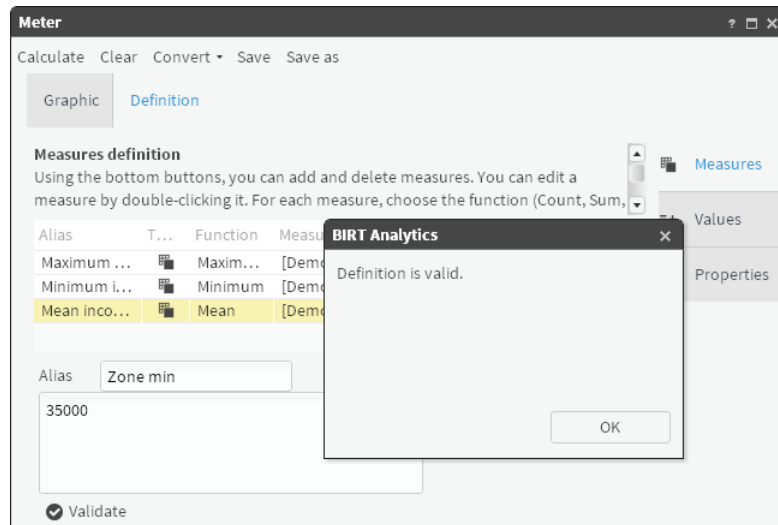


Figure 5-20 Validating the definition for a formula

- 4 To create the new formula, choose the check mark.
- 5 Repeat steps 1–3, entering a formula for a higher threshold.
- 6 Choose Values.
- 7 For each value to appear on the meter, select an appropriate measure from the list. For example, select measures that correspond to minimum, mean, and maximum income values, as shown in Figure 5-21.

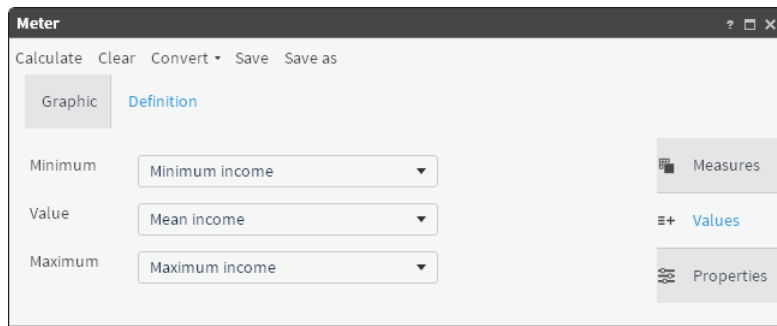


Figure 5-21 Assigning measures to values that will appear on a meter

How to add range properties to a meter

- 1 Choose Properties.
- 2 Choose Add range.
- 3 To indicate the minimum and maximum thresholds for the range, choose a measure or formula, as shown in Figure 5-22.
- 4 In Color, open the palette and select a color for the range. Choose OK.

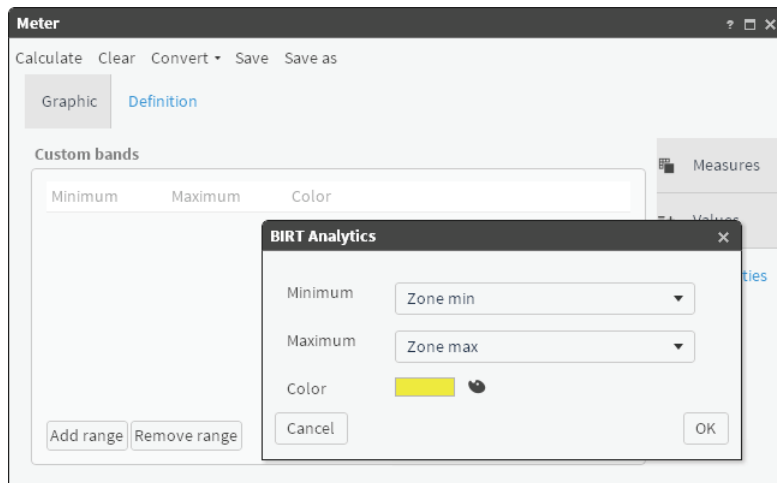


Figure 5-22 Configuring properties for a meter

- 5 Repeat steps 2–4 until you have created three ranges.

- 6 Choose Calculate to view the meter. Different color shades indicate different ranges, as shown in Figure 5-23.

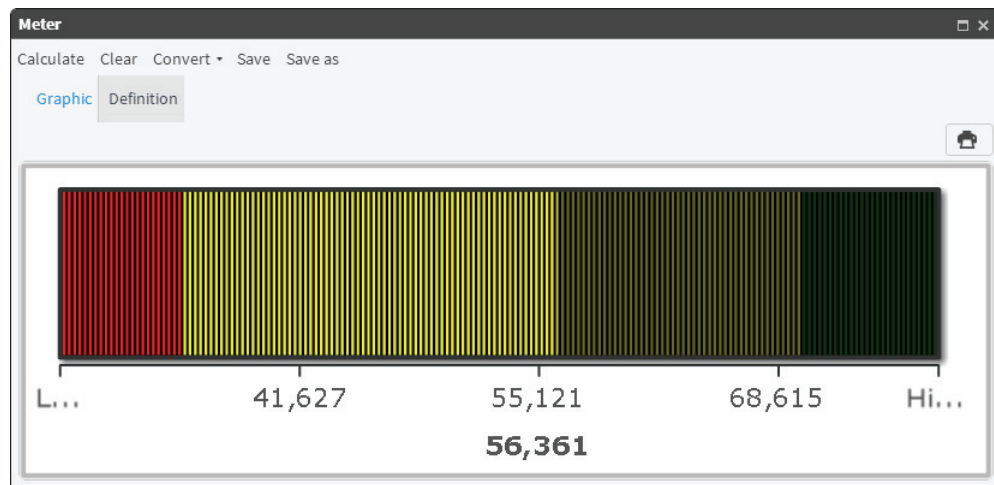


Figure 5-23 Viewing data measures on a meter

Using a label

A label indicator displays the values of defined measures as text. Use labels to create comments that can be included in the canvas.

How to create a measure for a label

- 1 In Gallery—Label, choose Definition—Measures.
- 2 Choose New measure.

- 3 In Alias, type a measure name.
- 4 From the list of function names, select a function name.
- 5 From My Data, drag an item on which the selected function will perform and drop it next to the selected function name. For Count, you must drag a table. For all other functions, drag a column.
- 6 From Discrete Values or Scratchpad, drag a segment and drop it in Filter, as shown in Figure 5-24.

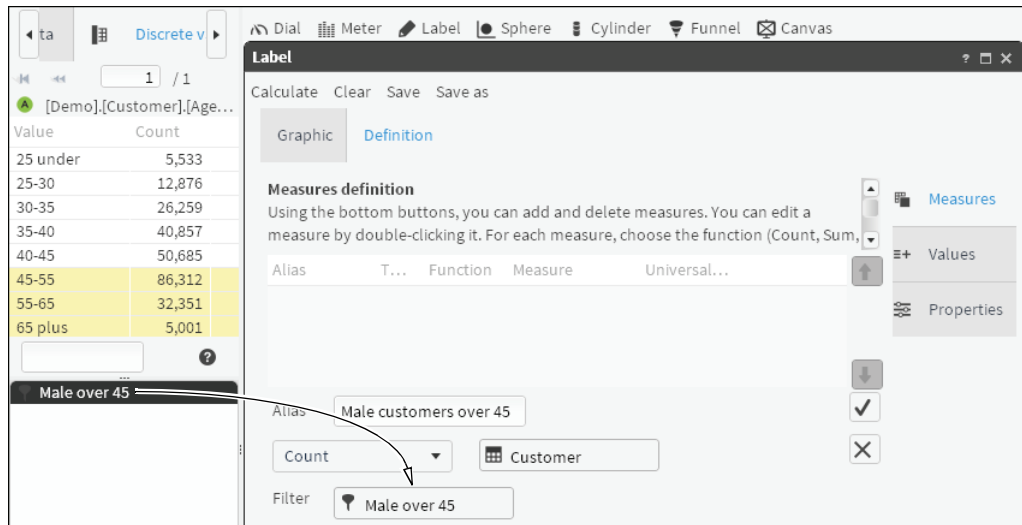


Figure 5-24 Adding a segment from Scratchpad as a filter for a label

- 7 Choose the check mark.

How to create a new, duplicate measure

- 1 Select an existing measure.
- 2 Choose Duplicate.
- 3 In Alias, type a unique name for the new measure.
- 4 Replace the function, column, or filter.
- 5 Choose the check mark.

How to define a label

- 1 In Label—Definition, choose Values.
- 2 In Text to show, type text that provides context for the label.
- 3 Right-click, then select a measure. The selected measure appears at the cursor location. The example in Figure 5-25 shows selecting a third measure for a text label.

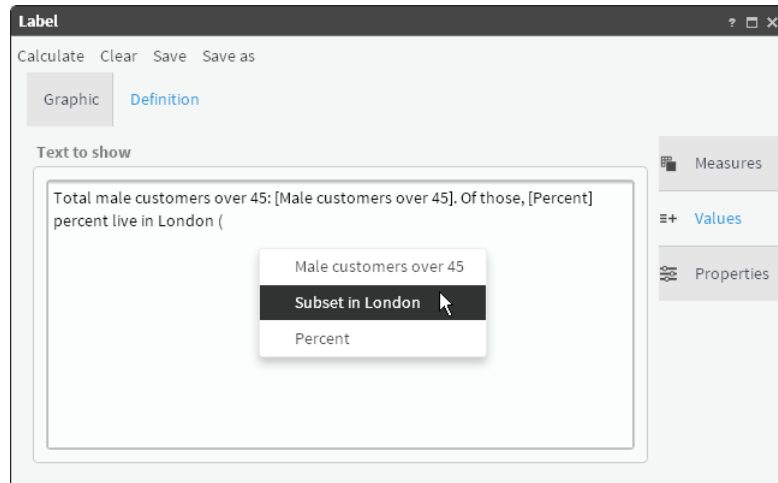


Figure 5-25 Selecting a measure for a label

- 4 Choose Properties. Then, select color, size, and font properties for the label text, as shown in Figure 5-26.

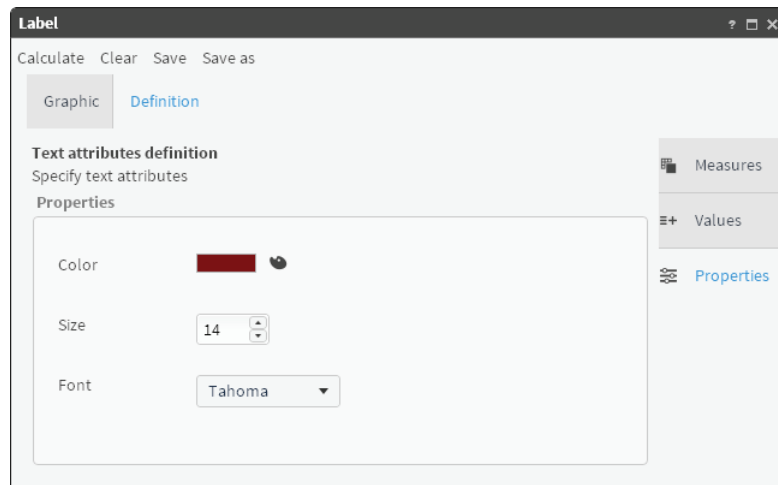


Figure 5-26 Selecting text properties for a label

- 5 To create the label, choose Calculate. The label appears in Graphic, as shown in Figure 5-27.

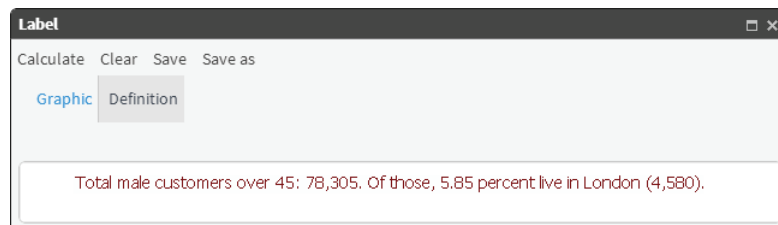


Figure 5-27 Previewing a complete label

- 6 To keep the label for future use, choose Save.

Using a sphere

A sphere displays data graphically using a round, three-dimensional shape that changes color based on the limit area of the value it represents. The following tabs are in the Definition tab:

- Measure tab: Use the buttons to add or delete measures and formulas. To edit an existing measure, double-click it.
- Values tab: Type the maximum and minimum values and the value to be represented by the measures.
- Properties tab: Select the values to be used as the limits of the areas to be represented. Choose the color.

You can convert a sphere into a dial, meter, or cylinder indicator.

How to add measures to a new sphere

- 1 In Gallery—Sphere, choose Definition.
- 2 Choose New measure.
- 3 Type a name for the new measure.
- 4 Select a function for the sphere. From Data Tree, drag a column on which to base the measure, as shown in Figure 5-28.

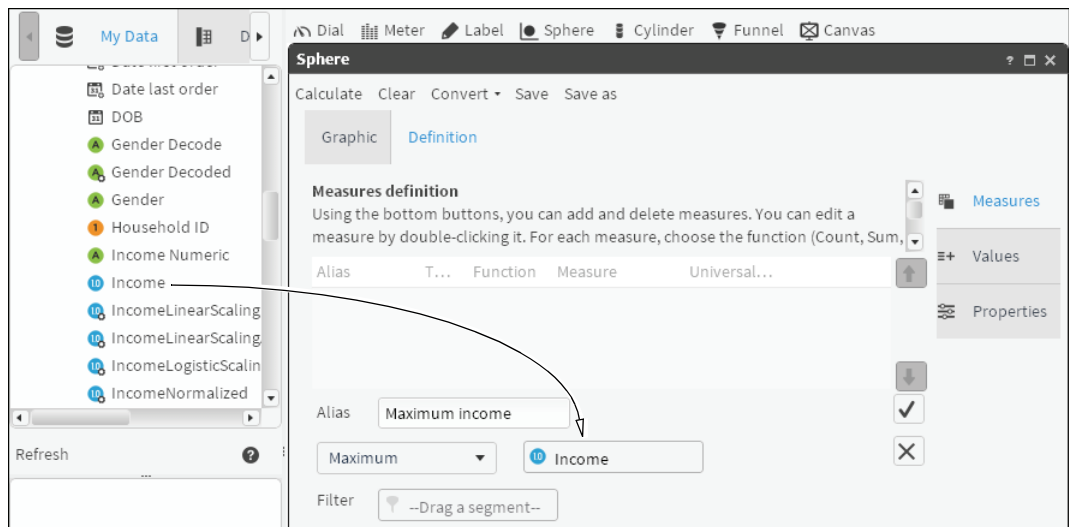


Figure 5-28 Defining a measure for a sphere

- 5 To create a filter, double-click one of the columns in Data Tree. In Discrete Values, drag a field and drop it in Filter, as shown in Figure 5-29.

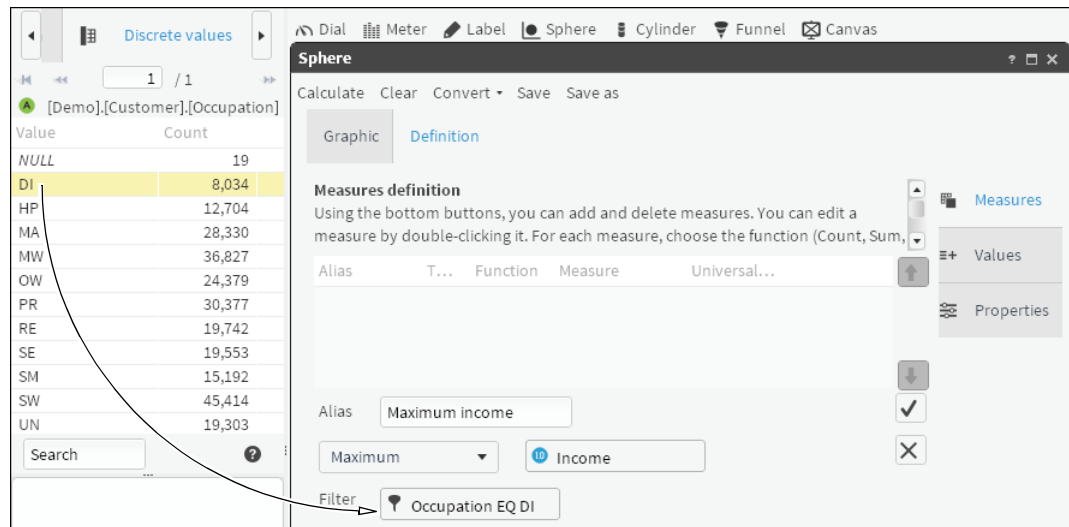


Figure 5-29 Adding a filter to a measure

- 6 To create the measure, choose the check mark.
- 7 To create another measure having properties similar to an existing one, select a measure and choose Duplicate.
- 8 Type a measure name and change the function to Minimum, as shown in Figure 5-30.

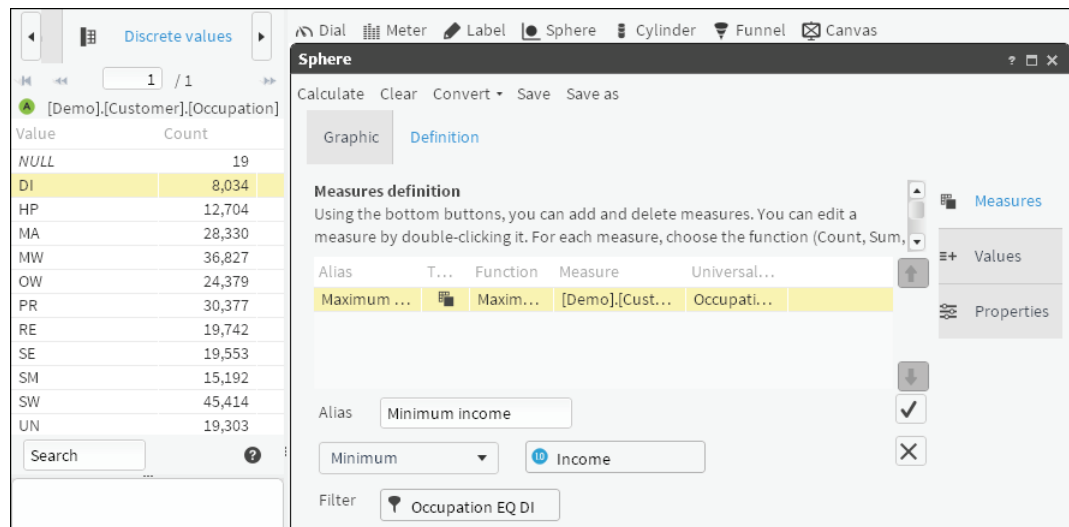


Figure 5-30 Creating a new measure having properties similar to an existing measure

9 To create the new measure, choose the check mark.

10 Repeat steps 7–9, changing the function to Mean.

How to add a formula to a sphere

1 Choose New formula.

2 Type a name for the formula, and enter a formula for a lower threshold in the dialog box below.

3 Choose Validate to ensure the function is valid. Choose OK.

4 To create the new function, choose the check mark.

5 Repeat steps 1–4, entering a formula for a higher threshold.

How to create ranges to display in a sphere

1 Choose Values.

2 Using the pull-down menus, select measures to indicate minimum, average, and maximum values to be used in the sphere, as shown in Figure 5-31.

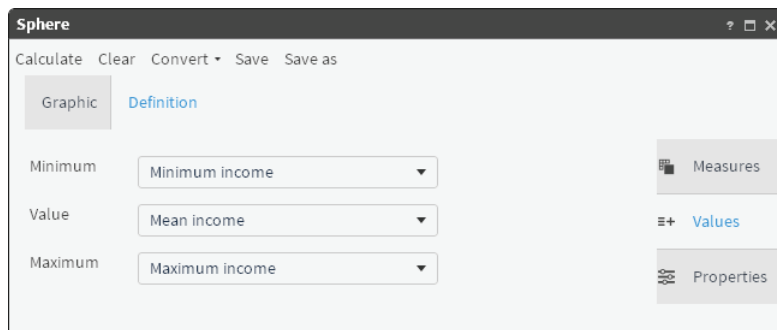


Figure 5-31 Assigning measures to values that appear in a sphere

3 Choose Properties.

4 Choose Add range.

5 Using the pull-down menus, select measures to indicate a minimum and maximum value to be used, as shown in Figure 5-32.

6 In Color, use the palette to apply a color to each range you create.

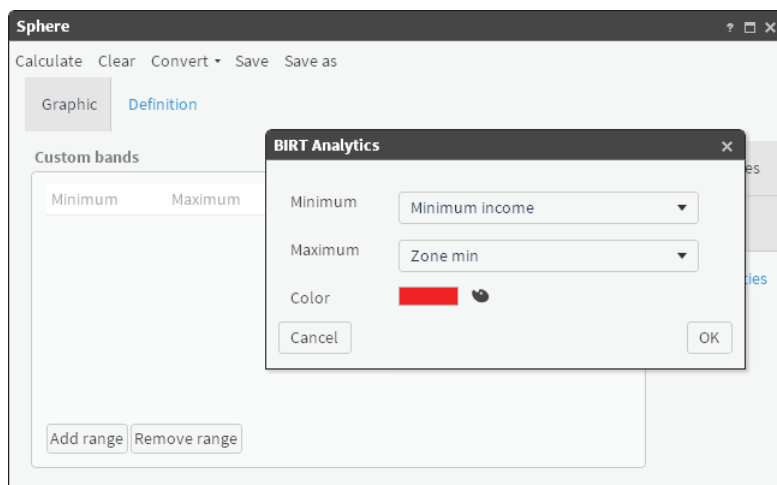


Figure 5-32 Configuring properties for a sphere

7 Repeat steps 4–6, until you create three ranges.

- 8 Choose OK.
- 9 Choose Calculate. The sphere appears on Graphic, as shown in Figure 5-33.

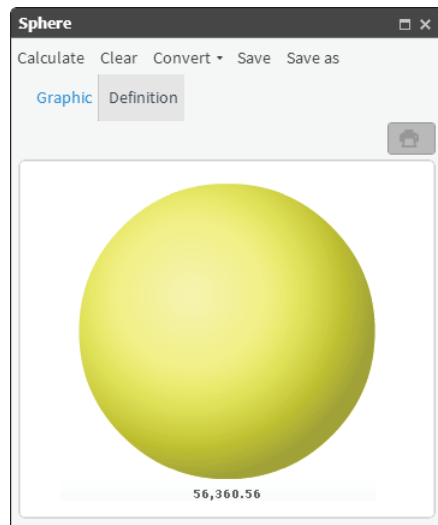


Figure 5-33 Visualizing a data measure using a sphere

Using a cylinder

A cylinder is an indicator that shows the volume of a cylinder depending on the value it represents. The following tabs are in the Definition tab:

- Measure tab: Use the buttons to add or delete measures and formulas. To edit an existing measure, double-click it.
- Values tab: Type the upper and lower limits of the indicator. Choose the measure to be represented.
- Properties tab: Choose the color of the cylinder.

You can convert a cylinder into a dial, meter, or sphere indicator.

How to add measures to a new cylinder

- 1 In Gallery—Cylinder, choose Definition.
- 2 Choose New measure.
- 3 Type a name for the new measure.
- 4 Select a function for the cylinder. From Data Tree, drag a column on which to base the measure, as shown in Figure 5-34.

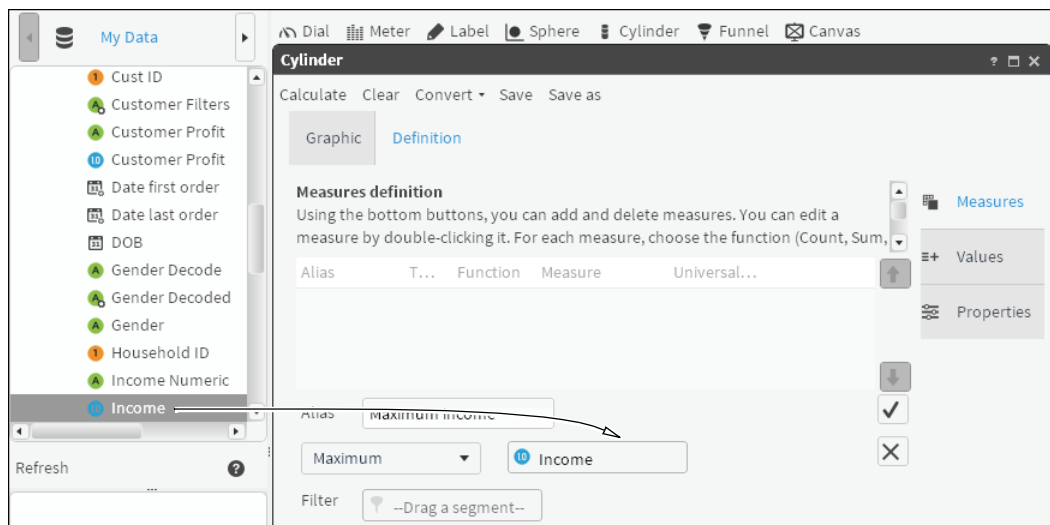


Figure 5-34 Defining a measure for a cylinder

- 5 To create a filter, double-click one of the columns in Data Tree. In Discrete Values, drag a field and drop it in Filter, as shown in Figure 5-35.

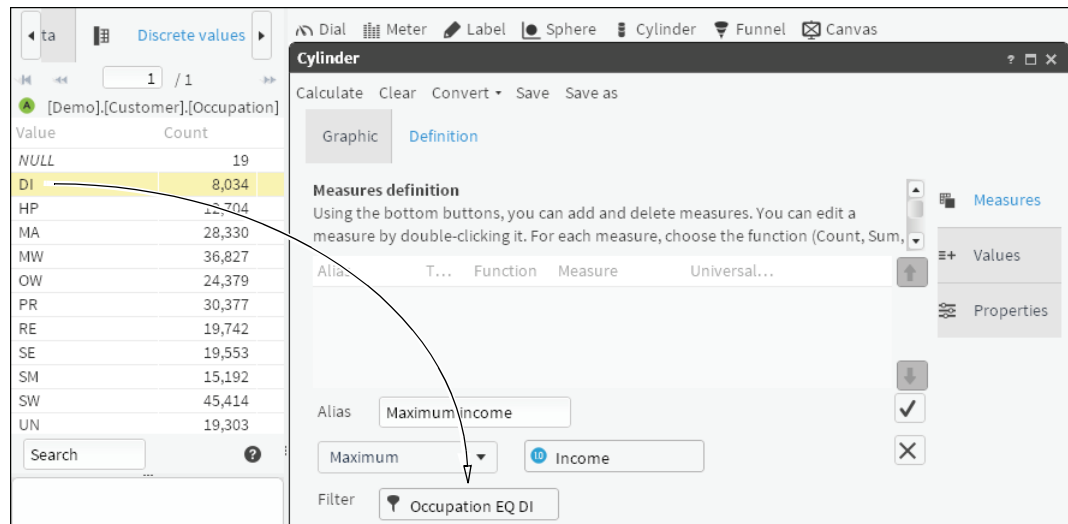


Figure 5-35 Adding a filter to a measure

- 6 To create the measure, choose the check mark.
- 7 To create another measure with properties similar to an existing one, select a measure and choose Duplicate.
- 8 Type a measure name and change the function to Minimum, as shown in Figure 5-36.

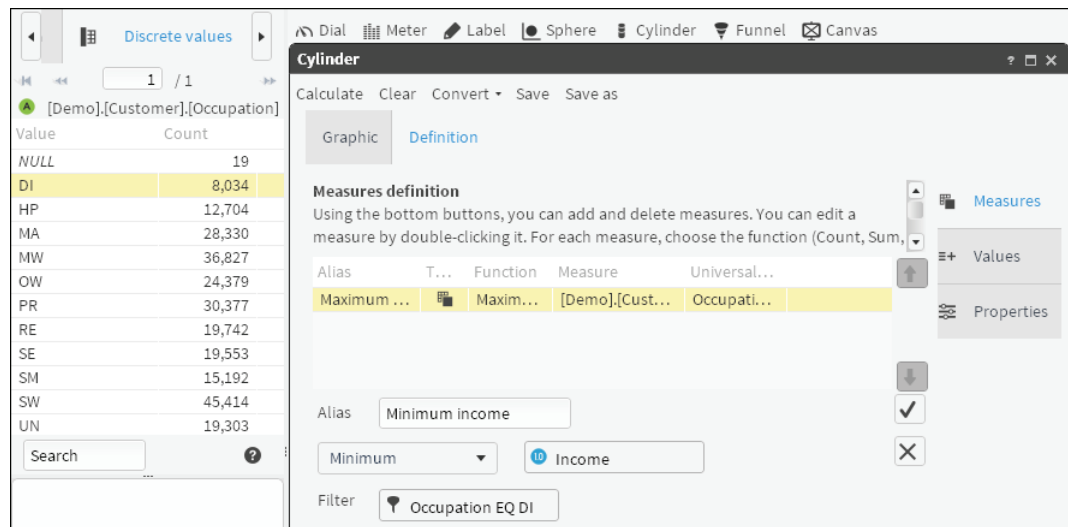


Figure 5-36 Editing the definition of a duplicate measure

- 9 To create the new measure, choose the check mark.
- 10 Repeat steps 7–9, changing the function to Mean.

How to add values and properties to the cylinder

- 1 Choose Values.
- 2 Using the pull-down menus, choose measures that indicate minimum, average, and maximum values to appear on the cylinder, as shown in Figure 5-37.

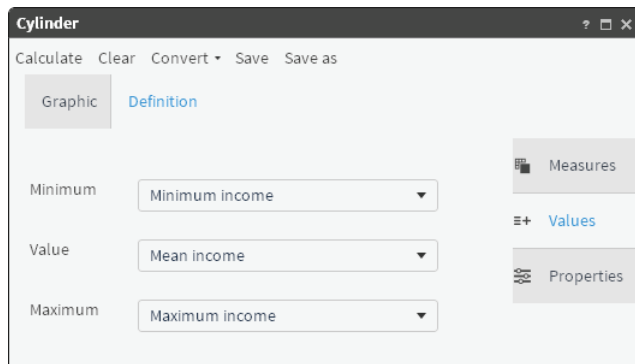


Figure 5-37 Assigning measures to values that appear on a cylinder

- 3 Choose Properties.
- 4 In Filling color, use the palette to apply a color to the range you created, as shown in Figure 5-38.

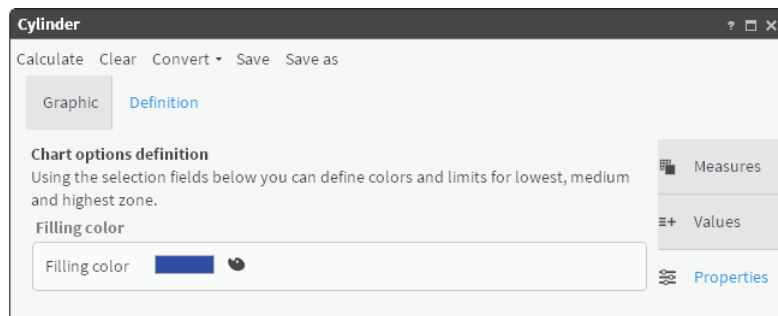


Figure 5-38 Configuring properties for a cylinder

- 5 Choose Calculate. The cylinder appears on Graphic, as shown in Figure 5-39.

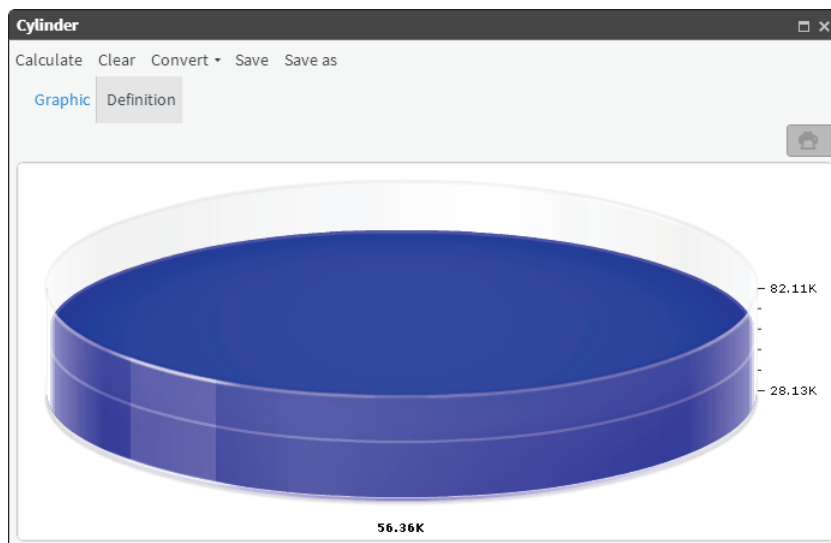


Figure 5-39 Visualizing data measures using a cylinder

Using a funnel

A funnel displays simplified data in graphic form using several, stacked slices. Each slice represents a group of data. A slice that you select on the funnel separates from the rest of the funnel to highlight the visualization. The following tabs appear in the Definition tab:

- Measure tab: Use the buttons to add or delete measures and formulas. To edit an existing measure, double-click it.
- Values tab: Provide the data to be used in the segmentation and measure.
- Properties tab: Select whether to display values or percentages.

How to add measures to a new funnel

- 1 In Gallery—Funnel, choose Definition.
- 2 Choose New measure.
- 3 Type a name for the new measure.
- 4 Select a function for the funnel. From Data Tree, drag a column on which to base the measure, as shown in Figure 5-40.

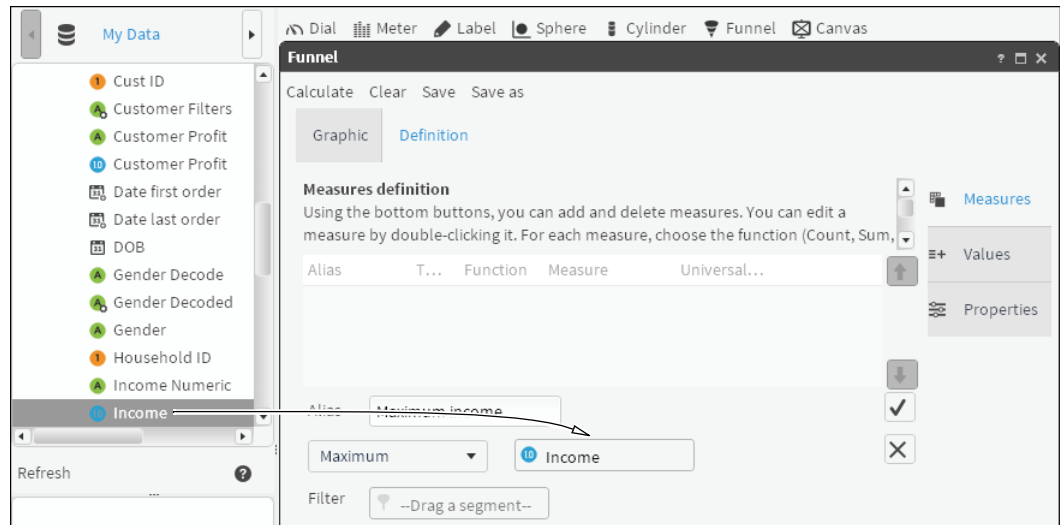


Figure 5-40 Defining a measure for a funnel

- 5 To create a filter, double-click one of the columns in Data Tree. From Discrete Values, drag a field and drop it in Filter, as shown in Figure 5-41.

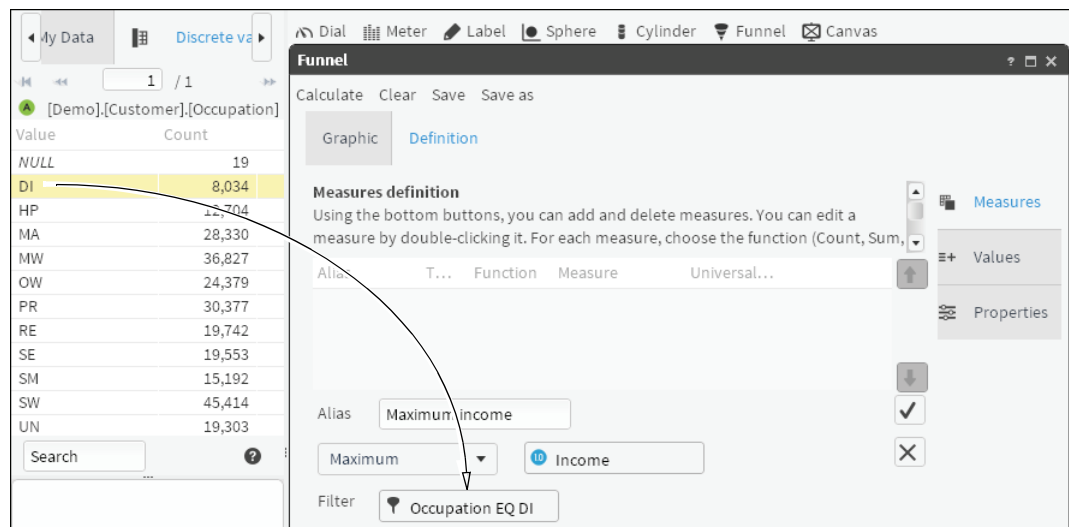


Figure 5-41 Adding a filter to a measure

- 6 To create the measure, choose the check mark.

How to add values and properties to a funnel

- 1 Choose Values.
- 2 Drag a column from My Data and drop the column in Section, as shown in Figure 5-42.

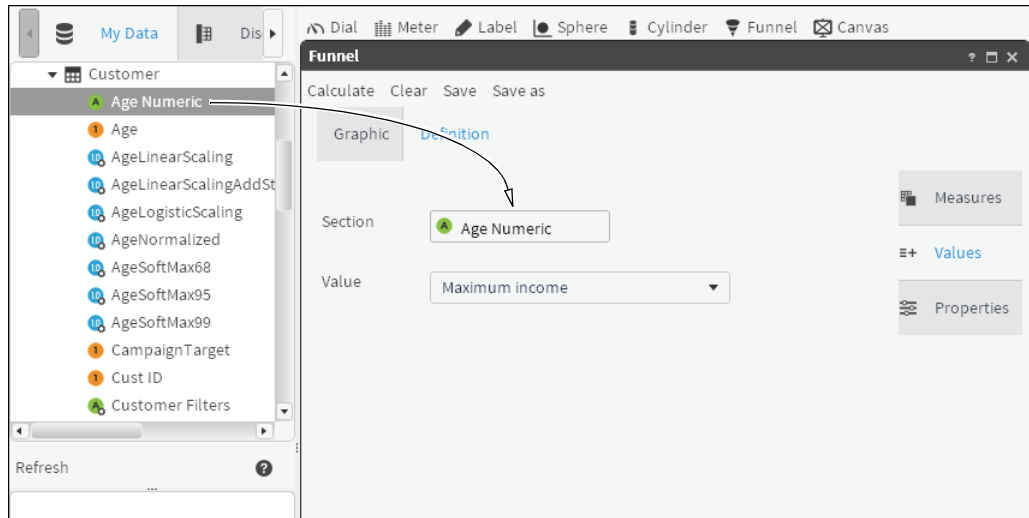


Figure 5-42 Assigning a measure to a value that appears on a funnel

- 3 Select a measure on which to base a funnel.
- 4 Choose Properties.
- 5 In View As, choose an option to represent the data. For example, Figure 5-43 shows choosing to view data as values.

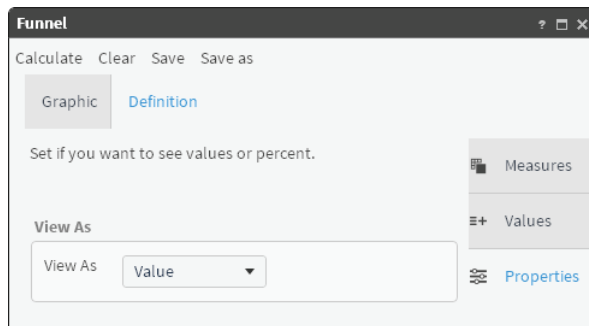


Figure 5-43 Configuring properties for a funnel

- 6 Choose Calculate. The funnel appears on Graphic. To see the values of a section, hover the cursor over the section, as shown in Figure 5-44.

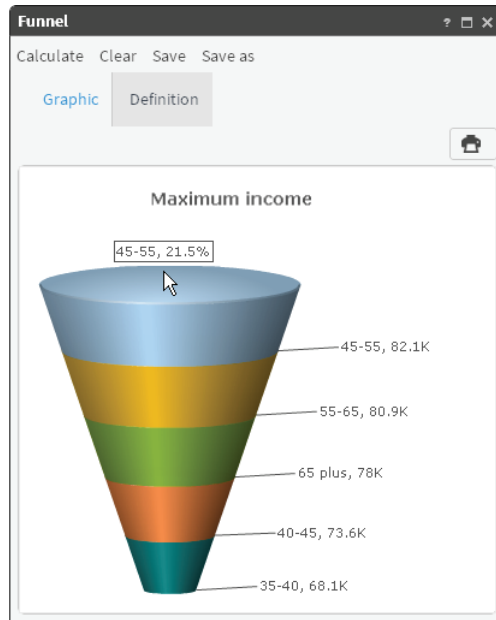


Figure 5-44 Visualizing a data measure using a funnel

Working with a Canvas

Canvas supports getting summaries of a series of previously defined indicators, for example a collection of indicators that simultaneously displays information about the number and amount of orders. Canvas has two modes, view and edit. The view mode, shown in Figure 5-45, does not allow you to make any modifications to content. The edit mode enables you to add or delete one or more components in the command box.

-  To configure an analysis, select the unlock icon to access the edit analysis mode. Then drag and drop the saved analyses or indicators from the analysis tree to the window to add them to the command box. These components are placed sequentially in the column.

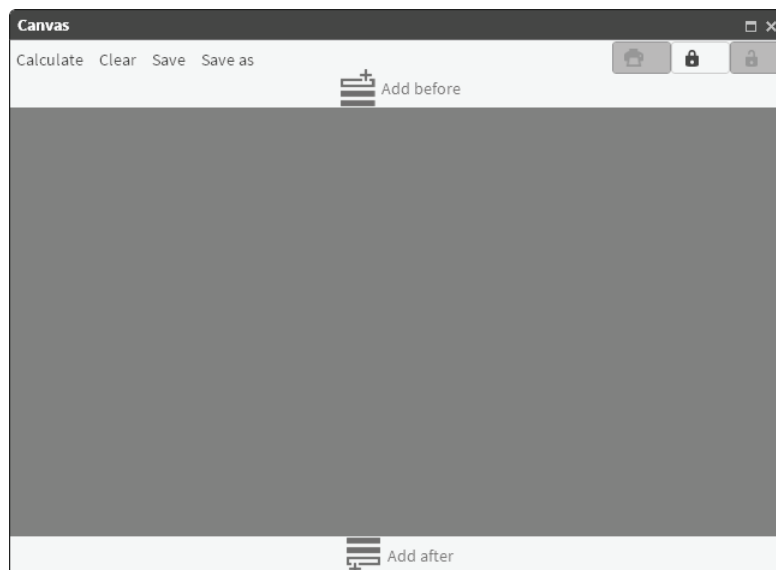


Figure 5-45 Canvas

When you drag a component, arrows appear indicating the area where you can drop it, as shown in Figure 5-46.

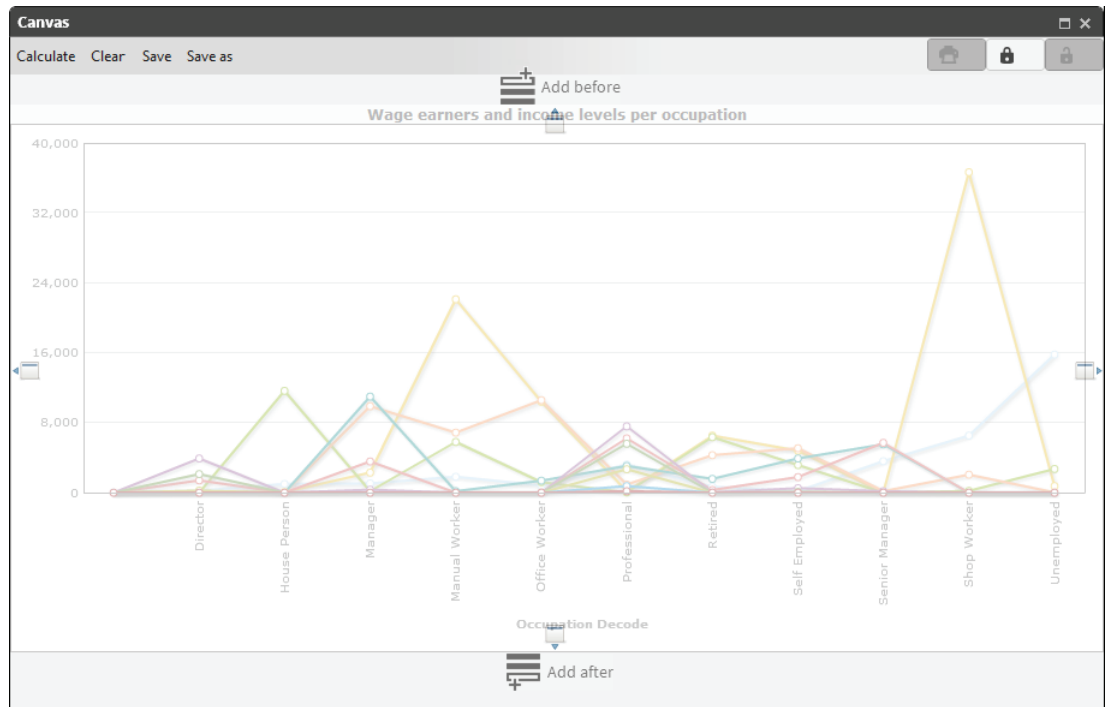


Figure 5-46 Arrows indicating where to drop a component

When you drag a new component to Add before or Add after, the analysis is dropped in a new row, as shown in Figure 5-47.

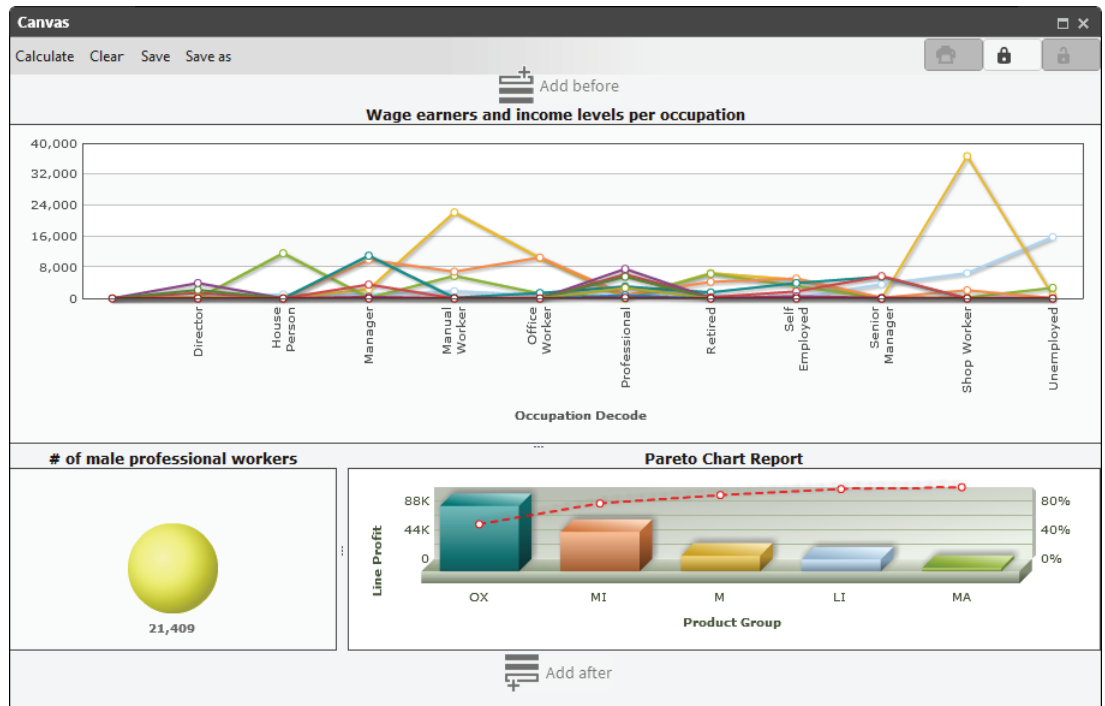


Figure 5-47 Analysis dropped in a new row

To change the size of the design areas, you use the sliders that appear in those areas. You can move canvas components by dragging them to another area or delete them by right-clicking the component and choosing Remove. In edit mode, you can modify components by right-clicking the component and choosing Open. Reopen the canvas to view saved changes.



When you finish placing components in the canvas and you have set its size, click the lock icon to change to query mode and finalize the canvas.

Any previously defined indicator analyses can be added to the canvas. It is also possible to add the graphic results from a crosstab, a bubble diagram, a Venn diagram, map, or Pareto. In the case of a crosstab or bubble, you can choose the display form as a table or graph.

6

Identifying and predicting data trends

This chapter contains:

- Understanding data mining and predictive analytics
- Preparing data for mining
- Understanding Clustering
- Understanding Forecasting
- Understanding decision trees
- Understanding association rules

Understanding data mining and predictive analytics

This information is a summary of the material available on Wikipedia. For more information about data mining and predictive analytics, see the following pages:

http://en.wikipedia.org/wiki/Data_mining

http://en.wikipedia.org/wiki/Predictive_analytics

Data mining is the process of finding patterns in large data sets. Data mining includes:

- **Anomaly detection:** The identification of unusual data records.
- **Association rule learning:** The identification of relationships between variables. For example, a supermarket gathers point-of-sale data. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This type of data mining is called market basket analysis.
- **Clustering:** Discovering groups and structures in the data that are similar.
- **Classification:** Determining the class of an object based on its attributes. For example, an email program classifies e-mails as legitimate or as spam.
- **Regression:** Finding a function that models the data with the least error.
- **Summarization:** Providing a more compact representation of the data set, including visualization and report generation.
- **Sequential pattern mining:** Finding sets of data items that occur together frequently in some sequences. Sequential pattern mining is the basis for web user analysis, stock trend prediction, DNA sequence analysis, finding language or linguistic patterns from natural language texts, and using a history of symptoms to predict disease.

BIRT Analytics supports association rule learning, clustering, and classification (decision trees). BIRT Analytics also supports using time-series prediction to produce short-term demand forecasts. For example, your sales data may contain a trend or a seasonal pattern.

Patterns identified by data mining can be further analyzed. A type of analysis that is of particular interest in business applications is predictive analytics. Predictive analytics is the process of analyzing data to make predictions about future or unknown events. In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Predictive analytics is used in actuarial science, marketing, financial services, insurance, telecommunications, retail, travel, healthcare, pharmaceuticals, and other fields. In financial services, for example, credit scoring is a very common application of predictive analytics. Scoring models analyze a borrower's credit history in order to rank borrowers by the likelihood that they will repay loans on time. One of the most widely used credit scores is the FICO score. Other applications of predictive analytics include:

- **Customer relationship management (CRM):** CRM uses predictive analytics in applications for marketing campaigns and customer services. These applications predict customers' buying habits and identify issues that may result in the loss of customers. For example, promotional activities are based on predictive analytics.
- **Clinical decision support systems:** Clinicians use predictive analytics to determine which patients are at risk of developing conditions such as diabetes, asthma, and heart disease.
- **Collections:** Every portfolio has delinquent customers who do not make their payments on time. The financial institution has to undertake collection activities on these customers to recover the amounts due. Collection resources are wasted on customers who are difficult or impossible to recover. Predictive analytics can help optimize the allocation of collection resources by identifying the most effective collection agencies, contact strategies, and legal actions against each customer, thus significantly increasing recovery and reducing collection costs.

- **Cross-selling:** Businesses collect and maintain data on customers and sales transactions. Identifying relationships in the data can provide a competitive advantage. Predictive analytics can analyze customers' spending, usage, and other behavior, leading to efficient cross sales. This results in greater profitability per customer and stronger customer relationships.
- **Customer retention:** Businesses must maintain customer satisfaction by rewarding loyalty and minimizing attrition. Businesses tend to respond too late to customer attrition, acting only after the customer has initiated the termination of service. At this stage, it is very unlikely that the customer will change their decision. Predictive analytics can prompt a more proactive retention strategy. By frequent examination of a customer's past service usage, service performance, spending, and other behavior patterns, predictive models can determine the likelihood that a customer will terminate service some time in the near future. Generous offers can increase the chances of retaining the customer. Silent attrition, the behavior of a customer to slowly but steadily reduce usage, is another problem that many businesses face. Predictive analytics can also predict this behavior, so that the company can take action to increase customer activity.
- **Direct marketing:** Predictive analytics can identify the most effective combination of products, marketing material, communication channels, and timing that should be used to target a consumer. The goal of predictive analytics is to lower the cost per order or cost per action.
- **Fraud detection:** Predictive analytics can reduce a business's exposure to fraud. Fraud includes inaccurate credit applications, fraudulent transactions (both offline and online), identity theft, and false insurance claims. Credit card issuers, insurance companies, retail merchants, manufacturers, business-to-business suppliers, and service providers are all potential victims of fraud. In the United States, the Internal Revenue Service uses predictive analytics to mine tax returns and identify tax fraud. Web fraud detection utilizes heuristics to study normal web user behavior and detect anomalies indicating fraud attempts.
- **Portfolio, product, or economy-level prediction:** Often the focus of analysis is not the consumer but a product, portfolio, firm, industry, or even the economy. For example, a retailer might want to predict store-level demand for inventory management purposes. Or the Federal Reserve Board might want to predict the unemployment rate for the next year. These types of problems can be solved with predictive analytics.
- **Risk management:** For example, the Capital asset pricing model (CAP-M) predicts the optimal portfolio to maximize return.
- **Underwriting:** Many businesses have to account for risk exposure and determine the cost to cover the risk. For example, auto insurance providers must accurately determine the premium to charge to cover each automobile and driver. A financial company must assess a borrower's ability to pay before granting a loan. For a health insurance provider, predictive analytics can be used to analyze several years of past medical claims data, as well as lab, pharmacy, and other records, to predict how expensive an enrollee is likely to be in the future. Predictive analytics can help underwrite these quantities by predicting the chances of illness, default, bankruptcy, and so on. Predictive analytics in the form of credit scores has reduced the time it takes for loan approvals, especially in the mortgage market where lending decisions are now made in a matter of hours rather than days or even weeks.

Preparing data for mining

BIRT Analytics supports normalization, scaling, and remapping operations that prepare a data column to meet conditions required by each data mining algorithm. Preprocessing applies a mathematical operation to values in a data column. Preprocessing a column of data values having a distribution that differs from a standard, or normal, distribution before applying a data mining algorithm can produce a more useful result. For example, you can

compare data sets that have different scales and units by standardizing the data so that it falls in the 0 to 1 range. Test scores are often calibrated by percentile, with most scores falling in the 25th to 75th percentile.

Figure 6-1 shows the distribution of raw, or non-standardized, data for age and income. Ages fall in the 19 to 93 range, while incomes fall in the 479.79 to 111571.4 range. To compare these distributions, you must standardize the data.

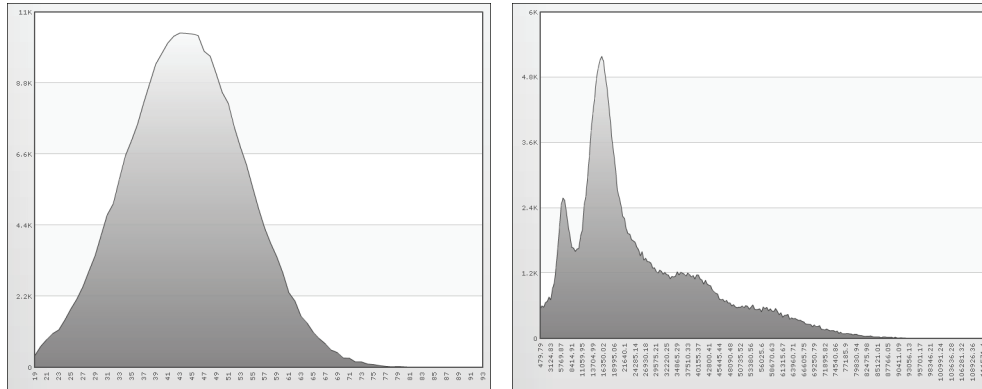


Figure 6-1 Distribution of raw data for age (left) and income (right)

Standardizing data in a column

There are four ways to standardize the data in a column:

- Normalization
- Linear scaling
- Logistic scaling
- Softmax scaling

In each case, a new column is created to contain the standardized data.

Understanding normalization

Normalization calculates the mean of all values in a column. Each value in the new column compares in the same way to values in a standard, normal distribution. The operation subtracts the mean value from each value in the column, then divides the difference by the standard deviation. The formula is:

$$y = (x - \text{mean}\{x_1, x_N\}) / (\text{stdv}\{x_1, x_N\})$$

Standard deviation shows how much variation there is from the average (mean), or expected value. A low standard deviation indicates that the data points tend to be very close to the mean. A high standard deviation indicates that the data points are spread out over a large range of values.

Figure 6-2 shows normalized data for age and income. The values on the horizontal axis represent the number of standard deviations from the mean. The standard deviation of the mean is 0.

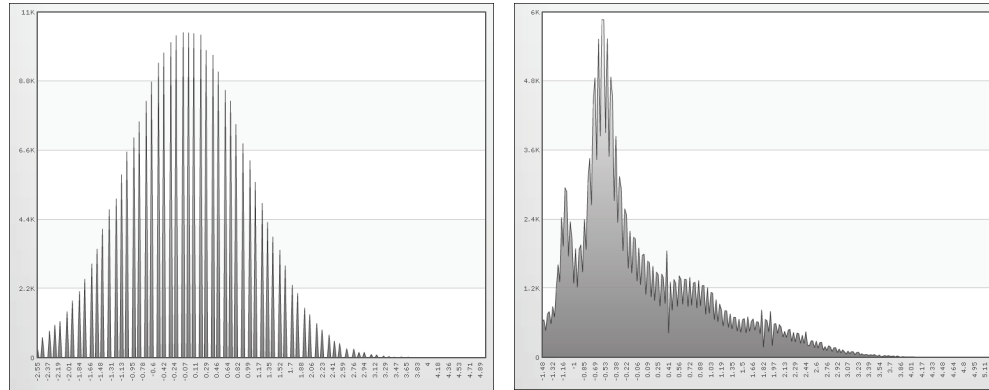


Figure 6-2 Normalized data for age (left) and income (right)

Understanding linear scaling

Standardization by linear scaling is useful when values in a column have the following characteristics:

- Many similar values
- Very extreme minimum and/or maximum values that are highly concentrated

The more the data are clustered, the better the result obtained by linear regression study. Maximum and/or minimum values are calculated to be suitable to perform linear regression analysis. Linear scaling supports two options:

- Use the original minimum and maximum values.
- Add a stretch to the minimum and maximum values. This option adds, according to the percentage indicated, a minimum and maximum value before the data is preprocessed.

The formula is:

$$y = (x - \min\{x_1, x_N\}) / (\max\{x_1, x_N\} - \min\{x_1, x_N\})$$

Figure 6-3 shows linear scaling with the original minimum and maximum values for age and income.

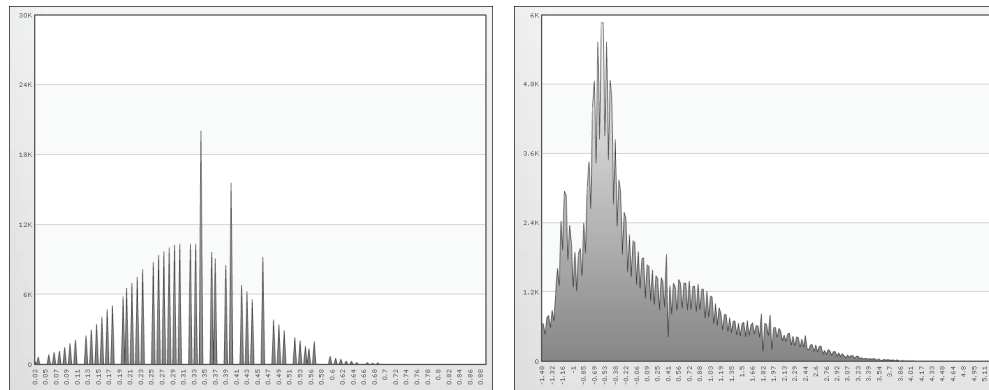


Figure 6-3 Linear scaling with original minimum and maximum values for age (left) and income (right)

Figure 6-4 shows linear scaling with a stretch to the minimum and maximum values for age and income.

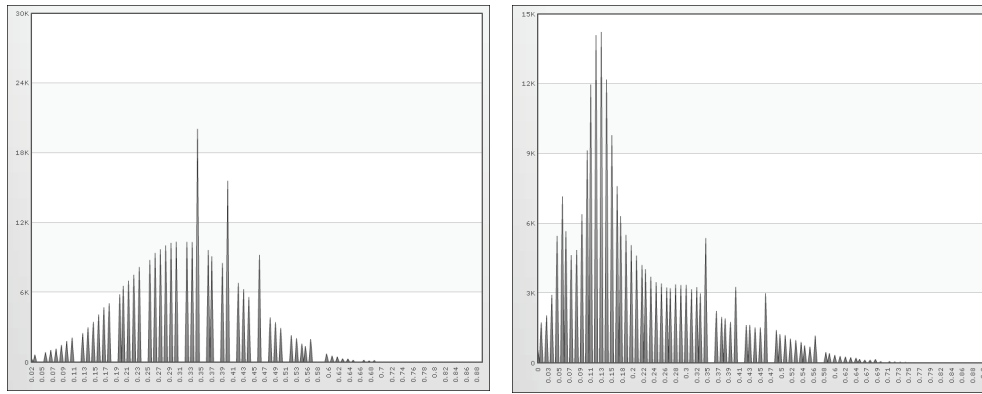


Figure 6-4 Linear scaling with a stretch to the minimum and maximum values for age (left) and income (right)

Understanding logistic scaling

Standardization by logistic scaling is used to recode the variable of study for use in a logistic regression. Logistic regression is a type of regression analysis used for predicting the outcome of a categorical dependent variable (a dependent variable that can take on a limited number of values) based on one or more predictor variables. The equation used is:

$$P(n) = 1 / (1 + e^{-n})$$

where n represents the values in the column. This equation analyzes the values to form a logistic model.

Understanding Softmax scaling

Softmax scaling standardization is a nonlinear transformation that reduces data ranges for the values in a column as much as possible. The objective is to achieve the minimum and maximum values asymptotically. In other words, the low-end and high-end values gradually approach the minimum and maximum values without ever reaching them.

If you choose Softmax scaling, you can set the confidence level to 68%, 95%, or 99%. The lower the confidence level is, the shorter the intervals and the greater the probability of error will be. The formula is:

$$x' = x - E(x) / \lambda (\sigma_x / 2\pi)$$

where

- x represents the values in the column.
- E(x) is the average value.
- λ is the confidence level.
- σ_x is the standard deviation of the study variable.
- π is the number 3.14159...

Figure 6-5 shows Softmax scaling at 68% for age and income.

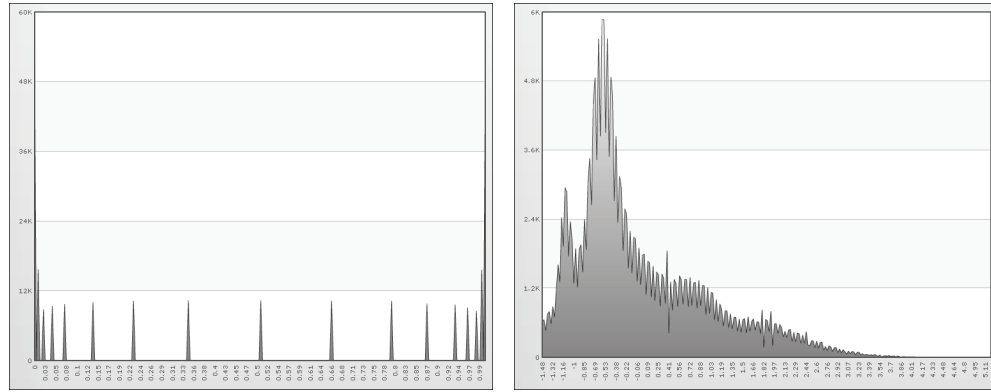


Figure 6-5 Softmax scaling at 68% for age (left) and income (right)

Figure 6-6 shows Softmax scaling at 95% for age and income.

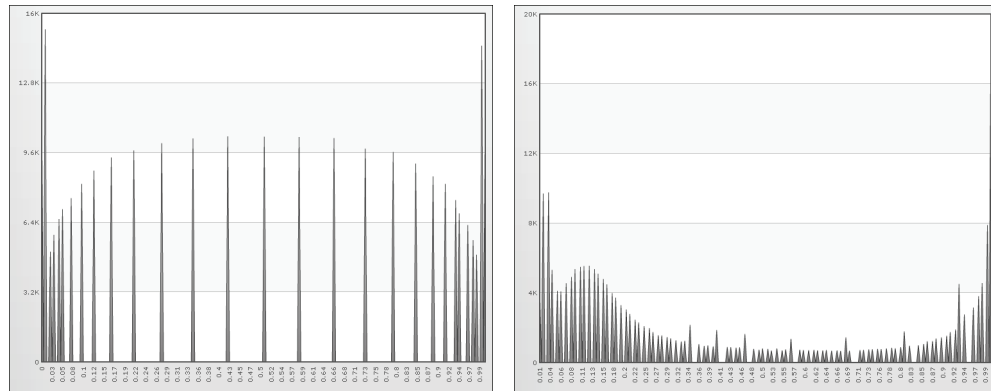


Figure 6-6 Softmax scaling at 95% for age (left) and income (right)

Figure 6-7 shows Softmax scaling at 99% for age and income.

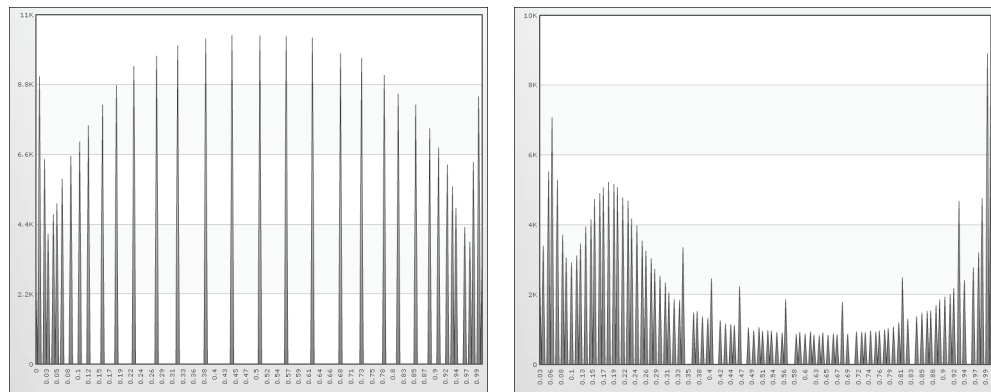


Figure 6-7 Softmax scaling at 99% for age (left) and income (right)

How to standardize the data in a column

- 1 In Analytics—Advanced, choose Preprocessing → Standardize Column.
- 2 Drag the column you want to standardize from My Data and drop it in Column in Standardize column, shown in Figure 6-8.
- 3 In New column, type the name of the new column.

- 4 Choose Normalize, or choose one of the scaling algorithms: linear scaling, logistic scaling, or SoftMax scaling. If you choose linear scaling or SoftMax scaling, you can customize the scaling.
- 5 Choose Create. The column appears in My Data.

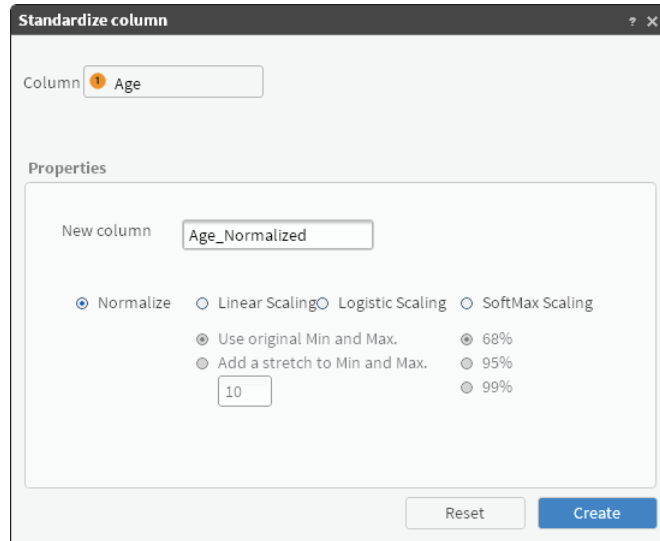


Figure 6-8 Standardizing the data in a column

Remapping a column

If you want to use a column that contains strings in an analysis, you must remap the strings to numbers. For example, Figure 6-9 shows the strings in the Occupation column on the left and the numbers in the remapped Occupation column on the right.

Value	Count
NULL	19
DI	8,034
HP	12,704
MA	28,330
MW	36,827
OW	24,379
PR	30,377
RE	19,742
SE	19,553
SM	15,192
SW	45,414
UN	19,303

Value	Count
-1	19
0	8,034
1	12,704
2	28,330
3	36,827
4	24,379
5	30,377
6	19,742
7	19,553
8	15,192
9	45,414
10	19,303

Figure 6-9 Occupation column (left) and remapped Occupation column (right)

How to remap a column

- 1 In Analytics—Advanced, choose Preprocessing → Remap Column.
- 2 Drag the column you want to remap from My Data and drop it in Column in Remap column, shown in Figure 6-10.
- 3 In New column, type the name of the new column.
- 4 Choose Create. The column appears in My Data.



Figure 6-10 Remapping a column

Understanding Clustering

Clustering enables you to organize data based on variables you specify. Usually, the clustering algorithm produces segments of data that help you identify groups with the largest number of attributes in common. In other words, clustering provides an idea about the similarity and differences between records in the same group.

After BIRT Analytics applies this algorithm, a new field is created in the selected table to group records into a specified number of clusters (N). Because each record is given a value for the clustering, you can see a count of records in each cluster in the Data Tree's Discrete Values view. For example, customers grouped in the same category or cluster may have common demographic features.

You must use continuous variables because clustering calculates the distance between values to set up a group, and only fields with continuous values work for clustering. Continuous means that there are many discrete values. Categorical variables, or fields with few discrete values like gender or occupation, do not work.

To set up a clustering model, create a training process.

How to set up a training process

1 Choose Parameters and specify the following:

- Domain: The segment of data from the database. All linked tables are automatically added.
- Confidence level: The representative sample size to create the groups.
- Clusters: The number of groups.
- The attributes or categories to create the groups. Only fields containing continuous values are available to choose. Add them by dragging from the list on the left to the area on the right.

2 Choose Train.

How to use the results

When training finishes, Results contains a list with all the groups that have been created, the number of records in each group, and the mean of each attribute used to set up the groups. Note that every mean value acts as a centroid of the group.

- 1 Review the results. After you save the cluster, you cannot train it again.
- 2 Save the trained cluster in My Folders, to an existing folder, or a newly created one.
- 3 Select the saved cluster in My Folders. Right-click and choose Open.
- 4 On K-Means, drag a new segment and drop it in Domain.

- 5 Type the new target column name.
- 6 Choose OK. The new column appears in the selected table.

Understanding Forecasting

Forecasting is a method of extrapolating or predicting data based on time. BIRT Analytics forecasting uses the Holt-Winters method, iteratively applying a formula to produce a time series and a forecast. This formula uses a weighted average of data prior to time t to provide a result for time t .

This method consists of three components: the level, trend, and seasonal component.

For example, to forecast the number of orders to be received during the next 12 months, you would perform the following tasks:

- Select the data you wish to observe, the y -axis of the time series.
- Select the time interval for the series, the x -axis of the time series.
- Execute pre-analysis, if necessary.
- Set model parameters. These may include:
 - Number of projections to make
 - Whether the time series has seasonality
 - Seasonal periodicity
 - Whether or not to replace outliers
 - Level, trend, and season smoothing parameters
- Execute forecast.

More about outliers

Outliers are observations that appear to deviate markedly from other members of the sample in which they occur. When running pre-analysis, values that are more than two standard deviations away from the sample mean are considered outliers. Outliers are replaced by the sample mean. Generally, outliers should be replaced so that they do not bias any projections.

The value of the previous observation replaces any null values encountered in the sample. If the null value encountered is the first observation of the sample, the value of the nearest non-null observation replaces the null value.

How to create and execute a forecast

To get a forecast of the number of orders to be received in the next 12 months, select Parameters and complete the following procedure:

- 1 From My Data, drag the Order table and drop it in Measure. Leave the default operation as Count.
- 2 Drag the Month column and drop it in Column, in the Dimension area.
- 3 To find out if your data contains a seasonal element, select Has seasonality and set Seasonal periodicity to 12. With monthly data, there are 12 observations in each cycle until the seasonal pattern is repeated.
- 4 Set other values, as required:
 - Filter: Constrain the data by the specified selection.
 - Number of projections: The number of observations to make into the future.

- Replace outliers: To eliminate outliers in the original data.
- Format: To specify the number of decimal places to use when displaying the forecasted data.
- Autocalculate: To calculate the smoothing parameters for the model automatically.
- Level, Trend and Season smoothing: If Autocalculate is not selected, you can set the smoothing parameters for the model manually. Smoothing parameters are values from 0.1 to 0.9. The smaller the smoothing parameter is, the less weight given to the most recent observations in the time series will be, making the series smoother.

5 Execute pre-analysis:

- 1 To analyze the time series before making a prediction, choose Pre-Analysis.
- 2 On Pre-Analysis, analyze the results and decide:
 - Whether to replace any outliers in the time series. The number of outliers appears next to Number of outliers. The coordinates of any outliers appears on the graph with a red line.
 - How much confidence you can place in the projections. Evidence for non-randomness is a 5-star assessment. An assessment of three or more stars is considered statistically significant. The less random the results are, the more confidence you can have in the reliability of any projections made. Keep in mind, however, that any randomness may be attributed to seasonal patterns in the time series.
 - Whether to account for seasonality in the forecast and, if so, the seasonal periodicity. Evidence for seasonality is a five-star assessment. An assessment with three or more stars is considered statistically significant. If seasonality is indicated, you select Has seasonality when making a forecast, using the same seasonal periodicity that you used to run the pre-analysis.

6 To run the forecast, choose Calculate.

How to use forecasting results

After calculating the results, you can analyze and save the forecast.

1 On Forecast, analyze the graphical results:

- A yellow line represents the forecast with a blue line representing the original data. Hover the cursor over a chart co-ordinate to display its value. All projections are labeled numerically on the *x*-axis, starting from 1.
- If outliers were replaced, then the original data shown displays the value used to replace the outlier.
- You can also zoom in to any part of the chart to get a more detailed perspective. Select the area of the graph on which to zoom. Zoom out by right-clicking in the chart and selecting Zoom Out Chart. Choose Reset Chart to show all data points.
- To make comparisons between different time periods, overlay different parts of the chart. Choose Switch to pin mode and select an area to float, then drag the floating area over another area of the chart to make the comparison. To return to the original state, choose Switch to Zoom Mode.
- The smoothing parameters used for the model, including those selected automatically, appear at the top of the tab.
- To export the forecast to a rich text format (RTF) document, choose the export icon. The generated document contains the chart exactly as shown in the Forecast tab, along with the data in tabular format.



- 2 On Table, review the forecasted data and the original data in tabular format. The forecasted values are labeled numerically, starting from 1.
- 3 To save the forecasting analysis, choose Save. On Save, you can choose to share the analysis with other BIRT Analytics users.

Understanding decision trees

A decision tree predicts the class of an object based on the object's attributes. For example, you can use a decision tree to predict whether a passenger on the *Titanic* survived, based on the passenger's gender, age, and number of siblings, as shown in Figure 6-11. The numbers under each leaf show the probability of survival for a passenger with the specified attributes and the percentage of passengers represented by the leaf. Figure 6-11 tells us:

- A female passenger had a 0.73 chance of surviving.
- A male passenger over the age of 9 had only a 0.17 chance of surviving.
- For male passengers age 9 or younger, the probability of survival depends on the number of siblings the passenger had. If the passenger had more than two siblings, he had only a 0.05 chance of surviving. If the passenger had two siblings or fewer, he had a 0.89 chance of surviving.

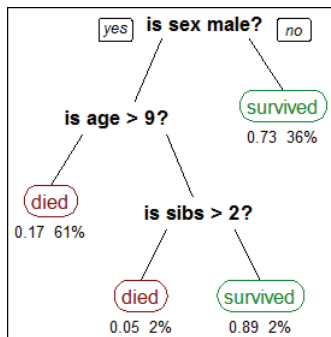


Figure 6-11 Decision tree showing survival probabilities for passengers on the *Titanic*

Figure 6-14 shows a decision tree that predicts whether a worker is low income, medium income, or high income, based on the worker's occupation and gender.

Training and testing a predictive model

A decision tree must be trained on sample data. The decision tree learns with each successive application of the predictive model. Some patterns found by data mining algorithms, however, are invalid. Data mining algorithms often find patterns in the training set that are not present in the general data set. This is called overfitting. To solve this problem, test the predictive model on a set of data different from the training set. The learned patterns are applied to the test set and the resulting output is compared to the desired output. For example, a data mining algorithm that distinguishes spam from legitimate e-mails is trained on a set of sample e-mails. Once trained, the learned patterns are applied to a test set of emails. The accuracy of the predictive model is measured from how many e-mails it classifies correctly.

Understanding the confusion matrix

A confusion matrix tabulates the results of a predictive algorithm. Each row of the matrix represents an actual class. Each column represents a predicted class. For example, consider a classification system that has been trained to distinguish between cats, dogs, and rabbits. Figure 6-12 shows how frequently the predictive algorithm correctly classifies each type of animal. The sample contains 27 animals: 8 cats, 6 dogs, and 13 rabbits. Of the 8 cats, the

algorithm predicted that three are dogs, and of the six dogs, it predicted that two are cats and one is a rabbit. The confusion matrix shows that the algorithm is not very successful distinguishing between cats and dogs. It is, however, successful distinguishing between rabbits and other types of animals, misclassifying only two of 13 rabbits. Correct predictions are tallied in the table's diagonal. Non-zero values outside the diagonal indicate incorrect predictions.



		Predicted class		
		Cat	Dog	Rabbit
Actual class	Cat	5	3	0
	Dog	2	3	1
	Rabbit	0	2	11

Figure 6-12 Confusion matrix showing the actual class and predicted class for animals

Understanding sensitivity and specificity

Sensitivity, also called the true positive rate, measures the proportion of actual positives that are correctly identified as such, for example the percentage of sick people who are correctly identified as having a condition. Specificity, also called the true negative rate, measures the proportion of negatives that are correctly identified as such, for example the percentage of healthy people who are correctly identified as not having a condition. A perfect predictor is 100% sensitive, in other words predicting that all people in the sick group are sick, and 100% specific, in other words not predicting that anyone in the healthy group is sick. For any test, however, there is a trade-off between the measures. For example, in an airport security setting where one is testing for potential threats to safety, scanners may be set to trigger on low-risk items such as belt buckles and keys, low specificity, in order to reduce the risk of missing objects that pose a threat to passengers and crew, high sensitivity.

How to create a decision tree

- 1 In Analytics—Advanced, choose Decision tree.
- 2 Drag the appropriate table from My Data and drop it in Domain in Decision tree, shown in Figure 6-13.
- 3 In the left pane of Domain columns, expand the database and the appropriate tables.
- 4 Drag the appropriate columns from the left pane and drop them in the right pane. The columns specify the attributes used to assign classifications in the decision tree.
- 5 Create the classifications:
 - 1 Choose New classifier.
 - 2 In Alias, type the name of the classifier.
 - 3  Choose the color picker button and select a color. The name and the color visually identify the classification in the decision tree. For example, you can define a High income classification identified by the color green.
 - 4 Drag a segment from Discrete values and drop it in Drag a segment. The segment specifies the condition for the classification. For example, you can define a High income classification as income over 80000.
 - 5  Choose the check mark. The classification appears in Classifications.
 - 6 Create the remaining classifications. Figure 6-13 shows three classifications: High income, Medium income, and Low income.

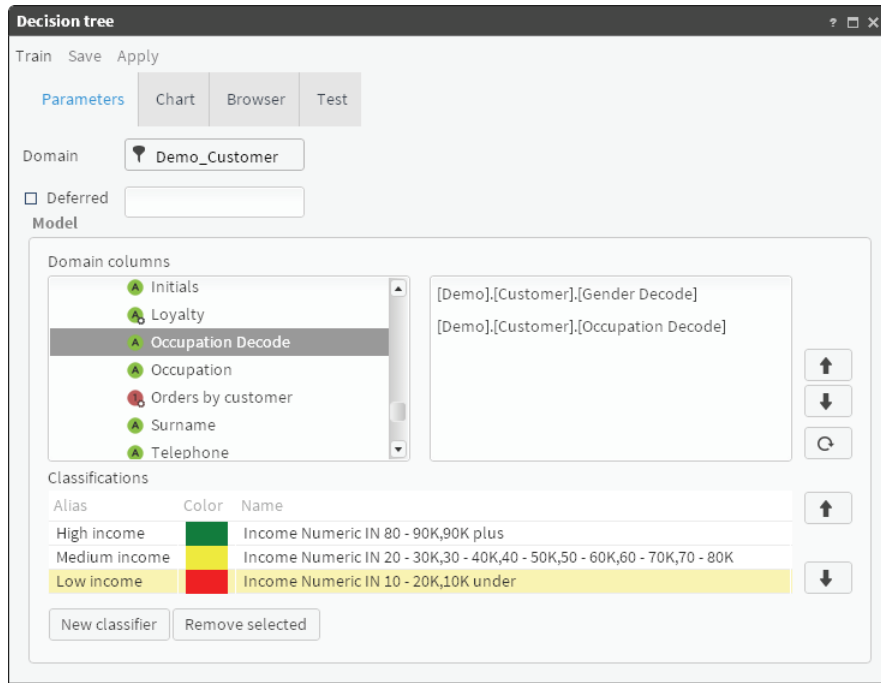


Figure 6-13 Specifying the parameters for a decision tree

- Choose Train. A graphical representation of the decision tree appears in Chart. The classifications appear with the names and colors you specified.

Figure 6-14 shows part of a decision tree for the Customer table domain with domain columns Occupation Decode and Gender Decode. The decision tree predicts whether a worker is low income, medium income, or high income, based on the worker's occupation and gender. For professionals, the income classification does not depend on gender. For office workers, however, the classification for men is medium income, while the classification for women is low income.

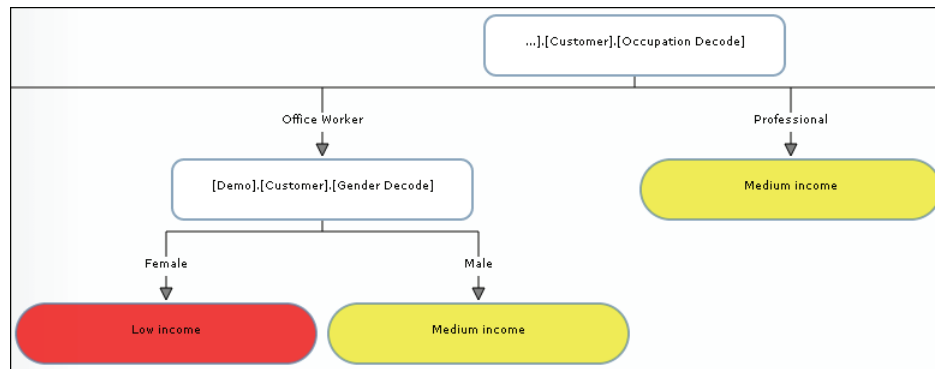


Figure 6-14 Graphical representation of a decision tree

- Choose Browser. A tabular representation of the decision tree appears.
- Expand the tree. Figure 6-15 shows a tabular representation of the decision tree, shown graphically in Figure 6-14.

Tree	Color	Class	Slice	Segment error	Total error
Total			100%	16.5155%	16.5155%
[Demo],[Customer],[Occupation Decode]					
Director	Yellow	Medium income	3.4598%	0.0257%	0.0009%
House Person	Red	Low income	5.2026%	0%	0%
Manager	Yellow	Medium income	12.1265%	8.9896%	1.0901%
Manual Worker	Red	Low income	15.5659%	20.1502%	3.1366%
Office Worker	Yellow	Medium income	10.478%	43.4491%	4.5526%
[Demo],[Customer],[Gender Decode]					
Female	Red	Low income	3.7145%	41.7953%	1.5525%
Male	Yellow	Medium income	6.7635%	44.3568%	3.0001%
Professional	Yellow	Medium income	12.0768%	4.1121%	0.4966%
Retired	Red	Low income	8.4796%	32.5067%	2.7564%
Self Employed	Yellow	Medium income	8.6174%	41.1959%	3.55%
Senior Manager	Yellow	Medium income	5.1448%	0%	0%
Shop Worker	Red	Low income	17.2922%	5.2706%	0.9114%
Unemployed	Red	Low income	1.5565%	1.3425%	0.0209%

Figure 6-15 Tabular representation of a decision tree

9 Choose Test.

10 Drag the appropriate table from My Data and drop it in Domain.

11 Choose Test. Figure 6-16 shows the test results.

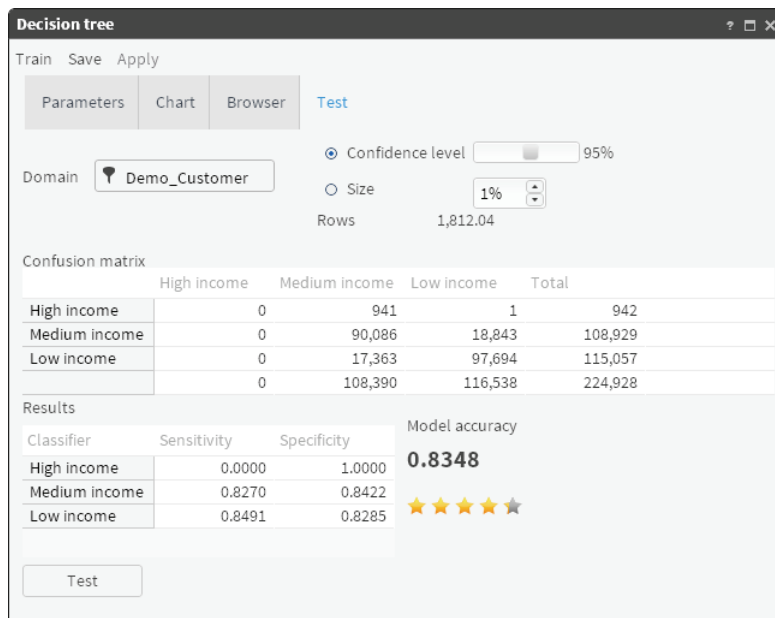


Figure 6-16 Test results, including the confusion matrix

Understanding association rules

Association rules find patterns in very large amounts of point-of-sale data. An association rule is of the form "If a shopper purchases Item A and Item B, the shopper also purchases Item C." For example, association rules might indicate that if a shopper buys onions and potatoes on a trip to the supermarket, she is likely also to buy hamburger meat. Such information can be used as the basis for decisions about marketing activities, such as promotional pricing and product placements.

To identify association rules, you must provide the transaction ID and the transaction items for a very large number of purchases. The Transaction Id and Transaction Item columns must belong to the same database table. You must also provide values for Minimum Support and Minimum Confidence:

- Minimum Support is the minimum percentage of transactions that must contain all the items in the rule. By default, it is set to 10%.

$$\text{Support (A,B)} = \text{Transactions (A,B)} / \text{Total transactions}$$

- Minimum Confidence measures how well a rule predicts the right-hand side after the “then” clause of the rule (If A,B then C) by comparing how often the right-hand side (consequent) appears when the condition on the left-hand side (antecedent) is met. By default, this is set to 65%.

$$\text{Confidence (A,B-> C)} = \text{Support (A,B,C)} / \text{Support (A,B)}$$

Use the default values for Minimum Support and Minimum Confidence for the first trial. Depending on the results, you can increase or decrease these values.

Optional parameters include:

- Filter: You can define a domain to restrict the number of transactions used in the trial.
- Antecedent Item Range: Minimum and maximum number of items that the left-hand side of the rule must contain.
- Filter for the antecedent items.
- Consequent Item Range: Minimum and maximum number of items that the right-hand side of the rule must contain.
- Filter for the consequent items.

The results of a trial are displayed in a table. The antecedent and consequent items appear in the first two columns. There are also several numeric columns that measure how well the rule predicts the consequent. The columns that appear in the Results tab are:

- Antecedent: Items on the left-hand side of the rule.
- Consequent: Items on the right-hand side of the rule.
- Support (%): Measures the percentage of transactions that contain all the items in the rule.
- Confidence (%): Measures how well the rule predicts the consequent by determining how often the consequent appears when the condition of the antecedent is met.
- Lift: Also called improvement. Measures how well the rule predicts the consequent. A lift greater than one indicates that the items in the rule appear together more than expected.
- Leverage: Measures how well the rule predicts the consequent. The calculation method is different from lift, but the purpose is the same.
- Opportunity: A star indicates that the lift is greater than one.

Items in the antecedent and consequent are enclosed in double quotation marks and separated by a comma. Numeric rule measures such as support, confidence, and lift are rounded to two decimal places.

How to find association rules

- 1 In Analytics—Advanced, choose Association Rules.
- 2 Drag the appropriate column from My Data and drop it in Column in Transaction ID.
- 3 Drag the appropriate column from My Data and drop it in Column in Transaction Item. Accept the default values for Minimum Support and Minimum Confidence, as shown in Figure 6-17.

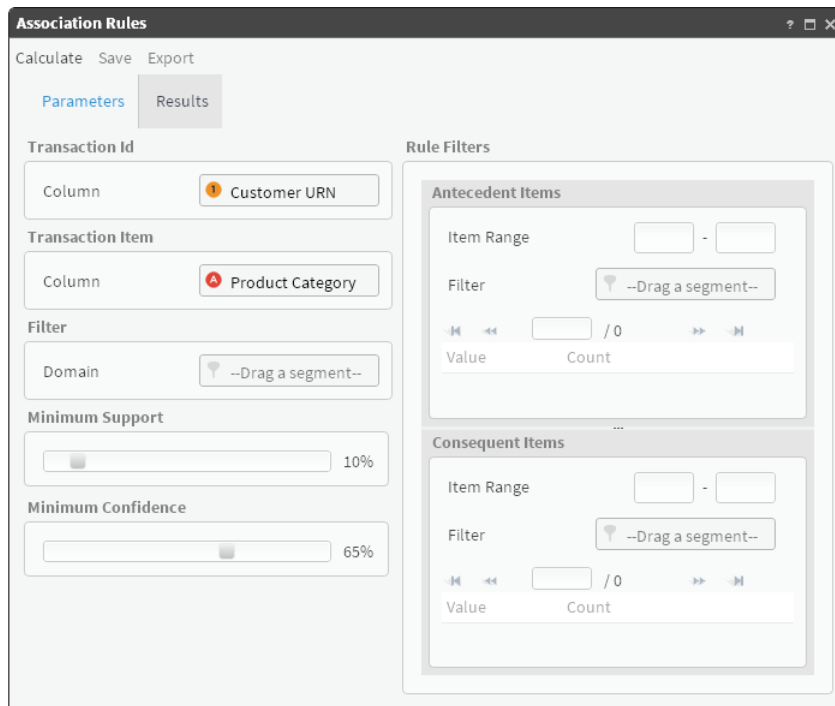


Figure 6-17 Default values for Minimum Support and Minimum Confidence

- 4 Choose Calculate. The association rules appear in the Results tab, as shown in Figure 6-18.

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Opportunity
"Savings Acc"	"Current Acc"	28.44	69.46	0.92	-0.03	
"Mortgage"	"Current Acc"	12.49	69.43	0.92	-0.01	
"Credit Card"	"Current Acc"	13.88	69.42	0.91	-0.01	
"Invesment"	"Current Acc"	21.17	69.36	0.91	-0.02	
"Loan"	"Current Acc"	14.92	69.27	0.91	-0.01	

5 of 5 rows

Figure 6-18 Results using default values for Minimum Support and Minimum Confidence

- 5 If the default values for Minimum Support and Minimum Confidence do not yield many opportunities, choose Parameters and decrease these values. In Figure 6-19, Minimum Support and Minimum Confidence are decreased to 5% and 30%, respectively.

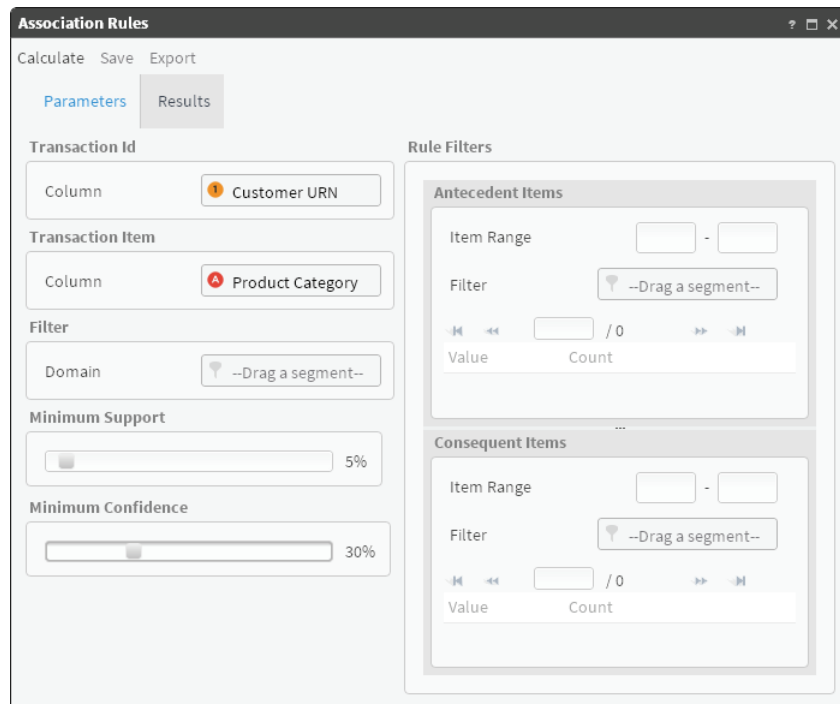


Figure 6-19 Decreasing the values for Minimum Support and Minimum Confidence

- Choose Calculate. The association rules appear in the Results tab, as shown in Figure 6-20. Decreasing the values for Minimum Support and Minimum Confidence yields several opportunities, indicated by a gold star.

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Opportunity
"Loan"	"Investment"	6.60	30.67	1.00	0.00	★
"Loan"	"Savings Acc"	8.83	41.01	1.00	0.00	★
"Current Acc", "Investment"	"Savings Acc"	8.67	40.96	1.00	0.00	★
"Current Acc", "Credit Card"	"Savings Acc"	5.68	40.93	1.00	-0.00	★
"Investment"	"Savings Acc"	12.49	40.93	1.00	-0.00	★
"Savings Acc"	"Investment"	12.49	30.52	1.00	-0.00	★
"Current Acc", "Loan"	"Savings Acc"	6.10	40.91	1.00	-0.00	★
"Credit Card"	"Investment"	6.10	30.50	1.00	-0.00	★
"Credit Card"	"Savings Acc"	8.18	40.90	1.00	-0.00	★
"Current Acc", "Savings Acc"	"Investment"	8.67	30.50	1.00	-0.00	★
"Current Acc", "Mortgage"	"Savings Acc"	5.11	40.87	1.00	-0.00	★
"Mortgage"	"Investment"	5.48	30.44	1.00	-0.00	★
"Mortgage"	"Savings Acc"	7.34	40.79	1.00	-0.00	★
"Savings Acc", "Mortgage"	"Current Acc"	5.11	69.58	0.92	-0.00	
"Savings Acc", "Credit Card"	"Current Acc"	5.68	69.47	0.92	-0.01	
"Savings Acc"	"Current Acc"	28.44	69.46	0.92	-0.03	
"Current Acc"	"Savings Acc"	28.44	37.48	0.92	-0.03	
"Mortgage"	"Current Acc"	12.49	69.43	0.92	-0.01	
"Credit Card"	"Current Acc"	13.88	69.42	0.91	-0.01	
"Savings Acc", "Investment"	"Current Acc"	8.67	69.41	0.91	-0.01	
"Investment"	"Current Acc"	21.17	69.36	0.91	-0.02	
"Loan"	"Current Acc"	14.92	69.27	0.91	-0.01	
"Savings Acc", "Loan"	"Current Acc"	6.10	69.09	0.91	-0.01	

Figure 6-20 Results using decreased values for Minimum Support and Minimum Confidence

7

Managing campaigns

This chapter contains:

- Understanding campaigns
- Configuring campaign elements
- Planning a campaign
- Running a campaign

Understanding campaigns

A campaign is a set of tasks, defined for specific population segment and completed during a defined time period to accomplish a specific goal. For example, a typical business campaign defines a set of communication tasks that channel information to a segment of customers or prospects. The most typical campaign generates advertising messages to customers in a selected market segment. Common goals of an advertising campaign are web site visits and online purchase transactions made by customers.

BIRT Analytics supports automating campaign tasks associated directly with events or conditions that occur in your database. For example, a forecast analysis predicts a seasonal percentage increase in purchases by customers. Also, association rules modeling shows what additional items a purchasing customer typically buys. By defining a campaign strategy that includes seasonal timing and targeted messaging, your company can effectively persuade a customer to buy an additional item or upgrade to one having a higher profit margin.

Further refining this idea, you can design specific messaging delivered to select market segments, based on data that you collect from that specific segment. For example, analyzing the profile of a customer who responds by purchasing one item enables your website to offer suggestions about similar products to other customers having a similar profile.

Through various supported media, you may also reach members of the following markets:

- Electoral voters
- Financial services clients
- Grocery store customers
- Hospital patients

This chapter explains how BIRT Analytics supports planning, creating, managing, and automating activities that comprise sophisticated business campaigns.

Configuring campaign elements

A campaign includes the following elements:

- Workflows that define campaign roles
- Permissions required to perform campaign tasks
- Stages that group tasks in a workflow
- Cells that define a campaign activity for a specific segment
- Properties that define specific activities
- Segments of data on which cells operate
- Media appropriate to communicate with the segment
- Resolution tables for history and response records
- Strategies that identify campaign goals

To configure a campaign using BIRT Analytics, use Campaign Workflow to define workflows, stages, permissions, and a resolution level for a campaign.

Creating a campaign workflow

A workflow defines a role required to complete campaign tasks and activities. A workflow includes one or multiple stages and users. In a workflow stage, each user is assigned permissions required to complete tasks. For example, the role Campaign Manager may

include two users, one with permission to initiate an activity such as generating messages to a group of customers. In the same workflow, another user may have permission to receive responses and initiate new messages.

How to create a workflow

- 1 Choose Workflow—Campaign workflow.
- 2 In Campaign Workflow, select Configuration, and choose Workflow.
- 3 In Workflow, choose Create, as shown in Figure 7-1.
- 4 In Workflow detail—Name, type a name of a role that completes campaign tasks, such as Campaign Manager.
- 5 In Workflow detail—Description, describe the purpose or goal of this workflow. For example, name and describe the goal of Campaign Manager.

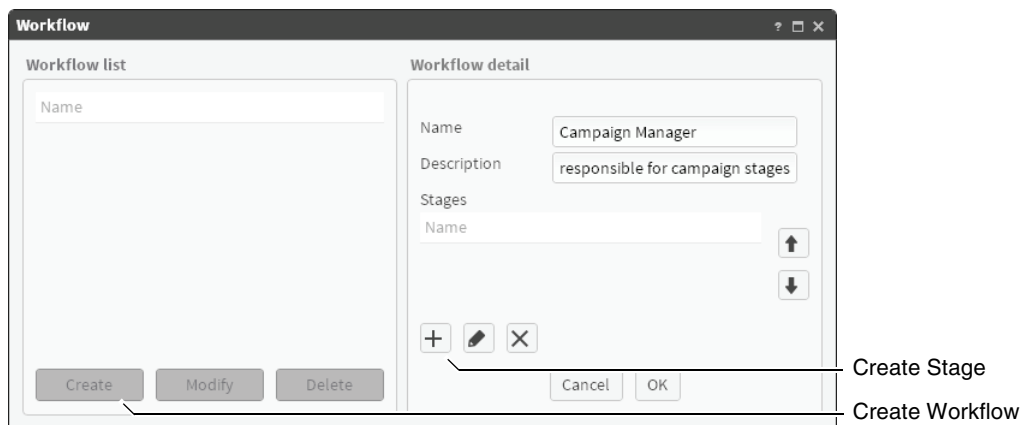


Figure 7-1 Creating a new workflow for a campaign

- 6 Create at least one stage in the workflow. You must create all stages in a workflow before executing a campaign.
- 7 In Workflow, choose OK.

Creating a stage

A workflow includes at least one stage. A stage supports defining users as task owners and assigning to each task owner the permissions required to perform tasks. Using multiple stages supports grouping sets of tasks and activities in a workflow. You must define all stages in a campaign workflow before executing the campaign. Executing a campaign locks the campaign workflow, preventing modifications.

Use stages to organize a workflow. For example, a typical workflow has two stages:

- Stage1, Define/Validate
- Stage2, Load/Response

Different users may have defined tasks in each stage. Organizing a workflow using a logical sequence of stages supports one role, such as Campaign Manager, which is responsible for completing tasks in multiple stages of a campaign.

For example, in a typical company, one group owns data collection tasks, another group owns data validation tasks, and another group owns data modeling and analysis tasks. A workflow uses stages to sequentially group tasks in a campaign. Stages support timely completion of tasks triggered by time and tasks triggered by events.

How to create a stage in a workflow



- 1 In Workflow detail, choose Create.

- 2 In New Stage—Name, type a name that identifies the stage in the workflow.
- 3 In Users, for at least one user, right-click and choose Permissions, as shown in Figure 7-2.

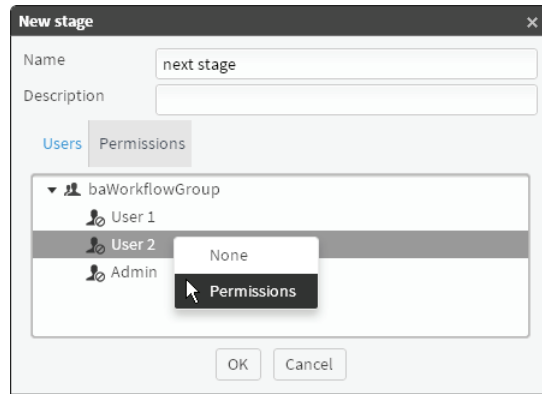


Figure 7-2 Defining a user in a stage
In New Stage, choose OK.

About assigning permissions

A workflow requires at least one stage. For practical purposes, you must define users and permissions sufficient to complete tasks in each stage. To demonstrate assigning permissions and users in different workflow stages, consider the following procedure, for example.

To create three stages in a workflow called Campaign Manager:

- 1 Create stage1, called generate messages.
- 2 In stage1, assign users Admin and User1 all permissions except Execution, Response, and Validation actions. To assign all permissions for a stage to one user:
 - 1 In New Stage, in Users, select a user, then choose Permissions.
 - 2 In Permissions, select any permission name, then choose Select.
 - 3 To deselect a permission, in Selected for a permission name, double-click the green check mark.
 - 4 To assign a user all permissions except three, select all permissions, then double-click each of those three, as shown in Figure 7-3.

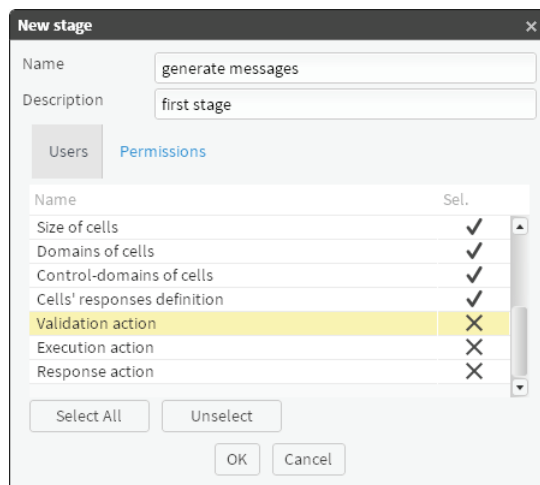


Figure 7-3 Adding user permissions to a new stage

- 3 Create stage2, called Validate.
- 4 In stage2, assign a user only the Validation action permission.
- 5 Create stage3, called Execution and Response.
- 6 In stage3, assign Execution action and Response action permissions to Admin and User2.
- 7 In Workflow, choose OK.

Defining a campaign resolution level

The resolution level for a campaign identifies a database column in which records used to generate campaign activities are stored. For example, a campaign that contacts all customers typically uses customer ID records stored in the Customers table to generate activities. Defining a resolution level for a campaign links history and response records generated by the campaign to the records used to generate campaign activities.

How to define a resolution level for a campaign

- 1 In Campaign Workflow, select Configuration and choose Resolution level.
- 2 In List of levels, choose Create.
- 3 In Level detail, type a name that describes a column in which to store the history and response data for the campaign.
- 4 From My Data, drag a column on which to resolve the campaign and drop it in Record Identifier, as shown in Figure 7-4. The column selected as Record Identifier must be discrete. In other words, each value must appear in only one column record.

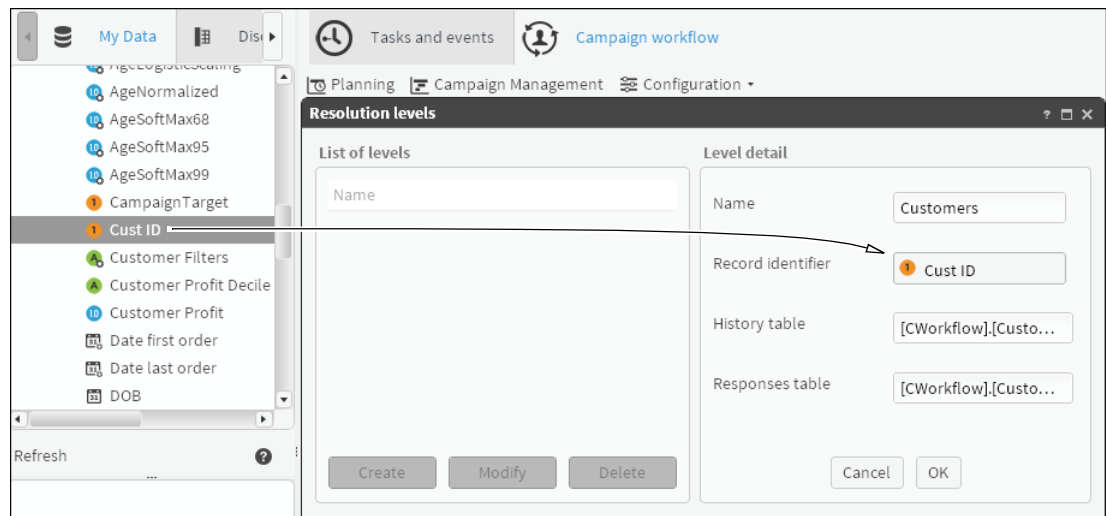


Figure 7-4 Creating a resolution level for a campaign

- 5 Choose OK.

Column names for storing campaign history and response records appear in a new table named CWorkflow, as shown in Figure 7-5.

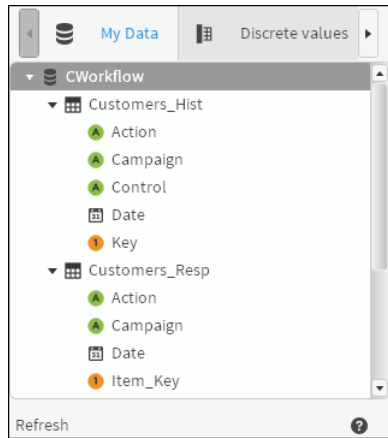


Figure 7-5 Creating a resolution table for a campaign

- 6 In Resolution levels, choose OK.

Defining a media condition

A media condition associates a media type and a selection or segment of records. Defining specific media conditions supports assigning a specific media condition to a campaign cell. For example, associating the email media type and a selection of valid email address records enables a cell to send email messages to valid recipients.

How to define a media condition

- 1 In Campaign Workflow, select Configuration→Media.
- 2 In Media list, choose Create.
- 3 Define the following media details, as shown in Figure 7-6:
 - In Name, type characters that establish a media name for selection in a campaign cell.
 - In Description, describe this media type and purpose.
 - Select a listed icon to associate with this media name.

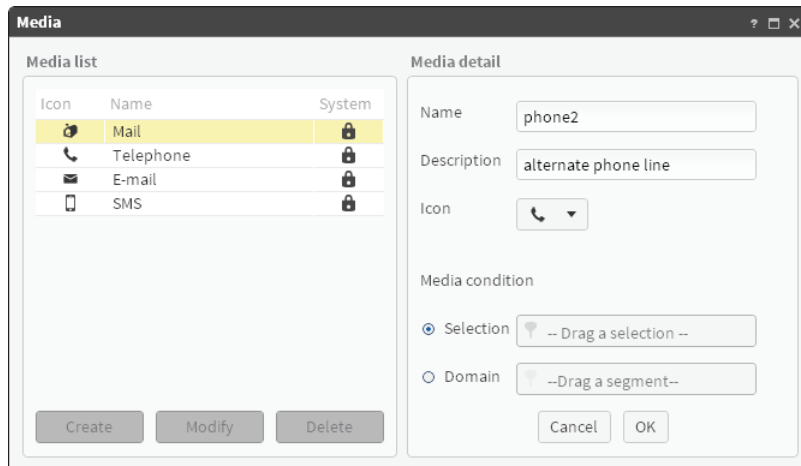


Figure 7-6 Defining media details

- 4 In Media condition, define a media condition using one of the following options:
 - Choose Selection. Drag a selection from Data Tree and drop it in Selection.
 - Choose Domain. Drag a segment from Data Tree and drop it in Domain.

For example, a media condition can limit a specific phone to call only the first one hundred phone records. Create a segment containing records 1 through 100 in [Customers].[Telephone]. Drag the First 100 Phone Numbers segment and drop it in Domain, as shown in Figure 7-7.

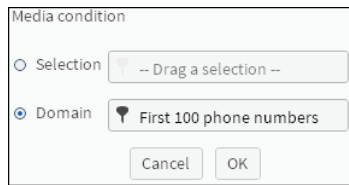


Figure 7-7 Defining a media condition

Choose OK.

Defining an action goal

You can associate an action with each campaign cell that describes a goal for the cell. In Campaign Workflow, action is a descriptive term only. Campaign Workflow includes the following default actions:

- Marketing
- Informative
- Collect

For example, assign the Informative action to a cell that has a goal of sending advertising messages. Alternatively, define a new action goal, such as Prospect.

How to define an action goal

- 1 In Campaign Workflow, select Configuration→Actions.
- 2 In Actions list, choose Create.
- 3 Define the following action details, as shown in Figure 7-8.
 - In Name, specify an action name for selection in a campaign cell.
 - In Description, specify a goal for this action.

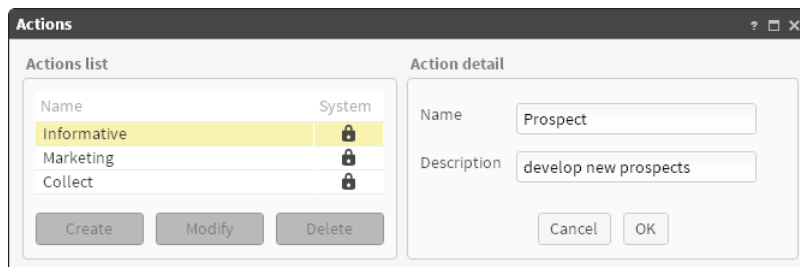


Figure 7-8 Defining action details

- 4 Choose OK.

Planning a campaign

Planning a campaign involves organizing a set of goal-oriented activities. Creating campaign structure by defining a strategy and cells establishes a manageable path toward completing the campaign. Defining campaign properties completes the campaign structure. Campaign workflow planning supports completing multiple activities toward a campaign goal.

This section describes how to define campaign strategies, properties, and cells.

About campaign properties

You define a campaign by setting the following properties:

- **General Data**
A code, name, and description that identify a campaign
- **Dates**
Calendar dates on which activities in a campaign begin and end
- **Resolution**
Database table that stores records on which a campaign runs
- **Domains**
A selected segment of data on which a campaign runs
- **Responses**
Records generated as a result of campaign activities

Creating a strategy

A campaign strategy names and describes one or multiple campaigns. A strategy supports grouping multiple campaigns. BIRT Analytics Campaign Workflow Planning supports creating a branched hierarchy containing strategies and campaigns.

How to create a strategy for a campaign

- 1 In Campaign Workflow, select Planning.
- 2 Right-click Strategies and choose New Strategy.
- 3 In New Strategy, type a name and description of a unique strategy, as shown in Figure 7-9.

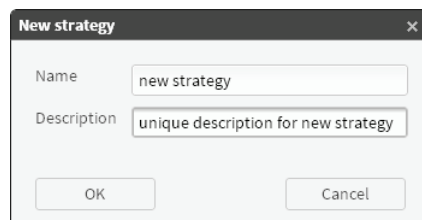


Figure 7-9 Creating a new strategy for a campaign

- 4 Choose OK.

Creating a campaign

Creating a campaign requires a strategy. As a best practice, establish a strategy, then create one, or multiple campaigns within the scope of that strategy. Modify the scope of a strategy as necessary, to accommodate expanding campaign activities.

How to create a new campaign

- 1 Choose Campaign Workflow—Planning.
- 2 In Planning, right-click a strategy and choose New Campaign.
- 3 In Campaign Detail—Campaign’s Properties—General Data, shown in Figure 7-10, type the following information:
 - In Code, specify only one campaign.

- In Name, type a name for this campaign code.
- Optionally, describe the campaign name and code.

Figure 7-10 Adding general properties for a new campaign

4 In Campaign's Properties—Dates, shown in Figure 7-11, select the following options:

- In Start Date and End Date, select values representing dates on which campaign activities begin and finish.
- Optionally, select Number of times. Then, select a number to set how many times to repeat all campaign activities between the defined start and end dates.

Figure 7-11 Adding a date range for a new campaign

5 In Campaign Properties—Resolution, select a resolution level, as shown in Figure 7-12.

Figure 7-12 Selecting a resolution level for a new campaign

6 Optionally, to limit the resolution level of a campaign, select De-duplicate and choose a column name from the list.

For example, consider a campaign designed to contact all customers by setting the resolution level to Customers table. To limit customers contacted to one-per-household, limit the campaign resolution by selecting De-duplicate and selecting Household ID.

7 In Campaign's Properties—Domains, define any selections to include or exclude in the campaign, as shown in Figure 7-13.

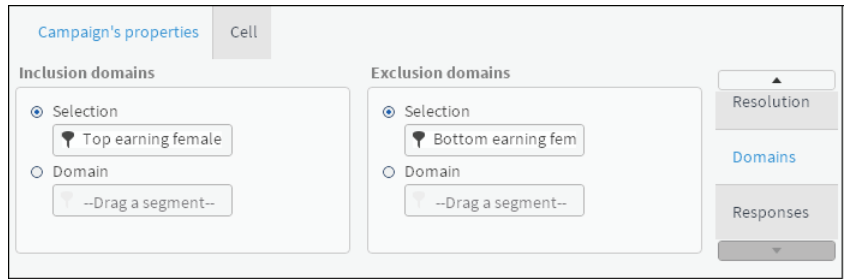


Figure 7-13 Defining a selection to exclude for a new campaign

- 8 In Campaign's Properties—Responses, select Deadline for Responses. Then, select a date value to set a deadline for responses to be included in this campaign, as shown in Figure 7-14.



Figure 7-14 Setting a response collection deadline for a new campaign

- 9 In Campaign detail, choose Save.

Saving a campaign locks property modification for all users except the user creating the campaign.

How to unlock a campaign

To unlock a campaign for editing by other users, choose Check in, as shown in Figure 7-15.

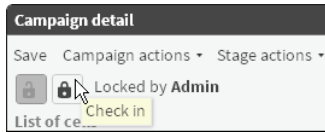


Figure 7-15 Unlocking a campaign for modification

No blocked appears for an unlocked campaign, as shown in Figure 7-16.



Figure 7-16 Examining an unlocked campaign

How to exclude a segment of records from a campaign

- 1 In Campaign Properties—Domains, select Domain in Exclusion domains, as shown in Figure 7-17.

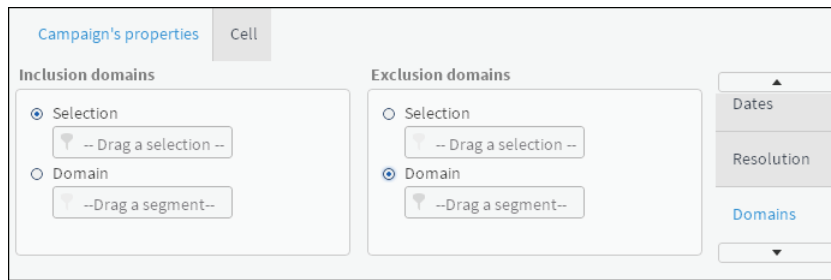


Figure 7-17 Selecting exclude domain for a campaign

- 2 From Scratchpad, drag a data segment and drop it in Exclusion domains—Domain.

For example, to exclude all male directors from a campaign, create a segment that includes only male directors. Then, drag the Male Directors segment from Scratchpad and drop it in Exclusion domains, as shown in Figure 7-18.

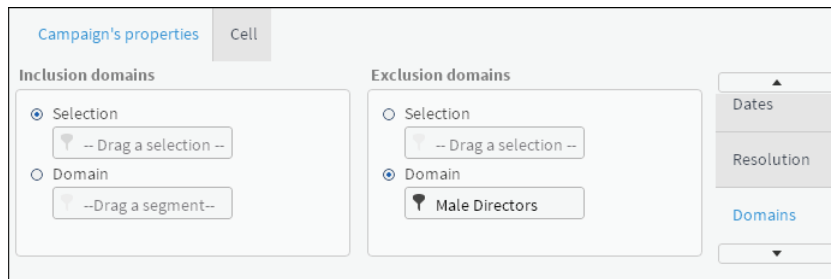


Figure 7-18 Excluding a segment from a campaign

In Campaign detail, choose Save.

How to define a segment as a campaign domain

- 1 In Campaign Properties—Domains, select Domain in Inclusion domains, as shown in Figure 7-19.

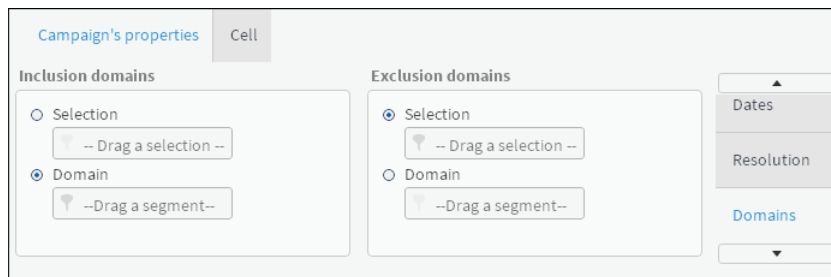


Figure 7-19 Selecting include domain for a campaign

- 2 From Scratchpad, drag a data segment and drop it in Inclusion domains—Domain.

For example, to define the domain of a campaign to include only those customers who buy product MA:

- 1 In My Data, find the Product Group MA value in the Order detail table.
- 2 Drag Product Group MA and drop it in Scratchpad, as shown in Figure 7-20.

Value	Count
HI	19,241
LI	157,449
M	14,403
MA	180,804
MI	120,114
OX	86,982

Figure 7-20 Selecting a segment for a campaign domain

- 3 Change the resolution of Product Group EQ MA segment to Customers, as shown in Figure 7-21.
 - In Scratchpad, right-click Product Group EQ MA and choose Change resolution level.

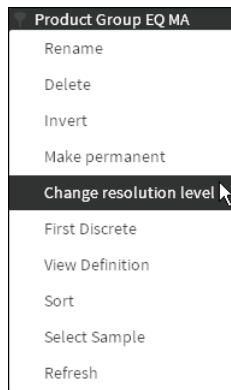


Figure 7-21 Changing the resolution level for a segment

- In Change resolution level, select Customer table and choose OK.
- Rename the segment Customers who buy MA, as shown in Figure 7-22.

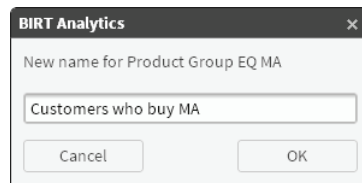


Figure 7-22 Renaming a segment

- 4 In Campaign Properties—Domains, in Inclusion Domain, select Domain.
- 5 Drag a segment from Scratchpad and drop it into Inclusion domains—Domain, as shown in Figure 7-23.



Figure 7-23 Including a segment as a campaign domain

- 3 In Campaign detail, choose Save.

About campaign cells

A campaign cell defines an action executed according to defined properties for all records in a segment. The action is defined to accomplish a specific goal within a campaign. For example, you can set cell properties that generate a specific text message to each male customer who purchased an Android phone. A cell may be set to execute at a specific time or coincident with a specific event during a campaign. A cell also may collect response data associated with records in a segment, according to business rules defined as cell properties.

To modify or delete an existing cell, in Campaign Detail—List of cells, select a cell and choose Modify or Delete, as shown in Figure 7-24.



Figure 7-24 Modifying a campaign cell

How to create a cell in a campaign

- 1 In Campaign Detail, in List of cells, choose Create.
- 2 In Cell—General data, provide descriptive information for the cell, as shown in Figure 7-25.
 - 1 In Code, specify a cell for the campaign.
 - 2 In Name, type a name for this campaign cell.
 - 3 In Description, describe this campaign cell.
 - 4 In Media, select a media type.
 - 5 In Action, select an action type.

Figure 7-25 Adding descriptive information to a campaign cell

- 3 In Cell—Sizes, limit the size of the campaign using the following steps, as shown in Figure 7-26:
 - 1 In Minimum count, set a low value in the range of records for which the cell executes.
 - 2 In Maximum count, set a high value in the range of records for which the cell executes.
 - 3 In Sample method, select the type of method by which the cell selects records.
 - 4 In Remaining, select one of the following options:
 - Pass on. This option saves cell result records for use in subsequently executed cells.
 - Exclude from others. This options prevents saving cell result records.

Figure 7-26 Adding size and sampling properties to a campaign cell

- 4 In Cell—Domains, define the campaign cell domain. To specify a selection or domain for the cell, drag a selection or a segment from Data Tree and drop it in Selection or Domain, as shown in Figure 7-27.

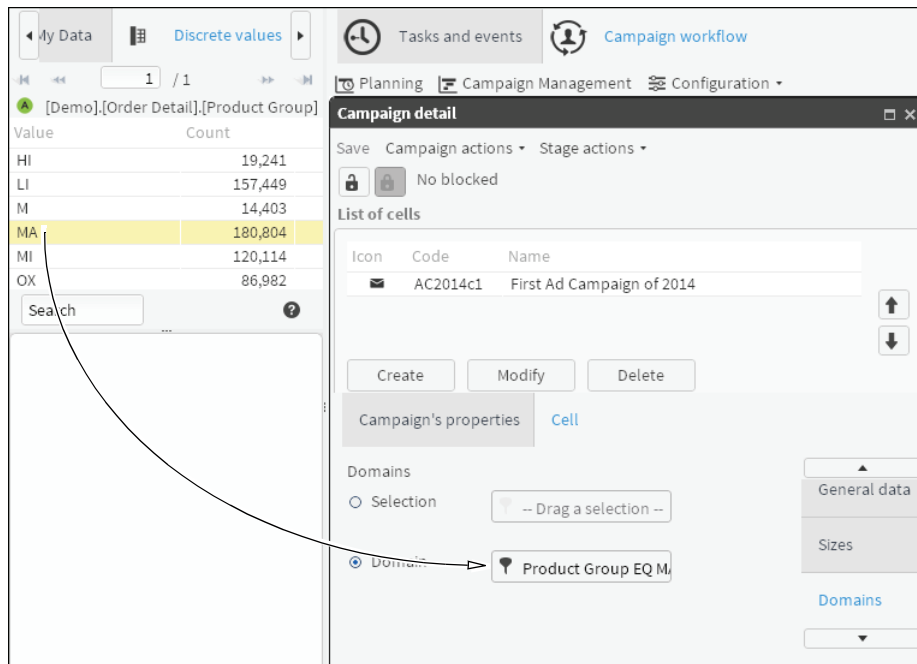


Figure 7-27 Defining the domain for a campaign cell

- In Cell—Control, define a control cell and set parameters. Select a control type, sample size, and sample method, as shown in Figure 7-28.



Figure 7-28 Defining a control cell

To determine the effectiveness of a cell, compare cell results to the control cell.

- Optionally, in Cell—Responses, select Implied response rules. Then, define a business rule using a segment from which responses generate, as shown in Figure 7-29.

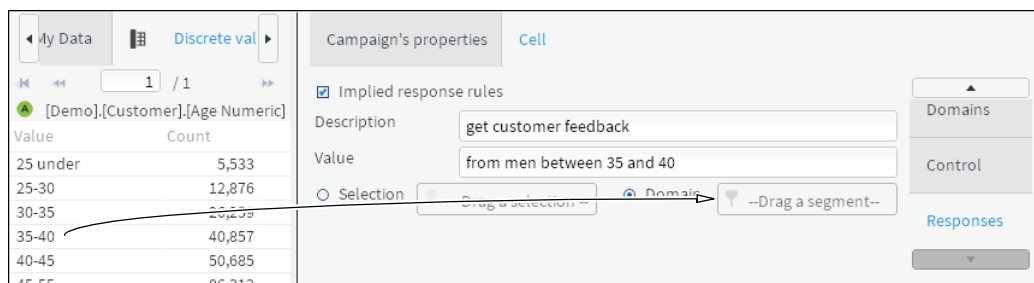


Figure 7-29 Adding a response rule to a campaign cell

- Choose Save.

Running a campaign

Running a campaign includes the following activities:

- Planning
- Configuring
- Starting
- Managing
- Executing
- Reviewing

This section describes how to start, manage, and execute a campaign that is already planned and configured.

Starting a campaign

Before starting a campaign, lock the campaign to block any changes to campaign properties. Then, start the stage and workflow that contains the first campaign activity or action.

How to start a campaign

- 1 Choose Campaign Workflow—Planning.
- 2 In Planning, expand Strategies.
- 3 Select and expand a specific strategy, if necessary.
- 4 In a specific strategy, right-click a campaign and choose Modify campaign.
- 5 To prevent changes to campaign properties, in Campaign detail, choose Lock.
- 6 Select Stage actions and choose Start campaign.
- 7 In Workflow, select the workflow that includes the first action in the campaign.
- 8 Choose OK.

The current workflow and stage appear in Campaign detail, as shown in Figure 7-30.

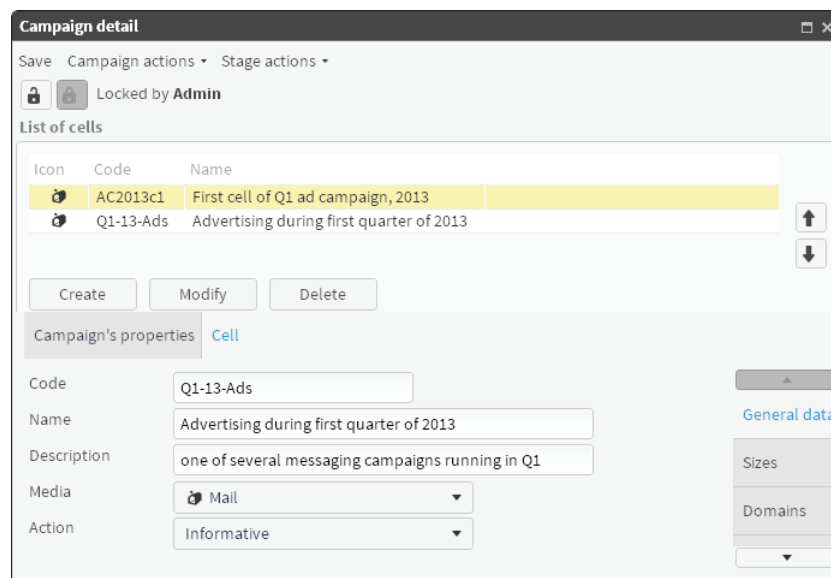


Figure 7-30 Examining the Campaign Manager workflow in stage 1 of a campaign

Managing campaign stages

You can manage and monitor campaign stages and workflows manually. You can also automate many actions in a campaign by definition. The following procedure summarizes the processes completed in a typical campaign.

- 1 Start campaign.
- 2 Start initial workflow.
- 3 Start initial stage.
Cells defined in a stage execute, according to defined properties.
- 4 Start next stage.
- 5 Start next workflow.
- 6 Repeat steps 4–5 until all tasks in all stages and workflows have completed.
- 7 Execute campaign.
- 8 Evaluate campaign results.

How to manage stages in a campaign

- 1 In Campaign detail, select Stage actions and make one of the following choices:
 - To start a subsequent stage, choose Next stage.
If you do not have permissions required to execute a stage, choose Accept in the message describing the permissions you lack.
 - To review a preceding stage, choose Previous stage.
 - To review a list of transitions between stages, choose Show History.
- 2 To close Campaign History, choose OK.

How to evaluate a campaign

- 1 Log in to BIRT Analytics as a user who has the Evaluate permission.
- 2 In Campaign detail, select Campaign actions and choose Evaluate.
- 3 In Results of the evaluation, examine the following levels of information:
 - To see details of records generated during completed stages, choose Table.
 - To see a summary of records generated by each completed cell, choose Chart.

Viewing campaign summaries

Campaign management supports viewing summary information using a Gantt chart display. This view presents a time-line background on which multiple campaigns appear. Options include viewing a summary of all campaigns, or only campaigns having current activity.

How to view campaigns on a Gantt chart

- 1 Choose Campaign Workflow—Campaigns Management.
- 2 In List of campaigns, select All campaigns, then choose Refresh.
- 3 To view information about a selected campaign, select one of the following options:
 - To limit campaigns appearing on the Gantt chart to only active ones, in List of campaigns, select Current.
 - To view information about a campaign, hover the cursor over a campaign icon on the chart.

- To view details for a campaign, double-click the image for a single campaign on the Gantt chart.

Executing a campaign

Executing a campaign completes the set of campaign activities. Executing a campaign loads all data generated to date by campaign activities into the history and response database tables defined for the campaign.

How to save campaign results and responses in a database table

- 1 After all stages in all workflows complete, select Campaign actions and choose Execute.
- 2 In Execution options, to manage loading data in the Workflow history table, select one of the following options:
 - To create a file containing all campaign result records, select Just generate file.
 - To load campaign result records in your database, select Generate file and load it on history table.
 - To delete cached result records after loading them in the history table, select Generate file, load it and remove it.
- 3 To generate response records in a file, select Campaign actions and choose Response.

How to review campaign results

- 1 To explore a column in CWorkflow, drag the column from My Data, or a segment from Discrete Values and drop it in Data Explorer.
- 2 To view summary of history records for this campaign in Data Explorer, choose Summary.
- 3 To view details about the response records for this campaign in Data Explorer, choose Record view.

8

Scheduling tasks

This chapter contains:

- Automating a task
- Managing scheduled tasks

Automating a task

BIRT Analytics supports automating tasks that initiate specific actions. A task initiates an action based on a trigger of one of the following types:

- Time trigger
A task having a time trigger runs at a set time.
- Event trigger
A task having an event trigger runs when a selected event occurs.

To automate a task, create a scheduled task to execute when the set trigger occurs.

About event types

Available events vary based on plug-ins installed. For each event, you specify a condition when you configure the event, including an attribute, an operation, and a value.

For example, you can create the following event condition:

```
Full Name = Michael
```

In this event, a task triggers each time the value Michael appears in the field Full Name.

Available events appear in the following groups:

- BIRT Analytics
 - Engineering
 - Modified definition
 - Delete repository item
- Data Mining
 - Apply model
- Campaign Workflow
 - Campaign executed
 - Cell executed
 - Step change

Typical event attributes are:

- User: User name
- Date CCYYMMDD
- Hour HHMMSS
- Engine data format
Date in analytical engine format. This attribute has to be used in SQL queries that invoke date or date-time fields.

About action types

BIRT Analytics supports communication, query, program execution, data analysis and data modeling actions.

Available actions appear in the following groups:

- Sending

- Send email

Sends a message via email to one or multiple recipients. To create a Send email action, you configure the following details: e-mail contents, event trigger, or time trigger attributes. Right-click in the Content editor to display available content options. Content options vary, based on the task trigger. Common examples of email task triggers include:
- Send execution results

Sends an email with campaign execution results. This mail can be delivered to each user who has been involved in a campaign execution process.
- Notify users via email

Sends an email to all the users related to a campaign triggering a task. This email will be sent to users having permissions for the current campaign stage and a valid email account on the database.
- Conditional

Query. A query action evaluates whether data matches a condition. Correct query syntax follows:

```
action, filter, operator, value
```

where

 - action is a function such as count, sum, or mean
 - filter is a database domain or SQL statement
 - operator is >, <, or =, for example
- Execute
 - Execute campaign

Runs a campaign execution.
 - Execute response load

Loads all campaign responses associated to an event in a new database table.
 - Export campaign cell

Exports all records from a selected cell in a campaign. Sets the event or campaign date of execution, the records to export (all the records, all the records except control, only control) and an exportation report that will be used as a template for exporting.
- Data Model
 - Delete column

Deletes a column. User must indicate entire column name [database].[table].[column].
 - Delete table

Deletes a table. User must type the full table name, for example, [database].[table].
- Apply
 - Apply Model

Applies the Decision Tree or Clustering data mining model. The Decision Tree or Clustering Analysis must be properly configured, trained and saved.
 - Apply model online

Set an action to apply the model in a deferred way.

In either case, the result is a column containing results of the applied model.

Creating a scheduled task

A scheduled task includes a trigger, task details, and an assigned action. The following procedure summarizes high-level steps for creating a scheduled task.

- 1 Create a new scheduled task.
- 2 Configure conditions that trigger the task.
- 3 Assign an action to the scheduled task.
- 4 Save the scheduled task.

How to create a task triggered by time

- 1 Choose Workflow—Tasks and events—New Scheduled Task.
- 2 In Scheduled task—Name, type characters that name the task.
- 3 Optionally, describe and activate the task, using the following options:
 - In Description, type characters that describe the task.
 - Select Starting date and Ending date. Then, use the date and time selector to specify a range of dates during which this task trigger is active.
- 4 In Trigger for task, expand Scheduled task on time.
- 5 Drag Time to the right and drop it in the box.
- 6 In Time, select values for each time parameter using the selector, as shown in Figure 8-1.

For example, double-click the selector icon for Month. In Month, select values for the minute, hour, day, month, and day-of-week that specify when the task triggers. Choose OK. Not every field requires a value.

Default values are: minute = 0, hour = 1, and current day. Default settings trigger a task on this day of each month at the hour and minute 0.

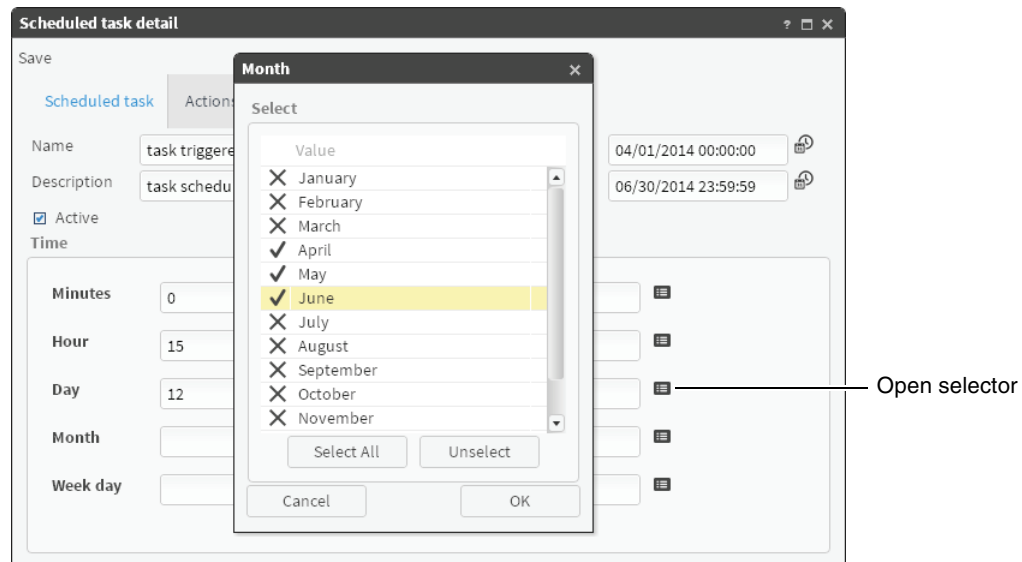


Figure 8-1 Setting time properties for a scheduled task

- 7 Choose Save.

How to create a condition for an event trigger

- 1 In Scheduled task detail, in Trigger for task, expand Scheduled task on event.
- 2 Drag an event and drop it as shown in Figure 8-2.

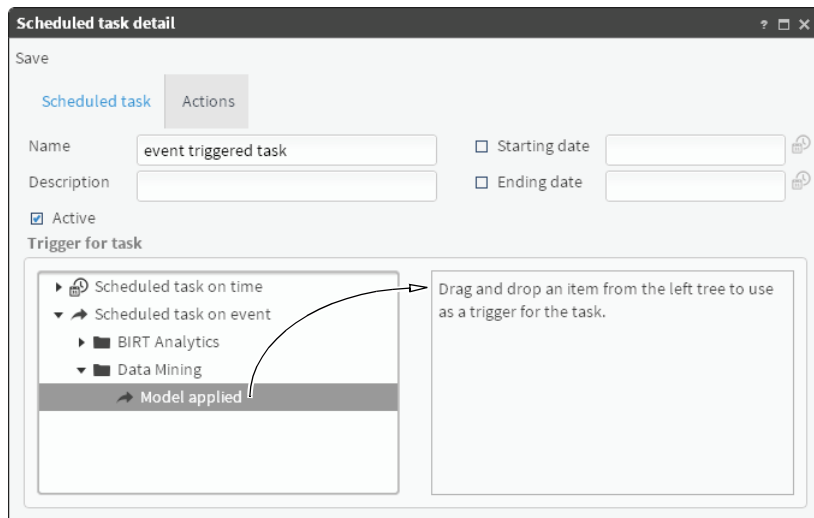


Figure 8-2 Creating a trigger for an event-driven task

- 3 Drag a listed attribute and drop it in Conditions, as shown in Figure 8-3.

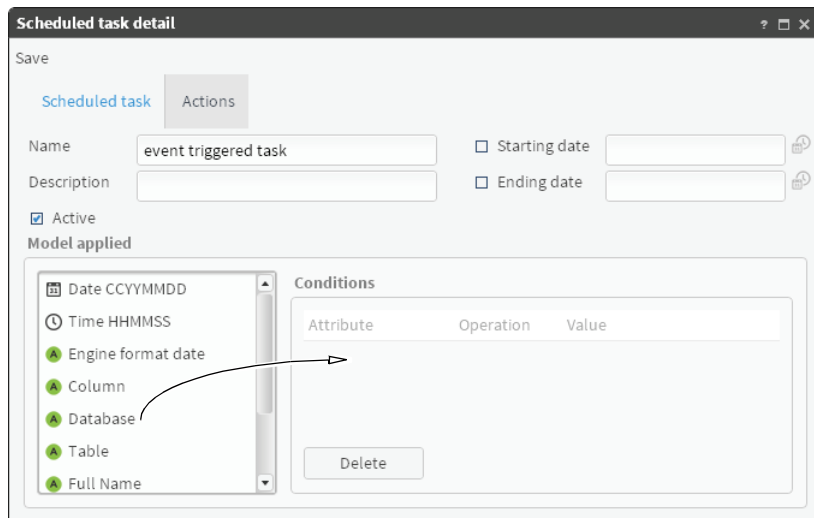


Figure 8-3 Creating a condition that triggers an event-driven task

- 4 In Conditions detail, select an operation and type a value for the selected attribute, as shown in Figure 8-4.

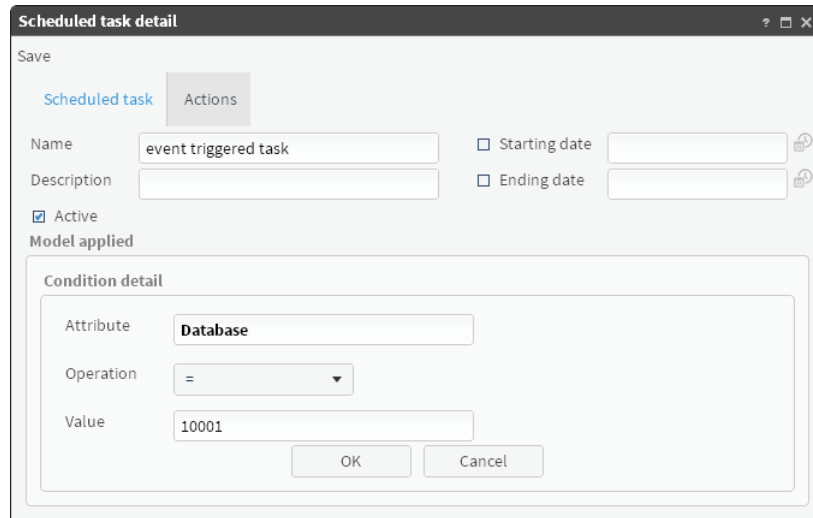


Figure 8-4 Configuring condition details for an event trigger

5 Choose OK.

How to assign an action to a scheduled task

- 1 In Actions, expand an action group.
- 2 Drag an action name and drop it in the Actions visual editor, as shown in Figure 8-5.

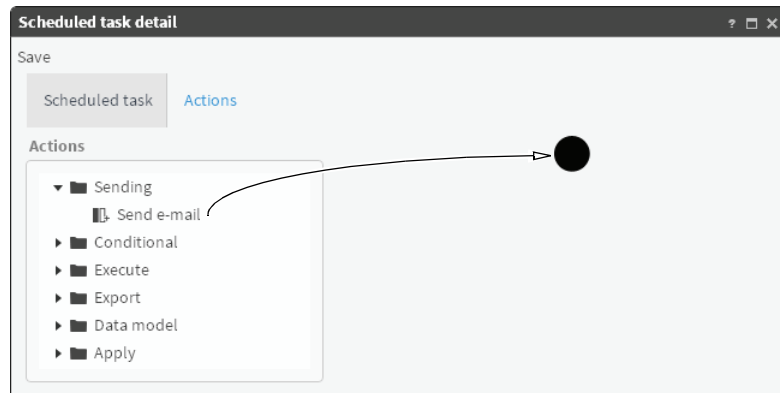


Figure 8-5 Dropping an action in Actions visual editor

3 In Action detail, define action properties using tools provided for a selected action type.

For example, to create a Send email action, provide values for the following properties, using the toolset provided in Action detail editor, as shown in Figure 8-6.

- In To..., type a valid email address.
- Choose the check mark to validate an email address.
- In Subject, type a subject for an email message.
- In Format, select HTML or Plain Text.
- In Content, type characters that comprise an email message.

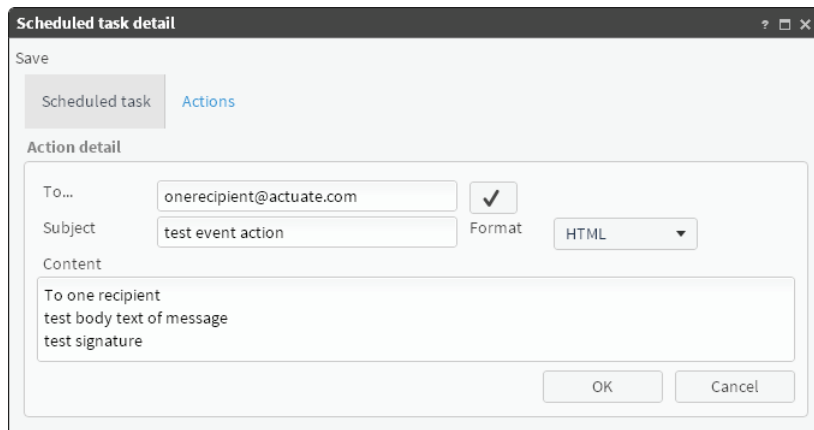


Figure 8-6 Defining properties for a Send email action

- 4 Choose OK. An action appears in the Actions visual editor, as shown in Figure 8-7.

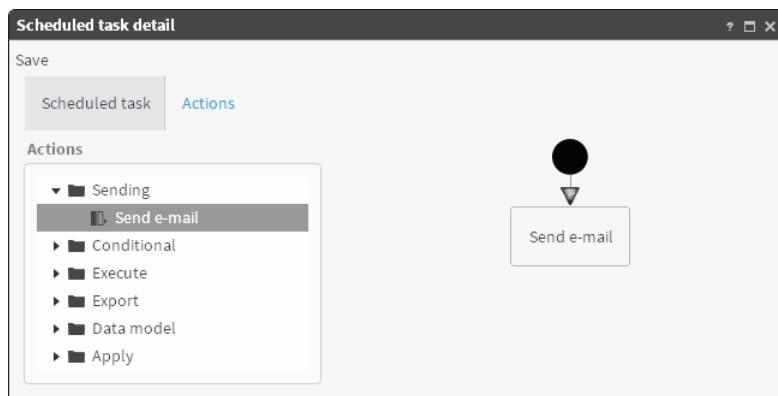


Figure 8-7 Examining the visual actions editor

How to edit action details

- 1 In Actions visual editor, double-click an action.
- 2 Action detail opens editing tools appropriate for the selected action.

How to remove an action from a scheduled task

- 1 In Actions, right-click an action in the visual editor.
- 2 Choose Delete, as shown in Figure 8-8.

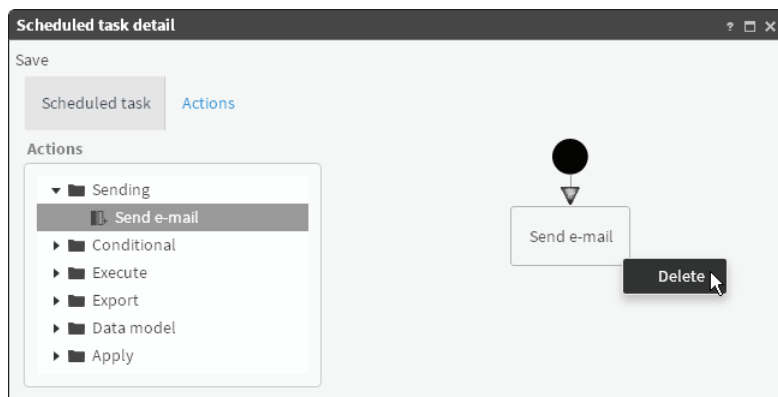


Figure 8-8 Removing an action from a scheduled task

Managing scheduled tasks

To view all scheduled tasks in BIRT Analytics, choose Workflow—Tasks and events—Scheduled tasks. Scheduled tasks appear, sorted by creation date and time in ascending order. A selected task appears highlighted, as shown in Figure 8-9.

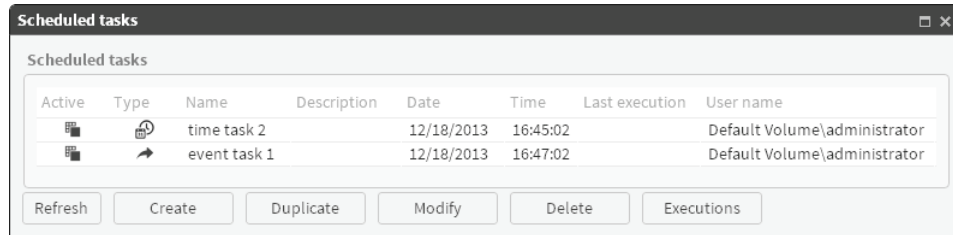


Figure 8-9 Selecting one scheduled task

Use Scheduled tasks to complete any of the following activities:

- To view current status of all scheduled tasks, choose Refresh.
- To create a new scheduled task having no properties, choose Create.
- To create a new scheduled task having the same properties as an existing one, select a task and choose Duplicate.
- To edit an existing scheduled task, select a scheduled task and choose Modify.
- To remove a scheduled task from BIRT Analytics, choose Delete.
- To view a list of scheduled tasks already completed, choose Executions.

Duplicating a scheduled task

Duplicate creates a new scheduled task having all properties of an existing, selected task. Duplicating a scheduled task, then modifying one or multiple properties in the duplicated task, saves time spent recreating identical properties for a new task. Only the time attribute, indicating the time at which a task is created, is unique for a new, duplicate task.

How to duplicate a scheduled task

- 1 In Scheduled tasks, select a task.
- 2 Choose Duplicate.
- 3 In Scheduled task detail, choose Save.

A new task having all properties of the original task appears in Scheduled tasks. The new task has a unique value in Time, as shown in Figure 8-10.

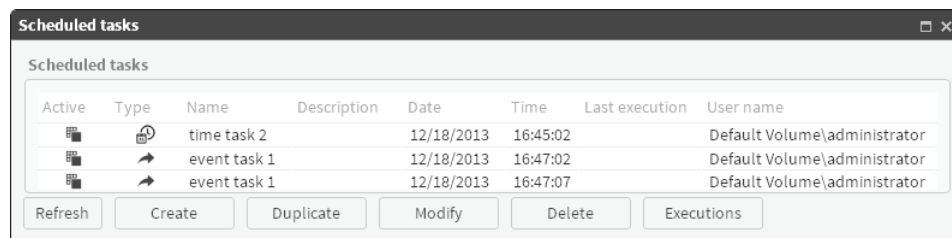


Figure 8-10 Examining a duplicated scheduled task

Modifying a scheduled task

You can modify properties of an existing, scheduled task before the time or event set as a trigger for that task occurs. Scheduled task details supports editing and saving changes to task properties.

How to modify a scheduled task

- 1 In Scheduled tasks, select a task from the list.
- 2 Choose Modify.
- 3 Using Scheduled task details, change at least one property of the task. For example, edit the name property.
- 4 Choose Save.

Using a conditional query to automate actions

Action properties support associating specific actions with defined trigger event conditions. Automating a notification process provides a basic example for associating event-triggered, scheduled tasks and unique trigger events. A basic example demonstrating how to define unique actions that result from evaluating a conditional query follows.

The following example task generates one of two unique email message actions based on evaluating an order quantity in a customer database.

How to specify actions based on query results

- 1 In a scheduled task, create an event trigger having conditions that monitor the customer order table:
 - 1 In Trigger for task, expand BIRT Analytics. Then drag and drop an Engineering event type in the scheduled task editor.
 - 2 In Scheduled task detail—Scheduled task, define conditions and attributes that identify a specific column, as shown in Figure 8-11.

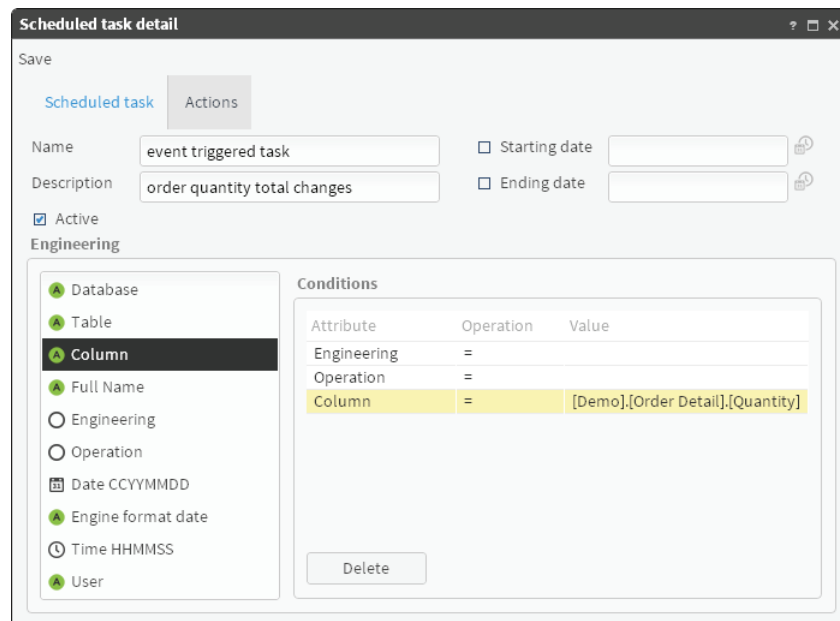


Figure 8-11 Defining specific conditions for a scheduled task

- 2 In Actions for the event-triggered task, expand Conditional. Drag Query and drop it in the visual editor.

- 3 Specify query conditions for an action using Action details. For example, the selections shown in Figure 8-12 compare a quantity in a specific field with a defined value. The result returned by the conditional query is Yes or No. Choose OK.

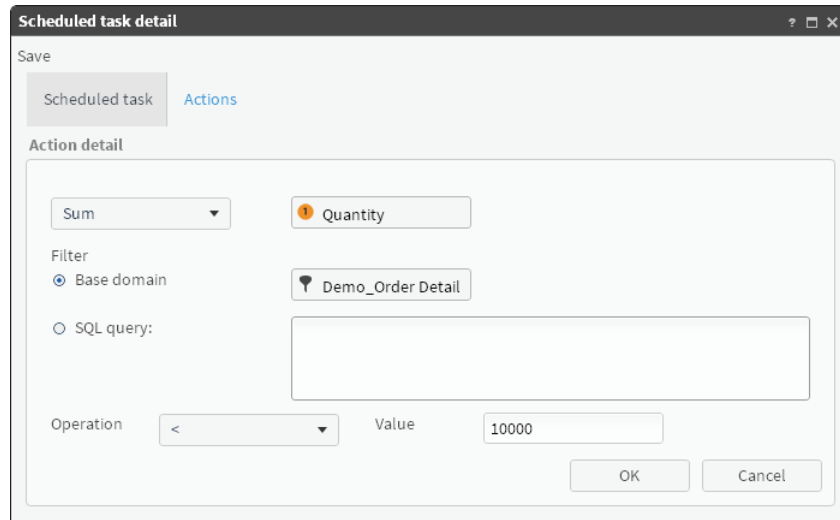


Figure 8-12 Specifying query conditions for an action

- 4 In Actions, add two Send email actions, using the visual editor, as follows:
 - 1 Expand Sending. Drag and drop a Send email action on Query. Choose Yes.
 - 2 Configure details for an email message action, appropriate for a Yes result.
 - 3 Drag and drop another Send email action on Query.
 - 4 Configure details for an email message action, appropriate for a No result.
- 5 Two Send email actions appear, one for each query condition, as shown in Figure 8-13.

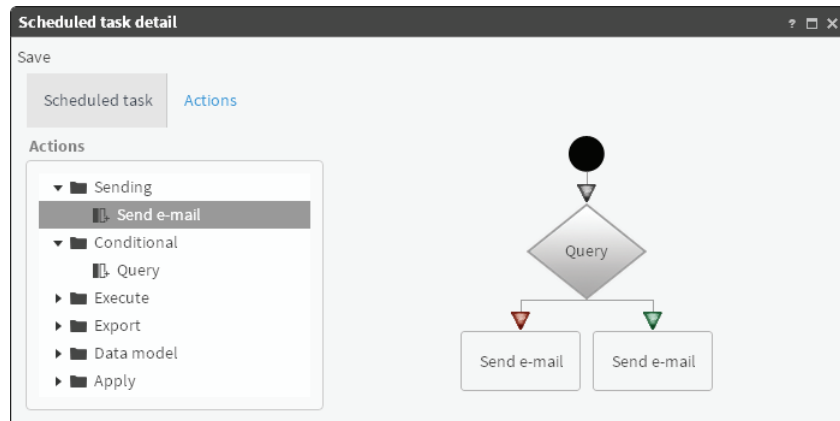


Figure 8-13 Examining a conditional query and associated actions

- 6 Choose Save.

A

access control list (ACL)

A group or set of users with access to a database object. Using the BIRT Analytics Administration tool, the administrator creates a security group or ACL that manages privileges for a database object.

Related terms

BIRT Analytics Administration, column, database, group, security role, table

action

An action is an event executed by a manual or task trigger. Example actions include send e-mail, query action, delete column, and apply model.

Related terms

scheduled task, trigger

Aggregates

A tool that supports grouping data from multiple tables in one table. Aggregates supports defining a function and filter as properties.

Related terms

filter, table

analysis

A tool that provides a specific view of data stored in FastDB. BIRT Analytics supports multiple analyses.

Related terms

Bubble analysis, Crosstab analysis, Evolution analysis, FastDB, Map analysis, Pareto analysis, Profile analysis, Venn analysis

antecedent

Terms representing the left-hand, or If... clause of an association rule. The antecedent clause of an association rule contains discrete data items.

Related terms

association rules, consequent

Association Rules

A predictive analytics tool that uses association rules to identify an If...Then relationship between data values stored in an information repository. For example, an association rule may show the following relationship: If a customer buys products A and B, then the customer also buys product C.

Related terms

association rules, predictive analytics

association rules

A predictive analytics technique that analyzes data for frequent If...Then patterns and calculates support and confidence criteria that identify the most important relationships. Support indicates how frequently the items appear in the database. Confidence indicates the number of times the If...Then relationships evaluate true.

An association rule has two parts, an antecedent and a consequent. The antecedent represents one or multiple data items. The consequent represents an item found in combination with the antecedent. An association rule returns a lift and a leverage value that measure how well the rule predicts the consequent.

Related terms

antecedent, Association Rules, confidence, consequent, lift, leverage, predictive analytics, support

Association Rules

A predictive analytics tool that uses association rules to identify an If...Then relationship between data values stored in an information repository. For example, an association rule may show the following relationship: If a customer buys products A and B, then the customer also buys product C.

Related terms

association rules, predictive analytics

B

baseline filter

A filter that returns a group of records to serve as a basis for comparison. For example, use the year 2012 as a baseline filter for profit, to compare profit earned in another year with profit earned in 2012.

Related terms

filter, record

big data analysis

The practice of analyzing, exploring, filtering, loading, segmenting, and studying massive quantities of data. Big data analysis uses statistics to describe qualities and predict trends in these data repositories.

Related terms

analysis, BIRT Analytics, data repository

BIRT Analytics

An application, including a data repository, data loader, and web service, that supports big data analysis.

Related terms

big data analysis, BIRT Analytics Administration

BIRT Analytics Administration

A BIRT Analytics system administration tool that runs as a browser-based application. The administrative user has full permission to modify all configurable features of the BIRT Analytics system.

Related term

BIRT Analytics

BIRT Analytics Loader module

A tool that extracts, transforms, and loads records from an external data source to FastDB.

Related terms

BIRT Analytics, BIRT Analytics Administration, FastDB

Bubble analysis

A tool that supports viewing a spatial distribution of data with respect to two axes.

Related terms

analysis, Crosstab analysis, Evolution analysis, Map analysis, Pareto analysis, Profile analysis, Venn analysis

C

calculated field

A data field that displays the result of an expression.

campaign

A set of tasks, defined for specific population segment. A campaign is completed during a defined time period to accomplish a specific goal.

Related term

segment

Canvas

A workspace for data analysis gadgets. Canvas supports arranging, assembling, and saving a collection of data visualization gadgets.

Related term

gadget

cell

A set of properties that defines campaign actions to be performed for all records in a segment.

Related terms

action, campaign, record, segment

Clustering

A predictive analytics tool that uses k-means cluster analysis. Clustering identifies groups of similar data values in large segments stored in a big data repository.

Related terms

k-means, cluster analysis, predictive analytics

cluster analysis

A data analysis task that iterates estimating of values assigned to common data attributes. Common attributes identify groups of similar items, called clusters. Comparing clusters highlights similar and different groups in big data.

Related terms

analysis, big data analysis, Clustering

column

1 A named field in a database table or query. For each data row, the column can have a different value, called the column value. The term column refers to the definition of the column, not to any particular value.

2 A vertical sequence of cells in a crosstab, grid element, or table element.

Related terms

column-oriented DBMS, database, data field, query, table

column-oriented DBMS

A column-oriented DBMS is a database management system (DBMS) that stores data tables as sections of columns of data rather than as rows of data. A column-oriented DBMS serializes all of the values of a column together, then the values of the next column, and so on.

Related terms

database, column

confidence

An expression used to identify an association rule. Confidence compares how often the consequent appears when the antecedent is met. The confidence expression has the following syntax:

Confidence (A,B-> C) = Support (A,B,C)/Support (A,B)

Related terms

association rules, support

consequent

Terms representing the left-hand, or ...Then clause of an association rule. The consequent clause of an association rule contains items found in combination with items in the antecedent.

Related terms

antecedent, association rules

Convert A BIRT Analytics option that displays results from one data analysis using a different type of data analysis. For example, an analysis created using Crosstab converts to a Bubble, Evolution, or Map analysis.

Related term
analysis

count The total number of records in a field.

Related terms
field, record

Crosstab analysis

A tool that supports analyzing data using cross-tabulation, or pivoting of different fields.

Related terms
analysis, Bubble analysis, Evolution analysis, Map analysis, Pareto analysis, Profile analysis, Venn analysis

Cylinder A data visualization gadget that displays numeric values and boundaries in ranges. A Cylinder displays defined data measures as colored slices that comprise one cylinder shape.

Related terms
Dial, Funnel, gadget, Gallery, Label, Meter, Sphere

D

data analysis A process including acquiring, organizing, transforming, and modeling data to support decision-making.








Data Explorer

A tool that displays records from a database stored in FastDB. Data Explorer provides a summary view for a table and a detail view for records, tables, selections, and segments.

Related terms
Data Tree, FastDB, record, table

data field A location storing data having a specific type. A data field typically contains data from a database or other data source. A data field appears as a column when viewing a table in Data Explorer. For example, the BIRT Data Analytics Demo database includes the data field types listed in Table G-1.

Table G-1 Data field types

Icon	Field type	Description
	Calculated	Displays a value result from an expression
	Date	Contains numbers that represent day, month, and year
	Date and time	Contains numbers that represent day, month, year, and time of day
	Full numeric	Contains whole, or integer numbers, such as 1 or 1000
	Real numeric	Contains real, or partial numbers such as 1.05 or 0.003
	Time	Contains a value representing time of day
	Text	Contains a string of alphabetic characters

Related terms

record, Data Explorer, Data Tree, data types

data integration

A process through which data in varied sources is combined.

data mining

A computational process used to extract and transform data to prepare it for analysis.

Related term

analysis

data repository

A physical or virtual location for storage and retrieval of data.

Related term

FastDB

Data Tree

A tool that supports viewing and working with databases, tables, and records stored in FastDB. Data Tree includes Discrete Values, My Data, and My Folders viewers.

Related terms

database, Discrete Values Viewer, My Data Viewer, My Folders Viewer, record, table

data types

A data type defines the limits of a data field in a BIRT Analytics database. For example, the BIRT Data Analytics demo database includes the data types listed in Table G-2.

Table G-2 Data types in BIRT Analytics Loader

Data type	Description
Date	Contains numbers that represent day, month, and year. The default format is mm_dd_yyyy.
Datetime	Date and time data from January 1, 1753, through December 31, 9999, providing accuracy to three-hundredths of a second, or 3.33 milliseconds. The default format is yyyy_mm_dd_hh_MM_ss.
Integer	Integer data from -2^{31} (-2,147,483,648) through $2^{31}-1$ (2,147,483,647).
Longint	Integer data from -2^{63} (-9,223,372,036,854,775,808) through $2^{63}-1$ (9,223,372,036,854,775,807).
Real	Floating precision number data with the following valid values: $-3.40E+38$ through $-1.18E-38$, 0 and $1.18E-38$ through $3.40E+38$.
String	A sequence of ASCII characters.
Time	Contains a value representing time of day. The default format is hh_MM_ss.
Unicode	A sequence of characters based on consistent encoding, representation, and handling of text as expressed in global writing systems.

Related terms

Data Explorer, data field, Data Tree, record

database

1 An integrated collection of logically related records that provides data for information application platforms, such as BIRT. The database model most commonly used is the relational model. Other typical models are entity-relationship, hierarchical, network, object, and object-relational.

2 An integrated set of logically related records stored in FastDB.

Related terms

record, table

decision tree

A predictive analytics technique that predicts the value of a target variable, based on values of multiple input variables. For example, use a decision tree to predict a survival rate, based on characteristics of the population that may survive.

Related terms
Decision Tree, predictive analytics

Decision Tree

A predictive analytics tool that uses the decision tree technique to predict an outcome, based on values of multiple input variables. For example, use Decision Tree to predict the product a customer will purchase, based on customer, purchase, gender, occupation, and income data.

Related terms
association rules, predictive analytics

Decodes A tool that supports renaming a data field stored in FastDB.

Related terms
data analysis, data field, FastDB

Dial A data visualization gadget that uses a needle-shaped pointer to display defined measures and numeric values in a range.

Related terms
Canvas, Cylinder, Funnel, gadget, Gallery, Label, Meter, Sphere

Discrete Values Viewer

A tool that supports viewing discrete values in a data record, selection, or segment.

Related terms
My Data Viewer, My Folders Viewer, record, segment, selection

Downloads A tool that supports writing FastDB records to an external database.

Related terms
database, Export file, FastDB, record

Dubnium.exe The file that runs the BIRT Analytics data repository, FastDB.

Related term
FastDB

E

Evolution analysis

A tool that supports viewing a time-progression view of data values.

Related terms
analysis, Bubble analysis, Crosstab analysis, Map analysis, Pareto analysis, Profile analysis, Venn analysis

Export Analytic DB

A tool that supports creating a new database field based on a segment defined in the database. The new field is stored in FastDB.

Related terms
Export file, FastDB, segment

Export file A tool that supports creating a new text file based on a segment defined in the database. The file is stored in FastDB.

Related terms
Downloads, FastDB, segment

Expressions A tool that supports creating a logical relationship, using data fields, functions, and operators. Results of the relationship appear as a calculated field in FastDB.

Related terms
calculated field, data field, FastDB

F

- FastDB** The BIRT Analytics data repository. FastDB is a web service that caches data and supports executing data analysis and forecasting algorithms.
Related terms
Data Tree, database, data repository, record, table
- field** See data field.
- filter** A function that limits the number of records included a segment or selection. BIRT Analytics supports the following three filter types: baseline, target, and universal.
Related terms
baseline filter, target filter, universal filter
- Forecasting** A predictive analytics tool that uses the Holt-Winters, iterative method. Forecasting predicts a future trend in data exhibiting a seasonal pattern.
Related terms
Holt-Winters, predictive analytics
- functionalities**
The system privileges an administrator grants to a security role.
Related terms
BIRT Analytics Administration, security role
- Funnel** A data visualization gadget displaying numeric values and boundaries that represent groups in a range, using colored bands that display on a funnel shape.
Related terms
Cylinder, Dial, gadget, Gallery, Label, Meter, Sphere

G

- gadget** A computer program that provides services without requiring an application for each one. BIRT Analytics provides multiple gadgets that support data visualization.
Related terms
Cylinder, Dial, Funnel, Gallery, Label, Meter, Sphere
- Gallery** A tool that supports running multiple data visualization gadgets. Use the Gallery to assemble, arrange, and save gadgets on the Canvas.
Related terms
Canvas, Cylinder, Dial, Funnel, gadget, Label, Meter, Sphere
- group** A set of users belonging to the same organizational unit who share the same permissions for performing tasks. Using the BIRT Analytics Administration tool, the administrator creates a group from the list of available users on the system.
Related term
BIRT Analytics Administration

H

- has seasonality**
User-selected option that recognizes a seasonal trend in a data set.
Related terms
Holt-Winters, seasonal periodicity, seasonality
- Holt-Winters** A popular numerical estimation method used to forecast values in data that exhibit seasonal trends. The Holt-Winters method repeats and refines a time-series formula that includes a

level, trend, and seasonal component. The formula calculates forecast values valid for time t using a weighted average for all data prior to time t .

Related term

Forecasting

I

Import

A tool that supports adding a field to a database by uploading records from an external database. The field is stored in FastDB.

Related terms

database, FastDB, field

indexed field

A data field having an associated key. An indexed field appears in a summary table used for data retrieval.

Related terms

data field, field, table

J-K

k-means

An iterative method of cluster analysis that groups large data sets into clusters of similar data. A k-means method forms clusters around data values having the nearest mean.

Related terms

analysis, Clustering, cluster analysis, mean

kurtosis

A coefficient that describes the degree of concentration for a distribution of values, based on a mathematical average. The kurtosis coefficient is a value between -0.5 and 0.5. Colloquially, the Kurtosis coefficient is an average that indicates how sharp a distribution is with respect to a standard normal distribution.

Related terms

skewness, standard normal distribution

L

Label

A data visualization gadget that associates specific alphanumeric characters with a defined measure. A Label displays a text description of a measure in the BIRT Analytics Gallery.

Related terms

Canvas, Cylinder, Dial, Funnel, gadget, Meter, Sphere

leverage

A value that indicates how well an association rule predicts the consequent. The method used to calculate leverage differs from the method used to calculate lift.

Related terms

association rules, lift

lift

A value that indicates how well an association rule predicts the consequent. A lift value greater than one indicates that the items in the rule appear together more than expected. The method used to calculate lift differs from the method used to calculate leverage.

Related terms

association rules, leverage

Links

A tool that supports maintaining links binding columns and tables in a database stored in FastDB.

Related terms

column, table

M

make permanent

A field operation that creates a new data field from either a calculated field or a current segment. The data field appears in FastDB.

Related terms

calculated field, FastDB, field, segment

Map analysis

A tool that supports plotting data values and regions on a geographic map. For example, a map analysis shows geographic regions and the number of high-net-worth customers in each region.

Related terms

analysis, Bubble analysis, Crosstab analysis, Evolution analysis, Pareto analysis, Profile analysis, Venn analysis

maximum

The highest registered value in a set of values.

Related term

minimum

mean

An arithmetic mean of all registered values in the field.

Related terms

median, mode

median

A value that divides a field into two symmetrical parts.

Related terms

mean, mode

Meter

A data visualization gadget that uses colored bars to display numeric values and boundaries in a range.

Related terms

Canvas, Cylinder, Dial, Funnel, Gallery, Label, Sphere

minimum

The lowest registered value in a set of values.

Related term

maximum

mode

The values having the most frequent number of occurrences in a field.

Related terms

mean, median

My Data Viewer

A tool that supports viewing fields and tables in multiple databases stored in FastDB.

Related terms

Data Explorer, database, Discrete Values Viewer, field, My Folders Viewer, table

My Folders Viewer

A tool that supports viewing reports, selections, and gadgets by a user or, if shared, by other users. My Folders appears as a tab in Data Tree and in the Start pane.

Related terms

Data Explorer, Discrete Values Viewer, gadget, My Data Viewer, selection

N

NetScaler Web Logging (NSWL) query

A type of SQL query that tracks HTTP data traffic and writes information to a log file in a standard format such as the following example:

```
Select * from [Demo].[Household]
where [Demo].[Household].[Town]='LONDON';
```

Related terms

BIRT Analytics Administration, query, security filter, SQL (Structured Query Language)

normal distribution

A bell-shaped, single-peaked, symmetric distribution of data. In a normal distribution, the mean, mode, and median coincide at the center.

Related term

standard normal distribution

Numeric Ranges

A tool that supports creating a calculated field that includes a series of ranges into which data from numeric fields is grouped. For example, Numeric Ranges supports defining the following age ranges: Young - for age values less than 21, Adult - for age values 21 through 67, and Old - for age values greater than 67.

Related terms

calculated field, field

O-P

parameter

A variable expression that accepts a defined set of values.

Related term

filter

Parametric

A tool that supports creating a field based on a defined condition, for use as a filter on a measure.

Related term

field

Pareto analysis

A tool that supports comparing data using the Pareto principle, a commonly accepted rule which implies a data distribution with a numeric ratio of 80% to 20%. For example, the Pareto principle implies that 80% of sales result from 20% of customers.

Related terms

analysis, Bubble analysis, Crosstab analysis, Evolution analysis, Map analysis, Profile analysis, Venn analysis

predictive analytics

A subject encompassing a variety of techniques used to analyze current and historical facts to make predictions about future, or otherwise unknown events. Credit scoring is a well-known application that uses predictive analytics techniques to generate a score for an individual, based on credit history data for that individual.

Related terms

Association Rules, Clustering, Decision Tree, Forecasting

Profile analysis

A tool that supports identifying a set of similar characteristics in a group. A profile analysis compares z-score values calculated for each set of characteristics.

Related terms

analysis, Bubble analysis, Crosstab analysis, Evolution analysis, Map analysis, Pareto analysis, Venn analysis, z-score

profile

A set of associated security roles, groups, filters, and users. Using the BIRT Analytics Administration tool, the administrator creates a profile from the lists of roles, groups, filters, and users available on the system. From the BIRT Analytics security options list, choose Profiles, specify a profile name, provide a description, then select the roles, groups, filters, and users to include in the profile.

Related terms

BIRT Analytics Administration, group, security filter, security role

prompted filter

A data set filter that supports user entry of parameter values.

Related terms

filter, parameter

Q

Quantile

A tool that supports creating a new calculated field by grouping values in a numeric field, using multiple groups that contain an equal number of values. For example, use Quantile to group a field containing 2400 values into four quartiles having 600 values each.

Related terms

calculated field, field

query

A statement specifying the data rows to retrieve from a data source. For example, a query that retrieves data from a database typically is a SQL SELECT statement.

Related terms

database, SQL (Structured Query Language)

R

Ranking

A tool that supports ordering a table by generating a column of calculated values that correspond to a sorted column. The calculated values represent an ordered list of ranks.

Related terms

column, table

record

A set of related, indexed data fields in a database. A record often appears as a row shown in a table. For example, a customer record could include a numeric field for customerID, a character string field for customer name, and an alphanumeric field for age group.

Related terms

field, row

row

See record.

Related terms

field, record

S

scheduled task

A scheduled task includes a trigger, task details, and an assigned action.

Related terms

action, campaign, stage, trigger, workflow

Scratchpad	<p>A BIRT Analytics work area that supports temporary caching of multiple segments. Scratchpad also supports creating new fields based on segments or selections.</p> <p>Related terms Data Explorer, Data Tree, segment, selection</p>
seasonal periodicity	<p>A value indicating the number of periods in a cycle. Input a value for seasonal periodicity to initiate a forecast that predicts a seasonal pattern in a data set.</p> <p>Related terms Forecasting, Holt-Winters</p>
seasonality	<p>In a data set, a periodic trend that corresponds to monthly, quarterly, or semi-annual periods such as seasons.</p> <p>Related terms Forecasting, Holt-Winters</p>
security role	<p>A set of functionalities that an administrator uses to configure permissions in the BIRT Analytics system.</p> <p>Related terms BIRT Analytics Administration, functionalities, query, security filter</p>
security filter	<p>A type of query that an administrator uses to limit access to data in the BIRT Analytics system.</p> <p>Related terms BIRT Analytics, BIRT Analytics Administration, group, NetScaler Web Logging (NSWL) query</p>
segment	<p>A segment is a group of records sharing at least one common characteristic.</p> <p>Related terms record, selection</p>
selection	<p>A selection is a user-specified request that returns a segment from a database.</p> <p>Related terms record, segment</p>
skewness	<p>A value that reflects the distribution of values in a data set. Skewness values can be positive, zero, or negative. A positive value reflects a data set in which more values lie to the left of the mean value. A negative value reflects a data set in which more values lie to the right of the mean. A zero value indicates values distributed evenly around the mean, typically implying a symmetric distribution.</p> <p>Related terms kurtosis, mean</p>
Sphere	<p>A data visualization gadget that uses a colored sphere shape to display numeric values and boundaries in a range.</p> <p>Related terms Canvas, Cylinder, Dial, Funnel, gadget, Label, Meter</p>
SQL (Structured Query Language)	<p>A language used to access and process data in a relational database.</p> <p>Related term database</p>
stage	<p>A tool that supports defining users as task owners and assigning to each task owner the permissions required to perform tasks. Define a stage to identify part of a campaign.</p> <p>Related terms action, campaign, scheduled task, trigger, workflow</p>

standard deviation

The value equal to the positive square root of variance calculated for a data set.

Related term
variance

standard normal distribution

The normal distribution in which the mean is zero and the standard deviation is one.

Related term
normal distribution

Standardize column

A tool for preprocessing data values having a distribution different from a standard normal distribution. Multiple options support value sets distributed closely, clustered, spread, or having many repeated values.

Related term
standard normal distribution

sum

The cumulated sum of all the values in a field.

Related term
sum-of-squares

sum-of-squares

The sum of all of the squared values in a set.

Related term
sum

support

An expression that calculates a ratio measuring how many transactions contain all items in an association rule. The support expression has the following syntax:

`Support (A,B) = Transactions (A,B)/Total transactions`

Related term
association rules

T**table**

A named set of records in a database.

Related terms
database, record

target filter

A filter that returns a group of records for comparison with an established baseline. For example, use the year 2010 as a target filter for profit, to compare profit earned in 2010 with profit earned in another, baseline year.

Related terms
baseline filter, universal filter

temporal file

A temporary data file generated and stored in the system cache. Using the BIRT Analytics Administration tool, the administrator can remove the accumulated temporal files and records created by an application to optimize performance.

Related term
BIRT Analytics Administration

trigger

A trigger is a time or event that starts a scheduled task.

Related terms
action, campaign, scheduled task, stage, workflow

U

universal filter

A filter that is always applied at a lower resolution level, before changing resolution.

Related terms

baseline filter, target filter

V-Y

value

1 The content of a constant, parameter, symbol, or variable.

2 A specific occurrence of an attribute. For example, blue is a possible value for an attribute color.

Related term

parameter

variance

A value equal to the squared average of the distances between each value and the arithmetic mean.

Related term

mean

Venn analysis

A tool that supports data analysis based on crossing more than two fields. A Venn analysis identifies coincident values in multiple data segments. For example, use a Venn analysis to show how many customers buy the same three products.

Related terms

Bubble analysis, Crosstab analysis, Evolution analysis, Map analysis, Pareto analysis, Profile analysis, segment

W

workflow

A role responsible for completing tasks or stages in a campaign.

Related terms

action, campaign, scheduled task, stage, trigger

Z

z-score

A value describing whether a quantifiable difference between two groups is statistically significant.

Related term

Profile analysis

Index

A

- access control list 143
- Access database files 25
- access permissions 18
 - See also* privileges
- accessing
 - BIRT Analytics 2
 - BIRT Analytics features 2
 - saved analyses 18
- ACLs. *See* access control lists
- action details, changing 137
- action names 119
- action properties 136, 139
- action types 125, 132
- actions 119, 125, 132, 143
- Actions visual editor 136, 137
- activating tasks 134
- adding
 - aggregate values 26, 27
 - blocks 22
 - bubble diagrams 46
 - calculated fields 30, 153
 - calculated measures 41
 - campaign cells 125
 - canvas components 90, 92
 - chart labels 46
 - chart legends 46
 - charts 40
 - clustering models 101
 - comments 78
 - crosstabs 39
 - cylinder indicators 85, 86
 - data fields 26, 31, 33, 152
 - data filters 19, 20, 24, 27, 46
 - default numeric ranges 32
 - dial indicators 66, 68
 - expressions 30, 31
 - funnel indicators 88, 89
 - label indicators 79
 - links 25
 - media conditions 118
 - meter indicators 74, 76
 - quantile range fields 33
 - ranking columns 35
 - segments 14
 - sphere indicators 81, 83
 - subfolders 8
 - tables 25
 - Venn diagrams 43
- addresses (email) 136
- Admin tools v, 144
- Advanced Analytics icon 2
- Advanced tab (bubble diagrams) 47, 48, 49
- Advanced tab (Crosstab) 40
- Advanced tab (Pareto) 62
- Advanced tab (Profile) 56
- Advanced tab (Venn diagrams) 44, 45
- aggregate functions 26
- Aggregates tool 26, 143
- analysis (defined) 143
 - See also* analyzing data
- analysis definitions 18, 38
- analysis gadgets. *See* gadgets
- Analysis icon 2
- analysis types 38
- analytical events 132
- analytical model 4
- analytical reports. *See* reports
- analytical tools 2, 18
- Analytics application. *See* BIRT Analytics
- Analytics icon 2
- analyzing data v, 2, 26, 38
- AND function 9
- AND operations 9, 21
- animation 54
- antecedent 143
- Antecedent Item Range parameter 108
- Apply model actions 133
- Apply model online actions 133
- Association Rules 144
- association rules 107, 143
- Association Rules tool 108
- associations 4
- attributes (events) 132
- Autocalculate option 42, 103
- automating tasks 114, 132, 139
- averages 69, 102
- axis labels (charts) 46
- axis labels (crosstabs) 42
- axis values 42, 45, 52

B

- Band properties (Numeric Ranges) 32
- base segments 55
- baseline filters 19, 20, 144
- baseline values 19
- binomial expressions 56
- BIRT Analytics Administration tool v, 144
- BIRT Analytics application v, 144
- BIRT Analytics features 2
- BIRT Analytics interface 2, 3, 6, 18
- BIRT Analytics Loader v, 144
- BIRT Analytics reports. *See* reports
- Block icon 22
- block selections 22
- Boolean values 30
- bubble analysis (defined) 144
- bubble analysis type 19, 42

- bubble charts 46, 48
- Bubble diagram editor 46
- bubble diagrams 45–50
- bubble size 45
- buttons (control) 51

C

- Calculate tool 18, 42
- calculated field icon 7
- calculated field type 146
- calculated fields 7, 19, 30, 144, 153
- calculated measures 41
 - See also* crosstabs
- calculated values 28, 153
- calculations 42, 55, 96, 101
- campaign cells 125
- campaign events 132
- Campaign History view 129
- Campaign Workflow editor 114, 119
- campaigns
 - changing property settings for 122
 - changing resolution levels for 124
 - creating 114, 120, 129
 - defining elements of 114
 - defining strategies for 120
 - evaluating 129, 130
 - excluding records from 122
 - exporting records for 133
 - grouping multiple 120
 - implementing actions for 119, 125
 - implementing tasks for 114, 115, 116
 - limiting scope 126
 - loading responses for 133
 - locking 128
 - managing stages 129
 - planning 119
 - repeating activities for 121
 - running 115, 128, 130, 133
 - saving 123, 130
 - sending execution results for 133
 - setting domains for 123, 126
 - setting media types for 118
 - setting properties for 120
 - setting response rules for 127
 - starting 128
 - storing records for 117, 130
 - unlocking 122
- Canvas modes 90
- Canvas tool 90–92, 145
- categorical data 6, 8, 45, 51
- categorical variables 45, 51, 55, 62
 - clustering and 101
 - logistic regression and 98
- categories. *See* categorical data
- category names 8, 54
- Change resolution level icon 10
- changing
 - action details 137
 - calculated fields 7
 - campaign cells 125
 - campaign properties 128
 - campaigns 128
 - canvas components 92
 - chart configurations 46
 - chart type 12, 13
 - crosstab analysis types 42
 - crosstabs 42
 - data 38
 - field names 28, 29
 - indicator type 72
 - indicators 90
 - measure names 48
 - measures 41
 - passwords 3
 - resolution 15, 22, 44
 - resolution levels 10, 124
 - scheduled tasks 139
 - segment names 10
 - segments 9
 - user preferences 3
- character separators 25
- character strings 100
- chart labels 46
- chart legends 42, 46
- Chart tab (bubble diagrams) 48
- Chart tab (Crosstab) 39
- Chart tab (Venn diagrams) 43
- chart types 12, 13, 40
- Chart view 12
- charts
 - changing type 12
 - comparing time periods and 103
 - converting crosstabs to 42
 - displaying 12, 13
 - hiding labels in 46
 - redrawing 42
 - showing bubble diagrams as 48
 - showing crosstabs as 39, 42
 - showing decision trees as 106
 - showing Venn diagrams as 43
 - viewing campaign summaries and 129
 - viewing forecasts and 103
 - viewing Pareto results in 63
- choropleth maps 58
- choose Selections icon 20
- Clear icon 11
- Clear option 18
- clearing workareas 18
- clock icon (evolution) 54
- cluster analysis 101, 133, 145
- clustering 101
- clustering algorithm 101
- clustering models 101
- Clustering tool 145
- clusters parameter 101
- coincidences 43, 44

- Collect action 119
- color bars (meters) 78
- color palette
 - cylinder indicators 87
 - decision tree classifications 105
 - dial indicators 69
 - meter indicators 77
 - sphere indicators 83
- colors (indicators) 66
- colors (maps) 58
- column bindings. *See* links
- column character separators 25
- column definitions (import) 25
- column headings 38
- column names 133
- column selector icon 11
- column-oriented DBMS 145
- columns
 - See also* calculated fields; fields
 - creating parametric 33
 - creating ranking 35
 - defined 145
 - defining event triggers and 139
 - deleting 133
 - linking to data in 25
 - plotting evolution and 52
 - preprocessing values in 95
 - reducing range of values in 98
 - remapping text strings in 100
 - sorting 7
 - standardizing data in 95, 96, 99
 - storing campaign data and 117
 - viewing crosstab 38
 - viewing database 14
- command box (Canvas) 90
- comma-separated file formats 18
- comments 78
- comparative analysis 19
- comparisons 55, 66, 95, 103
- computed fields. *See* calculated fields
- concatenation 30
- concentration (values) 150
- conditional expressions 30, 132
- confidence 145
- confidence level parameter 101
- confidence levels 98
- configuration options
 - See also* preferences
- configuration tools 2
- configuring
 - analysis operations 90
 - events 132
- confusion matrix 104
- consequent 145
- Consequent Item Range parameter 108
- constants 30
- context menus 3, 7, 10
- continuous values 101
- continuous variables 45, 101
- control buttons 51
- control types 127
- controls 51
- Convert tool 18, 72, 145
- converting indicator types 72, 74, 81, 85
- copying. *See* duplicating
- Count function 79
- count operations 11, 19, 63, 146
- creating
 - aggregate values 26, 27
 - blocks 22
 - bubble diagrams 45–50
 - calculated fields 30, 153
 - calculated measures 41
 - campaign action goals 119
 - campaign cells 125
 - campaign strategies 120
 - campaigns 120, 129
 - crosstabs 39
 - cylinder indicators 85, 86
 - data fields 26, 31, 33, 152
 - data filters 19, 27, 46
 - decision trees 105, 108
 - default numeric ranges 32
 - dial indicators 66, 68
 - event triggers 139
 - evolution analyses 51–55
 - expressions 30, 31
 - folders 8
 - forecasting models 102
 - funnel indicators 88, 89
 - label indicators 79
 - links 25
 - map analyses 58–61
 - meter indicators 74, 76
 - Pareto analyses 61–63
 - profile analyses 55–58
 - quantile range fields 33
 - queries 21, 33
 - ranking columns 35
 - scheduled tasks 134, 138
 - segments 14
 - sphere indicators 81, 83
 - tables 25
 - training processes 101
 - transitions 52
 - Venn diagrams 43–45
 - workflow stages 115
 - workflows 115
- Crosstab analysis (defined) 146
- crosstab analysis types 19, 42
- crosstab configurations 43
- crosstab help utility 55
- Crosstab tool 38, 39, 42
- crosstabs
 - adding calculated measures to 41
 - analyzing data and 38, 54

- changing resolution and 19
- clearing contents 42
- creating 39
- filtering 42
- setting display options for 42
- setting properties for 39
- viewing data in 40, 41
- CSV formats 18
- Cumulative Count option 63
- cumulative percentages 61
- current operation 52
- customizing
 - BIRT Analytics tabs 3
 - BIRT Analytics. *See* preferences
 - numeric ranges 32
- Cylinder editor 85
- cylinder gadgets 146
 - See also* cylinder indicators
- cylinder indicator definitions 85
- cylinder indicators
 - assigning values to 85, 86
 - converting dial graphs to 73
 - creating 85
 - displaying 87
 - setting properties for 85, 87
 - visualizing data and 85

D

- data
 - See also* values
 - aggregating 26
 - analyzing v, 2, 26, 38
 - deleting 11
 - determining if within limits 66, 74
 - determining relationships between 25, 55, 62
 - deviating from sample 102
 - displaying 6, 11, 66
 - extracting patterns from 107
 - filtering 14, 18–20
 - grouping 20, 26, 33, 45, 55
 - limiting access to 154
 - linking geographic 60
 - navigating through 6
 - organizing 101
 - plotting 12, 13
 - retrieving 15
 - sorting 24, 42
 - standardizing 95, 96, 99
 - testing conditions for 30
 - updating 40, 42
 - viewing changes to 38
 - viewing geographical representation of 58, 61
 - viewing over time 51, 54, 55, 102
 - viewing sample 11, 24
 - viewing value of 51, 54
 - visualizing 66
- data analysis 146
- data analysis gadgets. *See* gadgets

- data categories 6, 8, 45, 51
- data category names 54
- data controls 51
- data engineering tools 2, 26
- data enrichment operations 2, 26
- data exploration tools 2
- Data Explorer 11–13, 146
- data exports. *See* export operations
- data field types 146
- data fields. *See* fields
- data filters. *See* filters
- data imports. *See* import operations
- data items 6
- data mining 95
- data mining events 132
- data mining models 133
- data mining tools 2
- data repository 148
- data rows 153
 - See also* records
- data segments. *See* segments
- data set filters. *See* filters
- data sources. *See* databases
- data structures 9
- Data Tree 6, 147
- data type icons 6
- data types 147
- database icon 7
- database model 147
- databases
 - defined 147
 - exporting to 25
 - filtering data in 14
 - importing data from 25
 - navigating through 6
 - retrieving data in 15
 - selecting data in 20
 - storing campaign records and 117, 130
 - viewing analytical model of 4
 - viewing data in 7, 13
 - viewing information about 11
 - viewing specific columns in 14
 - viewing tables in 6, 14
 - writing calculated fields to 7
- date and time field type 146
- date attribute 132
- date data type 147
- date field icon 7
- date field type 146
- date fields 7
- date/time field icon 7
- Dates property 120, 121
- datetime data type 147
- decimal fields 30
- decimal places 30, 47, 103
- Decision Tree 133, 148
- decision tree 147
- decision tree classifications 105

- Decision Tree tool 105
- decision trees 104, 105, 108
- Decodes tool 28, 148
- De-duplicate option 121
- default numeric ranges 32
- default sort order 8
- Definition tab (Gallery) 66
- Definition tab (Map) 59
- Delete column actions 133
- Delete field command 7
- Delete icon 10
- Delete table actions 133
- deleting
 - actions 137
 - campaign cells 125
 - canvas components 92
 - categories 51
 - columns 133
 - data 11
 - fields 7
 - measures 42
 - segments 10
 - tables 7, 133
- delimiters 25
- demo database 4
- design areas (Canvas) 92
- Dial editor 66
- dial gadgets 148
 - See also* dial indicators
- dial graphs 73
- dial indicator definitions 66
- dial indicators
 - assigning values to 66
 - converting to meters 72
 - creating 66
 - displaying 69
 - filtering values for 67
 - reconfiguring values for 70
 - setting properties for 66, 69
 - setting range boundaries for 68, 69
 - visualizing data and 66
- difference measures (filters) 19
- differences 43, 44, 55, 101
- dimensions 40
 - See also* crosstabs
- Dimensions tab (Crosstab) 40
- discrete values 7, 12, 15, 24
- discrete values control 52
- Discrete Values Viewer 6, 7–8, 12, 148
- display options
 - Analytics tabs 3
 - crosstabs 42
 - Venn diagrams 45
- displaying
 - calculated values 28
 - campaign summaries 129
 - categorical data 45
 - category names 54
 - chart labels 46
 - chart legends 46
 - charts 12, 13
 - clustering models 101
 - current operation 52
 - cylinder indicators 87
 - data 6, 11, 66
 - data fields 6, 11
 - database columns 14
 - database tables 6, 14
 - decision trees 106
 - dial indicators 69
 - discrete values 7, 12, 15
 - files and folders 8
 - forecasts 103
 - funnel indicators 89
 - geographical data 58, 61
 - indicator summaries 90
 - label indicators 80
 - measures 19
 - meter indicators 78
 - multiple segments 48
 - outliers 103
 - records 11, 14, 63
 - scheduled tasks 138
 - smoothing parameters 103
 - sphere indicators 84
 - transaction data 108
 - visual indicators 66
 - Z-score values 58
- documentation v
- domain parameter 101
- Domains property 120, 121, 122, 123, 126
- download operations 9, 25
- Downloads tool 25, 148
- dubnium.exe 148
- duplicate records 35
- duplicating
 - measures 49, 79
 - scheduled tasks 138

E

- Edit command 7
- edit mode (Canvas) 90
- editing. *See* changing
- editors (text) 18
- email 133, 136, 139
- email addresses 136
- end-of-line characters 25
- engine data format attribute 132
- Engineering icon 2
- engineering tools 26
- Enrichment icon 2
- enrichment tools 2, 24, 26
- estimation method 149
 - See also* forecasting
- event attributes 132
- event triggers 132, 134, 139

- event types 132
- events 132, 143
- evolution analyses 51–55
- evolution analysis (defined) 148
- evolution analysis type 42
- Evolution editor 52
- Excel spreadsheets 18
- Exclude function 10
- exclude operations 10
- Execute campaign actions 133
- Execute response load actions 133
- exploding pie sectors 46
- Explore icon 2
- Export Analytic DB tool 148
- Export campaign cell actions 133
- Export Database tool 25
- export definitions 25
- export file formats 25
- Export File tool 25, 148
- export files 25
- Export icon 2, 18
- export modes 18
- export operations
 - analysis data 18
 - campaign records 133
 - crosstabs 42
 - Data Explorer and 11, 12, 13
 - data structures 9
 - forecasts 103
 - segments 25
- export tools 2, 25, 148
- export-to-database tool 25
- expressions
 - See also* formulas
 - blocking operands in 22
 - creating 30, 31
 - defined 148
 - displaying help for 30
 - forecasting and 102
 - setting conditions for 30
 - triggering events and 132, 134
 - unequal variances and 56

F

- FastDB web service 149
- features 2
- field (defined) 146
- field names 28, 29
- field operations 7
- field types 146
- fields
 - See also* columns
 - adding calculated 30, 153
 - as categorical variables 45, 51, 101
 - as transitional variables 52
 - building crosstabs and 38, 39, 40
 - building summary tables and 150
 - clustering and 101

- creating 26, 31, 33, 152
- crossing different 43
- deleting 7
- displaying 6
- exporting 25
- grouping data in 26, 33
- making permanent 7
- pivoting 38
- renaming 28, 29
- replacing values in 30
- retrieving 21
- selecting values in 8, 24
- sorting values in 24
- viewing categories in 8
- viewing contents of 7
- viewing information about 11, 12
- viewing specific 11
- file export operations 25
- file import operations 25
- file navigation views 6
- files 8, 25
- fill colors (cylinders) 87
- Filter parameter 108
- filter types 19, 149
 - See also* specific type
- filtering
 - bubble diagrams 46, 47
 - crosstabs 42
 - cylinder indicators 85
 - data 14, 18–20
 - forecast models 102
 - funnel indicators 88
 - indicator values 67
 - maps 60
 - meter indicators 75
 - Pareto values 62
 - ranked columns 35
 - record groups 144
 - sphere indicators 81
 - transactions 108
 - Venn diagrams 45
- filters 19, 20, 24, 27, 46, 149
- First discrete icon 10
- floating areas (charts) 103
- folders 8, 18
- font properties 80
- forecasting model parameters 102, 103
- forecasting models 102
- Forecasting tool 102, 103, 149
- forecasts v, 102, 103
- format options. *See* display options
- formula definitions 68, 76, 83
- formulas 68, 76, 83
 - See also* expressions
- Frequency view 13
- full numeric field icon 7
- full numeric field type 146
 - See also* numeric fields

- functionalities 149
- functions
 - aggregate data 26
 - cylinder indicators 85
 - dial indicators 66, 67
 - displaying help for 30
 - evolution 52
 - funnel indicators 88
 - label indicators 79
 - meter indicators 75
 - sphere indicators 81
- functions list 79
- fundamental data types. *See* data types
- Funnel editor 88
- funnel gadgets 149
 - See also* funnel indicators
- funnel indicator definitions 87
- funnel indicators
 - assigning values to 88, 89
 - creating 88
 - displaying 89
 - setting properties for 88, 89
 - visualizing data and 87

G

- gadget types 149
- gadgets 149
- Gallery 66, 149
- Gallery icon 2
- Gantt charts 129
- General Data property 120
- geographic data 58, 60, 61
- geographic maps 151
- global percentages 45
- goal-oriented activities 119
- goals (campaigns) 119, 125
- Graphic tab (Gallery) 66
- Graphic tab (Pareto) 63
- graphic visualizations 66
- graphics 2
- graphs. *See* charts
- green boxes 51
- grouping
 - data 20, 26, 33, 45, 55
 - multiple campaigns 120
 - records 101, 154
 - sets of tasks 115
- groups (users) 149

H

- has seasonality option 102, 149
- help utilities 55
- hiding chart labels 46
- Holt-Winters forecasting method 102
- Holt-Winters method 149
- home page (BIRT Analytics) 2
- hour attribute 132

I

- icons (BIRT Analytics) 2
- image file types 12
- images. *See* graphics
- Implied response rules option 127
- Import icon 2
- import operations 9, 25, 150
- import tool 2, 25
- incorrect predictions 105
- index measures 19
- index values 19
- indexed fields 150
 - See also* summary tables
- indicator summaries 90
- indicator types 72
- indicators (BIRT Analytics) 66, 90
- indicators (segments) 55
- Informative action 119
- inner selection icon 24
- inner selections 24
- input 2, 20
 - See also* prompts
- integer data type 147
- integers. *See* numeric data
- inversions 24
- Invert icon 10, 24

J

- join functions 9
- joins 9
- JPG files 12

K

- Kurtosis 12, 150
- Kurtosis coefficient 150

L

- Label editor 78, 79
- label gadgets 150
 - See also* label indicators
- label indicators
 - creating 79
 - displaying 80
 - setting properties for 80
 - visualizing data and 78
- labels (charts) 46
- labels (crosstabs) 38, 42
- language preferences 3
- legends (charts) 42, 46
- leverage 150
- lift 150
- linear regression analysis 97
- linear scaling 97, 100
- links 60
- Links tool 25, 150
- Loader module v, 144
- loading data structures 9

- localization preferences 3
- lock icon 92
- locking campaigns 128
- logistic regression 98
- logistic scaling 98, 100
- Longint data type 147

M

- Make permanent command 7
- Make permanent icon 10
- make permanent operations 151
- managing campaign workflow 2
- map analyses 58–61
- map analysis (defined) 151
- map analysis type 19, 42
- map definitions 59
- Map editor 59
- Marketing action 119
- marketing campaigns 107
- marketing campaigns. *See* campaigns
- master tables 25
- mathematical operations 22, 30
- Maximum function 66, 75
- maximum values 12, 42, 151
 - linear scaling and 97
 - reducing range of values for 98
- mean values 12, 151
 - clustering and 101
 - normalization and 96
 - outliers and 102
- measure definitions 66
- measure names 47, 48, 79
- measures
 - adding calculated 41
 - adding comments and 78
 - building crosstabs and 41
 - converting indicator type and 72
 - creating 42, 79
 - deleting 42
 - displaying 19, 45, 52
 - duplicating 49
 - editing 41
 - linking geographic data to 60
 - renaming 48
 - reusing 41
 - viewing as text 78
 - viewing Pareto curve for 63
 - viewing progression of 51, 54, 55
- Measures tab (Crosstab) 41, 42
- media conditions (campaigns) 118
- media lists 118
- media types 118, 125
- median values 12, 151
- menus 3, 7, 9
- messages 133, 136, 139
- Meter editor 74
- meter gadgets 151
 - See also* meter indicators
- meter indicator definitions 74
- meter indicators
 - assigning values to 74, 75
 - converting dial indicators to 72
 - creating 74
 - defining thresholds for 76, 77
 - displaying 78
 - filtering values for 75
 - setting properties for 74, 77
 - visualizing data and 74
- minimum confidence values (transactions) 108
- Minimum function 67, 75
- minimum support values (transactions) 108
- minimum values 12, 42, 151
 - linear scaling and 97
 - reducing range of values for 98
- mode values 12, 151
- moving canvas components 92
- multidimensional bubbles 46
- My Data Viewer 6–7, 151
- My Folders icon 18
- My Folders Viewer 8, 151

N

- naming
 - calculated measures 42
 - campaign cells 125
 - data fields 29
 - measures 47, 79
 - segments 10
 - workflow stages 116
- NAND function 10
- needles (dials) 66
- negative values 154
- NetScaler Web Logging queries 152
- New query icon 21, 22
- non-discrete values 12
- nonlinear transformations 98
- non-pivoted crosstabs 38, 40
- non-random projections 103
- NOR function 10
- normal distribution 152
- normalization 96, 100
- notifications 133, 136, 139
- Notify users via email actions 133
- NSWL queries 152
- null values 102
- numeric data 45, 100
- numeric field types 7, 146
- numeric fields 12, 30, 31
- Numeric Ranges tool 31, 152
- numeric variables 45, 62
- numerical estimation method. *See* Holt-Winters method

O

- observations 102
- online documentation v

- opening Selections windows 20
- operations 52
- operators 21, 30, 41
- Options tab (Crosstab) 42
- OR function 9
- OR operations 9, 21
- outliers 102, 103
- output 2
- output formats 12, 18

P

- pagination bar 11
- paging through records 11
- parameters
 - adding to queries. *See* parametric queries
 - applying as filters. *See* parametric filters
 - calculating automatically 103
 - clustering operations and 101
 - defined 152
 - defining selections and 20
 - entering values for 153
 - forecasting models and 102, 103
 - setting association rules and 108
- parametric columns 33
- parametric filters 20, 46, 152
- parametric queries 21, 33
- parametric query icon 21
- Parametric tool 33, 152
- Pareto analyses 61–63
- Pareto analysis (defined) 152
- Pareto editor 62
- Pareto principle 152
- partial filters 46
- passwords 3
- patterns 2, 107
- percentages 45, 61, 108
- percentile ranking 96
- periodic trends 154
- permanent segments 10
- permissions 18, 116
 - See also* privileges
- Permissions page 116
- personalizing BIRT Analytics 3
- pie sectors (charts) 46
- pin mode (charts) 103
- pivoted analysis 19
- pivoted crosstabs 39, 40
- Play icon (evolution) 53
- point-of-sale data 107
- positions options (tabs) 3
- positive values 154
- pre-analysis 103
- predefined maps 59, 61
- predictions 103, 105, 108
- predictive algorithms 104
- predictive analytics 152
- predictive models 104
- preferences 2, 3

- Preferences dialog 3
- preprocessing data values 95
- Preprocessing tool 99, 100
- previewing data 66
- printing forecast analyses 103
- privileges *v.* 143, 149
 - See also* access permissions
- probability 104
- profile (users) 153
- profile analyses 55–58
- profile analysis (defined) 152
- Profile editor 56
- profile segments 55
- projections 102
 - See also* Forecasting tool
- prompted filter 153
- prompts 20, 21
- properties 66

Q

- quantile range fields 33
- quantile rank 63
- Quantile tool 33, 153
- quantile types 33
- queries
 - adding parameters to 21
 - creating 21, 33
 - defined 153
 - limiting data access and 154
 - retrieving data with 16, 33, 153
 - specifying actions with 139
 - tracking HTTP data traffic and 152
 - viewing output for 15
- query actions 133
- query language 154
- query mode (Canvas) 92
- query operators 21

R

- random projections 103
- range boundaries (dials) 66
- range of values 7, 31, 69, 77, 83
 - reducing 98
- ranking columns 35
- ranking order 35
- Ranking tool 35, 153
- real data type 147
- real numeric field icon 7
- real numeric field type 146
 - See also* numeric fields
- recalculation 15
- Record identifier option 117
- Record View 11
- records 6
 - See also* segments
 - defined 153
 - displaying 11, 63
 - excluding from campaigns 122

- exporting campaign 133
- filtering 19
- grouping 101, 154
- including in campaigns 123
- incrementing 35
- limiting number displayed 14
- linking campaign 117
- paging through 11
- ranking subset of 35
- retrieving 15
- storing campaign 130
- updating count 11
- viewing number of 101
- redrawing charts 42
- Refresh icon 11
- regression analysis 97, 98
- relationships 25, 55, 62
- relative measures 52
- removing. *See* deleting
- Rename icon 10
- renaming
 - data fields 28, 29
 - measures 48
 - segments 10
- replacing outliers 103
- reports 7, 8
- repository 148
- resolution 4, 22, 25, 44
- resolution icon 22
- resolution levels 10
- resolution levels (campaigns) 117, 121, 124
- Resolution property 120, 121
- Resolution Table option 44
- resolution tables 10, 15, 39, 44
- response deadlines (campaigns) 122
- response rules (campaigns) 127
- Responses property 120, 122, 127
- rest values 55
- result measures 19
- result sets 15, 25
 - See also* queries
- results tables (export) 18
- rich-text formats 18
- role names 115
- roles 114, 154, 156
- row headings (crosstabs) 38
- rows 38, 153
 - See also* records
- RTF formats 18
- running campaigns 115, 128, 130, 133

S

- sample data 11, 24, 104
- sample mean 102
- sample selection icon 24
- Save As option 18
- Save option 18
- saving

- analysis definitions 18
- analysis reports 8
- calculated fields 7
- campaigns 123, 130
- charts as images 12
- crosstab configurations 43
- data structures 9
- export definitions 25
- export output 25
- forecasting models 104
- label indicators 80
- segments 9
 - user preferences 3
- scaling algorithms 100
- scheduled tasks 153
 - assigning actions to 136, 137
 - automating 139
 - creating 134, 138
 - displaying 138
 - duplicating 138
 - editing 139
 - removing actions from 137
 - triggering events for 134
- Scheduled tasks editor 138
- scheduling tasks and events 2
- Scratchpad 9, 154
- seasonal elements 102, 103
- seasonal patterns 154
- seasonal periodicity 154
- seasonal trends 149
- seasonality 154
- security filters 154
- security groups. *See* access control lists
- security roles 154
- segment definitions 10
- segment filters 57
- segment indicators 55
- segmentation tools 2
- segments
 - building 2, 20
 - changing 9
 - comparing 19
 - defined 154
 - defining as campaign domain 123
 - defining as decision tree condition 105
 - deleting 10
 - displaying multiple 48
 - examining values in 8
 - excluding from campaigns 123
 - exporting 25
 - filtering data and 14, 57
 - identifying coincidences between 43, 44
 - identifying common attributes in 101
 - identifying differences between 43, 44, 55
 - interacting with 2, 21
 - joining contents of 9
 - making permanent 10
 - placing in Data Explorer 11

- placing in Scratchpad 9
- profiling groups of values in 55, 56
- renaming 10
- saving 9
- selecting data for 20
- sorting 11
- Segments window 9
- Select sample icon 11
- selected tasks 138
- Selection command 7
- Selection sort icon 24
- selections 8, 20–24, 154
- Selections command 20
- Selections icon 2
- Selections window 21
- Send email actions 133, 136, 140
- Send execution results actions 133
- sensitivity 105
- separator characters 25
- shapes (indicators) 66
- shared files 8
- significance 56
- similarities 101
- skewness operations 154
- skewness values 12, 154
- smoothing parameters 103
- Softmax scaling 98, 100
- Sort icon 11
- sort order 8, 24
- sorting
 - bubble diagrams 46
 - columns for reports 7
 - data 24, 42
 - data categories 8
 - segments 11
- specificity 105
- Sphere editor 81
- Sphere gadgets 154
 - See also* sphere indicators
- sphere indicator definitions 80
- sphere indicators
 - assigning values to 81, 83
 - converting dial graphs to 73
 - creating 81
 - defining thresholds for 83
 - displaying 84
 - filtering values for 81
 - setting properties for 81
 - visualizing data and 80
- sphere measures 52
- spheres (evolution) 54
- spreadsheets 18
- SQL (defined) 154
- SQL statements. *See* queries
- squared values 13, 155
- stage 154
- stages (campaigns) 115, 116, 129
- standard deviation 13, 96, 102, 155

- standard normal distribution 155
- Standardize column tool 155
- Start icon 2
- starting
 - BIRT Analytics 2
 - campaigns 128
- statistical tools 2
- Statistics view 12
- strategies (campaigns) 120
- string data 30, 100
- string data type 147
- subfolders 8, 18
- sum (defined) 155
- summary values 13, 22, 155
- Summary view 11
- sum-of-squares (defined) 155
- support 155
- switch statements 30
- symmetric distribution 154
- system privileges 149

T

- table icon 7
- table names 133
- table resolution 15
- Table tab (bubble diagrams) 46
- Table tab (Crosstab) 39
- Table tab (Pareto) 63
- Table tab (Profile) 58
- Table tab (Venn diagrams) 43
- tables
 - adding fields to 26, 31, 33, 152
 - binding columns to. *See* links
 - creating 25
 - crosstabs and 19, 38, 39
 - defined 155
 - deleting 7, 133
 - displaying 6, 14, 42
 - expanding or collapsing items in 6
 - filtering data in 14, 19, 20
 - grouping data in 26
 - grouping records in. *See* segments
 - importing 25
 - interacting with 2
 - monitoring 139
 - permanent fields and 7
 - ranking 35
 - retrieving records from 15
 - updating data in 40, 42
 - viewing associations among 4
 - viewing decision trees as 106
 - viewing discrete values in 7, 12, 15
 - viewing forecast data in 104
 - viewing information about 11
 - viewing records in 11
 - viewing transaction data in 108
 - viewing Venn diagrams and 43, 44
- tabs 3

- target filters 19, 155
- target groups (profiles) 55
- target values 19
- task owners 115
- tasks 114, 132
 - See also* scheduled tasks
- temporal files 155
- temporary files 155
- temporary variables. *See* transitional variables
- text 78, 79
- text editors 18
- text field icon 7
- text field type 146
- text fields 30
- text files 25
- text strings 100
- themes (BIRT Analytics) 4
- thresholds
 - dial indicators 68, 69
 - meter indicators 76, 77
 - sphere indicators 83
- time data type 147
- time field icon 7
- time field type 146
- time fields 7
- time icon (transitions) 55
- time parameters 134
- time periods 103
- time series predictions 102, 103
- time triggers 132, 134
- timelines 51, 54, 55
- toolbars 18
- tools menu 3
- totals 19, 27, 45
- training processes (clustering) 101
- transaction IDs 107
- transactions 107
- transformations 98
- transient files 155
- transition variables 52, 54
- transitions 52, 54
- trends 2
- triangle icon 6
- triggers 155
- triggers (events) 132, 134, 139
- true negative rate 105
- true positive rate 105
- types. *See* data types

U

- unequal variance 56
- Unicode data type 147
- Uniform Resource Names 25
- unindexed field icon 7
- unindexed fields 6
- unique values 7
 - See also* discrete values
- universal filters 19, 60, 156

- unlock icon 90
- unlocking campaigns 122
- unreliable results 25
- updating
 - data 40, 42
 - record counts 11
- URNs 25
- user attribute 132
- user groups 149
- user preferences 2, 3
- user-specified requests. *See* selections

V

- validating formula definitions 68, 76, 83
- value definitions 66
- value labels (charts) 46
- values
 - See also* data
 - calculating distance between 101
 - calculating mean 96
 - categorical variables and 55, 98
 - comparing 19, 55, 66, 95, 103
 - customizing numeric range 32
 - determining variation from mean 96
 - displaying 28, 51, 54
 - inverting 24
 - measuring significance of 26, 56
 - preprocessing 95
 - ranking 35, 153
 - reducing range of 98
 - selecting 7, 8, 20, 24
 - setting chart axis 42
 - setting range of 69, 77, 83
 - showing average 69
 - showing differences between 43, 44, 55, 101
 - showing progression of 51, 54, 55
 - showing similarities between 101
 - tabulating predictive algorithms and 105
 - viewing discrete 7, 12, 15
 - viewing distribution of 12, 45, 102
 - viewing frequency of 12, 13
 - viewing unique 7
- variables 45, 51, 62, 101
- variance 56, 96, 156
- Venn analysis (defined) 156
- Venn diagram editor 43
- Venn diagram formats 45
- Venn diagram resolution tables 44
- Venn diagrams 43–45
- View definition icon 10
- view mode (Canvas) 90
- viewers 151
- viewing
 - calculated values 28
 - campaign summaries 129
 - categorical data 45
 - category names 54
 - chart labels 46

- chart legends 46
- charts 12, 13
- clustering models 101
- current operation 52
- cylinder indicators 87
- data 6, 11, 66
- data fields 6, 11
- database columns 14
- database tables 6, 14
- decision trees 106
- dial indicators 69
- discrete values 7, 12, 15
- files and folders 8
- forecasts 103
- funnel indicators 89
- geographical data 58, 61
- indicator summaries 90
- label indicators 80
- measures 19
- meter indicators 78
- multiple segments 48
- outliers 103
- records 11, 14, 63
- scheduled tasks 138
- smoothing parameters 103
- sphere indicators 84
- transaction data 108
- visual indicators 66
- Z-score values 58
- views (Data Explorer) 11
- views (Data Tree) 6
- visual indicators 66

W

- web services. *See* BIRT Analytics application; FastDB
 - web service
- web-based marketing campaigns. *See* campaigns
- weighted averages 102
- Word documents 18
- work areas (BIRT Analytics) 6, 18
- workflow 156
- workflow events 132
- Workflow history tables 130
- Workflow icon 2
- workflow roles 114, 156
- workflow stages 115, 116, 129
- workflows (campaigns) 114, 119, 129

X

- x-axis labels (charts) 42
- x-coordinates (evolution) 52

Y

- y-axis labels (charts) 42, 46
- y-axis values (charts) 42
- y-coordinates (evolution) 52

Z

- zero values 154
- zoom options 103
- Z-score indicator 55
- Z-score operations 55, 156
- Z-score values 56, 58

