IBM® Analytical Decision Management
Version 17 Release 0

IBM Analytical Decision Management for Campaign Optimization

IBM

Contents

Preface

About IBM Business Analytics

IBM Business Analytics software delivers complete, consistent, and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance and strategy management, and analytic applications provides clear, immediate and actionable insights into current performance and the ability to predict future outcomes. Combined with rich industry solutions, proven practices, and professional services, organizations of every size can drive the highest productivity, confidently automate decisions, and deliver better results.

As part of this portfolio, IBM SPSS Predictive Analytics software helps organizations predict future events and proactively act upon that insight to drive better business outcomes. Commercial, government, and academic customers worldwide rely on IBM® SPSS® technology as a competitive advantage in attracting, retaining, and growing customers, while reducing fraud and mitigating risk. By incorporationg IBM SPSS software into their daily operations, organizations become predictive enterprises - able to direct and automate decisions to meet business goals and achieve measurable competitive advantage. For further information or to reach a representative, visit http://www.ibm.com/spss.

Technical support

Technical support is available to maintenance customers. Customers may contact Technical Support for assistance in using IBM Corp. products or for installation help for one of the supported hardware environments. To reach Technical Support, see the IBM Corp. web site at http://www.ibm.com/support. Be prepared to identify yourself, your organization, and your support agreement when requesting assistance.

Chapter 1. About IBM Analytical Decision Management

Product overview

IBM Analytical Decision Management brings the benefit of predictive analytics to real business problems, allowing you to build custom applications tailored to your customers or industry. While applications are typically configured to solve very specific problems, all are based on a common set of capabilities:

- Automate decisions using business rules.
- Add insight using predictive models.
- Use prioritization, optimization, or simulation to reach the best decision based on the above.

A number of packaged applications are available, tailored to solving specific business problems. Contact your sales representative for more information.

What's new in version 17

IBM Analytical Decision Management 17 (17.0.0.0) adds the following new features:

Following is a description of the main new features:

Expression Builder. The Expression Builder has been enhanced to allow more complex and powerful expressions. The Expression Builder in IBM Analytical Decision Management now better matches the Expression Builder in IBM SPSS Modeler client. See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_expressions.dita.

Derived tab. In most IBM Analytical Decision Management applications, you can add additional fields to your project data model such as expressions, model output, fields created from segment rules, or fields from secondary tables. The new **Derived** sub tab of the Data tab displays any fields that are to be included as part of the data. See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_app_derived.dita.

Secondary Tables section. A new Secondary Tables section has been added to most IBM Analytical Decision Management applications. This replaces the various **Add/Edit additional tables** links found in previous releases. See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_app_additional_fields.dita.

Multi-factor authentication. Functionality that supports multi-factor authentication (MFA) has been added to IBM SPSS Collaboration and Deployment Services 7, IBM SPSS Modeler 17, and IBM Analytical Decision Management 17. These enhancements can improve security for large-scale enterprise environments. For example, SPSS products can be integrated with a company's existing Kerberos Single Sign On infrastructure and with SmartCard infrastructures (when a company's employees use security badges and card readers for machine and application authentication, for example). See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dms_install_security.dita.

Data source mapping. When possible, IBM Analytical Decision Management will now automatically map compatible storage types when a user adds an additional data source. If the type is not compatible, the user will be allowed to convert types.

File storage types. Users can now set the storage type and the format when reading in fixed and variable format data files, to ensure that values are read correctly. For example, if data source type is **File** and the file format is **Variable length fields in plain text** when adding a new data source in IBM Analytical

Decision Management, the Data Source dialog will now have a new **Specify Data Format** section that allows the user to specify the input type and the format of each field. This is similar to the functionality available in IBM SPSS Modeler client.

Project summary. For applications with a Deploy tab, a new **Project Summary** section has been added. Use this section to validate the project and display important project summary information, including any errors found that may impact deployment. The project summary displays information such as all attribute references and all repository references, and can be exported as a *.csv file. You can also move to other tabs of the application to fix errors and then return to the project summary on the Deploy tab.

Model references. When you reference models from within an IBM Analytical Decision Management application, the behavior of the model link has changed. When you click a model reference link, the model viewer dialog opens, allowing you to view details about the model. Or you can now right-click the model reference to select the option **Reference** or **Build**. When you click a rule reference, the rule viewer will open (as in previous releases). Or you can right-click a rule reference and select **Reference**.

Analytic Server credential suffix. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to configure a suffix to append to Analytic Server credentials. This will derive the name of the credential that will be used for Analytic Server connections in IBM Analytical Decision Management.

IBM Cognos credentails. When importing IBM Cognos BI database data or single list reports for use in IBM Analytical Decision Management, you can now select an IBM Cognos credential stored in the IBM SPSS Collaboration and Deployment Services Repository rather than entering user name and password information or using an anonymous connection.

Previous versions

Version 8
IBM SPSS Analytical Decision Management 8 extended users' ability to deploy analytics into everyday business processes, supporting a wider range of decisions with greater insight and control than ever before.
- Fine-tune and visualize the combined impact of rules, models, and optimization on business outcomes, with enhanced simulation capabilities.
- Solve a wider range of problems with new prebuilt applications and new optimization techniques such as price optimization and assortment optimization.
- Improved integration with IBM Operational Decision Management, and big data support via IBM SPSS Analytic Server integration.

Following is a description of the main new features that were added in version 8:

New installation technology. The product installation has been simplified and improved. It now uses Installation Manager technology. For more information, see the *IBM SPSS Analytical Decision Management Installation Guide*.

User interface enhancements. Many changes have been made to the user interface to improve usability.

IBM Analytical Decision Management for Predictive Maintenance. IBM SPSS Analytical Decision Management for Predictive Maintenance is a new prebuilt application that uses optimization to anticipate and prevent equipment failures before they occur. You can minimize costly downtime and make best use of scarce infrastructure and resources by predicting failures and scheduling the right correcting actions at the right time. You might anticipate which parts are likely to fail first and replace them at the appropriate time. Predicting when machines will fail helps to maximize profits and avoid losses from machines failing and being out of action.

IBM Analytical Decision Management for Demand Optimization. IBM SPSS Analytical Decision Management for Demand Optimization is a new prebuilt application that uses optimization to avoid costly overstocks and shortfalls by allocating the right amount of each product to each store. You can maximize profits and avoid losses due to inventory shortfalls and restocking fees by allocating the right amount of inventory to each outlet.

Simulation enhancements. New simulation capabilities combine the functionality of the previous **Simulate** button (formerly on the Define tab) with the previous **What If?** functionality (formerly on the Optimize, Prioritize, or Combine tab) into a single feature. The new and improved simulation functionality has its own toolbar in the upper-right corner of the user interface, and allows for scenarios to be saved with the project and persist across sessions.

Deployment and scoring. Various enhancements have been made to the Deploy/Score tab.

Application URL. IBM SPSS Collaboration and Deployment Services has added the ability to customize the URL for web-based applications, including the URL for launching IBM SPSS Analytical Decision Management. Contact your administrator for details. More information is available in the *Application context roots* chapter of the *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide*.

Statistics file enhancements. IBM SPSS Statistics compressed, password-protected data files (*.sav* and *.zsav*) are now supported in IBM SPSS Analytical Decision Management.

Tree viewer enhancements When viewing individual model results in IBM SPSS Modeler Advantage, the name of each model is a link that opens a new dialog containing information about the model such as the target field and input fields. If the selected model is a *tree model*, then a tree viewer will also be displayed. The tree viewer has been enhanced. You can now perform such actions as zooming in and out of the tree, changing the tree orientation, navigating large trees, and expanding or collapsing individual nodes within the tree.

Advanced options on Optimize tab. For applications that use complex mathematical optimization (CPLEX), such as IBM SPSS Analytical Decision Management for Campaign Optimization, IBM SPSS Analytical Decision Management for Predictive Maintenance, and IBM SPSS Analytical Decision Management for Demand Optimization, advanced options are now available related to the CPLEX optimization algorithm. You can now export an Optimization Programming Language (OPL) model file and add custom settings.

Data View source. A new *Data View* node has been added to IBM SPSS Modeler 16 to include data in streams that was defined in an IBM SPSS Collaboration and Deployment Services **analytic data view**. The new Data View data source can also be used on the Data tab in IBM SPSS Analytical Decision Management. You can create a connection to use with IBM Operational Decision Management rules. For complex rules that use a hierarchical data model, note that administrators may restrict use of such rules because they can degrade performance.

Prior to version 8, external rules required flat input parameters that matched the IBM SPSS Analytical Decision Management project data model. But with this new analytic data view support, external rules can now have hierarchical input parameters. See the topics "Data View source" on page 37 and "Locking data options" on page 99 for more information, and *Chapter 5* of the *Application Designer's Guide*.

IBM SPSS Analytic Server integration. In IBM SPSS Modeler 16, IBM SPSS Analytic Server has been introduced, allowing the use of Hadoop Distributed File System (HDFS) as a data source. IBM SPSS Modeler users can create streams using the HDFS Analytic Server source node and build models based on the Hadoop big data system. IBM SPSS Analytical Decision Management users can open these model streams in IBM SPSS Modeler Advantage for viewing (they cannot create new projects using Analytic Server data sources).

New chapter in Application Designer's Guide. A new chapter has been added to the *IBM SPSS Analytical Decision Management Application Designer's Guide*. It provides an example configuration of a custom optimization application. The standard IBM SPSS Analytical Decision Management for Campaign Optimization application uses a preconfigured Optimization Programming Language (OPL) model file called *CampaignOptimization.mod*. In some cases, the standard Campaign Optimization application and its OPL model might not suit the purposes of the business problem at hand. The new chapter walks through an example of a custom OPL model built using IBM ILOG CPLEX Optimization Studio, and a custom IBM SPSS Analytical Decision Management optimization application, for a fictional company.

Tabular evaluation results. A new **Include tabular results** option is available when running evaluations. Deselecting the option improves performance because it takes more time to run evaluations that include tabular results.

New setting for maximum number of nominal fields. A new setting has been added to the Properties dialog for setting the maximum number of members allowed for *Nominal* data fields. This is used on the Data tab. For more information, see "Properties" on page 29.

Single Sign On. ODBC connections can now use Single Sign On (SSO) for IBM SPSS Modeler Server and ODBC database connections with IBM SPSS Analytical Decision Management. Some configuration steps are required. See the topic Security considerations in the installation documentation.

SQL push-back. When the source data originates in a database, execution is pushed back to the database whenever possible (in a process known as "SQL push-back"). In general, IBM SPSS Analytical Decision Management supports the same level of SQL push-back as IBM SPSS Modeler but with the following limitations:

- If the project includes additional fields or selection rules which have references to models or external rules, then these references will not push back to the database, and this may inhibit the degree of push-back achievable for the project stream as a whole.
- In applications other than IBM SPSS Modeler Advantage, push-back is enabled only up to selection rules (global selections, and member selections on the first dimension). The degree of push-back may be increased if you choose an Integer field as the entity ID in the project data model.

Version 7.0.0.1

IBM SPSS Decision Management 7 Fix Pack 1 (7.0.0.1) added the following features:

Data Overview. A new data overview feature is available on the Data tab of all applications, on the Modeling tab and the Score tab in IBM SPSS Modeler Advantage, and on the Deploy tab of applications that include the *Score Now* feature. See the topic "Data overview" on page 40 for more information. This feature is only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It isn't available with version 15.

Scoring Summary Report. After scoring a project, a new *Scoring Summary Report* section provides details such as where the data was scored to, file size (if applicable), database and table name (if applicable), model name or project name, and number of records scored. The Scoring Summary Report section is available in IBM SPSS Modeler Advantage, and on the Deploy tab of applications that include the Score Now feature. Note that the files size is only displayed if running on IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It's not available with version 15.

Download local copy of scored file. If you choose to send scoring output to a flat file, you can also download a local copy of the file if desired. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to enable or disable the feature, and to increase or decrease the maximum file size limit allowed. If the **Download File** button is inactive, either the file size exceeds the maximum size limit set by the administrator, or the administrator disabled the feature. The feature is disabled by default. This feature is only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It isn't available with version 15.

New optional settings for building models. When building models in IBM SPSS Modeler Advantage, new optional settings are available such as data partitioning options, choosing which modeling techniques (algorithms) to include when building the model, specifying the maximum number of models to retain and combine, and setting profit criteria prior to building the model. The options available depend on the type of model and the target selected. See the topic "Optional model settings" on page 62 for more information. Note that the following new features are only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). They aren't available with version 15.

- Selecting which modeling techniques to include when building the model
- Setting profit criteria
- Viewing high level details about individual models
- Viewing additional details about a selected model

Tree viewer. After building a model in IBM SPSS Modeler Advantage and viewing individual model results, the name of each model is a link that opens a new dialog containing information about the model such as the target field and input fields. If the selected model is a *tree model*, a tree viewer will be displayed now also. Note that the tree viewer is only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It isn't available with version 15.

Evaluation enhancements. Various enhancements have been made to the Evaluate dialog box in IBM SPSS Modeler Advantage. For example, a new **Modify Model Selection** button is available, allowing you to modify the models to include in the evaluation. After making the selections, you can also copy the current model selections back to the individual model results table, if desired. Some of the new options are only available for certain model types (most new options are only for the *Predictive* model type, and some are also for the *Predictive rule* model type). Note that the following new features in the Evaluate dialog box are only enabled with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). They aren't available with version 15.

- Performance charts for numeric target
- Profit simulation for a numeric target
- Coincidence matrix charts (for all targets)
- Table view of all charts
- The **Modify Model Selection** button
- The option to **Include individual model evaluations for performance charts**

Microsoft Windows 8 and Internet Explorer 10. IBM SPSS Collaboration and Deployment Services 5 Fix Pack 1 (5.0.0.1) adds support for Windows 8 and Internet Explorer 10. They are not supported with IBM SPSS Collaboration and Deployment Services 4.2.1 or 5.

User interface enhancements. Various minor changes have been made to the user interface to improve usability.

Version 7
IBM SPSS Decision Management 7 (7.0.0.0) added the following features:

Optimization. The goal of optimization is to identify the solution that best meets a specific goal, such as maximizing the revenue from a marketing campaign, or minimizing the risk of fraud or churn. The result of optimization is a solution that answers this question across all possible combinations within your population, in the manner that best maximizes the value you want to optimize (such as profit).

Clustering models. New in IBM SPSS Modeler Advantage, cluster modeling methods help discover groups of data records with similar values or patterns. These techniques are used in marketing (customer segmentation) and other business applications (records that fall into single-record clusters may contain errors or be instances of fraud). Clustering models are often referred to as unsupervised learning models, since they do not use a target field, and do not return a specific prediction that can be evaluated as true

or false. IBM SPSS Decision Management allows you to build cluster models automatically. You can also combine, rename, modify, or delete clusters, or add custom clusters based on prior business knowledge.

Association models. New in IBM SPSS Modeler Advantage, association modeling allow you to predict which items are most likely to appear together, and predict the strength of the relationship between them. Association rules associate a particular conclusion (the purchase of a particular product, for example) with a set of conditions (the purchase of several other products, for example).

Data. The Data tab now includes the ability to have multiple data inputs, define the project data model with an Application View, refer to model outputs as if they were data, upload local data files, and preview and count data.

Improvements to IBM Cognos data integration. As with IBM SPSS Modeler 15, IBM SPSS Decision Management now supports the import of Cognos pre-built reports, Cognos filters and calculations, and control of automatic disaggregation.

Multiple elements. Applications can now include more than one dimension hierarchy. When launching an application, if the selected application is configured to include more than one element, you will be prompted to select one or more of the available elements (hierarchies) to use in the application. The dimension tree on the Define tab specifies the range of possible decisions or recommendations that can be returned. If an application includes more than one dimension hierarchy, they will be included on separate tabs of the dimension tree area on the Define tab.

Dynamic allocation. In previous releases, to allocate something, it had to exist in the dimension tree. Now, for dimension items with no children, you can read items from a table linked by a key on the project data model. For example, you might have a dimension item called *Cars* with no children. You could set up dynamic allocations on *Cars* to reference fields in a table such as *BMW*, *Audi*, and *Ford*.

Multiple allocations. A new **Multiple allocation** option allows allocation of records to multiple dimension items (such as multiple offers) for each segment. Multiple allocations can be made for each rule instead of a single allocation per rule.

User interface improvements. Many changes and additions to the IBM SPSS Decision Management user interface have been made to improve the usability and scalability of all applications.

Gallery enhancements. Various enhancements have been made to the gallery, such as a **Clear Gallery** option for removing more than one gallery item at a time.

Rule editor enhancements. Depending on where the rule editor is launched from, users can now set the rule type for selection rules (include or exclude), assign risk points for aggregation rules, select a dimension for allocation rules, and use *OR* and *Split OR* controls to combine multiple expressions within a rule and create an *OR* expression. When creating rules, now you can also select a dimension from a hierarchy tree (for applications with more than one tree) or launch the expression editor to create an expression. This functionality was not available in the rule editor dialog box previously.

Extended IBM Operational Decision Management integration. In IBM SPSS Collaboration and Deployment Services Deployment Manager, you can use the Add New External Rule dialog box to define references to IBM Operational Decision Management rulesets that can then be used in IBM SPSS Decision Management. Previously, these rules could only be accessed through a web service. With this release, you can now reference local rules in the IBM Operational Decision Management database. For the new local option, the rule is executed via the IBM SPSS Modeler Server (the same server defined in Deployment Manager for use by IBM SPSS Decision Management).

WhatIf? enhancements. Various enhancements have been made to functionality in the WhatIf? dialog box. For example, for the simulation date you can now enter range of dates on which to perform the WhatIf? evaluation.

Upload stream files. You can now open any IBM SPSS Modeler stream file from the local file system and store it to the repository if desired.

Multiple IBM SPSS Modeler Server credentials. If you have multiple IBM SPSS Modeler Server credentials defined in IBM SPSS Collaboration and Deployment Services Deployment Manager, you can now also specify more than one for use with IBM SPSS Decision Management. Then when the IBM SPSS Decision Management user logs on, the server will try the first credential in the list to see if the user has authority, proceeding through the list in order until a credential is found with the proper authority. If the user does not have authority to any of the credentials in the ordered list, he or she cannot use IBM SPSS Decision Management. This provides better control over which users have access to which parts of the IBM SPSS Modeler Server file system.

Version 6.2

IBM SPSS Decision Management 6.2 added the following features:

Support for Cognos data. When used with IBM SPSS Modeler Server 14.2 or newer, IBM SPSS Decision Management 6.2 added the ability to read data from an IBM Cognos Business Intelligence server when defining data sources, and to write scoring results back to a Cognos BI Server. After defining a connection to a Cognos BI server, you can browse packages, dimensions, queries, and reports to choose the desired fields. See the topic "IBM Cognos BI source" on page 38 for more information.

ILOG Business Rule Management System integration. Rules created in a Business Rules Management System such as ILOG can be referenced and used in IBM SPSS Decision Management applications, provided they have been developed to support the same data model used in the current IBM SPSS Decision Management project. These external rules can be selected and used in the same manner as other models and rules within the IBM SPSS Decision Management application. See the topic "Using rules from IBM Operational Decision Management" on page 55 for more information. Note that ILOG is now referred to as IBM Operational Decision Management.

Support for Linux on System Z. SUSE Enterpise Server 10 is supported on 64-bit IBM System z10.

Chapter 2. Campaign optimization example

IBM Analytical Decision Management for Campaign Optimization

IBM Analytical Decision Management for Campaign Optimization is similar to the IBM Analytical Decision Management for Customer Interactions application in that it helps decide which promotions to offer customers when they contact your organization, and delivers recommendations to a channel such as a call center, retail web site, or store location. By combining the logic of business rules with the insight gained through predictive modeling, the application identifies the most profitable decision for each customer. IBM Analytical Decision Management for Campaign Optimization introduces **optimization** and has the added value of maximizing the return on your campaign investment by determining the best offers for individual customers within constraints such as your budget, channel capacity, and contact policies.

For applications that use complex mathematical optimization (CPLEX), such as IBM Analytical Decision Management for Campaign Optimization, IBM Analytical Decision Management for Operations, and IBM Analytical Decision Management for Demand Optimization, note that advanced options are available related to the CPLEX optimization algorithm used in the application. Using IBM ILOG CPLEX Optimization Studio, advanced users may develop additional applications that use this capability. IBM ILOG CPLEX Optimization Studio is available for separate purchase. Contact your sales representative for more information.

See the topic "Optimization" on page 75 for more information.

Optimizing campaigns

For this example, a large retail organization wants to leverage predictive analytics to maximize the value of their retail promotions by matching the right offer to the right customer, based on their behavior and characteristics.

Based on the models, rules, and other parameters you define, IBM Analytical Decision Management for Campaign Optimization maximizes profit by identifying the best campaign to offer each customer, and delivers recommendations to a channel such as a call center, retail web site, or store location.

This example uses the data files *customer_summary_data.txt* and *customer_contact_history.txt*, which are distributed with IBM Analytical Decision Management. A completed version of this example is also provided in the file *retail_campaign_history.str*. By default, the sample data files are installed in the IBM SPSS Modeler Server *Demos* folder accessible from your application (for example, *C:\Program Files\IBM\SPSS\ModelerServer\17\Demos*). Contact your administrator for assistance installing these files if necessary.

Along with the sample data, this IBM Analytical Decision Management for Campaign Optimization example chapter requires the sample file *customer_contact_aggregation.str*. Have your administrator use the instructions in the following "Sample files" section to add the file to your repository before proceeding with the example.

Sample files

Data and other sample files are included with IBM Analytical Decision Management that can be used to replicate the examples shown in the user guides.

Data

During installation, a number of sample data files are installed under the IBM SPSS Modeler Server *\Demos* directory (for example, *C:\Program Files\IBM\SPSS\ModelerServer\17\Demos*). This data may be used with the prebuilt sample applications, and include:

- *bank_customer_data.txt*
- *bank_response_data.txt*
- *customer_contact_history.txt*
- *customer_summary_data.txt*
- *insurance_claim_data.txt*
- *insurance_fraud_data.txt*
- *inventory_orders.txt*
- *inventory_stock.txt*
- *machine_log_data.txt*
- *machine_summary_data.txt*
- *retail_purchase_data.txt*

Sample application files, models, and rules that use these data files assume the default location indicated above. If data files are installed to a different drive or location, data source paths must be updated in the sample applications accordingly.

Samples

A number of sample application files, models, and rules are provided for use with the prebuilt IBM Analytical Decision Management applications. The samples are packaged in the *\demo* directory on the installation disk, or provided in a file available for electronic download. Complete the following steps to import sample files for one or more prebuilt applications into the IBM SPSS Collaboration and Deployment Services Repository.

1. **DVD.** Copy the file *DecisionManagementExamples.pes* from the *\demo* folder on the DVD to a temporary location.

   OR

   Downloaded file. Download the file *ADM_17_Demo.zip* to a temporary location and extract it.
2. Open IBM SPSS Collaboration and Deployment Services Deployment Manager and log on to your Content Server connection.
3. Right-click the Content Repository folder and select **Import**.
4. In the Import dialog box, click **Browse** and select the *DecisionManagementExamples.pes* file you copied or extracted to a temporary location in step 1. Click **OK**.
5. In the Import Conflicts dialog, click **OK**. When the import completes, you will see the following new folders containing example projects that can be opened in IBM Analytical Decision Management.
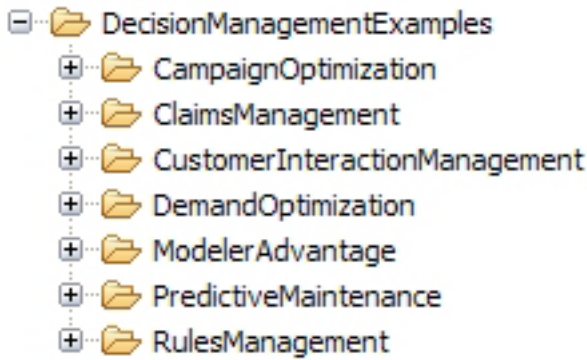
Figure 1. Folders for sample artifacts

Note that all sample application files, models, and rules assume that data files are installed under IBM SPSS Modeler Server's *Demos* directory (for example, *C:\Program Files\IBM\SPSS\ModelerServer\17\ Demos*). If data files are installed to a different location, the data source paths must be updated in the sample applications for the local configuration.

Launching the application

The Launch page includes links to available applications at your site. These links may be configured by your administrator; you may be able choose from additional applications by clicking the **Add application** link in the upper right-hand corner.

1. To access the Launch page, navigate to *http://hostname:port/DM*, where *hostname* is the IP address or name of the repository machine and *port* is the application server port number. If you don't know this information, contact your administrator. You can also add a bookmark to this page in your browser, or create a shortcut on your desktop.

   If you have problems accessing the IBM Analytical Decision Management URL or any IBM SPSS Collaboration and Deployment Services URLs, an administrator may have customized them. Contact your administrator. More information is available in the *Application context roots* chapter of the *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide*.

2. Log in with a valid user name and password. Contact your administrator for details if necessary.

3. To create a new project, select **New** from the drop-down list under IBM Analytical Decision Management for Campaign Optimization, and click **Go**.

4. Make sure both the *Campaign* and *Channel* elements are selected, with the *Campaigns* element appearing first in the list, and click **OK**. These will appear on the Define tab later in the project. For more information about elements, see "Project configuration" on page 24.

Figure 2. IBM Analytical Decision Management for Campaign Optimization Home page

Each icon on the Home page represents a different step in the process, from choosing data through defining, deploying, and reporting on the results. To get started, click the Data icon.

Setting up data sources

The Data tab allows you to configure a data source for the sample data file used in the example. Data sets can be configured by your administrator, or you can add your own.
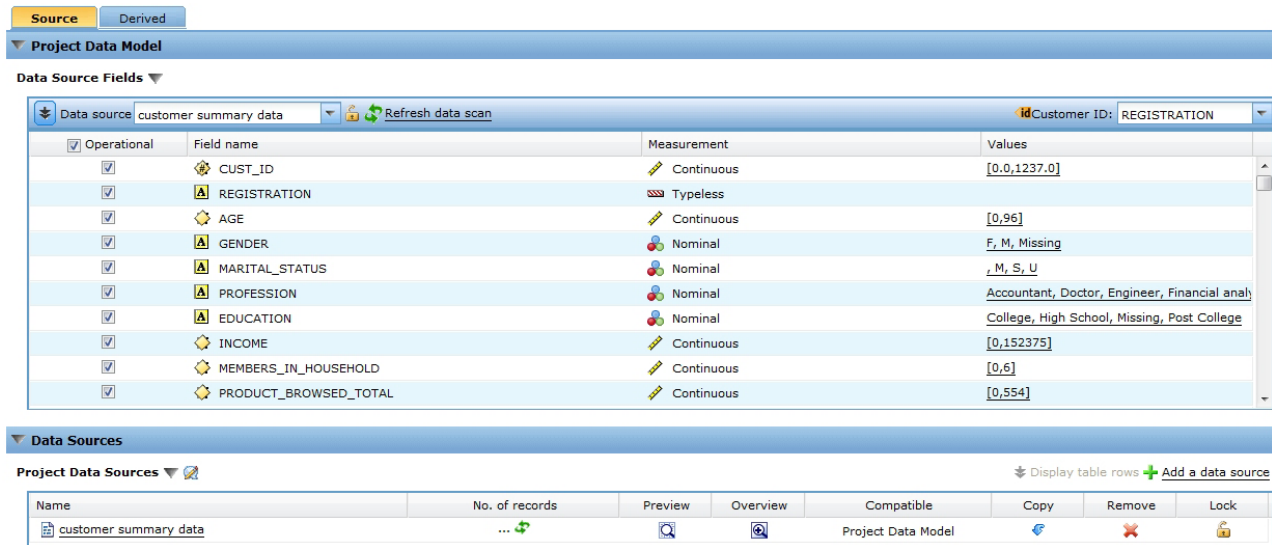

Figure 3. Data tab

The project data model defines the fields that are available throughout the application, and may be required for rules and models to evaluate. All other data sources must be mapped to the data model in order to be used with the application.

This example uses two data files, both of which are distributed with the application:

- **customer_summary_data.txt** includes information about customers such as demographic data, browsing history, and purchase transaction history. The project data model is based on this file.
- **customer_contact_history.txt** includes information about customer contact history, which will be valuable for purposes such as making sure customers aren't contacted with offers too often.

By default, the sample data files are installed in the IBM SPSS Modeler Server *Demos* folder accessible from your application (for example, *C:\Program Files\IBM\SPSS\ModelerServer\17\Demos*). Contact your administrator for assistance installing these files if necessary.

To define the project data model:
1. On the Project Data Sources panel, select **Add a data source** from the **Data source** drop-down.
2. Specify *customer summary data* as the data source name.
3. Select **File**, and select *customer_summary_data.txt*. Click **Save** to exit the Data Source Editor.
4. Select the **REGISTRATION** field from the Customer ID drop-down,
5. Under the Secondary Tables section at the bottom of the screen, click **Add/Edit Secondary Tables**. The Add/Edit Secondary Tables dialog box will be displayed.
6. Click **Add a new table**, type *Contact History* for the name, and click **Save**.
7. Select **Add a data source** from the Data source drop-down.
8. Specify *customer contact history* as the data source name.
9. Select **File**, and select *customer_contact_history.txt*. Click **Save** to exit the Data Source Editor.
10. Click **Links**. The Links dialog will be displayed.

11. Click **Add**, select **REGISTRATION** in both columns, and click **Add**. Then click **Save** to close the Links dialog and click **Save** again to close the Add/Edit Secondary Tables dialog.

12. Go to the Derived tab and click the **Derive field using a model or rule** icon. Browse to and select the file *\Content Repository\DecisionManagementExamples\CampaignOptimization\ customer_contact_aggregation.str*. Make sure **Contact History** is selected in the drop-down and select **CONTACT_COUNT** for the model output field.

13. Click **Save** to return to the Derived sub tab of the Data tab.

Defining campaigns, offers, and channels

Global Selections are applied to all records before any further processing occurs, and are useful for enforcing high-level policies. For example, you can exclude customers with a history of payment problems, or ensure that customers do not receive an offer within two months of a previous offer. Or you can exclude certain regains, certain age groups, etc.

The Global Selections tab is not used in this example, so we'll proceed to the Define tab.

The Define tab allows you to specify the available campaigns, offers, and channels, and the rules used to allocate customers accordingly.
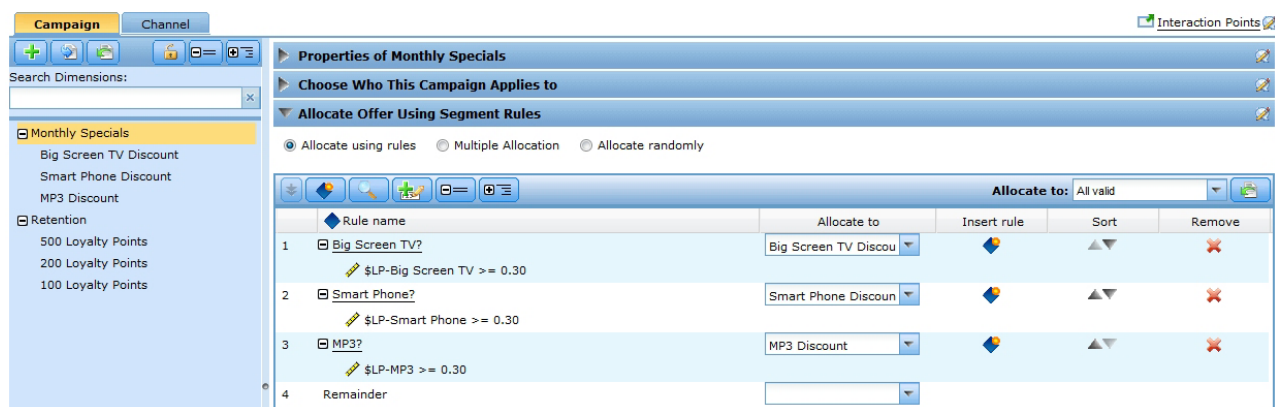
Figure 4. Define tab showing properties for Monthly Specials campaign

- The **Campaign** dimension tree in the left pane specifies the range of possible offers that can be returned, organized by campaign.
- The settings on the right determine how customers are selected and allocated for each campaign. These rules need not be mutually-exclusive; the same customer may be allocated to multiple campaigns or offers as appropriate.
- The **Channel** dimension tree in the left pane specifies the delivery channels available for sending offers.

To define campaigns and offers

When you first open the Define tab, a single default campaign is displayed (*My Campaign*, with child offer *My Offer*).

1. Right-click **My Campaign**, select **Rename**, and change its name to *Monthly Specials*. Click **Save**.
2. Under the *Monthly Specials* campaign, right-click **My Offer** and select **Rename**. Change its name to *Big Screen TV Discount* and leave the dialog open.
3. Click **Add a new Offer** to create a new offer called *Smart Phone Discount*, and another new offer called *MP3 Discount*. Click **Save**.

4. Click the **Add dimension tree members** icon, click **Add new Campaign**, and create a new campaign called *Retention*. Under the *Retention* campaign, add offers for *500 Loyalty Points*, *200 Loyalty Points*, and *100 Loyalty Points*. Click **Save**.

```
Monthly Specials
    Big Screen TV Discount
    Smart Phone Discount
    MP3 Discount
Retention
    500 Loyalty Points
    200 Loyalty Points
    100 Loyalty Points
```

5. With the *Monthly Specials* campaign dimension selected in the left pane, click **Properties of Monthly Specials** in the right pane.

6. Use the calendar chooser to specify a current active date as desired.

7. Under **Choose Who This Campaign Applies to**, do not add any selection rules. All customers are eligible for this campaign.

8. Under **Allocate Using Segment Rules**, create three rules named *Big Screen TV?*, *Smart Phone?*, and *MP3?* to allocate customers to offers as follows:
   - If *$LP-Big Screen TV >= 0.30* then allocate to *Big Screen TV Discount*.
   - If *$LP-Smart Phone >= 0.30* then allocate to *Smart Phone Discount*.
   - If *$LP-MP3 >= 0.30* then allocate to *MP3 Discount*.

9. Under **Allocate Using Segment Rules**, make sure **All valid offers** is selected for **Allocate to** (so a customer who is eligible for multiple monthly specials will receive more than one).

10. For the *Retention* campaign, specify properties in the same manner as for *Monthly Specials*, with the comments below.

11. Under **Allocate Using Segment Rules**, create three rules named *High LTV*, *Medium LTV*, and *Low LTV* to allocate customers to offers as follows:
    - If *LTV = HIGH VALUE* then allocate to *500 Loyalty Points*.
    - If *LTV = MEDIUM VALUE* then allocate to *200 Loyalty Points*.
    - If *LTV = LOW VALUE* then allocate to *100 Loyalty Points*.

12. Under **Allocate Using Segment Rules**, make sure **First valid** is selected for **Allocate to** (so a customer who is eligible for all the rewards offers will receive the *500 Loyalty Points* offer because rules are evaluated in the order listed).

To define channels

1. Click the **Channel** tab to view the channel tree. A single default channel is displayed (*My Channel*).
2. Right-click **My Channel**, select **Rename**, and change its name to *Mailshot*.
3. Click **Add new Channel** and create a new channel called *SMS*. Click **Save**.

Running a simulation

By running a simulation, you can see how the current rules apply to your sample data.

1. From the Define tab, click the **Create a new scenario based on current project settings** icon in the upper-right corner.

2. Select *customer summary data* for the simulation data source.

3. Specify a simulation date that falls within the active range for all campaigns and channels, as specified on the Define tab in the previous section. (By default, each campaign is active from the day it is created, and never expires.)

4. Click **Run**.

5. The scenario results will be displayed. Select **Define** from the drop down to display results for the Define tab.

Looking at the results, the Campaign graph shows that from all the allocations being made, 790 (24.2%) are for the *Monthy Specials* campaign and 2476 (75.8%) are for the *Retention* campaign. Note that these percentages are based on the total number of allocations being made, not on the number of records (customers) used from the simulation data set.
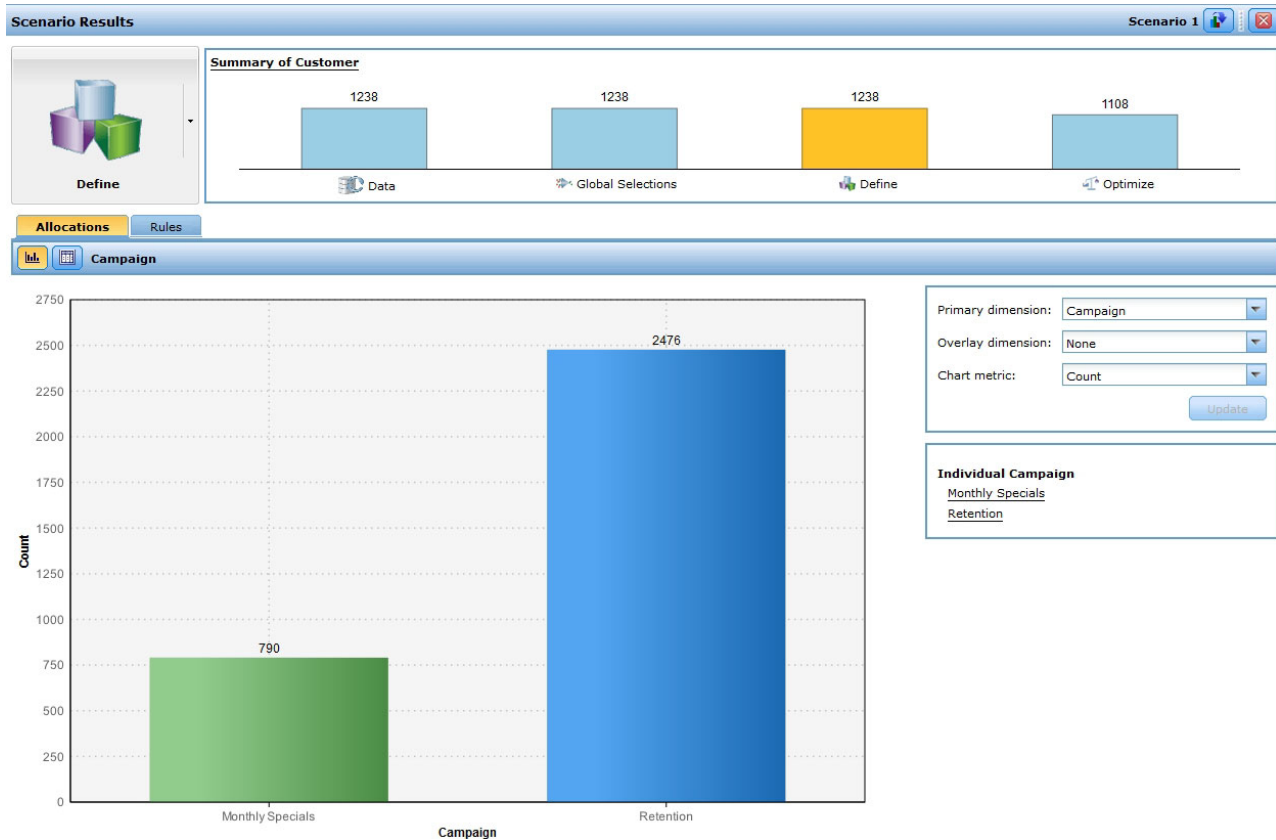
Figure 5. Simulation results showing campaign allocations

Select **Offer** from the **Primary dimension** drop-down menu and click **Update**. The Offer graph shows a similar breakdown by offer across all campaigns.
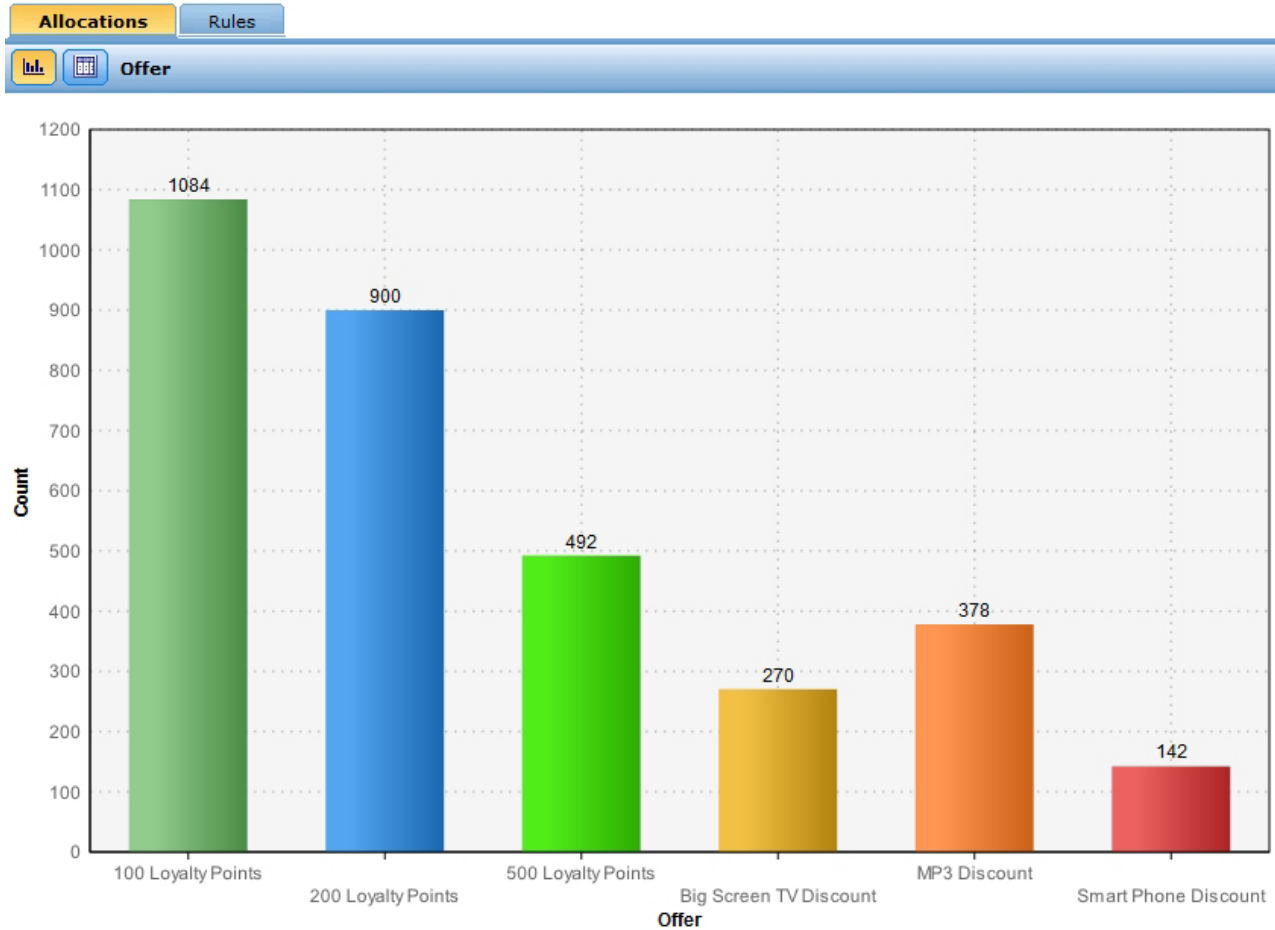
Offer

Figure 6. Simulation of offers across all campaigns

Select **Channel** from the **Primary dimension** drop-down menu and click **Update**. The Channel graph shows that an equal number of mailings and SMS messages (text messages) will be sent (1633 each).
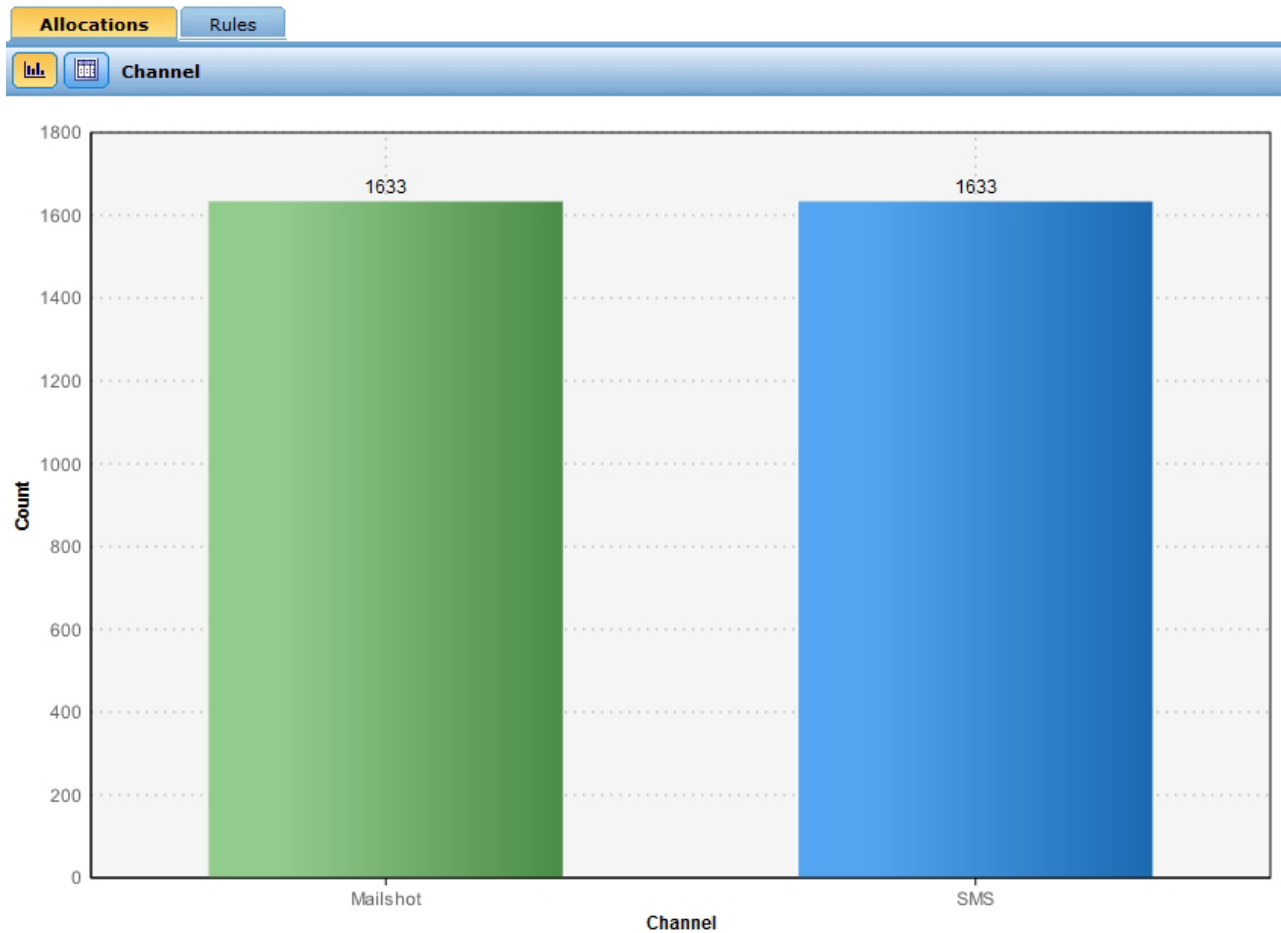
Figure 7. Simulation by of channels

Browse around through the other simulation results as desired. Based on the results, you may choose to adjust your allocation rules to push more or fewer records to one campaign or offer. When finished, close the scenario results to return to the Define tab.

Running a test

To see how specific records are processed by the application, click the **Perform a test on the application for a selection of records** icon in the upper-right corner of the Define tab.

1. Select *customer summary data* for the test data source.
2. Specify the number of records to return, such as 50, and select one or more fields to display.
3. Select **Define** for the step you wish to test and click **Run**. The test results will be displayed.

Figure 8. Running a test

The test summary shows the values of the selected fields for each record, along with the offers each customer is eligible for.

4. To see details on a specific record, select it and see the **Test Detail** section.

From the details, you can see how the customer was evaluated. For example, customer 21 qualifies for the *MP3 Discount* offer and *200 Loyalty Points* offer, and customer 22 qualifies only for the *100 Loyalty Points* offer. You can also click the information icon beside each record to see further details. When finished, close the test results to go to the Optimize tab.

Optimizing offers to reach a decision

The allocations on the Define tab determine which customers are eligible for which offers. But eligible doesn't necessarily mean profitable. And if a customer is eligible for multiple offers, which do you choose? This is what the Optimize tab is for.

Figure 9. Optimize tab

Optimization allows you to identify the best or most profitable offer for each individual within the constraints of budget and channel capacity. In other words, IBM Analytical Decision Management optimization defines the best way customers can be allocated to offers. The goal of optimization is to identify the solution that best meets a specific goal, such as maximizing the revenue from a marketing campaign. The result of optimization is a solution that answers this question across all possible combinations within your population, in the manner that best maximizes the value you want to optimize (such as profit).

See the topic "Optimization" on page 75 for more information.

This IBM Analytical Decision Management for Campaign Optimization example determines the most profitable offer for each customer, based on the following equation and constraints.

```
Total projected profit based on current project settings = (Prob. to Respond * Revenue) - (Offer
Cost + Channel Cost)
```

Where:
- `Prob. to respond` is a customer's propensity to respond to a particular offer.
- `Revenue` is the income revenue value or contribution expected from that response.
- `Offer Cost + Channel Cost` is the cost of making the offer.

Note: IBM Analytical Decision Management for Campaign Optimization projects created in previous releases will use a different equation. The equation was updated for version 17.

Table 1. Optimization constraints used in this example.

Constraint description	Constraint value name	Equation	Constraint
How many offers of this type can be made across all customers	Offers Available	>=	count (Offer)
Total $ available for all campaigns in this optimization window	Total Budget	>=	sum (Offer Cost + Channel Cost)
Maximum number of offers that can be made to each customer	Max. Offers	>=	count (Offer / Customer) + (Recent Offers)
Maximum number of offers that can be processed by each channel	Capacity	>=	count (Channel)

Optimization parameters

1. The top three parameters apply to all campaigns. Specify **20000** for *Total Budget* (total budget available across all campaigns), **3** for *Max. Offers* (the maximum number of offers that can be made to each customer), and select the **CONTACT_COUNT** field for *Recent Offers* (the number of recent offers each customer has received, to prevent customers from receiving an offer too many times).
2. To input parameters, click **Customize table** on the right side of the Optimize tab.
3. Select the **Offer** radio button for all parameters, and click **Save**. This allows you to specify inputs separately for each offer, rather than globally for each campaign. See the topic "Optimization" on page 75 for more information.
4. Specify inputs as shown in the table below. Inputs starting with **$LP** are fields selected from the customer summary data.

Table 2. Settings for Optimize tab.

Campaign/Offer	Prob. to Respond	Revenue	Offer Cost	Offers Available	Priority
Monthly Specials					

Table 2. Settings for Optimize tab (continued).

Campaign/Offer	Prob. to Respond	Revenue	Offer Cost	Offers Available	Priority
Big Screen TV Discount	$LP-Big Screen TV	330	0.50	50	Normal
Smart Phone Discount	$LP-Smart Phone	170	0.50	150	Normal
MP3 Discount	$LP-MP3	50	0.50	400	Normal
Retention					
500 Loyalty Points	0.50	2.5	0.1	500	Normal
200 Loyalty Points	0.50	1.0	0.1	500	Normal
100 Loyalty Points	0.50	0.5	0.1	500	Normal

5. On the Channel tab, for the *Mailshot* channel, set a *Capacity* of **10000** and *Channel Cost* of **0.1**. For the *SMS* channel, set a *Capacity* of **10000** and *Channel Cost* of **0.001**

Scenario analysis

You can create scenarios based on the current project settings to experiment with different combinations and see how changes impact your bottom line. Each scenario run is preserved, allowing you to compare each run.

1. To simulate your results based on the current inputs, click the **Create a new scenario based on the current project settings** icon in the upper-right corner of the Optimize tab.

2. In the New Scenario dialog box, select *customer summary data* for the data source and click **Run**.

Based on the results, you can see how many customers each offer will be sent to, the expected profit for each offer, and other details. In the upper-left corner, select **Optimize** instead of **Overall**. Then select **Channel** from the **Primary dimension** drop-down menu and click **Update**. The Channel results show how many mailings and text messages will be sent, the expected profit for each channel, and other details. Explore all the scenario results as desired and then close the results.
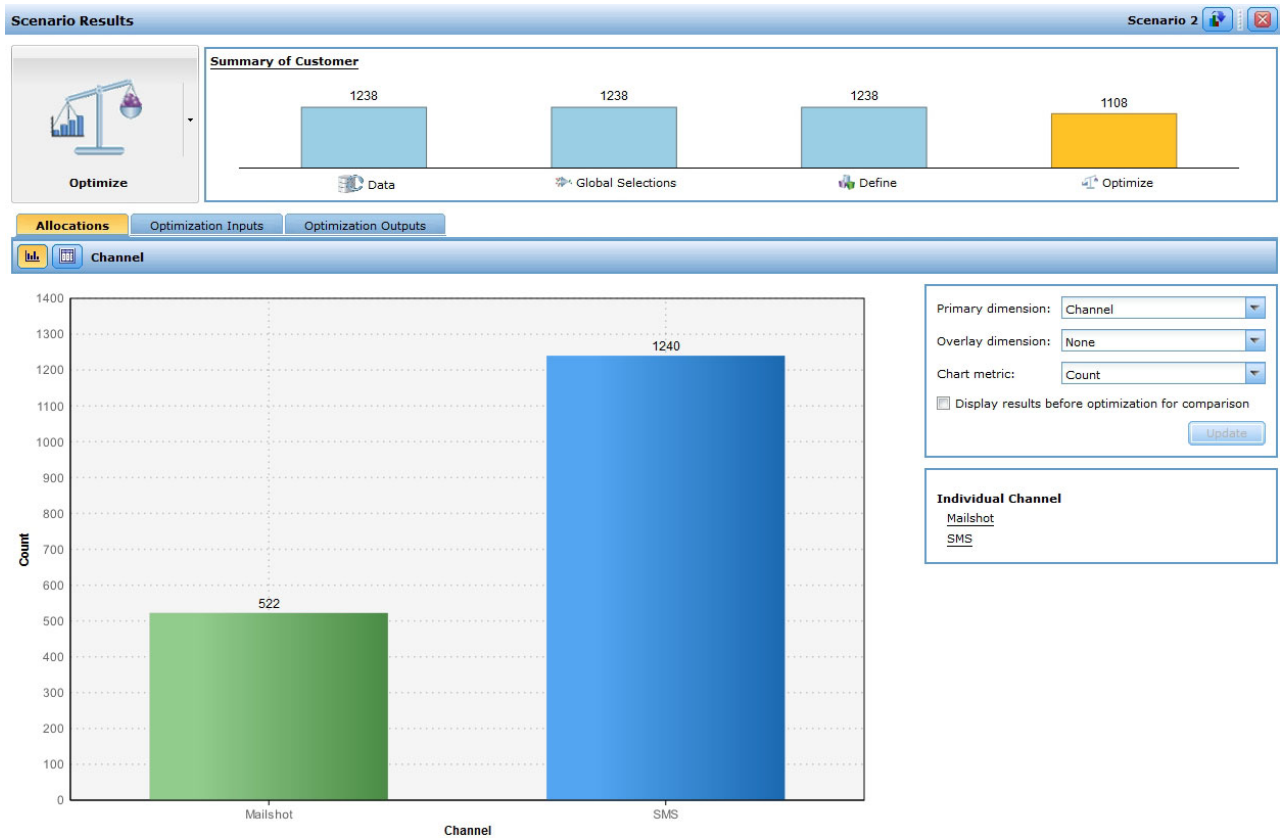
Figure 10. Scenario results

Deploying the application

After you update your project and are satisfied with the simulation results, you can apply a label to indicate this is the version of the project you want to put into production. For details, see "Deploying applications" on page 85.

Chapter 3. Navigating the workspace

Changing the displayed language

Each user can easily change the language of his or her application without having to restart any servers. For example, German-speaking users, English-speaking users, and Japanese-speaking users can all be using the same application in their own language.

From the **Tools** menu in your browser, go to **Internet Options** or **Options**, switch to the desired language, and then refresh the screen. If IBM Analytical Decision Management is not provided in your desired language, it will be displayed in English by default, or in your second language of choice if it is available.

The IBM Analytical Decision Management interface supports the following languages. Note that this list is subject to change from release to release, and that the documentation is not available in all the same languages as the user interface.

- English
- German
- Spanish
- French
- Italian
- Japanese
- Korean
- Portuguese (Brazilian)
- Russian
- Chinese (China/simplified)
- Chinese (Taiwan/traditional)

Launching applications

The Applications launch page provides configurable links to your installed applications. Each panel represents an installed application. Click the drop-down to choose from recently-saved projects, models, or rules. You can also add one or more custom shortcuts that point to specific, labeled versions.

- To access the Applications Launch page, navigate to *http://hostname:port/DM*, where *hostname* is the IP address or name of the repository machine and *port* is the application server port number. If you don't know this information, contact your administrator. You can also add a bookmark to this page in your browser, or create a shortcut on your desktop.

  If you have problems accessing the IBM Analytical Decision Management URL or any IBM SPSS Collaboration and Deployment Services URLs, an administrator may have customized them. Contact your administrator. More information is available in the *Application context roots* chapter of the *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide*.

- To create a new project, model, or rule, select **New** from the drop-down list for the desired application and click **Go**.

- To open a saved project, model, or rule, select a recently-used version or shortcut from the drop-down list in the application panel, or select **Browse** from the drop-down list to choose a different object or version. To quickly the see the repository location of an item in the list, hover over it in the drop-down menu.

- To add a new custom shortcut that will always be available in the drop-down list, click the triangular arrow icon in the upper right corner of the application panel. Optionally, you can specify the number of shortcuts to display.

Other ways to open an application

- Copy or type the URL into the address bar in your browser, or add the URL to the list of saved bookmarks or favorites in your browser.
- Create a desktop shortcut pointing to the URL.
- Link to an application URL from within your company's Intranet site or portal.

Customizing the launch page

The controls in the upper right corner of the Application Launch page allow you to add applications, and to toggle the display between panel and list view. You can also drag panels to rearrange them on the page.

- To add a new application panel, select **Add application** in the upper right corner of the Application Launch page. (If no additional applications are available, this link is not available.)

  A list of available applications is displayed. Optionally, you can create shortcuts to one or more labeled versions.

- To change the arrangement of application panels on the Launch page, click on the title bar of any panel and hold down the mouse button to drag the panel to a new location.
- To remove an application, click the remove icon (red X) in the upper right corner of the application panel. Any shortcuts that point to that application will also be removed. (The remove icon will not be available if the application panel has been locked on the page by the administrator.)
- Unsaved projects, models, and rules may be listed in the gallery on the launch page. This gallery displays models that are still building, as well as those that have completed building but are not yet saved. See the topic "Gallery" on page 25 for more information.

Project configuration

When launching an application that includes a Define tab with a dimension tree, if the selected application is configured to include more than one dimension hierarchy element, you will be prompted to select one or more of the available elements (hierarchies) to use in the application. The dimension tree on the Define tab specifies the range of possible decisions or recommendations that can be returned. If an application includes more than one dimension hierarchy, they will be included on separate tabs of the dimension tree area on the Define tab. For more information about the Define tab and its dimension tree, see Chapter 8, "Defining decisions," on page 71.

1. The **Project configuration** dialog displays all available dimension hierarchies defined in the application template. Select one or more dimension hierarchies (elements) you wish to use to define the project. A description of each element may be available as defined by the application designer in the application template.

   The element name is the name of the top-most dimension. Dimensions can only belong to one element.

2. If you select more than one element, you can sort the order in which they'll appear throughout the application and which elements can depend on each other.

   The order you specify here is important. The elements will appear in ordered tabs, and hierarchy members of previous tabs can be referred to in rules within subsequent tabs, but references cannot be made to subsequent tabs. For example, rules on a second and third tab can refer to hierarchy members in the first tab, but rules on the first tab cannot refer to members on the second or third tabs (and rules on the second tab cannot refer to members on the third tab).

3. When finished in the **Project configuration** dialog, click **OK**.

The Home page

The application provides a step-by-step workflow, as represented by the icons on the Home page. Click any icon to jump to that step.

Data. Define data sets for use by the application. Operational data pertains to current production -- such as current customers, or people you want to sell to in future -- while models are based on past data, such as results from previous campaigns.

Global Selections. Choose the records to include or exclude from all processing by the application. For example, you could enforce a mandatory exclusion to ensure that customers with bad credit are not considered for any campaign or offer.

Define. Define the range campaigns and offers that will be available, and the rules and models used to allocate customers accordingly.

Prioritize or Optimize. When multiple outcomes or recommendations are returned on the Define tab, choose the best based on the specified prioritization or optimization equation. For example, identify the best decision for each customer.

Deploy. Deploy the application for use in a test or production environment such as a call center, Web site, ATM, or store location as appropriate.

Scenarios. Run a scenario to preview the results based on current settings in the project. The latest scenario is displayed on the Home page.

Reports. Monitor the status of deployed applications.

Gallery

The gallery is displayed on the Applications Launch Page. It lists unsaved work, such as a model that has been left to build overnight, allowing you to start the model before you leave work, close your browser, and return to find your results in the morning. You can also use the gallery to retrieve your work after accidentally closing your browser. Items in the gallery may include projects, rules, models, data scans, evaluations, and other long-running tasks as appropriate to the application.

- If the name of the item is underlined, click the name to open the relevant item. For example, if you have an unsaved data scan, click the data scan title in the gallery to open the Data tab.
- If you do not require any item shown in the gallery, click the delete icon to remove it; a warning message is displayed to confirm you want to permanently delete the item. Or to remove several gallery items at once, click **Clear Gallery**. The Clear Gallery dialog box will be displayed allowing you to select which tabs to remove all objects from.
- On the Projects tab of the gallery, the **Created in** drop-down menu allows you to filter the items displayed on the tab for each application. For example, while working in an application you may not want to see gallery items for other applications.

Setting preferences

There are a number of preferences you can adjust to customize IBM Analytical Decision Management to your needs. Primarily, this customization consists of setting your own default options to save you having to repeatedly select the same actions when creating, testing, and saving your work. The preferences you set are unique to you and apply to any application you use.

To access the user preferences settings, from within any application, click the **Set user preferences** icon on the toolbar.

Figure 11. Set user preferences icon

You can specify your own preferred settings for each area as detailed in the sections that follow.

General settings

Default number of recently opened objects in application shortcuts. Select the number of items to be available from the drop-down list on the Application Launch page.

Date format. Select the format to use when displaying or recording dates.

Time format. Select the format to use when displaying or recording time.

Currency. Select the default currency for each new project. Note that this will not overwrite existing data brought in as a different currency.

Decimal places. Select the number of decimal places to display for numbers.

Number of rows to show on preview. Select the maximum number of records to show when displaying previews; for example, when checking a data source upon import.

Number of results per page (paginated tables). Select the maximum number of results to show on each page when displaying tables.

Maximum number of items displayed in a graph. Sometimes graph results do not work well when there are many items to display. Specify the maximum number of items to display in graphs. This will force a table to be displayed instead of a graph if the number of items exceeds this maximum. This setting only applies to categorical fields in data overviews and the scoring summary report graph for models with a categorical target.

Number of data segments (tiles) to be applied to charts and tables. Select the maximum number of data segments to use throughout the application for charts and tables, where applicable.

Test and simulation settings

Default number of records tested. Specify the default number of records to include when running a test.

Default test/simulation date. Specify whether the default simulation date is set to the current date, or a specific date.

Define step settings

If you are creating new dimensions, such as campaigns or claims depending on the application, you can specify the default **From** and **To** dates applied to each new item. Alternatively, for open-ended offers, you can choose **No expiration**.

Prioritize/Optimize step settings

Prioritization/Optimization enables you to compare all possible solutions for a given record and pick the one that best meets your objective; however, when you have more than one way of dealing with records you may set up different prioritization/optimization actions. For example, if you run a promotional campaign, you may have different priorities if a customer interacts with you by post or telephone.

You can specify whether to copy your existing prioritization/optimization settings to each new interaction point that you set up, or use the default values created by your system administrator.

Opening and saving projects, models, and rules

You can open from, delete from, or save to the IBM SPSS Collaboration and Deployment Services Repository, or download to the local file system. You can also upload IBM SPSS Modeler stream files from the local file system and save them to the IBM SPSS Collaboration and Deployment Services Repository. Saving to the IBM SPSS Collaboration and Deployment Services Repository provides important benefits including support for multiple versions, labeling, and security. Depending on the local configuration, not all options may be supported.

 Creates a new project.

 Saves the current project to its last known location in the IBM SPSS Collaboration and Deployment Services Repository. Optionally, you can apply a new label when saving.

 Saves and labels the current project, or saves under a new name. Where applicable, you can choose between saving to the IBM SPSS Collaboration and Deployment Services Repository, and downloading to the local file system.

 Opens a project from the IBM SPSS Collaboration and Deployment Services Repository or the local file system as applicable. Or, where applicable, you can also upload a project from the local file system. File system access is not supported for all applications. If uploading an IBM SPSS Modeler stream, you can choose to open the file or save it to the repository..

Opening from and saving to the repository

The IBM SPSS Collaboration and Deployment Services Repository provides important benefits including support for multiple versions, labeling, and security. Each time you save you create a new version, and you can revert to an earlier version at any time. You can also manage versions using labels, and apply security settings to determine who can access, view, or remove an object.

The tree view in the left pane displays folders in the IBM SPSS Collaboration and Deployment Services Repository you have authority to see. Navigate the folder tree or perform a search to find what you're looking for. You can also filter the file list by labels available in the lower-left pane. For example, you might click a *production* label to only display files marked with that label.

To see more details about a file, when available, click the preview icon or the information icon beside it.

To delete a file, click the delete icon beside it. This is especially useful for users who do not have access to the IBM SPSS Collaboration and Deployment Services Deployment Manager client software.

Folder. Displays available folders and allows you to navigate the hierarchy. Choose the folder to save to. You can also create a new folder or rename an existing folder using the icons above the folder list.

Filter by Label. Limits the display to objects with the selected label. When browsing folders, only objects with this label are shown.

Export as. When exporting rules used in an application, a **Save as** or **Export as** drop-down is also available. Select **Single ruleset** or **Individual rules** from the drop-down to specify how to save the rule. If only referenced rules are available, the drop-down will be disabled and the rule will be saved as a single ruleset (referenced rules cannot be saved individually).

If you choose to save as individual rules, a table will be displayed to enter details for each individual rule you wish to save.

If you choose to save as a ruleset, you will be given the options to **Store with Selections** or **Store with Risk points**. If the rule you are saving is a Selection rule, you can select the **Store with Selections** option to also save whether it's an inclusion or exclusion rule. If the rule you are saving is an Aggregation rule, you can select the **Store with Risk points** option to save the aggregation rule with its aggregation points.

Name. Enter a unique name for the object you are saving.

Keywords. Enter any keywords to quickly identify the saved object when searching for it. For example these might be *Sales* and *1-2009*.

Description. Enter a longer description of the object saved. For example, this might be *Analysis of retail sales; 1st quarter of 2009*.

Label(s) If you are likely to have several versions of the same saved object, you can identify each with a unique label. Either select an existing label or create a new one.

Working with labels

Labels are used to identify a specific version of a project, model, or rule. When you share a rule between multiple projects, for example, you can specify that you want to use the version labeled *production*. If a newer version of the rule is created, the one used in your project won't change unless the *production* label is explicitly moved to that version.

When referencing shared objects, using a labeled version is strongly recommended, particularly in a production environment, as any changes to the latest version may otherwise have unintended consequences. In certain cases, such as when referencing a rule from within a rule, a labeled version must be used. See the topic "Exporting and reusing rules" on page 54 for more information.

If no version is specified, the latest version is used, which means that subsequent changes to the rule will automatically be picked up. (Think of *LATEST* as a special label that is automatically rolled to each new version as it is created.)

Each time you save a new object or version, you can choose the labels to apply to that version. See the topic "Opening and saving projects, models, and rules" on page 27 for more information.

Restrictions on label names. When defining labels for use with IBM Analytical Decision Management, the characters **? # &** and **:** (colon) cannot be included in label names, as they may prevent the application from correctly interpreting references to objects that use these labels. For example, the following are not valid label names:

Test & Deploy

DEPLOY:PHASE1

Locking projects

When you open a project, model, or rule, a Lock check box is available in the upper right corner allowing you to lock the object in the IBM SPSS Collaboration and Deployment Services Repository so that other users cannot edit it.

- When a project is locked, other users can open it for viewing, but cannot save changes.
- When you close a locked project (for example by closing the browser) the lock is released so that others may edit the object.

- If you choose to unlock a project without closing it (by deselecting the Lock check box) you will be prompted to save any changes. Unsaved changes will be lost after the object is unlocked.
- After unlocking a project, you can continue to make changes to it, unless another user happens lock it while you have it unlocked (in which case you will no longer be able to save changes).

Properties

The Properties dialog includes a **Project** tab and a **Repository** tab. To display the Properties dialog, click the Properties icon in any application.

Figure 12. Properties icon

On the **Project** tab, you can specify the maximum members allowed for *Nominal* fields. This is used on the Data tab. Select **Maximum members for nominal fields** and specify a value. The default is **250**. This is useful in situations such as the following:

1. A user creates a new data source and runs a data scan.
2. A number of fields that contain useful categorical information appear in the project data model as *Typeless* because the number of distinct values in the data exceeds the default maximum size limit set in the Properties dialog.
3. The user clicks the Properties icon and increases the **Maximum members for nominal fields** value to an appropriate size for his data.
4. Back on the Data tab, the user clicks **Refresh data scan** but the fields remain as *Typeless* because the refresh action doesn't alter the measure, to preserve any changes the user has made. In other words, a refresh retains the existing measure for a field and only recomputes the values because any changes the user has made in the data model should be preserved.
5. To get the values into the project data model, the user must manually change the measure of each field from *Typeless* to *Categorical* prior to running the refresh. Note that if there are many fields affected, it may be faster to recreate the data source.

Note that the Maximum members for nominal fields setting does not apply to data sources of the file type *.sav.*. Many *.sav* files have user customized files and values, so they will not be affected by this setting. The user must manually change the fields.

On the **Repository** tab, you can add keywords and a description to help identify a project, model, or ruleset when browsing for it at a later date:

- **Keywords.** Enter any keywords to quickly identify the saved object when searching for it. For example these might be *Sales* and *Q1-2013*.
- **Description.** Enter a longer description of the object saved. For example, this might be *Analysis of retail sales; first quarter of 2013*.

Chapter 4. Managing data sources

Use the Data tab to define data sources for analysis, simulation and testing, scoring, or other use. Data sets can be configured by your administrator, or you can add your own.

The tab includes a **Source** tab and a **Derived** tab.The Source tab is divided into these main areas:

- **Project Data Model.** Defines the fields required by the application. All other data sets are mapped relative to this source. The **Data Source Fields** section lists the input fields and types your project uses. If desired, use the **Derived** tab to add expressions or model output from different data sources to extend your project data model.
- **Data Sources.** The **Project Data Sources** section lists the data sources that have been saved for use with the current project, model, or rule. The **My Data Sources** section lists data sources you have defined or saved. Data sets from this list can be copied to or from different projects, and then promoted for use as the project data model, making it possible to share data across applications.
- **Secondary Tables.** Use this section to add tables from different data sources to extend your project data model, if desired.

Project data model

The project data model acts like a template listing the input fields and types your project will use.

Data source. Displays the data source selected for the application.

Customer ID. If desired, you can select a field to use as the *entity* ID, to enable SQL push-back. This can improve performance when the data originates in a database and if the selected field is an integer.

Operational. A check in this box shows that the data in the field is available for use in rules and expressions, and for predictions. Remove the check box if the field is to be used as the target field for analysis. For example, if you are running a query to see how well customers will respond to a direct mail campaign, the results field would be analytical since you do not know the response values in advance.

Note: If you select an IBM SPSS Collaboration and Deployment Services Enterprise View data source as the project data model, the boxes in this column are disabled.

Field name. Displays the name of each field, as listed in the data source, along with an icon that identifies its storage type.

Measurement. Displays the measurement type of each field in the data source. If you change any of the measurement types, you must refresh the data scan to ensure data compatibility. See the topic "Measurement levels" on page 41 for more information.

Values. Lists the values for each field in the data source. For example, maximum and minimum values in a range. For flag fields, hold the mouse pointer over the value to display a tool tip that indicates the "true" and "false" values defined for the field. You can also click a value to edit it.

Typically, the project data model defines the inputs as a set of fields coming from a single logical table. This is always the case for IBM SPSS Modeler Advantage. However, for other applications, the project data model may consist of a primary table along with 0 or more secondary tables. These secondary tables are defined by the secondary data sources that are associated with the primary data source.

Project data sources

Project data sources may include those predefined by the administrator, or added by users. Optionally, the administrator may have locked one or more data sources to prevent users from modifying or removing them, or locked all data options so users can't create new data sources.

If a data source's inputs don't directly match those of the project data model, you can map the former to the latter and fix the discrepancy. For example, if the project data model requires a field named *purchase* with values *Yes* and *No* (measurement level flag), then any data source used must have a comparable field that can be mapped accordingly.

Name. Displays the data source name and shows an icon that identifies its file type.

No. of records. Click the icon in this column to show the record count for any data source in the table. The count is shown beside the icon.

Preview. Click the icon in this column to preview a sample of the data contained in the source. For more information, see "Previewing data" on page 40.

Overview. Click the icon in this column for an overview of the data source. For more information, see "Data overview" on page 40.

Compatible. Either displays a note that the data source is used as the project data model, or shows a green, orange, or red ball to indicate how compatible the data source is with the project data model.
- A green ball shows that the data source is operationally compatible with the project data model data source. An operationally compatible data source is one that includes all the operational fields of the project data model, but can have additional fields. This data source is suitable for rules, scoring, simulation, and test operations.
- An orange ball shows that the data source has at least one field that is compatible with the project data model, with the same name and type. This data source may also have additional fields, and is suitable for building and evaluating models.
- A red ball shows that the data source is incompatible with the project data model, and fields must be mapped before it can be used in the application. An incompatible data source is one which has at least one field whose type is incompatible with the equivalent project data model type.

Map fields. This option allows you to compare the data source fields with those in the project data model and map or unmap any compatible fields to match those required by the project data model. For more information, see "Mapping fields" on page 42.

Copy. Copies the data source to the **My Data Sources** area.

My data sources

Data sources on this list are saved with your user account, so they are available whenever you log in, and can be copied to any project, model, or rule that you open (assuming your administrator gives you authority to do so). Fields in this part of the tab work in the same way as those in the **Project Data Sources** area; although there is no **Compatibility** column.

The **Copy** column enables you to copy the data source into the **Project Data Sources** area.

Secondary tables

To add tables from different data sources, click **Add/Edit Secondary Tables**. Secondary tables can be used for dynamic table-based allocation, or for extending your project data model, enabling model outputs to be added with different data sources. For more information, see "Adding secondary tables" on page 43.

Note: This **Secondary Tables** section is not available in the IBM SPSS Modeler Advantage application. The primary reason for building models in IBM SPSS Modeler Advantage is to reuse them elsewhere in IBM Analytical Decision Management. If a model is deployed with multiple inputs, it cannot be used elsewhere in IBM Analytical Decision Management, so this functionality for adding tables is not included in the application.

Working with data sources

- To add a new data source to either the **Project Data Sources** or **My Data Sources** lists, select **Add a data source**. For more information, see "Creating a new data source."
- To change the measurement level, or type, of a field in the data source in the **Project Data Model**, select the relevant level. For more information, see "Measurement levels" on page 41.
- To copy data sources to or from the **Project Data Sources** list, click the appropriate arrow in the Copy column.
- To map field names for a data source to the project data model, click the appropriate link under the Compatible column. (Once field names are mapped, the link is no longer displayed.) For more information, see "Mapping fields" on page 42.
- To preview a data source, click the **Preview** icon. For more information, see "Previewing data" on page 40.
- To add secondary tables to extend your project data model, click **Add/Edit Secondary Tables** in the Secondary Tables section. For more information, see "Adding secondary tables" on page 43.
- To add additional fields (expressions, fields using segment rules, or model output from different data sources) to extend your project data model, click the **Derived** tab. For more information, see "Deriving fields" on page 43.

Creating a new data source

1. To access the Data Source Editor, select **Add a data source** under either the Project Data Sources or My Data Sources area.
2. Type a name for the data source.
3. Select the source type. Note that the list of available databases is from those available on the IBM SPSS Modeler machine.
    - **File.** Many file types are supported, such as text (*.txt*, *.csv*, etc.), Excel, the *.sav* and *.zsav* formats used by IBM SPSS Statistics, and SAS files. See the topic "File source" for more information.
    - **Database.** Any database that supports ODBC, including Microsoft SQL Server, DB2, Oracle, and others. See the topic "Database source" on page 36 for more information.
    - **Enterprise View.** Creates a connection to an IBM SPSS Collaboration and Deployment Services Enterprise View as defined in IBM SPSS Collaboration and Deployment Services. See the topic "Enterprise View source" on page 37 for more information.
    - **Data View.** Creates a connection to be used with IBM Operational Decision Management rules. For complex rules that use a hierarchical data model, note that administrators may restrict use of such rules because they can degrade performance. See the topic "Data View source" on page 37 for more information.
    - **IBM Cognos BI server.** Any database that supports IBM Cognos BI server. See the topic "IBM Cognos BI source" on page 38 for more information.
4. Specify options as appropriate for the source type.

    You can click the **No. of records** icon to display the number of records in the current data source, or click **Preview** to display the first ten records of the data source.

File source

The application supports several data file types. When you enter the filename or browse for the file, the application automatically detects the type and expands the Data Source Editor to display additional fields

that relate to that type. If the file type is incorrect, or requires a specific version (for example, Microsoft Excel 1997-2003), you can change the type to the one required.

You can click **Browse** to select a file from the repository, or click **Upload local file** to select a file from your local file system (if enabled).

If you upload a local file, it will be uploaded to the IBM SPSS Modeler Server. In the Upload file dialog, browse to and select the local file you want to upload and then browse to and select the IBM SPSS Modeler Server upload destination. When you select the destination, you can change the file name if desired.

The **Upload local file** capability is disabled by default. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to enable or disable it, and to increase or decrease the maximum file size limit allowed.

Text-based data files

When you select a text-based data source, you are prompted to enter further details.

First row has column names. Select this if the names of each column are included as a heading row in the data source.

Encoding. Specifies the text-encoding method used. You can choose between the system default or UTF-8.

Decimal symbol. Specifies how decimals should be represented in the data.
* **Stream default.** The decimal separator defined by the current stream's default setting will be used. This will normally be the decimal separator defined by the computer's locale settings.
* **Period (.).** The period character will be used as the decimal separator.
* **Comma (,).** The comma character will be used as the decimal separator.

Delimiters. Using the check boxes listed for this control, you can specify which characters, such as the comma (,), define field boundaries in the file. You can also specify more than one delimiter, such as ", |" for records that use multiple delimiters. The default delimiter is the comma.

Note: If the comma is also defined as the decimal separator, the default settings here will not work. In cases where the comma is both the field delimiter and the decimal separator, select **Other** in the Delimiters list. Then manually specify a comma in the entry field.

Select **Allow multiple blank delimiters** to treat multiple adjacent blank delimiter characters as a single delimiter. For example, if one data value is followed by four spaces and then another data value, this group would be treated as two fields rather than five.

Advanced options

EOL comment characters. Specify characters, such as # or !, to indicate annotations in the data. Wherever one of these characters appears in the data file, everything up to but not including the next new-line character will be ignored.

Specify input fields. Specify the number of input fields to be used from each record.

Specify data format For **File** data source types that are of type **Variable length fields in plain text**, you can use this section to set the input storage type and the format of each field to ensure that values are read correctly. This is similar to the functionality available in IBM SPSS Modeler client. The **Override** option indicates whether the default is overridden. Selecting **Override** enables the **Storage** and **Input**

Format controls. Deselecting **Override** will change the values back to their original defaults. **Input Format** only applies real, date, time, and timestamp storage types.

Skip header characters. Specify how many characters you want to ignore at the beginning of the first record.

Lines to scan for type. Specify how many lines to scan for specified data types.

Strip lead and trail spaces. Select to discard leading and trailing spaces in strings on import. You can choose to strip from the left, right, both sides, or none.

Invalid characters. Select **Discard** to remove invalid characters from the data source. Select **Replace with** to replace invalid characters with the specified symbol (one character only). Invalid characters are null characters or any character that does not exist in the encoding method specified.

Quotes. Using the drop-down lists, you can specify how single and double quotation marks are treated on import. You can choose to **Discard** all quotation marks, **Include as text** by including them in the field value, or **Pair and discard** to match pairs of quotation marks and remove them. If a quotation mark is unmatched, you will receive an error message. Both **Discard** and **Pair and discard** store the field value (without quotation marks) as a string.

Note: When using **Pair and discard**, spaces are kept. When using **Discard**, trailing spaces inside and outside quotes are removed (for example, `' " ab c" , "d ef " , " gh i " '` will result in `'ab c, d ef, gh i'`). When using **Include as text**, quotes are treated as normal characters, so leading and trailing spaces will be stripped naturally.

Specify input fields. See the topic "Selecting input fields" on page 40 for more information.

Excel data files

When you select an Excel data source, you are prompted to enter further details:

First row has column names. Select this if the names of each column are included as a heading row in the data source.

Named range. Enables you to select a named range of cells as defined in the Excel worksheet. If you use a named range, other worksheet and data range settings are no longer applicable and are disabled as a result.

Choose worksheet. Specifies the worksheet to import, either by name or by index.
- **By name.** Select the name of the worksheet you want to import.
- **By index.** Specify the index value for the worksheet you want to import, beginning with 0 for the first worksheet, 1 for the second worksheet, and so on.

Range on worksheet. You can import data beginning with the first non-blank row or with an explicit range of cells.
- **Range starts on first non-blank row.** Locates the first non-blank cell and uses this as the upper left corner of the data range.
- **Explicit range of cells.** Enables you to specify an explicit range by row and column. For example, to specify the Excel range A1:D5, you can enter A1 in the first field and D5 in the second (or alternatively, R1C1 and R5C4). All rows in the specified range are returned, including blank rows.

On blank rows. If a blank row is encountered, you can choose whether to skip and ignore the row or choose **Return blank rows** to continue reading all data to the end of the worksheet, including blank rows.

Specify input fields. See the topic "Selecting input fields" on page 40 for more information.

IBM SPSS Statistics data files

When you select an IBM SPSS Statistics data source (*.sav* or *.zsav* file), you are prompted to enter further details. If the file is password protected, you will also be prompted to enter the password.

Variable names. Select a method of handling variable names and labels upon import from a IBM SPSS Statistics *.sav* or *.zsav*file.

- **Read names and labels.** Select to read in both variable names and labels; this is the default option. Labels may be displayed in charts, model browsers, and other types of output.
- **Read labels as names.** Select to read in the descriptive variable labels from the IBM SPSS Statistics *.sav* file rather than the short field names, and use these labels as variable names.

Values. Select a method of handling values and labels upon import from a IBM SPSS Statistics *.sav* or *.zsav*file.

- **Read data and labels.** Select to read in both actual values and value labels; this is the default option.
- **Read labels as data.** Select if you want to use the value labels from the *.sav* or *.zsav* file rather than the numerical or symbolic codes used to represent the values. For example, selecting this option for data with a gender field whose values of 1 and 2 actually represent *male* and *female*, respectively, will convert the field to a string and import *male* and *female* as the actual values.

  It is important to consider missing values in your IBM SPSS Statistics data before selecting this option. For example, if a numeric field uses labels only for missing values (0 = *No Answer*, –99 = *Unknown*), then selecting the option above will import only the value labels *No Answer* and *Unknown* and will convert the field to a string. In such cases, you should import the values themselves.

Specify input fields. See the topic "Selecting input fields" on page 40 for more information.

Database source

When you select a Database source, you are prompted to enter further details:

Database. Select the database you want to use.

Connect. Select whether to connect to the database using database credentials or credentials stored in the IBM SPSS Collaboration and Deployment Services Repository. Then either Enter your database login name and password or select a credential from the repository and click **Connect** to check that you have access to the database.

Note that database login name and password are not saved. So the next time you open and run the project you will be prompted to log on, unless your administrator configured Single Sign On (SSO) for IBM SPSS Modeler Server and ODBC database connections. See the *Security Considerations* section of the IBM Analytical Decision Management installation guide for more information.

Table Name. Select the table from which data is to be used.

Strip lead and trail spaces. Select to discard leading and trailing spaces in strings on import. You can choose to strip from the left, right, both sides, or none.

Quote table and column names. Specify whether you want table and column names to be enclosed in quotation marks when queries are sent to the database (if, for example, they contain spaces or punctuation).

- The **As needed** option will quote table and field names *only* if they include nonstandard characters. Nonstandard characters include non-ASCII characters, space characters, and any non-alphanumeric character other than a full stop (.).

- Select **Always** if you want *all* table and field names quoted.
- Select **Never** if you *never* want table and field names quoted.

Specify input fields. See the topic "Selecting input fields" on page 40 for more information.

IBM Analytical Decision Management supports the same database data sources as IBM SPSS Modeler for reading, writing, and pushing data. For details, see the product compatibility report for IBM SPSS Modeler 17. Note that this is different from the databases supported by IBM SPSS Collaboration and Deployment Services 7, which is the database that hosts the repository.

Enterprise View source

When you select an Enterprise View source, you are prompted to enter further details:

Application View. The drop-down list displays the selected Application View, if any. If connections have been made to other Application Views in the current session, these also appear on the drop-down list. Click **Browse** to search for other Application Views in the repository.

Table. Select the Application View table from which data is to be used.

Data provider. The drop-down list displays the names of Data Provider Definitions that reference the selected Application View. Select a Data Provider Definition or select the **-no data source-** option. Selecting the **-no data source-** option means the data source is effectively unusable in the application and will not be available in Test, Simulation, etc.

Environment. The environment value, which is set in IBM SPSS Collaboration and Deployment Services, determines which Data Provider Definitions are available.

Specify input fields. See the topic "Selecting input fields" on page 40 for more information.

For more details about the IBM SPSS Collaboration and Deployment Services Enterprise View, and complete details about how this data is configured, see the IBM SPSS Collaboration and Deployment Services documentation. Note that the IBM SPSS Collaboration and Deployment Services Enterprise View will be deprecated in a future release.

Data View source

A Data View source creates a connection to be used with IBM Operational Decision Management rules. For complex rules that use a hierarchical data model, note that administrators may restrict use of such rules because they can degrade performance. For details, see "Locking data options" on page 99.

A *Data View node* can be configured in IBM SPSS Modeler to include data in streams that was defined in an IBM SPSS Collaboration and Deployment Services analytic data view. An analytic data view defines a structure for accessing data that describes the entities used in predictive models and business rules. The view associates the data structure with physical data sources for the analysis.

For more information, see the IBM SPSS Modeler documentation and the *Source, Process, and Output Nodes* section of the IBM SPSS Modeler help.

When you select a Data View source, you are prompted to enter further details:

Analytic data source. Select the analytic data source from the repository.

Table name. Select the table from which data is to be used. This is a table from the data model in the analytic data view. Each table in the data model represents a concept or entity involved in the predictive analytic process. Fields for the tables correspond to attributes of the entities represented by the tables. For example, if you are analyzing customer orders, your data model could include a table for customers and

a table for orders. The customers table might have fields for customer identifier, age, gender, marital status, and country of residence. The orders table might have fields for the order identifier, number of items in the order, total cost, and the identifier for the customer who placed the order. The customer identifier field could be used to associate the customers in the customers table with their orders in the orders table.

Data Access Plan. Select the data access plan from which data is to be used. A data access plan associates the data model tables in an analytic data view with physical data sources. An analytic data view typically contains multiple data access plans. When you change the data access plan in use, you change the data used by your project. For example, if the analytic data view contains a data access plan for training a model and a data access plan for testing a model, you can switch from training data to testing data by changing the data access plan in use.

Optional Attributes. If available and if desired, select optional attributes to add to the data source.

Specify input fields. See the topic "Selecting input fields" on page 40 for more information.

IBM Cognos BI source

You can import IBM Cognos BI database data or single list reports for use in IBM Analytical Decision Management.

Note: If the IBM SPSS Modeler Server in use with IBM Analytical Decision Management does not support IBM Cognos BI server, then this option is not available.

When you select **IBM Cognos BI server** and click **Connect** you are prompted to enter the server connection details.

1. Type the **Server URL** of the IBM Cognos server from which to import or export data. Contact your Cognos system administrator if you are not sure which URL to use.
2. Select the **Mode** in which you will connect. To log in as a specific user (for example, as an administrator), select **Set credentials** and enter the IBM Cognos namespace, user name, and password:
   - Enter the IBM Cognos security authentication provider **Namespace** used to log on to the server. The authentication provider is used to define and maintain users, groups, and roles, and to control the authentication process.
   - Enter the IBM Cognos **User name** with which to log on to the server.
   - Enter the **Password** associated with the specified user name.

   OR

   Browse to and select an IBM Cognos credential stored in the IBM SPSS Collaboration and Deployment Services Repository.
3. Alternatively, to log in with no user credentials, in which case you cannot specify the namespace, user name, and password fields, select **Use Anonymous Connection**.

   Note: Some server connections do not allow anonymous connections.

Selecting Cognos object details

After setting up the IBM Cognos server connection, select the mode. **Data** mode allows you to browse to and select a package from which to import data. **List report** mode allows you to browse to and select a predefined list report from which to import data.

A package contains an IBM Cognos model and all of the folders, queries, reports, views, shortcuts, URLs, and job definitions associated with that model. An IBM Cognos model defines business rules, data descriptions, data relationships, business dimensions and hierarchies, and other administrative tasks.

1. To select a package, select **Data** mode and click **Browse**. A dialog box is displayed listing all available packages from which you can import content.

2. Or to select a single list report, select **List report** mode and click **Browse**. A dialog box is displayed listing packages and folders from which you can select a report. *Note*: Only single list reports can be imported; multiple lists are not supported.

3. Where necessary, expand the list to find the package or report you require. Select the item and click **OK**. The package or report details are shown in the Data Source Editor.

4. After you select a package, you then choose the objects to import. For example, you select the package object query subjects (which represent database tables) or individual field query items (which represent table columns) to use from a package. *Note*: The data must be in UTF-8 format.

5. Select the **Aggregate data before importing** option if you want to import aggregated data rather than raw data. Aggregated data generally has many fewer records than raw data.

Select Objects. Displays the objects (also known as namespaces) you can select from the package or list report. Expand one or more objects to display the input fields you can import. Selecting an object automatically selects all of its input fields.

Review Parameters. If parameters have been defined, either for a data object or a report, you can specify values for these parameters before importing the object or report. This table contains the list of parameters required for filters and calculations. These may have default values. A warning is provided if one or more parameters does not have a default. Set parameters as appropriate in the **Value** column. You can use parameters to make adjustments (for example, to perform a parameterized calculation) before importing the data.

Review Filters. If a package has filters defined, you can import one or more of them. If a filter you import is associated with imported data, that filter is applied before the data is imported. Filters may or may not be parameterized.

Review Input Fields. Lists the input fields (also known as database objects) that will be imported into IBM Analytical Decision Management for processing. If you do not require a particular field, deselect the check box next to its field name.

Cognos object icons: The various types of objects you can import from a Cognos BI database are represented by different icons, as the following table illustrates.

Table 3. Cognos object icons.

Icon	Object
	Package
	Namespace
	Query subject
	Query item
	Measure dimension
	Measure
	Dimension
	Level hierarchy
	Level

Table 3. Cognos object icons (continued).

Icon	Object
![filter icon]	Filter
![report icon]	Report
![calculation icon]	Standalone calculation

Selecting input fields

When you have a large amount of data, you may want to reduce or fine-tune the data imported from your source (for example, to only use the fields that are relevant to your area of interest).

1. At the bottom of the Data Source Editor dialog, click **Specify Input Fields**; the dialog expands to display all of the data fields in your source. In addition, all expressions you create are also displayed. For details about creating expressions, see "Expression Builder" on page 44.

2. Select each of the fields and expressions you want to use.

In addition to selecting data you can, where required, change the measurement level for one or more of the data fields. For more information, see "Measurement levels" on page 41.

Previewing data

To verify that you selected the correct data source, click the **Preview** icon to display a sample of the data that is contained in the source.

Data overview

The data overview feature is available on the Data tab of all applications, on the Modeling tab and the Score tab in IBM SPSS Modeler Advantage, and on the Deploy tab of applications that include the Score Now feature.

1. To run an overview of a data source, click the Data Overview icon available throughout the application.

Figure 13. Data Overview icon

2. The Data Overview dialog will appear. If desired, select an overlay field from the drop-down in the Overview Options section.

   Then after running the overview, tabs will be available to display results for the primary selected field only, or to overlay the primary selected field with the field specified in this drop-down. For example, in the results you may want to view information about the **Age** of customers, and then overlay it with another field such as **Gender**.

3. Select the data fields to include in the overview and click **Run Overview**. All fields available in the data source are listed. All fields are selected by default.

   The data overview will run and the results will appear. You can sort the information and choose which columns to display.

4. Click any field to see its details. A new results tab will open for each field you select, allowing you to view charts and tables and select overlay fields, if available.

Measurement levels

The measurement level helps characterize the information represented by each data field, and may determine how a given field is used in rules, modeling, or other applications. The measurement level can be specified in the Project Data Model for a data source; for example, you may want to set the measurement level for an integer field with values of 0 and 1 to *Flag*, to indicate that 0 = *False* and 1 = *True*. Alternatively, you can change the level in the **Data Source Editor** dialog when you specify the input fields to be used. For more information, see "Selecting input fields" on page 40.

The following measurement levels are available:

- **Default** Data whose storage type and values are unknown (for example, because they have not yet been read) are displayed as **<Default>**.
- **Continuous** Used to describe numeric values, such as a range of 0–100 or 0.75–1.25. A continuous value can be an integer, real number, or date/time.
- **Categorical** Used for string values when an exact number of distinct values is unknown. This is an **uninstantiated** data type, meaning that all possible information about the storage and usage of the data is not yet known. Once data have been read, the measurement level will be *Flag*, *Nominal*, or *Typeless*, depending on the maximum number of members for nominal fields specified in the Project Properties dialog box.
- **Flag** Used for data with two distinct values that indicate the presence or absence of a trait, such as `true` and `false`, `Yes` and `No` or `0` and `1`. The values used may vary, but one must always be designated as the "true" value, and the other as the "false" value. Data may be represented as text, integer, real number, date, time, or timestamp.
- **Nominal** Used to describe data with multiple distinct values, each treated as a member of a set, such as `small/medium/large`. Nominal data can have any storage—numeric, string, or date/time. Note that setting the measurement level to *Nominal* does not automatically change the values to string storage. For information about setting the maximum members allowed for nominal fields, see "Properties" on page 29.
- **Ordinal** Used to describe data with multiple distinct values that have an inherent order. For example, salary categories or satisfaction rankings can be typed as ordinal data. The order is defined by the natural sort order of the data elements. For example, `1, 3, 5` is the default sort order for a set of integers, while `HIGH, LOW, NORMAL` (ascending alphabetically) is the order for a set of strings. The ordinal measurement level enables you to define a set of categorical data as ordinal data for the purposes of visualization, model building, and export to other applications (such as IBM SPSS Statistics) that recognize ordinal data as a distinct type. You can use an ordinal field anywhere that a nominal field can be used. Additionally, fields of any storage type (real, integer, string, date, time, and so on) can be defined as ordinal.
- **Typeless** Used for data that does not conform to any of the above types, for fields with a single value, or for nominal data where the set has more members than the defined maximum. It is also useful for cases in which the measurement level would otherwise be a set with many members (such as an account number). When you select **Typeless** for a field, the role is automatically set to **None**, with **Record ID** as the only alternative. The default maximum size for sets is 250 unique values. This number can be adjusted or disabled in the Project Properties dialog box, which can be accessed from the toolbar icon.
- **Collection** Used to identify non-geospatial data that is recorded in a list. A collection is effectively a list field of zero depth, where the elements in that list have one of the other measurement levels.
- **Geospatial** Used with the List storage type to identify geospatial data. Lists can be either List of Integer or List of Real fields with a list depth that is between zero and two, inclusive.

Editing data field values

In data tables, the **Values** column lists the values for each field in the data source (for example, maximum and minimum values in a range). For flag fields, hold the mouse pointer over the value to display a tool tip that indicates the "true" and "false" values that are defined for the field. You can also click a value to edit it. A dialog box opens, and its contents vary depending on the type of data field selected. For example:

- If the field type is **Continuous**, you can enter upper and lower boundary values.
- If the field type is **Flag**, you can swap the true and false values.
- If the field type is **Nominal**, you can add, remove, reorder, and rename values.

Mapping fields

When adding a new data source, you may need to map all of its fields to ensure compatibility with the project data model. For example, if the project data model requires a field named *purchase* with values *Yes* and *No* (measurement level flag), then any data source used must have a compatible field. If the field names are not identical, they can be mapped accordingly. Note that the input and associated mapped field must have the same data type.

Alternatively, if the project data model changes, then it may be that many data sources no longer map to it correctly. If that is the case, then the compatibility indicator ball in the **Project Data Sources** area will change to the appropriate color for each of those data sources; you will then have to remap all of them to the new project data model.

1. In the **Project Data Sources** area, click **Map fields**. A dialog opens, and the system makes a best guess at mapping fields for you. At any point, you can return to the system's best guess by clicking **Reset**.
2. Select a field in the **Available Inputs** column, and its corresponding field in the **Required Input Fields** column, and click **Map** to map the two fields. Repeat this until all fields in the **Available Inputs** column have been mapped.
3. When finished, click **OK**.

When possible, IBM Analytical Decision Management will automatically map compatible storage types when a user adds an additional data source. If the type is not compatible, the user will be allowed to convert types.

The following table lists the allowed mappings.

Table 4. Types that can be converted

To → From ↓	Int	Real	String	Date	Time	timestamp
Int	n/a	Auto	Yes	Not allowed	Not allowed	Not allowed
real	Yes (decimal digits truncated)	n/a	Yes	Not allowed	Not allowed	Not allowed
String	Yes	Yes	n/a	Yes	Yes	Yes
Date	Not allowed	Not allowed	Yes	n/a	Not allowed	Yes (time as 00:00)
Time	Not allowed	Not allowed	Yes	Not allowed	n/a	Yes (base date)
Timestamp	Not allowed	Not allowed	Yes	Auto (time truncated)	Auto (date truncated)	n/a

Adding secondary tables

In most IBM Analytical Decision Management applications, you can add additional tables to your project data model. The **Secondary Tables** section at the bottom of the Data tab displays any tables that are to be included as part of the data.

Note: The **Secondary Tables** section is not included in IBM SPSS Modeler Advantage or the IBM SPSS Rules Management application. The primary reason for building models in IBM SPSS Modeler Advantage is to reuse them elsewhere in IBM Analytical Decision Management. If a model is deployed with multiple inputs, it cannot be used elsewhere in IBM Analytical Decision Management, so this functionality for adding secondary tables is not included in IBM SPSS Modeler Advantage.

To add a table:

1. On the Data tab, in the Secondary Tables section, click **Add/Edit Secondary Tables**.
2. To extend the project data model by enabling model outputs to be added from different data sources, click the **Add a new table** icon.
3. Type a name for the new table. This will add an empty table and require you to select a data source and key after you exit this dialog box. If the project data model uses an Application View, you can save time by selecting from a list of existing Application View tables.

   OR

   If your project data model uses a Data View data source, you can select an existing analytic data source table from the repository. Then, after you add the table, a **Relationship** drop-down is available so you can change the relationship between the additional table you added and the project data model, if desired.
4. When finished in the **Add a New Table** dialog, click **Save**. A new tab is added for every additional table defined.
5. Select a data source and then click the **Links** button. Use the Links dialog to copy linked fields from a relationship or select an individual link field. Or you can select not to link tables (to read dimension members from the entire secondary table when using dynamic allocation, for example), or select the option for a one-to-one relationship between rows in your data. By specifying the one-to-one option, you can then use the Derived tab to add fields from the secondary table to your project data model.
6. When finished in the Links dialog, click **Save** to return to the **Add/Edit Secondary Tables** dialog.
7. If you need to edit any tables you add, click the **Edit tables** icon.

Deriving fields

In most IBM Analytical Decision Management applications, you can add additional fields to your project data model such as expressions, model output, fields created from segment rules, or fields from secondary tables. The **Derived** tab of the Data tab displays any fields that are to be included as part of the data, and the fields can be ordered so they can be used cumulatively (their order in the table will match the order in which they are calculated; for example, the second derived field can use the first derived field as an input).

The **Derived** tab is not included in the IBM SPSS Rules Management application.

- From the Data tab, click the **Derived** tab.

To add an expression:

- Click the **Derive field using an expression** icon. The Expression Builder will be displayed. For more information, see "Expression Builder" on page 44.

To add model output:

1. Click the **Derive field using a model or rule** icon then browse for and select a model. The **Inputs and Outputs** dialog for the selected model will be displayed.

2. If the model input fields are compatible, the **Model Input Mapping** section of the dialog will be collapsed. If the model input fields are not compatible, the section will be expanded and you must map available table inputs to inputs of the selected model.

3. In the **Model Outputs** section, all available model outputs for the selected model are listed. Select the model output fields you want to add to your data. The **Model output name** column displays the name of the output field as generated by the model.

   Any fields the model outputs can be selected. The model output fields are all the fields coming out of the stream which are not also inputs to the stream (all fields added by the stream). The input fields may not include everything in the project data model, so it's possible for a model output field to have the same name as a project data model field. So you can also change the data field names here if desired.

   The **Measurement** column displays the measurement type of the model output prefixed with the appropriate type icon, and the **Values** column displays the values of the model output. Note that these columns may display results for some of the fields and not others. Typically the measurement types and values are known for actual model outputs (predicted value, probability, etc.) without a data scan, so if known it will be displayed. But the measurement types or values may not be known for other output fields from a stream (for example, a derived field) without a data scan.

4. A **Model Parameters** section will be included at the bottom of the dialog if the selected stream includes scoring parameters. Enter values for the model parameters.

5. Click **Save** when finished in the dialog. Any selected model outputs will be added to the **Additional Fields** section of the Derived tab.

To create a field using segment rules:

1. Click the **Derive field using segment rules** icon.

2. Create a new segment rule. Segments are defined by one or more expressions that return a true or false result. For more information, see "Defining segment rules" on page 51.

3. Click **OK** when finished in the dialog. The new segment rule will be added to the **Additional Fields** section of the Derived tab.

To add a field from a secondary table:

1. If you defined a secondary table with a one-to-one relationship on the Data tab, click the **Add field from a secondary table** icon.

2. Select a secondary table from the drop-down and then select one or more table fields.

3. Click **Save** when finished in the dialog. The field(s) will be added to the **Additional Fields** section of the Derived tab.

Expression Builder

You can use the expression builder to derive additional fields or attributes for use by the application. For example, if you are using banking data, you may want to create an expression that shows the ratio between a customer's income and the number of loan accounts they have.

Expressions are available throughout the application in the same way as operational data (for example, in rules and in model predictors).

To display existing expressions, or create a new expression, go to the Data tab and click the **Derived** tab.

Creating an expression

1. In the Expression Builder dialog, enter a unique name for the expression.

2. Type in the **Expression Equation** field, using the function and field lists as references

   or

3. Select the desired fields and functions from the lists.

4. Double-click or click the blue arrow to add the field or function to the **Expression Equation** field.
5. Use the operand buttons in the center of the dialog to insert operations into the expression.
6. Enter a **Description** that explains what the expression is used for.
7. Click **Check** to validate the expression, then click **Save**.

Note: You can copy and paste expressions from IBM SPSS Modeler client, but note that the results may not always be as expected if the expression is stream-dependent (for example, if you paste a supported @ function or row function, it may not produce the desired results because the rest of the stream may not be set up to support it).

The only expressions that are officially supported are the ones provided in the IBM Analytical Decision Management user interface.

Selecting functions

The function list displays all available functions and operators. Scroll to select a function from the list, or, for easier searching, use the drop-down list to display a subset of functions or operators. Available functions are grouped into categories for easier searching.

Following are details about the categories. For more detailed information, refer to the IBM SPSS Modeler documentation.

Table 5. Function categories

Function Type	Description
Recently Used Functions	Contains a list of functions used during the current session.
Operators	Lists all the operators you can use when building expressions. Operators are also available from the buttons in the center of the dialog
Information Functions	Used to gain insight into field values. For example, the function `is_string` returns true for all records whose type is a string.
Conversion Functions	Used to construct new fields or convert storage type. For example, the function `to_timestamp` converts the selected field to a timestamp.
Comparison Functions	Used to compare field values to each other or to a specified string. For example, `<=` is used to compare whether the values of two fields are lesser or equal.
Logical Functions	Used to perform logical operations, such as `if`, `then`, `else` operations.
Numeric Functions	Used to perform numeric calculations, such as the natural log of field values.
Random Functions	Used to randomly select items or generate numbers.
String Functions	Used to perform various operations on strings, such as `stripchar`, which allows you to remove a specified character.
Datetime Functions	Used to perform various operations on date, time, and timestamp fields.
Blanks And Null Functions	Used to access, flag, and frequently fill user-specified blanks or system-missing values. @BLANK is not supported (because its behavior is dependent on other nodes like the Type node). `@NULL(FIELD)` and `undef` are supported. `@NULL(FIELD)` is type Boolean and returns true if the value of `FIELD` is the system-missing `$null$`. Returns false for all other values, including user-defined blanks. If you want to check for both, use `@BLANK(FIELD)` and `@NULL(FIELD)`. `undef` is type Any and is used generally to enter a `$null$` value—for example, to fill blank values with nulls in the Filler node.
All Functions	Contains a complete list of available functions.

Selecting fields and values

The field list displays all fields available. Scroll to select a field from the list. Double-click or click the blue arrow to add a field to the expression.

Viewing and selecting values

Field values can be viewed from a number of places in the application. To view values for a field from the Expression Builder, select the desired field and click the expression inputs button to open a dialog box listing values for the selected field. You can then select a value and click **Insert** to paste the value into the current expression.

Figure 14. Expression inputs button

For flag and nominal fields, all defined values are listed. For continuous (numeric range) fields, the minimum and maximum values are displayed.

Checking expressions

Click **Check** in the Expression Builder (lower right corner) to validate the expression. Expressions that have not been checked are displayed in red. If errors are found, a message indicating the cause is displayed.

The following items are checked:
- Correct quoting of values and field names
- Correct usage of parameters and global variables
- Valid usage of operators
- Existence of referenced fields
- Existence and definition of referenced globals

If you encounter errors in syntax, try creating the expression using the lists and operator buttons rather than typing the expression manually. This method automatically adds the proper quotes for fields and values.

Note the following limitations when building expressions in IBM Analytical Decision Management. Expressions cannot contain any of the following items:
- A reference to an IBM SPSS Modeler stream parameter
- A reference to an IBM SPSS Modeler stream global
- A reference to a database function
- A reference to one of the following special field or field value @ functions:
  - @TARGET
  - @PREDICTED
  - @FIELD
  - @PARTITION_FIELD
  - @TRAINING_PARTITION
  - @TESTING_PARTITION
  - @VALIDATION_PARTITION

Using expressions that return a Boolean

IBM SPSS Modeler and IBM Analytical Decision Management do not have a Boolean data type. Boolean expressions cannot be used to directly produce a new field. The use of Boolean expressions is limited to expression constructs that expect a boolean expression (such as the `if` statement). Following are three examples:

```
if ('RFM Score' > 5) then "T" else "F" endif
if ('RFM Score' < 5) then "Low" elseif  ('RFM Score'  >= 5 and 'RFM Score' < 10) then "Medium" else "High" endif
hassubstring('Phone type','Coolphone')
```

In the first two examples above, the "RFM Score" expressions (`'RFM Score' > 5`, `'RFM Score'  >= 5`, and `'RFM Score' < 10`) can be thought of as boolean.

These expressions cannot be used by themselves to derive a field. An error such as `Invalid type for an expressions which is not a condition: Boolean` would be displayed. But expressions like these can be used in conditional expressions, as in the full examples, and conditional expressions can be sued to derive fields with categorical measurements such as a Flag type field. Flag is the default measurement for a field with two values.

The second example above could generate a field with three distinct values (as long as records which fall into each of the Low, Medium, and High categories are found during the data scan). This would likely result in the derived field's measurement level defaulting to Nominal. The third example above is unlikely to return a field with two distinct values.

See "Measurement levels" on page 41 for more information about measurement levels.

Chapter 5. Global selections

Global selections allow you to choose the records to be included or excluded from processing by the application. They provide an effective means to enforce corporate-wide policies, such as excluding customers with poor credit or payment history from all future campaigns.

Global selections are implemented using standard selection rules, which can be defined applied at multiple points in an application. See the topic "Defining selection rules" on page 53 for more information.

- You can toggle the selection between include and exclude by clicking the icon in the appropriate column for each rule.
- Because global selections are applied before other processing occurs, they may help to speed processing by reducing the number of records to be evaluated.
- An administrator or application designer may have already included one or more rules in your application that you cannot remove, even though you can add additional rules. For example, if your organization only targets people over 18 years old, a rule may be in place to enforce this policy.

Using shared rules

Global selections can be particularly effective when used in combination with shared rules, which are saved as separate objects that can be referenced (used) by multiple applications. If the shared rule changes, all applications that use the rule can then be updated. To search for rules by name or label, click the **Find an existing rule** icon.

Optionally, you can export the rules you create for use in other applications. See the topic "Exporting and reusing rules" on page 54 for more information.

Chapter 6. Working with rules

Rules are used to automate decisions based on business logic, or on the output from predictive models. Depending on the application, you might use rules to exclude customers based on payment history, to refer risky claims to a special investigations unit, or to target those with the highest propensity to churn, buy, or commit fraud as identified by a predictive model. While rules bring automation, models provide predictive power, and the ability to combine them is one of the key features that distinguish predictive applications from other analytical tools.

Rules can be created, applied, and saved at different points within each application. Alternatively, you can create and edit shared rules using IBM SPSS Rules Management.

IBM SPSS Rules Management

IBM SPSS Rules Management provides a central tool for creating and editing shared rules, which can be used throughout applications to select and process records, and to automate decisions accordingly. While rules can also be created and saved within each application, IBM SPSS Rules Management allows you to save rules as separate objects that can be referenced from multiple applications. For example, a rule to exclude customers who are below legal age could be shared across all applications to enforce a global policy, and any changes to the shared rule would be applied across all applications.

Note that IBM SPSS Rules Management may not be installed at all sites or available to all users. As an alternative, rules can also be defined and saved locally for use within each application.

Creating rules

- At the point in the application where you want to add a rule, click the **Create a new rule** icon.
- Alternatively, click the **Find an existing rule** icon to browse for existing rules which can be added or reused in the current application. See the topic "Exporting and reusing rules" on page 54 for more information.
- To create a shared rule, launch IBM SPSS Rules Management from the Application Launch page.

Defining segment rules

Segment rules are used throughout the application to select and allocate records in different applications as appropriate. Segments are defined by one or more expressions that return a true or false result. For example, you might have a rule with a single segment that returns true if the value of the age field is less than 18. Segments can be defined based on fields or expressions defined in the data, or on the propensity scores or predictions returned by a model. In applications with a Define tab that are configured to have multiple dimension hierarchies (referred to as *elements*), you can also choose from other dimension members. See the topic "Project configuration" on page 24 for more information.

Each segment is defined by one or more expressions, and returns a value of true if all conditions are met. For example, a segment defined by age and gender would only include records that met both conditions, and the rule would return *false* for all others.

To create or edit a segment rule

1. In the application where you want to add the rule, click the **Create a new rule** icon.
2. Specify a name for the rule.
3. For the input column, click the down-arrow icon to specify a field, rule, or model, as desired.

Table 6. Input column options.

Icon	Usage
	If enabled, allows you to select a dimension member from a previously defined element.
	Allows you to choose a field from the current data set.
	Allows you to choose an existing model or rule from the repository.
	Allows you to build a new model.
	Allows you to build an expression.

4. Choose the function you want to use. See the following table for available options.

Table 7. Available operators.

Value type	Available operators
Time stamp, Date, Date time	• = (equals) • != (not equal to) • > (greater than) • < (less than) • >= (greater than or equal to) • <= (less than or equal to) • BETWEEN • ISNIL • IS NOT NIL
A range	• = (equals) • != (not equal to) • > (greater than) • < (less than) • >= (greater than or equal to) • <= (less than or equal to) • BETWEEN • ISNIL • IS NOT NIL • ONE OF • NOT ONE OF
A set	• = (equals) • != (not equal to) • ISNIL • IS NOT NIL • ONE OF • NOT ONE OF

Table 7. Available operators (continued).

Value type	Available operators
A flag	• = (equals) • != (not equal to) • ISNIL • IS NOT NIL

5. Specify a value as appropriate for the input type. For non-numeric values, a dropdown list of choices is provided.

Alternatively, click the **Find an existing rule** icon to browse for rules that have already been defined. See the topic "Exporting and reusing rules" on page 54 for more information.

Note that a segment rule such as Age BETWEEN 50 and 20 will not return results. It is the same as the expression x <= value and value < y, which can never be true. This is consistent with SQL Server behavior. Age BETWEEN 20 and 50 is the correct format.

Also note that a segment rule such as Gender = $null$ will not return results either. If you want to filter out the fields for which the Gender value is empty, you can instead use an expression like Gender Not One Of (Male, Female).

Segment rule sets

- Segment rules can also be combined into sets. A segment ruleset returns a value of true if *any* of the individual segments is true. Note this differs from the expressions *within* a segment, all of which must be true.
- You can export segment rulesets from any application that uses rules, or you can do this in IBM SPSS Rules Management if available. See the topic "Exporting and reusing rules" on page 54 for more information.

Defining selection rules

A selection rule is a set of one or more segment rules or rule sets with an *Include* or *Exclude* attribute assigned to each. Selection rules are used throughout the applications to select subsets of records for modeling, or scoring, or other processing.

1. In the application where you want to add the rule, click the **Create a new rule** icon. This icon is available at points where selection rules can be applied.
2. Specify a name for the rule, select whether to use the rule for inclusion or exclusion (or do this later), and add one or more expressions as desired. See the topic "Defining segment rules" on page 51 for more information.
3. Click **OK** to save the rule.
4. Click the icons in the **Include/Exclude** column if you want to toggle this setting for each rule. For referenced selection rulesets, this setting cannot be changed. You cannot combine inclusions and exclusions in the same rule.

Alternatively, click the **Find an existing rule** icon to browse for rules that have already been defined. See the topic "Exporting and reusing rules" on page 54 for more information.

Evaluating selection rules

Click the **View selection rule logic** icon to see details about how selections are made (logical AND versus OR).

- When evaluating selection rules, if multiple include segments are specified, records matching *any* of the segments are included. For example, customers might be included based on credit rating OR tenure.

- If multiple exclude segments are specified, records matching any of the segments are excluded. For example, customers might be excluded based on age OR credit rating.
- In case of a conflict, exclude rules take precedence. For example, a customer that falls into any excluded segment will be excluded regardless of any segments they would otherwise qualify for.
- When creating selection rule sets in IBM SPSS Rules Management, all rules must be of the same type—either *Include* or *Exclude.* You cannot combine inclusion and exclusion rules in the same rule set.

Exporting and reusing rules

By default, rules created in an application are saved with that application. Alternatively, rules can be exported or saved to IBM SPSS Collaboration and Deployment Services Repository for use in multiple applications. In this case each application stores a reference to a specific, labeled version of the shared rule; any changes to the rule are automatically applied to all applications that use that version.

Exporting rules

From any application that contains a local rule you want to share, click the **Export** icon to save the rule to IBM SPSS Collaboration and Deployment Services Repository where it can be reused as needed.
- When exporting multiple rules, specify whether you want to export them all in a single ruleset, or individual rules. In the case of individual rules, each segment is exported as a separate rule.
- When exporting multiple selection rules, if you choose **Store with Selections**, a selection ruleset is exported. In this case you can export either inclusions or exclusions, but not both. If you need a mix of inclusions and exclusions, these need to be exported in separate rulesets. If **Store with Selections** is not selected, a segment ruleset is exported.
- When exporting from an aggregation rule editor, if you choose **Export with risk points**, an aggregation ruleset is created. If this option is not selected, a segment ruleset is exported.

Using existing rules

- In any application where you want to make use of a shared rule, click the **Find an existing rule** icon to browse IBM SPSS Collaboration and Deployment Services Repository for existing rules that can be reused in the current application.
- To reference a specific labeled version, click the arrow next to the rule name and choose the desired version. If no version is selected, the latest is used. (Think of *LATEST* as a special label that is automatically rolled to each new version as it is created.) Using a specific labeled version is recommended to avoid unexpected results that can occur when a new version of a rule is created.

  Referenced rules can be combined with local rules in many cases, but are visually distinguished from local rules by bold text, and by an icon that indicates the type of rule (or ruleset). Referenced rules cannot be edited directly in the application where they are used, but can be edited in IBM SPSS Rules Management, which can be access from the Application Launch page. Note IBM SPSS Rules Management may not be installed at all sites or available to all users.
- Alternatively, referenced rules can be converted to local rules and saved in the current application if desired. To do this, click the name of any referenced rule, and then select Convert to single rules in the Rule Preview dialog box.

Guidelines for sharing rules

When browsing and working with shared rules, the following guidelines apply:
- Rules can be exported for reuse from any application, but once exported, can be edited in IBM SPSS Rules Management, and cannot be edited in any other application. Note that IBM SPSS Rules Management may not be installed at all sites or available to all users.

- Using a specific labeled version (other than *LATEST*) is strongly recommended, particularly in a production environment, as ongoing changes the latest version may otherwise have unintended consequences. In certain cases, such as when referencing a rule from within a rule, a labeled version must be used.
- Single rules can be added from IBM SPSS Collaboration and Deployment Services Repository by reference, or converted to local rules from the Rule Preview dialog box.
- Single rules can be referenced and combined with local rules in any rule editor. If exported as a rule set, nested references to single rules will be preserved.
- Segment and selection rule sets can be referenced and combined like single rules in most cases, but the combined rule set cannot be exported without first converting any nested rule sets to single rules. Thus an exported rule set can contain references to nested rules, *but not to nested rule sets*.
- Aggregation rule sets can also be referenced, but cannot be combined with other rules, and cannot contain nested references to other rules. When referencing an aggregation rule set, the referenced rule replaces the entire contents of the aggregation rule editor. (Alternatively, if not referencing an aggregation rule set, the aggregation editor can contain a mix of local rules, referenced segment rules, and segment rule sets.)
- When using IBM SPSS Rules Management, the same restrictions apply as for exported rules. Nested rules are allowed, nested rule sets are not. This is because IBM SPSS Rules Management is designed to create and edit shared rules (similar to exporting from any of the other applications).
- To avoid any of the restrictions on referenced rule sets, whether in IBM SPSS Rules Management or elsewhere, a rule set can always be converted to separate, local rules.
- Shared rules cannot contain references to external IBM Operational Decision Management rules.

Adding annotations

In an aggregation or allocation, click the **Add Annotations** icon to add a column of free text fields into which you can enter a note that will be included with the rule output. For example, this may be extra information that tells a call center operator who to contact when an insurance claim is flagged for referral.

Annotations can be used in combination with the **Return with** value specified for a dimension, to provide additional detail. While the **Return with** value is associated with the dimension or decision that is returned, the annotation tells you which rules fired as input to that decision. See the topic "Defining campaigns and offers" on page 71 for more information.

Using rules from IBM Operational Decision Management

Rules created in a Business Rule Management System such as IBM Operational Decision Management be referenced and used in IBM Analytical Decision Management applications. These external rules can be selected and used in the same manner as other rules by clicking the **Find an existing rule** icon at the desired point in the application. External rules are distinguished by their icon, but otherwise usable in the standard manner.

For information on creating external rules for use in IBM Analytical Decision Management, see the *Application Designers Guide*.

The rule is inserted into the application in the same manner as any other rule. Note that rule outputs must match the context in which the rule is used; for example allocations require a true/false or yes/no (boolean) value as input to the decision.

As with other rules, click the name of any external rule used in the application to view details on the rule including input and fields, along with the rule description and other properties.

If an external rule has multiple output fields, a prompt is displayed allowing you to choose the desired field.

If an external rule is used in place of a model, no target field is shown, and all configured outputs are shown in the drop-down list.

Tips for using external rules

- Rules that return a value of *true* or *false* (or *yes/no* or *0/1*) can be used in any selection or allocation, as input to any rule in the rule editor, or as a **Return with** field associated with a specific dimension item.
- Rules that return categorical outputs (*red*, *white*, or *blue*; *north*, *south*, *east*, *west*, etc.) can be used in the **Use a model to decide** section on the Define tab, as an input in the rule editor, or as a **Return with** field associated with a dimension item.
- Rules that return numerical outputs can be used in the **Use a model to decide** section on the Define tab, as input to a prioritization, as an input in the rule editor, or as a **Return with** field associated with a dimension item.
- Analytic data view data sources support the use of complex (hierarchical) external rules. To reference a complex external rule, the project data model must be a *Data View* source and the functionality must be enabled (on the Data tab, an administrator must select the appropriate **Allow use of ODM rules based on hierarchical data** option).
- You cannot create a shared rule in IBM Analytical Decision Management that contains a reference to an external rule.

Chapter 7. Building predictive models

About predictive models

Models can be used to predict what is likely to happen in the future, based on patterns in past data. For example, you might use a model to predict which customers are least likely to churn, or most likely to respond to a particular offer, based on characteristics such as income, age, and the organizations and memberships they subscribe to.

Models can be used in the same way as rules, but while rules may be based on corporate policies, business logic, or other assumptions, models are built on actual observations of past results, and can discover patterns that may not otherwise be apparent. While rules bring common business logic to applications, models lend insight and predictive power.

Data for building and scoring models

Two different types of data are used in the modeling process:

- To build the model, you need information about the thing you want to predict. For example, if you want to predict churn, you need information about customers who have churned in the past. This is often referred to as historical or analytical data, and must contain some or all of the fields in the project data model, plus an additional field that records the outcome or result you want to predict. This extra field is used as the **target** for modeling.
- To use the model in predicting future results, you need data about the group or population you are interested in, such as potential customers or incoming claims, for example. This is often referred to as **operational** data, or **scoring** data. The project data model is typically based on this data.
- If the target field is included in the data model, the **Operational** column should not be selected for this field, because it is not available when scoring the model. Having built the model using historical data, the goal in scoring is to apply the model to new data where the outcome is not already known.

Building models

You can build a new model using any of the following methods:

- From the Define tab, click the **Build a model** icon under any panel where applicable. See the topic Chapter 8, "Defining decisions," on page 71 for more information.
- From the Prioritize or Optimize tab, click the **Build a new model** icon from any input field toolbar.
- When creating or editing a rule, click the **Build a new model** icon from the input field toolbar in the rule editor.

Figure 15. Build model icon

Building a predictive model

When building a predictive model, the only required specifications are a data source and target field. You can also choose options for preparing and partitioning data, and for selecting the fields and records to use. The model is generated by applying a range of methods as appropriate to the target type, and automatically identifying the specific technique or combination that performs best on the selected data.

When scored, the model returns one or more fields containing predictions, as well as confidence values associated with those predictions. For example, if the target field is a flag field named *Response*, the

model returns a field named *XF-Response* containing predicted values for that field. The predictions or scores returned by the model can then be used as input to rules, prioritization, and other points in the application as appropriate.

1. Specify a data source. This can be any analytical data source that records the outcome you want to predict.

2. Specify a target field that records the result you want to predict. For example, the target field might indicate which customers have churned, responded to past offers, raised fraudulent claims, and so on.

   All other fields will automatically be included as inputs that may be useful in predicting this value.

3. Specify optional settings as desired. See the topic "Optional model settings" on page 62 for more information.

4. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 40 for more information.

5. Click **Build Model**.

   You can close the browser or work on other tasks while the model builds. The model can be accessed from the Gallery once building is complete. See the topic "Gallery" on page 25 for more information. If the model is taking longer to build than expected, click the **Stop** button to revert to the previously-saved version.

   When model building is complete, results are displayed. You can choose to view the **Combined model results** (performance charts for the combined model that was currently built) or the **Individual model results** (high level details of the individual models that make up the combined model).

   When viewing *individual model results*, the name of each model is a link that opens a new dialog containing information about the model such as the target field and input fields. If the selected model is a *tree model*, then a tree viewer will also be displayed.

6. Optionally, use the Evaluate and Test features to see how the model performs on your sample data. See the topic "Evaluating models" on page 66 for more information.

7. Save the model before closing the model builder or returning to the application.

8. Click **Use Model**, and select the model field you want to use. For example if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

Predictive modeling results

When a predictive model completes, the results section allows you to switch between viewing **Combined model results** and **Individual model results**. The combined model results include distribution and variable importance charts for the combined model that was built, providing a quick visual assessment of model performance and of the importance of different predictors in obtaining the final result. The individual model results list information about all the individual models that make up the combined model, allowing you to click each individual model to view its details.

Distribution charts

The model distribution charts display observed values on the horizontal axis, with predicted values on the vertical axis or overlay. This allows you to see how often each value or response was predicted, and how often that prediction was correct, providing a quick visual assessment of model accuracy.

• For flag or set targets, the distribution is shown using a bar chart, with one bar for each category of the target field. The height of each bar indicates how many times that result is predicted by the model. The colored regions within each bar indicate how many times the matching result was actually observed in the testing data. If the model is reasonably accurate, then the largest area within each bar should match the corresponding response.

• For range targets, the distribution is represented by a binned scatterplot. Again, predicted values are plotted on the horizontal axis, and observed values are plotted on the vertical axis. Ideally the dots should cluster along a diagonal line. The more scattered the dots (results) and the farther from a linear distribution, the less accurate the model is.

Predictor importance charts

Typically, you will want to focus your modeling efforts on the predictor fields that matter most and consider dropping or ignoring those that matter least. The predictor importance chart helps you do this by indicating the relative importance of each predictor in estimating the model. Since the values are relative, the sum of the values for all predictors on the display is 1.0. Predictor importance does not relate to model accuracy. It just relates to the importance of each predictor in making a prediction, not whether or not the prediction is accurate.

Calculating predictor importance may take significantly longer than model building, particularly when using large datasets. It takes longer to calculate for SVM and logistic regression than for other models, and is disabled for these models by default. If using a dataset with a large number of predictors, initial screening using a Feature Selection node may give faster results (see below).

If the option to automatically partition data was selected when the model was built, then the testing partition will automatically be selected on the Evaluate page.

If partitioning was not turned on, you will need to select a data source for evaluating the model. Selecting a data source other than the one used to build the model is recommended. Note that partitioning is on by default, so this should not be necessary unless you have changed this setting.

Individual model results

Individual model results show information about each individual model that makes up the combined model. The name of each model is a link that opens a new dialog containing full details about that model, such as the target field and input fields. If the selected model is a *tree model*, then a special tree viewer will also be displayed. The tree viewer is interactive and provides the following functionality:

- Zoom in and out on the tree
- Show the tree items as tables, graphs, or tables and graphs combined
- Change the tree's orientation to top-down, left-right, or right-left
- Use the *tree map window* and the arrow icons to navigate large trees more easily

Building a predictive rule model

Predictive rule models allow you to identify segments that show a higher or lower likelihood relative to a specific outcome. For example, you might look for customers who are least likely to churn, or most likely to say yes to a particular offer. The resulting model resembles a list of selection rules, with an **Include** or **Exclude** attribute that specifies how each segment is scored.

Predictive rule models resemble selection rules, but differ in how they are scored. In the case of a selection rule, excluded records are removed, resulting in fewer records in the output. In the case of a model, no records are removed, but additional fields are added indicating the prediction and propensity values for each record. Included segments are scored as *True* (or *False*, or whatever outcome you are trying to predict). Excluded segments are scored as *NULL*, as are any records that fall to the remainder because they do not match any rule.

Each record is also given a propensity score that indicates the likelihood of the target response, making it possible to identify the customers who are most likely to respond. As such, predictive rule models are well-suited for generating mailing lists, and they are widely used in customer relationship management, including call center or marketing applications.

- Predictive rule modeling is available for categorical targets only.
- You can view, modify, and reorganize segments and to choose which to include or exclude for purposes of scoring. For example, you can choose to exclude one group of customers from future offers and include others and immediately see how this affects your overall hit rate.

- As with selection rules, you can manually add or edit segments as needed. It is often easier to let the model find them for you, however, and a number of options are provided to help you fine-tune the result.
- Rules are applied in the order listed, with the first matching rule determining the outcome for a given record. Taken independently, rules or conditions may overlap, but the order of rules resolves ambiguity. If no rule matches, the record is assigned to the remainder rule.

To obtain a predictive rule model

1. Specify a data source. This can be any analytical source that records the outcome you want to predict.
2. Specify a target field that records the result you want to predict.
3. Specify the specific value or response you want to find. For example, if the target field is named *response*, you could choose *True* or *False* to find customers who responded--or who didn't--accordingly.

   For example, the target might indicate which customers have churned, responded to past offers, raised fraudulent claims, etc. All other fields will automatically be included as inputs that may be useful in predicting this value.
4. Specify optional settings as desired. See the topic "Optional model settings" on page 62 for more information.
5. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 40 for more information.
6. Click **Start Build**.
7. Specify whether you want to find segments with high or low probability relative to the specified target. For example, you could look for groups or customers who are mostly likely to churn, or least likely, and include them accordingly.
8. Specify the maximum number of segments to be found. Typically this number is kept small in the interest of simplicity and performance.
9. Click **Grow Model** to obtain a list of one or more segment rules that identify groups with higher or lower probability.
10. Click the **Include**/**Exclude** icons to specify how the rules are to be used and how the relevant segments are scored.

    Included segments are given a score that matches the target value (for example *response=true*). Excluded segments are scored as *NULL,* but are not removed from the file.
11. Optionally, you can manually add or edit segments as desired. See the topic "Defining segment rules" on page 51 for more information.
12. Optionally, use the Evaluate and Test features to see how the model performs on your sample data.
13. Save the model before closing the model builder or returning to the application.
14. Click **Use Model**, and select the model field you want to use. For example, if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

Tips:
- Use the up/down arrows to move rules up or down. Each record is scored based on the first hit, so the order of rules will change the result. For example, a record that matches multiple rules will be scored based on the first that applies.
- Change one or more options and click **Grow Model** again to find additional rules. For example, you might lower the threshold for minimum segment or confidence interval in order to identify additional segments.
- Clicking **Grow Model** does not clear the table or rebuild the model from scratch, it only finds more segment rules to add to the table.

Building a clustering model

Clustering models allow you to categorize records into a certain number of clusters. This can help you identify natural groups in your data.

Clustering models focus on identifying groups of similar records and labeling the records according to the group to which they belong. This is done without the benefit of prior knowledge about the groups and their characteristics. In fact, you may not even know exactly how many groups to look for. This is what distinguishes clustering models from the other machine-learning techniques—there is no predefined output or target field for the model to predict. These models are often referred to as **unsupervised learning** models, since there is no external standard by which to judge the model's classification performance. There are no *right* or *wrong* answers for these models. Their value is determined by their ability to capture interesting groupings in the data and provide useful descriptions of those groupings.

Clustering methods are based on measuring distances between records and between clusters. Records are assigned to clusters in a way that tends to minimize the distance between records belonging to the same cluster.

Clustering models are often used to create clusters or segments that are then used as inputs in subsequent analyses. A common example of this is the market segments used by marketers to partition their overall market into homogeneous subgroups. Each segment has special characteristics that affect the success of marketing efforts targeted toward it. If you are using data mining to optimize your marketing strategy, you can usually improve your model significantly by identifying the appropriate segments and using that segment information in your predictive models.

To obtain a clustering model

1. Specify a data source. This can be any analytical source that records the outcome you want to predict.
2. Specify optional settings as desired. See the topic "Optional model settings" on page 62 for more information.
3. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 40 for more information.
4. Click **Find Clusters**.
5. Optionally, you can add manual clusters. See the topic "Using manual clusters" for more information.
6. Optionally, use the Evaluate and Test features to see how the model performs on your sample data.
7. Save the model before closing the model builder or returning to the application.
8. Click **Use Model**, and select the model field you want to use. For example, if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

Using manual clusters

If desired, you can manually define clusters, then use automated clustering to find clusters of records not identified by any of the manually-defined clusters. Create manual clusters via a "first hit" segment-based allocation rule set. The rules you create will define clusters and the remainder will be used for auto cluster modeling.

Note that manual clusters are not included in results summary charts.

1. In the **Manual Cluster** section, click the **Create a new rule icon** and add one or more rules to define manual clusters.
2. If desired, change cluster label names.
3. Click the **Records in manual clusters** icon to perform a record count of the total number of records included in the manual clusters. The **Count** column will be updated for each manual cluster defined.

Building an association model

Association models allow you to predict which items are most likely to appear together, and predict the strength of the relationship between them.

Association rules associate a particular conclusion (the purchase of a particular product, for example) with a set of conditions (the purchase of several other products, for example). For example, the rule

```
beer <= cannedveg & frozenmeal (173, 17.0%, 0.84)
```

states that *beer* often occurs when *cannedveg* and *frozenmeal* occur together. The rule is 84% reliable and applies to 17% of the data, or 173 records. Association rule algorithms automatically find the associations that you could find manually using visualization techniques, such as the Web node.

The advantage of association rule algorithms over the more standard decision tree algorithms (C5.0 and C&R Trees) is that associations can exist between *any* of the attributes. A decision tree algorithm will build rules with only a single conclusion, whereas association algorithms attempt to find many rules, each of which may have a different conclusion.

The disadvantage of association algorithms is that they are trying to find patterns within a potentially very large search space and, hence, can require much more time to run than a decision tree algorithm. The algorithms use a **generate and test** method for finding rules--simple rules are generated initially, and these are validated against the dataset. The good rules are stored and all rules, subject to various constraints, are then specialized. **Specialization** is the process of adding conditions to a rule. These new rules are then validated against the data, and the process iteratively stores the best or most interesting rules found. The user usually supplies some limit to the possible number of antecedents to allow in a rule, and various techniques based on information theory or efficient indexing schemes are used to reduce the potentially large search space.

To obtain an association model

1. Select a data source. This can be any analytical source that records the outcome you want to predict.
2. Specify optional settings as desired. See the topic "Optional model settings" for more information.
3. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 40 for more information.
4. Click **Build Model**.
5. Optionally, use the Test features to see how the model performs on your sample data. Note that Evaluation is not available for the Association model type.
6. Save the model before closing the model builder or returning to the application.
7. Click **Use Model**, and select the model field you want to use. For example, if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

With other model types, you can choose between different modeling techniques. For association models, the *Apriori* model type is used.

Optional model settings

When building models, you can choose from the following optional settings. Note that depending on the type of model and the target selected, you may not see all of these settings. If changing any of these settings for an existing model, the model must be rebuilt for changes to apply.

Automatically clean up and prepare data for reliable model building. Identifies and repairs data issues to make modeling faster, more predictable, and more reliable. Screens fields that are problematic or not likely to be useful, for example by handling missing and extreme values, deriving new attributes when appropriate, and improving performance through intelligent screening and sampling techniques. The first time a model is built with a new data source, the analysis is done to identify issues and fixes, which may slow performance on the first pass only. On subsequent runs, the fixes are applied, but the analysis is not

repeated unless the data source changes. This setting may be disabled for some models, including those with custom data preparation settings specified by an expert user.

Note: Automatic data preparation settings only apply when creating a new model, and binning is not performed.

Automatically partition data to enable model evaluation on build data source for evaluation and testing. Selecting this option splits the data into separate subsets or samples for training and testing the model. By building the model on one subset and testing it on another, you can get an idea how it will generalize to other data sets. You can also specify a percentage value for the randomly generated *training partition size* and *testing partition size*.

The minimum training partition size is 1 and the maximum 100. The minimum testing partition size is 0 and the maximum is 100. The minimum validation partition size is 0 and the maximum is 100. The total percentage must be 100 or less for the model to build successfully.

These partitioning options are available for all model types except association modeling.

Choose model techniques for model building. If desired, click **Select** to choose which modeling techniques to include when building the model. Based on the target you selected, all available modeling techniques will be listed. The modeling techniques available (and those selected by default) will vary depending on the target you selected. For example, the *Decision List* algorithm will only appear for *flag* targets. This feature is only available for predictive models.

For complete details about the modeling techniques, see the *IBM SPSS Modeler Algorithms Guide* and other documentation shipped with the IBM SPSS Modeler product and available on the Web.

Maximum # of models to be combined. Allows you to set the maximum number of models to be retained and combined. If you set this option to 1, a single model will be built and retained. If there are less models built than the value entered here, then all models built will be retained and combined. This option is only available for predictive models.

Profit Criteria (used to rank models with binary targets). Allows you to set profit criteria prior to building the model. The values set here will be set as the default for evaluating the model. This option is only available for predictive models with a flag target selected.

Specify inputs to use. Allows you to select the fields you want to use. Typically these would be those that have some practical relationship to the thing you are trying to predict, such as age or income. If you have a large data set, limiting the number of fields is one way of simplifying the model. If the data contains fields such as customer id or contact number, these would not typically be useful in modeling and should not be selected. Fields that duplicate other data may also be excluded.

Clicking on a linked input field (an expression) opens the expression viewer for that expression. To edit an expression, see the Data tab. See the topic "Expression Builder" on page 44 for more information.

Specify selections to use. Specifies which records to include or exclude when modeling. You can search for existing rules, or create new ones as appropriate. See the topic "Defining selection rules" on page 53 for more information.

In addition, if global selections have been defined, they will be displayed here, and you can specify whether they should also apply during modeling. You can either choose to use all or none of the global selection rules; you cannot choose a subset.

Segmentation options
When building predictive rule models, a number of segmentation options are available. These are in addition to the options available with other types of models.

Minimum segment size. These settings dictate the minimum segment size, with the larger of the two values taking precedence. For example, if the percentage value equates to a number higher than the absolute value, the percentage setting takes precedence.

- **Percentage of previous segment.** Specifies the minimum group size as a percentage of records. The minimum allowed setting is 0; the maximum allowed setting is 99.9.
- **Absolute value.** Specifies the minimum group size as an absolute number of records. The minimum allowed setting is 1; there is no maximum setting.

Maximum number of attributes. Specifies the maximum number of conditions per segment rule. The minimum allowed setting is 1; there is no maximum setting.

Allow attribute re-use. When enabled, each cycle can consider all attributes, even those that have been used in previous cycles. The conditions for a segment are built up in cycles, where each cycle adds a new condition. The number of cycles is defined using the **Maximum number of attributes** setting.

Confidence interval for new conditions. Specifies the confidence level for testing segment significance. This setting plays a significant role in the number of segments (if any) that are returned as well as the number-of-conditions-per-segment rule. The higher the value, the smaller the returned result set. The minimum allowed setting is 50; the maximum allowed setting is 99.9.

Auto cluster options

When building clustering models, a number of auto cluster options are available. These are in addition to the options available with other types of models.

Evaluation field. Specifies the field to use for evaluation. If an evaluation field is selected, the ranking method in the IBM SPSS Modeler auto cluster build node will be set to *Importance of evaluation*. If no evaluation field is selected, the ranking method will be left as the default. Note that an analytical field can be selected as the evaluation field here.

Set a desired range for the number of clusters found. Specifies the desired minimum and maximum range for the number of clusters found by the auto cluster model.

Association rule and scoring options

When building association models, a number of association rule and scoring options are available. These are in addition to some of the options available with other types of models.

Association Rule Options.

- **Coverage (%).** Specify the minimum coverage percentage to control how generally applicable the rules will be. For example, setting a value of 80% means that all rules will apply to at least 80% of cases. Note that higher values will yield fewer rules, or potentially none at all.
- **Confidence (%).** Specify the minimum confidence percentage.
- **Maximum number of preconditions.** Specify the maximum number of preconditions.
- **Data are already sorted by ID (transactional format only).** Select this option to disable the sort that takes place by default at the start of model building to ensure IDs are contiguous. Selecting this option will reduce processing time. This option has no impact if the data format is tabular.

Scoring Options.

- **Maximum number of predictions.** Specify the maximum number of predictions included. This setting is used in conjunction with the rule criterion drop-down to produce the "top" predictions, where top indicates the highest level of confidence, support, lift, and so on.
- **Rule criterion.** Select the measure used to determine the strength of rules. Rules are sorted by the strength of criteria selected here in order to return the top predictions for an item set. Available criteria are:

– **Confidence** displays the ratio of rule support to rule condition support. This indicates the proportion of records with the specified rule condistion(s) for which the rule prediction(s) is/are also true. For example, if 50% of the training data contains bread (indicating rule condition support) but only 20% contains both bread and cheese (indicating rule support), then the prediction for the rule bread -> cheese would be `Rule Support / Rule Condition Support` or, in this case, 40%.

– **Coverage (%)** displays rule condition support--that is, the proportion of IDs for which the rule conditions are true, based on the training data. For example, if 50% of the training data includes the purchase of bread, then the `rule bread -> cheese` will have a rule condition support of 50%.

– **Rule Support (%)** displays the proportion of records for which the entire rule, rule condition(s), and rule prediction(s), are true. For example, if 20% of the training data contains both bread and cheese, then rule support for the rule `bread -> cheese` is 20%.

– **Lift** displays the ratio of confidence for the rule to the prior probability of having the rule prediction. For example, if 10% of the entire population buys bread, then a rule that predicts whether people will buy bread with 20% confidence will have a lift of `20/10 = 2`. If another rule tells you that people will buy bread with 11% confidence, then the rule has a lift of close to 1, meaning that having the rule condition(s) does not make a lot of difference in the probability of having the rule prediction. In general, rules with lift different from 1 will be more interesting than rules with lift close to 1.

– **Deployability (%)** is a measure of what percentage of the training data satisfies the rule condition(s) but does not satisfy the rule prediction. In product purchase terms, it basically means what percentage of the total customer base owns (or has purchased) the condition(s) but has not yet purchased the rule prediction(s). The deployability statistic is defined as `((Rule Condition Support in # of Records - Rule Support in # of Records) / Number of Records) * 100`, where `Rule Condition Support` means the number of records for which the rule conditions are true and `Rule Support` means the number of records for which both rule conditions and the prediction are true.

- **Check entire transaction.** Select from the following options:

  – **Ensure predictions not present** ensures that rule predictions are not also present. For example, if the purpose of scoring is to make a home furniture product recommendation, then it is unlikely that a basket that already contains a dining room table will be likely to purchase another one. In such a case, you should select this option. On the other hand, if products are perishable or disposable (such as cheese, baby formula, or tissue), then rules where the consequent is already present in the basket may be of value. In the latter case, the most useful option might be **Do not check for predictions**.

  – **Ensure predictions are present** ensures that rule predictions are also present in the basket. This approach is useful when you are attempting to gain insight into existing customers or transactions. For example, you may want to identify rules with the highest lift and then explore which customers fit these rules.

  – **Do not check for predictions** includes all rules when scoring, regardless of the presence or absence of rule predictions in the basket.

- **Allow repeat predictions.** Select this option to include multiple rules with the same rule prediction when scoring. For example, selecting this option would allow the following rules to be scored:

  ```
  bread & cheese -> wine
  cheese & fruit -> wine
  ```

  Turn off this option to exclude repeat predictions when scoring.

- **Ignore unmatched items.** Select this option to ignore the presence of additional items in the item set. For example, when this option is selected for a basket that contains [`tent & sleeping bag & kettle`], the rule `tent & sleeping bag -> gas stove` will apply despite the extra item (`kettle`) present in the basket.

Evaluating models

After building or opening a completed model, click **Evaluate** to obtain charts such as distribution, error distribution, coincidence matrix, gains, lift, response, profit, and ROI, and to simulate the maximum profit possible. The chart types and other options available will vary depending which model type was selected. The Evaluate dialog is only available if the selected model is complete. Note that evaluation is not available for the Association model type.

1. In the upper panel, select an evaluation data set. As with model building, this must be an analytical data set for which actual responses are known. (Otherwise you have no way to compare actual responses to those predicted by the model.)

2. If you want to modify the models to include in the evaluation, click **Modify Model Selection**. After making selections, you can also click the following **Copy current model selections back to individual model results table** button, if desired. These options are only available for the Predictive model type.

Figure 16. Copy current model selections back to individual model results table

3. Select the graphs you want to build.

4. If simulating profit, select the **Simulate maximum profit** box and enter the population on which to base the profit calculation, the budget, and the minimum value. See the topic "Simulating maximum profit" on page 67 for more information.

5. If requesting a profit or ROI chart, or if simulating profit, specify cost and revenue values. These values indicate the cost for each offer, and the amount of revenue you expect to make from each favorable response, respectively.

6. Select the response value that is considered as a "hit." This must be one of the values for the target field used to build the model, such as *response = true* to indicate a positive or favorable response.

7. If requesting a performance chart, you can also select the **Include individual model evaluations for performance charts** option.

8. If desired, select **Include tabular results**. This requires more time for the evaluation results to be returned.

9. Click **Run** to view the results.

Distribution charts (all targets)

Distribution charts plot observed versus expected values, and are displayed by default when the model completes. The Evaluate dialog box allows you to obtain distribution charts for a data set other than the testing data set. For range targets, the distribution chart is the only chart type available.

You can choose from the following additional charts:

Error Distribution charts

The error distribution charts displays a histogram of the errors in the model outcomes, where the error is the difference between the predicted value and the actual value. A positive error arises when the predicted value is greater than the actual value, and a negative error arises when it is less. This chart is only available for continuous targets with the Predictive model type.

Coincidence matrix charts

A coincidence matrix plots the counts of *predicted* versus *actual* outcomes. For a numeric target, the range is divided into the number of bins you specified.

Gains charts

Gains are used to show the gain or "lift" that can be obtained by using the model, and are defined as the proportion of hits in each increment relative to the total number of available hits. The flat diagonal line plots the expected response for the entire sample, if no model were used. In this case, the response rate would be constant, since one person is just as likely to respond as another. To double your yield, you would need to ask twice as many people. The curved line indicates how much you can improve this result by targeting those who are most likely to respond. The steeper the curve, the higher the gain.

Lift charts

The lift chart plots the percentage of records in each increment that are hits with the overall percentage of hits in the training data set.

Response charts

The response chart plots the percentage of records in the increment that are hits.

Profit charts

Profit equals the revenue for each record minus the cost for the record. Profits for a quantile are simply the sum of profits for all records in the quantile. Revenues are assumed to apply only to hits, but costs apply to all records. See the topic "Simulating maximum profit" for more information.

ROI charts

ROI (return on investment) is similar to profit in that it involves defining revenues and costs. ROI compares profits to costs for each interval.

Simulating maximum profit

Based on the response rates predicted by the model, you can simulate the maximum profit to be made for a given population. To do this you need to specify cost and revenue values, as well as the size of the population.

To predict the maximum profit:

1. Specify the population, or total number of records, on which the prediction should be based. *Note*: When scoring a model, this is automatically set to the total number of records in the data source or subset.
2. Specify a value for budget if you want the budget parameter to be included in the profit simulation, and a minimum value.
3. Specify the cost associated with each record. For example, this could be the print and mailing cost, per person, of promotional literature for a campaign.
4. Specify the expected revenue that a positive response would generate. For example, the retail cost of an item in a campaign.
5. Select the response you are interested in.
6. Click **Run**.

The predicted results graph shows the best result for maximum profit, and identifies the percentage of records you would need to target to achieve this, based on propensity to respond as identified by the model.

Beneath the graph you can move a slider bar from zero to 100% of the records to target to see how this would affect your profit.

Profit comparison

To verify that the chosen records are a good selection, you can compare the likely profit to be made from those records, a random group of the same percentage of records, and the whole set of records.

Details shown include the number of records targeted, the likely response rate, costs, income, and overall profit. You can export this information in HTML format for use in other applications.

Profit chart

You can display the predicted cumulative or non-cumulative profit from the selected records as a graph.

Note: This graph is generated when you first run the evaluation and is not automatically updated with any changes you make via the slider in the profit simulation section.

Using model scores in applications

To score a model means to apply it to some data or population in order to obtain a result or prediction that can be used as input to decisions. For example, you might score all customers in your database to identify prospects for a direct mail campaign, or score customers one at time as they contact your call center in order to decide which products or services to offer them. Depending on the application, scoring results can be written to a database table or flat file, or used as inputs to the segment, selection, and allocation rules that drive decisions in an application.

Model scores can be used in the following ways:
* As input to rules. See the topic "Creating rules" on page 51 for more information.
* To allocate decisions on the Define tab.
* As inputs to the prioritization.

Model output fields

By convention, the names of the scoring output fields are based on the name of the target field, but with a custom prefix added. For example if the target field is named *Response*, the output fields will have names like *$XF-Response* and *$XFRP-Response*. The specific fields depend on the type of model, and the measurement level for the selected target field, as detailed below. See the topic "Measurement levels" on page 41 for more information.

Flag targets. Automated models with flag targets output the following fields:
* **$XF-<target>.** Predicted value for each record, reported in terms of the "true" and "false" values defined for the field.
* **$XFRP-<target>.** Propensity score; indicates the likelihood of the "true" value for each record. Propensity scores differ from confidence scores, which are typically reported in terms of the prediction for that record. For example a "false" prediction with a high confidence will translate to a low propensity as this would indicate a high likelihood *not* to respond.

Categorical targets. Automated models with categorical targets (nominal or ordinal) output the following fields:
* **$XS-<target>.** Predicted value for each record.
* **$XSC-<target>.** Confidence value associated with the prediction

Continuous targets. Automated models with continuous numeric targets output the following fields:
* **$XR-<target>.** Predicted value for each record.
* **$XRE-<target>.** Standard error for the prediction.

Predictive rule (interactive) models. Predictive rule models (flag or categorical targets only) output the following fields:

- **$D-<target>.** Predicted outcome for each record, with the "true" value defined for the field used to indicate the desired response, and null indicating any other value.
- **$DP-<target>.** Probability of desired outcome for each record; matches the probability displayed for the segment in the results table when building the model. See the topic "Building a predictive rule model" on page 59 for more information.
- **$DI-<target>.** Index number used to identify the segment into which the record falls (displayed in the left-most column in the modeling results).
- **$DRP-<target>.** Propensity score indicating the likelihood of the desired response. Available for flag targets only.

Chapter 8. Defining decisions

The Define tab allows you to specify the range of possible decisions or recommendations that can be returned by the application, as well as the models and rules that determine how they are allocated.

The rules need not be mutually-exclusive; the same customer may be allocated to multiple campaigns or offers, or none, as appropriate. The final determination is made in the prioritization/optimization step.

- The dimension tree at left specifies the range of possible decisions or recommendations that can be returned.
- The settings on the right determine how records are selected and allocated to specific dimensions using models and rules. The rules need not be mutually-exclusive; the same record may be allocated to multiple items, or none, as appropriate.
- In cases where multiple recommendations are returned on the Define tab, a prioritization equation can be used to choose which items are actually allocated.
- You can also set up dynamic allocation to items not listed in the dimension tree. This can only be enabled on dimension members that have no child dimensions. See the topic "Dynamic allocation" on page 73 for more information.
- The output from the rules is returned as either a single decision or multiple recommendations.
- There is limitation of one child dimension for a given parent.

Defining campaigns and offers

The tree view at left defines the campaigns and offers available in the application. These items define the possible decisions, or results, that can be returned. Note that some applications may be configured to include more than one dimension hierarchy tree, displayed on separate tabs. The right pane displays properties, selections, and allocations for the selected item.

If the tree is too large to be visible, you can drag the right border to resize the panel.

- Click the *Add a new* icon at the top of the tree to add dimension items.
- Click the *Expand all* or *Collapse all* icons to show or hide items.
- Right-click individual items in the tree to display a context menu that allows you to perform various tasks such as adding, renaming, moving, and duplicating items:
  - To move an item, select **Move up** or **Move down** from the context menu, or select **Move...** and select the new location from the drop-down list.
  - To reuse an existing item, select **Add / Modify** from the context menu, click **Add an existing**, and choose the item you want to reuse.
  - To specify one or more field, text, or model outputs to be returned with an item, select **Return with** from the context menu. Return fields are extremely useful in tailoring recommendations to specific cases, or providing additional information that may be useful in processing the result. Along with a loan offer, for example, an application might return a model score that determines the recommended interest rate, or a bit of text with additional details about the loan. The dialog allows you to set the type of data the return field contains (string, integer, or real, for example), change the name of return fields, set a default value to override a dimension's default, specify a member value, or remove return fields. Note that the default value field is always enabled (the field can be used for any member that does not have a member-specific value).
  - To dynamically allocate items from a table to a dimension member, select **Enable dynamic allocation** from the context menu. This option is only available for members with no children. A special icon is displayed beside any members in the tree that have dynamic allocation enabled.

If the application includes more than one dimension hierarchy element, they will be included on separate tabs. The tab order you specified when you opened the application is important. Hierarchy members of previous tabs can be referred to in rules within subsequent tabs, but references cannot be made to subsequent tabs. For example, rules on a second and third tab can refer to hierarchy members in the first tab, but rules on the first tab cannot refer to members on the second or third tabs (and rules on the second tab cannot refer to members on the third tab).

Importing and exporting dimension tree items

You can export dimension tree information to an external file, and import for reuse as needed. For example, this would allow you to restore the items used in a previous version of the tree. Dimensions are imported from the top-most level down. For example you can import campaigns, including any child offers. But you cannot import only offers within an existing campaign.

* To export the current tree, click the *Export* icon displayed just above the tree, then specify a folder and filename.
* To import dimension items, click the *Import* icon, and specify an XML file containing the items you want to import. This should be a file exported from a compatible application that uses dimension types matching the current project (such as customers and offers, claims and actions, etc.). Attempting to manually edit the XML source is not recommended.

Imported dimensions are added to the current tree. If the file contains more dimension levels than the current project allows, you can choose to import only the levels that are allowed. Dimensions that cannot be imported are listed in gray to indicate this.

Name restrictions for dimension tree items

The following three characters cannot be used in dimension member names: > < &

Campaign properties

Active. Specify start and end dates to indicate when the item is valid.

Interaction points. Specifies where the item applies. The available options are configured for each application, and may include call center, web site, ATM, in-store locations, etc. Interaction point settings also constrain child dimensions, for example an offer cannot be valid longer than the campaign to which it belongs.

Note: Interaction points must be configured by the administrator before they will appear in this section. See the topic "Defining interaction points" on page 100 for more information.

Campaign selections

Selections specify which records a given set of recommendations can be applied to. They are similar to global selections, but applied to specific dimensions.

You can choose from existing rules, or create your own. Optionally, any new rules can be exported to the IBM SPSS Collaboration and Deployment Services Repository for reuse in other applications. See the topic "Defining selection rules" on page 53 for more information.

Defining allocations

Recommendations can be allocated using the following methods:

* Based on segment rules or random percentages (which can use model scores as inputs)
* Dynamically allocated to items not in the dimension tree (group dimension items independent of the tree by referencing fields)

Allocation using segments or random percentages

Recommendations can be allocated using rules, or based on random percentages. For example you might allocate customers to different offers based on attributes such as gender, or propensity to respond as determined by a predictive model.

Allocate using segment rules

Each row in the table displays a rule that returns true for records that match, or false otherwise.

- Select **First valid** to allocate records to the first valid option based on the order listed, or select **All valid** to use all valid allocations. For example, if a customer is eligible for one offer based on their credit rating, and another based on tenure (loyalty), they may be allocated to one or both as determined by this setting. However, note that if you select **All valid** and a remainder offer is returned, the remainder offer will not be allocated unless no other rules match.

- Click the rule name to edit the rule, or to see a preview in the case of a referenced rule.

- Specify how records are allocated to dimension items (such as offers) for each segment. You can choose from the items that have been defined, select **New**, or select **None** to indicate that no allocation should be made.

- Specify how records that fall under the remainder are allocated (those that don't match any of the preceding segments). These can be allocated to a specific offer or dimension if desired. This will always be displayed as the last row in the table.

- Optionally, you can use the arrows to change the order in which rules are displayed. When **First valid** is selected, this also determines the order in which rules are evaluated. For example, a record that matches the first segment will not be considered for subsequent segments.

Multiple allocation

The **Multiple allocation** option functions in the same basic way as the **Allocate using rules** option, except you can allocate records to multiple dimension items (such as multiple offers) for each segment. Multiple allocations can be made for each rule instead of a single allocation per rule.

Allocate randomly

Allocates records randomly to offers or other dimension items. You specify the overall probability for each item and records are chosen at random to meet the specified percentages.

- Select **Weighted probability** to specify a different probability for each allocation. For example you could allocate 60% of records to one offer, 30% to another, and the remaining 10% to no offer. You can specify any number of probabilities as long as they don't add up to more than 100%.

- Select **Equal probability** to allocate records evenly across specified items. Again, you can specify as many items as you want to divide into halves, thirds, quartiles, etc.

- Select **Add / Remove** to choose the items to allocate. There will be a row added to the table for each selected item.

- Optionally, you can use the arrows to change the order in which rules are displayed. For random allocations, the order has no effect on rule execution.

- Click the **Add annotations** icon to add a new Annotations column. You can enter free text into the fields in this column that relates to each allocation.

Dynamic allocation

In early versions of the product, to allocate something, it had to exist in the dimension tree. For dimension items with no children, you can now group dimension items independent of the tree. For example, you might have a dimension item called *Cars* with no children. You could set up dynamic allocations on *Cars* to reference fields in a table such as *BMW*, *Audi*, and *Ford*.

1. Right-click an item in the tree that has no children and select **Enable dynamic allocation**.

2. In the Dynamic Allocation section, type a value into the text box or use the input toolbar to define the allocation.

Chapter 9. Optimizing decisions

While the selections and allocations on the Define tab may return multiple possible decisions or recommendations for each record, the Prioritize or Optimize tab specifies how the final decision is made.

Optimization

The selections and allocations on the Define tab determine which customers are candidates for which offers , and may return many possible solutions. If a customer is eligible for multiple offers , which do you choose? Optimization allows you to identify the best or most profitable offer for each individual via the right channel within the constraints of budget and channel capacity. In other words, IBM Analytical Decision Management optimization defines the best way that entities can be allocated to dimensions.

The goal of optimization is to identify the solution that best meets a specific goal, such as maximizing the revenue from a marketing campaign, or minimizing the risk of fraud or churn. The result of optimization is a solution that answers this question across all possible combinations within your population, in the manner that best maximizes the value you want to optimize (such as profit).

Optimization supports a number of deployment scenarios. For example:

- **Real-time optimization.** Optimizing allocations for a customer on a real-time interaction point (for example, best offers to make to a customer who is looking at the web site or who has called in to the call center).
- **Batch optimization.** Optimizing allocations across multiple customers on a batch interaction point (for example, best offers to make to each customer by the best outbound channel).
- **Pre-scoring real-time offers ahead of time on a batch interaction point.** For example, rather than performing a real-time optimization when the customer visits the web site, perhaps you would rather pre-calculate what offers should be made to the customer if they visit the web site, or if they call the call center, etc. If the customer does not arrive, then those precomputed offers could expire and can be recomputed on more recent data.

Optimization is helpful for solving problems such as allocating customers to marketing campaigns or offers, for example.

Applications configured to use complex mathematical optimization (CPLEX), such as IBM Analytical Decision Management for Campaign Optimization, IBM Analytical Decision Management for Operations, and IBM Analytical Decision Management for Demand Optimization, have an **Optimize** tab.

Note that CPLEX optimization scales in a non-linear fashion. For example, the increase in time and resources required to optimize over 1000 records will more than double when 2000 records are used – possibly taking four times as long or more. This is because optimization is trying to find combinations of records, and the number of available combinations increases non-linearly.

1. Specify optimization parameters in the table as desired. You can simply type in values, or click the icon next to each input to choose a field, browse for an existing model, build a new model, or create an expression which can then be used as input to the optimization. If the application is configured to include more than one decision hierarchy element, each hierarchy will have its own tab containing a parameters table.

   You can also reorder the parameters in the table, but this has no impact on the optimization.

   The following table describes some of the available items in the Optimization Parameters section:

Table 8. Optimization parameters

Item	Description
Total Budget	Specify the total budget available across all campaigns.
Max. Offers	Specify the maximum number of offers that can be made to a customer.
Recent Offers	Specify a value for the number of recent offers a customer has already received.
Campaign/ Offer	Lists the available campaigns and offers, such as monthly specials and TV discount.
Prob. to Respond	The probability that the customer will respond to an offer.
Revenue	The expected revenue if the customer accepts the offer.
Offer Cost	Cost of making the offer.
Offers Available	The number of this offer available to be made.
Channel	Lists the available channels (mediums used to communicate with the customer), such as e-mail.
Capacity	The maximum number of offers that can be made through this channel, such as the maximum number of offers sent via mail.
Channel Cost	Specifies the cost associated with making an offer through the channel.

2. Use the **Priority** drop-down menu to give priority to certain items. For example, offers given high priority will be considered before those that are normal or low priority. The prioritization equation is still computed for all offers, but the selection is done in different groups based on priority order. In other words, all solutions of a given priority are considered before any solution of a lower priority. The priority will take precedence over the object value. For example, if *Offer A* has a high priority and *Offer B* has a lower priority but higher expected profit, then *Offer A* will be selected.
3. Select **Customize table** to choose at which level you wish to define each parameter. For example, whether to specify parameters at the campaign level, or separately for each offer within a campaign .
4. The **Optimize** tab **maximizes revenue** or **maximizes return on investment**, based on the optimization equation used. Click **Select equation** and choose which optimization equation defines your goal.

   ```
   Total projected profit based on current project settings = (Prob. to Respond * Revenue) - (Offer Cost + Channel Cost)
   
   OR
   
   Return on investment =  ( (Prob. to Respond * Revenue) - (Offer Cost + Channel Cost) ) / (Offer Cost + Channel Cost)
   ```

   Where:
   - `Prob. to Respond` is a customer's propensity to respond to a particular offer.
   - `Revenue` is the income value or contribution expected from that response.
   - `Offer Cost + Channel Cost` is the cost of making the offer and the cost of using channels such as e-mail, mail, or telephone.

   Note: IBM Analytical Decision Management for Campaign Optimization projects created in previous releases will use a different equation. The equation was updated for version 17.

   Equations can be modified for different applications by the application designer. While the specific equation may change, many business problems may be modeled in this manner.
5. In the **Constraints** section you can select from available constraints to define the boundaries the optimization solution must work within, such as total budget. For example, you likely do not want the amount spent to exceed the total budget.

   The application designer configures the optimization equations and the constraints. For more information, see the *Application Designer's Guide*.

6. To experiment with different parameter combinations and see how each impacts your result, click the **Create a new scenario based on current project settings** icon.

7. If desired, click the **Perform a test on the application for a selection of records** icon. Note that optimization global constrains are applied at the individual customer level (not globally), ROI is calculated at the individual customer level (not based on all customers), and if a record is excluded by the optimization, then its objective value is always shown as zero.

For applications that use complex mathematical optimization (CPLEX), such as IBM Analytical Decision Management for Campaign Optimization, IBM Analytical Decision Management for Operations, and IBM Analytical Decision Management for Demand Optimization, advanced options are available to administrators. See "Advanced options" for more information.

Note that for applications with multiple dimension hierarchy elements, when users open the application they can choose which elements to include and reorder them. If an element has been excluded, any objective functions or constraints will also be excluded. See the topic "Project configuration" on page 24 for more information.

Advanced options

For applications that use complex mathematical optimization (CPLEX), such as IBM Analytical Decision Management for Campaign Optimization, IBM Analytical Decision Management for Operations, and IBM Analytical Decision Management for Demand Optimization, advanced options are available related to the CPLEX optimization algorithm used in the application. Using IBM ILOG CPLEX Optimization Studio, advanced users may develop additional applications that use this capability. IBM ILOG CPLEX Optimization Studio is available for separate purchase. Contact your sales representative for more information.

1. On the Optimize tab, in the **Optimization Model** section, click **Advanced Options**. This button is only available to users with administrative authority. The Advanced Options dialog will be displayed.

2. If you want to export an Optimization Programming Language (OPL) model file for the current project, select **Export model to file**, browse to the desired location, select a file format, and specify a file name. You must select either **CPLEX LP format files** (*.lp*) or **MPS standard files** (*.mps*) for the file format.

3. To add custom settings, select **Use custom settings** and then click **Add new setting**. Type a name for each custom setting you add, select a type, and specify a value.

4. When finished, click **Save**.

Administrators may want to export CPLEX *.lp* or *.mps* files for troubleshooting purposes or to control certain CPLEX parameters.

Advanced users can tune the CPLEX parameters for IBM Analytical Decision Management to attain better performance on large data sets. We also recommend minimizing the number of possible allocations presented to CPLEX (the data volume). In other words, you should reduce the number of combinations (for example, the number of alternative offers allocated to each customer).

Following is a list of CPLEX parameters that can be used for the CPLEX node in IBM SPSS Modeler client. These are the valid custom names that are recognized by the CPLEX integration with IBM Analytical Decision Management.

Important: Refer to the IBM ILOG CPLEX Optimization Studio documentation for details about these settings. Note that they may have and adverse affect on IBM Analytical Decision Management's performance.

Table 9. CPLEX parameters

Boolean Settings	Floating Point Settings	Integer Settings	String Settings
CloneLog	EpMrk	AdvInd	WorkDir

Table 9. CPLEX parameters (continued)

Boolean Settings	Floating Point Settings	Integer Settings	String Settings
PreInd	EpOpt	RootAlg	FileEncoding
MIPOrdInd	EpPer	NodeAlg	APIEncoding
MPSLongNum	EpRHS	MIPEmphasis	IntSolFilePrefix
LBHeur	NetEpOpt	AggInd	
PerInd	NetEpRHS	ClockType	
PreLinear	TiLim	CraInd	
DataCheck	DetTiLim	DepInd	
QPmakePSDInd	TuningTiLim	PreDual	
MemoryEmphasis	BtTol	PrePass	
NumericalEmphasis	CutLo	RelaxPreInd	
	CutUp	RepeatPresolve	
	EpGap	Symmetry	
	EpInt	DPriInd	
	EpAGap	PriceLim	
	EpRelax	SimDisplay	
	ObjDif	NetFind	
	ObjLLim	PerLim	
	ObjULim	PPriInd	
	PolishTime	ReInv	
	PolishAfterEpAGap	ScaInd	
	PolishAfterEpGap	Threads	
	PolishAfterTime	AuxRootThreads	
	ProbeTime	ParallelMode	
	RelObjDif	SingLim	
	CutsFactor	Reduce	
	TreLim	ColReadLim	
	SolnPoolGap	RowReadLim	
	SolnPoolAGap	WriteLevel	
	WorkMem	SiftDisplay	
	BarEpComp	SiftAlg	
	BarQCPEpComp	BrDir	
	BarGrowth	Cliques	
	BarObjRng	CoeRedInd	
	EpLin	Covers	
		MIPDisplay	
		NodeFileInd	
		NodeSel	
		VarSel	
		BndStrenInd	
		FPHeur	

Table 9. CPLEX parameters (continued)

Boolean Settings	Floating Point Settings	Integer Settings	String Settings
		MIPOrdType	
		FlowCovers	
		ImplBd	
		Probe	
		GUBCovers	
		StrongCandLim	
		FracCand	
		FracCuts	
		PreslvNd	
		FlowPaths	
		MIRCuts	
		DisjCuts	
		ZeroHalfCuts	
		MCFCuts	
		AggCutLim	
		EachCutLim	
		DiveType	
		MIPSearch	
		MIQCPStrat	
		SolnPoolCapacity	
		SolnPoolReplace	
		SolnPoolIntensity	
		PopulateLim	
		BarAlg	
		BarColNz	
		BarDisplay	
		BarOrder	
		BarCrossAlg	
		BarStartAlg	
		NetPPriInd	
		NetDisplay	
		ConflictDisplay	
		FeasOptMode	
		TuningMeasure	
		TuningRepeat	
		TuningDisplay	
		MIPKappaStats	
		SolutionTarget	

Chapter 10. Running simulations and tests

The **Whatif?** functionality available on the Optimize, Prioritize, or Combine tab and the **Simulate** functionality available on the Define tab in pre-8.0 versions of the product have been combined, enhanced, and added to the main toolbar. The new functions are available at any point in the application, and scenarios can be saved and compared. After running a scenario, a preview of scenario results also appears on the Home page and persists across browser sessions.

The Test functionality has also been moved to the main toolbar and is available at any point in the application.

Creating scenarios

From any point in an application that includes simulation, you can create a scenario to see how records will be allocated based on the current settings. Each scenario you run will be saved with the project. You can select from the following options available on the toolbar:

 Creates a new scenario based on the current project settings. The New Scenario dialog will be displayed.

 Opens an existing scenario. A drop-down displays available scenarios, accompanied by an icon that indicates whether the scenario ran successfully, is still running, or failed.

 Opens the Manage Scenarios dialog, allowing you to delete or rename one or more scenarios that exist in the current project, and add or edit their descriptions. The dialog also shows the date and time each scenario was created.

 Compares up to three existing scenarios. The Compare Scenarios dialog is displayed.

 Updates the project settings based on the settings in the scenario that is currently open.

1. To create a new scenario in your project, click the **Create a new scenario based on current project settings** icon on the toolbar. The New Scenario dialog will be displayed.
2. Select the data source you want to use. Typically this is a sample of a few thousand records (though it can be larger or smaller) that is representative of your operational data.
3. Type a name for this scenario run, or accept the default. If desired, also type a description.
4. Specify the simulation date, if desired. This setting depends on the current application, but can have significant impact on the results.
5. In the Content section, select the project step (tab) through which to run the scenario.
6. If you want to include fields in the scenario results, click **Fields** and make selections.
7. If you want to include fields and associated aggregation methods to include as results metrics in the scenario results, click **Metric**.
8. Click **Run**. Note that while the scenario runs, you can go to a different screen or close the application altogether. The scenario will be saved and you can come back and open it later. You can also run multiple scenarios. The last run scenario will be displayed on the application Home page.

Scenario results

After running a scenario, results are presented in various forms. The upper section displays a summary report, and the bottom section displays interactive results.

In the upper-left corner, you can select which step (tab) to display results through. For example, results are included for all allocations and rules, or you can display results for each individual allocation or rule. For applications with an Optimize/Prioritize tab, you can also view optimization inputs and outputs. And for applications with a Combine tab, you can view the matrix.

The simulation displays separate results for each dimension, indicating how records are allocated across the items in that dimension. The numbers indicated in each dimension are the sum of those in the sub dimensions. For example, the total number of people eligible for a campaign is the sum of those eligible for offers within the campaign.

If the application has multiple dimension hierarchies, the element-specific results will be displayed on separate tabs.

Select from options such as the following to customize the view.
- **Primary dimension.** Select the primary dimension for which to display results.
- **Overlay dimension.** If desired, you can include an *overlay* from another hierarchy member, which will include each dimension available in the selected overlay. This will create a matrix-style table and charts with the count for each possible variation of the overlay displayed. For example, you may want to overlay offers across channels to see the offers made on each channel.
- **Chart metric.** Select whether to display the count or percent in the results.
- **Graphical format or tabular format.** Select the **Display scenario results in graphical format** icon or the **Display scenario results in tabular format** icon to switch between charts and tables.

Testing the application

The Test dialog box allows you to see how one or more sample records will be scored. Results are specific to the context in which the test is run, and may reflect the scores returned by a model or rule, or the recommendations returned by the current application. Following are some examples of the types of results you may see, depending on your application type:
- You might look at a handful of specific customers and see which offers are allocated to each of them.
- You might look at a handful of customer records and see which customers are likely to respond to a direct mail campaign.
- You might look at a handful of specific insurance claims and see which actions are allocated to each of them.
- You might look at a handful of specific cases and see how each rule evaluates.

In addition, you can view details to see how specific selections and rules are applied as factors in the decision.

1. Select the data source you want to use, or select **Specify custom data** in the Records section to enter test records manually.
2. Specify the test date to use where applicable. This setting depends on the current application, but can have significant impact on the results.
3. Specify selections to use, if any. For example, you may want to include or exclude certain records from the test. You can search for existing rules, or create new ones as appropriate. See the topic "Defining selection rules" on page 53 for more information.
4. If using a data source, specify the number of records you want to return. You can return the first N records, up to a maximum of 100, read from the top of the data or from a random sample of records.

5. If using custom data, select **Specify custom data** and click **Add custom record**. All fields defined in the project data model are listed, allowing you to specify a value for each. Click **Save** when finished.

6. Under **Select Fields to Display**, select the fields you want to appear in the test output.

7. Select which step of the project you want to test and click **Run**. The test summary displays a list of successful allocations for each record.

8. Click the information icon at the beginning of each row to obtain details for that row.

9. Where available, select **Test Detail** to see additional details as available. You can choose to display successful allocations, unsuccessful allocations, or all attempted allocations for records selected in the Test Summary section. Note that the contents of the table will vary depending on which tab Test is run for and whether the application has multiple elements in its hierarchy tree. If an application's hierarchy tree has more than two dimension levels, additional columns will be displayed for each level.

Note: For applications that use CPLEX optimization algorithm (IBM Analytical Decision Management for Campaign Optimization, IBM Analytical Decision Management for Demand Optimization and IBM Analytical Decision Management for Operations), Test does not disable or ignore any constraints. But global constraints are unlikely to take effect when there is only one customer, for example. Of course, if you set the total budget to an extremely small value, then it will have an effect even for one customer and rule out some of the more expensive offers.

Chapter 11. Deploying applications

Deploying applications

The Deploy tab enables you to check that all parts of a project have been correctly set up and then label the project as ready for use within your organization. The project could be for testing purposes, batch processing, or deployed in a real-time production environment such as a call center, Web site, ATM, or store location as appropriate.

Note: When deploying for real time scoring or scheduled jobs, the project data model defines the inputs that will be required. When deploying for real time scoring, the project data model defines the input fields that will be required. These can either be passed by the caller of the scoring web service or supplied by an Analytic Data View with Real Time Data Access Plan.

When scheduling a job with IBM SPSS Collaboration and Deployment Services Deployment Manager client, the user may override the source by supplying an alternative source that has the fields (including field type) matching those in the project data model.

To view project summary information

In the **Project Summary** section, click **Generate**. The current project will be validated. If no errors are found, a message is displayed and the project is ready to be deployed. All attribute references and repository references are listed.

If errors are found, details about each error are displayed. Scroll through the information to troubleshoot the errors. You can move to other tabs in the application, fix errors, and return to the project summary on the Deploy tab. To refresh the project summary report, click **Generate** or **Update** again.

If desired, you can also download the project summary information to a `ProjectSummary.csv` file. The file will contain the following columns:
- OK
- ErrorMessage
- ReferenceType
- ReferencedFrom
- Member
- Dimension
- Title
- Label
- Path
- ObjectURI
- Attribute
- Table

These columns are a union of the table columns shown in the user interface. If a field does not apply, it will be empty.

To label the project for deployment
1. From the **Deploy As** list, select the label to apply to the project. Available labels are configured by the local administrator, but would typically include options such as *Test*, *Pre-production*, and *Deploy*. See the topic "Defining labels for use in production process" on page 100 for more information.

2. To apply the label, click **Deploy**. If a previous version of the project is already deployed a warning is displayed; you can choose whether to deploy your latest version or keep the existing deployed version.

When a project is labeled for deployment, it can be configured for use with batch or real-time scoring, and integration with existing IT systems (such as a call center or Web site). Typically this is done by a consultant as part of a service engagement. If scoring has already been configured, labeling a new version will cause this version to be used in place of the previous one. Any external models or rules referenced in the project must also be labeled.

To specify Real Time Scoring options

The Real Time Scoring panel includes an **Output Fields** section and an **Interactive Questions** section.

The **Output Fields** section displays the current list of fields configured to be included in the output when the project is scored in real time. You can add or remove fields from the scoring output, or rename them:

1. Click **Add/Edit output fields**. The dialog lists all the variables in the project stream that are available for scoring, and all the fields that will appear in the scoring output.
2. Add fields to or remove fields from the **Output Fields** area, as desired. At least one output field must be selected to be able to score the project.
3. In the **Output name** column, modify names to display in the scoring output, as desired.
4. When finished, click **Save**.

The **Interactive Questions** section can be used to prompt users for additional information when needed. Each question is linked to a specific field in the data model. For example, you may have a promotional campaign for the retired, in which case you need either their age or date of birth; if your data doesn't have this detail, you can create a question that prompts a call center operator to ask for it and enter the response. Your administrator will need to link these questions to whichever system the operator uses when dealing with customers to ensure the correct questions are presented at the right time.

1. For each question you want to specify in the **Interactive Questions** section, select the **Enable** check box for the appropriate field, and specify the prompt text to be used.
2. If interaction points are enabled, you can use the same questions for all interaction points, or specify different questions for each. For example, the phrasing of a question may be different if you are setting it up for use by a call center operator or for display on a web site. See the topic "Defining interaction points" on page 100 for more information.

To specify Batch Scoring options

In the Batch Scoring panel, the basic steps in scoring a project are as follows.

1. Select the interaction point to use for scoring, if applicable.
2. Select the data source, and any subsets of that data, to score.
3. Select the destination for the scored data.
4. Set a deploy date, if desired. You can either specify a single date or specify a date range. If you specify a single date, any campaigns that are valid on that date will be included in the optimization. If you specify a date range, any campaigns that are valid at some point in the range will be included. If you do not set a deploy date, the current date will be used. Note that start and end dates are inclusive (in other words, if the last day of a month is selected as the end date, that date will be included rather than ending at the day before the last day of the month).

   The deploy date specified here is compared to the Active date you specified on the Define tab. See the topic Chapter 8, "Defining decisions," on page 71 for more information.

   Note that setting a deploy date also creates parameters in the project (in the IBM SPSS Modeler stream). This may be useful to administrators who use IBM SPSS Collaboration and Deployment Services Deployment Manager to schedule the projects to run, because additional stream parameters

will be available called *DEPLOY_USE_DATES*, *DEPLOY_START_DATE*, and *DEPLOY_END_DATE*. For example, the administrator may choose to override values set for these parameters when he or she actually runs the stream or schedules it to run. These parameters do not affect real-time scoring.

5. In the **Output Fields** section, click **Add/Edit output fields** if you want to add fields to the scoring output. The dialog lists all the variables in the project stream that are available for scoring, and all the fields that will appear in the scoring output. Add, remove, rename, or reorder fields as desired. To return to the default settings, click **Reset to default**. This resets the output selections to include the ID/Key field from the Project Data Model (if there is one) and any fields with an associated output attribute. When finished, click **Save**.

6. Set other scoring options as desired, then click **Score Now** to run scoring against the project using your settings. When complete, review the Scoring Summary Report section. See the topic "Scoring summary report" on page 94 for more information.

Scoring models to a database table, file, Analytic Server, or IBM Cognos BI server

To score a model means to apply it to the data sample or population you are interested in. For example, you might use your existing customers to generate a model that predicts which customers are most likely to respond to a campaign, and then use the model to score records for people who are not currently your customers, but whom you might want to include in a promotional mailing.

The basic steps in scoring a model are:

1. Select the data source, and any subsets of that data, to score.
2. Select the destination for the scored data.
3. Select the information to output and how to score it.
4. Specify which records to score. From the Score drop-down, select whether to score **All records**, the **Top N%**, the **Top N records**, **Min/Max propensity** (only available for *flag* targets), or **Minimum score** (only available for *continuous* targets).
5. Preview the expected results and run a data overview, if desired.
6. Click **Score Now** to run scoring against the model using your settings. When complete, review the Scoring Summary Report section. See the topic "Scoring summary report" on page 94 for more information.

Scores are written to one or more fields which are added to the data. By convention, the names of these fields are based on the target field, but with a prefix added. For example, if the target field is named *Response*, the output fields will have names like *$XF-Response* and *$XFRP-Response*. See the topic "Selecting output fields" on page 92 for more information.

Note: When the source data originates in a database, execution is pushed back to the database whenever possible (in a process known as "SQL push-back"). In general, IBM Analytical Decision Management supports the same level of SQL push-back as IBM SPSS Modeler but with the following limitations:

- If the project includes additional fields or selection rules which have references to models or external rules, then these references will not push back to the database, and this may inhibit the degree of push-back achievable for the project stream as a whole.
- In applications other than IBM SPSS Modeler Advantage, push-back is enabled only up to selection rules (global selections, and member selections on the first dimension). The degree of push-back may be increased if you choose an Integer field as the entity ID in the project data model.

IBM SPSS Modeler Advantage

For **IBM SPSS Modeler Advantage**, the saved stream will contain two branches: the build branch and the score branch.

The source node of the **build** branch will always be based on the project data model data source. It will include fields that are checked as either operational or not operational.

The source node for the **score** branch will also always be based on the project data model data source. However, it will only include fields that are marked as operational. Fields marked as not operational will not be included as they are not required for scoring. For example, the target field of a predictive model is typically flagged as not operational and is not required as a scoring input.

By default, the score branch of the saved stream will have a Table terminal node. This table node will be used to indicate the score branch in the stream's deployment settings. The IBM SPSS Collaboration and Deployment ServicesScoring Service will use this information to determine the terminal node to use when performing real time scoring.

If configured, the **Score** tab will have a section for **Batch Scoring**. Under **Batch Scoring**, there is a section for **Score Now options**. These settings only apply when scoring within the IBM Analytical Decision Management user interface by clicking the **Score Now** button. They will not be reflected in the generated stream that is saved. As a result, they do not apply when scoring via IBM SPSS Modeler client nor when scheduling a job using IBM SPSS Collaboration and Deployment Services Deployment Manager client.

Batch scoring also has controls for selecting the destination and the output fields. Upon saving the stream, an additional terminal node will be created on the score branch. The destination and the output field settings will be used to create an Export terminal node on the scoring branch. These settings apply when running the stream from IBM SPSS Collaboration and Deployment Services Deployment Manager client or when scheduling a job using IBM SPSS Collaboration and Deployment Services Deployment Manager client.

Non-IBM SPSS Modeler Advantage projects

For projects **other than IBM SPSS Modeler Advantage**, the saved stream will only have the score branch. Just like the IBM SPSS Modeler Advantage score branch, the source node for the score branch will always be based on the project data model data source. Again, it will only include fields that are marked as operational. If secondary tables are used, additional source nodes will be created. These secondary source nodes are defined by the secondary data sources that are associated with the primary data source.

The **Deploy** tab, if configured, may have sections for **Real Time Scoring** and **Batch Scoring**. These settings are independent of each other.

Upon saving the stream, the **Real Time Scoring** settings will be used to create a Table terminal node on the scoring branch. This table node will used to indicate the score branch in the deployment settings of the saved stream. The IBM SPSS Collaboration and Deployment Services Scoring Service will use this information to determine the terminal node to use when performing real time scoring.

Under **Batch Scoring**, there is a section for **Score Now options**. These settings only apply when scoring within the IBM Analytical Decision Management user interface by clicking the **Score Now** button. They will not be reflected in the generated stream that is saved. As a result, they do not apply when scoring via the IBM SPSS Collaboration and Deployment Services Scoring Service nor when scheduling a job using IBM SPSS Collaboration and Deployment Services Deployment Manager client.

Batch scoring also has controls for selecting the destination and the output fields. Upon saving the stream, these settings will be used to create an Export terminal node on the scoring branch. These settings apply when scheduling a job using IBM SPSS Collaboration and Deployment Services Deployment Manager client or when running the stream from IBM SPSS Modeler client.

All projects

Both IBM SPSS Modeler Advantage and non-IBM SPSS Modeler Advantage projects may construct distinct terminal nodes for real time and batch scoring. When scheduling jobs using IBM SPSS Collaboration and Deployment Services Deployment Manager client, be sure to configure the job step so that just the one export node designated for batch scoring is executed.

Selecting data and subsets

From the **Data source to score** list, select the required data source.

Data subsets

If you don't want to score the entire data source, you can select a subset of the data by using an existing rule, or creating a new rule that will only be used for this model. See the topic "Defining selection rules" on page 53 for more information.

Selecting records to be scored

For models that return propensity scores, you can use these values to select the records to be scored. For example, you could score the 30% of customers with the highest propensity to respond, and use these to produce a mailing list. Propensity scores indicate the likelihood of the "true" value for the target, and are available for predictive models with flag targets, as well as predictive rule (interactive) models. (In the case of predictive rule models, the selected response is the "true" value and all others are false. For other types of models, these options are not available.)

There are three ways you can choose to select records based on propensity:

- **The top percentage.** By default this selects the top ten percent. This option enables you to simulate your scoring results based on profit. See the topic "Simulating maximum profit" on page 67 for more information.
- **The top number of records.** By default this selects the top 1000 records, where available.
- **By specifying minimum and maximum propensity values.** For example, all records that have a propensity to respond between 75% and 100%.

Selecting the scoring destination

You can choose to output the scoring data to a database, file, Analytic Server, or Cognos BI server. Supported file types include text (*.txt, *.csv), Excel, the *.sav format used by IBM SPSS Statistics, SAS, and those supported by IBM SPSS Data Collection. See the topic "File destination" for more information.

Database destination

When you choose to send scoring output to a database you are prompted to enter further details.

1. Enter the **Database name** or browse for the one you require.
2. Specify how scores are to be stored. You can choose to create a new table, in which case you are prompted to enter a name for the table. Alternatively, you can either append them to an existing table or overwrite an existing data source, in which case you are prompted to select the relevant table or data source.
3. Change the default string size, if desired.

When you append to, or overwrite, an existing database you can map the scoring fields to fields in the existing destination. See the topic "Map scoring fields" on page 92 for more information.

File destination

Scores can be output to several different file types. For each type, you either browse for an existing file or create a new file to which the scoring data is added.

When you append to, or overwrite, an existing file you can map the scoring fields to fields in the existing destination. See the topic "Map scoring fields" on page 92 for more information.

Excel file

Type. Select the Excel file type that you are exporting.

Include field names. Specifies whether field names should be included in the first row of the worksheet.

Flat file

When you choose to send scoring output to a flat file you can enter further details to control the file contents.

Write mode. If **Overwrite** is selected, any existing data in the specified file will be overwritten. If **Append** is selected, output will be added to the end of the existing file, preserving any data it contains.
* **Include field names.** If this option is selected, field names will be written to the first line of the output file. This option is available only for the **Overwrite** write mode.

New line after each record. If this option is selected, each record will be written on a new line in the output file.

Field separator. Specifies the character to insert between field values in the generated text file. Options are **Comma**, **Tab**, **Space**, and **Other**. If you select **Other**, enter the desired delimiter character(s) in the text box.

Symbol quotes. Specifies the type of quoting to use for values of symbolic fields. Options are **None** (values are not quoted), **Single (')**, **Double (")**, and **Other**. If you select **Other**, enter the desired quoting character(s) in the text box.

Encoding. Specifies the text-encoding method used. You can choose between the system default or UTF-8.

Decimal symbol. Specifies how decimals should be represented in the data.
* **Stream default.** The decimal separator defined by the current stream's default setting will be used. This will normally be the decimal separator defined by the computer's locale settings.
* **Period (.).** The period character will be used as the decimal separator.
* **Comma (,).** The comma character will be used as the decimal separator.

IBM SPSS Statistics data file

Export field names. Specifies a method of handling variable names and labels upon export to a IBM SPSS Statistics *.sav* file. Note that any invalid names are automatically corrected by replacing invalid characters.
* **Names and variable labels.** Select to export both field names and field labels. Names are exported as IBM SPSS Statistics variable names, while labels are exported as IBM SPSS Statistics variable labels.
* **Names as variable labels.** Select to use the field names as variable labels in IBM SPSS Statistics. To prevent possibly creating invalid IBM SPSS Statistics names, select **Names and variable labels** instead.

SAS data file

Type. Select the SAS file format to be created. You can choose from three SAS file formats: SAS for Windows/OS2 (*.sd2), SAS for UNIX (*.ssd), or SAS Version 7/8/9 (*sas7bdat).

Export field names. Select options for exporting field names and labels for use with SAS.

- **Names and variable labels.** Select to export both field names and field labels. Names are exported as SAS variable names, while labels are exported as SAS variable labels.
- **Names as variable labels.** Select to use the field names as variable labels in SAS. To prevent possibly creating invalid SAS names, select **Names and variable labels** instead.

Analytic Server destination

When you choose to send scoring output to an Analytic Server you are prompted to enter further details.

1. Specify the **Database source**.
2. Specify how scores are to be stored. You can either append them to an existing table or overwrite an existing data source, in which case you are prompted to select the relevant table or data source.
3. Select output fields to include with the scoring output.
4. Specify which records to score. From the Score drop-down, select whether to score **All records**, the **Top N%**, the **Top N records**, **Min/Max propensity** (only available for *flag* targets), or **Minimum score** (only available for *continuous* targets).

Cognos BI server destination

When you choose to send scoring output to a Cognos BI server you are prompted to enter further details.

Note: You can export only relational data, not OLAP data.

To export data to Cognos BI, you must specify the following:
- **Cognos connection.** The connection to the Cognos BI server.
- **ODBC connection.** The connection to the Cognos data server that the Cognos BI server uses.

The connections must point to the same database, and the username and password for the Cognos server to connect to ODBC must be the same as the ODBC details.

You export the actual data to the data server, and the package metadata to the Cognos BI server.

1. Select the Cognos BI Server **Destination** and click **Connect**; you are prompted to enter the server connection details.
   a. Type the **Server URL** of the IBM Cognos server from which to import or export data. Contact your Cognos system administrator if you are not sure which URL to use.
   b. Select the **Mode** in which you will connect. To log in as a specific user (for example, as an administrator), select **Set credentials** and enter the IBM Cognos namespace, user name, and password:
      - Enter the IBM Cognos security authentication provider **Namespace** used to log on to the server. The authentication provider is used to define and maintain users, groups, and roles, and to control the authentication process.
      - Enter the IBM Cognos **User name** with which to log on to the server.
      - Enter the **Password** associated with the specified user name.

      OR

      Browse to and select an IBM Cognos credential stored in the IBM SPSS Collaboration and Deployment Services Repository.
   c. Alternatively, to log in with no user credentials, in which case you cannot specify the namespace, user name, and password fields, select **Use Anonymous Connection**.

      Note: Some server connections do not allow anonymous connections.
2. Select the **Data source**.
3. Type the path and name of the **Folder** on the Cognos BI server where the export package is to be created. Alternatively, you can browse for the one you require.

4. Type the **Package name** of the package that is to contain the exported metadata. This must be a new package; you cannot export to an existing package. See the topic "Selecting Cognos object details" on page 38 for more information.

5. To select the ODBC Destination, enter the **Database name** or browse for the one you require.

6. Specify how scores are to be stored. You can choose to create a new table, in which case you are prompted to enter a name for the table. Alternatively, you can either append them to an existing table or overwrite an existing data source, in which case you are prompted to select the relevant table or data source.

When you append to, or overwrite, an existing database you can map the scoring fields to fields in the existing destination. See the topic "Map scoring fields" for more information.

Map scoring fields

When you append or overwrite scores to an existing database table or file, you may need to map the scoring fields to fields in the existing table or file. Note that the output and associated mapped field must have the same data type.

- When writing scores to an existing database table, all fields must be mapped to existing fields in order for scoring to proceed, whether overwriting or appending the new scores. If necessary, click **Map Fields** to specify outputs for any fields that remain.

- When writing to an existing file, mapping is only necessary if **Append** is selected under output file options. If overwriting an existing file, mapping is not required as the new file will simply replace the old one.

1. Click **Map Fields**. A dialog opens, and the system makes a best guess at mapping fields for you. At any point, you can return to the system's best guess by clicking **Reset**.

2. Select from the available output destinations and click **Map** to map them to the required scoring fields.

3. When finished, click **Save**.

Selecting output fields

On the Score tab, select **Specify output fields** to choose the fields to include with the scoring output. In addition to the actual scores, you will typically want at least one field that allows you to identify each record, such as a customer id field. Optionally, some or all of the input fields used in modeling may also be included.

Model output fields

By convention, the names of the scoring output fields are based on the name of the target field, but with a custom prefix added. For example if the target field is named *Response*, the output fields will have names like *$XF-Response* and *$XFRP-Response*. The specific fields depend on the type of model, and the measurement level for the selected target field, as detailed below. See the topic "Measurement levels" on page 41 for more information.

Flag targets. Automated models with flag targets output the following fields:

- **$XF-<target>.** Predicted value for each record, reported in terms of the "true" and "false" values defined for the field.

- **$XFRP-<target>.** Propensity score; indicates the likelihood of the "true" value for each record. Propensity scores differ from confidence scores, which are typically reported in terms of the prediction for that record. For example a "false" prediction with a high confidence will translate to a low propensity as this would indicate a high likelihood *not* to respond.

Categorical targets. Automated models with categorical targets (nominal or ordinal) output the following fields:

- **$XS-<target>.** Predicted value for each record.

- **$XSC-<target>.** Confidence value associated with the prediction

Continuous targets. Automated models with continuous numeric targets output the following fields:
- **$XR-<target>.** Predicted value for each record.
- **$XRE-<target>.** Standard error for the prediction.

Predictive rule (interactive) models. Predictive rule models (flag or categorical targets only) output the following fields:
- **$D-<target>.** Predicted outcome for each record, with the "true" value defined for the field used to indicate the desired response, and null indicating any other value.
- **$DP-<target>.** Probability of desired outcome for each record; matches the probability displayed for the segment in the results table when building the model. See the topic "Building a predictive rule model" on page 59 for more information.
- **$DI-<target>.** Index number used to identify the segment into which the record falls (displayed in the left-most column in the modeling results).
- **$DRP-<target>.** Propensity score indicating the likelihood of the desired response. Available for flag targets only.

Scoring models

Once you have specified your scoring details, you can preview a sample of the model scores. This is especially useful when you have a large dataset since it enables you to quickly check that the scoring results appear as you expect, without having to wait for the entire dataset to be scored.

Note: If the data is found to be invalid or incomplete, a warning is displayed when you attempt to preview it.

The preview shows data as a table that displays records by row and variables by columns.

Scoring

To establish how many records will be scored, click the refresh icon. This provides a count of the number of records that will be used to build the model, taking into account the total number of records in the data source and any optional settings you specify.

When you click **Score Now**, a progress bar is displayed across the bottom of the tab. You can stop scoring at any time while this is displayed.

Notes:

If the data is found to be invalid or incomplete, a warning is displayed when you attempt to score it.

When scoring to a database, the execution is pushed back to the database whenever possible (often referred to as "SQL pushback"). Note that if the project includes a rule that contains a reference to a model, the execution will not push back to the database.

Once the data is successfully scored, a message informs you where the results have been stored--either in the database or file you specified--and the Scoring Summary Report section provides a summary. See the topic "Scoring summary report" on page 94 for more information. If you chose to send the scoring output to a flat file, and if an administrator has enabled the functionality, you can download a local copy of the file if desired.

Scoring summary report

After scoring a project, the Scoring Summary Report section provides details such as where the data was scored to, file size (if applicable), database and table name (if applicable), model name or project name, and number of records scored. The Scoring Summary Report section is available in IBM SPSS Modeler Advantage, and on the Deploy tab of applications that include the Score Now feature.

If you chose to send the scoring output to a flat file, you can also download a local copy of the file if desired. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to enable or disable the feature, and to increase or decrease the maximum file size limit allowed. If the **Download File** button is inactive, either the file size exceeds the maximum size limit set by the administrator, or the administrator disabled the feature. The feature is disabled by default.

Chapter 12. Reporting results

Report types

IBM Analytical Decision Management is able to read reports from many formats, such as:
- Text files (*.txt*, *.csv*)
- Spreadsheets (*.xls*, *.xlsx*)
- Images (*.png*, *.jpeg*, *.gif*, *.bmp*)
- HTML
- PDF

In addition to these file types, IBM Analytical Decision Management supports the reporting functionality enabled by BIRT (Business Intelligence and Reporting Tools), an open-source package distributed by Eclipse Foundation under the Eclipse Public License. BIRT provides core reporting features, such as report layout, data access, and scripting. For more information about BIRT, see the IBM SPSS Collaboration and Deployment Services documentation and the BIRT project page. BIRT Report Designer for IBM SPSS is a standalone application that can be used in conjunction with IBM Analytical Decision Management. It provides a rich user interface with a number of advanced features for creating reports and must be installed separately.

Note: If a BIRT Report Designer for IBM SPSS report requires a JDBC-based database connection, a corresponding JDBC driver must be installed with the IBM SPSS Collaboration and Deployment Services Repository. For application server-specific information on the location of the JDBC drivers, see the corresponding section of the IBM SPSS Collaboration and Deployment Services Repository installation instructions.

Accessing a report

Depending on how your system has been set up, you can access reports from a link on the main application page, from a dedicated report tab, or both.

Any reports that you have already viewed are displayed.

To retrieve a report:
1. Click **Find a Report** to display a browser window where you can search for the report.
2. Select the report.
3. Click **Open**; the report is added to the list of those you can access from within the application.
4. To view the report's contents, click the icon in the **Action** column. The report opens in a new window.

Chapter 13. Administering applications

Administrators can control the way business users interact with IBM Analytical Decision Management applications. This level of control makes business users' jobs easier so they can more quickly and efficiently use the application to help solve business problems. It also enables administrators to restrict access to certain functionality.

Use IBM SPSS Collaboration and Deployment Services Deployment Manager to grant administrative authority to specific users or groups as needed. Only users who are granted the *Decision Management Administration* action have access to the features described in this section.

1. Open IBM SPSS Collaboration and Deployment Services Deployment Manager.
2. From the Tools menu, select **Server Administration**.
3. On the Server Administration tab, connect to your server.
4. Under **Users and Groups**, select **Local User Repository**.
5. Click **New Group** to create a new group for administrators and a new group for business users. For example:
   - Decision Management Administrators
   - Decision Management Users

   You can also create new users at this time, if desired. While creating the groups, add the appropriate users to each group. For example, members of the administrators group will be granted authority to administer IBM Analytical Decision Management applications at your organization.
6. Go to Roles and click **New Role** to create a new administrative role and a new business user role. For example:
   - IBM Analytical Decision Management administrator
   - IBM Analytical Decision Management user

   While creating the new roles, assign them the following actions, at a minimum.

Table 10. Role actions.

Role	Actions Required
Decision Management administrator	Access Contents and Folders
	Decision Management Administration
	Job Edit
	Run Report Dynamically
	Score Model
	Submit Work
	Show All Versions
Decision Management user	Access Contents and Folders
	Job Edit
	Run Report Dynamically
	Score Model
	Submit Work
	Show latest *OR* Show All Versions*

Note: Either the *Show latest* or *Show All Versions* action is required if users need access to project versions they did not create. Otherwise projects will open in read-only mode if the current user does not have access to the latest version.

7. Select the Decision Management administrator role and click **Edit Users and Groups**. Add the Decision Management Administrators group you just created.
8. Select the Decision Management user role and click **Edit Users and Groups**. Add the Decision Management Users group you just created.

General options

Certain general options are available to administrators on all tabs of the IBM Analytical Decision Management user interface. These options allow administrators to hide or lock tabs from business users.

* The top of each tab has the option **Hide <tab name> tab from non-administrators**, where **<tab name>** is the name of a tab. This hides the entire tab from business users. For example, if you don't want users to see the Data tab for an application, log on to the application, go to the Data tab, and select **Hide Data tab from non-administrators**.
* The top of each tab has the option **Lock all <tab name> options**, where **<tab name>** is the name of the tab. This locks all controls on the tab so they're read-only for the business user. For example, if you want business users to be able to view current data settings but not change them, you can set this option on the Data tab of an application. This setting is also very useful for situations when you want to set certain things up in advance for the business user to simplify their job.
* Administrators can also lock and unlock individual settings on each tab by clicking the padlock icon when available.

For example, in applications with a Define tab, administrators can lock the dimension hierarchy tree so business users cannot change it. Clicking the padlock icon locks or unlocks its associated control. Business users will see the locked padlock icon for controls the administrator locks, but will never see an unlocked padlock icon.

Individual items that are locked may also be displayed with a grayed-out background or muted colors to indicate that they are unavailable for selection; an error message is displayed if a user attempts to select them. When an entire tab is locked a message to that effect is displayed at the top of the tab.

Controlling application shortcuts

The Applications Launch Page provides an easy way for business users to open applications and add custom shortcuts to specific, labeled versions of applications. Administrators can specify which application shortcuts will show up on the launch page for all users. After logging on with administrator authority:

* If desired, click the padlock icon on an application shortcut so business users cannot remove it from their launch page. Note that business users can still add and remove other applications shortcuts. Users will see the locked padlock icon instead of the close icon for shortcuts that are locked. See the topic "Launching applications" on page 23 for more information.

Hiding the Home page

Each application provides a step-by-step workflow, as represented by the graphics on the Home page. Clicking a graphic jumps to that section of the user interface. However, business users may not always need access to every step in an application's workflow.

Administrators can hide the Home page from business users. For example, in a certain application business users might only use the Reports tab. In this case an administrator could hide the Home page and all other tabs from business users. So when the business user logs on he is taken directly to the Reports tab and never sees any other tabs.

1. On the Home page, select **Hide Application home page**.
2. In the drop-down, specify which tab to display when users first open the application.

   Note that if you subsequently hide this selected tab, the next unhidden tab in the application is displayed.

Locking data options

Data

The Data tab defines the data sets the application uses for analysis, simulation and testing, scoring, or other operational use. These data sources are defined relative to your server installation; hence all paths to data files, ODBC sources, and so on are all relative to the server.

As an administrator, you may want to set up some or all data before business users start using their applications. Administrators can:

- Select **Lock Datasource** when creating a new data source to lock it after it's created.
- Lock and unlock individual data sources by clicking the padlock icon next to a data source in the table, or in the Data Source Editor dialog.
- Set and lock the data model used in the application by selecting a data model in the **Project Data Model** drop-down and clicking the padlock icon.

  Note that business users can copy a locked data source from Project Data Sets to their My Data Sets section, but it remains locked in both sets. See the topic Chapter 4, "Managing data sources," on page 31 for more information.

- If the project data model is a *Data View*, use the **Allow use of ODM rules based on hierarchical data** drop-down to control the usage of IBM Operational Decision Management (ODM) rules that have a hierarchical data model. Use of such rules can result in poor performance. The following options are available in the drop-down:
  - **Don't allow** is recommended for best performance. Disallows ODM rules with a complex data structure from being used anywhere in the project. An error is displayed if a user selects a complex rule at any point in the application.
  - **On Data tab only** allows the complex rules to be referenced only on the Data tab, generating an additional field to use elsewhere in the project. This is done using the **Add/Edit additional fields and tables** link). This allows added flexibility, with a performance hit. This option isn't available in the IBM SPSS Modeler Advantage application, which only supports complex rules on the Modeling tab
  - **Anywhere in project** allows the complex rules to be referenced anywhere throughout the project. This provides maximum flexibility, with the slowest performance. This option isn't available in the IBM SPSS Modeler Advantage application, which only supports complex rules on the Modeling tab.

Enforcing global selections

Global selections allow you to choose the records to include or exclude from an application. For example, you could exclude customers with bad credit, or select certain types of claims for special processing. As an administrator, you may want to set up some global selections and enforce them for all users of an application.

For example, to exclude customers under a certain age, you can create a rule like *Age<18* and set it to *Exclude*. You can then lock this rule on the Global Selections tab to ensure it is always enforced by the application.

Administrators can:

- Lock and unlock individual selections by clicking the padlock icon next to a selection in the table.
- Select **Lock Rule** when creating a new rule to lock it after it's created.

Defining interaction points

Interaction points specify where an item such as a campaign or offer applies. Options might include a call center, web site, ATM, or in-store location. Administrators use the Define Interaction Points dialog available on the Define tab to predefine the interaction points, which are then available for selection by business users.

Administrators can:

- Create, modify, or remove interaction points available in the application.
- Enable or disable interaction points. If enabled, an interaction point is selected for business users by default. Business users can select which available interaction points apply to the item they're defining.

To work with interaction points, click the **Define Interaction Points** link in the upper-right section of the screen. After defining interaction points, they'll appear in the dimension properties section.

Locking constraint and interaction point options

Interaction points can be used in both the prioritization and combination steps; these are effectively two versions of the same step, but with different ways of calculating a result.

For applications that use prioritization, the best decision is determined by using a prioritization equation that defines your goal, such as maximizing profit. The goal of prioritization is to maximize (or minimize) the value of this function in order to make the best decision for each record. On the Prioritize tab, as an administrator you can:

- Lock all prioritization options to prevent business users from specifying different settings.
- Set and lock the interaction point option **Use same settings for all interaction points** if you don't want business users to be able to specify different settings for each interaction point.
- Set and lock the **Max number of offers** value for constraints) if you don't want business users to change the maximum number of offers you specify.

For applications that use combination, the best decision is made by taking rule and model outputs from the Define tab and combining them to result in a single output. On the Combine tab, as an administrator you can:

- Lock all combination options to prevent business users from specifying different settings.

For complete details about the Prioritize or Combine step, see the *Application User's Guide*.

Defining labels for use in production process

Business users can use the Deploy tab to deploy applications for use in a test or product environment such as a call center, web site, ATM or store location as appropriate. During deployment, business users can select a specific version of the application to deploy. When deployed, the selected version and all its artifacts are given a label to identify them as the deployed version. As an administrator, you can create labels to be used during this process. For example, you may want to create a label called *Deploy*, a label called *Test*, and a label called *Pre-production*, all with different colors assigned to them for quick identification purposes.

On the Deploy tab, the Deploy As Labels section is only visible to administrators. Business users don't have access to these options.

Administrators can set, modify, or remove production process labels available to business users in an application by:

- Selecting from a list of labels that already exists in the IBM SPSS Collaboration and Deployment Services Repository.
- Creating new labels in the IBM SPSS Collaboration and Deployment Services Repository.

To add or remove a label:

1. Click **Add/Remove existing label**. The **Add/Remove production process labels** dialog opens.
2. Select labels in the Available Labels column and move them to the Production Process Labels column to make them available for use in the production process. Move labels back to the Available Labels column to remove them from the production process. Only labels appearing in the Production Process Labels column can be used in the production process.

   Note: When you create a new label, it is applied to a temporary object in the IBM SPSS Collaboration and Deployment Services Repository (labels cannot exist without being assigned to an object). Until the new label is used in an application, it will be deleted if it is removed from the Production Process Labels section.
3. Assign a color to each label used in the production process. When finished, click **OK**.

To create a new label:

1. Click **Create New Label**. The **New Label** dialog opens.
2. Type a name for the new label, assign a color to it, and click **OK**.

Hiding interactive questions

Business users can use interactive questions to request additional data that may not be in the data source. For example, a promotional campaign for the retired would need a person's age or date of birth to confirm their eligibility; the interactive question can be set up to prompt a telesales operator to request this information.

On the Deploy tab, administrators can select **Hide Interactive Questions** to hide the section from business users.

Locking scoring destination options

For applications configured to use the Score Now section (Deploy or Score tab), there is a section that allows business users to specify the destination for storing score data (to database or file). For example, if storing scores in a database, you specify the database name and the method for storing the scores (such as creating a new table in the database).

As an administrator, you can lock the destination options for scoring. For example, you may want scores to only be appended to an existing table in a certain database—instead of giving business users the option to overwrite existing data, append new data to an existing table, create a new table, or store data in a file.

- To lock all scoring destination options, click the padlock icon next to the Destination drop-down on the Deploy tab. Business users will be able to see the destination options, but not edit them.

See the topic "Deploying applications" on page 85 for more information.

Locking report options

On the Reports tab, business users can view reports available in an application, specify which report is used on the Home page of the application, and give a title to the report used on the Home page. The report on the Home page is displayed in a small panel on the bottom of the screen, providing a quick summary of current results. If desired, administrators can lock report options:

- To lock the report used on the Home page of an application, go to the Reports tab and click the padlock next to the drop-down. Locking this option also prevents business users from removing the associated report from the table.
- To lock the title used for the Reports section on the Home page of an application, go to the Reports tab and click the padlock icon next to the text field.

Downloading project metadata for use in creating external rules

Rules created in a Business Rules Management System such as IBM Operational Decision Management can be referenced and used in IBM Analytical Decision Management applications, provided they have been developed to support the same data model used in the current IBM Analytical Decision Management project, and can be deployed as a Web service for use by the application. To accomplish this, an IBM Analytical Decision Management administrator must download the relevant metadata. There are two ways to download the metadata, and the complexity of the data model and the potential resultant rule complexity differ accordingly:

Downloading project metadata from IBM Analytical Decision Management

An administrator can click the **Download metadata** icon in IBM Analytical Decision Management to download a *.ZIP* file containing metadata for the current project. The downloaded file is saved locally on the system used by the administrator to trigger the download.

Figure 17. Download metadata icon

The metadata downloaded this way supports the creation of rules based on the project data model. The project data model will be flat (not hierarchical), so any rules authored based on this metadata can only utilize a flat data model. A rule authored based on this flat data model may be referenced in an IBM Analytical Decision Management project no matter the type of the project data model data source.

The following files are included in the *.ZIP* file:

- **XML schema definition (*.XSD).** Contains definitions of the field types in the data model for the current IBM Analytical Decision Management project. This file can be imported into an external development tool such as IBM Operational Decision Management Rule Designer in order to develop rules for use with the current project.
- **Web Service Description Language template (*.WSDL).** Included to assist in development of Web services that use the data model. The *.WSDL* file is provided as a template that can be used to build Web services responsible for mediating messages between the IBM Analytical Decision Management application and an external rule service. Details on the mediating Web service can be manually supplied by the integrator.
- **Include dimension metadata.** If a dimension tree has been defined for the project on the Define tab, dimension tree metadata can optionally be included.

Downloading data model metadata for an IBM SPSS Collaboration and Deployment Services analytic data view

In IBM SPSS Collaboration and Deployment Services Deployment Manager, you can export metadata for a particular analytic data view.

Metadata downloaded in this way supports the creation of rules based on the logical data model of an analytic data view. An analytic data view's logical data model may be hierarchical. An IBM Analytical Decision Management Data View source supports generation of an XML representation of the hierarchical data for each record, which can then be passed to external rules during execution. Rules authored based on this metadata may be complex and utilize the hierarchical data model.

In IBM Analytical Decision Management, an external rule based on a hierarchical data model may only be referenced in a project where the project data model uses a Data View source. To ensure compatibility, the Data View source and the external rule must be based on the same analytic data view's logical data model. Only the Data View source can generate the hierarchical data based on the analytic data view instance it references. In other words, to use external rules, the project data model in the IBM Analytical Decision Management project needs to match that used to create the rule, and the analytic data view makes that possible.

To download a *.ZIP* file with metadata for the analytic data view's logical data model:

1. In IBM SPSS Collaboration and Deployment Services Deployment Manager, open an analytic data view from the repository and go to the **Logical** tab.
2. Right-click a table and export the XOM metadata.
3. Select a location to save the file to and click **Save**.

The following files are included in the *.ZIP* file:

XML schema definitions (*.*XSD*). Two or more XML schema definitions are included. One contains a definition which may be useful for a rule response type. The other schema(s) contain the definitions of the tables, including hierarchical relationship detail and field types in the logical data model of the analytic data view.

You can import these schema files into an external development tool such as IBM Operational Decision Management Rule Designer in order to develop rules for use with a project whose project data model uses a Data View source based on the same analytic data view instance.

For information on creating external rules for use in IBM Analytical Decision Management, see the *Application Designers Guide*.

Note that administrators may restrict use of these complex external rules because they can degrade performance. For details, see "Locking data options" on page 99.

Editing the XML schema definition

The *.XSD* file downloaded from IBM Analytical Decision Management contains two complex data types: one representing the IBM Analytical Decision Management project data model, which should be used for the input parameter of IBM Operational Decision Management rules projects; and one representing a rules return type, which should be used for the output parameter of IBM Operational Decision Management rules projects.

The IBM SPSS Collaboration and Deployment Services Deployment Manager analytic data view metadata contains two or more.*XSD* files: one relating to the logical data model of an analytic data view, which should be used for the input parameter of IBM Operational Decision Management rules projects; and one or more containing a complex type representing a rules return type, which should be used for the output parameter of IBM Operational Decision Management rules projects.

Expert users may choose to edit the generated *.XSD* schema definition to tailor them for the specific rule or rules they wish to create. There are circumstances when the user wants to edit the input and/or the response related types defined in the related *.XSD* schema definition prior to creating their rules. For instance, the expert user may need to edit the input parameter related schema to permit their authored rules to handle the potential of input fields with null values. The expert user can do this by indicating via the *.XSD* schema definition which input parameter fields/elements are required and which are optional.

Editing the input parameter type

The number of possible occurrences of each field-related element is not explicitly set in the *.XSD* schema (i.e., the attributes `minOccurs` and `maxOccurs` are not being set, and the default value for both is 1). As such, if the schema is not edited before using it to generate IBM Operational Decision Management rules, all fields in the input parameter type will be required by the generated rules.

Editing the project data model related input parameter type

If the schema defining the input parameter type was generated by downloading an IBM Analytical Decision Management project's metadata, the input parameter's data model will be flat. In this case, IBM Analytical Decision Management will directly create the payload required by the rule service on execution. This gives the expert user more freedom when editing the input parameter type and still maintaining compatility (i.e., the expert user can remove fields/elements they will not require for rule authoring, or indicate fields/elements are optional).

Editing the analytic data view logical data model related input parameter type

If the schema defining the input parameter type was generated by exporting an analytic data view's logical data model from IBM SPSS Collaboration and Deployment Services Deployment Manager client, the input parameter's data model may be hierarchical. If it is hierarchical, then IBM Analytical Decision Management will *not* directly create the payload required by the rule service on execution. The main part of the payload will be generated by the Analytic Data View source. This limits how the expert user may edit the input parameter type and maintain compatibility (for example, the expert user can only indicate fields/elements are optional).

General information

Prior to invoking a rule created in a Business Rules Management System such as IBM Operational Decision Management, IBM Analytical Decision Management asks the rule's service whether its input parameter data model is hierarchical or flat. If the input parameter is flat, it also checks which inputs are required by the rule's service.

If the input parameter's data model is flat, and the record to score contains a null value for any of the rule service's required fields, IBM Analytical Decision Management will not invoke the rule, and the output/response fields of the rule are populated with default values. The value depends on the type of the output/response field as follows:

Table 11. Values for response field types.

Response field type	Value
xsd:string	""
xsd:boolean	""
xsd:int	0
xsd:long	0
xsd:byte	0
xsd:double	0.0

Table 11. Values for response field types (continued).

Response field type	Value
xsd:float	0.0
xsd:date	0.0
xsd:time	0.0
xsd:datetime	0.0

This can impact the results of scoring records via a rule service, and care should be taken in editing the input parameter in the *.XSD* schema.

If the input parameter's data model is hierarchical, and the record to be scored contains a null value for any of the rule service's required fields, IBM Analytical Decision Management will invoke the rule, but the field or fields with a null value will be omitted from the payload created by the Data View source node. If the omitted field was a required field, this will cause the rule invocation to fail, so again care should be taken when editing the input parameter in the *.XSD* schema.

Editing the schema requires knowledge of the rule services that will be created. If a particular input field is not required by the rule, and the rule's input parameter data model is **flat**, then the field-related element may either be **removed** from the input parameter in the schema or **marked as optional**. If a particular field is not required by the rule, and the rule's input parameter is **hierarchical**, the field-related element **cannot be removed**; but it may be **marked as optional**. This is because IBM Analytical Decision Management does not directly generate the payload for rules based on a hierarchical data model.

When you remove an input parameter field, the field value is ignored when the rule is called and the rule cannot refer to the field. When you make an input parameter field optional, a null value is allowed and the rule must be written defensively to check for that. For example, if the rule's input parameter data model is flat and you know the field Age will not be used by the rule, you can remove the related element to indicate it is not a required field. In the following *.XSD* schema example, if the field Age is not required, the entire section <xsd:element name="Age"> can be removed from the schema:

```
<xsd:complexType name="drugs">
<xsd:annotation>
  <xsd:appinfo>
   <dmname>drugs</dmname>
  </xsd:appinfo>
 </xsd:annotation>
<xsd:sequence>
  <xsd:element name="Age">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Age</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:long">
     <xsd:minInclusive value="15"/>
     <xsd:maxInclusive value="74"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  <xsd:element name="Sex">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Sex</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:string">
     <xsd:enumeration value="F"/>
     <xsd:enumeration value="M"/>
```

```
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  ...
 </xsd:sequence>
</xsd:complexType>
```

Or if the rule's input parameter data model is flat or hierarchical, you might want to make an input field element optional if it's not required by the rule(s)--by setting the element's `minOccurs` attribute value to `0`:

```
<xsd:complexType name="drugs">
 <xsd:annotation>
  <xsd:appinfo>
   <dmname>drugs</dmname>
  </xsd:appinfo>
 </xsd:annotation>
<xsd:sequence>
  <xsd:element name="Age" minOccurs="0">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Age</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:long">
     <xsd:minInclusive value="15"/>
     <xsd:maxInclusive value="74"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  <xsd:element name="Sex">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Sex</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:string">
     <xsd:enumeration value="F"/>
     <xsd:enumeration value="M"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  ...
 </xsd:sequence>
</xsd:complexType>
```

Hierarchical data model example:

```
<xsd:complexType name="Customer">
 <xsd:sequence>
  <xsd:element name="FirstName" type="xsd:string" minOccurs="0"/>
  <xsd:element name="Sirname" type="xsd:string" minOccurs="0"/>
  <xsd:element name="Age" type="xsd:long"/>
  <xsd:element name="Sex" type="xsd:string"/>
  <xsd:element name="CustID" type="xsd:string"/>
  <xsd:element maxOccurs="unbounded" minOccurs="0" name="Orders" type="xom:Order"/>
 </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="Order">
 <xsd:sequence>
  <xsd:element name="CustID" type="xsd:string"/>
  <xsd:element name="NoOfItems" type="xsd:long"/>
  <xsd:element name="TotalCost" type="xsd:long" minOccurs="0"/>
  <xsd:element name="Date" type="xsd:date"/>
 </xsd:sequence>
</xsd:complexType>
```

The rules can also be written defensively in order to handle null values, in which case the field can be marked as `optional` in the schema. Following is an example of an IBM Operational Decision Management rule called *currentOfferIsReplacePhone* that has been written to deal with null values:

```
if
    the CURRENT OFFER of 'the customer' is not null
    and the CURRENT OFFER of 'the customer' is "Replace Phone"
then
    make it true that 'the response' is value ;
else
    make it false that 'the response' is value ;
```

Editing the response parameter type

Note that it is possible to create and utilize IBM Operational Decision Management rules that return a variety of types to IBM Analytical Decision Management (*boolean* for use in Selections, *numeric* for use in aggregate rules for example, and *string*). However, to create anything other than a *string* data response rule, the XML schema definition (*.XSD*) which contains the response type must be manually edited.

Also, a rule project can contain multiple rules, each returning different types. It is necessary to maintain multiple versions of the *.XSD* to support this (for example, a version where the data response variable has been set to numeric, another for string, another for boolean, etc.).

The *.XSD* file downloaded from IBM Analytical Decision Management contains two complex data types: one representing the IBM Analytical Decision Management project data model which should be used for the input parameter of IBM Operational Decision Management rules projects, and one representing a rules return type which should be used for the output parameter of IBM Operational Decision Management rules projects.

The response-related *.XSD* file downloaded from IBM SPSS Collaboration and Deployment Services Deployment Manager contains the response type definition and will only contain the one complex data type.

In both of the above cases, the return type (whose name will always have a "response" suffix) defines a `value` element which defaults to a *string*. If you want to create a rules project with a different kind of return value, you must manually edit the downloaded *.XSD* before importing the IBM Operational Decision Management rules project so that the return type's `value` element is defined as desired. Following are a few examples:

```
<xsd:complexType name="claim_dataresponse">
    <xsd:annotation>
        <xsd:appinfo>
            <dmname>claim_dataresponse</dmname>
        </xsd:appinfo>
    </xsd:annotation>
    <xsd:sequence>
        <xsd:element name="value" type="xsd:string"/>
    </xsd:sequence>
</xsd:complexType>


<xsd:complexType name="claim_dataresponse">
    <xsd:annotation>
        <xsd:appinfo>
            <dmname>claim_dataresponse</dmname>
        </xsd:appinfo>
    </xsd:annotation>
    <xsd:sequence>
        <xsd:element name="value" type="xsd:boolean"/>
    </xsd:sequence>
</xsd:complexType>



<xsd:complexType name="claim_dataresponse">
```

```
    <xsd:annotation>
        <xsd:appinfo>
            <dmname>claim_dataresponse</dmname>
        </xsd:appinfo>
    </xsd:annotation>
    <xsd:sequence>
        <xsd:element name="value" type="xsd:long"/>
    </xsd:sequence>
</xsd:complexType>
```

Chapter 14. Getting help

Three levels of user-assistance are provided:

- Custom help is provided for each prebuilt application packaged with IBM Analytical Decision Management. To launch the help, click the help icon on any screen or dialog:

- Coach text, which is displayed in a small pop-up window, providing details on the current application or business problem. Coach text provides an additional layer of customized, context-sensitive help that can be easily configured for a specific site or application. Coach text appears throughout the user interface when you click the small coach text icon:

- Standard tool tips, which are provided for individual controls within a page or dialog. To display a tool tip, hover the button, link, or other control.

Available documentation

IBM Analytical Decision Management documentation is available in two formats. When you access help in the application, the online IBM Knowledge Center is launched, with the relevant help topic displayed. If you have trouble launching the help, contact your administrator in case you do not have authority to access the external Knowledge Center URL.

The Knowledge Center is an online, centralized collection of all documentation. It is always available, even when the product is not running. We recommend bookmarking it in your browser. It also contains documentation for other IBM products such as IBM SPSS Modeler, IBM SPSS Collaboration and Deployment Services, and Statistics.

At times when you are without an Internet connection, or when the Knowledge Center may be temporarily unavailable for maintenance, all documentation is also available in Portable Document Format (*.pdf*). The PDF files are available on the installation media, and are also installed into the IBM SPSS Collaboration and Deployment Services installation directory (for example, *C:\Program Files\IBM\SPSS\Deployment\7.0\Server\documentation\DecisionManagement*).

PDF documentation is also available on the Web at http://www-01.ibm.com/support/docview.wss?uid=swg27043687, and release notes are available at http://www-01.ibm.com/support/docview.wss?uid=swg27043686.

Appendix A. Sharing streams

Models and projects created in IBM Analytical Decision Management are saved as stream files, and can be modified and used in IBM SPSS Modeler if available. For example, you may want to create a model or project stream using IBM SPSS Modeler Advantage or another IBM Analytical Decision Management application, then customize the stream in IBM SPSS Modeler before deploying through IBM Analytical Decision Management. Alternatively, streams created in IBM SPSS Modeler can be opened in IBM Analytical Decision Management, provided deployment settings have been specified as described in this appendix.

Models that adhere to the format described in this appendix can be viewed, edited, scored, and refreshed using IBM SPSS Modeler Advantage. Other models can also be scored in IBM SPSS Modeler Advantage, but may not display properly and cannot be refreshed.

Opening streams in IBM Analytical Decision Management

When opening streams in IBM Analytical Decision Management or IBM SPSS Modeler Advantage:
- If deployment type is not specified (none), the stream can be opened but only the Data tab is available.
- If deployment type is **Scoring Only**, the Score tab may also be available, provided the scoring branch starts with a single data source node of a type supported by IBM SPSS Modeler Advantage or IBM Analytical Decision Management. In other words, all source types supported by IBM SPSS Modeler are also supported by IBM SPSS Modeler Advantage and IBM Analytical Decision Management as long as there is only one source node in the stream (no merged data sets).
- If deployment type is **Model Refresh,** the Modeling tab may also be available, provided the modeling branch starts with a single data source node of a type supported by IBM SPSS Modeler Advantage or IBM Analytical Decision Management, and the selected modeling node is either an automated modeling node (supported for predictive modeling), a decision list node (supported for predictive rule modeling), a clustering node (supported for cluster modeling), or an association rule node (supported for association modeling). In addition, a Type node must be present that defines the target field as well as other information used in modeling.

Specifying deployment settings in IBM SPSS Modeler

In IBM SPSS Modeler, specify deployment settings in the Stream Properties dialog box, as follows:
1. From the IBM SPSS Modeler menus, choose:

   Tools > **Stream Properties** > **Deployment**
2. From the **Deployment type** list, choose **Scoring Only** or **Model Refresh** as appropriate.
3. Select a **Scoring node**, **Modeling node**, and **Model node** as needed.
4. Select **Check** to confirm that all required options have been specified, or select **Store** to store the stream in IBM SPSS Collaboration and Deployment Services Repository.

For streams created in IBM SPSS Modeler Advantage or IBM Analytical Decision Management, deployment information is set automatically.

Tips for using IBM SPSS Modeler streams in IBM Analytical Decision Management
- It may be easier to create streams using IBM SPSS Modeler Advantage or IBM Analytical Decision Management and then modify in IBM SPSS Modeler, rather than building from scratch using IBM SPSS Modeler. Streams created in this manner will be structured in a manner that allows them to be fully functional in IBM Analytical Decision Management.

- For all streams created in IBM SPSS Modeler Advantage, supernodes are included at certain locations within the branch that can be customized by an expert user. It is strongly recommended that any customizations or additions to the stream be added within these supernodes to ensure they are preserved in the stream and recognized by IBM SPSS Modeler Advantage.
- Any local rules nodes containing an expression rule immediately downstream of each source node will have their fields included in the Add/Edit Additional Fields dialog. Note that local rules can only be created within IBM Analytical Decision Management or IBM SPSS Modeler Advantage and not by IBM SPSS Modeler.
- If there is a Partition node upstream of the modeling node (and downstream of the Type node) then **Automatically partition data to enable model evaluation on build data source for evaluation and testing** is enabled in IBM SPSS Modeler Advantage, and the check box state is set depending on whether the partition node is enabled or not.
- If there is an Automated Data Preparation (ADP) node downstream of the Type node, then **Automatically clean up and prepare data for reliable model building** is enabled and the check box state is set depending on whether the ADP node is enabled or not.
- If Rule or Select nodes are present upstream of the modeling node, these are used to populate the **Specify selections to use** section.
- If stream parameters have been defined (Parameters tab in the Stream Properties dialog box), a Parameters link will be displayed on the Modeling and Score tabs in IBM SPSS Modeler Advantage. The user can click this link to specify these values.

Launching IBM SPSS Modeler Advantage from IBM SPSS Modeler

From the IBM SPSS Modeler splash screen, you can choose to create a model with IBM SPSS Modeler Advantage.

Appendix B. Accessibility

Accessibility features help users who have a disability, such as restricted mobility or limited vision, to use information technology products successfully. IBM strives to provide products with usable access for everyone, regardless of age or ability. This product uses standard web browser navigation keys.

Because IBM Analytical Decision Management is a browser-based application, accessibility options for disabled users, such as those who are visually impaired, are controlled from your web browser settings. For details about user interface actions, including keyboard shortcuts, screen readers, and so on, refer to your web browser's documentation.

Vendor software

IBM Analytical Decision Management may require use of some vendor software that is not covered under the IBM license agreement. IBM makes no representation about the accessibility features of these products. Contact the vendor for the accessibility information about its products.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment IBM has to accessibility.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who want to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Other product and service names might be trademarks of IBM or other companies.

Index