IBM SPSS Decision Management for Customer Interactions 6.2

IBM®

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 incorporating 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.

Contents

5   Global selections   53

6   Working with rules   55

7   Building predictive models   67

8   Defining decisions   89

9 *Prioritizing decisions*   **99**

10 *Deploying applications*   **103**

11 *Reporting results*   **105**

12 *Administering applications*   **112**

13 Help and accessibility 123

Appendices

A *Sharing streams between IBM SPSS Decision Management and IBM SPSS Modeler* 124

B *Notices* 126

Index 128

About IBM SPSS Decision Management

What's new

IBM® SPSS® Decision Management 6.2 adds the following features:

Support for Cognos data. When used with IBM® SPSS® Modeler Server 14.2 or newer, SPSS Decision Management adds the ability read data from a 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.

ILOG Business Rule Management System integration. Rules created in a Business Rules Management System such as ILOG can be referenced and used in SPSS Decision Management applications, provided they have been developed to support the same data model used in the current SPSS Decision Management project. These external rules can be selected and used in the same manner as other models and rules within the SPSS Decision Management application.

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

About IBM SPSS Decision Management

IBM® SPSS® 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 or simulation to reach the best decision based on the above.

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

Customer interaction example

IBM SPSS Decision Management for Customer Interactions

IBM® SPSS® Decision Management for Customer Interactions decides which promotions to offer customers when they contact your organization, and delivers recommendations to a call center, Web site, or store location in real time. By combining the logic of business rules with the insight gained through predictive modeling, the application identifies the most profitable decision for each customer.

Figure 2-1
IBM SPSS Decision Management for Customer Interactions

Managing customer interactions

For this example, a large bank wants to increase customer retention and overall profits through several overlapping marketing campaigns:

- The retention campaign offers complimentary tickets to selected sporting or theater events, and is limited to high-value customers.
- The cross-sell campaign offers services including credit cards, personal loans, and home-equity loans.

Some customers may be eligible for several offers, while others may be excluded based on payment record or other factors. In the case of overlap, which offer is most profitable? Is profit the only factor, or should other factors be considered in order to balance the interests of different business units or objectives?

Based on the models, rules and other parameters you define, IBM® SPSS® Decision Management for Customer Interactions identifies the best offer for each customer, and delivers recommendations to the call center or Web site in real time.

This example uses the data files *bank_customer_data.txt* and *bank_response_data.txt*, which are distributed along with SPSS Decision Management for Customer Interaction. A completed version of this example is also provided in the file *bank_customer_interaction.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\Modeler\14.2Demos*. Contact your administrator for assistance installing these files if necessary.

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 link in the upper right-hand corner.

Figure 2-2
Applications launchpad
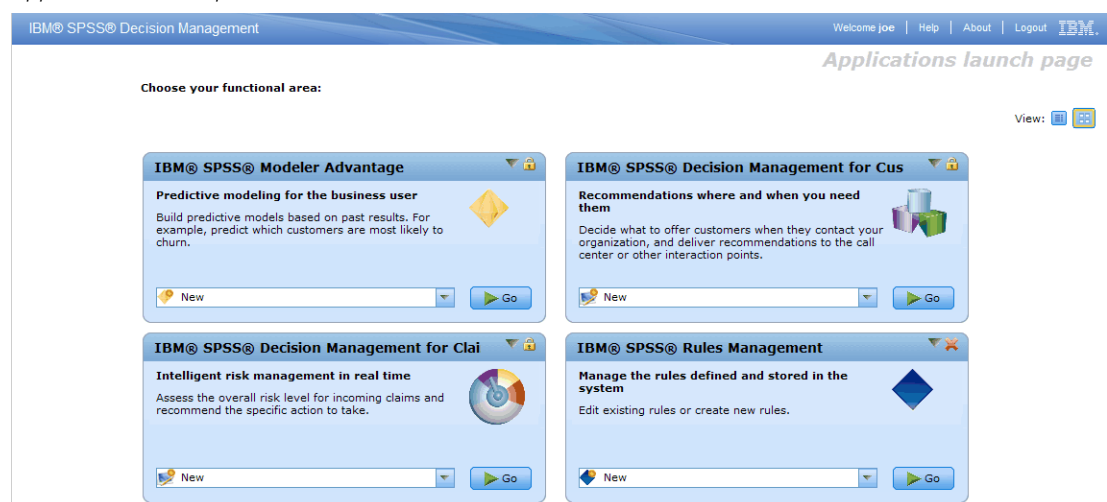
▶ To access the Launch page, navigate to *http://hostname:port/DM*, where *hostname* is the IP address or name of the repository machine and *port* is the application server port number. If you don't know this information, contact your administrator. You can also add a bookmark to this page in your browser, or create a shortcut on your desktop.

▶ Log in with a valid username and password. Contact your administrator for details if necessary.

▶ To create a new project, select New from the drop-down list under IBM® SPSS® Decision Management for Customer Interactions, and click Go.

4

Figure 2-3
IBM SPSS Decision Management for Customer Interactions Home page
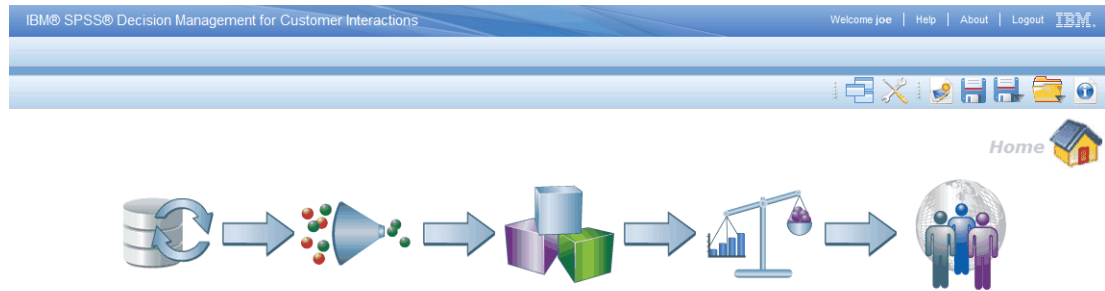
Each icon on the Home page represents a different step in the process, from choosing data through defining, deploying, and reporting on the results. To get started, click the Data icon.

Setting up data sources

The Data tab allows you to configure data sources for the two data files used in the example. Data sets can be configured by your administrator, or you can add your own.

Figure 2-4
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:

■ **bank_customer_data.txt** includes information about current customers, including customer ID, payment history, the number of products they own, the number of months they have been a customer, as well as demographic information like age and income. The project data model is based on this file.

■ **bank_response_data.txt** contains information about customers who have been offered one or more services in previous campaigns, and records their responses. By analyzing this data, the application can identify which customers are most likely to respond to similar offers in future campaigns.

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\Modeler\14.2Demos*. Contact your administrator for assistance installing these files if necessary. For more information, see the topic Sample files in Chapter 12 on p. 121.

To define the project data model

▶ On the Project Data Sources panel, select Add a data source.

Figure 2-5
Customer data source
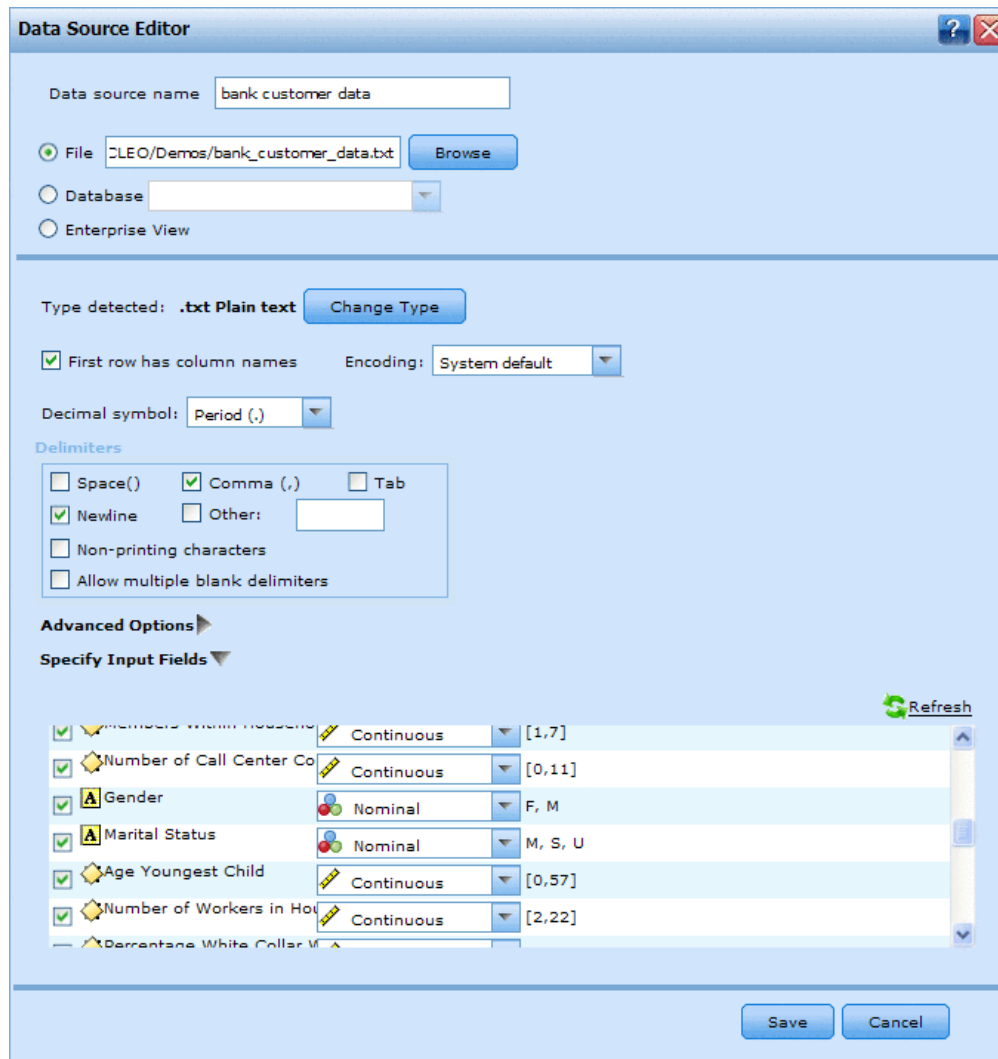
▶ Specify *bank customer data* as the data source name.

▶ Select File, and select *bank_customer_data.txt*.

▶ Under Specify input fields, make sure the measurement type for the *Has Bad Payment Record* field is set to Flag. The flag setting indicates that the field can have two possible values, indicating *true* and *false* (typically coded as *T* and *F,* or *0* and *1* in a numeric field), to indicate the presence or absence of that attribute.

▶ Under Specify input fields, make sure the measurement type for the *Gender* field is set to Nominal. (The values of this field simply indicate gender as male or female, and not *true/false* as would be the case with a flag.)

▶ Select Save.

Since this is the first data source defined for the project, it is automatically selected for the project data model, and used to define the fields required by the application. All other data sets are mapped relative to this source.

To define the response data source

The response data set contains many of the same fields used in the project data model, as well as two additional fields named *Campaign ID* and *Response*. These fields indicate which offer was made to each customer, and whether or not they responded, and can be used to build a model that predicts which customers are most likely to respond in future.

▶ On the Project Data Sources panel, select Add a data source.

Figure 2-6
Response data source

▶ Specify *bank response data* as the data source name.

▶ Select File, and select *bank_response_data.txt*.

▶ Under Specify input fields, make sure the measurement type for the *Response* field is set to Flag.

▶ Under Specify input fields, deselect the check box next to the *Campaign ID* field to indicate this field should not be used as an input when modeling. (This field can be used to indicate which campaigns a customer was selected for in the past, but is not useful in predicting future response.)

▶ Select Save.

Because some of the fields in the data model are not found in the response data set, an orange icon is displayed for this data source under the Compatible column on the Data tab, indicating that some fields match, but not all. The data set is suitable for modeling, but not for simulation, scoring, or deployment. (To be used for these purposes, all fields from the data model need to be present.)

Global selections

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.

Figure 2-7
Global selections
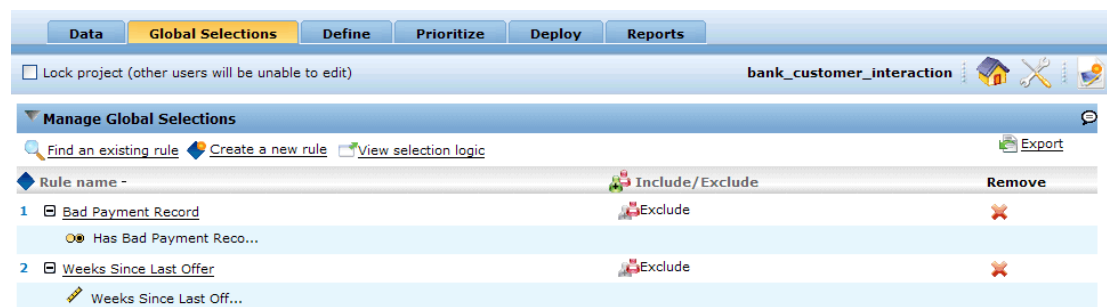
▶ On the Global Selections tab, select Create a new rule.

▶ For the rule name, specify *Bad Payment Record*.

Figure 2-8
Creating a rule
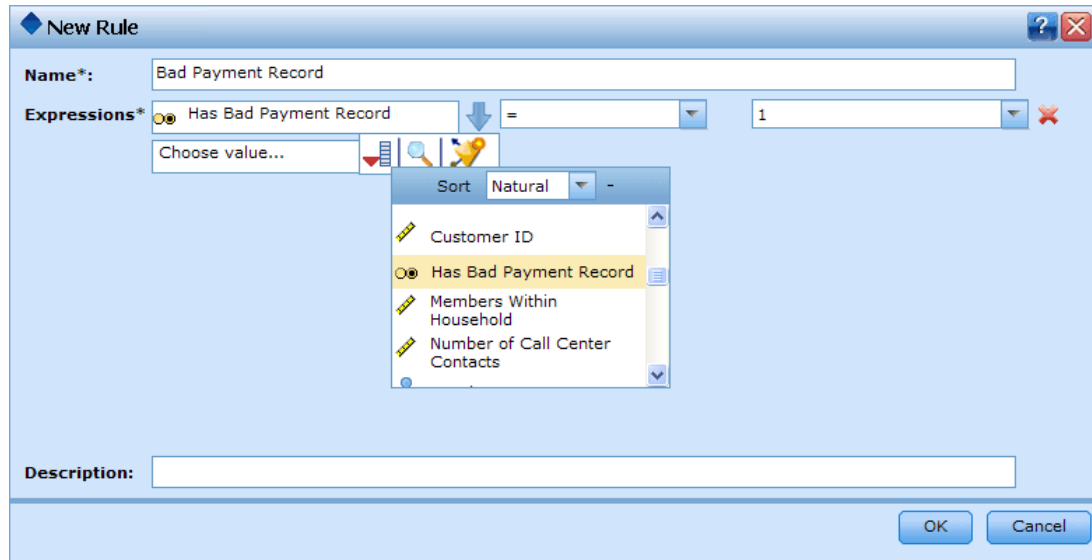
▶ For the expression, choose *Has Bad Payment Record* as the input field. (Click the down arrow to choose between field, rule, and model inputs, then choose the desired field.)

▶ To complete the expression, specify 1 as the value to match (where 1 is the "true" value).

▶ On the Global Selections tab, click the Include/Exclude icon to change the state to *Exclude*.

▶ Add an additional rule to exclude customers if *Weeks Since Last Offer < 8*.

Defining campaigns and offers

The Define tab allows you to specify the available campaigns and offers, and the rules used to allocate customers accordingly.

Figure 2-9
Define tab showing properties for the cross-sell campaign
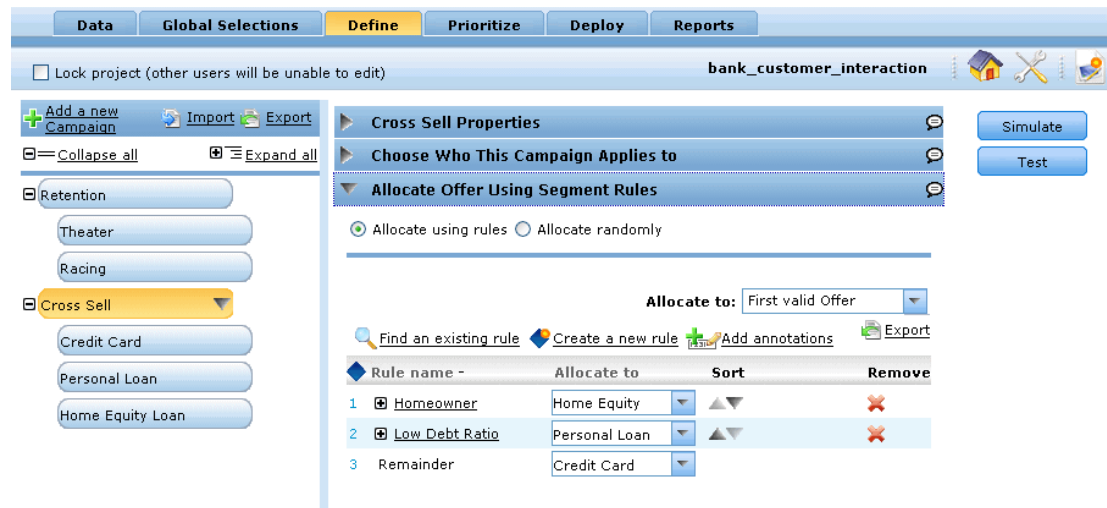
- The dimension tree in the left pane specifies the range of possible offers that can be returned, organized by campaign.

- 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.

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*).

► Change the name of *My Campaign* to *Retention*. (With the mouse pointer on *My Campaign*, click the arrow icon to display the context menu, and choose Rename.)

Figure 2-10
Renaming a campaign
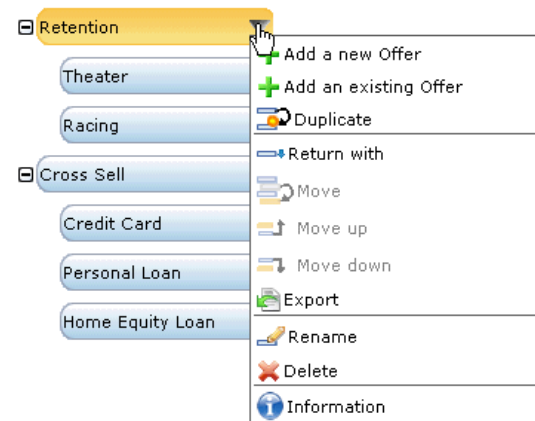
▶ Under the *Retention* campaign, add offers for *Theater* and *Racing.* (Select Add a New Offer from the context menu for the campaign.)

▶ Add a second campaign named *Cross Sell*, with offers for *Credit Card*, *Personal Loan*, and *Home Equity Loan.*

▶ With the *Retention* dimension selected in the left pane, click on Retention Properties in the right pane.

Figure 2-11
Retention campaign properties

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

▶ Under Choose Who This Campaign Applies to, add a selection rule named *High Value Customers*. This rule should have two conditions, to select customers with *Months Current Account > 12* and *Number of Products > 3*.

▶ Under the Include/Exclude column, make sure the Include icon is displayed.

Because both conditions are specified in the same rule, customers will only be selected for the campaign if both conditions are met. (Alternatively, you could specify these conditions in separate rules, and customers would be included if either condition was met. When in doubt, click View selection logic in the rule editor for details.)

▶ Under Allocate Offer Using Segment Rules, create rules to allocate males (*Gender = M*) to the *Racing* offer, and allocate females to the *Theater* offer.

▶ For the *Cross Sell* campaign, specify properties in the same manner as for *Retention*, with the comments below.

Figure 2-12
Cross sell campaign properties

▶ Specify a current active date as needed.

▶ Under Choose Who This Campaign Applies to, do not add any selection rules. All customers are eligible for this campaign (unless excluded by global selections).

▶ Under Allocate to, specify rules to allocate customers to offers as follows:

■ If *Homeowner = True* then *Home Equity Loan*.

■ If *Personal Debt to Equity Ratio* < 30 then *Personal Loan*.

■ For the remainder then *Credit Card*.

▶ Under Allocate to, make sure that First valid offer is selected. Thus a customer who is eligible for both the home equity and personal loan offers will receive the home equity offer because rules are evaluated in the order listed.

Running a simulation

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

▶ From the Define tab, select Simulate.

▶ Select *bank customer data* for the simulation data source.

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

▶ Select Run.

Looking at the results, the Campaigns tab shows that out of 2000 customers in the simulation data set, roughly 1800 will be allocated to the *Cross Sell* campaign, and just under 200 will be allocated to the *Retention* campaign.

Figure 2-13
Simulation results showing campaign allocations
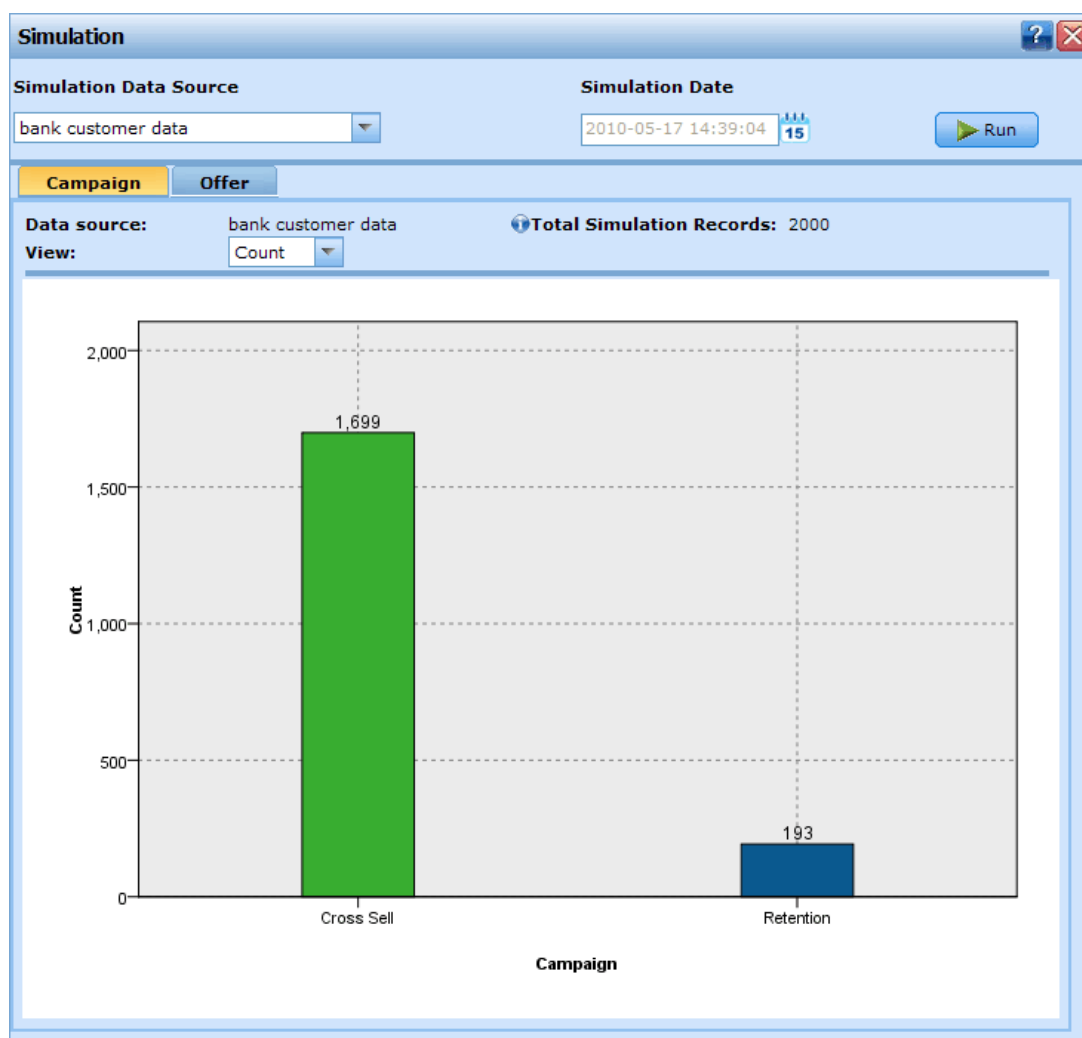
The *Offer* tab shows a similar breakdown by offer across both campaigns.

Figure 2-14
Simulation by rule
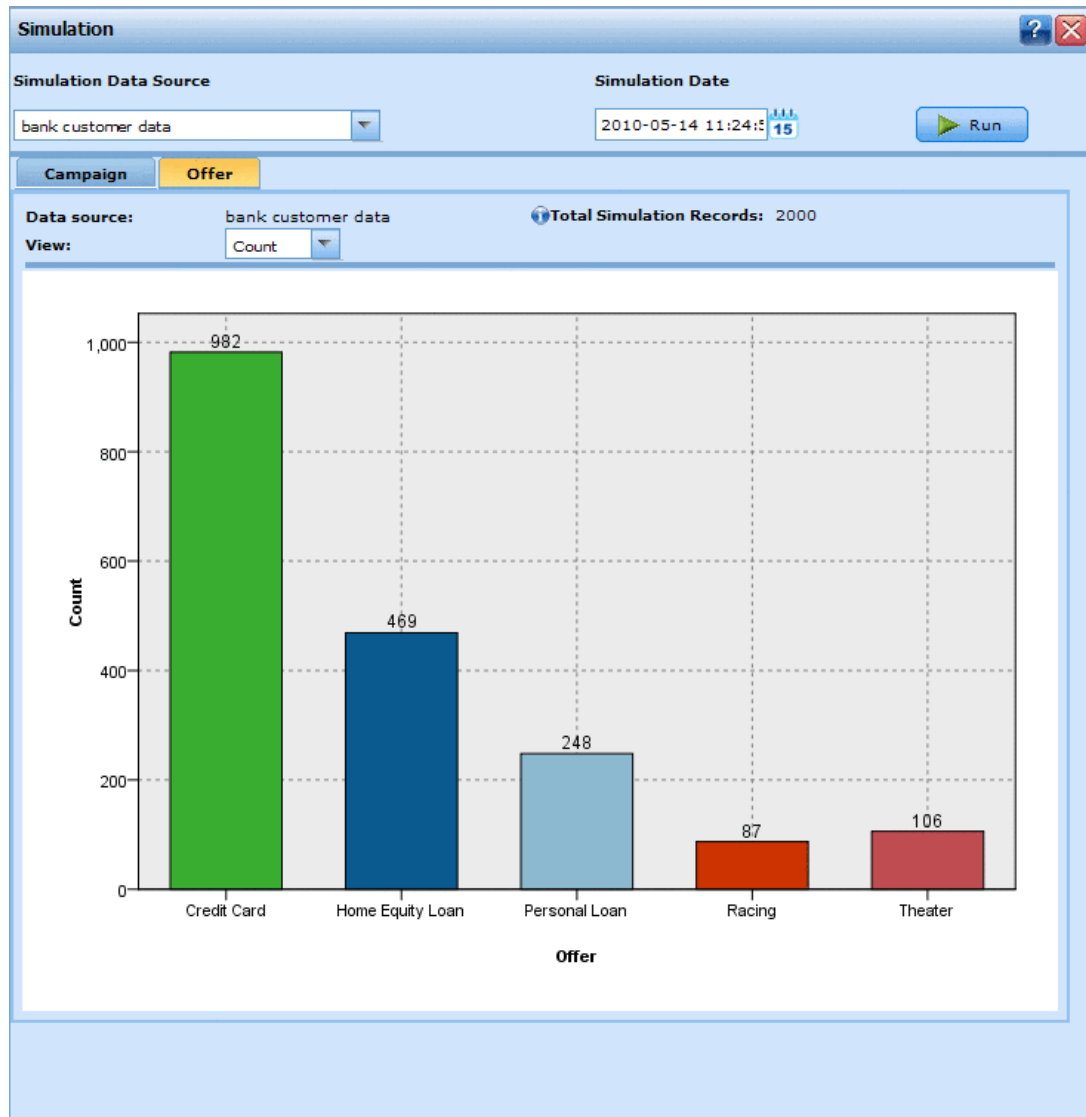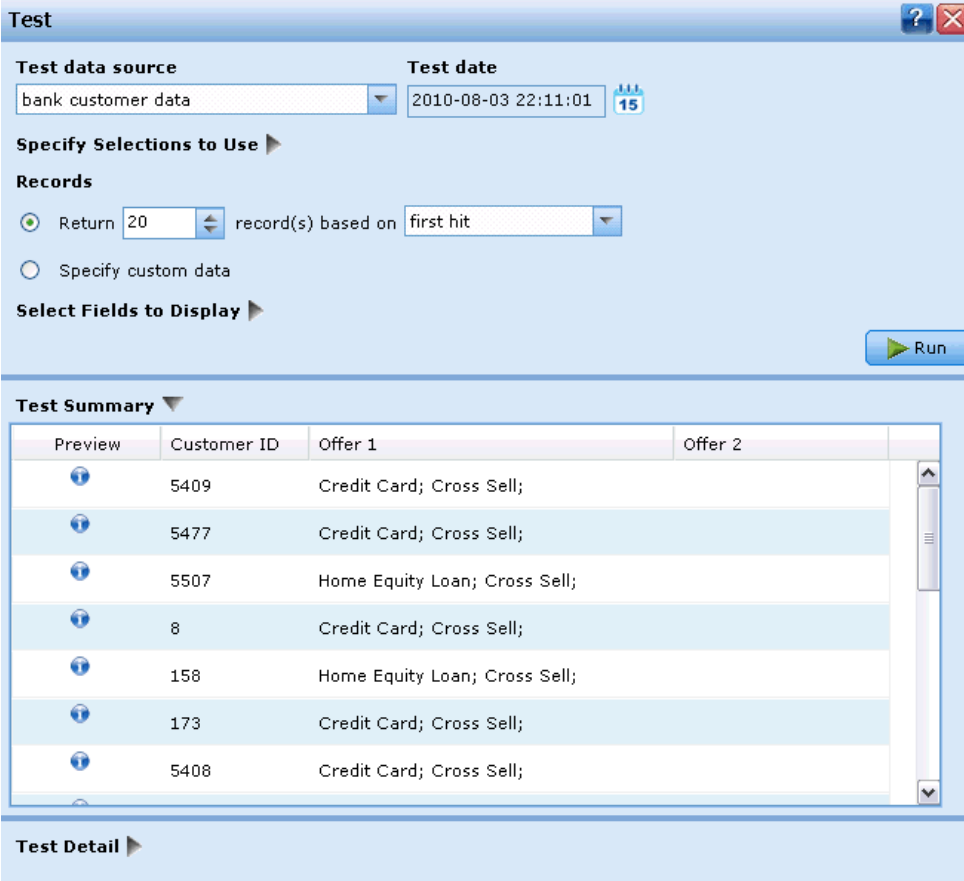
Based on the simulation results, you may choose to adjust your selections or allocations to push more or fewer records to one campaign or offer.

Running a test

To see how specific records are processed by the application, select Test from the Define tab.

▶ For the data source, select bank customer data.

▶ Specify 20 as the number of records to return, select one or more fields to display if desired, and click Run.

Figure 2-15
Test summary

The test summary shows the values of the selected fields for each record, along with the offers each customer is eligible for. Most customers are only candidates for one offer, but a handful qualify for more than one (though you may have to scroll down through the results to see them).

▶ To see details on a specific record, click on it and select Test Detail.

Figure 2-16
Test details

From the details you can see how the customer was evaluated at each stage. For example, customer 5803 qualifies for the credit card offer, but was excluded from the retention campaign at the Define step. You can also click the information icon to see further details on the that record.

Figure 2-17
Additional test details

Prioritizing offers to reach a decision

The selections and 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 Prioritize tab is for.

Figure 2-18
Prioritize tab

The Prioritize determines the most profitable offer for each customer, based on the following equation.

```
Expected profit = ( [probability to respond] * [income] ) - [cost:]
```

Where:

- Probability is a customer's propensity to respond to a particular offer.
- Income is the value or contribution expected from that response .
- Cost is the cost of making the offer.

Prioritization inputs

▶ To input parameters for each offer, click Customize table in the upper right corner of the Prioritize tab.

Figure 2-19
Customizing the table

▶ Select the Offer button for all parameters, and click Save. This allows you to specify inputs separately for each offer, rather than globally for each campaign.

▶ Specify inputs as shown in the above figure, with the following comments:

- For the retention campaign, the probability to respond is 0.1, based on results from previous campaigns. The revenue number of 180 is based on the average value of a customer in a typical year. The costs are high for this campaign, because you are providing complimentary tickets in return for loyalty.

- For the cross-sell campaign, the probability to respond is 0.04, again based on previous campaigns. The cost and revenue values depend on the service being offered.

WhatIf? analysis

The WhatIf? dialog box allows you to experiment with different combinations and see how changes impact your bottom line. To try a new combination, edit the input values, type a name, and click Run. Each run adds a new column in the results table at lower left, allowing a side-by-side comparison.

▶ To simulate your results based on the current inputs, click WhatIf?... on the Prioritize tab.

Figure 2-20
WhatIf? dialog box

▶ In the WhatIf? dialog box, type Run1 as the simulation name, and click Run.

Based on the results, you can see that cross-sell offers will be made to 469 customers, and the retention campaign will not be made to any customers.

Retention campaign

With the initial values entered, the expected profit for this campaign will always be negative:

```
Expected profit = (0.1 * 180) - 21 = -3
```

To make the campaign profitable, you need to either improve the response rate, or increase the revenue. One way to do this is to target the most profitable customers based on annual value.

▶ Under the Revenue column, change the input for both retention offers to use the *Annual value* field as an input. (Click the down arrow icon, and select the field icon from the toolbar.)

Figure 2-21
Prioritization parameters

Campaign/Offer	Prob.to Respond	Min.Profit	Revenue	Cost
⊟Retention				
Theater	0.1	10	✎ Annual va	22
Racing	0.1	10	180	21
⊟Cross Sell				

This value is computed differently for each customer using a formula that factors in the number of accounts they own, number of transactions, and other factors, allowing you to identify and target only the most profitable customers.

Cross-sell campaign

For the cross-sell campaign, you will build a model to predict which customers are most likely to respond.

▶ Under the *Prob. to Respond* column, click the down arrow icon next to the *Credit Card* offer, and select the icon to build a new model.

Figure 2-22
Using a model in the prioritization

⊟Retention				
Theater	0.1	10	✎ Annual va	22
Racing	0.1	10	180	21
⊟Cross Sell				
Credit Card	0.04	10	60	3
Personal Loan		10	200	3

Figure 2-23
Model tab
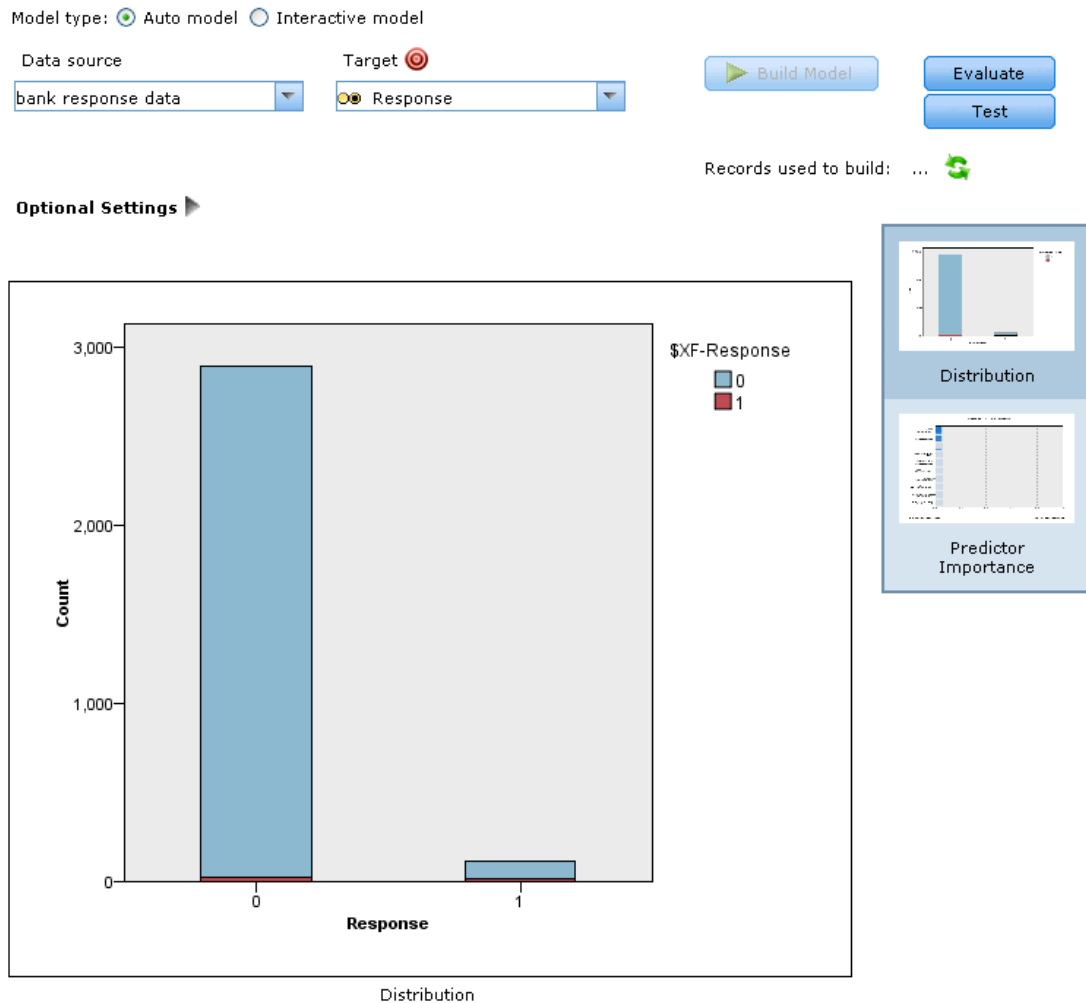
Distribution

▶ Select *bank response data* for the data source, and select *Response* as the target field.

▶ Select Build Model.

When the model completes, a distribution chart is displayed. The chart shows observed values on the horizontal axis, with a value of 1 indicating a positive response.

- The height of the bars reflects the overall response rate. Out of 3000 customers in the response data set, 111 responded, for an overall hit rate of about 3.8%.

- The colored areas within each bar represent the responses predicted by the model, with red indicating a positive prediction, and blue indicating a negative response, as shown in the legend in the upper right corner. Due to the low overall response rate, there are relatively few predicted responses (red areas). However the false bar (*Response = 0*) is overwhelmingly blue, indicating that the overwhelming majority of those predicted not to respond actually did not.

▶ Save the completed model as *bank_response_model.str*.

▶ Click Use Model.

▶ When prompted to choose an input field, select Propensity.

▶ Use the same model propensity field for the other two offers in the cross sell campaign (*Personal Loan* and *Home Equity Loan*). (To do this, click the down arrow next to each offer and choose the icon to find an existing model, and browse to the model you just saved.)

▶ Click Run to repeat the simulation. The result now shows a mix of offers across the two campaigns.

Figure 2-24
WhatIf? dialog box

Deploying the application

After you update your models and rules and are satisfied with the simulation results, you can apply a label to indicate this is the version of the application you want to put into production.

Figure 2-25
Deploy tab
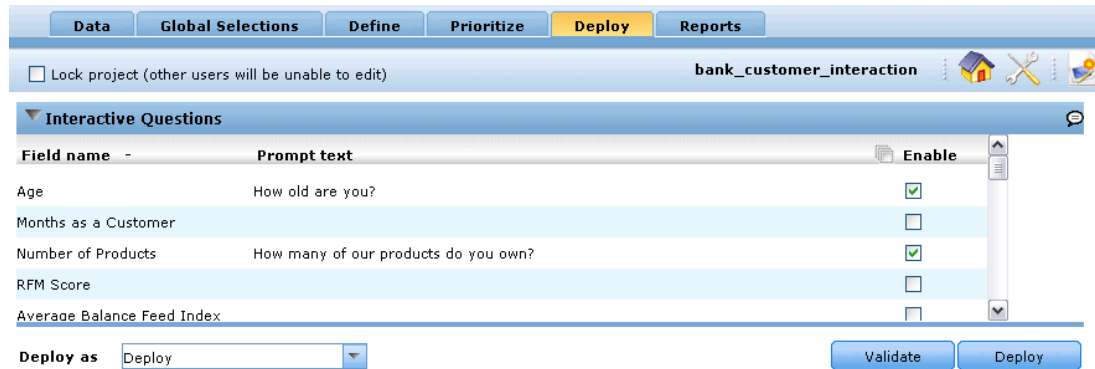
▶ On the Deploy tab, select the label you want to apply. Depending on your process, you may have a testing or preproduction phase that precedes the actual deployment.

▶ Click Validate to check for errors in the project, such as references to rules or models that cannot be resolved. This can happen, for example, if a label is moved or updated by another user.

▶ Optionally, under Interactive Questions, you can specify the text used to prompt for any data values that may be missing after the application is deployed. For example, this can provide your call center employees with the right questions to ask when data is missing. To add text for a particular field, select the check box at right to enable that field.

▶ To apply the label, click Deploy.

Once the label is applied, a services engagement is typically required to integrate the application with your call center, web site, or other IT infrastructure. For more information, see the *Application Designer's Guide* and contact your sales or support representative.

Navigating the workspace

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.

Figure 3-1
Applications launch page

▶ 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.

▶ 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 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.

Figure 3-2
Adding a custom shortcut
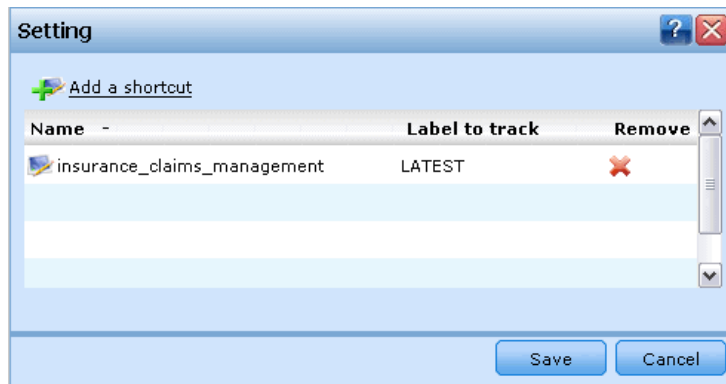
Other ways to open an application

- 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.)

Figure 3-3
Link to add an application
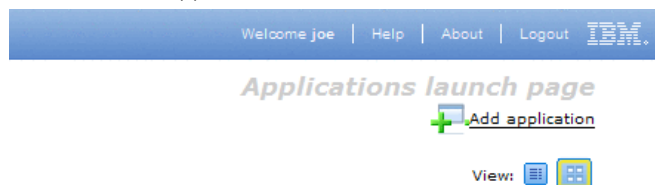
A list of available applications is displayed. Optionally, you can create shortcuts to one or more labeled versions.

Figure 3-4
Adding a new application to the launch page

▶ 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.

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.

Figure 3-5
Home page
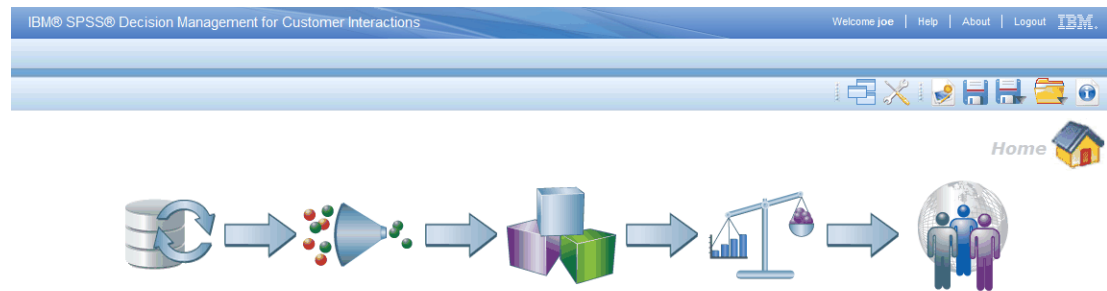
Data. Define data sets for use by the application. Operational data pertains to current production — such as current customers, or people you want to sell to in future — while models are based on past data, such as results from previous campaigns.

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. When multiple outcomes or recommendations are returned on the Define tab, choose the best based on the specified prioritization 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.

Report. Monitor the status of deployed applications.

Tracking step completion. While working through the application, you can mark major sections as done. These sections are highlighted with a green check mark for easy reference. Optionally, your administrator may complete one or more sections for you (such as data sources).

Gallery

The gallery may be displayed on the Applications Launch Page, as well as each application home page, depending on how your administrator has set up your system. It list lists unsaved work, such as a model that has been left to build overnight—you can 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.

Figure 3-6
Gallery on Applications Launch Page
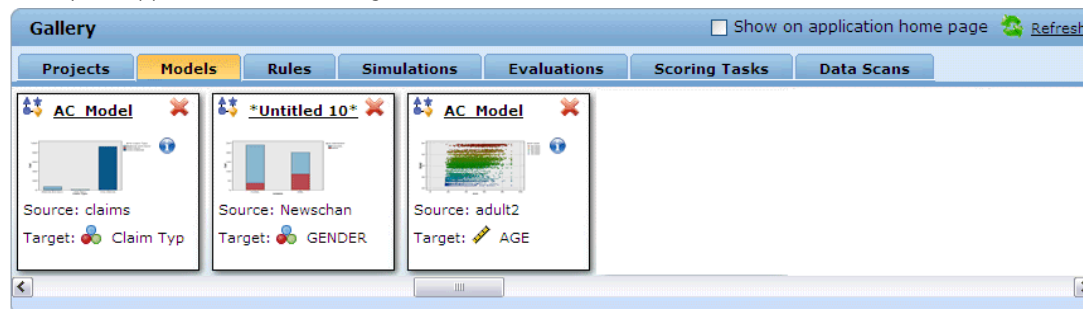
If the name of the item is underlined, click the name to open the relevant item. For example, if you have an unsaved data scan, click on the data scan title in the gallery to open the Data tab.

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.

Setting preferences

There are a number of preferences you can adjust to customize IBM® SPSS® 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 that you use.

To access the User Preferences settings dialog, from within any application, click the User Preferences icon on the toolbar.

Figure 3-7
User preferences icon

Figure 3-8
User Preferences settings dialog

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

User settings

Figure 3-9
User Settings

General settings

Figure 3-10
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.

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.

Always prompt before deleting. Select this to override any deletion prompts that have been removed within individual applications.

Model settings

Figure 3-11
Model Settings
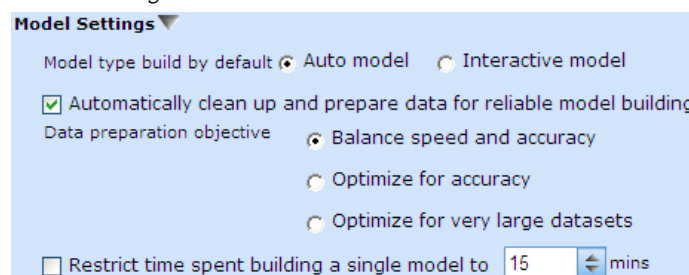
Model type build by default. Specify whether by default you want the application to build the best model automatically from the available data, or if you want to fine tune the building options each time a model is built.

Automatically clean up. Selecting this option enables you to let the application do a lot of the data preparation. For example, repairing data issues, screening less useful data, or deriving new attributes.

You can choose whether the clean up operation should give an equal balance to both speed and accuracy, give priority to accuracy over speed, or give priority to speed such as when you are working with very large datasets or are looking for a quick result.

Restrict time spent building. Use this option to specify the default amount of time to spend building each model. This can be especially useful when processing large datasets, and can have a significant impact on the time needed for modeling building to complete.

Test and simulation settings

Figure 3-12
Test and Simulation Settings
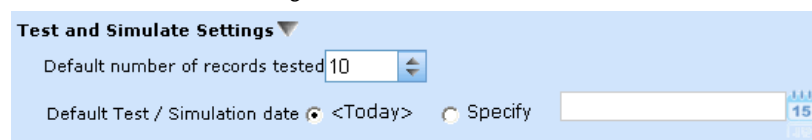
Default number of records tested. Specify the default number of records to include when running a test.

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

Define step settings

Figure 3-13
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 step settings

Figure 3-14
Prioritize Step Settings

Prioritization 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 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 settings to each new interaction point that you set up, or use the default values created by your system administrator.

Rules settings

Figure 3-15

Aggregation. When using aggregation rules, specify whether by default all records have the value of the Remainder value points added to them or if the points are only added to those records to which no rules apply.

Allocation. When using allocation rules, specify whether the allocation type is established by using rules or just applied randomly. In addition, you can select whether rules-based allocation allocates records to the first valid option or all valid allocations.

Saving projects, models, and rules

You can save your work to the IBM® SPSS® Collaboration and Deployment Services Repository, or download to the local file system. 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, model, or rule. The specific type of object depends on the application.

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

Saves the current project, model, or rule 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, model, or rule from the IBM SPSS Collaboration and Deployment Services Repository and/or local file system as applicable.

Saving to IBM SPSS Collaboration and Deployment Services Repository

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. Support for IBM® SPSS® Collaboration and Deployment Services is required to access these features.

Figure 3-16
Save As dialog box
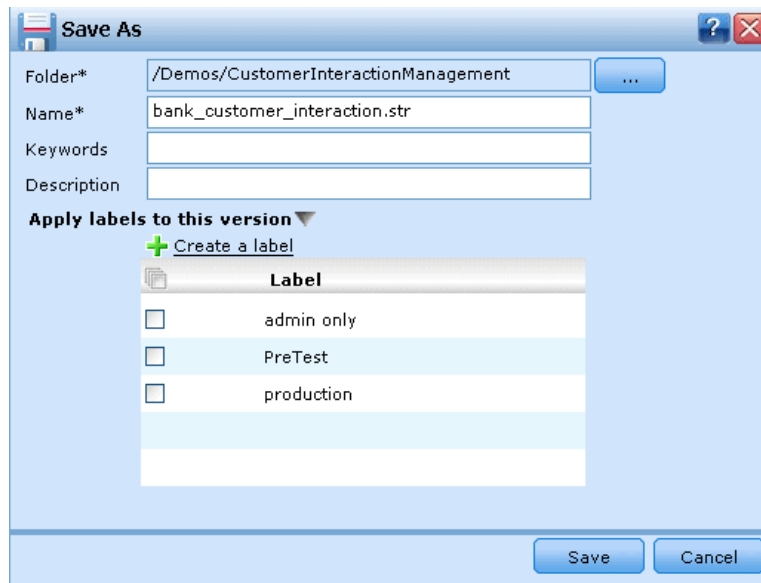
Folder. Choose the folder to save to (see below).

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".

Apply labels to this version. 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.

Choosing a folder

Figure 3-17
Choosing the folder to save to
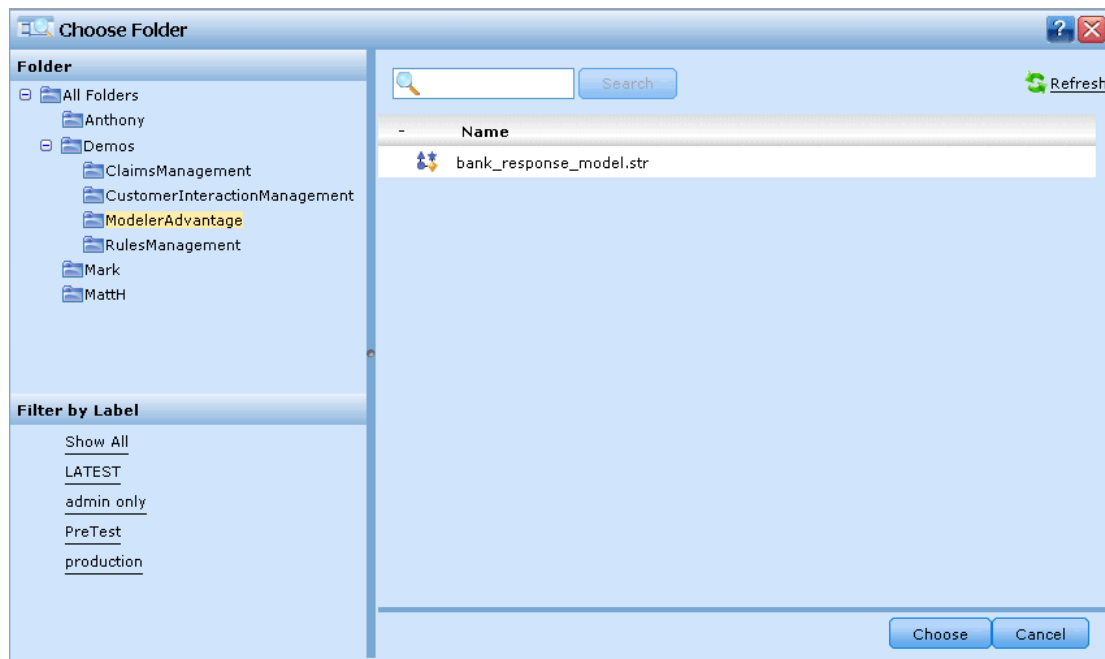
The tree view in the left pane displays folders in the IBM SPSS Collaboration and Deployment Services Repository you have authority to see. Navigate the folder tree or perform a search to find what you're looking for. You can also filter the file list by labels available in the lower-left pane. For example, you might click a *production* label to only display files marked with that label.

To see more details about a file, click the information icon beside it. A properties panel is displayed, with details such as keywords, labels, author, and date.

Folder. Displays available folders and allows you to navigate the hierarchy.

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

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.

Figure 3-18
Finding an existing rule
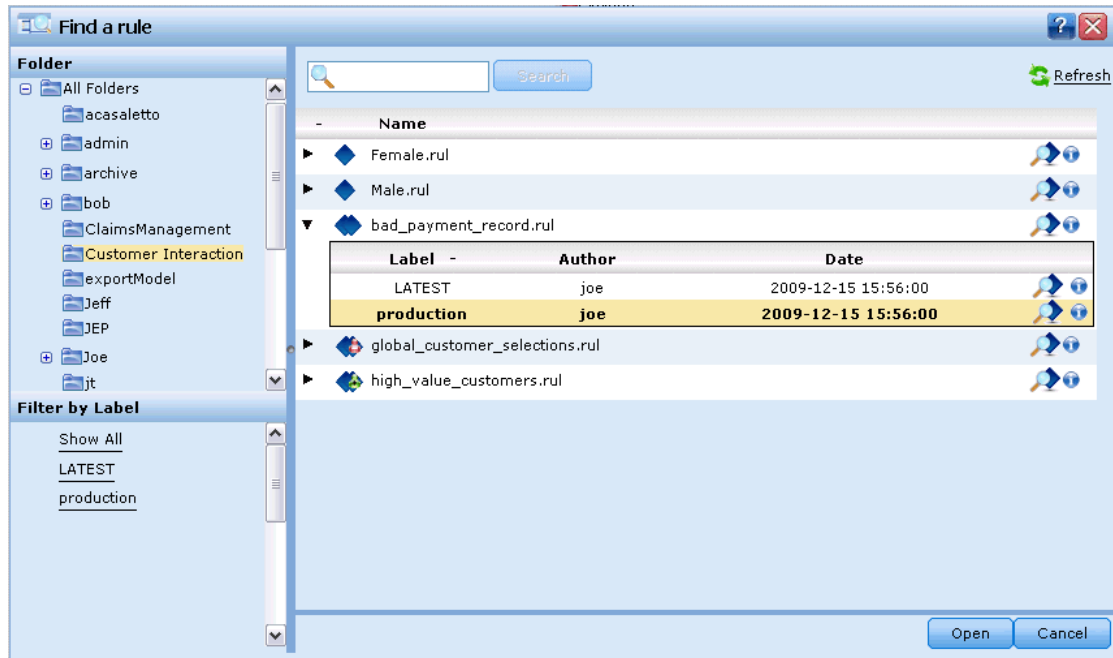
When referencing shared objects, using a labeled version is strongly recommended, particularly in a production environment, as any changes to the latest version may otherwise have unintended consequences. In certain cases, such as when referencing a rule from within a rule, a labeled version must be used. For more information, see the topic Exporting and reusing rules in Chapter 6 on p. 59.

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. For more information, see the topic Saving projects, models, and rules on p. 31.

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

Test & Deploy

DEPLOY:PHASE1

Locking projects

When you open a project, model, or rule for editing, the object is locked in the IBM® SPSS® Collaboration and Deployment Services Repository so that other users cannot edit it. This is indicated by a check box in the upper right corner of the workspace. Alternatively, you can deselect this check box to allow others to edit the object.

Figure 3-19
Locking a workspace
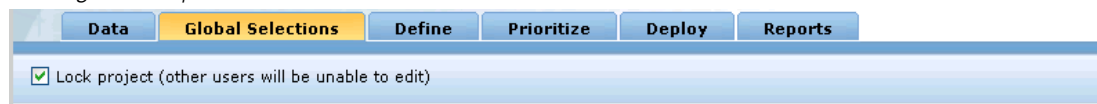
- When a project is locked, other users can open it for viewing, but cannot save changes.

- 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 browse the workspace, but you can no longer save changes.

Object properties

When editing a ruleset, model, or project, you can add keywords and a description to help identify the object when browsing for it at a later date. To display the Edit Properties dialog, click the Properties icon in any application.

Figure 3-20
Properties icon

Figure 3-21
*Properties for a model, rule, or project*
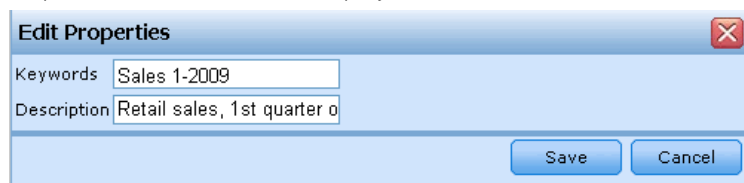
Keywords. Enter any keywords to quickly identify the saved object when searching for it. For example these might be "Sales" and "1-2009"

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

Managing data sources

The Data tab enables you 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.

Figure 4-1
Data tab
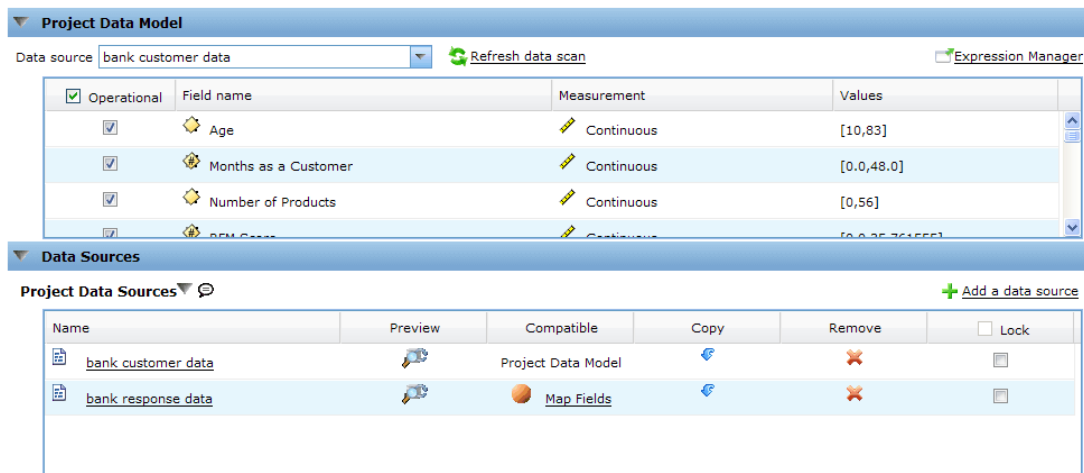
The tab is divided into three main areas:

■ **Project Data Model.** Defines the fields required by the application. All other data sets are mapped relative to this source.

■ **Project Data Sources.** Lists the data sources that have been saved for use with the current project, model, or rule.

■ **My Data Sources.** Lists data sources you have defined or saved. Data sets from this list can be copied to or from different projects, making it possible to share data across applications.

Project data model

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

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. For more information, see the topic Measurement levels on p. 39.

Value. 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.

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.

Preview. Enables you to preview a sample of the data contained in the source. For more information, see the topic Previewing data on p. 38.

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 with the project data model the data source is.

- 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 the topic Mapping fields on p. 52.

Copy. Enables you to copy 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 has given 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.

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 the topic Creating a new data source on p. 40.

▶ 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 the topic Measurement levels on p. 39.

▶ 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 have been mapped, the link is no longer displayed.) For more information, see the topic Mapping fields on p. 52.

▶ To preview a data source, click the Preview icon. For more information, see the topic Previewing data on p. 38.

Previewing data

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

Figure 4-2
Data preview

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.

The following measurement levels are available:

■ **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*.

■ **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. In IBM® SPSS® Decision Management, the first value from the category list will be used as "false" value and the second will be the "true" value, as indicated by a tool tip when hovering over the values when displayed on the Data tab. 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.

■ **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).

Creating a new data source

Figure 4-3
Data Source Editor
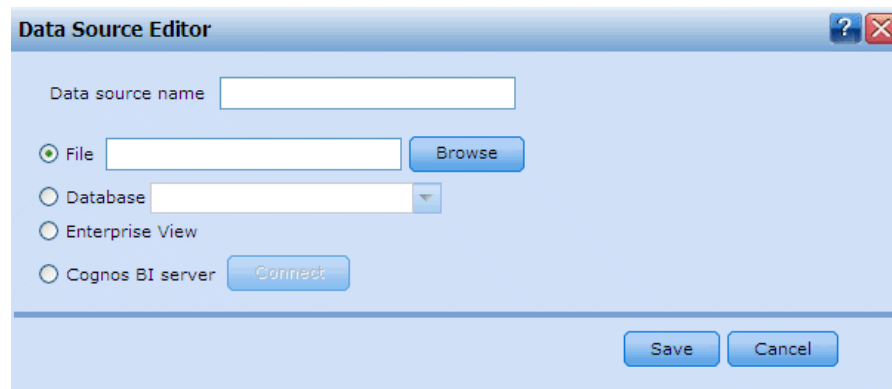
▶ To access the Data Source Editor, select Add a data source under either the Project Data Sources or My Data Sources area.

▶ Type a name for the data source.

▶ Select the source type:

■ **File.** Supported file types include text (**.txt*, **.csv*), Excel, and the **.sav* format used by IBM® SPSS® Statistics.

■ **Database.** Any database that supports ODBC, including Microsoft SQL Server, DB2, Oracle, and others.

■ **IBM SPSS Collaboration and Deployment Services Enterprise View.** Creates a connection to an IBM® SPSS® Collaboration and Deployment Services Enterprise View as defined in IBM® SPSS® Collaboration and Deployment Services. For more information, see the topic IBM SPSS Collaboration and Deployment Services Enterprise View source on p. 46.

■ **Cognos BI server.** Any database that supports Cognos BI server, version 14.1 or lower. For more information, see the topic Cognos BI source on p. 47.

▶ Specify options as appropriate for the source type.

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.

Text-based data files

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

Figure 4-4
Text-based data source settings

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.

■ **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

Figure 4-5
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.

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.

Specify input fields. For more information, see the topic Selecting input fields on p. 50.

Excel data files

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

Figure 4-6
Excel data source settings

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.

IBM SPSS Statistics data files

When you select a IBM® SPSS® Statistics data source, you are prompted to enter further details.

Figure 4-7
IBM SPSS Statistics data source settings

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

Database source

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

Figure 4-8
Database data source settings

Test connection. Enter your database username and password and click to check that you have access to the database.

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 Never if you *never* want table and field names quoted.

- Select Always if you want *all* table and field names quoted.

Specify input fields. For more information, see the topic Selecting input fields on p. 50.

IBM SPSS Collaboration and Deployment Services Enterprise View source

When you select an IBM® SPSS® Collaboration and Deployment Services Enterprise View source, you are prompted to enter further details.

Figure 4-9
IBM SPSS Collaboration and Deployment Services Enterprise View data source settings

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 the adjacent Browse button to search for other Application Views in the repository.

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

Data provider. The drop-down list displays the names of up to ten Data Provider Definitions (DPDs) for the selected Application View. Only DPDs that reference the selected Application View are shown.

Environment. The environment value, which is set in IBM® SPSS® Collaboration and Deployment Services, determines which DPDs are available.

Specify input fields. For more information, see the topic Selecting input fields on p. 50.

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.

Cognos BI source

Note: If the IBM® SPSS® Modeler Server in use with IBM® SPSS® Decision Management does not support Cognos BI server, version 14.1 or lower, then this option is not available.

Figure 4-10
Data Source Editor - Cognos BI server selected

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

Figure 4-11
Cognos server selection

▶ Type the Server URL of the Cognos server from which to import or export data. Contact your Cognos system administrator if you are not sure which URL to use.

▶ Select the Mode in which you will connect.

▶ To log in as a specific user (for example, as an administrator), select Set Credentialsand enter the Cognos namespace, username, and password:

- Specify the 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 Cognos User name with which to log on to the server.

- Enter the Password associated with the specified user name.

▶ Alternatively, to log in with no user credentials, in which case you can not fill in the namespace, username, and password fields, select Use Anonymous connection.

Note: Some server connections do not allow Anonymous connections.

Selecting Cognos package details

When you have set up a Cognos server connection, you first select a package from which to import data and then select the query subjects or individual query items that you want to import from the selected package.

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

To select a package, click Browse, the Choose Package dialog box is displayed and lists all available packages from which you can import content.

Figure 4-12
Cognos package selection
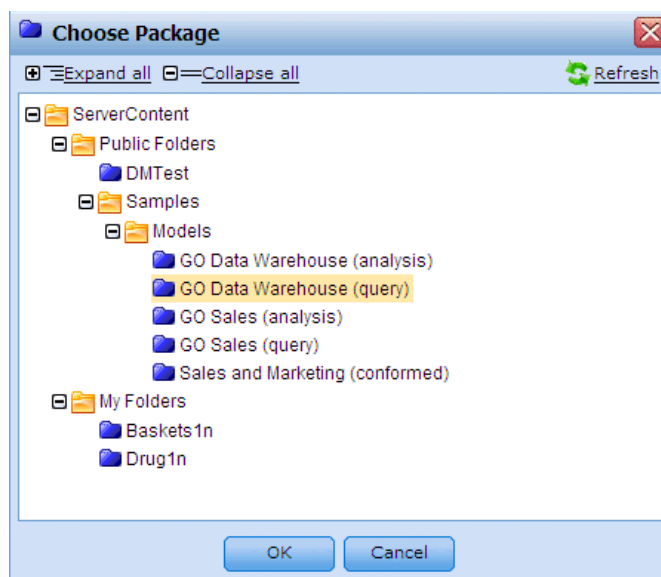
Where necessary, expand the list to find the package you require; click on the package name to select it, and click OK. The package details are shown in the Data Source Editor.

After you have selected a package you then choose the package object query subjects (which represent database tables) or individual field query items (which represent table columns) that you want to use from the package. *Note*: The data must be in UTF-8 format.

Figure 4-13
Cognos input field selection
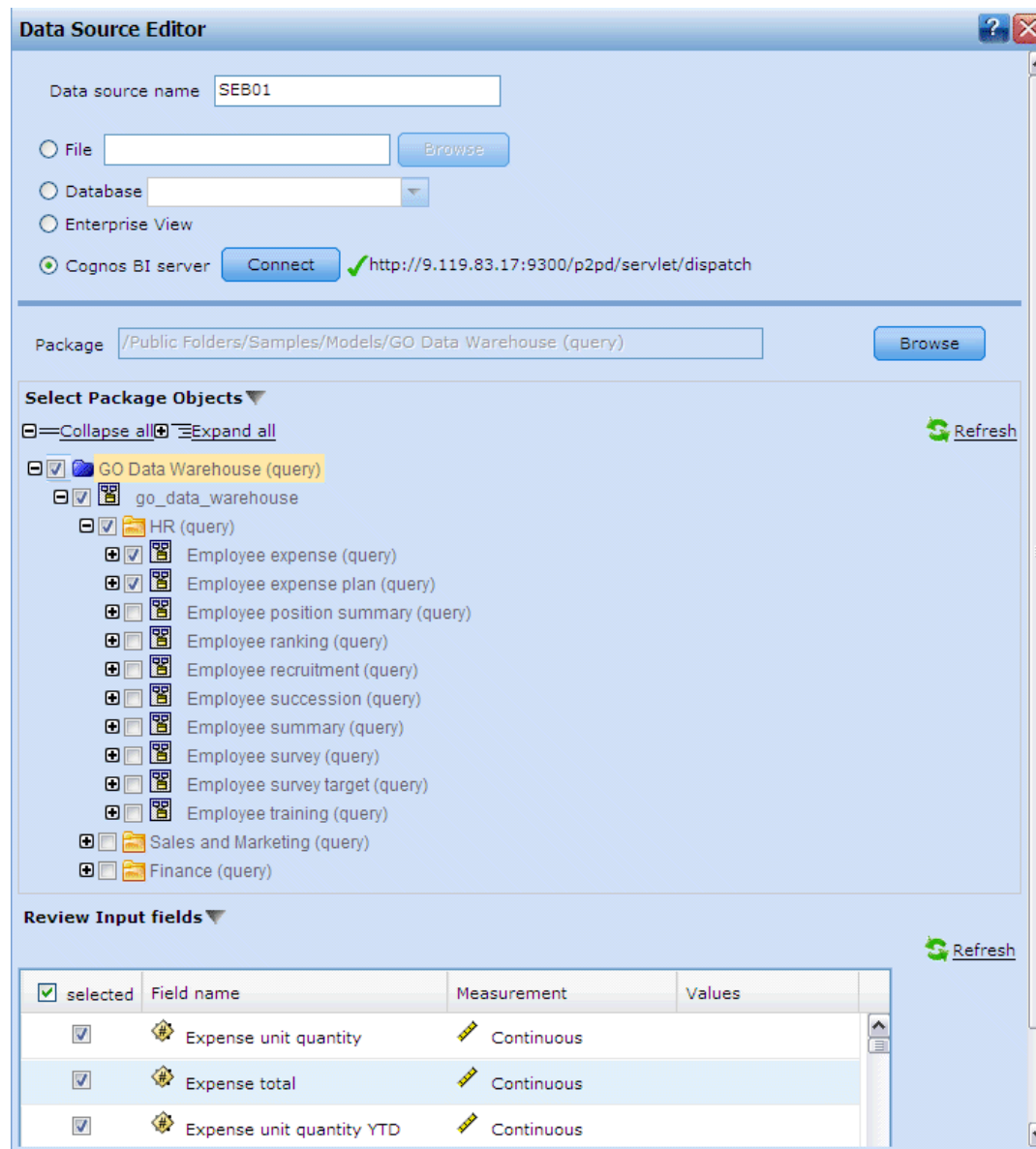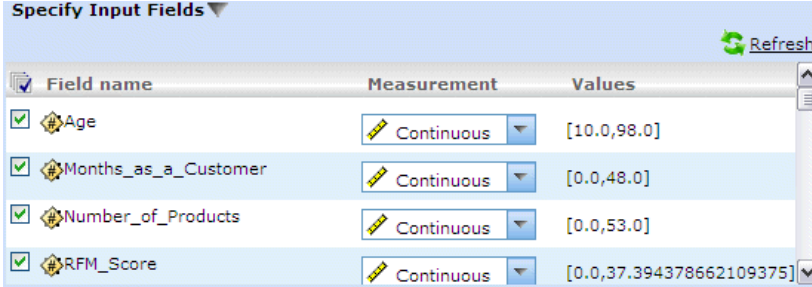
Select Package Objects. Displays the name of the selected package, together with the package objects (also known as namespaces) associated with the package that you can select. Expand one or more package objects to display the input fields that you can import. Selecting a package object automatically selects all of its input fields.

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

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.

Figure 4-14
Data input fields

▶ 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 that you have defined are also displayed.

▶ Select each of the fields and expressions that 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.

Expression manager

You can use the expression manager 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 always numeric with a measurement of Continuous; this can not be changed.

Note: Expressions are available throughout the application in the same was as operational data; for example, in rules and model predictors.

Figure 4-15
Expression manager
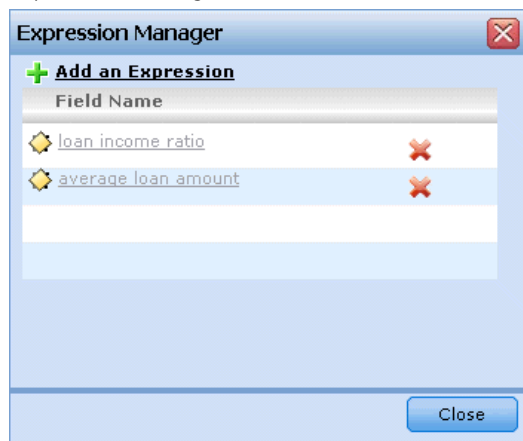
▶ To display existing expressions, click Expression Manager on either the Data tab or the Data Source Editor.

Creating an expression

Figure 4-16
Creating a new expression
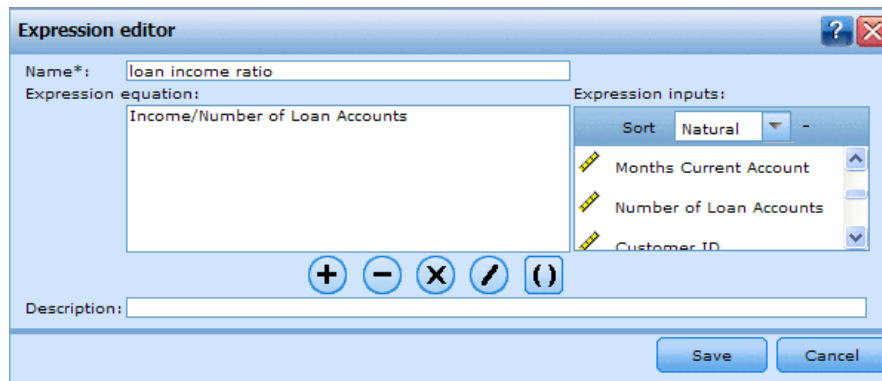
▶ In the Expression Manager dialog, click Add an Expression; the Expression Editor dialog is displayed, showing all available inputs from the data source.

▶ Enter a unique Name for the expression.

▶ Either double-click or drag the required inputs and mathematical symbols to the Expression Equation field to build your expression.

▶ Enter a Description that explains what the expression is used for.

▶ Click Save to return to the Expression Manager dialog.

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.

Figure 4-17
Mapping data fields
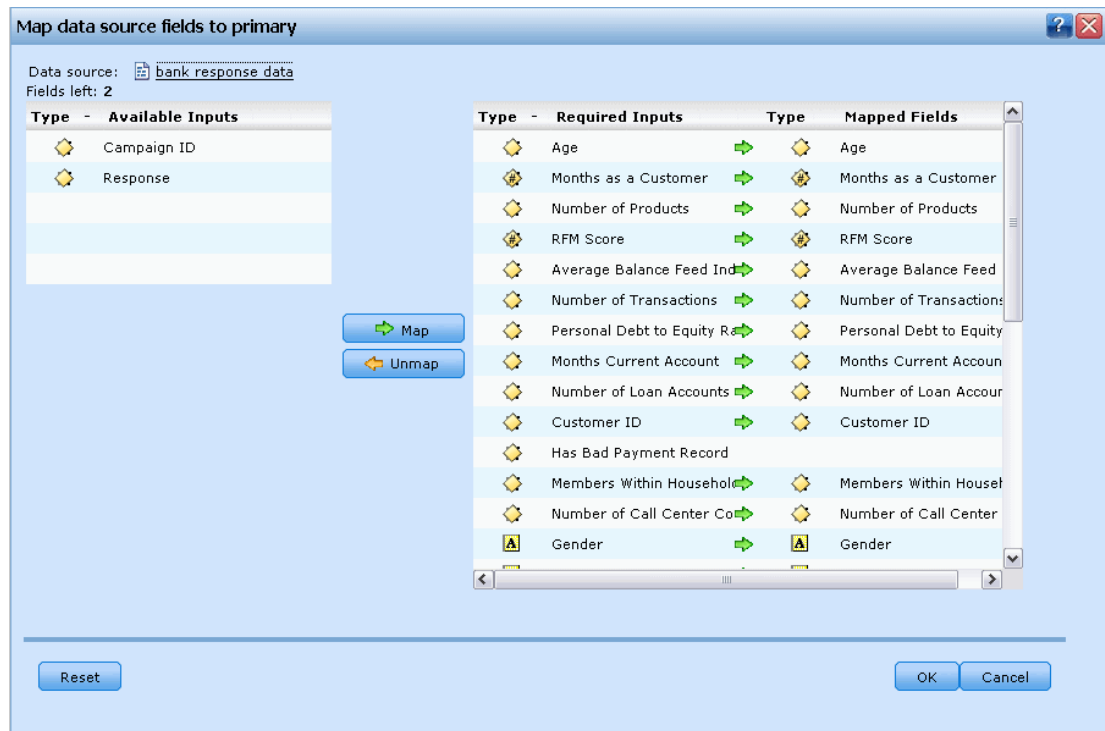
▶ In the Project Data Sources area, click Map fields. A dialog opens, and the system makes a best guess at mapping fields for you. At any point, you can return to the system's best guess by clicking Reset.

▶ 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.

▶ When finished, click OK.

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.

Figure 5-1
Global Selections tab
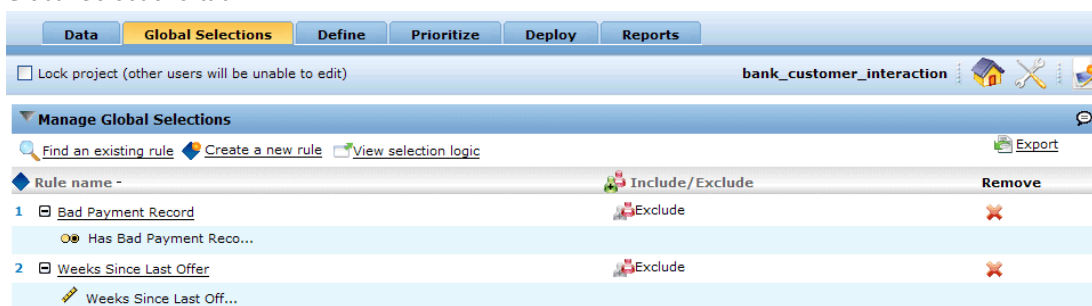
Global selections are implemented using standard selection rules, which can be defined applied at multiple points in an application. For more information, see the topic Defining selection rules in Chapter 6 on p. 58.

- 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, select Find an existing rule.

Figure 5-2
Finding a rule
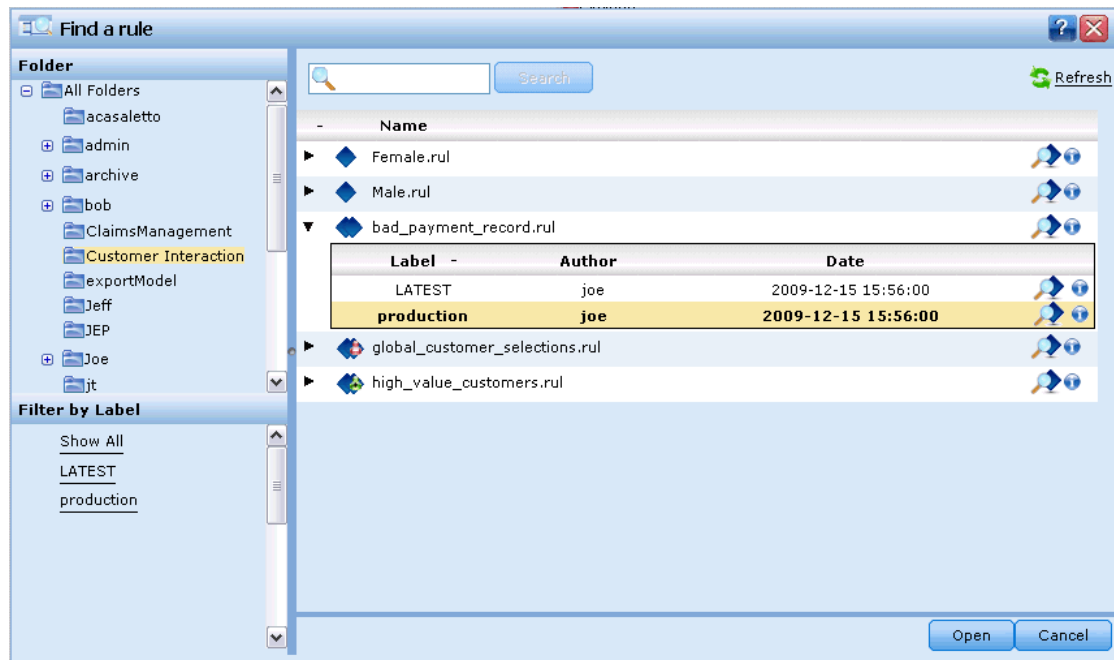
Optionally, you can export the rules you create for use in other applications. For more information, see the topic Exporting and reusing rules in Chapter 6 on p. 59.

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, 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.

Figure 6-1
IBM SPSS Rules Management
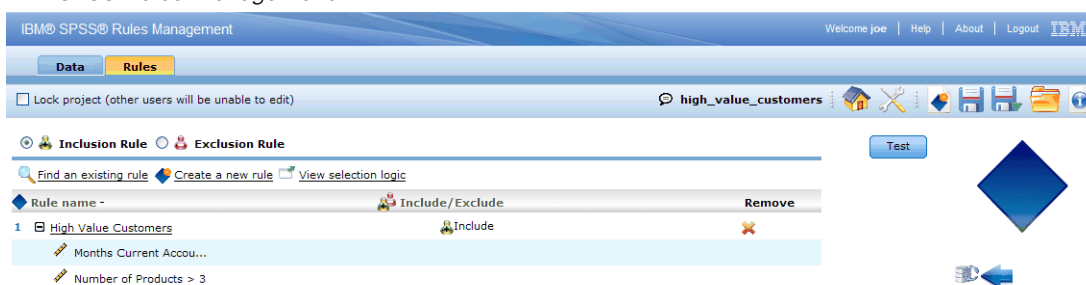
Note that SPSS Rules Management may not be installed at all sites or available to all users. As an alternative, rules can also be defined and saved locally for use within each application.

Creating rules

▶ At the point in the application where you want to add a rule, select Create a new rule.

▶ Alternatively, select Find an existing rule to browse for existing rules which can be added or reused in the current application.

55

▶ 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 returned true if the value of the age field was 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.

Figure 6-2
Segment rule editor

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

▶ In the application where you want to add the rule, select Create a new rule.

Figure 6-3
Creating a new segment rule
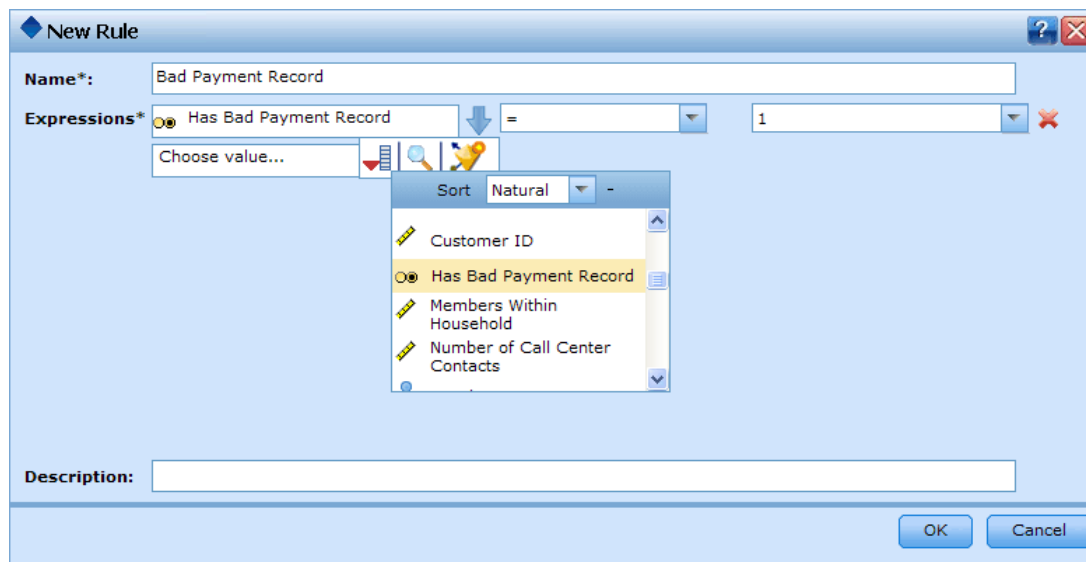
▶ Specify a name for the rule.

▶ For the input column, click the down-arrow icon to specify a field, rule, or model, as desired.

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.

▶ Choose the function you want to use. For numeric range inputs, options include the following:

- = (equals)
- != (not equal to)
- > (greater than)
- < (less than)
- >= (greater than or equal to)
- <= (less than or equal to)
- BETWEEN
- ISNIL

For categorical values:

- = (equals)
- != (not equal to)

- ONE OF
- ISNIL

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

Alternatively, select Find an existing rule to browse for rules that have already been defined. For more information, see the topic Exporting and reusing rules on p. 59.

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. For more information, see the topic Exporting and reusing rules on p. 59.

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.

Figure 6-4
Selection rule

▶ In the application where you want to add the rule, select Create a new rule. This link is available at points where selection rules can be applied.

▶ Specify a name for the rule and add one or more expressions as desired. For more information, see the topic Defining segment rules on p. 56.

▶ Click OK to save the segment rule.

▶ In the selection rule editor, click the icons in the Include/Exclude column to toggle this setting for each rule. For referenced selection rulesets, this setting cannot be changed.

Alternatively, select Find an existing rule to browse for rules that have already been defined. For more information, see the topic Exporting and reusing rules on p. 59.

Evaluating selection rules

Select View selection logic to see details on how selections are made (logical AND versus OR).

Figure 6-5
Selection rule logic

- 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, choose Export to save the rule to IBM SPSS Collaboration and Deployment Services Repository where it can be reused as needed.

Figure 6-6
Exporting a selection rule
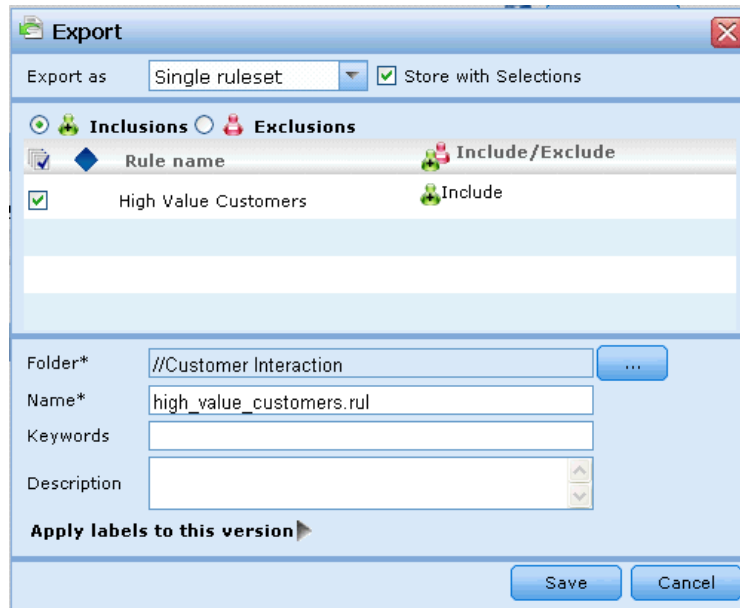
- When exporting multiple rules, specify whether you want to export them all in a single ruleset, or individual rules. In the case of individual rules, each segment is exported as a separate rule.

- 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, select Find an existing rule to browse IBM SPSS Collaboration and Deployment Services Repository for existing rules that can be reused in the current application.

Figure 6-7
Finding an existing rule
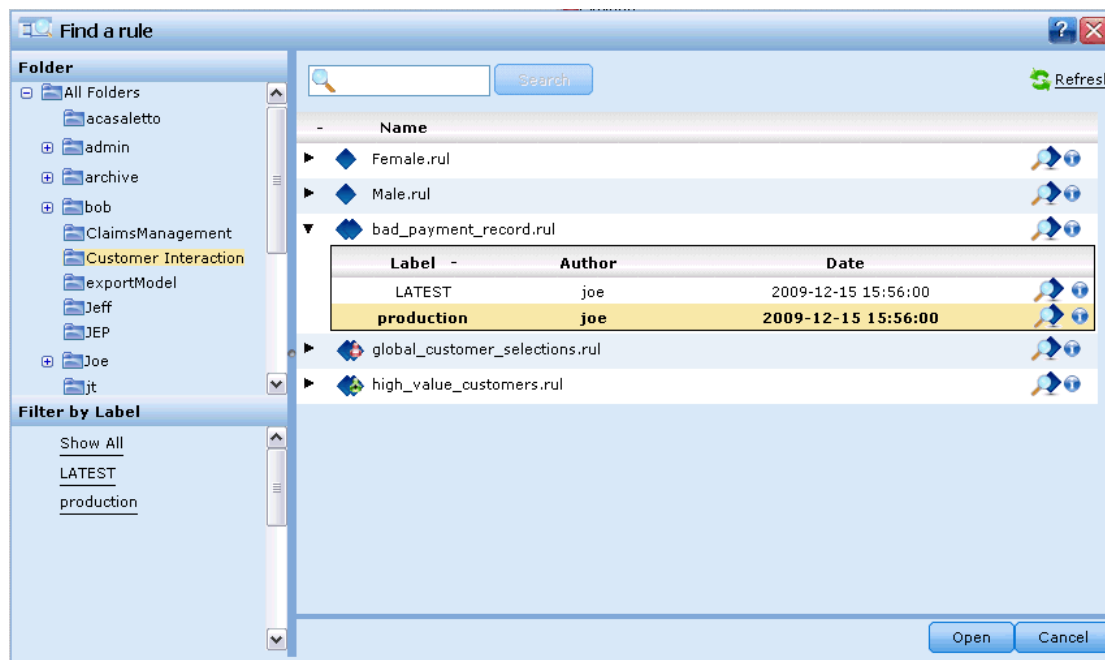
▶ To reference a specific labeled version, click the arrow next to the rule name and choose the desired version. If no version is selected, the latest is used. (Think of *LATEST* as a special label that is automatically rolled to each new version as it is created.) Using a specific labeled version is 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 SPSS Rules Management may not be installed at all sites or available to all users.

Figure 6-8
Selection rule editor with both shared and local rules
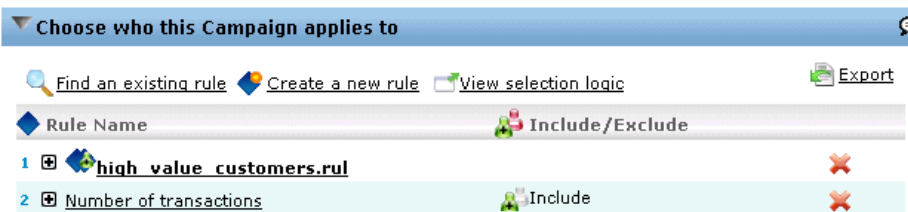
▶ Alternatively, referenced rules can be converted to local rules and saved in the current application if desired. To do this, click on the name of any referenced rule, and then select Convert to single rules in the Rule Preview dialog box.

Figure 6-9
Rule preview
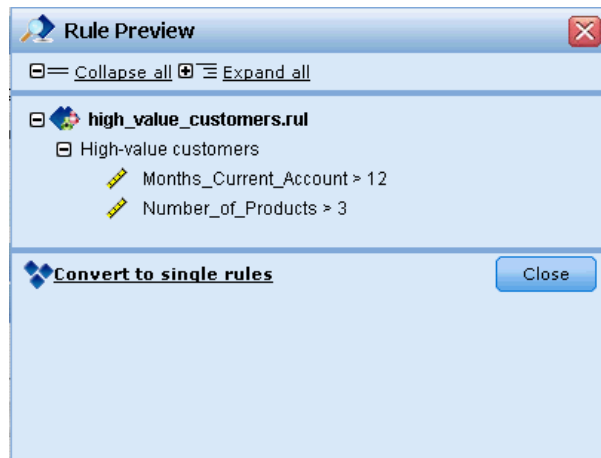
Guidelines for sharing rules
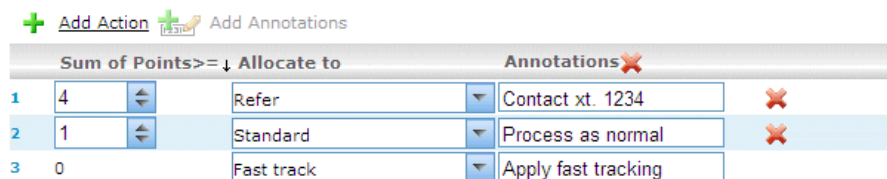
When browsing and working with shared rules, the following guidelines apply:

■   Rules can be exported for reuse from any application, but once exported, can be edited in SPSS Rules Management, and cannot be edited in any other application. Note that 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 SPSS Rules Management, the same restrictions apply as for exported rules. Nested rules are allowed, nested rule sets are not. This is because 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 SPSS Rules Management or elsewhere, a rule set can always be converted to separate, local rules.

Adding annotations

In an aggregation or allocation, select Add Annotations 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.

Figure 6-10
Adding annotations

	Sum of Points>=↓	Allocate to		Annotations✖	
1	4 ⇕	Refer	▼	Contact xt. 1234	✖
2	1 ⇕	Standard	▼	Process as normal	✖
3	0	Fast track	▼	Apply fast tracking	

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. For more information, see the topic Defining campaigns and offers in Chapter 8 on p. 89.

Using rules from ILOG Business Rule Management System

Rules created in a Business Rules Management System such as ILOG can be referenced and used in IBM® SPSS® Decision Management applications. These external rules can be selected and used in the same manner as other rules, by selecting Find an existing rule 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 SPSS Decision Management, see the *Application Designers Guide*.

Figure 6-11
Using an external rule
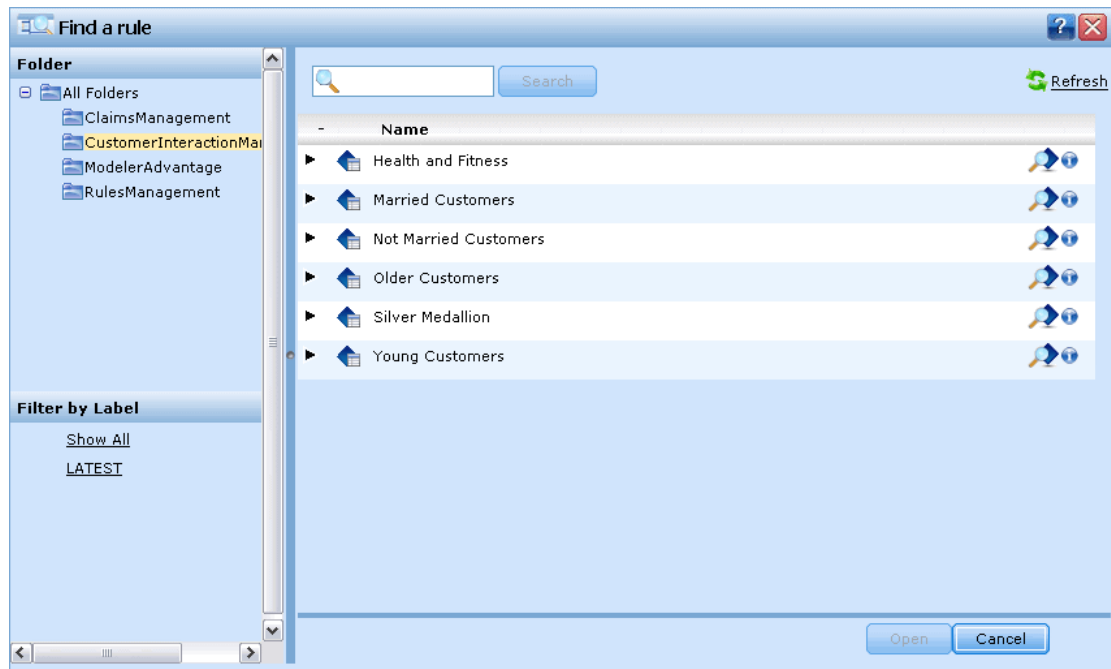
The rule is inserted into the application in the same manner as any other rule. Note that rule outputs must match the context in which the rule is used; for example allocations require a true/false or yes/no (boolean) value as input to the decision.

Figure 6-12
Using an external rule in an allocation
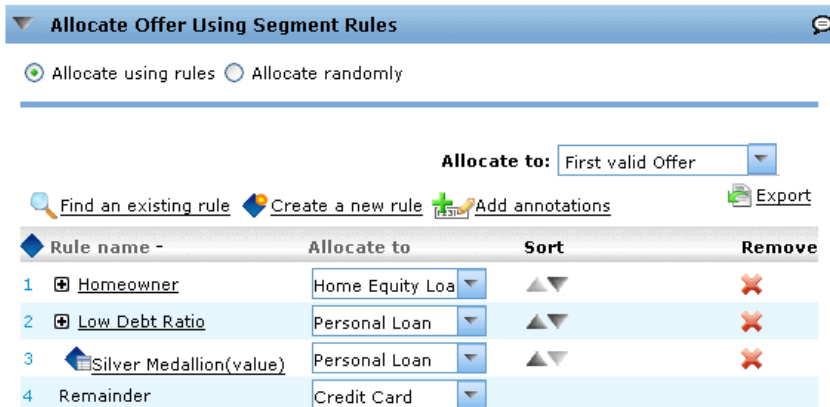
As with other rules, click the name of any external rule used in the application to view details on the rule including input and fields, along with the rule description and other properties.

Figure 6-13
External rule preview
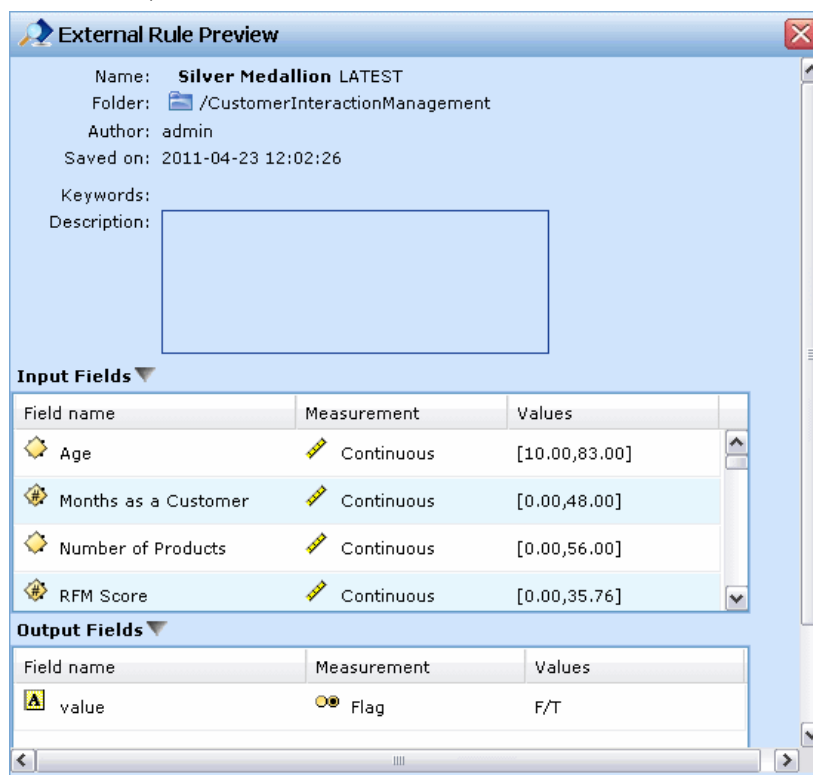
If an external rule has multiple output fields, a prompt is displayed allowing you to choose the desired field.

Figure 6-14
Selecting an external rule field
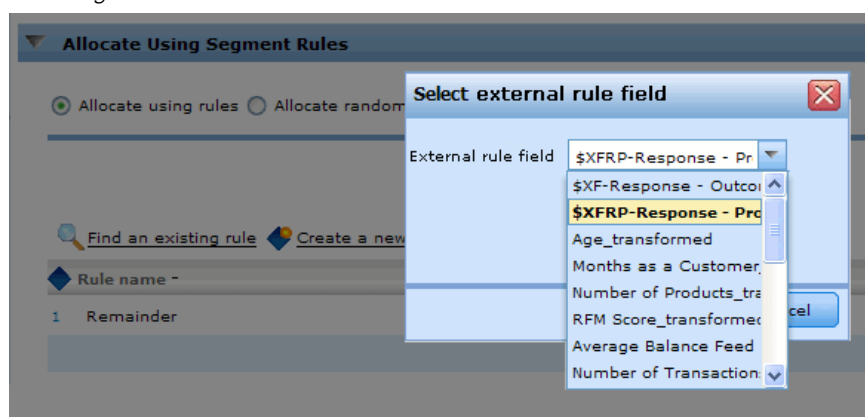
If an external rule is used in place of a model, no target field is shown, and all configured outputs are shown in the drop-down list.

Figure 6-15
Using an external rule in place of a model
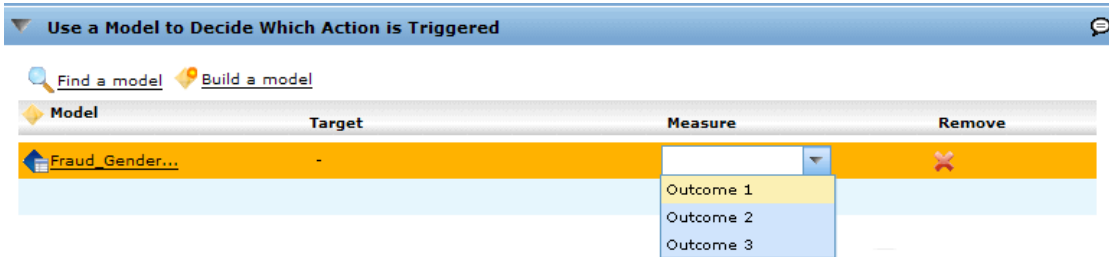
Tips for using external rules

- 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.

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, select Build a model under any panel where applicable.

- From the Prioritize tab, select the build new model icon from any input field toolbar.

- When creating or editing a rule, select the build new model icon from the input field toolbar in the rule editor.

Figure 7-1
Build model icon

Figure 7-2
Building a model from the rule editor
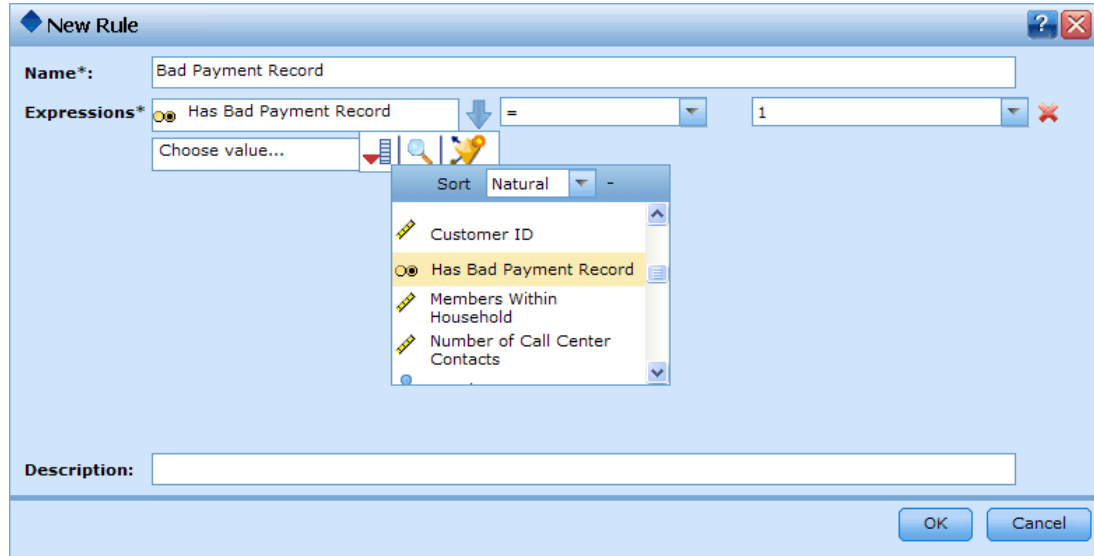
To build an automated model

When building an automated model, the only required specifications and 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.

Figure 7-3
Model settings
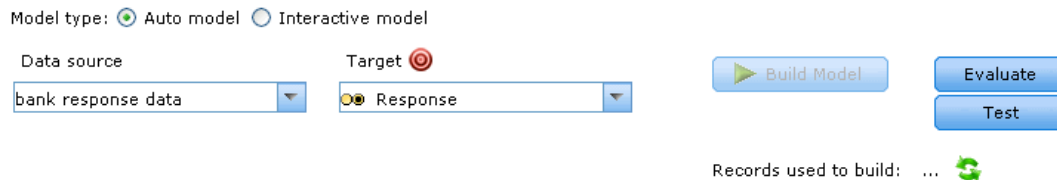
▶ Specify a data source. This can be any analytical data source that records the outcome you want to predict.

▶ Specify a target field the 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.

▶ Specify optional settings as desired. For more information, see the topic Optional model settings on p. 75.

▶ Select 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. For more information, see the topic Gallery in Chapter 3 on p. 26. 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, distribution and predictor importance charts are displayed. For more information, see the topic Automated modeling results on p. 69.

▶ Optionally, use the Evaluate and Test features to see how the model performs on your sample data. For more information, see the topic Evaluating models on p. 77.

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

▶ 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.

Automated modeling results

When an automated model completes, distribution and variable importance charts are displayed, providing a quick visual assessment of model performance, and of the importance of different predictors in obtaining the final result.

Figure 7-4
Distribution of observed and predicted values
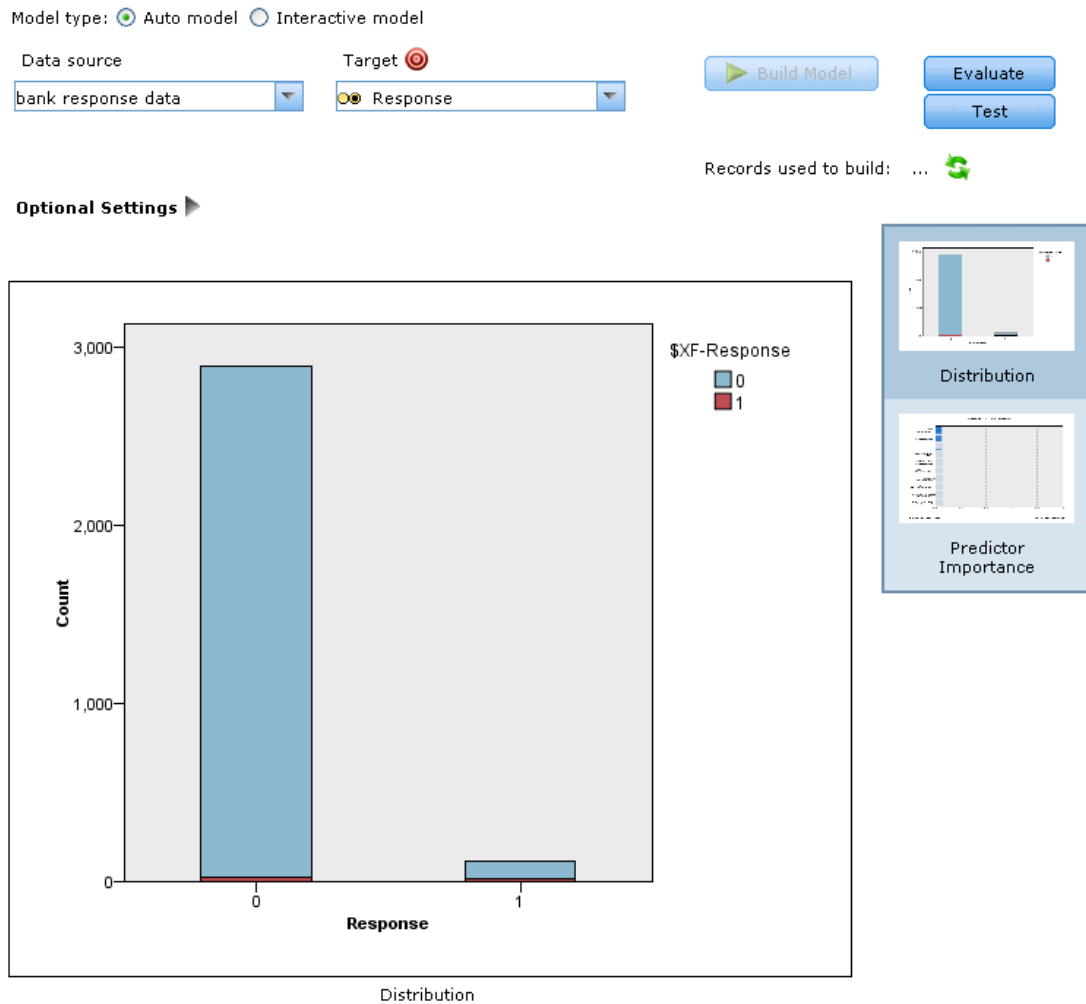
Distribution

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.

Figure 7-5
Variable Importance chart
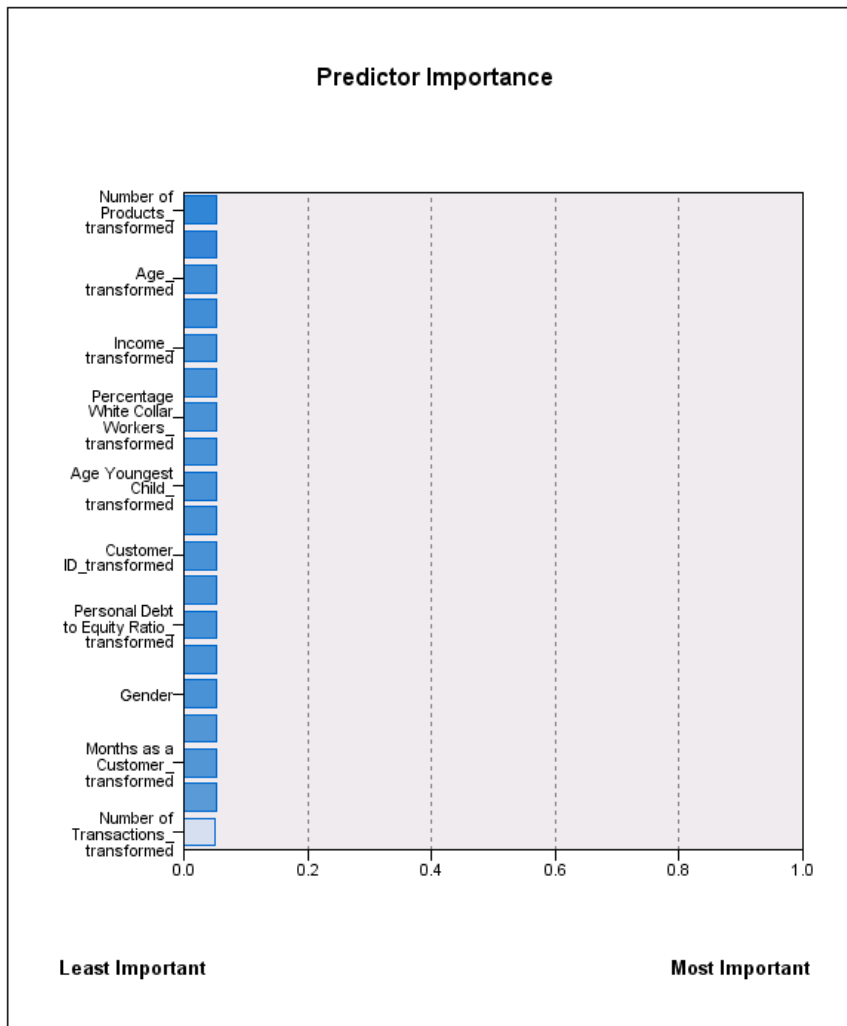
If the option to automatically partition data was selected when the model was built, then the testing partition will automatically be selected on the Evaluate page.

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.

Interactive models

Interactive 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.

Interactive 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, interactive models are well-suited for generating mailing lists, and they are widely used in customer relationship management, including call center or marketing applications.

■ Interactive 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 a 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 an interactive model

Figure 7-6
Building an interactive model

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

▶ Specify a target field that records the result you want to predict.

▶ Specify the specific value or response you want to find. For example if the target field is named *response*, you cold 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.

▶ Select Start Build.

Figure 7-7
Building an interactive model

▶ 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.

▶ Specify the maximum number of segments to be found. Typically this number is kept small in the interest of simplicity and performance.

▶ Select Grow Model to obtain a list of one or more segment rules that identify groups with higher or lower probability.

▶ 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.

▶ Optionally, you can manually add or edit segments as desired. For more information, see the topic Defining segment rules in Chapter 6 on p. 56.

▶ Optionally, use the Evaluate and Test features to see how the model performs on your sample data.

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

▶ 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 select 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.

Optional model settings

When building models, you can choose from the following optional settings. If changing any of these settings for an existing model, the model must be rebuilt for changes to apply.

Figure 7-8
Specifying optional settings

Automatically partition data, to enable model validation on build data source. 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.

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.

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. For more information, see the topic Expression manager in Chapter 4 on p. 50.

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. For more information, see the topic Defining selection rules in Chapter 6 on p. 58.

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

Figure 7-9
Segmentation options

When build interactive 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.

Evaluating models

After building or opening a completed model, select Evaluate to obtain distribution, gains, lift, response, profit, and ROI charts, and to simulate the maximum profit possible. This dialog is only available if the selected model is complete.

Figure 7-10
Gains chart

▶ 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.)

▶ Select the graphs you want to build.

▶ If simulating profit, select the Simulate maximum profit box and enter the population on which to base the profit calculation.

▶ 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.

▶ 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.

▶ Select 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.

For flag and set targets, you can choose from the following additional charts:

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.
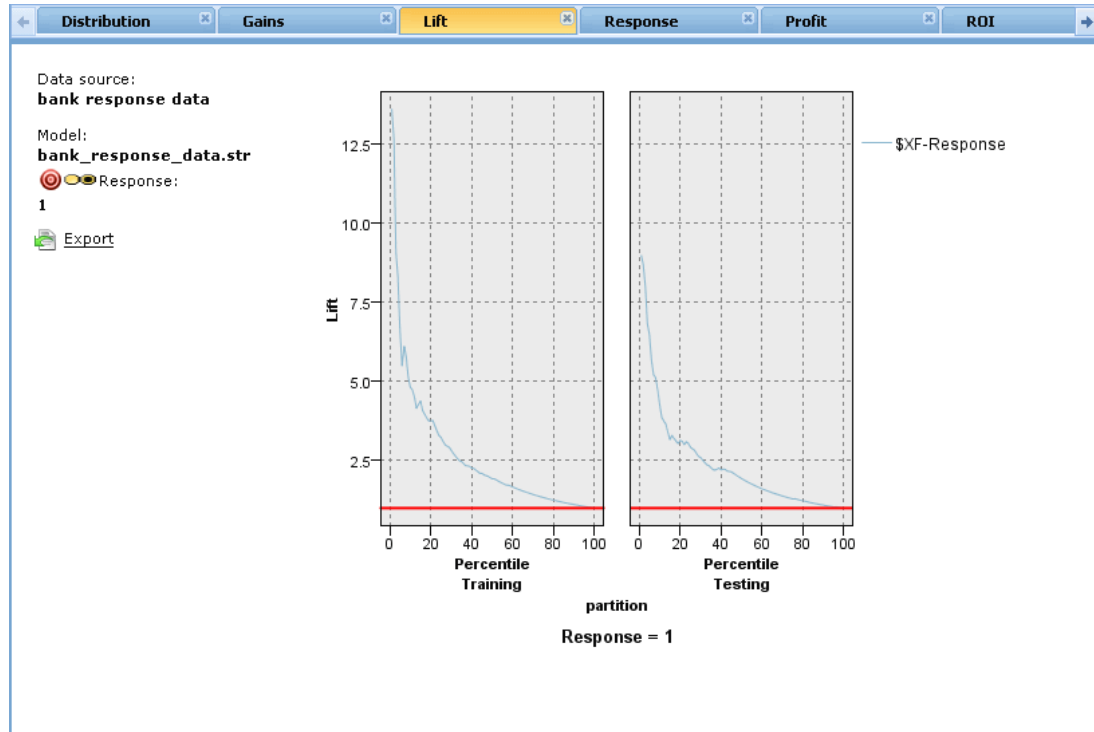
Figure 7-11
Gains chart
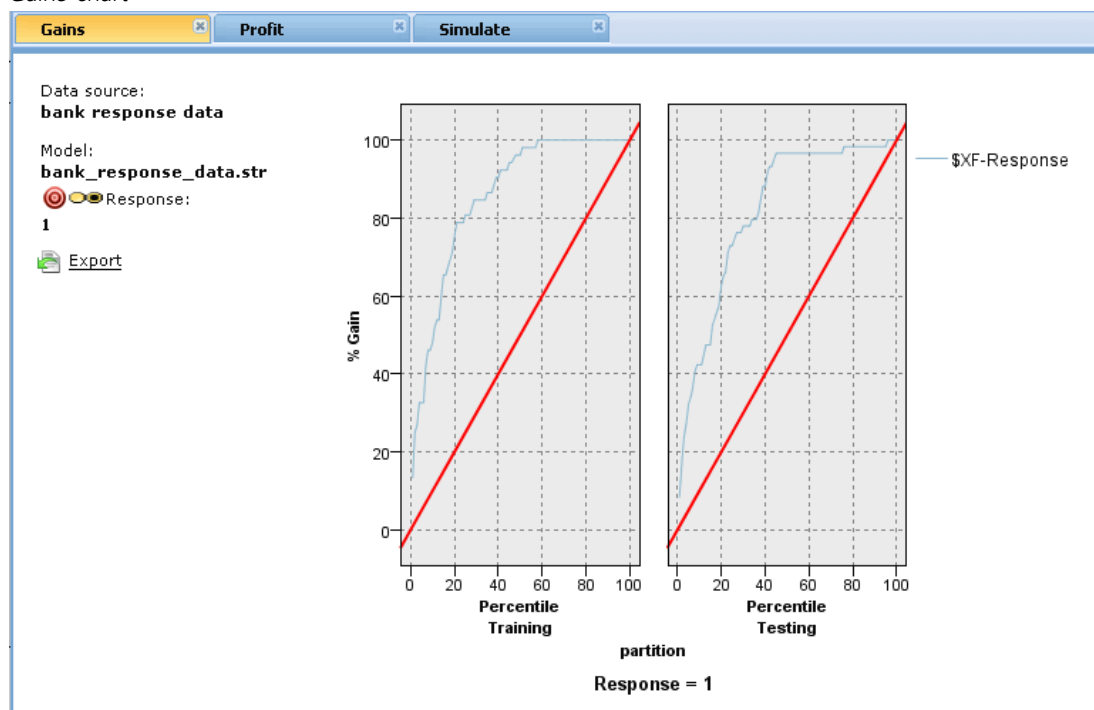
Lift charts

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.
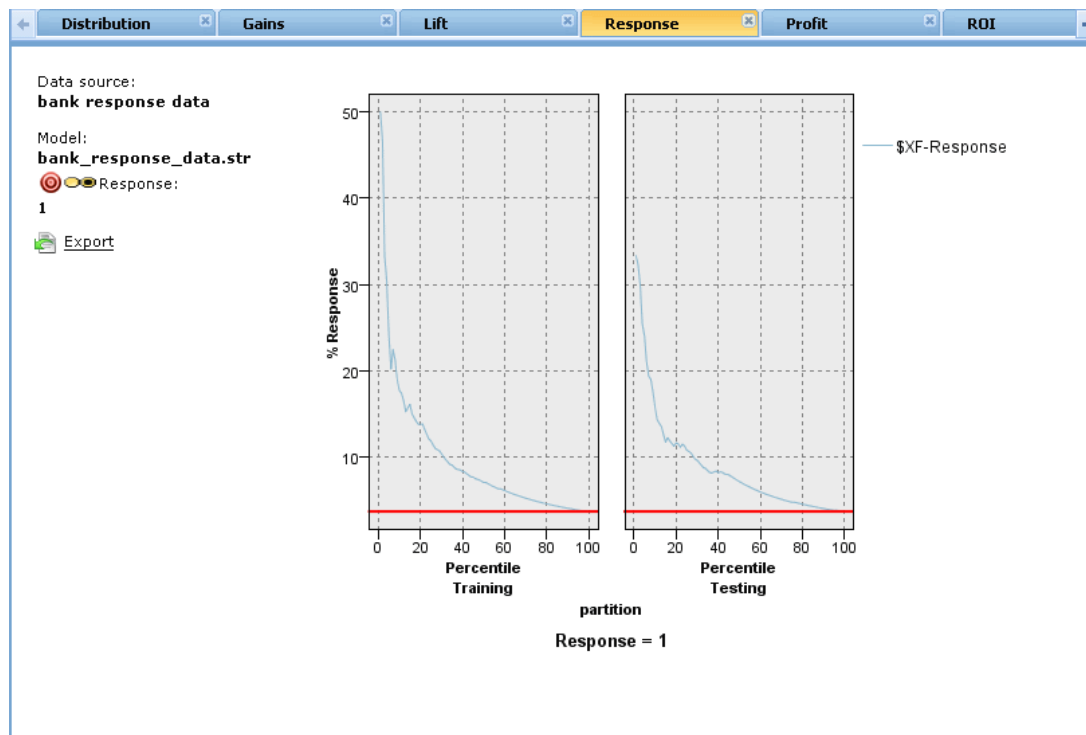
Figure 7-12
Lift chart

Response charts

The response chart plots the percentage of records in the increment that are hits.
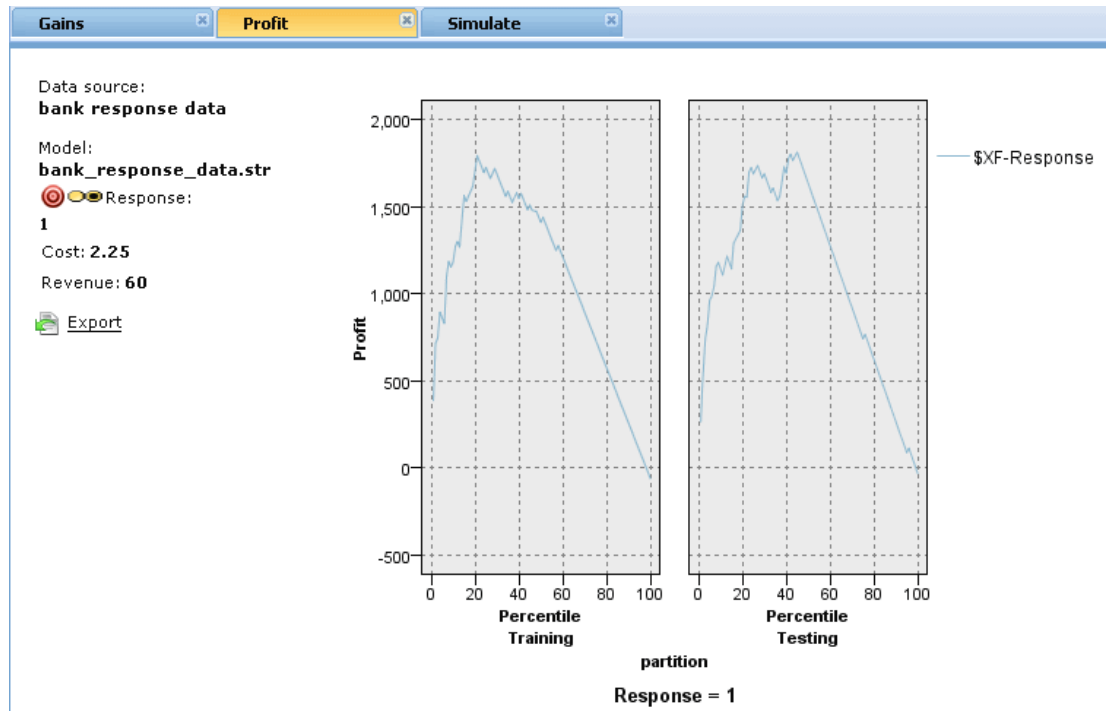
Figure 7-13
Response chart

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. For more information, see the topic Simulating maximum profit on p. 83.

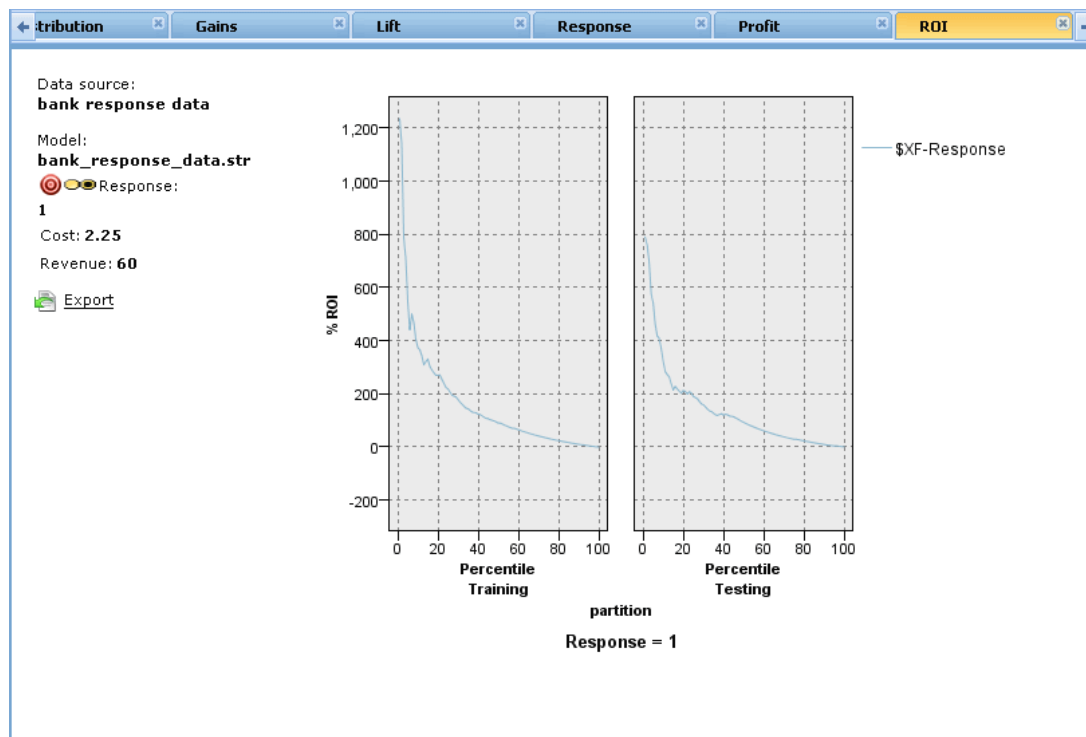Figure 7-14
Profit chart

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.
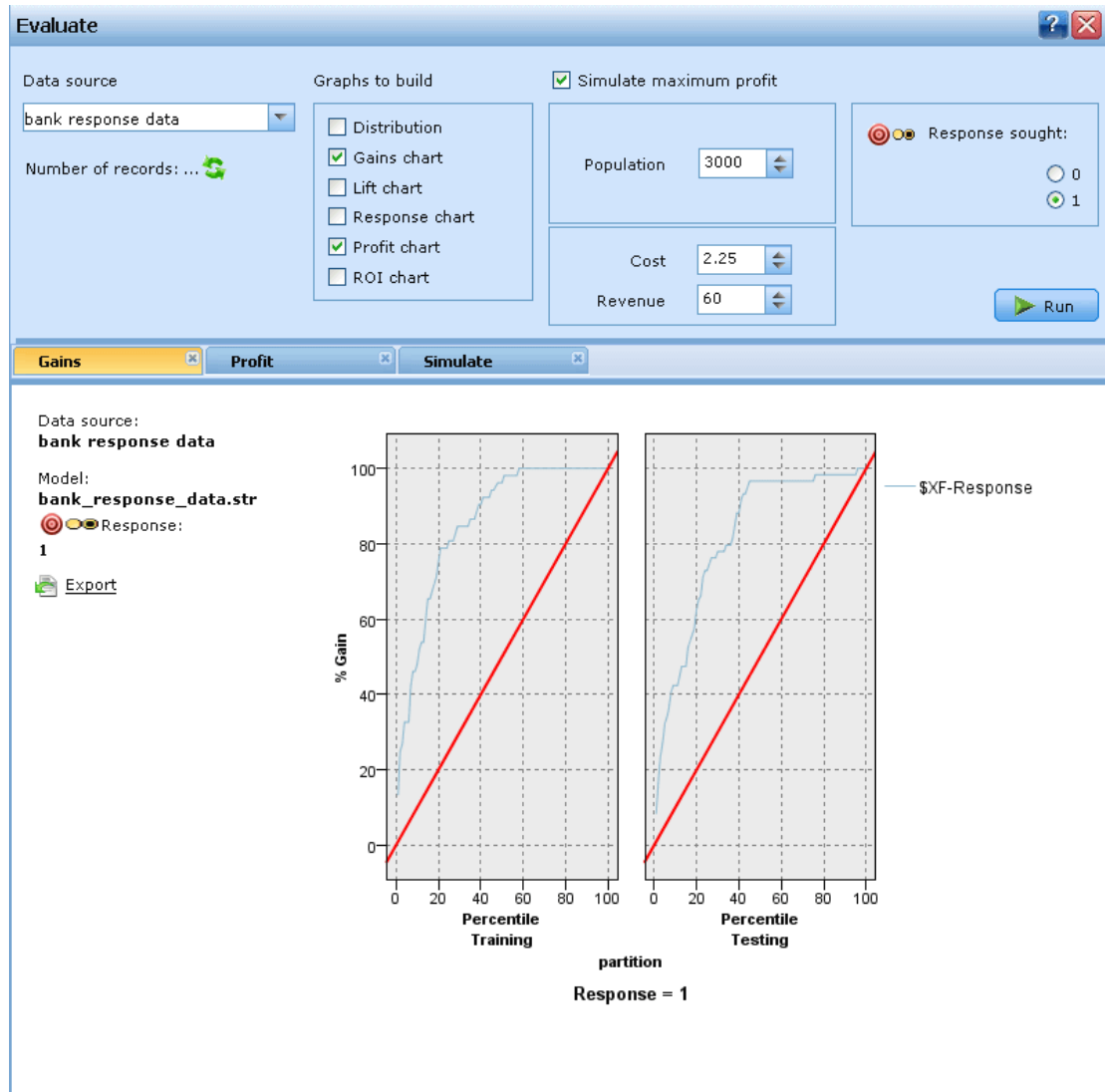
Figure 7-15
ROI chart

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.

Figure 7-16
Simulation settings

To predict the maximum profit:

▶ 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.

▶ 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.

▶ Specify the expected revenue that a positive response would generate. For example, the retail cost
of an item in a campaign.

▶ Select the response you are interested in.

▶ Click Run.

Figure 7-17
Simulated profit prediction
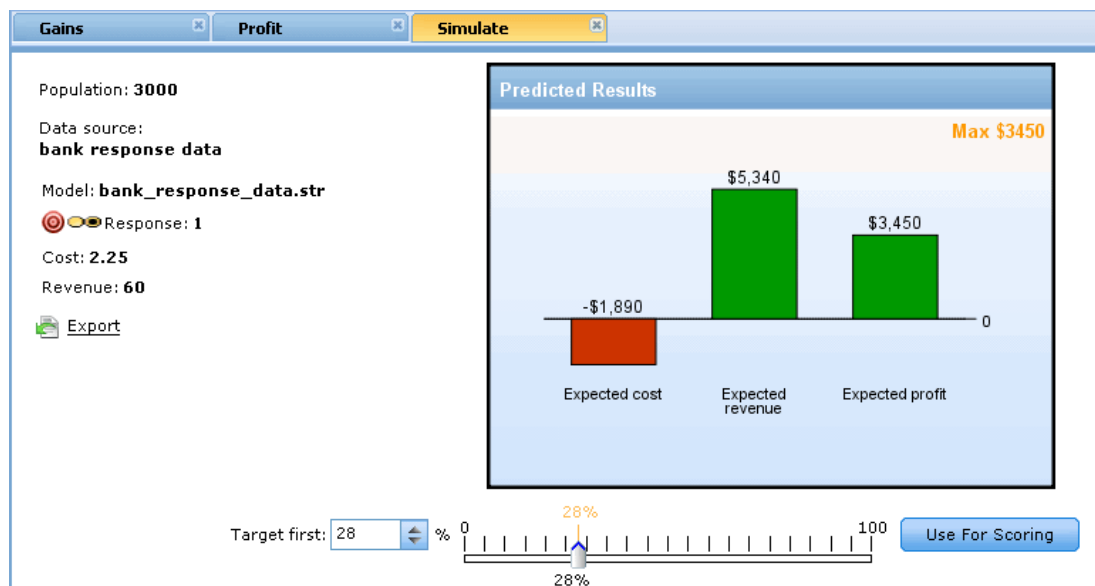
The predicted results graph shows the best result for maximum profit, and identifies the percentage of records you would need to target to achieve this, based on propensity to respond as identified by the model.

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.

Figure 7-18
Profit comparison

Compare Profit ▼

|	Targeted28%	Random28%	Everyone
Offers made	840	840	3000
Expected response	89	31.08	111
Expected response(%)	10.6	3.7	3.7
Expected cost	$1890	$1890	$6750
Expected revenue	$5340	$1864.8	$6660
Expected profit	**$3450**	**$-25.2**	**$-90**

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.

Figure 7-19
Cumulative profit chart
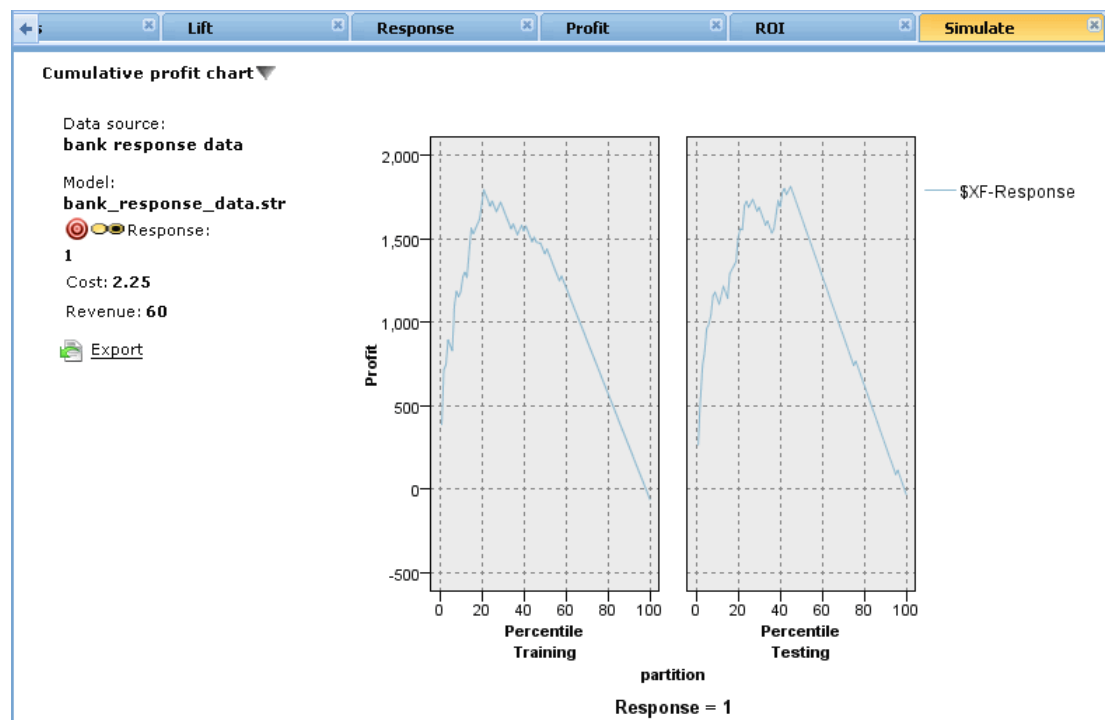
Using model scores in applications

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.

Figure 7-23
Creating a new segment rule
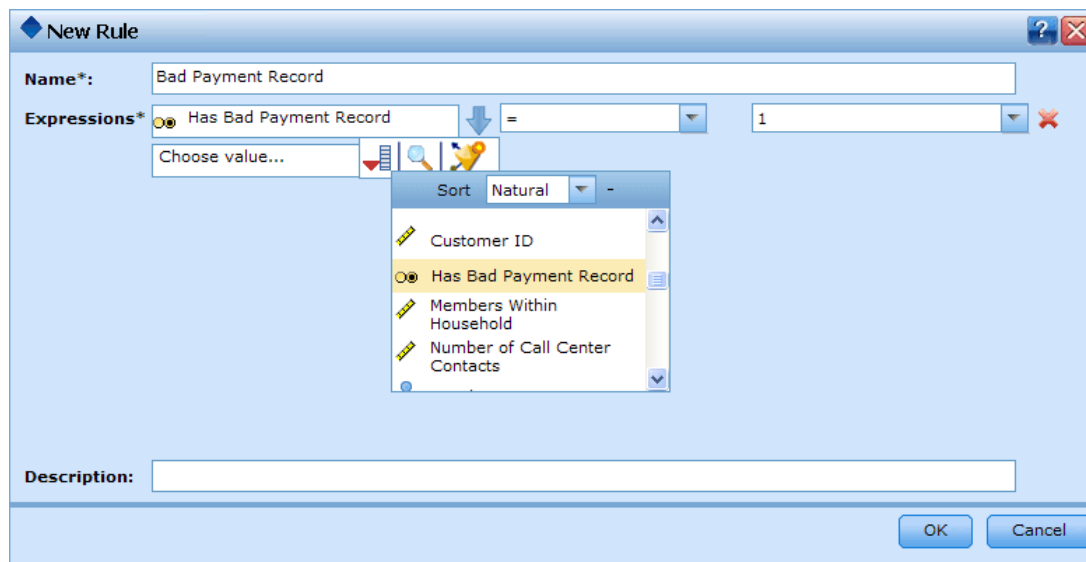
Model scores can be used in the following ways:

■ As input to rules. For more information, see the topic Creating rules in Chapter 6 on p. 55.

■ To allocate decisions on the Define tab.

■ As inputs to the prioritization. For more information, see the topic Prioritization in Chapter 9 on p. 99.

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. For more information, see the topic Measurement levels in Chapter 4 on p. 39.

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.

Interactive models. Interactive 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. For more information, see the topic Interactive models on p. 72.
- **$DI-<target>.** Index number used to identify the segment into which the record falls (displayed in the left-most column in the modeling results).
- **$DRP-<target>.** Propensity score indicating the likelihood of the desired response. Available for flag targets only.

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 step.

Figure 8-1
Defining campaigns

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.

- The output from the rules is returned as either a single decision or multiple recommendations.

## *Defining campaigns and offers*

The tree view at left defines the campaigns and offers that are available in the application. These items define the possible decisions, or results, that can be returned. The right pane displays properties, selections, and allocations for the selected item.

Figure 8-2
Dimension tree
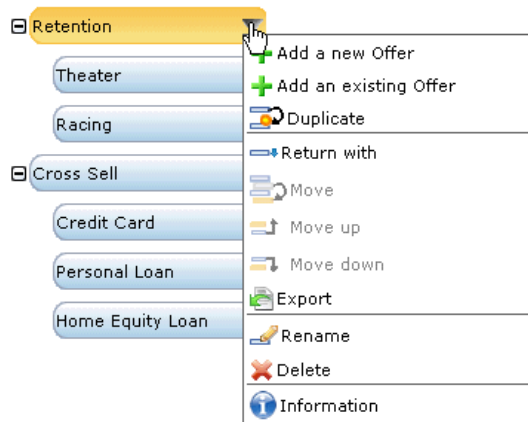
▶ Click Add... at the top of the tree to add dimension items.

▶ Click Expand all or Collapse all to show or hide items.

▶ Click on individual items to display a context menu that allows you to add, rename, move, and duplicate items.

▶ To move an item, select Move from the context menu, and select the new location from the drop-down list.

▶ To reuse an existing item, select Add an existing... from the context menu, and choose the item you want to reuse.

Figure 8-3
Reusing an existing item
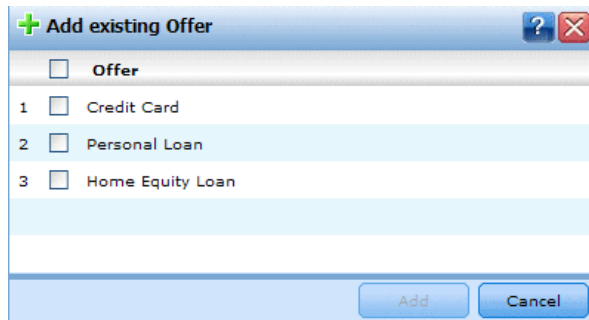
▶ To specify a field, text, or model output to be returned with an item, select Return with from the context menu. Return fields are extremely useful in tailoring recommendations to specific cases, or providing additional information that may be useful in processing the result. Along with a loan offer, 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.

Figure 8-4
*Specifying a return field or value*
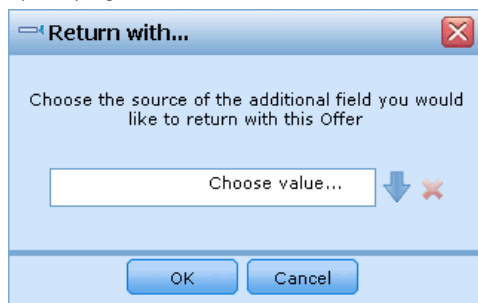
▶ If the tree is too large to be visible, you can drag the right border to resize the panel.

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.

Figure 8-5
Importing dimension items
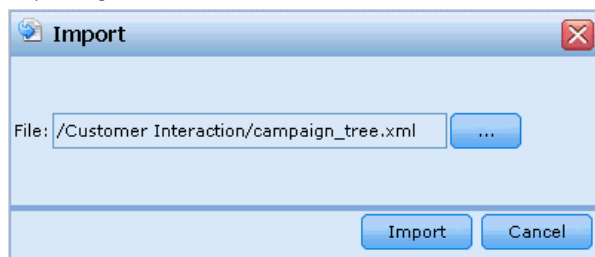
▶ To export the current tree, select the Export link displayed just above the tree, then specify a folder and filename.

▶ To import dimension items, select Import, and specify an XML file containing the items you want to import. This should be a file exported from a compatible application that uses dimensions 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.

Campaign properties

Figure 8-6
Dimension 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. For more information, see the topic Defining interaction points in Chapter 12 on p. 116.

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.

Figure 8-7
Campaign selections

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. For more information, see the topic Defining selection rules in Chapter 6 on p. 58.

Defining allocations

Recommendations can be allocated using the following methods:

■ Based on segment rules or random percentages (which can use model scores as inputs)

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.

Figure 8-8
Allocation using segment rules

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.

▶ 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.

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.

Figure 8-9
Random allocation

▶ 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 affect on rule execution.

▶ Select Add Annotations to add a new column to the right of the Order column. You can enter free text into the fields in this column that relate to each allocation.

Simulate

From the Define tab, you can run a simulation to see how records will be allocated based on the current settings.

Figure 8-10
Simulation results
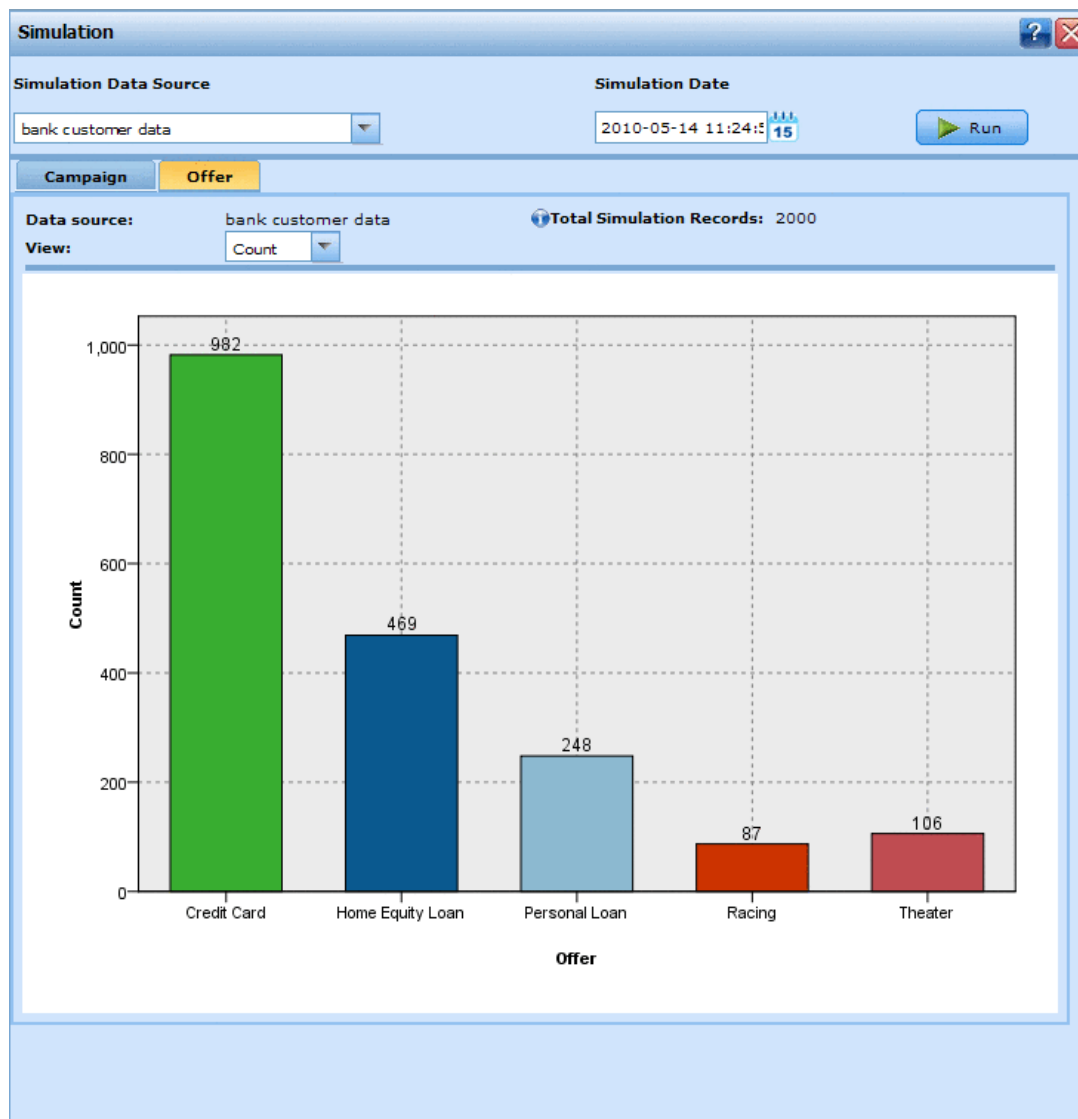
▶ Specify the data source you want to use. This can be any data source that has Simulation selected under as the usage type. Typically this would be a sample of a few thousand records (though it can be larger or smaller) that is representative of your operational data.

▶ Specify the simulation date and interaction point to use when running the simulation. These settings depend on the current application, but can have significant impact on the results.

The simulation results display a separate chart 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.

▶ Choose Count or Percent to customize the view.

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.For example, you can look at a handful of specific customers and see which offers are allocated to each of them. In addition, you can view details to see how specific selections and rules are applied as factors in the decision.

Figure 8-11
Test results

▶ Select the data source you want to use, or select Specify custom data in the Records section to enter test records manually.

▶ Specify the test date and interaction point to use where applicable. These settings depend on the current application, but can have significant impact on the results.

▶ 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. For more information, see the topic Defining selection rules in Chapter 6 on p. 58.

▶ 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.

▶ 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.

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

▶ Click Run.

▶ The test summary displays results for each record. Values for the selected input fields will be displayed.

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

Figure 8-12
Test record preview

Field name -	Value
Age	26
Months as a Customer	12.0
Number of Products	1
RFM Score	**7.879**
Average Balance Feed Index	**179**
Number of Transactions	**1**
Personal Debt to Equity Ratio	26
Months Current Account	8

Where available, select Test Detail to see additional details as available. For example you can see how recommendations are determined for the current record based on rules or models that fired at each step.

Figure 8-13
Test details

Figure 8-14
Additional details

Prioritizing decisions

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

Prioritization

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? Prioritization allows you to identify the best or most profitable offer for each customer, and to experiment with different parameters to see how each affects your bottom line.

Figure 9-1
Prioritize tab

▶ Specify 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, or build a new model which can then be used as input to the prioritization.

▶ Select Override to give priority to certain campaigns or offers. Offers marked with override will be considered before those that are not so marked. The prioritization equation is still computed for all offers, but the selection is done in two groups (first for override items, then for others).

Figure 9-2
Choosing a field or other input type

Campaign/Offer	Prob.to Respond	Min.Profit	Revenue	Cost
⊟ Retention				
Theater	0.1	10	✏ Annual va	22
Racing	0.1	10	180	21
⊟ Cross Sell				

▶ Select Customize table to choose whether to specify parameters at the campaign level, or separately for each offer within a campaign.

Figure 9-3
Customizing the table

Parameter -	Campaign	Offer
Prob.to Respond	○	◉
Min.Profit	○	◉
Revenue	○	◉
Cost	○	◉

Customize Prioritization Table
Choose at which level you wish to define each parameter
[Save] [Cancel]

▶ To experiment with different parameter combinations and see how each impacts your result, click WhatIf?...

▶ Optionally, you can to use the same settings for all interaction points, if available. If not, you can deselect this option and choose interaction points separately from the drop-down list. For example, the cost of an offer may differ when made through a call center rather than a retail store location, while a Web page might allow multiple offers to be displayed to the same customer. Note that interaction points must be configured by the administrator.

Prioritization equation

Figure 9-4
Prioritization equation

▽ Prioritization Equation(Value to be maximized)

(**Prob.to Respond** ✕ **Revenue**) ⊖ **Cost**

The Prioritize tab determines the most profitable offer for each customer, based on the following equation.

```
Expected profit = ( [probability to respond] * [income] ) - [cost:]
```

Where:

- Probability is a customer's propensity to respond to a particular offer.
- Income is the value or contribution expected from that response .
- Cost is the cost of making the offer.

Note this function can be modified for different applications by the application designer. While the specific function may change, many business problems may be modeled in this manner.

WhatIf? analysis

Select WhatIf?… on the Prioritize tab to experiment with different parameters and see how changes impact the final result. Each simulation run adds a new column in the results table at lower left, allowing a side-by-side comparison.

Figure 9-5
WhatIf? dialog box

▶ Select a data source, and specify a date for the simulation. Depending on the active dates defined for each campaign, this can have a big impact on the result.

▶ Select an interaction point to use for the simulation, if available. Interaction points are configured by the administrator. For more information, see the topic Defining interaction points in Chapter 12 on p. 116.

▶ To try a new combination, edit the input values, type a name for the simulation, and click Run.

Simulation results are displayed in the panel at lower left. Each simulation adds a new column. Results are displayed for each campaign or offer, as well as the total across all dimensions.

▶ Select the measure to display from the drop-down list. The following measures are available:

- **Count.** The number of records allocated to that item.
- **Percent.** The percentage of total records accounted for by the current item.
- **Predicted Profit.** Expected profit based on the prioritization equation.
- **Probability to Respond.** Average across all offers made.
- **Cost.** Based on the number of offers.

▶ Click on any simulation column to make it current.

▶ To apply settings from the current simulation to the Prioritize tab, click Update Settings. Alternatively, you can close the WhatIf? dialog without updating settings to restore the original state.

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.

Figure 10-1
Deploy tab
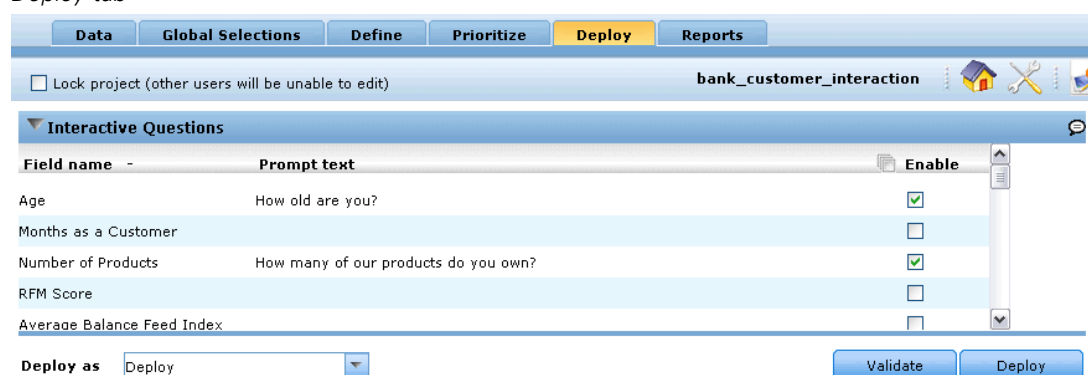
To specify interactive questions

Interactive questions 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.

▶ For each question you want to specify, select the Enable check box for the appropriate field, and specify the text to be used.

▶ 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.

To label the project for deployment

▶ From the Deploy As list, select the label to apply to the project. Available labels are configured by the local administrator, but would typically inlcude options such as *Test*, *Pre-production*, and *Deploy*.

103

▶ Click Validate; any errors or missing steps are listed for you to address.

Figure 10-2
Validation errors
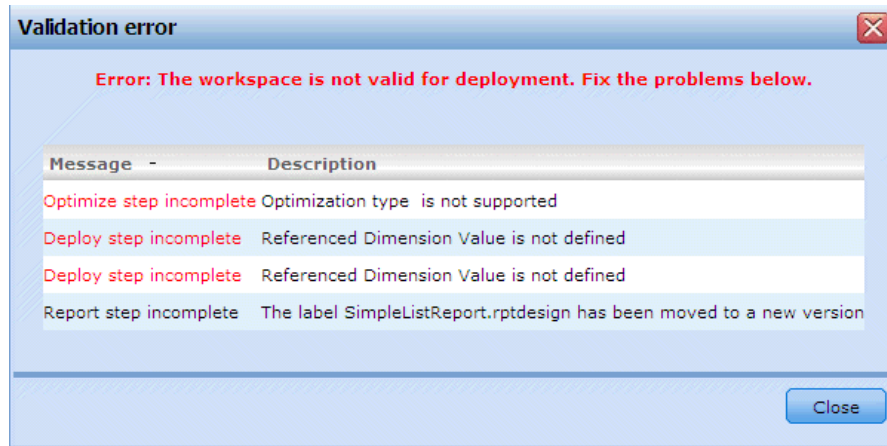
▶ To apply the label, click Deploy. If a previous version of the project is already deployed a warning is displayed; you can choose whether to deploy your latest version or keep the existing deployed version.

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.

Reporting results

Report types

IBM® SPSS® Decision Management is able to read reports from many formats, these include:

- Text files (.txt, .csv)

- Spreadsheets (.xls, .xlsx)

- Images (.png, .jpeg, .gif, .bmp

- HTML

- PDF

In addition to these file types, SPSS 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 BIRT project page (*http://www.eclipse.org/birt*). BIRT Report Designer for IBM® SPSS® is a standalone application that can be used in conjunction with SPSS 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.

Figure 11-1
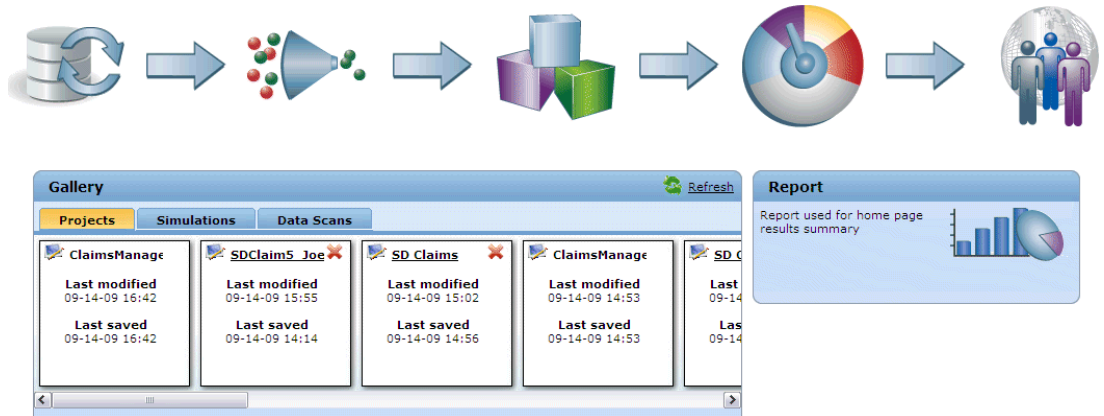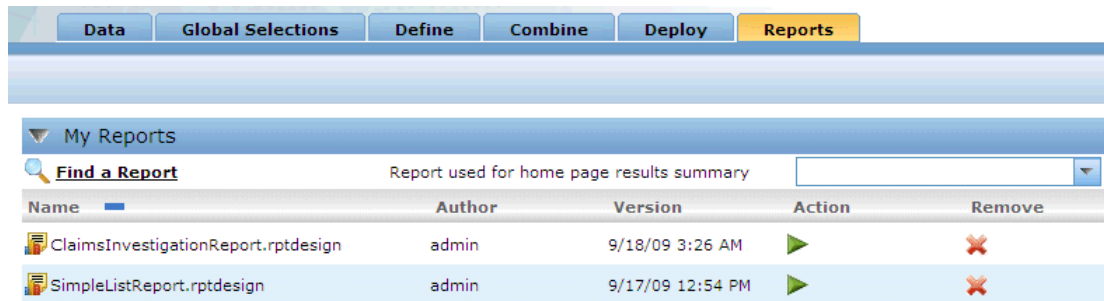*Report link on main application page*
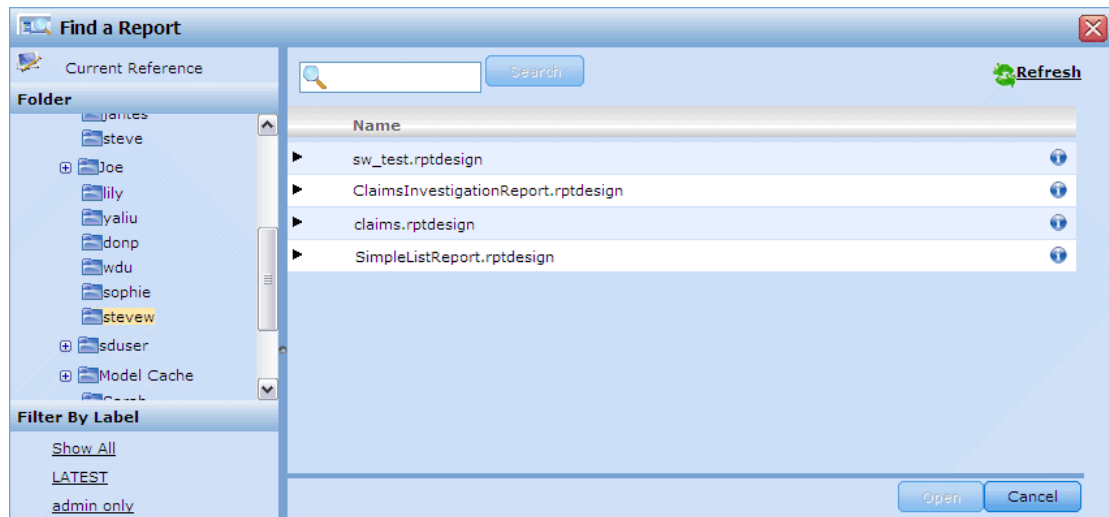

Figure 11-2
Reports tab within an application

Any reports that you have already viewed are displayed.

To retrieve a report:

► Click Find a Report to display a browser window where you can search for the report.
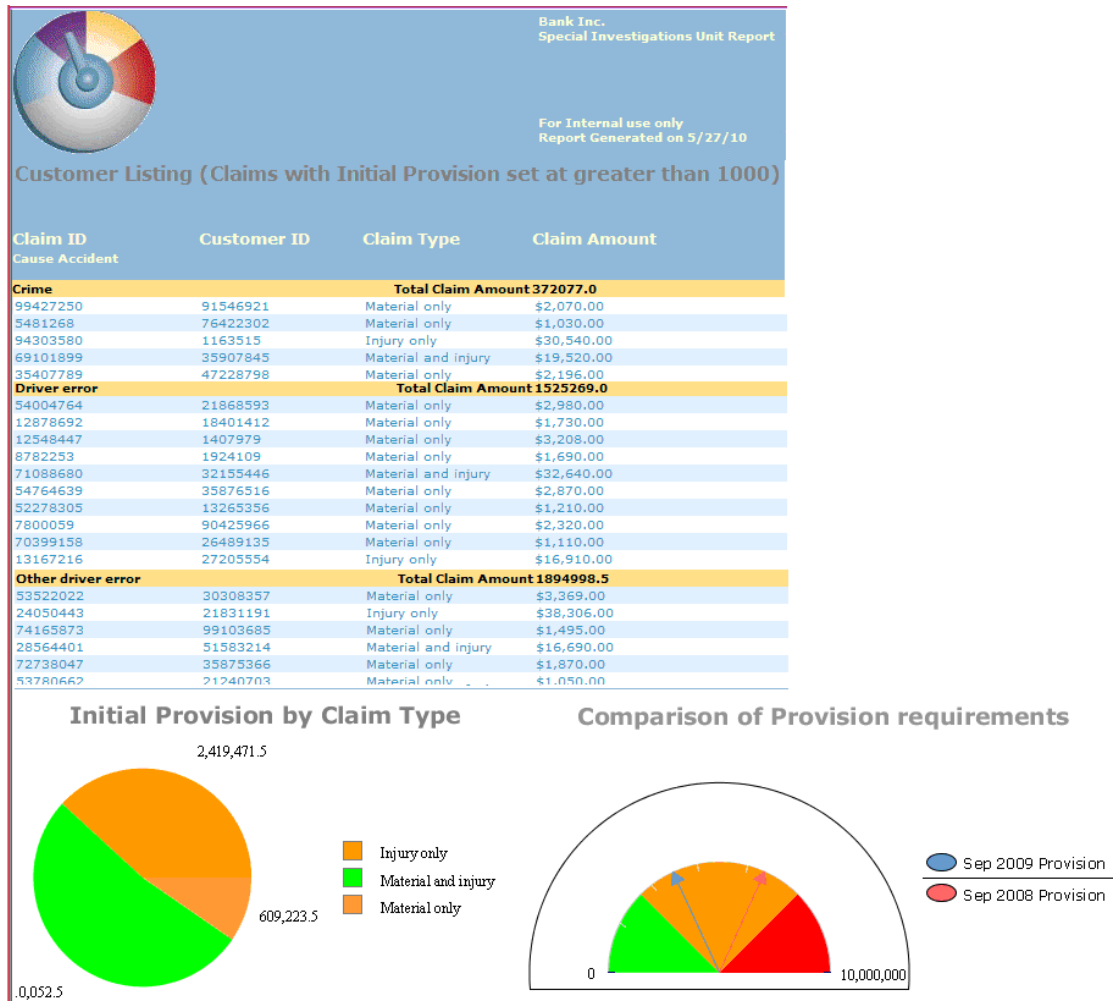
Figure 11-3
Browse for existing reports

▶ Select the report.

▶ Click Open; the report is added to the list of those you can access from within the application.

▶ To view the report's contents, click the icon in the Action column. The report opens in a new window.
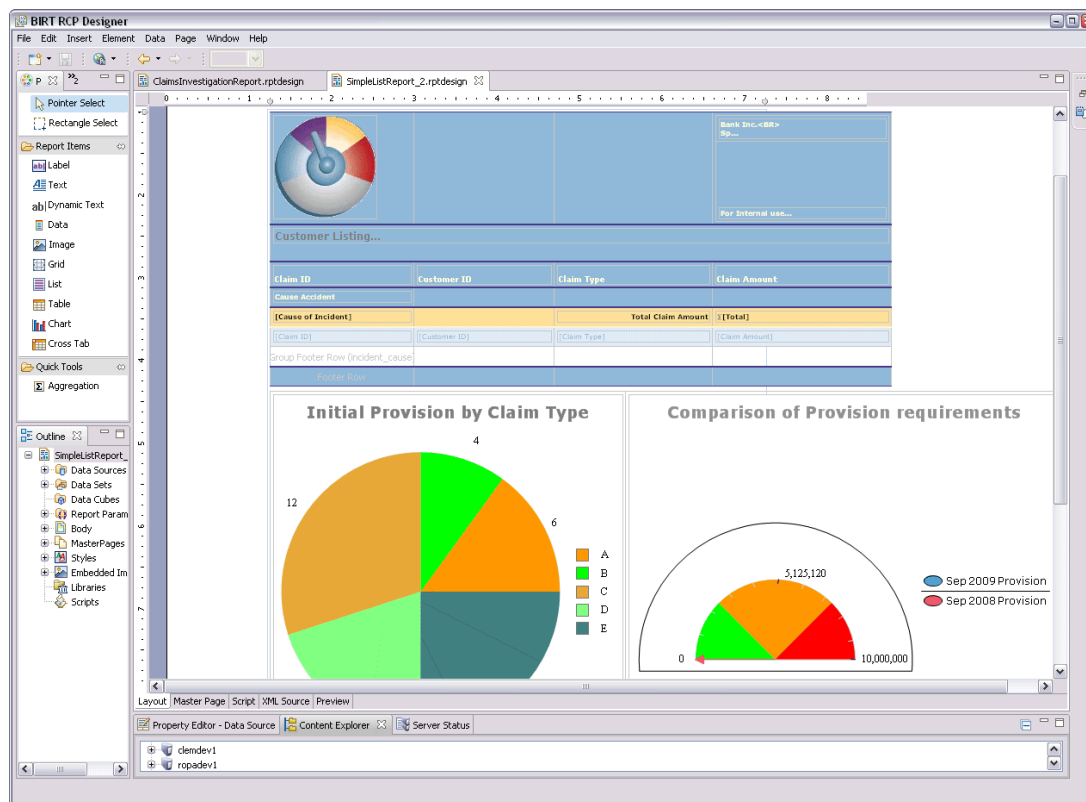
Example report

An example of a report produced using BIRT reporting tools is shown below.

Figure 11-4
BIRT-produced report example

To produce this, the basic report layout was created in BIRT Report Designer for IBM® SPSS® and then accessed and run from within the Reports tab.
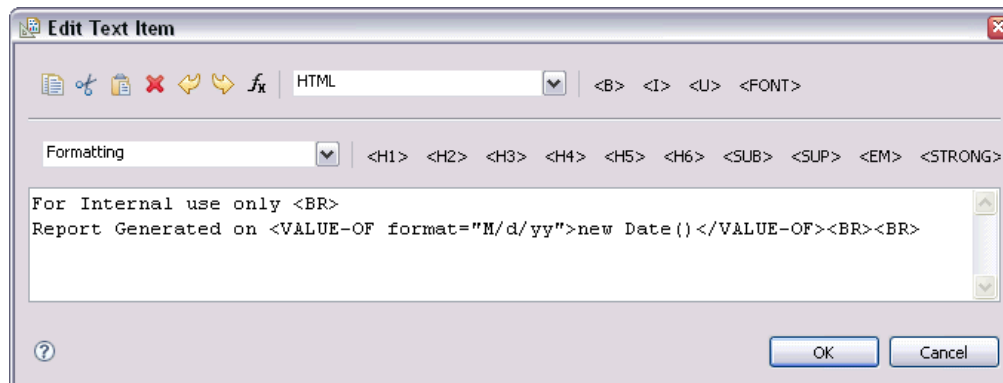
Figure 11-5
BIRT RCP Designer

For full details of how to use BIRT Report Designer for IBM SPSS, please see the documentation supplied with it. Some of the main parts that were used to create the report are listed below.
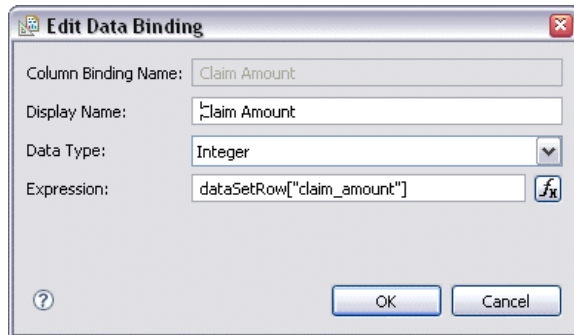
► Basic text is entered to give the report title details.

Figure 11-6
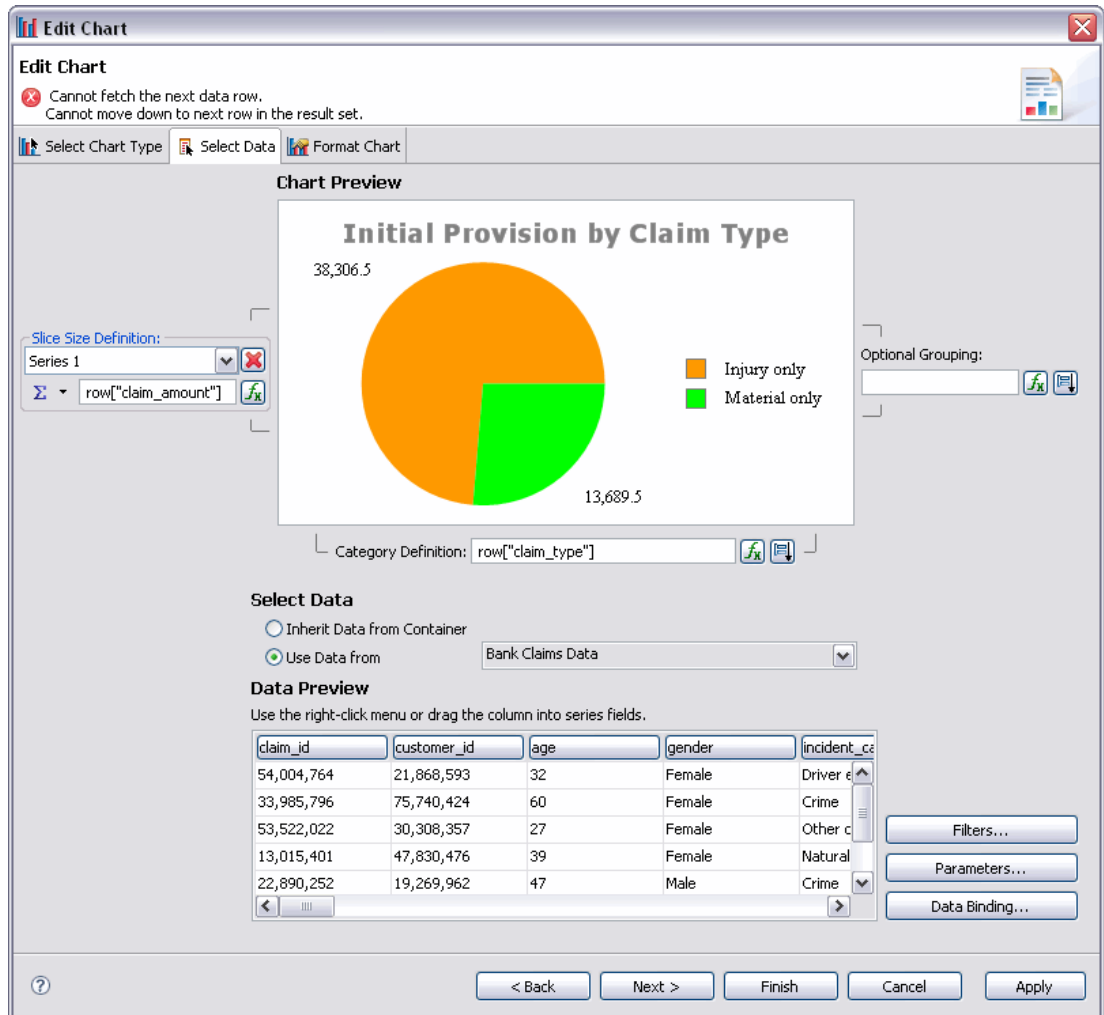*Text entry, including automatic date entry*

► The relevant data items are dragged from the data set and dropped onto the report layout.
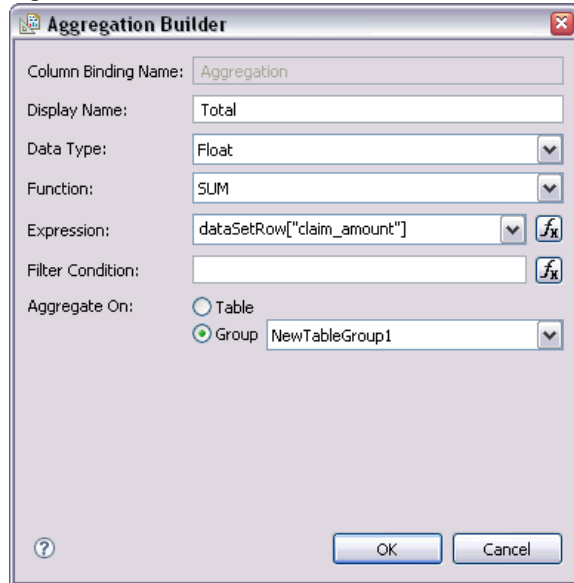
Figure 11-7
Data item used in the report

▶ Graphic chart elements are added to the report layout; the data to be displayed in them is selected.

Figure 11-8
Chart data selection

▶ Expressions and mathematical calculations are added as required. For example, to aggregate the values of each insurance claim type.
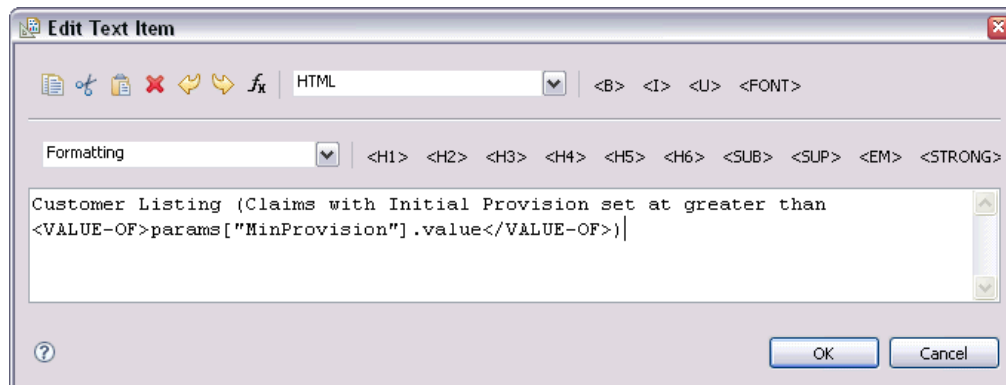
Figure 11-9

▶ Any required parameters are created and added to the report. For example, when the report is run the user is prompted to enter the value of financial provision that they are interested in.

Figure 11-10
Parameter details added to user prompt

Administering applications

Administrators can control the way business users interact with IBM® SPSS® 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.

▶ Open Deployment Manager.

▶ From the Tools menu, select Server Administration.

▶ On the Server Administration tab, connect to your server.

▶ Under Users and Groups, select Local User Repository.

▶ 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 SPSS Decision Management applications at your organization.

▶ Go to Roles and click New Role to create a new administrative role and a new business user role. For example:

■ SPSS Decision Management administrator.

■ SPSS Decision Management user.

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

Role	Actions Required
Decision Management administrator	Access Contents and Folders Decision Management Administration Run Report Dynamically Score Model Submit Work Show All Versions
Decision Management user	Access Contents and Folders 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.

▶ Select the Decision Management administrator role and click Edit Users and Groups. Add the Decision Management Administrators group you just created.

▶ Select the Decision Management user role and click Edit Users and Groups. Add the Decision Management Users group you just created.

Figure 12-1
Decision Management Administration action in IBM SPSS Collaboration and Deployment Services Deployment Manager
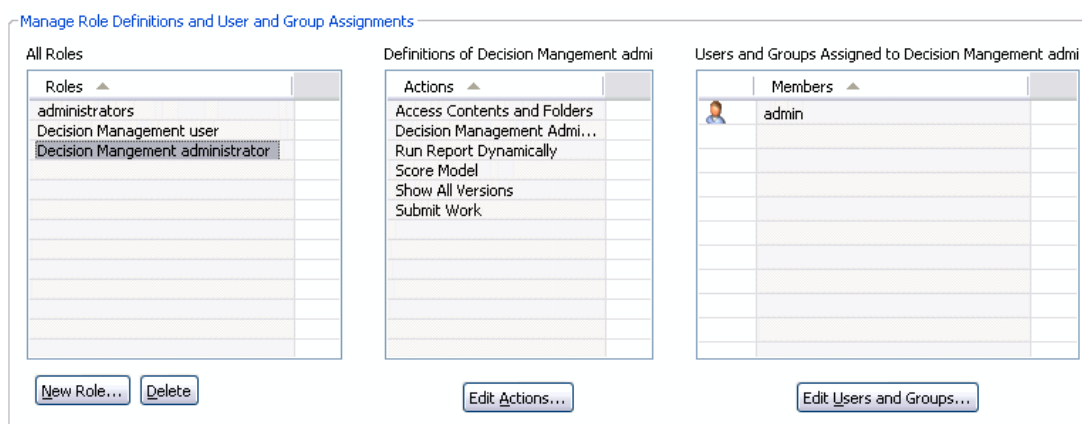
General options

Certain general options are available to administrators on all tabs of the IBM® SPSS® 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> within the Application, where <tab name> is the name of a tab. This hides the entire tab from the business user. 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 within the Application.

■ 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. Clicking the 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:

▶ Select In default launcher for applications you want all users to see on their launch page.

▶ 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. For more information, see the topic Launching applications in Chapter 3 on p. 23.

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.

Figure 12-2
Hiding an application home page

☑ Hide application home page

Default tab visible after opening | Data ▾
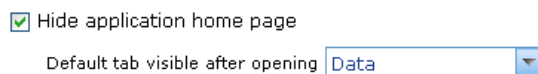
▶ On the Home page, select Hide Application home page.

▶ 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.

Figure 12-3
Administrator options on the Data tab
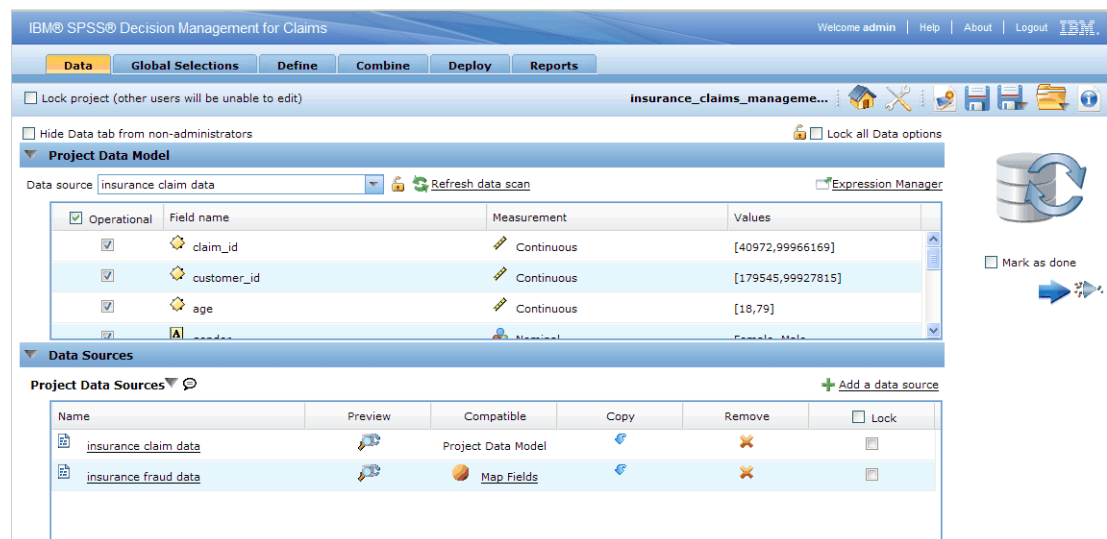
Note that business users can copy a locked data source from Project Data Sets to their My Data Sets section, but it remains locked in both sets. For more information, see the topic Managing data sources in Chapter 4 on p. 36.

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.

Figure 12-4
Administrator options on the Global Selections tab

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 section of 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.

Figure 12-5
Defining interaction points
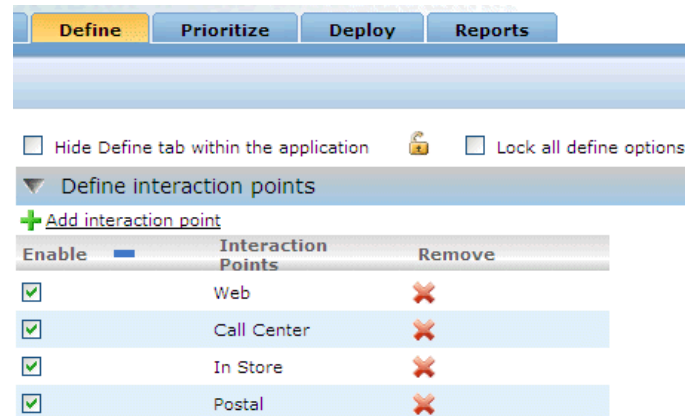
Locking constraint and interaction point options

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 Production Process Labels section is only visible to administrators (as indicated by the gray title bar). Business users don't have access to these options.

Figure 12-6
Labels for use in production process
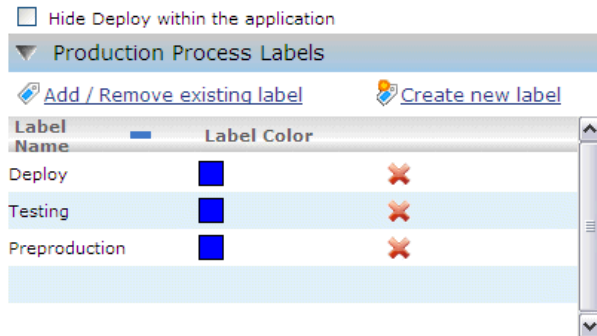
Administrators can set, modify, or remove production process labels available to business users in an application by:

■ 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:

► Click Add/Remove Existing Label. The Add/Remove Production Process Labels dialog opens.

Figure 12-7
Add/Remove production process labels
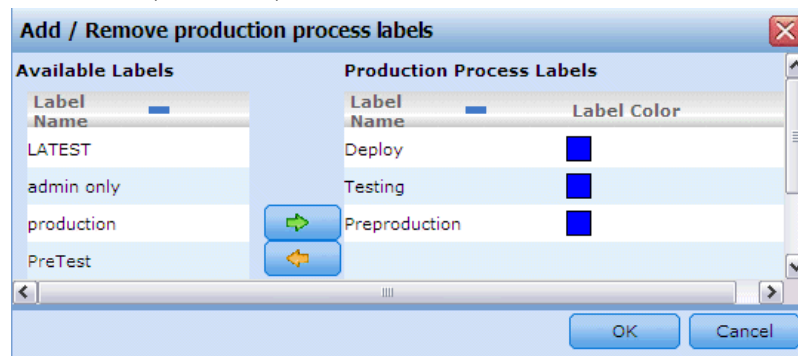
► Select labels in the Available Labels column and move them to the Production Process Labels column to make them available for use in the production process. Move labels back to the Available Labels column to remove them from the production process. Only labels appearing in the Production Process Labels 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.

▶ Assign a color to each label used in the production process. When finished, click OK.

To create a new label:

Click Create New Label. The New Label dialog opens.

Figure 12-8
New label
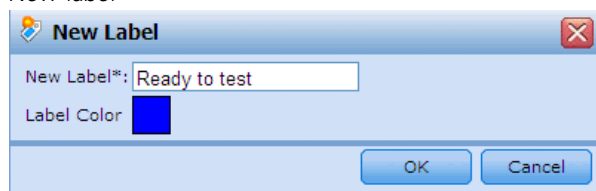
Type a name for the new label, assign a color to it, and click OK.

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.

Figure 12-9
Interactive questions
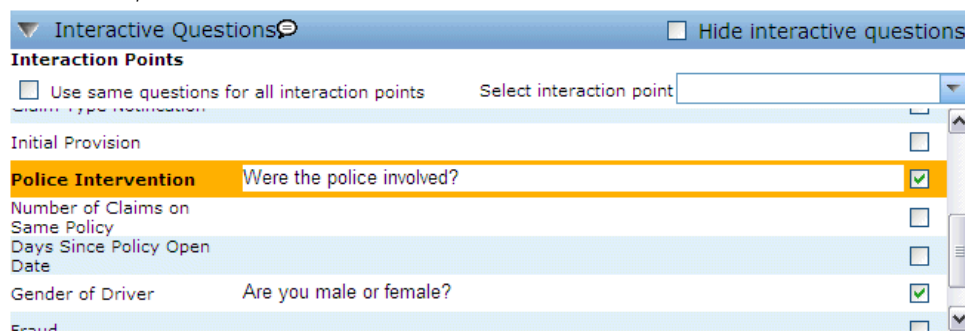
Locking scoring destination options

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.

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.

Report used for homepage results summary | Campaign Details Report | ▼ | 🔓

▶ 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.

Title of homepage results summary panel | Current results | 🔓

Downloading project metadata for use in creating external rules

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 SPSS Decision Management project, and can be deployed as a Web service for use the application. To accomplish this, the SPSS Decision Management administrator can download a *.ZIP* file with metadata for the current project. The downloaded file is saved locally on the system used by the administrator to trigger the download.

Figure 12-10
Download metadata icon

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 SPSS Decision Management project. This file can be imported into an external development tool such as ILOG Rules Studio 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 SPSS 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.

Figure 12-11
Downloading project metadata
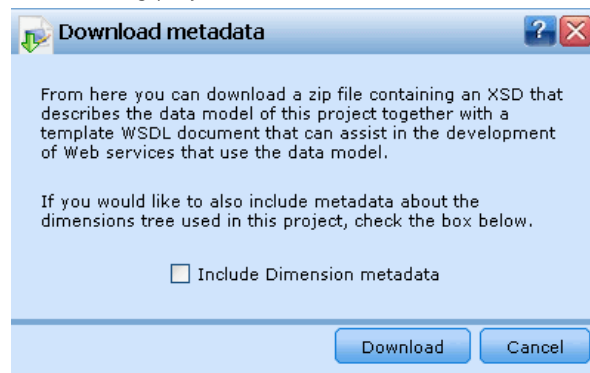
For information on creating external rules for use in SPSS Decision Management, see the *Application Designers Guide*.

Sample files

Many IBM® SPSS® Decision Management applications ship with data and other sample files 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\Modeler\14.2Demos*). These files may be used with prebuilt sample applications, and include:

- *insurance_claim_data.txt*
- *insurance_fraud_data.txt*
- *bank_customer_data.txt*
- *bank_response_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, the data source paths will need to be updated accordingly.

Samples

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

► **DVD.** Copy the files from the *\Demo* folder on the DVD to a temporary location, for example *c:\temp\ClaimsManagement*. If samples are packaged in a *.zip* file, be sure to extract all files before proceeding.

   or

► **Downloaded file.** Download the desired sample file for the application, for example *SPSS_Decision_Mgmt_61_Claims_demo.zip*. Extract all files before proceeding.

► Open IBM® SPSS® Collaboration and Deployment Services Deployment Manager and log on to your Content Server connection.

► In Deployment Manager, create a new folder for the application, for example *ClaimsManagement*.

► In Deployment Manager, right-click on the folder you created in the previous step and select Add File to Repository.

► Browse to the temporary location where you copied the sample files, for example *c:\temp\ClaimsManagement*. Select all files in the folder, and then click Open. The files will be added to the folder in IBM SPSS Collaboration and Deployment Services Repository. (If demo files were packaged in a *.zip* file, be sure to extract all files before uploading.)

Note that all sample application files, models, and rules assume that data files are installed under the IBM® SPSS® Modeler Server   *\Demos* directory (for example, *C:\Program Files\IBM\SPSS\Modeler\14.2Demos*). If data files are installed to a different drive or location, the data source paths will need to be updated for the local configuration.

Help and accessibility

Getting help

Three levels of user-assistance are provided:

■ Custom help is provided for each application packaged with IBM® SPSS® 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.

Accessibility

Because IBM® SPSS® Decision Management is a browser-based application accessibility options for disabled users, such as those who are visually impaired, are controlled from your browser settings. For details about user interface actions, including keyboard shortcuts, screen readers, and so on, please refer to your browser's help or user guide.

Help accessibility

An alternate version of the user help system is available to provide additional support for screen readers. To switch to the alternate version you will need to run a batch file to convert the existing help system into a format more efficiently read by a screen reader. In the alternate version the table of contents is on the right and help topics display on the left.

To enable support for screen readers

▶ Navigate to the root directory of the help on the server (for example, *C:\Program Files\IBM\SPSS\Collaboration and Deployment Services\4.2\help\en\DecisionManagement\ClaimsManagment\userhelp*).

▶ To switch to the screen reader-friendly format double-click the batch file *format-for-screen-readers.bat* to run it.

To switch back to the default format run the batch file *restore-default-format.bat*.

Sharing streams between IBM SPSS Decision Management and IBM SPSS Modeler

Models and projects created in IBM® SPSS® 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 SPSS Decision Management application, then customize the stream in SPSS Modeler before deploying through SPSS Decision Management. Alternatively, streams created in SPSS Modeler can be opened in SPSS Decision Management, provided deployment settings have been specified as described below.

Opening streams in IBM SPSS Decision Management

When opening streams in SPSS 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 SPSS Decision Management (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 SPSS Decision Management, and the selected modeling node is either an automated modeling node (supported for automated modeling) or a Decision List node (supported for interactive model building). 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 SPSS Modeler, deployment settings are specified in the Stream Properties dialog box, as follows:

▶ From the SPSS Modeler menus, choose:
Tools > Stream Properties > Deployment

▶ From the Deployment type list, choose Scoring Only or Model Refresh as appropriate.

▶ Select a Scoring node, Modeling node, and Model node as needed.

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.

124

For streams created in IBM SPSS Modeler Advantage or SPSS Decision Management, deployment information is set automatically.

Tips for using IBM SPSS Modeler streams in IBM SPSS Decision Management

- It may be easier to create streams using IBM SPSS Modeler Advantage or SPSS Decision Management and then modify in SPSS Modeler, rather than building from scratch using SPSS Modeler. Streams created in this manner will be structured in a manner that allows them to be fully functional in SPSS 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 Expression Manager. Note that local rules can only be created within SPSS Decision Management or IBM SPSS Modeler Advantage and not by 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 SPSS Modeler splash screen, you can choose to create a model with IBM SPSS Modeler Advantage.

Notices

This information was developed for products and services offered worldwide.

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 character set (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 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 wish 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, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

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, ibm.com, and SPSS are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at *http://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.

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 of Sun Microsystems, Inc. in the United States, other countries, or both.

SAS is a registered trademark of SAS Institute Inc. in the United States, other countries, or both.

Other product and service names might be trademarks of IBM or other companies.

Index