



SQL Server™ 2005: Data Mining

Table of Contents

SQL Server™ 2005 Data Mining	3
Lab Setup	4
Exercise 1 Creating Decision Tree and Naïve Bayes Data Mining Models	6
Exercise 2 Viewing Mining Accuracy Charts	19
Exercise 3 Creating a Prediction Query.....	24

SQL Server™ 2005 Data Mining

Objectives

Note: This lab focuses on the concepts in this module and as a result, may not comply with Microsoft® security recommendations.

Note: All SQL Server 2005 labs are based on beta builds of the product. The intent of these labs is to provide you with a general feel of some of the planned features for the next release of SQL Server. As with all software development projects, the final version may differ from beta builds in both features and user interface. For the latest details on SQL Server 2005, please visit <http://www.microsoft.com/sql/2005/default.asp>.

After completing this lab, you will be able to:

- Create Decision Tree and Naïve Bayes Data Mining Models
- View Mining Accuracy Charts
- Create a Prediction Query
- Model Time Series

Scenario

Prerequisites

Estimated Time to Complete This Lab

75 Minutes

Lab Setup

In this part, you will set up the views you will work with in the rest of the lab.

Tasks	Detailed Steps
<p>Task 1: Log in.</p>	<p>Log on with user name Administrator and password Pass@word1.</p>
<p>Task 2: Create the Views</p>	<ol style="list-style-type: none"> From the Microsoft Windows® task bar, select Start All Programs Microsoft SQL Server 2005 CTP SQL Server Management Studio. In the Connect to Server dialog box, make sure that in the Server type drop-down list, Database Engine is selected. Enter localhost in the Server name textbox and select Windows Authentication in the Authentication drop-down list, as in Figure 1. Click Connect.  <ol style="list-style-type: none"> Select File Open File. Navigate to the C:\SQL Labs\Lab Projects\Data Mining Lab\DM Setup directory, and select the ViewCreation.sql file. Click Open. Click Connect in the Connect to Server dialog box. Execute the script by pressing F5, or by clicking on the Execute icon in the toolbar, as shown in Figure 2.

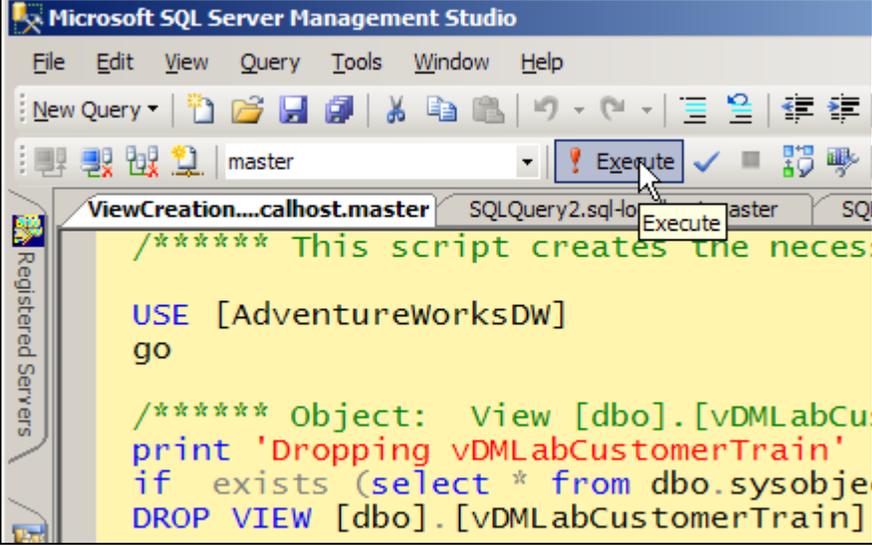
Tasks	Detailed Steps
	 <p>The screenshot shows the Microsoft SQL Server Management Studio interface. The menu bar includes File, Edit, View, Query, Tools, Window, and Help. The toolbar contains various icons for file operations and execution. The server name 'master' is displayed in the top right. The main window shows a script with the following content:</p> <pre>/****** This script creates the neces USE [AdventureWorksDW] go /****** Object: View [dbo].[vDMLabCu print 'Dropping vDMLabCustomerTrain' if exists (select * from dbo.sysobje DROP VIEW [dbo].[vDMLabCustomerTrain]</pre> <p>A mouse cursor is hovering over the 'Execute' button in the toolbar, with a tooltip that also says 'Execute'. The script text is highlighted in yellow.</p>

Figure 2: Execute Script

7. When the script has executed successfully, select the **File | Exit** menu item to close SQL Server Management Studio.

Exercise 1

Creating Decision Tree and Naïve Bayes Data Mining Models

The management at Adventure Works wants to analyze purchasing decisions based on customer demographics. Analysis Services has improved data mining functionality, providing the following data mining techniques:

- Microsoft Association Rules
- Microsoft Clustering
- Microsoft Decision Trees
- Microsoft Naïve Bayes
- Microsoft Neural Network
- Microsoft Sequence Clustering
- Microsoft Time Series

In this exercise, you will develop an Analysis Services solution using the Microsoft Business Intelligence Development Studio environment. The Business Intelligence Development Studio is an environment based on the Microsoft Visual Studio® 2005 environment.

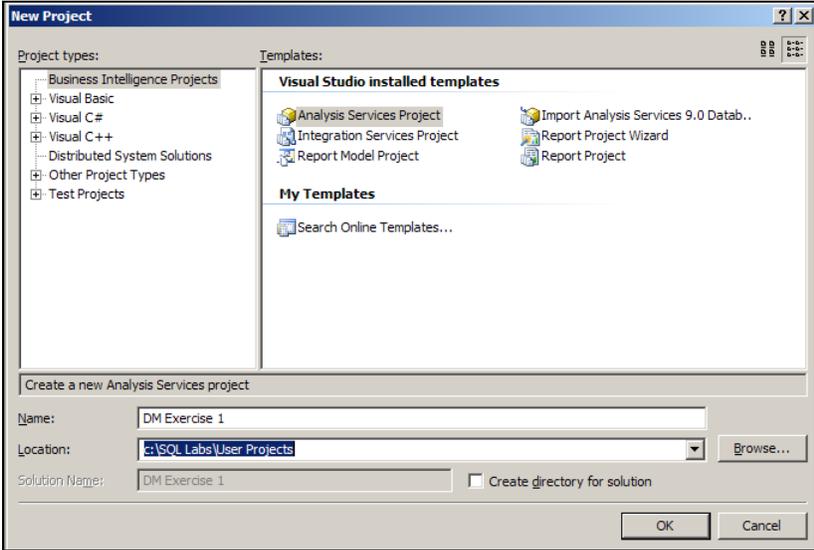
Business Intelligence Development Studio provides you with an integrated development environment for designing, testing, editing, and deploying projects to the Analysis Server. You will create and view a data mining structure with Decision Trees and Naïve Bayes data mining models using the AdventureWorksDW customer data.

To create and view data mining models, you will:

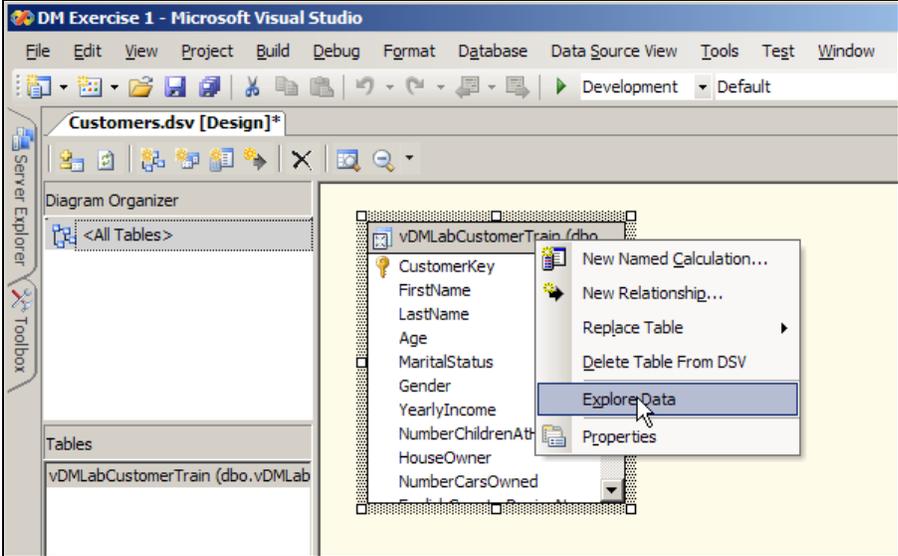
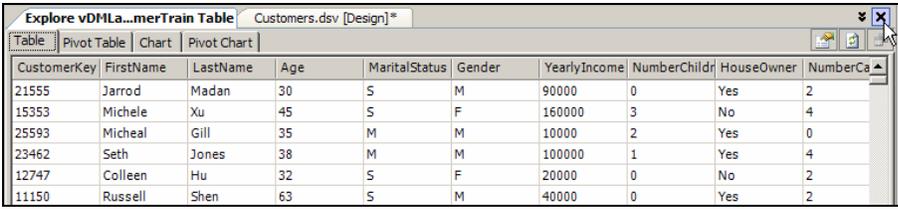
- Create an Analysis Services project in the Business Intelligence Development Studio environment.
- Create a data source and data source view.
- Create a data mining structure and decision trees data mining model using the Mining Model Wizard.
- Create a related mining model (Naïve Bayes) in the Mining Models view.
- Deploy the Analysis Services solution.

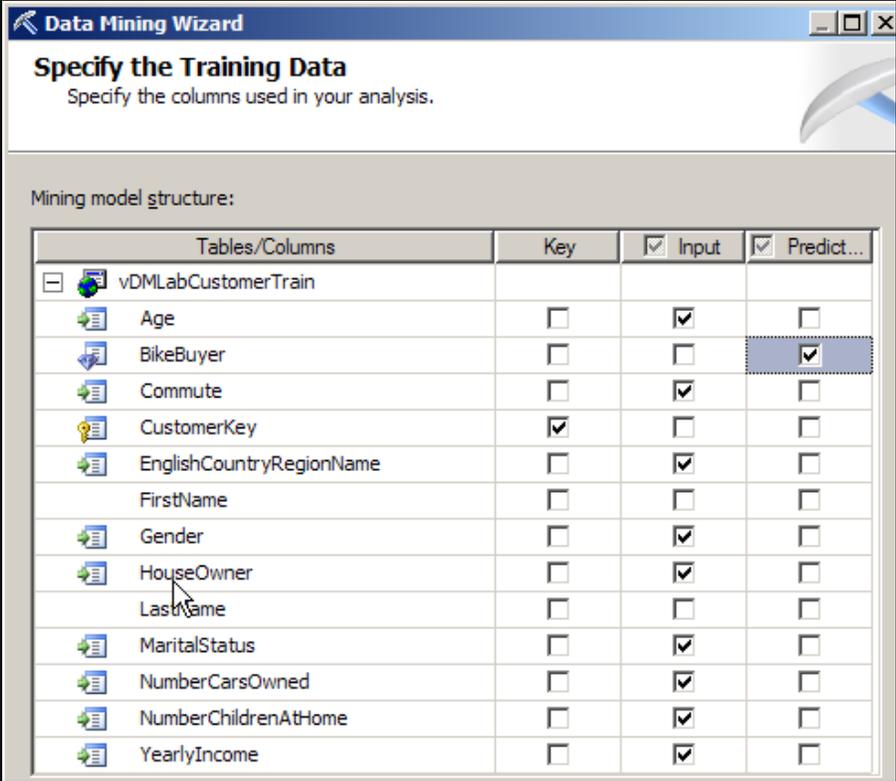
Explore the data mining models using the Mining Model Viewer

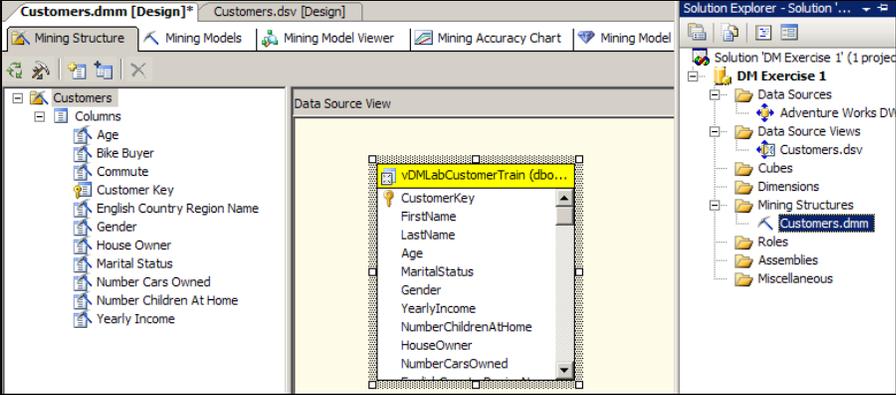
Tasks	Detailed Steps
<p>Task 1: Create an Analysis Services Project</p>	<ol style="list-style-type: none"> 1. From the Windows task bar, select Start All Programs Microsoft SQL Server 2005 CTP Business Intelligence Development Studio. 2. Select File New Project. 3. In the New Project dialog box, in the Project Types pane, click the Business Intelligence Projects folder. 4. In the Templates pane, click the Analysis Services Project icon. 5. In the Name text box, type DM Exercise 1. 6. In the Location text box, enter C:\SQL Labs\User Projects. 7. Uncheck the Create directory for Solution checkbox. Figure 1 shows how the New Project dialog box should look once you're done.

Tasks	Detailed Steps
	<p>8. Click OK.</p>  <p>Figure 1: New Project dialog box</p> <p>The project is created in a new solution; the solution is the largest unit of management in the Business Intelligence Development Studio environment. Each solution contains one or more projects. An Analysis Services Project is a group of related files containing the XML code for all of the objects in an Analysis Services database.</p> <p>You can view the solution and its projects in the Solution Explorer pane on the right-hand side in the Business Intelligence Development Studio. If the Solution Explorer is not visible, you can view it by selecting the View Solution Explorer menu item (or the keyboard shortcut Ctrl + Alt + L).</p>
<p>Task 2: Set the Deployment Mode Property</p>	<ol style="list-style-type: none"> 1. In the Solution Explorer window, right-click the DM Exercise 1 project, and select Properties from the context menu. 2. In the DM Exercise 1 Property Pages dialog box, under the Configuration Properties folder, click Deployment. 3. In the right pane, click the Deployment Mode property. In the Deployment Mode drop-down list, click DeployAll, and then click OK. <p>You can configure the build, debugging, and deployment properties of an Analysis Services project.</p>
<p>Task 3: Create a Data Source</p> <p>NOTE: If the Data connections pane already includes localhost.AdventureWorksDW, skip to step 11.</p>	<ol style="list-style-type: none"> 1. In the Solution Explorer pane, under the DM Exercise 1 project, right-click the Data Sources folder, and then select New Data Source from the context menu. 2. In the Data Source Wizard dialog box, on the Welcome to the Data Source Wizard page, click Next. 3. On the Select how to define the connection page, make sure the Create a data source based on an existing or new connection radio button is chosen. Click New ...

Tasks	Detailed Steps
	<ol style="list-style-type: none"> 4. In the Connection Manager dialog box, select the SqlClient Data Provider from the .Net Providers folder in the Provider drop-down combo box at the top of the page. 5. In the Server name drop-down list type “localhost”. 6. Under Log on to the server, click Use Windows Authentication. 7. In the Select or enter a database name drop-down list, click AdventureWorksDW. 8. Click Test Connection. 9. Click OK to dismiss the message box 10. In the Connection Manager dialog box, click OK. 11. In the Data Source Wizard dialog box, on the Select how to define the connection page, verify that localhost.AdventureWorksDW is selected, and click Next. 12. On the Completing the Data Source Wizard page, leave the default Data source name Adventure Works DW unchanged, and then click Finish. <p>You have now set up the information how to connect to the database you are working with. It is now time to define the schema information you want to use in the solution. You do this through the Data Source View.</p>
<p>Task 4: Create a Data Source View</p> <p>NOTE: At Step 3, Analysis Services may take a few moments to read the database schema.</p> <hr/> <p>NOTE: You may need to expand the Name column, and/or the entire dialog box, in order to be able to see and select vDMLabCustomerTrain.</p> <hr/> <p>NOTE: Analysis Services may take a few moments to read the data after Step 7.</p> <hr/>	<ol style="list-style-type: none"> 1. In the Solution Explorer pane, under the DM Exercise 1 database, right-click the Data Source Views folder, and then select New Data Source View from the context menu. 2. In the Data Source View Wizard dialog box, on the Welcome to the Data Source View Wizard page, click Next. 3. On the Select Data Source page, in the Relational data sources pane, verify that Adventure Works DW is selected, and then click Next. 4. In this project, your Data Source View is not going to be based on a table; instead, it will be based on a view. On the Select Tables and Views page, double-click vDMLabCustomerTrain to add this table to the Included objects list. 5. Click Next. 6. On the Completing the Wizard page, in the Name text box, type Customers and then click Finish. The Data Source View Designer will open. The Data Source View Designer is a graphical representation of the data schema you have defined. 7. Right-click the vDMLabCustomerTrain table and then click Explore Data, as in Figure 2.

Tasks	Detailed Steps
	 <p>Figure 2: Explore Data</p> <ol style="list-style-type: none"> This opens a new tab in which you can view the data for the table. If you like, you can make the tab into a dockable floating window instead. You do this by right-clicking on the tab header and choose Floating or Dockable. In the Explore vDMLabCustomerTrain Table window, scroll to view the data, and then click on the X in upper right hand corner, as in Figure 3, to close the window.  <p>Figure 3: Explore Table Window</p> <p>A Data Source View contains data source schema information. As shown here, you do not have to base the Data Source View on table(s): You can use views as well.</p>
<p>Task 5: Create a Data Mining Structure</p> <p>NOTE: The CustomerKey, FirstName, and LastName columns should not be selected as Input or Predictable columns.</p>	<ol style="list-style-type: none"> In the Solution Explorer pane, under the DM Exercise 1 database, right-click the Mining Structures folder, and then select New Mining Structure from the context menu. In the Data Mining Wizard, on the Welcome to the Data Mining Wizard page, click Next. <p>The Mining Model Wizard is the starting point for all data mining operations.</p> <ol style="list-style-type: none"> On the Select the Definition Method page, click From existing relational database or data warehouse and then click Next. On the Select the Data Mining Technique page, in the Which data mining technique do you want to use? drop-down list, verify that Microsoft

Tasks	Detailed Steps
	<p>Decision Trees is selected, and then click Next.</p> <ol style="list-style-type: none"> On the Select Data Source View page, in the Available data source views pane, verify that the Customers data source view is selected, and then click Next. On the Specify Table Types page, in the Input tables pane, in the vDMLabCustomerTrain row, verify that the Case check box is selected, and then click Next. On the Specify the Training Data page, in the Mining model structure pane, select or deselect each cell by clicking on the check box as shown in Figure 4.  <p>Figure 4: Specifying Columns for Analysis</p> <p>Because CustomerKey is the primary key of the source table, the Data Mining Wizard has automatically selected it as the key. The key identifies the cases in the mining model.</p> <ol style="list-style-type: none"> Click Next. On the Specify Columns' Content and Data Type page, click Next. On the Completing the Wizard page, in the Mining structure name text box, type Customers and check the Allow drill through check box, and then click Finish. <p>The Mining Structure designer will open as in Figure 5.</p>

Tasks	Detailed Steps
	 <p>Figure 5: The Mining Structure</p> <p>A data mining structure may contain multiple data mining models. Each data mining model uses a subset of the data referenced by the data mining structure. When the data mining structure is processed, the source data is queried once and then all of the data mining models are processed in parallel.</p>
<p>Task 6: Add and edit columns in the Mining Structure</p>	<ol style="list-style-type: none"> 1. In the Mining Structure tree view on the left side of the designer window, right-click Columns, and then click Add a Column. 2. In the Select a Column dialog box, in the Source column tree view, select the Age column, and then click OK. 3. An alert will appear indicating that you already have an Age column selected. Click Yes to approve and dismiss the dialog box. 4. In the Mining Structure tree view, right-click the Age 1 column, and then click Properties. 5. In the Properties window, in the Content property drop-down list, select Discretized. <p>By changing the Content property to Discretized, the server will automatically determine discrete ranges for the column.</p> <ol style="list-style-type: none"> 6. In the Properties window, in the Name property text box, type Age Discretized, and then press <Enter>. 7. An alert will appear confirming that you want to change the name for all related columns. Click Yes to approve and dismiss the dialog box.
<p>Task 7: Rename the Mining Model</p> <p>NOTE: The column next to the Structure column may be called something other than Customers.</p> <hr/> <p>NOTE: Step 3 renames the Decision Tree mining model, but does not rename the mining model structure.</p>	<ol style="list-style-type: none"> 1. Select the Mining Models tab to view information about the model, as in Figure 6.

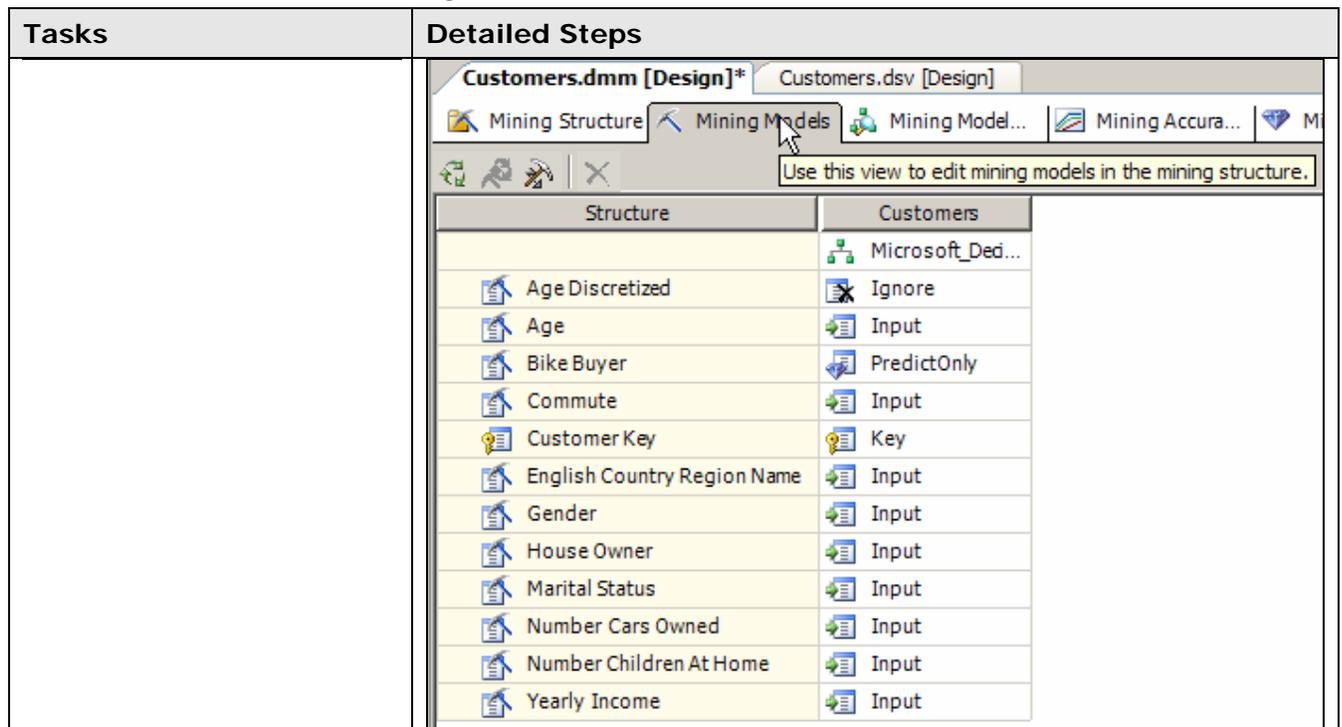


Figure 6: The Mining Models View

2. In the Mining Models grid, right-click on the second column's (the **Customers** column in Figure 6) column heading, and then click **Properties**.
3. In the **Properties** window, in the **Name** property text box, type **Customers DT** to rename the mining model, and then press **<Enter>**.

Task 8: Create a Related Mining Model

1. Click on the **Create a Related Mining Model** icon on the Mining Models icon bar, as shown in Figure 7.

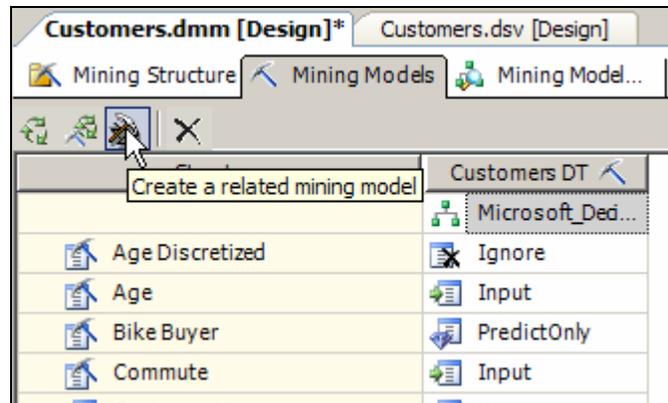
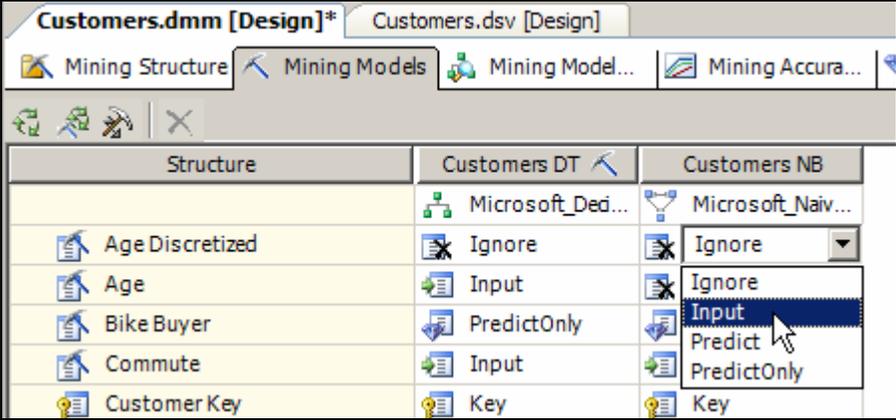
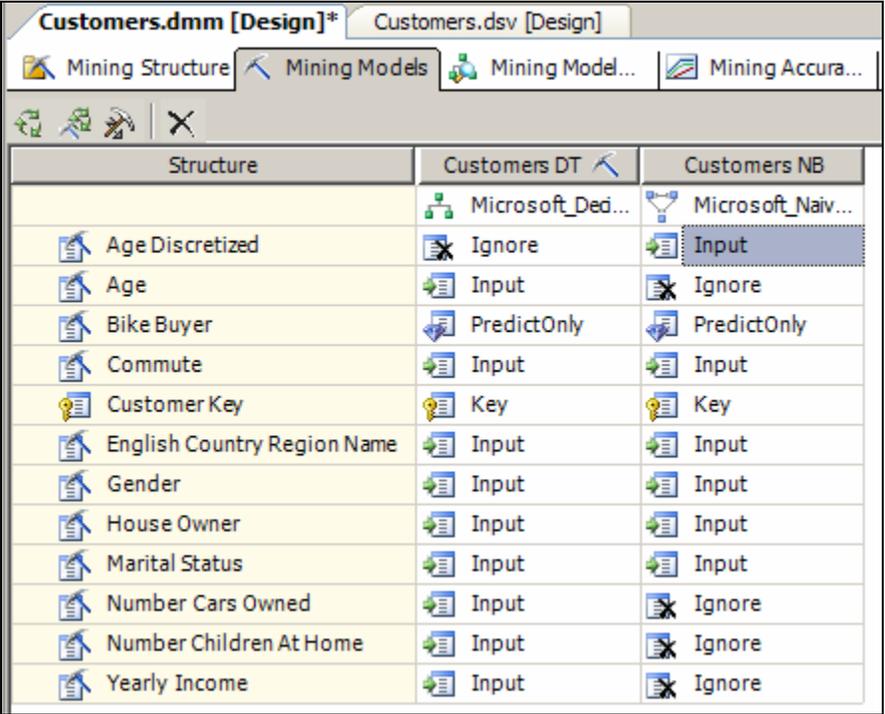
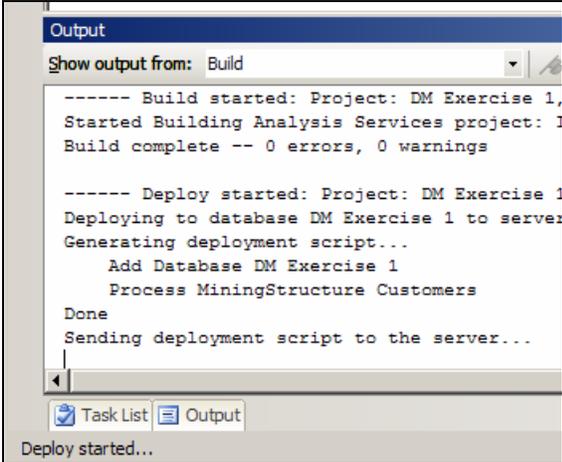
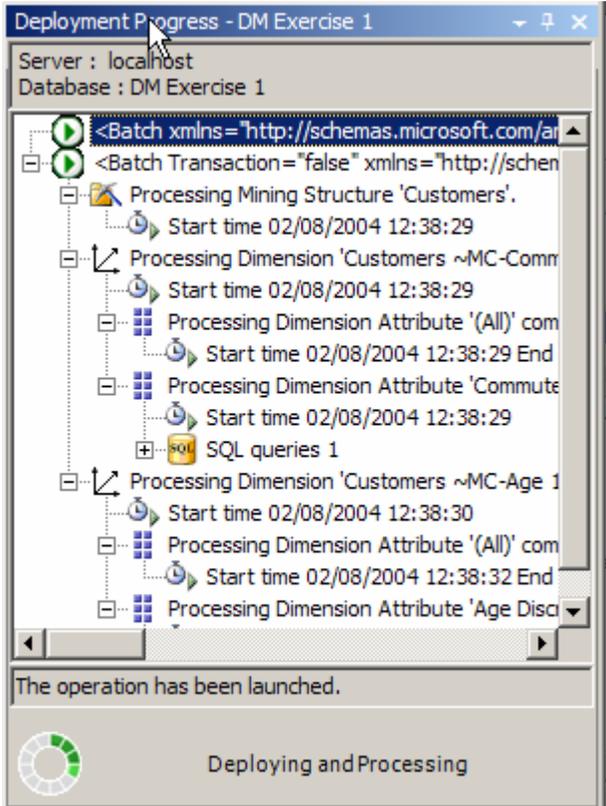
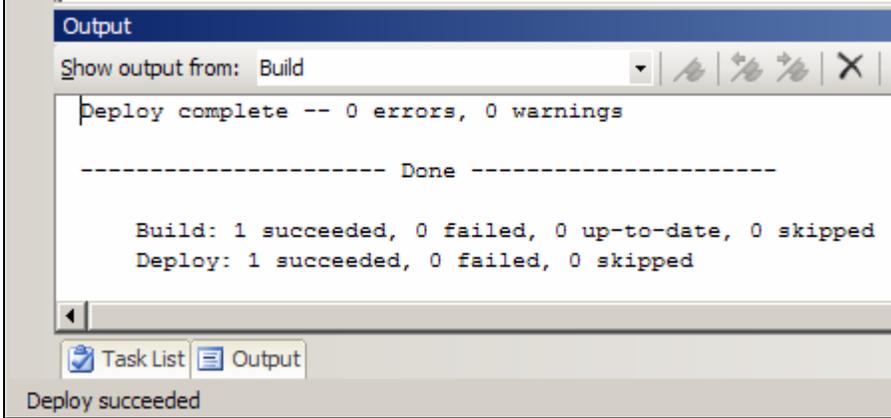


Figure 7: The Create a Related Mining Model icon

2. In the **New Mining Model** dialog box, in the **Model Name** text box, type **Customers NB**.
3. In the **Algorithm Name** drop-down list, click **Microsoft Naïve Bayes**, and click **OK**.
4. When the alert appears confirming that you want to use the Microsoft Naïve Bayes algorithm and that some columns will be ignored, click **Yes** to

Tasks	Detailed Steps
	<p>approve and dismiss the dialog box.</p> <p>The Naïve Bayes algorithm does not support continuous columns. Therefore, the Age column will be ignored in this mining model. Instead, you will use the Age Discretized column.</p> <p>5. Click in the Age Discretized cell in the Customers NB column (the content is currently Ignore) in the cell drop-down list, select Input as in Figure 8.</p>  <p>Figure 8 Changing Usage of a Mining Model Column</p> <p>6. You should now have an end result as shown in Figure 9.</p>  <p>Figure 9: The Customers Mining Model</p>
<p>Task 9: Deploy the Analysis Services Solution</p> <p>NOTE: Analysis Services may</p>	<p>1. Select the Build Deploy Solution menu item.</p> <p>An Output window will open showing you deployment progress, as shown in Figure 10. The deployment progress is also shown in the Deployment Progress pane on the right hand side of the</p>

Tasks	Detailed Steps
<p>take quite a while to process the models after Step 1.</p> <hr/> <p>Note: The Analysis Services project is automatically saved when you deploy the solution in Step 2.</p> <hr/>	<p>Business Intelligence Development Studio, as in Figure 11. The Deployment Progress pane can give you more detailed information about what happens during deployment than what you see in the Output window.</p>  <p>Figure 10: The Output Window during Deployment</p>  <p>Figure 11: The Deployment Progress Pane</p> <p>2. Once deployment is complete (as indicated by the Output window in Figure 12), close the Output and Deployment Progress windows by clicking the X in the upper-right corner of the window. The Deployment Progress pane also gives you information about completion of the deployment process.</p>

Tasks	Detailed Steps
	 <p>Figure 12: Successful Deployment</p> <p>In the above procedures, various wizards and editors have been creating XML documents based on your input. Deployment sends the XML data to the Analysis Server and which then processes the Analysis Services database.</p>

<p>Task 10: View the Customers DT Mining Model Decision Tree</p> <p>NOTE: If an alert appears indicating that changes have been made, click No.</p>	<ol style="list-style-type: none"> 1. On the tabs above the designer window, click the Mining Model Viewer tab. 2. In the Mining Model drop-down list, select Customers DT. 3. Press LeftAlt+Shift+Enter to view the designer window in full screen. (You can press LeftAlt+Shift +Enter again to return to normal view later.) 4. In the Tree drop-down list, make sure Bike Buyer is selected; Figure 13 shows the result.
--	---

Note: If accidentally closed, the Mining Model Viewer of the Mining Model Designer can be re-opened. Select the **View | Solution Explorer** menu item. In the Solution Explorer window, under the Mining Models folder, right-click **Customers.dmm**, and then select **Browse** from the context menu.

Note: The **Mining Legend** window on the right side of the display may be relocated and resized to improve the display of the decision tree. If you accidentally close the **Mining Legend** window, select the Mining Model tab and then reselect the Mining Model Viewer tab, and the **Node Legend** window will re-appear

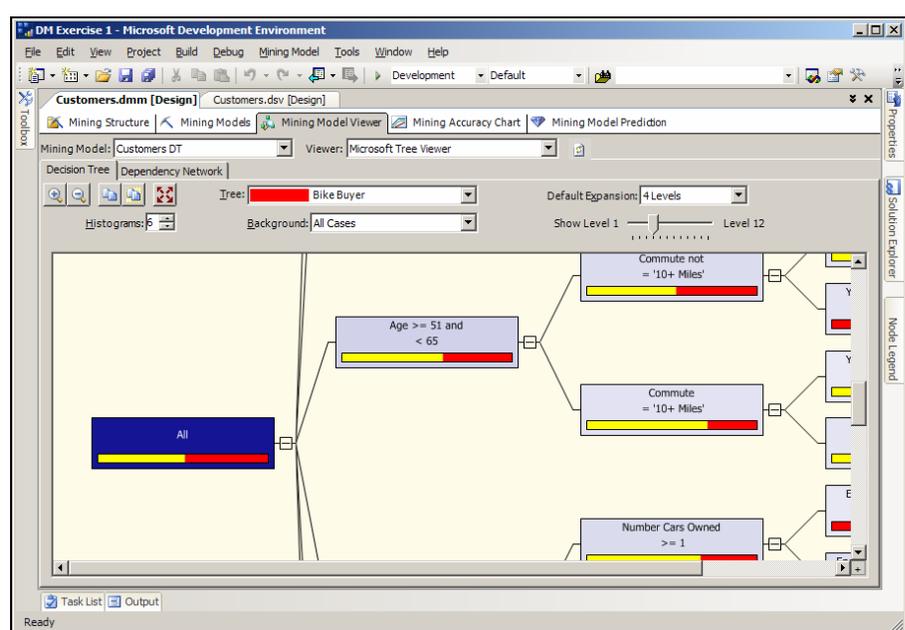
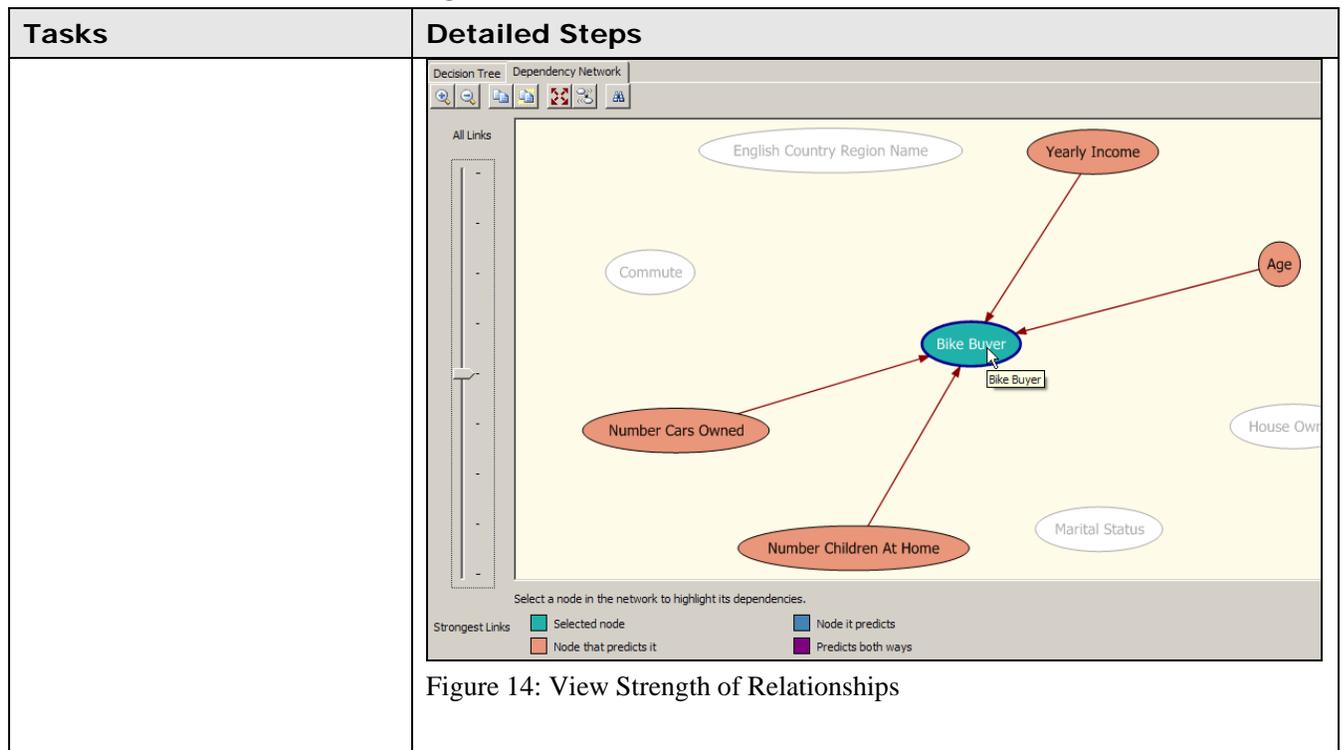


Figure 13: Browsing the Mining Model

5. If you are not in full-screen mode, in the lower-right corner of the Mining Model Viewer, click and hold on the small + icon. The mouse

Tasks	Detailed Steps
<p>when the viewer is redisplayed.</p>	<p>pointer will change to a cross-arrow icon and the Navigation window will appear. You may drag the mouse to navigate within the Mining Model Viewer.</p> <ol style="list-style-type: none"> On the Show Level slider control, drag the pointer to the left so that only one level of the decision tree is displayed. Click the All node. <p>The All node contains a histogram with red representing bike buyers and yellow representing non-bike buyers.</p> <p>Information about all customers is displayed in the Mining Legend window. Notice that 49.39% of the 18,484 customers are bike buyers. (You may need to widen the Mining Legend window in order to be able to see the percentages.)</p> <ol style="list-style-type: none"> On the Show Level slider control, drag the pointer to the right so that two levels of the decision tree are displayed. <p>Age is most predictive of a customer's bike buying behavior.</p> <ol style="list-style-type: none"> Click on each node of level 2. The Mining Legend window will display detailed information for each node. In the Background drop-down list, click Yes. <p>The shade of each node indicates the concentration of the value in the Background drop-down list. Expand and contract nodes in the diagram in order to investigate the predicting factors for each group.</p>
<p>Task 11: View the Customers DT Mining Model Dependency Network</p>	<ol style="list-style-type: none"> Within the designer, click the Dependency Network tab. <p>The Dependency Network viewer displays the strength of the relationships between the attributes in a decision tree model.</p> <ol style="list-style-type: none"> On the Links slider control, drag the pointer to the bottom. In the Dependency Network diagram, click the Bike Buyer node. <p>The color of each node indicates the attribute's relationship to the Bike Buyer attribute.</p> <ol style="list-style-type: none"> On the Links slider control, slowly drag the pointer up to the top. As you drag the pointer upward, the relationships within the data are displayed, as shown in Figure 14.



<p>Task 12: View the NB Naïve Bayes Mining Model Attribute Profile display</p>	<ol style="list-style-type: none"> 1. In the Mining Model drop-down list, click Customers NB to view the Naïve Bayes mining model. 2. Select the Attribute Profiles tab. 3. In the Predictable drop-down list, ensure that Bike Buyer is selected. <p>The Attribute Profiles tab displays the other attributes that impact the state of the predictable value selected.</p>
---	---

<p>Task 13: View the Attribute Characteristics display</p>	<ol style="list-style-type: none"> 1. Click the Attribute Characteristics tab. 2. In the Attribute drop-down list, ensure that Bike Buyer is selected. In the Value drop-down list, select Yes. <p>The characteristics of bike buyers, ordered by their frequency, are displayed.</p> <ol style="list-style-type: none"> 3. In the Value drop-down list, select No. <p>Notice that the characteristics of non-bike buyers are different than the characteristics of bike buyers.</p>
---	--

<p>Task 14: View the Attribute Discrimination display</p>	<ol style="list-style-type: none"> 1. Click the Attribute Discrimination tab. 2. In Attribute drop-down list, ensure that Bike Buyer is selected. 3. In the Value1 drop-down list, select Yes. 4. In the Value 2 drop-down list, select No. <p>The attribute values that impact a customer's bike buying decisions are displayed. The attribute values are ordered by how strongly they favor bike buyers or non-bike buyers.</p>
--	--

<p>Task 15: View the Dependency</p>	<ol style="list-style-type: none"> 1. Click the Dependency Network tab.
--	---

Tasks	Detailed Steps
Network	<ol style="list-style-type: none">2. On the Links slider control, drag the pointer to the bottom.3. In the Dependency Network diagram, click the Bike Buyer node. <p>The color of each node indicates the attribute's relationship to the Bike Buyer attribute.</p> <p>On the Links slider control, slowly drag the pointer up to the top. As you drag the pointer upward, the relationships within the data are displayed.</p>
Task 16: Close the Analysis Services Project	<ol style="list-style-type: none">1. Select File Close Solution. If prompted to save changes, select Yes.2. Select File Exit.

Exercise 2

Viewing Mining Accuracy Charts

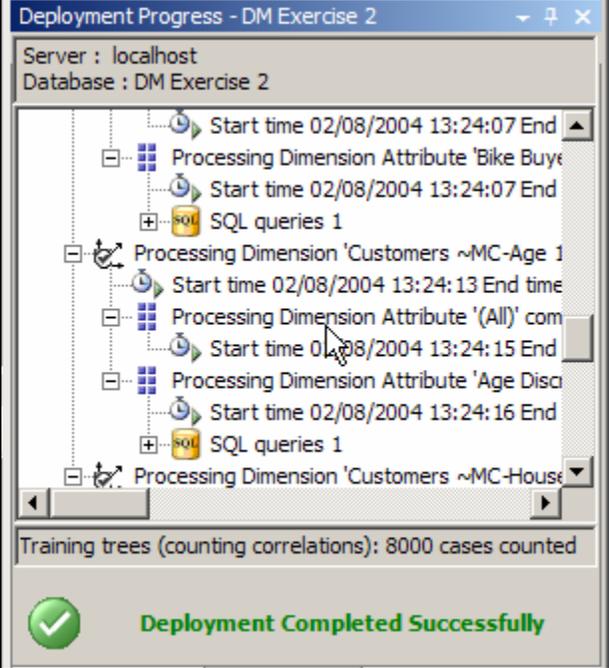
The management team at Adventure Works wants to determine the accuracy of their data mining models. Using a validation data set that was kept out of the training set, they create Mining Accuracy Charts to visually identify which model is performing most accurately.

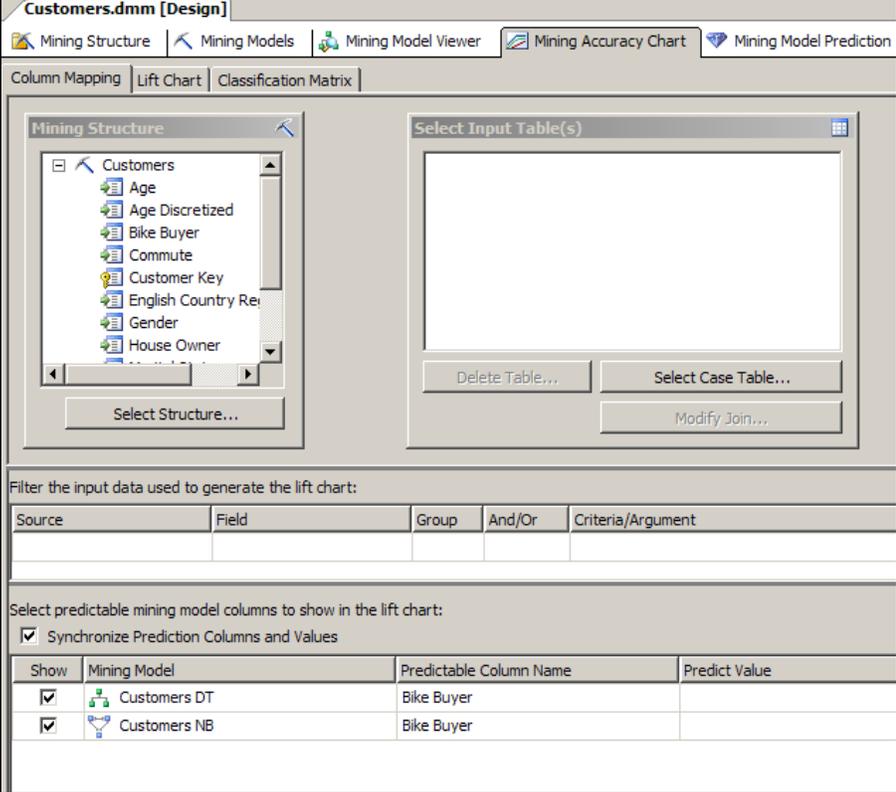
In this exercise, you will validate the mining models created in Exercise 1 by using the Mining Accuracy Chart view of the Mining Structure Designer.

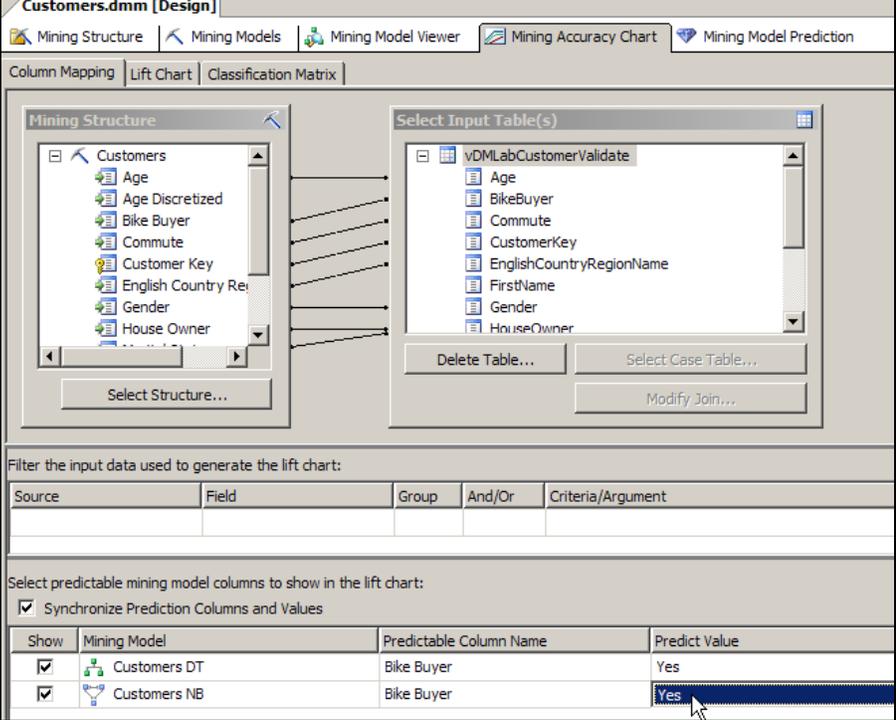
To view the Mining Accuracy Chart, you will:

- Create a prediction query by selecting an input table and mapping the columns of the data mining model to the columns in the validation data set.
- View and interpret a Lift Chart

Tasks	Detailed Steps
<p>Task 1: Open an existing project</p> <p>NOTE: The solution used in Exercise 2 is different from the solution created in Exercise 1.</p>	<ol style="list-style-type: none"> 1. From the Windows task bar, select Start All Programs Microsoft SQL Server 2005 CTP Business Intelligence Development Studio. 2. Select File Open Project/Solution. 3. In the Open Project dialog box, navigate to the C:\SQL Labs\Lab Projects\Data Mining Lab\DM Exercise 2 folder, click DM Exercise 2.sln, and then click Open.
<p>Task 2: Deploy the Analysis Services Solution</p> <p>NOTE: Analysis Services may take quite a while to process the models.</p>	<ol style="list-style-type: none"> 1. Select Build Deploy Solution. The Output window will show you deployment progress, as shown in Figure 1. The deployment progress is also shown in the Deployment Progress pane on the right-hand side of the Business Intelligence Development Studio, as in Figure 2. The Deployment Progress pane can give you more detailed information about what happens during deployment than what you see in the Output window. <div data-bbox="513 1335 1256 1717" data-label="Image"> </div> <p>Figure 1: Output Window</p>

Tasks	Detailed Steps
	 <p>Figure 2: Deployment Progress Pane</p> <ol style="list-style-type: none"> Once deployment is complete, close the Output window and the Deployment Progress pane.
<p>Task 3: Create a Prediction Query</p> <p>NOTE: If the Solution Explorer window is not visible, select the View Solution Explorer menu item.</p> <hr/> <p>NOTE: Analysis Services may take a while to process the prediction query.</p> <hr/>	<ol style="list-style-type: none"> In the Solution Explorer window, in the Mining Structures folder, double-click Customers.dmm. From the list of tabs above the designer window, select the Mining Accuracy Chart tab. The Mining Accuracy Chart view will open, displaying the Column Mapping page, as shown in Figure 3.

Tasks	Detailed Steps
	 <p>Figure 3: Column Mapping Page</p> <p>You use the Column Mapping page to design a Prediction Query. The query executes in order, comparing the mining model's predicted values to the validation data set's actual values.</p> <ol style="list-style-type: none"> On the Column Mapping tab, in the Select Input Table(s) window, click the Select Case Table button. In the Select Table window, make sure Customers is selected in the Data Source drop-down list, click the vDMLabCustomerValidate table, and then click OK. <p>Relationships between the mining structure and the input table are automatically created between columns with the same name. Relationships can be added or deleted by the user.</p> <ol style="list-style-type: none"> In the table at the bottom of the Column Mapping page, verify that the Show check box is selected for both the Customers DT and Customers NB mining models. In the Predictable Column Name column, verify that Bike Buyer is selected for both mining models. <p>In the Predictable Column Name drop-down lists, the mining model column names are restricted to columns that have the usage type set to Predict or Predict Only.</p> <ol style="list-style-type: none"> In the Predict Value column, in the drop-down list, click Yes for both mining models, as shown in Figure 4.

Tasks	Detailed Steps
	 <p>Figure 4: Predict Value</p> <p>8. Click the Lift Chart tab.</p> <p>The prediction query designed on the Column Mapping page executes. The prediction query returns a prediction for each case in the validation data set.</p> <p>The mining model will have greater accuracy in its prediction for some cases than for others. For each case, the prediction query will also return the probability that the prediction is correct.</p> <p>The cases are sorted by the probability that the prediction is correct, and then the percentages of correct predictions are displayed on the lift chart, as in Figure 5.</p>

Tasks	Detailed Steps
	<p>Figure 5: Data Mining Lift Chart</p> <p>9. Point and click at any of the lines on the chart. The Mining Legend pane will open together with a tool tip and other display information.</p> <p>As you move along different line points on a line or point to a different line, the values displayed in the Mining Legend pane and tool tip change.</p> <p>10. Point at the locations where each line reaches a Target Population of 90%, and wait for the Mining Legend pane to update.</p> <p>The example in Figure 5 shows that the Customers DT model (the blue line) only needs to select 71% of the customers in order to identify 90% of the bike buyers.</p> <p>The Ideal Model (the red line) only needs to select 45% of the customers in order to identify 90% of the bike buyers.</p> <p>The Customers NB model needs to select 85% of the customers in order to identify 90% of the bike buyers.</p> <p>The Random Guess model (the yellow line) needs to select 90% of the customers in order to identify 90% of the bike buyers.</p> <p>Because the Customers DT mining model needs to select fewer cases in order to identify a given percentage of bike buyers, it is deemed more accurate than the Customers NB model.</p>
<p>Task 4: Close the Analysis Services Project</p>	<ol style="list-style-type: none"> 1. Select File Close Solution. If prompted to save changes, select Yes. 2. Select File Exit.

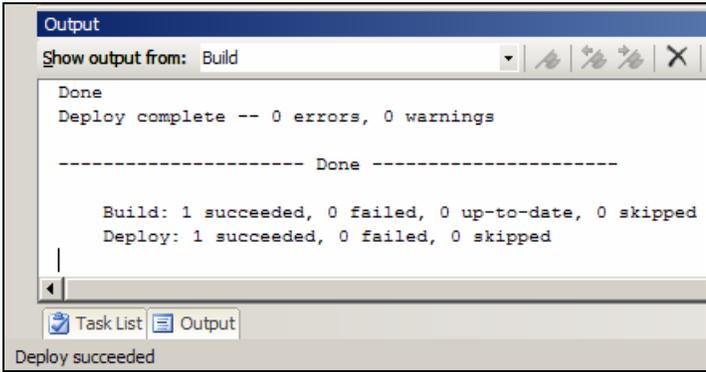
Exercise 3 Creating a Prediction Query

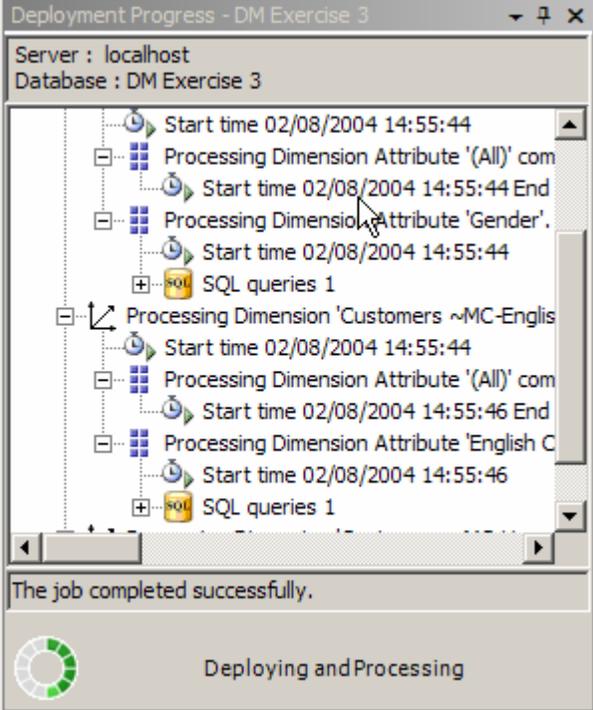
The sales and marketing department at Adventure Works received a potential customer list containing demographic data. The department will use the Customers DT data mining model to predict the likelihood that an individual in the list is a bike buyer.

In this exercise, you will make predictions by using the Mining Model Prediction view of the Mining Structure Designer.

To make predictions, you will:

- Select an input table and map the columns of the data mining model to the columns in the prediction data set.
- Create a prediction query in the Prediction Query Builder Design pane.
- View the prediction query results.

Tasks	Detailed Steps
<p>Task 1: Open an existing project</p> <p>NOTE: The solution used in Exercise 3 is different than the solutions examined in the previous exercises.</p>	<ol style="list-style-type: none"> 1. From the Windows task bar, select Start All Programs Microsoft SQL Server 2005 CTP Business Intelligence Development Studio. 2. Select File Open Project/Solution. 3. In the Open Project dialog box, navigate to the C:\SQL Labs\Lab Projects\Data Mining Lab\DM Exercise 3 folder, click DM Exercise 3.sln, and then click Open.
<p>Task 2: Deploy the Analysis Services Solution</p> <p>NOTE: Analysis Services may take quite a while to process the models after Step 1.</p>	<ol style="list-style-type: none"> 1. Select Build Deploy Solution. The Output window will indicate deployment progress, as shown in Figure 1. The deployment progress is also shown in the Deployment Progress pane on the right hand side of the Business Intelligence Development Studio, as in Figure 2. The Deployment Progress pane can give you more detailed information about what happens during deployment than what you see in the Output window.  <p>Figure 1: Output Window</p>

Tasks	Detailed Steps
	 <p>Figure 2: Deployment Progress Pane</p> <ol style="list-style-type: none"> Once deployment is complete, close the Output window and the Deployment Progress pane.
<p>Task 3: Create a Prediction Query using the Decision Tree Mining Model</p> <p>NOTE: If the Solution Explorer pane is not visible, select View Solution Explorer.</p>	<ol style="list-style-type: none"> In the Solution Explorer window, in the Mining Structures folder, double-click Customers.dmm. From the list of tabs above the designer window, select the Mining Model Prediction tab. In the Mining Model window, click the Select Model... button. In the Select Mining Model dialog box, expand Customers, click Customers DT, as shown in Figure 3, and then click OK.

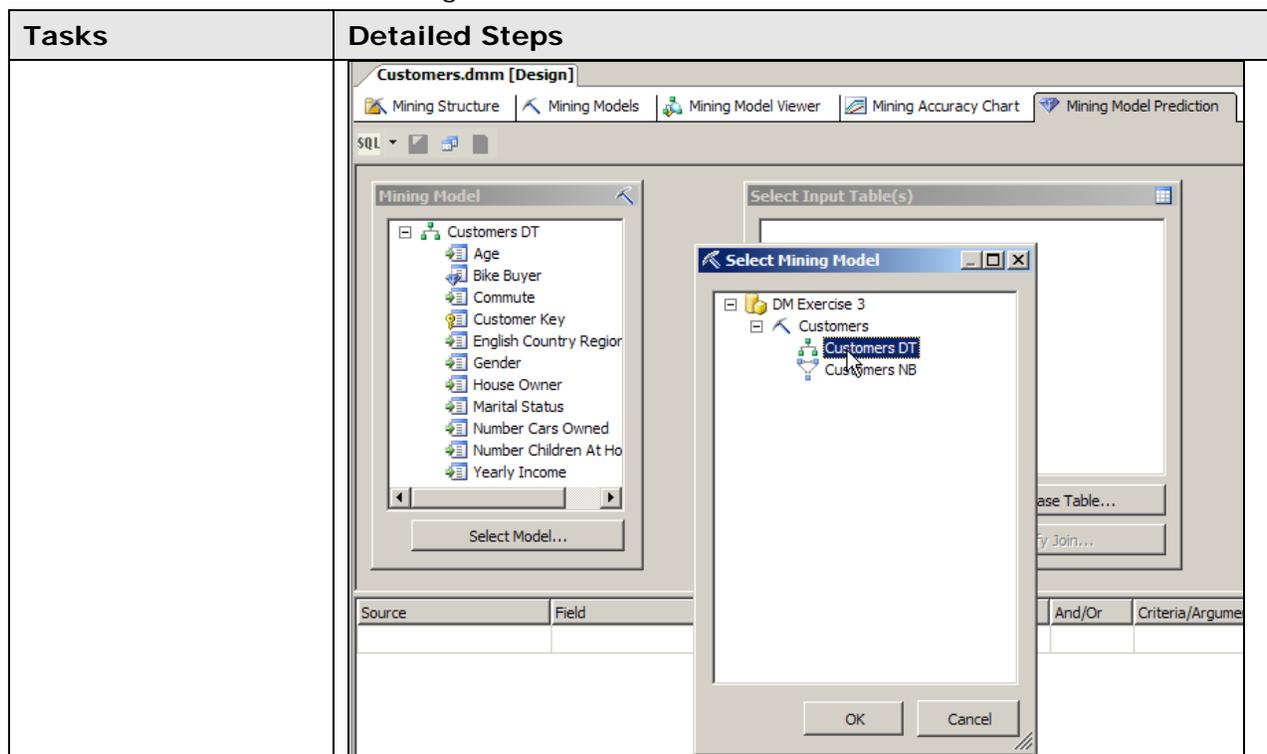


Figure 3: Select Mining Model in Mining Model Prediction View

5. In the **Select Input Table(s)** window, click the **Select Case Table...** button.
6. In the **Select Table** window, expand **Customers**, click the **vDMLabCustomerPredict** table, and then click **OK**.

Relationships between the mining structure and the input table are automatically created between columns with the same name. Relationships can be added or deleted by the user.

Task 4: Build the Decision Tree Prediction Query

NOTE: The columns of the table may be re-sized by dragging the dividing line between the column headings.

NOTE: The value in the Source column will change from Customers DT mining model to Customers DT.

NOTE: You can also do Step 4 by clicking on the down arrow in the **Switch to query result view** icon

1. Enter values into the first row of the table at the bottom of the designer as shown below:

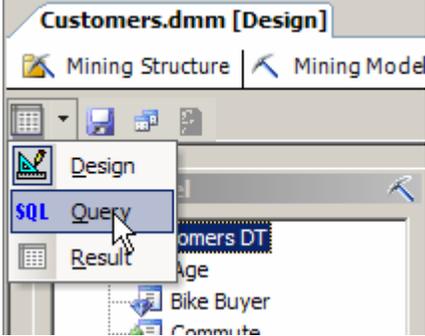
Column	Value
Source	vDMLabCustomerPredict
Field	CustomerKey
Show	(Checked)

2. Input values into the second row of the table as shown below:

Column	Value
Source	Customers DT mining model
Field	Bike Buyer
Show	(Checked)

3. Enter values into the third row of the table as shown below:

Column	Value
Source	Prediction Function

Tasks	Detailed Steps																																											
<p>in the upper left corner of the Mining Model Prediction View as in Figure 5, and then click Query.</p>	<table border="1" data-bbox="513 199 1422 367"> <tr> <td>Field</td> <td>PredictProbability</td> </tr> <tr> <td>Alias</td> <td>Confidence</td> </tr> <tr> <td>Show</td> <td>(Checked)</td> </tr> <tr> <td>Criteria/Argument</td> <td>[Customers DT],[Bike Buyer]</td> </tr> </table> <p>Figure 4 shows what you should see after having entered the values as described in the previous tables.</p> <table border="1" data-bbox="513 464 1422 604"> <thead> <tr> <th>Source</th> <th>Field</th> <th>Alias</th> <th>Show</th> <th>Group</th> <th>And/Or</th> <th>Criteria/Argument</th> </tr> </thead> <tbody> <tr> <td>vDMLabCustomerPre...</td> <td>CustomerKey</td> <td></td> <td><input checked="" type="checkbox"/></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Customers DT</td> <td>Bike Buyer</td> <td></td> <td><input checked="" type="checkbox"/></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Prediction Function</td> <td>f PredictProbability</td> <td>Confidence</td> <td><input checked="" type="checkbox"/></td> <td></td> <td></td> <td>[Customers DT],[BikeBuyer]</td> </tr> <tr> <td></td> <td></td> <td></td> <td><input type="checkbox"/></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>Figure 4: Values for the Decision Tree Prediction Query</p> <p>4. Select Mining Model Query to view the SQL view of the query.</p>  <p>Figure 5: Switch to SQL Query View</p> <p>In the bottom pane of the Mining Model Prediction view, the text of the prediction query is displayed.</p>	Field	PredictProbability	Alias	Confidence	Show	(Checked)	Criteria/Argument	[Customers DT],[Bike Buyer]	Source	Field	Alias	Show	Group	And/Or	Criteria/Argument	vDMLabCustomerPre...	CustomerKey		<input checked="" type="checkbox"/>				Customers DT	Bike Buyer		<input checked="" type="checkbox"/>				Prediction Function	f PredictProbability	Confidence	<input checked="" type="checkbox"/>			[Customers DT],[BikeBuyer]				<input type="checkbox"/>			
Field	PredictProbability																																											
Alias	Confidence																																											
Show	(Checked)																																											
Criteria/Argument	[Customers DT],[Bike Buyer]																																											
Source	Field	Alias	Show	Group	And/Or	Criteria/Argument																																						
vDMLabCustomerPre...	CustomerKey		<input checked="" type="checkbox"/>																																									
Customers DT	Bike Buyer		<input checked="" type="checkbox"/>																																									
Prediction Function	f PredictProbability	Confidence	<input checked="" type="checkbox"/>			[Customers DT],[BikeBuyer]																																						
			<input type="checkbox"/>																																									
<p>Task 5: Display the Decision Tree query result</p> <p>NOTE: As with the SQL view, you can also view the result by clicking the down arrow in the Switch to query result view icon in the upper left corner of the Mining Model Prediction View, as in Figure 4, and then click Result.</p>	<ul style="list-style-type: none"> You view the result of the Decision Tree Query result by selecting Mining Model Result. <p>The results of the prediction query are displayed.</p> <p>The CustomerKey column identifies each record from the input table.</p> <p>The Bike Buyer column contains the mining model's prediction of the customer's bike buying behavior.</p> <p>Larger values in the Confidence column mean the mining model has more accuracy in the prediction contained in the Bike Buyer column.</p> <p>By contacting potential customers predicted to be bike buyers with a high Confidence value, Adventure Works can use the results of the prediction query to promote their bikes to those individuals most likely to be bike buyers. Adventure Works marketing expenses will be reduced because potential customers who are not likely to be bike buyers will not be contacted.</p>																																											
<p>Task 6: Close the Analysis Services</p>	<ol style="list-style-type: none"> 1. Select File Close Solution. If prompted to save changes, select Yes. 2. Select File Exit. 																																											

