About this tutorial

This tutorial was extracted from the manual Using the Intelligent Miner for Data, which is delivered with the IBM DB2 Intelligent Miner for Data Version 6.1. The author slightly altered the original text. For example, references to other sections in the manual were deleted. To order the Using the Intelligent Miner for Data manual separately, contact your IBM representative.
## Contents

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>About this tutorial</td>
<td>iii</td>
</tr>
<tr>
<td>Tutorial</td>
<td>1</td>
</tr>
<tr>
<td>Before you start</td>
<td>1</td>
</tr>
<tr>
<td>The business problem</td>
<td>1</td>
</tr>
<tr>
<td>The mining run tasks</td>
<td>2</td>
</tr>
<tr>
<td>Starting the Intelligent Miner in demo mode</td>
<td>3</td>
</tr>
<tr>
<td>Using the demonstration data on AIX servers</td>
<td>3</td>
</tr>
<tr>
<td>Using the demonstration data on OS/390 servers</td>
<td>4</td>
</tr>
<tr>
<td>Using the demonstration data on AS/400 servers</td>
<td>4</td>
</tr>
<tr>
<td>Using the demonstration data on Sun Solaris servers</td>
<td>6</td>
</tr>
<tr>
<td>Using the demonstration data on Windows NT servers</td>
<td>7</td>
</tr>
<tr>
<td>The Intelligent Miner main window</td>
<td>8</td>
</tr>
<tr>
<td>Defining a data object</td>
<td>9</td>
</tr>
<tr>
<td>Setting the mode parameters</td>
<td>14</td>
</tr>
<tr>
<td>Specifying the input fields</td>
<td>15</td>
</tr>
<tr>
<td>Specifying advanced parameters</td>
<td>16</td>
</tr>
<tr>
<td>Specifying other parameters</td>
<td>16</td>
</tr>
<tr>
<td>Specifying the result object name</td>
<td>17</td>
</tr>
<tr>
<td>Interpreting the results generated</td>
<td>18</td>
</tr>
<tr>
<td>Applying the model</td>
<td>20</td>
</tr>
<tr>
<td>Specifying the settings object and name</td>
<td>21</td>
</tr>
<tr>
<td>Specifying the input data</td>
<td>21</td>
</tr>
<tr>
<td>Setting the mode parameters</td>
<td>22</td>
</tr>
<tr>
<td>Specifying the input fields</td>
<td>22</td>
</tr>
<tr>
<td>Specifying advanced parameters</td>
<td>22</td>
</tr>
<tr>
<td>Specifying parallel parameters</td>
<td>22</td>
</tr>
<tr>
<td>Specifying output fields</td>
<td>23</td>
</tr>
<tr>
<td>Specifying the output data object name</td>
<td>23</td>
</tr>
<tr>
<td>Creating a sequence</td>
<td>25</td>
</tr>
<tr>
<td>Creating a statistics function</td>
<td>26</td>
</tr>
<tr>
<td>Specifying the statistics function and name</td>
<td>27</td>
</tr>
<tr>
<td>Specifying the input data for statistics function</td>
<td>27</td>
</tr>
<tr>
<td>Computing statistics, quantiles, or a sample</td>
<td>28</td>
</tr>
<tr>
<td>Specifying output fields</td>
<td>29</td>
</tr>
<tr>
<td>Specifying the output data object name</td>
<td>29</td>
</tr>
<tr>
<td>Specifying the result object name</td>
<td>29</td>
</tr>
<tr>
<td>Running the statistical function</td>
<td>30</td>
</tr>
<tr>
<td>Interpreting the results</td>
<td>30</td>
</tr>
</tbody>
</table>

© Copyright IBM Corp. 1996, 1999
Tutorial

This Intelligent Miner mining tutorial consists of several mining tasks. The tutorial starts with data in a flat file, details the process of defining Intelligent Miner data objects, running Intelligent Miner functions, and viewing results using the Intelligent Miner’s visualizers.

This Intelligent Miner tutorial consists of an abbreviated data mining scenario with five phases: Defining data, building a model, applying the model, automating the process, and analyzing the results. By following the steps in this tutorial, you will learn how to use the Intelligent Miner wizards to define data objects, run mining functions, and view results in the Intelligent Miner.

This tutorial and the sample data used in this tutorial are designed to support the learning objectives. As such, they do not represent actual or recommended methods for using the Intelligent Miner. To shorten the time it takes to complete the tutorial, the data file is small and can be processed quickly. Additionally, the five phases represent an important subset of the activities at the core of many mining projects. Finally, the tutorial uses the Demographic Clustering function to accomplish its goals. There are other functions within the Intelligent Miner that can be used to accomplish the same end. Typical mining investigations would compare the results of more than one function.

Before you start

To use this tutorial, you need:

- The Intelligent Miner server installed on AIX, OS/400, OS/390, Sun Solaris, or Windows NT
- The Intelligent Miner client installed on AIX, OS/2, Windows NT, or Windows 95

The business problem

Imagine that you work for a bank that sells several products, including Regular Checking, Premier Checking, and Exclusive Checking accounts and option packages for each account. The bank already knows that Premier Checking is their most profitable product, and wants to increase the number of customers who have this type of checking account. The marketing department wants to identify different groups based on demographic data, such as age and income, within the Premier Checking customers so that the department can prepare different ad campaigns for each of the groups.
Additionally, the department wants to identify customers who are not currently Premier Checking customers who have similar demographics to the customers who are Premier Checking customers.

You have obtained some customer data from corporate headquarters to solve this business problem. This data is named banking.txt. It contains information about customers from all branches of the bank. You can use the Intelligent Miner to mine this data and provide demographic information to the marketing department. Your customer data includes information about customers who already have the Premier Checking account, so you can use the Demographic Clustering mining function to identify different groups, based on demographic data, among customers who already have Premier Checking.

The mining run tasks

This tutorial will demonstrate five phases of data mining tasks:

Defining the data

Define a data object that points to a flat file containing your customer data file banking.txt. The data object will be named Customers.

You must specify which properties of your customers are contained in the data, their data types, and the columns in the flat file that they occupy.

The Intelligent Miner data objects simply point to the location of your data, so that the Intelligent Miner can process this data. You will not actually be changing the contents of the banking.txt file. See "Defining a data object" on page 9 for instructions on how to complete this step.

Building the model

Define a Demographic Clustering settings object named Build model. This settings object uses the Customers data object as the input data. It runs in clustering mode, and produces a results object named Model. This model contains information that describes the clusters identified during the mining run. See "Building a model" on page 13 for instructions on how to complete this step.

Applying the model

Define a Demographic Clustering settings object named Apply model. This settings object uses the Customers data object as the input data. It runs in application mode using the Model results object and produces an output data object named Scored customers and a flat file named scored.txt. This output file identifies the subgroup
associated with a customer record. See “Applying the model” on page 20 for instructions on how to complete this step.

Automating the process

To automate the process you create a sequence object Target Marketing containing the Build model settings object and the Apply model settings object. A sequence is an object containing several other objects in a specific sequential order. You can run a sequence, which runs each of the objects within the sequence in the order that you specified. This allows you to combine several mining tasks into one step. See “Creating a sequence” on page 25 for instructions on how to complete this step.

Analyzing the results

Define a Bivariate Statistics function named Analyze. This statistical function analyzes the data object Scored customers and an produces an output data object Target customers, a flat file target.txt, and a result object Target customer demographics. See “Creating a statistics function” on page 26 and “Interpreting the results” on page 30 for instructions on how to complete this step.

Starting the Intelligent Miner in demo mode

The Intelligent Miner includes a sample flat file named banking.txt that contains fictitious customer data. To access this sample data, you must start the Intelligent Miner in demo mode.

Running the Intelligent Miner in demo mode provides all the same functionality as running Intelligent Miner in regular mode, but also includes sample mining bases and sample data. The interface that you use to create objects, define data, and perform other data mining tasks for this tutorial is the same interface that you use in regular mode.

Using the demonstration data on AIX servers

To start the Intelligent Miner in demo mode, first determine if you are running in AIX local mode or if you are using remote access to an AIX server.

If you are running in AIX local mode: Start the Intelligent Miner server in demo mode using these commands:

1. cd /usr/lpp/IMiner/bin
2. ./idmstart
3. ./imdemo
If you are using remote access to an AIX server running the Intelligent Miner in demo mode:

Start the Intelligent Miner server using these commands:

1. cd /usr/lpp/IMiner/bin
2. ./idmstartdemo

Start the Intelligent Miner client as usual. Clients connecting to a server running the Intelligent Miner in demo mode share the sample mining bases. The mining bases are opened in read-only mode because many users might access the same sample mining base at the same time. Each client must sign on to the Intelligent Miner server using the appropriate user ID and password on the Server Logon page of the Preferences notebook.

Restarting the Intelligent Miner in demo mode: To remove all additional mining bases and reset them to their initial state, issue these commands:

1. idmstop
2. idmstartdemo to start the Intelligent Miner in demo mode, or idmstart to start the Intelligent Miner in regular mode

Note: In demo mode, the IDM_MNB_DIR environment variable points to a temporary directory rather than to the home directory on the server. This temporary directory is defined as /tmp/dmtnsample.xxxxx, where xxxxx is a five-digit number.

Using the demonstration data on OS/390 servers

To make the demonstration data available for particular client users, you must install this data in each user’s home directory on the server. To do so, perform the following steps:

1. Run the IDMDEMO job. The IDMDEMO job is provided with the installation sample library (SIDMSAM1).
2. Adjust the job to your needs before you submit it by following the instructions.
   You can now open the sample mining bases at the client side and use them to get familiar with the product.
3. Repeat this action for any user who needs access to the demonstration data.

Using the demonstration data on AS/400 servers

To make the demonstration data available for particular client users, you must install this data in each user’s home directory on the server. Thus you must repeat the following steps for each user who needs access to the
demonstration data. The description below uses a home directory profile named MYPROF as an example. Replace this name with an actual profile name when you perform these steps:

1. Enter DSPUSRPRF MYPROF and record the value of the HOMEDIR parameter for that directory.
2. Make sure that the directory exists by entering WRKLNK '/HOME/MYPROF'. If the directory does not exist, create it by entering CRTDIR '/HOME/MYPROF'.
3. You might want to create a subdirectory in the home directory so that the demonstration data is kept separate from the "real" data. This example uses a subdirectory named IMDEMO.
4. Copy the demonstration data to this directory. Enter:
   CALL QIDM/QYDMDEMOC '/HOME/MYPROF/IMDEMO'

   Note: Intelligent Miner Version 1 places all files in a subdirectory called dmtksample. Intelligent Miner Version 2 places these files in the directory you specify.
   This command creates the following directories as subdirectories of the
   HOME/MYPROF/IMDEMO directory:
   DATA       Contains the data stream files
   IDMMNB     Contains mining bases
   IDMRES     Contains results
   Files in the IDMMNB and IDMRES directory contain explicit links to this directory and do not work if you copy them to another directory without using the QIDM/QYDMDEMOC command.
5. Change the profile of the home directory so that you can use the demonstration data. Enter the following command:
   CHGPRF HOMEDIR('/HOME/MYPROF/IMDEMO') CCSID(37)
   Set the Coded Character Set Identity (CCSID) to 37 because the demonstration data is available in English only.
6. Grant access to this user profile unless the user has *ALLOBJ authority. To do so, enter the following commands and specify a valid client user ID for USERID:
   a.
   CHGAUT OBJ('/HOME/MYPROF/IMDEMO')
   USER(USERID)
   DTAAUT(*Rwx OBJAUT(*ALL))
   b.
Repeat this step for the DATA, IDMMNB, and IDMRES subdirectories.

7. Undo the changes to a home directory profile if the user does not need the demonstration data anymore.

**Using the demonstration data on Sun Solaris servers**

When you install the Intelligent Miner, demonstration data in English is copied to your server. To replace the English demonstration data with demonstration data in one of the supported languages, enter the appropriate command from the command line:

```
pkgadd -a ./admin -d . IMdemoES
To install the demonstration data in Spanish.
```

```
pkgadd -a ./admin -d . IMdemoFR
To install the demonstration data in French.
```

```
pkgadd -a ./admin -d . IMdemoGE
To install the demonstration data in German.
```

```
pkgadd -a ./admin -d . IMdemoHU
To install the demonstration data in Hungarian.
```

```
pkgadd -a ./admin -d . IMdemoIT
To install the demonstration data in Italian.
```

```
pkgadd -a ./admin -d . IMdemoJP
To install the demonstration data in Japanese.
```

```
pkgadd -a ./admin -d . IMdemoKR
To install the demonstration data in Korean.
```

```
pkgadd -a ./admin -d . IMdemoBR
To install the demonstration data in Portuguese.
```

```
pkgadd -a ./admin -d . IMdemoRU
To install the demonstration data in Russian.
```

```
pkgadd -a ./admin -d . IMdemoTW
To install the demonstration data in Traditional Chinese.
```

```
pkgadd -a ./admin -d . IMdemoCN
To install the demonstration data in Simplified Chinese.
```

During the installation process, several files must be overwritten. Confirm the overwriting of files when prompted to do so.

To reinstall the English demonstration data, enter the following command from the command line:
pkgadd -a ./admin -d . IMdemoEN

To start the Intelligent Miner in demo mode on the server, enter idmstartdemo.

A set of sample mining bases with predefined mining and statistics settings is created in a temporary directory. All clients connecting to this server share these mining bases.

To protect the sample mining bases against accidental deletion or overwriting, they are locked when a user opens them. Thus to save any changes or modifications, select Save as from the File menu to save the mining base in question under a different name.

To remove all additional mining bases and reset the original sample mining bases to their initial state, enter these commands:

1. idmstop
2. idmstartdemo

To return to regular (non-demo) mode, enter the following commands:

1. idmstop
2. idmstart

Note: In demo mode, the IDM_MNB_DIR environment variable points to a temporary directory rather than to the home directory on the server. This temporary directory is defined as /tmp/dmtksample.xxxxx, where xxxxx is a five-digit number.

Using the demonstration data on Windows NT servers

Before you start the Intelligent Miner in demo mode, check whether you use the Intelligent Miner in local mode, in stand-alone mode, or in client/server mode. Then follow the instructions in the appropriate section.

Local mode and stand-alone mode

To start the Intelligent Miner in demo mode:

1. Make sure that you started the Intelligent Miner server.
2. Enter imdemo from an MS-DOS window.
   This creates temporary sample mining bases and starts the graphical user interface (GUI) with predefined mining and statistics settings. All functions available with the regular program are also supported in demonstration mode. When you close the GUI, the temporary mining bases are deleted.
   The demonstration mode does not affect any Intelligent Miner service started in regular mode on the server. Other users who are connected to
the same server can work on their own mining bases while a stand-alone user is running the Intelligent Miner in demonstration mode.

**Client/server mode**

To make the temporary sample mining bases accessible to a remote client, enter `idmstartdemo` in an MS-DOS window.

A set of sample mining bases with predefined mining and statistics settings is created in a temporary directory. All clients connecting to this server share these mining bases. The mining bases are locked to prevent accidental overwriting or deletion. Use **Save as** from the **File** menu to save any modifications to the sample mining bases.

To remove all additional mining bases and reset the original sample mining bases to their initial state, follow these steps:

1. Enter `idmstopdemo` to run the idmstopdemo batch file. This file switches the server back to regular mode and removes the sample mining bases.
2. Enter `idmstartdemo` again.

To return to regular (non-demo) mode, enter `idmstopdemo`.

In demo mode, the IDM_MNB_DIR environment variable points to a temporary directory rather than to the home directory on the server. This temporary directory is defined as `%TEMP%\dmtkdemo\idmmnb`, where X is a number from zero to nine.

**Attention:** While the server is in demonstration mode, users cannot access mining bases created in regular mode. Before you start the Intelligent Miner in demonstration mode, make sure that no other client user is connected to the server. Otherwise, the other client users might lose data or results. This applies similarly to the **idmstopdemo** command: Verify that no client user is connected to the server before you stop the demonstration mode. If other users also work in demonstration mode, they cannot access the sample mining bases any longer.

---

**The Intelligent Miner main window**

The Intelligent Miner main window helps you manage mining bases and perform data mining tasks. A mining base is a collection of the mining objects needed to conduct a mining run.

After starting the Intelligent Miner client on AIX or Windows:

1. Select **Options + Preferences**. The Preferences notebook opens.
2. Click the Miscellaneous tab to display the Miscellaneous page.
3. Check the **After mining function is run, automatically visualize a result** box.
4. Under Autosave Mining Base, click the **Save after creating** radio button.
5. Click **OK** to save your preferences.

**Figure 1** shows the Intelligent Miner main window.

**Defining a data object**

The first step in this tutorial is to define an Intelligent Miner data object that points to the raw customer data you want to mine. For this tutorial, you are using data from a flat file banking.txt that resides on the Intelligent Miner server. To define a data object:

1. Click the **Create Data** icon on the toolbar to start the Data wizard from the Intelligent Miner main window.
2. From the Welcome page of the Data wizard, click **Next** to continue.

**Specifying the data format and object name**

On the **Data format and settings** page of the Data wizard, you must specify what kind of data you are using, and the name of the Intelligent Miner data object.

1. Select **Flat files** from the list, if it is not already selected.
2. Enter the name of the data object in the Settings name field: Customers. Optionally, you can also specify a comment associated with this data object, for example, you might type the comment: Data about banking customers.

3. Ensure that the **Show the advanced pages and controls check box** is not checked.

4. Click **Next** to continue.

### Specifying the location of the data

On the Flat files page of the Data wizard, you specify the name and location of the data for this Intelligent Miner data object.

1. Double-click on the folder `dmtksample.n` in the list on the left, where `n` is a number. The contents of this folder are displayed in the list.

2. Double-click on the `Data` folder. The contents of this folder are displayed in the list on the right.

3. Scroll through the list and select `banking.txt`.
   
   If you do not see the `banking.txt` file in the list of available files, make sure that you provided the correct user ID and password for the Intelligent Miner server on the Server Logon page of the Preferences notebook.

4. Click **Add file**.

5. Make sure that the use mode **Read only** is selected.

6. Click **Next** to continue.

### Specifying the field parameters

On the Field Parameters page, you specify the begin and end positions, field names, field types, and optional name mappings for fields in this data object. The flat file display shows you a few lines of the flat file data, to help you enter the begin and end positions for each field. You do not have to specify the field parameters for all fields in the flat file, only the ones that you want to use for the mining run.

Table 1 shows the field parameters for the flat file `banking.txt`. The begin and end positions are the numerical positions of the columns representing each field in the flat file. For example, the Flat file display field in Figure 2 on page 11 shows columns 1–6 contain values for the field `gender`, which is of the categorical field type.

<table>
<thead>
<tr>
<th>Begin and end position</th>
<th>Field name</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-6</td>
<td>gender</td>
<td>Categorical</td>
</tr>
<tr>
<td>10-16</td>
<td>age</td>
<td>Continuous</td>
</tr>
<tr>
<td>24-25</td>
<td>siblings</td>
<td>Continuous</td>
</tr>
</tbody>
</table>
Table 1. Field parameters for the Customers data object (continued)

<table>
<thead>
<tr>
<th>Begin and end position</th>
<th>Field name</th>
<th>Field type</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-36</td>
<td>income</td>
<td>Continuous</td>
</tr>
<tr>
<td>38-44</td>
<td>type</td>
<td>Categorical</td>
</tr>
<tr>
<td>45-45</td>
<td>product</td>
<td>Categorical</td>
</tr>
</tbody>
</table>

To enter the field parameters:

1. Type 1–6 in the **Begin and end position** entry field.
2. Type gender in the **Field name** entry field.
3. Select **Categorical** in the **Field type** entry field.
4. Click **Add**.
5. Repeat the previous steps define the field parameters for this data object, as shown in [Table 1 on page 10].

When you finish defining the field parameters, the window should look like [Figure 2].

**Tip:** You can use a name mapping to substitute a character string for a numeric code. For example, you might find it convenient to use a name mapping for the product field, which has values ranging from 1–8. You can map each numeric value to the name of the product, for example, product 1 is the Premier Checking account, product 2 is the Regular Checking account, and so on. This tutorial does not use name mappings for any fields in this data object.
6. Click **Next** to continue.

### Defining computed fields

Because there is no need for a computed field in this tutorial, you do not have to define any computed fields for this data object.

Click **Next** to continue.

### Saving the data object

The next page of the Data wizard is the Summary page. The Summary page provides a summary of the parameters you defined for the current object you are creating.

Click **Finish** to complete the wizard. If you receive an error message after clicking **Finish**, you can use the **Back** button to return to any page and make corrections based on the message. At this point, you have defined the data object for your mining run. By selecting the Data folder in the mining base container, you can see the icon representing this data object.

Now that you have defined the first object for your mining run, you should save the mining base.

1. To save the mining base, click on the **Save mining base as** icon from the main window toolbar.
2. Enter **Target Marketing** as the mining base field. Optionally, enter a descriptive comment about the purpose of that mining base.
3. Click **Save** to save the mining base.

Note that, in demonstration mode, the sample mining bases are deleted when you stop the Intelligent Miner server. Therefore, to save the results of this tutorial permanently, export the sample mining base by following these steps:

   a. Click **Mining Base → Export Mining Base** in the Intelligent Miner main window.
   b. To specify an export path on your current drive, navigate through the tree view and select a suitable folder.
      
      To export the sample mining base to another drive, enter the full path, including the drive letter, in the **Mining base and supporting files filename stem** field.
   c. Enter a name for the mining base in the **Mining base and supporting files filename stem** field or append a mining-base name to the path you specified.
   d. Click **OK** to export the mining base.
Building a model

You can use the Demographic Clustering mining function to obtain information about customers who already have the Premier Checking account. You run this function in clustering mode to produce a model as a result object.

This function generates clusters from your input data. Information about these clusters is stored in a result object that you can view with the clustering visualizer.

To create the settings object for building the model:
1. Click on the Create mining button from the Intelligent Miner main window tool bar. The Intelligent Miner displays the Welcome page of the Mining wizard.
2. Click Next to continue.

Specifying the mining function and name

On the Mining functions and settings page, you select the type of settings object that you want to create and specify the name and an optional comment for this settings object:
1. Select Clustering – Demographic from the list of mining functions.
2. Type the settings name Build Model. Optionally, you can specify a more descriptive comment associated with this settings object.
3. Check the Show the advanced pages and controls check box. The advanced pages and controls of this wizard allow you to use additional options when defining your settings object. For example, using the advanced pages and controls, you can specify a filter condition for the records of the input data. The Intelligent Miner filters the records based on the condition you specify while it is running the mining function.
4. Click Next to continue.

Specifying the input data

On the input data page of the wizard, you specify the input data for this mining function:
1. From the list of available input data, select the Customers data object that you created in a previous step.
2. Select Disk space under Optimize mining run for.
3. Under the Advanced options, click the ... button that is next to Filter records condition. You can filter the records of the input data for this settings object.
For this tutorial, you want to include all input data records for customers who have the Premier Checking account. Because the code for the Premier Checking account is 1, the filter condition you want to specify is product=1.

1. Click the AND push button. The expression builder creates a template for the expression, which displays as \((\text{Arg1} = \text{Arg2})\).
2. In the Category list, click on Field Names. The Value list displays all the available fields that you can include in this expression.
3. From the Value list, select the field product.
4. Click on the Arg1 button. This sets the field product as the first argument in the expression.
5. Select Constants from the Category list.
6. Double-click on <new constant> in the Value list.
7. Type in the new constant value 1.
8. Press Enter. The new constant is added to the list of constants.
9. Select the constant 1 from the Value list.
10. Click the Arg2 button. This sets the constant value of 1 as the second argument in the expression.
11. Click OK to return to the Input data page of the Mining wizard.
12. Click Next to continue.

Setting the mode parameters

On the mode parameters page of the wizard, you specify the mode parameters for the mining function. You can run the Demographic Clustering mining function in two modes: clustering mode or application mode. For more information on the mode parameters, see the online help for the Demographic Clustering mining function.

For this tutorial, you will run this mining function in clustering mode.

Clustering mode: In clustering mode, the function identifies groups of similar records called clusters. The function has parameters that you can use to control the results, including the maximum number of clusters, maximum number of passes, accuracy, and similarity threshold.

Use the default values for these parameters for this tutorial.

Maximum number of passes
The default value is 2.

Maximum number of clusters
The default value is 9.

Accuracy
The default value is 2.
Similarity threshold
The default value is 0.5.

To set the mode parameters, ensure that the Clustering mode radio button is selected.

Click Next to continue.

Specifying the input fields

On the Input fields page of the Mining wizard, you specify the active fields and supplementary fields for this settings object.

The Demographic Clustering mining function searches the input data for records with similarities, and places similar records into clusters. The active fields you specify are used to determine whether or not the input data records are similar. Statistics about the supplementary fields you specify are included in the result, but are not used to determine similarities.

For example, one of the fields in the customer data is gender. If you specify gender as an active field, the Demographic Clustering mining function uses this as a criterion in determining whether or not two customers are similar. You do not use gender as an active field because of the bank’s policy not to include gender information in marketing decisions. Instead, you use demographic information like income, age, and siblings. To see how gender is distributed within the clusters, you add it as a supplementary field.

Because you have filtered out all customers who do not have the Premier Checking account, you do not want to use product as an active field in the clustering process.

1. Select age, income, siblings, and type from the list of Available fields. Click the > push button to add them to the list of Active fields.

2. Select gender from the list of Available fields. Click the > push button to add it to the list of Supplementary fields.

After you have specified the active fields and supplementary fields, the input fields page should look like Figure 3 on page 13.
3. Click Next to continue.

Specifying advanced parameters

The next few pages of the wizard are advanced pages. For this tutorial, you accept the default values.
1. On the Field parameters page of the wizard, click Next to continue.
2. On the Additional field parameters page of the wizard, click Next to continue.
3. On the Outlier treatment page of the wizard, click Next to continue.
4. On the Similarity matrix page of the wizard, click Next to continue.

Specifying other parameters

You should see the Output fields page now. However, if you are connected to a server on which the parallel version is installed, you see the Parallel parameters page. In this case, proceed as follows:
1. Make sure that the Run the serial mode of the function radio button is selected.
2. Click Next to proceed to the Output fields page.

When you reach the Output fields page, proceed as follows:
1. Make sure that the Create output data radio button is not selected.
2. Click Next to continue.
Specifying the result object name

Each time that you run the Demographic Clustering mining function in clustering mode, it creates a result object and saves it with the name you that specify on this page. Because the mining process is an iterative one, you will probably run a settings object more than once. You can choose to allow this settings object to replace a result object with the same name, because an existing result object with the same name may have been generated by a previous iteration of this settings object.

To specify the results object name:
1. Type Model as the name of the result object. Optionally, specify a comment associated with this result object.
2. Check the If a result with this name exists, overwrite it check box.
3. Click Next.
4. On the Summary page of the Mining wizard, review the parameters for the settings object that you are defining. Select the Run this settings immediately check box.
5. Click Finish to complete this task.

After you completed each step in the Mining wizard, the Intelligent Miner runs the settings object and displays a progress indicator that allows you to monitor the status of the mining function. After successfully running the mining function, the Intelligent Miner displays the result object generated by this settings object.

This settings object generates a result object named Model, which describes clusters of customers who have the Premier Checking account. Each cluster contains customers with similar characteristics for the fields income, age, type, and siblings.

Whether the Intelligent Miner displays the result object immediately after generating it depends on a preference setting on the Miscellaneous page of the Preferences notebook. You can specify whether you want to visualize results immediately after they are generated. If the Intelligent Miner does not display the result object immediately, you can view the result object by double-clicking on the result object in the Results folder from the Intelligent Miner main window or by pressing View Results in the Progress Indicator window.
Interpreting the results generated

The results generated by the mining function are shown in Figure 4. The multiple rows of graphs are designed to give you an understanding of the clusters described in the result.

![Figure 4. Results of the build model settings object](image)

The display shows nine rows, each representing one of the nine clusters identified by the mining run. Within each cluster, the pie charts and the bar charts represent active and supplemental fields used in the cluster. In this case, fields that had the greatest influence on forming the cluster are displayed on the left, while fields with the least influence are displayed on the right. The numbers down the left side represent the cluster size as a percentage; for example, the top cluster represents 36% of the data, the next lower cluster represents 24%, and so on. The numbers down the right side identify the cluster ID.

The top row is the cluster with the largest number of customers, accounting for 36% of the customers. Each bar or pie chart shows the distribution of the field for the cluster and for the entire set of Premier Checking (product 1) customers. Supplementary fields are indicated with square brackets around the field names.
You can display more detail about a cluster by double-clicking on a chart in the row. Figure 5 shows the display of the top cluster, the largest cluster. This cluster contains the account type, age, gender, and siblings fields. Assume that account type indicates the options package the customer purchased, specified at the bank by colors. Double-clicking on any of the graphs shows a single graph.

Figure 5 includes a pie chart for gender. Each pie chart produced by the Intelligent Miner shows two distributions. The outside ring shows the distribution for the entire sample. The inside ring shows the distribution for the associated cluster. For example, Figure 5 shows a pie chart in which the outside ring represents the distribution of male and female customers for all Premier Checking customers; the inside ring represents the distribution of male and female customers in this cluster. The chart indicates that this cluster has a slightly greater percentage of males in it than the group of Premier Checking customers as a whole. Looking back at Figure 4 on page 18 you can quickly see that cluster 2 is mostly male and cluster 7 is mostly female.

Figure 6 on page 20 shows the distribution of age for the first cluster. The solid bars represent the percentage of people in each age group for all the data and the red transparent bars represent the distribution in cluster 6. As
shown by the graph, a higher percentage of the customers represented by
cluster 6 are below the age of 15 when compared to the entire population of
Premier Checking customers.

Viewing the clustering results can provide insights about the characteristics of
a subgroup of Premier Checking customers. The result object contains detailed
statistical information, which you use as a model to apply to a new set of data
in a subsequent step of this tutorial.

### Applying the model

The next step in the tutorial is to apply the clustering model created in the
previous step. The result object named Model contains descriptions of nine
clusters of customers who already have the Premier Checking account.

To create a settings object for applying the model:
1. Click on the Create mining button from the Intelligent Miner main
   window toolbar. The Intelligent Miner displays the Welcome page of the
   Mining wizard.
2. Click Next to continue.

Specifying the settings object and name

For this step in the tutorial, you again run the Demographic Clustering mining function to apply the model created previously.
1. Select Clustering – Demographic from the list of mining functions.
2. Type the settings name **Apply Model**. Optionally, specify a more descriptive comment associated with this settings object.
3. Check the Show the advanced pages and controls check box.
4. Click Next to continue.

Specifying the input data

For this step in the tutorial, you use the same customer data used by the previous settings object. In most situations, you will build a model using one set of data, and apply the model to a different set of data. The flat file used in this tutorial contains customers with Premier Checking accounts and customers without Premier Checking accounts. In this case, you use the Intelligent Miner filtering feature to separate the customer records into two sets.
1. From the list of available input data, select the **Customers** data object.
2. Ensure that you are optimizing the mining run for disk space.
3. Under Advanced parameters, click the ... button next to Filter records condition. You will be filtering the records of the input data for this statistics function.

   The Intelligent Miner displays the Expression Builder for filtering records. In this step, you use the filter to select only the customers who do not have premier checking. The expression looks like this: `((product<>1))`. To create this expression:
   a. Click the AND push button. The expression builder creates a template for the expression, which displays as `((Arg1 = Arg2))`.
   b. Click on Field Names in the Category list. The Value list displays all the available fields that you can include in this expression.
   c. From the Value list, select the field **product**.
   d. Click the Arg1 button. This sets the field product as the first argument in the expression.
   e. Click the <> button. This sets the operand as “not equal to.”
   f. Select Constants from the Category list.
   g. Select the constant 1 from the Value list.
   h. Click on the Arg2 button. The constant value of 1 is set as the second argument in the expression. The expression you defined looks like this: `((product<>1))`.  

   Tutorial 21
4. Click OK to return to the Input data page of the Mining wizard.
5. Click Next to continue.

**Setting the mode parameters**

In a previous step, you ran the Demographic Clustering mining function in clustering mode to create a model. In this step, you apply the model to the customer data by running this mining function in application mode.

In application mode, the mining function scores how similar each customer is to the two most similar clusters of customers with Premier Checking.

For more information on the mode parameters, see the online help for the Demographic Clustering mining function.

To set the mode parameters:
1. Click the Application mode radio button.
2. Select the result object Model from the application mode container under the Application mode group box.
3. Click Next to continue.

**Specifying the input fields**

In this step, you score customers using the same fields that were used to define the clusters in the Build model.

1. Select age, income, siblings, and type from the list of Available fields. Click the > push button to add them to the list of Active fields.
2. Click Next to continue.

**Specifying advanced parameters**

The next few pages of the wizard are advanced pages. For this tutorial, you accept the default values by clicking the Next button on the the next four pages.

1. On the Field parameters page of the wizard, click Next to continue.
2. On the Additional field parameters page of the wizard, click Next to continue.
3. On the Outlier treatment page of the wizard, click Next to continue.
4. On the Similarity matrix page of the wizard, click Next to continue.

**Specifying parallel parameters**

You should see the Output fields page now. However, if you are connected to a server on which the parallel version is installed, you see the Parallel parameters page. In this case, proceed as follows:
1. Make sure that the **Run the serial mode of the function** radio button is selected.

2. Click **Next** to proceed to the Output fields page.

### Specifying output fields

On the output fields page of the Mining wizard, you must select the fields that the output data contains. The output data will also contain the cluster ID, the record score, and the confidence values, which are generated by the mining function. In this tutorial, the record score value is a measure of how similar the customers are to the clusters that they are part of.

1. Select `>>` to add all the available fields to the list of output fields.
2. Type `clusterID` in the **Cluster ID field name** entry field.
3. Type `score` in the **Record score field name** entry field.
4. Type `conf` in the **Confidence field name** entry field.

The output fields page now looks like this [Figure 7](#).

5. Click **Next** to continue.

### Specifying the output data object name

On the output data page of the wizard, you specify the name of the output data object for this settings object. This page shows the existing data objects in the current mining base. Because you want to preserve the Customers data
object, you need to create a data object that contains the output data named Scored customers. To do this, you open the Data wizard from this page of the Mining wizard, define the data object, and then return to this page of the Mining wizard.

1. Click on **Create data**. The Data wizard’s Welcome page opens.
2. Click **Next** to continue.
3. Select **Flat files**.
4. Type **Scored customers** in the settings name field. Optionally, type a comment to describe this data object.
5. Click **Next**.
6. On the Flat files page, change to the directory that contains the file banking.txt.
7. In the **Path and file name** entry field append scored.txt to the path.
8. Click on **Add file**.
9. Select the **The specified flat file does not yet exist** check box.
10. Click **Next** to continue.
11. On the Summary page of the Data wizard, click **Finish** to continue.

After defining the output data object, return to the Output data page of the mining wizard to continue the process of defining the mining object. You see the data object in the **Available output data** container, as shown in Figure 8.

![Figure 8. Output data page](image-url)
1. Select the data object **Scored customers**.
2. Click **Next**. The Summary page opens.
3. Select the **Run this settings immediately** check box.
4. Click **Finish** to continue. The Intelligent Miner will run the mining function and display a progress indicator that allows you to monitor the status of the mining function.
5. Click **OK** in the progress indicator window after the mining function has stopped running.

You now have a flat file that contains a list of customers and scores of how similar these customers are to the Premier Checking customers.

### Creating a sequence

Now that we have created a few of the functions for our mining run, we can create a sequence that runs each of the functions in the order we specify.

The benefit of using a sequence object is being able to combine several steps into one step. If you combine several functions into a sequence object, you need to run only the sequence object, which then runs each of the objects within it.

To create the sequence for this tutorial:

1. Click the **Create sequence** button on the Intelligent Miner main window toolbar.
2. On the **Welcome** page of the Sequence wizard, click **Next**.
3. In the **Setting name** field type **Target Marketing** as the name of this sequence object.
4. Click **Next**.
5. On the Parameters page of the Sequence wizard, you can use the mining base tree view to navigate to the objects you want to include in the sequence. The Parameters page is shown in Figure 9 on page 26. Because the first object you want to run is the Build model settings object, click on the + next to the **Mining** folder.
6. Click the **Clustering** folder. The contents of the folder are displayed.
7. Select the settings object **Build model** and drag and drop it to the **Sequence** work area.

   The settings object build model is added to the sequence as the first object to run.
8. In the **Contents of folder** area, select the mining object **Apply model** and drag and drop the object to the **Sequence** work area.

   This adds the settings object **Apply model** to the sequence as the second object to run, after **Build model**.
9. Click Next.

10. On the Additional parameters page of the Sequence wizard, check the If a settings object in the sequence fails, continue running the sequence check box.

11. Click Next.

12. On the Summary page of the Statistics wizard, select the Run this settings immediately check box.

13. Click Finish.

Creating a statistics function

In this step of the mining tutorial, you will create a Bivariate Statistics function. This function will produce descriptive statistics about the fields in scored data.

To create a statistical function:

1. Click the Create statistics button from the toolbar on the Intelligent Miner main window. The Statistics wizard opens.

2. Click Next to continue.
Specifying the statistics function and name

On the Statistics functions and settings page, you select the type of statistics function that you want to create, and specify the name and comment for this statistics function:

1. Ensure that Bivariate Statistics in the list of statistics functions is selected.
2. Type the settings name Analyze. Optionally, type a descriptive comment.
3. Check the Show the advanced pages and controls check box.
4. Click Next to continue.

Specifying the input data for statistics function

To specify the input data for this statistics function:

1. From the list of available input data, select the Scored customers data object that you created in a previous step.
2. Under the Advanced options, click the ... button next to Filter records condition. You will be filtering the records of the input data for this statistics function.

   The Intelligent Miner displays the Expression Builder for filtering records.

   In a previous step of this tutorial, you created a score for each record in the input data. The score ranges from 0 to 1, and a higher score means a greater similarity. For this tutorial, you include all input data records for customers whose score is greater than 0.7.

   The expression you want to create looks like this: ((score > 0.7)). To create this expression:
   a. Click the AND push button. The expression builder creates a template for the expression, which appears as ((Arg1 = Arg2)) on the first line.
   b. Click on Field Names in the Category list.
   c. Scroll through the list of field names and select the field score.
   d. Click the Arg1 button. The field score is set as the first argument in the expression.
   e. Click the > button.
   f. Click Constants in the Category list.
   g. Double-click <new constants> from the Value list.
   h. Type the constant value 0.7.
   i. Press Enter.
   j. Select 0.7 from the list of constants.
   k. Click on the Arg2 button.

   The expression looks like this: ((score > 0.7)).
3. Click **OK** to return to the Input Data page of the Statistics wizard, which now looks like Figure 10.

![Input data page of the Statistics wizard](image)

**Figure 10. Input data page of the Statistics wizard**

4. Click **Next**.

5. On the Parallel parameters page of the Statistics wizard, make sure that you are running this function in serial mode, on the Intelligent Miner server node.

6. Click **Next**.

**Computing statistics, quantiles, or a sample**

On the Statistics page of this wizard, you specify whether you want to compute statistics. For this tutorial, you will compute univariate statistics for the selected input data fields:

1. Select the **Compute statistics** radio button.
2. Select **age, clusterID, conf, gender, income, product, score, and siblings**, and click the > button to compute the univariate statistics for these fields.
3. Click **Next** to continue.
4. You will not compute quantiles in this tutorial. Click **Next**.
5. You will not create a sample in this tutorial. Click **Next**.
Specifying output fields

On this page of the Statistics wizard, you can choose whether to create output data, and what fields to include in the output data. Only input data records that meet the filter condition that you specified will be included in the output data. In this case, the output data will contain customers whose scores are greater than 0.7.

1. Click Create an output table.
2. Click >> to add all the available fields to the list of output fields.
3. Click Next to continue.

Specifying the output data object name

On this page of the wizard, you specify the name of the output data object for this statistics function. This output data object must exist before you can proceed. Because you have not yet defined the output data object, you will need to define the output data object from this page of the Statistics wizard:

1. Click on Create data. The Data wizard opens.
2. Click Next to continue.
3. Select Flat files.
4. Enter Target customers as the name of the data object. Optionally, type a comment that describes this data object.
5. Click Next to continue.
6. On the Flat files page, change to the directory that contains the file banking.txt.
7. In the Path and file name entry field, append target.txt to the path.
8. Click Add file.
9. Click The specified flat file does not yet exist.
10. Click Next.
11. On the Summary page of the Data wizard, click Finish. You now return to the Statistics wizard to continue the process of defining the statistics object.
12. On the output data page of the Statistics wizard, select the data object Target customers from the Available output data field.
13. Click Next.

Specifying the result object name

On the Results page of the Statistics wizard, enter the name of the result object that is generated by this statistical function:

1. Type Target customer demographics in the Results name field.
2. Check the If a result with this name exists, overwrite it check box.
3. Click **Next** to continue to the Summary page of the wizard.

**Running the statistical function**

To run the statistical function, check the **Run this settings immediately** check box.

Click **Finish** to continue.

The Intelligent Miner now runs this statistical function. A progress indicator shows the status of the function. After completion, the Intelligent Miner automatically displays the results of the statistical function.

Save the mining base from the Intelligent Miner main window. Click on the Save Mining Base icon.

You have produced an output data file named Target customer demographics. This file contains the customers identified as having high scores of similarity with typical customers of the Premier Checking account. You can now analyze the results.

**Interpreting the results**

The result generated by the Analyze statistical function is shown in Figure 11 on page 31. The multiple graphs show the distribution of the fields that you selected for statistics. The visualizer allows you to display more detail by double-clicking on any of the graphs. The Bivariate Statistics function provides you with statistics about the customers that have been targeted for your ad campaign. A quick look at the product graph shows that the customers similar to typical Premier Checking account customers are distributed fairly evenly over several other products.
The detailed statistics computed by mining and statistics object are shown on the details page. The top portion of the details page for the results generated by the Analyze statistical function is shown in Figure 12 on page 32.

**Figure 11. Results of the Analyze statistical function**

The detailed statistics computed by mining and statistics object are shown on the details page. The top portion of the details page for the results generated by the Analyze statistical function is shown in Figure 12 on page 32.
To access this page, use the Details for all partitions menu item under the View menu on the menu bar. In Figure 12, you can see that of the possible 1792 customers who do not have Premier Checking, 315 had a score of 0.7 or greater. These are the customers who have similar demographics to those who have purchased Premier Checking. You could adjust the filter to 0.6 to include more customers in the analysis.