

CHAPTER

1

Introduction to SAS Enterprise Miner Software

<i>Data Mining Overview</i>	1
<i>Layout of the SAS Enterprise Miner Window</i>	2
<i>Using the Application Main Menus</i>	3
<i>Using the Toolbox</i>	8
<i>Using the Pop-Up Menus</i>	8
<i>Unavailable and Dimmed Items</i>	9
<i>Enterprise Miner Documentation</i>	9

Data Mining Overview

SAS defines *data mining* as the process of uncovering hidden patterns in large amounts of data. Many industries use data mining to address business problems and opportunities such as fraud detection, risk and affinity analyses, database marketing, householding, customer churn, bankruptcy prediction, and portfolio analysis. The SAS data mining process is summarized in the acronym SEMMA, which stands for sampling, exploring, modifying, modeling, and assessing data.

- *Sample* the data by creating one or more data tables. The sample should be large enough to contain the significant information, yet small enough to process.
- *Explore* the data by searching for anticipated relationships, unanticipated trends, and anomalies in order to gain understanding and ideas.
- *Modify* the data by creating, selecting, and transforming the variables to focus the model selection process.
- *Model* the data by using the analytical tools to search for a combination of the data that reliably predicts a desired outcome.
- *Assess* the data by evaluating the usefulness and reliability of the findings from the data mining process.

You might not include all of these steps in your analysis, and it might be necessary to repeat one or more of the steps several times before you are satisfied with the results. After you have completed the assessment phase of the SEMMA process, you apply the scoring formula from one or more champion models to new data that might or might not contain the target. The goal of most data mining tasks is to apply models that are constructed using training and validation data to make accurate predictions about observations of new, raw data.

The SEMMA data mining process is driven by a process flow diagram, which you can modify and save. The GUI is designed in such a way that the business analyst who has little statistical expertise can navigate through the data mining methodology, while the quantitative expert can go “behind the scenes” to fine-tune the analytical process.

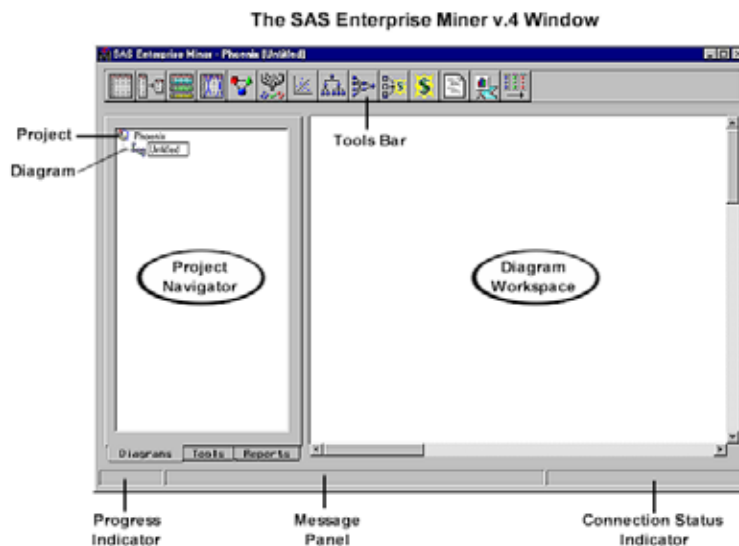
Enterprise Miner contains a collection of sophisticated analysis tools that have a common user-friendly interface that you can use to create and compare multiple

models. Statistical tools include clustering, self-organizing maps / Kohonen, variable selection, trees, linear and logistic regression, and neural networking. Data preparation tools include outlier detection, variable transformations, data imputation, random sampling, and the partitioning of data sets (into train, test, and validate data sets). Advanced visualization tools enable you to quickly and easily examine large amounts of data in multidimensional histograms and to graphically compare modeling results.

Enterprise Miner is designed for PCs that are running under Windows 95, 98, NT, 2000, XP, or subsequent releases of those operating environments. The figures and screen captures presented in this document were taken on a PC running under Windows NT 4.0.

Layout of the SAS Enterprise Miner Window

To start Enterprise Miner, start SAS and then type `miner` on the SAS command dialog bar. The SAS Enterprise Miner window opens:



The SAS Enterprise Miner window contains the following interface components:

- Project Navigator — You use the Project Navigator to manage projects and diagrams, add tools to the Diagram Workspace, and view HTML reports that are created by the Reporter node. Note that when a tool is added to the Diagram Workspace, it is referred to as a node. The Project Navigator has three tabs:
 - Diagrams tab — lists the current project and the diagrams within the project. By default, Enterprise Miner creates a local project that uses your user ID as the name (for this example, **Phoenix**) and a process flow diagram that is named **Untitled**. When you reopen Enterprise Miner, the last project you were working with is loaded.
 - Tools tab — contains the Enterprise Miner Tools Palette. The Tools Palette presents all the data mining tools that are available for constructing process flow diagrams. The tools are grouped according to the SEMMA data mining methodology.
 - Reports tab — contains the HTML report entries for the project. You generate reports using Enterprise Miner's Reporter node.

- **Diagram Workspace** — enables you to build, edit, run, and save process flow diagrams.
- **Tools Bar** — contains a customizable subset of Enterprise Miner tools that are commonly used to build process flow diagrams in the Diagram Workspace. You can add or delete tools from the Tools Bar.
- **Progress Indicator** — displays a progress indicator bar that indicates the execution status of an Enterprise Miner task.
- **Message Panel** — displays messages about the execution of an Enterprise Miner task.
- **Connection Status Indicator** — displays the remote host name and indicates whether or not the connection is active for a client/server project.

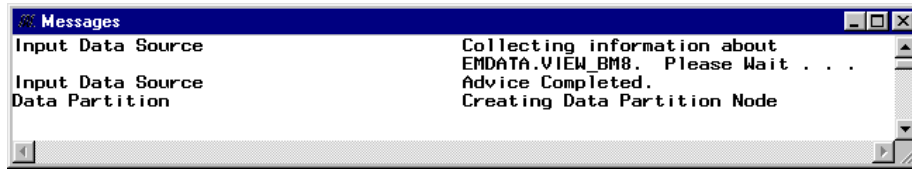
Using the Application Main Menus

You use Enterprise Miner's application menus to perform common tasks. Different windows have different menus. Here is a brief summary of the SAS Enterprise Miner window pull-down menus:

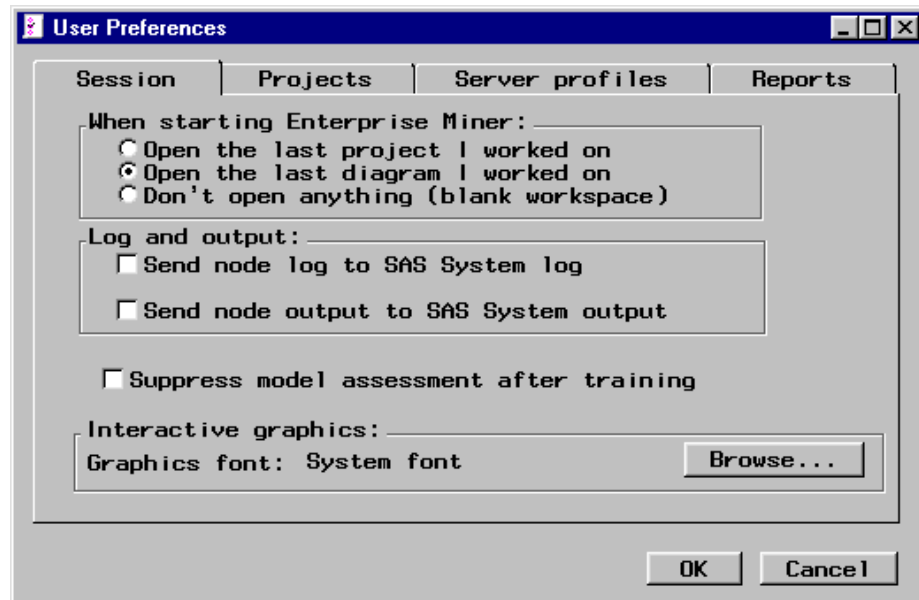
- **File**
 - **New**
 - **Project** – creates a new project.
 - **Diagram** – creates a new diagram.
 - **Open** – opens a new or existing diagram within the current project.
 - **Save Diagram** – saves the current diagram to the project.
 - **Save Diagram as** – names and saves the current diagram to the project.
 - **Print Setup** – specifies printing options.
 - **Print** – prints the contents of the SAS Enterprise Miner window.
 - **Delete current project** – deletes the current project. If you select this menu item, then a message appears that asks you to verify the deletion.
 - To delete the project, select **[Yes]**. All files in the project folder and any subfolders will be deleted. You cannot delete a project that someone else has opened.
 - **Close diagram** – closes and saves the diagram.
 - **Close project** – closes the project.
 - **Exit Enterprise Miner** – ends the Enterprise Miner session.
- **Edit**
 - **Copy diagram to clipboard** – copies the diagram to the clipboard.
 - **Undelete** – restores the last deleted node.
 - **Copy** – copies a node or object to the paste buffer. You must first select the node or object that you want to copy.
 - **Delete** – deletes the selected node or connection from the Diagram Workspace.
 - **Clone** – copies the current node and adds it to the Custom folder of the Tools Palette.
 - **Paste** – pastes the node, diagram, or object from the clipboard.
 - **Select all** – selects all nodes in the Diagram Workspace.
 - **Create subdiagram** – creates a subdiagram that you use to collapse a set of selected nodes and connections into a single Subdiagram node icon.

- **View**

- **Messages** – opens the Messages window that displays collected messages for this diagram.



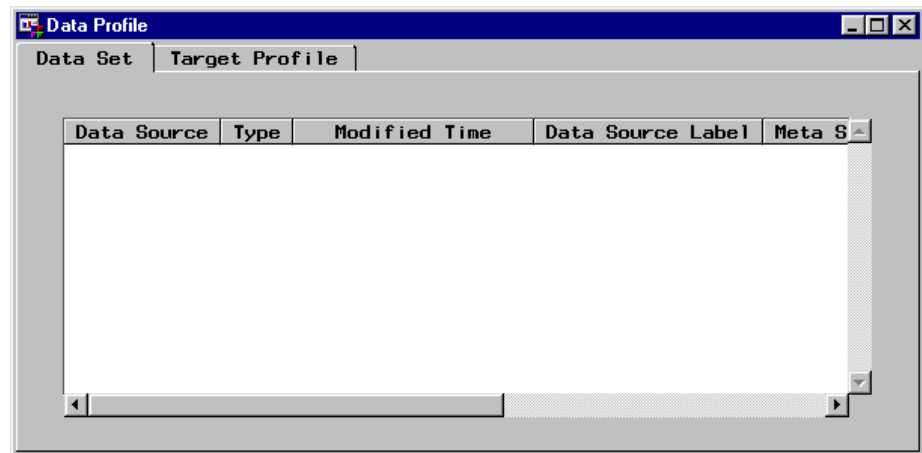
- **Refresh** – refreshes the Project Navigator and the Diagram Workspace view.
 - **Up One Level** – displays the next higher level of the process flow diagram. If there are no subdiagrams in your process flow diagram, then only one level can be displayed. If there are subdiagrams in your process flow diagram, then the subdiagrams can be displayed either in their condensed form (hiding the internal structure of the subdiagram) or in their expanded form (showing their internal structure). The higher level displays subdiagrams in condensed form.
 - **Top Level** – displays the process flow diagram in its most condensed form. All subdiagrams are condensed.
- **Options**
 - **User preferences** – opens the User Preferences window that enables you to specify start-up options, set the default directories for new projects, set up server profiles for client/server projects, and specify the HTML report items that you want to display when you run the Reporter node.



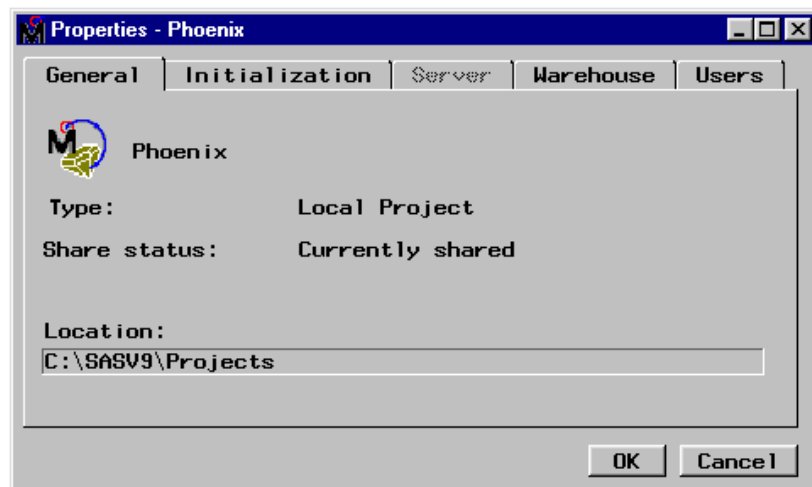
- **Project**

- **Data Profiles** – opens the Data Profile window that you use to define target profile information for a data set. For details, see the Target Profiler Help.

Help ► EM Reference Target Profiler

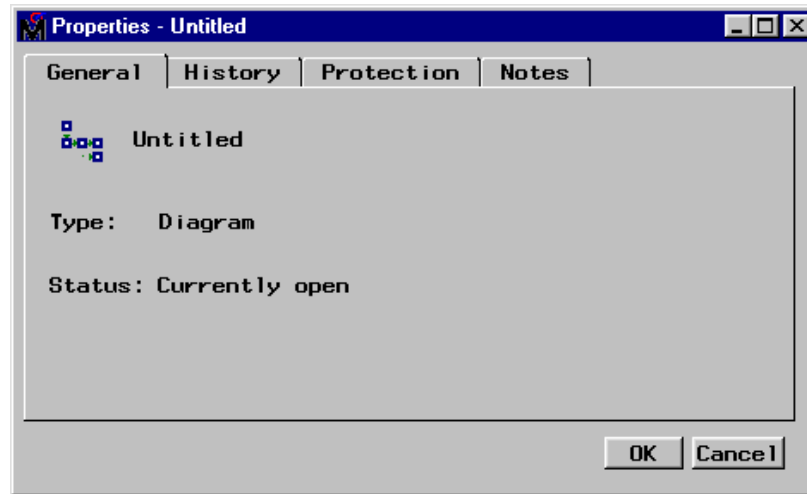


- **Properties** – shows project properties, such as the project name and type, share status, project location, server information, and a list of users that have the project open. You can also specify start-up and exit code that runs when the project is opened and closed, specify the SAS warehouse path, or update the project server profile. These settings can be specified if the project is opened by one user.



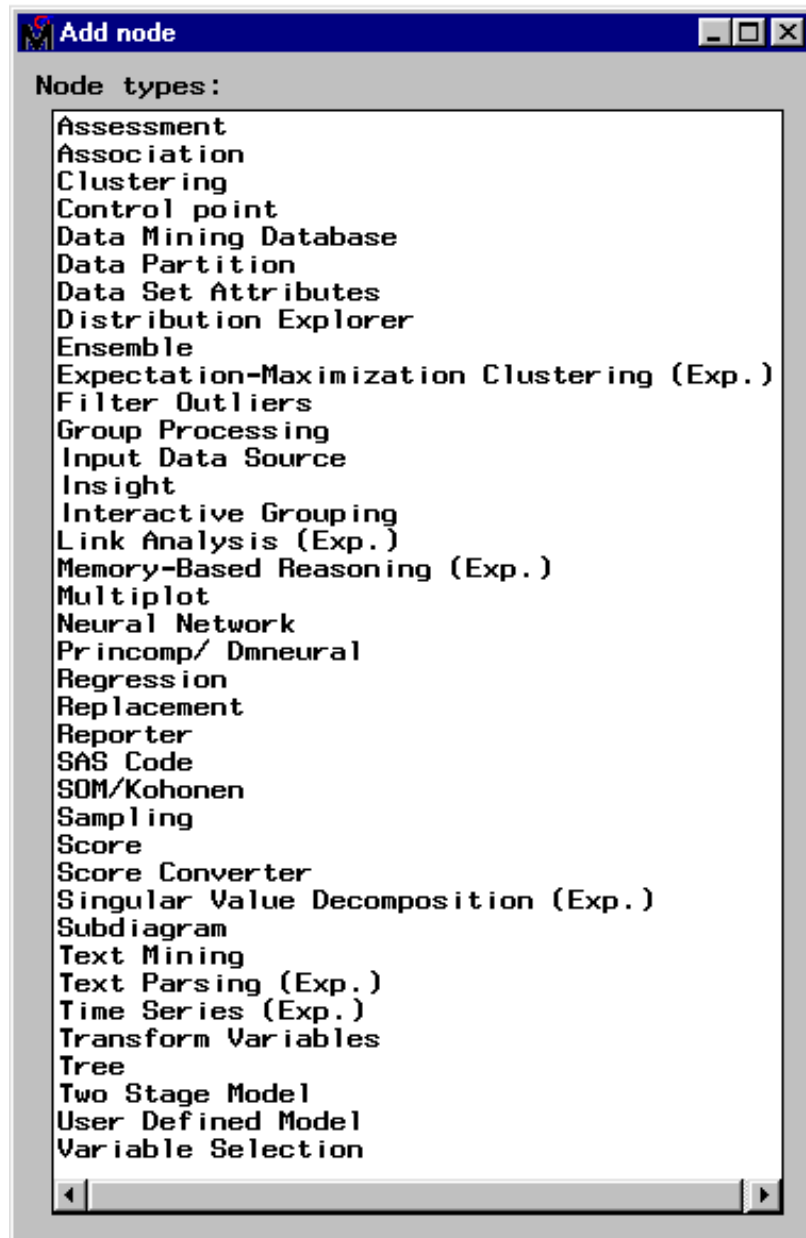
- **Diagram**
 - **Connect items** – a mode of editing where node icons are fixed in location, so that you can more easily make connections among them.
 - **Move items** – a mode of editing where node icons cannot be connected, so that you may more easily move the icons where you want them in the process flow diagram.
 - **Move and Connect**– (default) a mode of editing where node icons can be moved and connected in the Diagram Workspace.
 - **Large icons** – displays large node icons in the Diagram Workspace. This item is dimmed if a node is selected in the Diagram Workspace.
 - **Small icons** – (default) displays small node icons in the Diagram Workspace. This item is dimmed if a node is selected in the Diagram Workspace.
 - **Properties** – opens the Properties window that enables you to view diagram properties, such as the name, type, status, date created, and

date last modified. You can also apply password protection to the diagram and store notes about it.



Actions

- Open** – opens the selected node.
- Run** – runs the selected node and any predecessor nodes in the process flow that have not been previously run.
- Results**– opens the Results Browser for those nodes that generate results.
- Add Node** – opens the Add node window that enables you to add a new node to the Diagram Workspace.



- **Add Endpoints** – adds endpoints to the process flow.
- **Help**
 - **Getting Started with Enterprise Miner Software** – an interactive Enterprise Miner Tutorial that provides an overview of data mining and the Enterprise Miner software.
 - **Help** – Enterprise Miner Reference Help. This is a hierarchical Help listing. You can also access Help for a given window in Enterprise Miner by clicking the **what's This?** icon from the Enterprise Miner Toolbox. After your mouse pointer changes into a question mark, click inside the window frame you would like Help on.
 - **About Enterprise Miner** – Production release information about your copy of Enterprise Miner.

Using the Toolbox

In Enterprise Miner, the toolbox at the top of the SAS window appears as follows:



Tool tips are displayed when you hold your mouse pointer over a tool icon. The icons represent the following:

- 1 Open – opens a diagram within a project. You can also perform this task by using the main menu:

File ► Open

- 2 Save Diagram – saves the current diagram. You can also perform this task by using the main menu

File ► Save diagram

- 3 Delete Current Project – deletes the current project. You can also perform this task by using the main menu:

File ► Delete current project

- 4 User Preferences – opens the User Preferences window that you use to specify start-up options, set the default directories for new projects, set up server profiles for client/server projects, and specify the HTML report items that you want to display when you run the Reporter node. You can also open the User Preferences window from the main menu:

Options ► User preferences

- 5 Print Diagram – prints the current process flow diagram. You can also perform this task by using the main menu:

File ► Print

- 6 Copy Diagram to Clipboard – copies the diagram to the clipboard. You can also perform this task by using the main menu:

Edit ► Copy diagram to clipboard

- 7 What's This? – turns the mouse pointer into a question mark, indicating that Enterprise Miner is in Help mode. Click the question mark inside an Enterprise Miner window to get Help on the contents of the selected window frame.

- 8 Getting Started – launches the Enterprise Miner 4.3 online tutorial.

Using the Pop-Up Menus

Many of the functions that are found in the Enterprise Miner main menus are available in pop-up menus. Right-click a GUI object within Enterprise Miner to get pop-up menus.

For example, instead of using the main menu to insert a diagram, follow these steps:

- 1 Select the Diagrams tab of the Project Navigator and right-click either the Project folder, the diagram icon, or anywhere in the background of the tab. A pop-up menu appears.
- 2 Select the **New Diagram** pop-up menu item. A new diagram is added to the project.

Unavailable and Dimmed Items

At times particular window objects and menu items are unavailable for use. When a window object or menu item is unavailable, it appears dimmed and does not respond to mouse clicks. When the data mining settings are changed so that the object's use is contextually appropriate, then the objects become active and will accept user input.

When you configure statistical settings within a node, dimmed objects prevent you from selecting inappropriate option combinations. Selecting some options enables additional option settings that you should specify to completely configure your data mining node.

Enterprise Miner Documentation

For more information, see

- *Enterprise Miner Reference Help* – describes how to use each Enterprise Miner node as well as other features of Enterprise Miner, such as the Target Profiler. Help is available through the Enterprise Miner main menu:

Help ► EM Reference

Note: You can search for text strings in the EM Reference Help, using the menu choices:

Edit ► Find

and typing in a search string. △

- *Frame Help* – Enterprise Miner nodes have Help available at the window frame level. For Help on a field or object that is contained in a particular window frame:
 - 1 Click the **what's This?** tool icon in the Enterprise Miner Toolbox. The mouse pointer changes into a question mark.
 - 2 Click inside the window that you would like help on.
- *Add-In Tools for SAS/Warehouse Administrator Software* – To download information about these tools, go to the SAS/Warehouse Administrator product page: <http://support.sas.com/rnd>.
- *Data Mining Using SAS Enterprise Miner: A Case Study Approach, Second Edition* – provides simple directions about how to begin accessing data and how to complete an analysis.

