SAS Enterprise Miner: An Overview

Introduction:

SAS Enterprise Miner support the entire data mining process. It streamlines the data mining process to create highly accurate predictive and descriptive models based on analysis of vast amounts of data from across an enterprise. Data mining is applicable in a variety of industries and provides methodologies for such diverse business problems as fraud detection, householding, customer retention and attrition, database marketing, market segmentation, risk analysis, affinity analysis, customer satisfaction, bankruptcy prediction, and portfolio analysis. SAS Enterprise Miner provides the data access to relational and detail data stores. The base SAS language provides unrivalled power in aggregating and transforming data. Together, SAS/STAT and Enterprise Miner can support virtually any modeling need. Enterprise Miner functions are organized into a process known as SEMMA: **S**ample,**E**xplore,**M**odify,**M**ode,**A**ssess. Within each of the SEMMA categories, Enterprise Miner provides a number of tools to promote the process of data mining.

Sample the data by creating one or more data sets. The sample should be large enough to contain significant information, yet small enough to process. This step includes the use of data preparation tools for data import, merge, append, and filter, as well as statistical sampling techniques.

Explore the data by searching for relationships, trends, and anomalies in order to gain understanding and ideas. This step includes the use of tools for statistical reporting and graphical exploration, variable selection methods, and variable clustering.

Modify the data by creating, selecting, and transforming the variables to focus the model selection process. This step includes the use of tools for defining transformations, missing value handling, value recoding, and interactive binning.

Model the data by using the analytical tools to train a statistical or machine learning model to reliably predict a desired outcome. This step includes the use of techniques such as linear and logistic regression, decision trees, neural networks, partial least squares, LARS and LASSO, nearest neighbor, and importing models defined by other users or even outside SAS Enterprise Miner.

Assess the data by evaluating the usefulness and reliability of the findings from the data mining process. This step includes the use of tools for comparing models and computing new fit statistics, cutoff analysis, decision support, report generation, and score code management.

You might or might not include all of the SEMMA steps in an 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 SEMMA steps, you can apply a scoring formula from one or more champion models to new data that might or might not contain the target variable. Scoring new data that is not available at the time of model training is the goal of most data mining problems. Furthermore, advanced visualization tools enable you to quickly and easily examine large amounts of data in multidimensional histograms and to graphically compare modeling results. Scoring new data that is not available at the time of model training is the goal of most data mining problems. SAS Enterprise Miner includes

tools for generating and testing complete score code for the entire process flow diagram as SAS Code, C code, and Java code, as well as tools for interactively scoring new data and examining the results.

How Does SAS Enterprise Miner Work?

In SAS Enterprise Miner, the data mining process is driven by a process flow diagram that you create by dragging nodes from a toolbar that is organized by SEMMA categories and dropping them onto a diagram workspace.
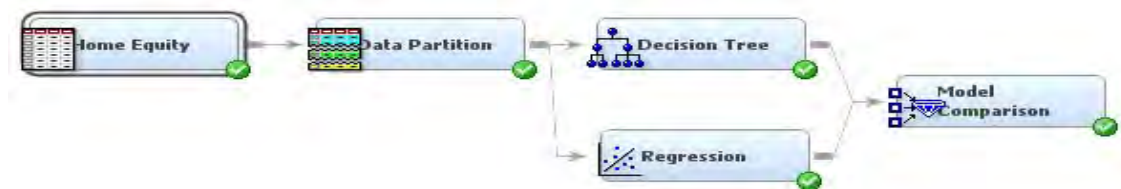

Example Process Flow Diagram

The graphical user interface (GUI) is logically designed in such a way that the business analyst who has little statistical expertise can navigate through the data mining methodology, and the quantitative expert can explore each node in depth to fine-tune the analytical process.

SAS Enterprise Miner automates the scoring process and supplies complete scoring code for all stages of model development in SAS, C, Java, and PMML. The scoring code can be deployed in a variety of real-time or batch environments within SAS, on the Web, or directly in relational databases.

The SAS Enterprise Miner GUI
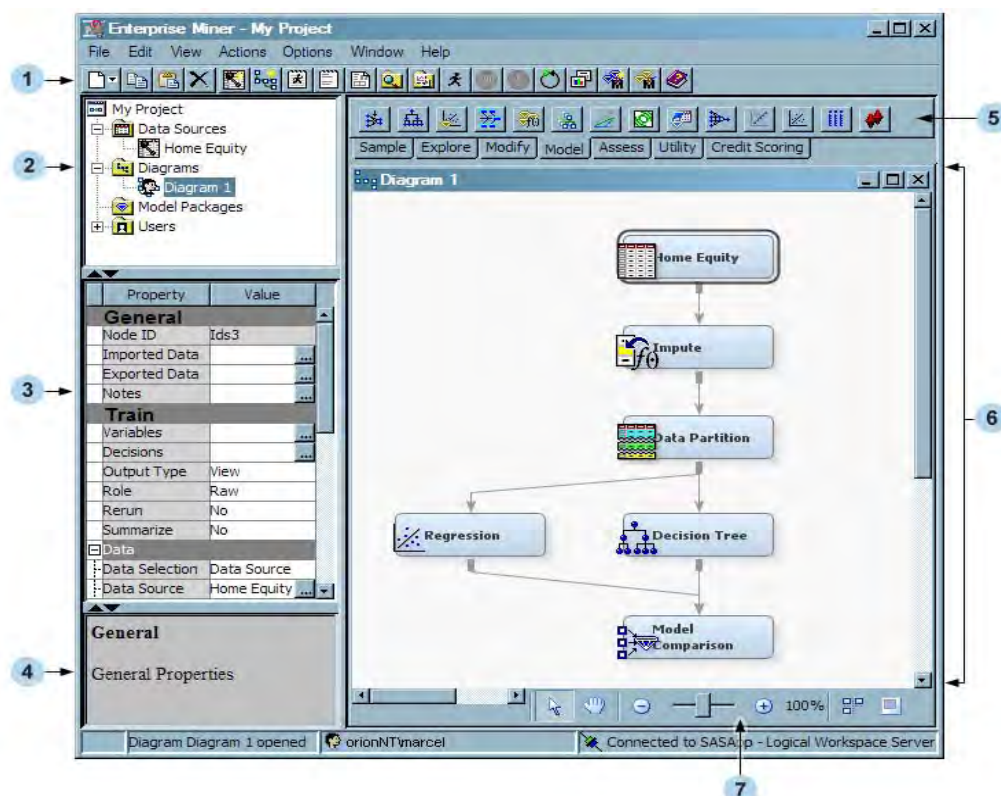
1. **Toolbar Shortcut Buttons** — Use the toolbar shortcut buttons to perform common computer functions and frequently used SAS Enterprise Miner operations. Move the mouse pointer over any shortcut button to see the text name. Click on a shortcut button to use it.

2. **Project Panel** — Use the Project Panel to manage and view data sources, diagrams, results, and project users.

3. **Properties Panel** — Use the Properties Panel to view and edit the settings of data sources, diagrams, nodes, and users.

4. **Property Help Panel** — The Property Help Panel displays a short description of any property that you select in the Properties Panel. Extended help can be found from the Help main menu.

5. **Toolbar** — The Toolbar is a graphic set of node icons that you use to build process flow diagrams in the Diagram Workspace. Drag a node icon into the Diagram Workspace to use it. The icon remains in place in the Toolbar, and the node in the Diagram Workspace is ready to be connected and configured for use in the process flow diagram.

6. **Diagram Workspace** — Use the Diagram Workspace to build, edit, run, and save process flow diagrams. In this workspace, you graphically build, order, sequence, and connect the nodes that you use to mine your data and generate reports.

7. **Diagram Navigation Toolbar** — Use the Diagram Navigation Toolbar to organize and navigate the

Working with E-Miner:

When we start the SAS Enterprise Miner Client, we have to choose either the Personal Workstation or Enterprise Client configuration. We select Personal Workstation when we are using SAS services from our personal computer or laptop computer. On other hand if we select the Enterprise Client then we are connect to remote SAS servers using the SAS Analytics Platform server.
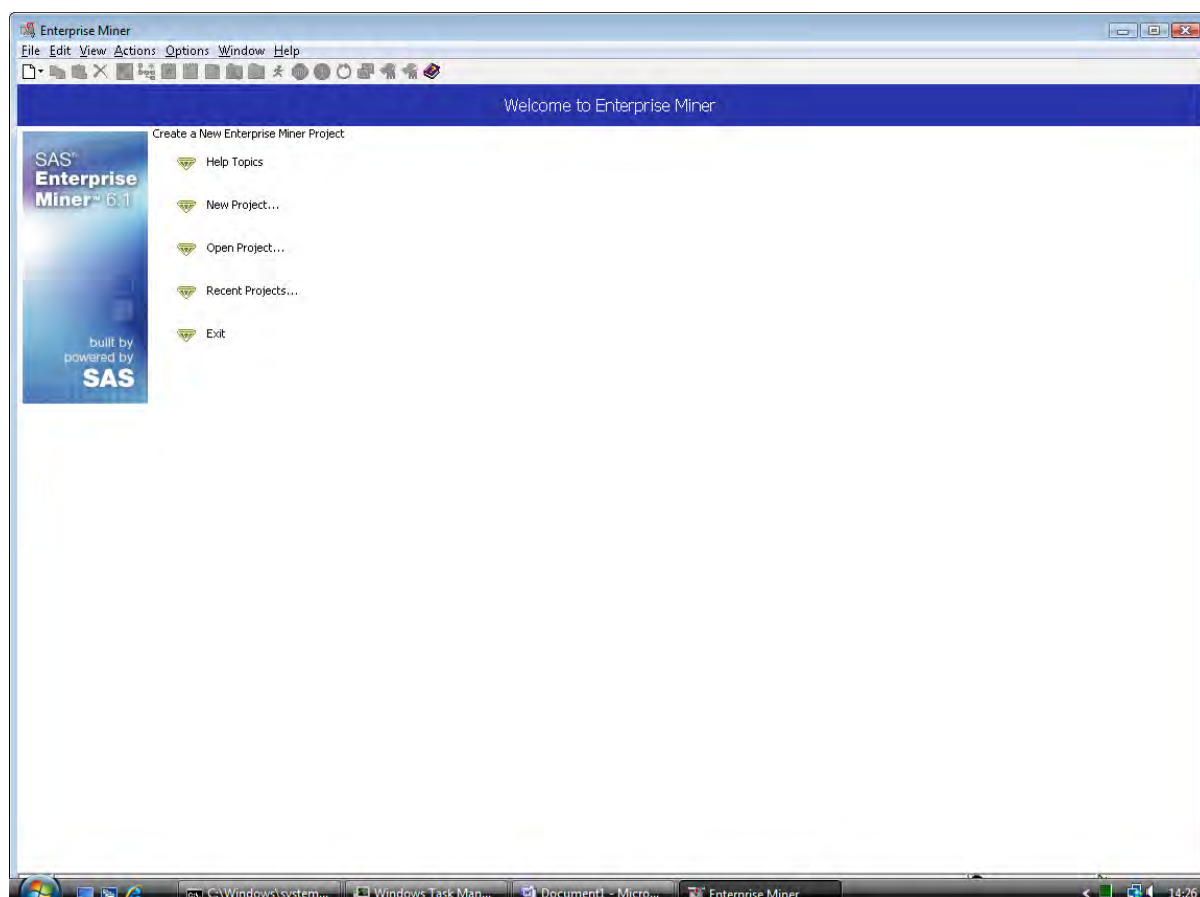
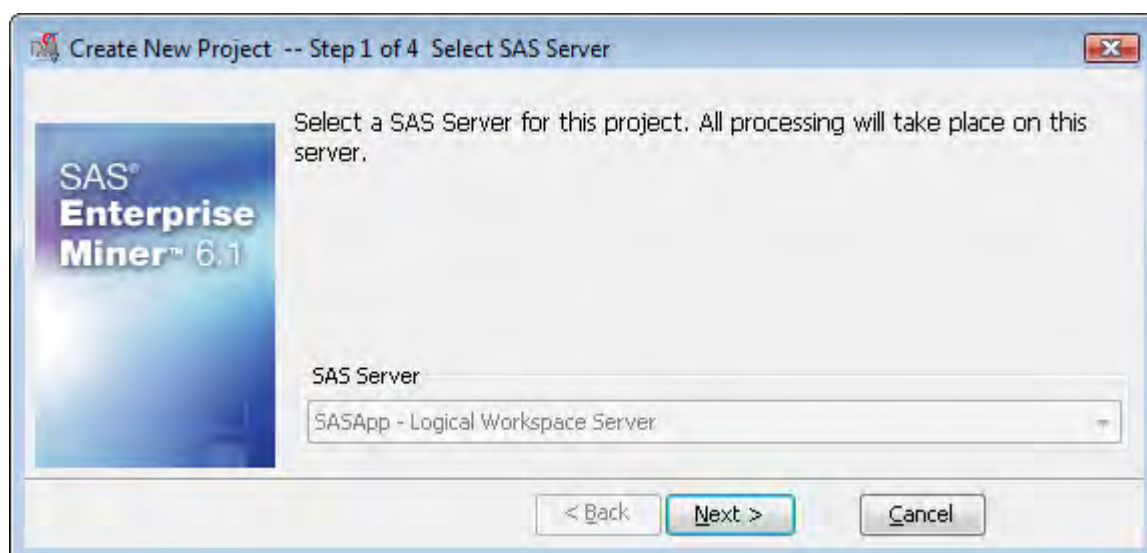Launch Enterprise Miner: SAS→Analytics→ SAS Enterprise Miner Client 6.1

We can check the checkbox "Remember password" only if the machine is used exclusively by the user. If it's a shared/public computer then do not select the checkbox "Remember Password".
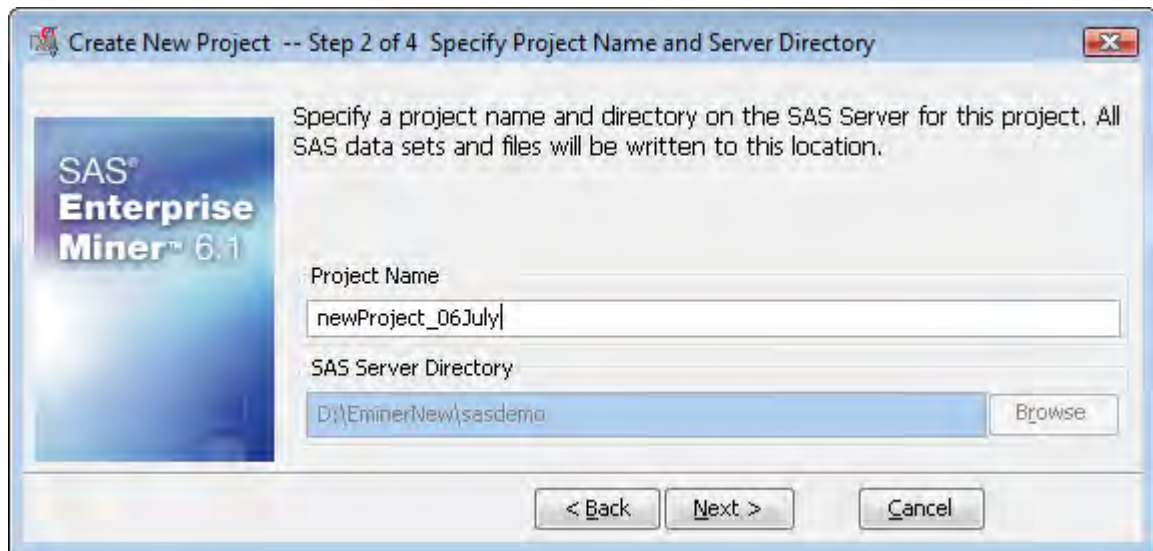
Create New Project

To create a new project click on "New Project" link on the screen.
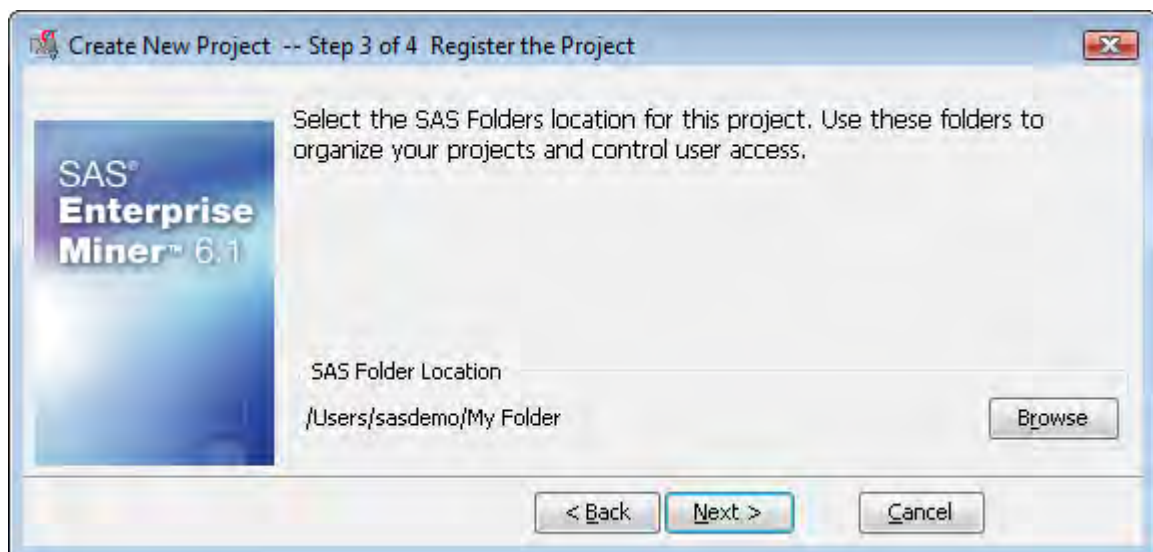
Leave the screen as default .
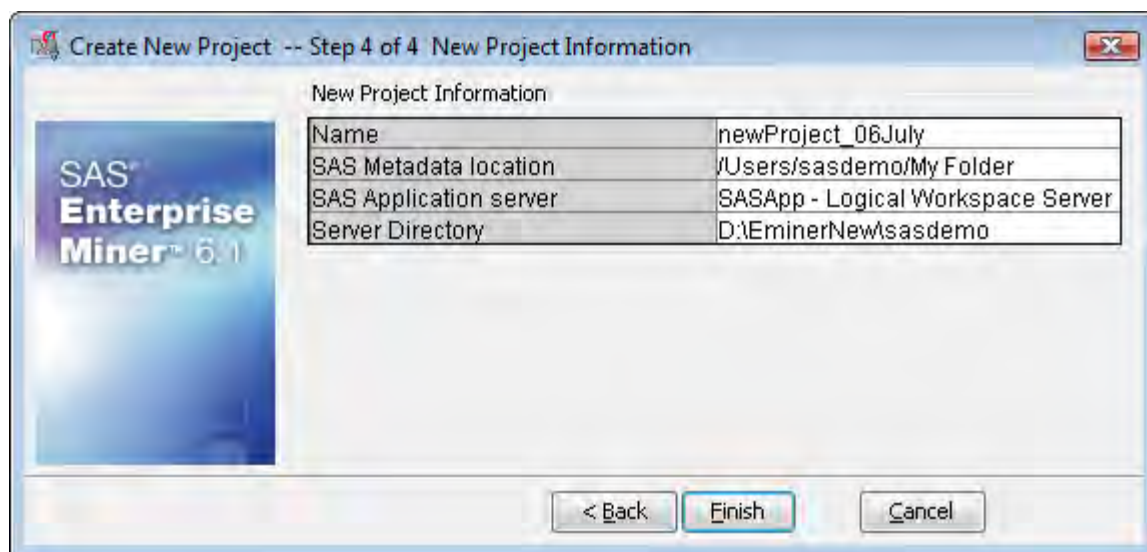

Click on "Next" Button. Give a name to the project.

The project gets created in the path D:\EminerNew\sasdemo

The above mentioned path cannot be edited and also the project path will change as per his login credentials. Click on "Next" button .

The path mentioned in the above screen need not be changed. Now Click on "Next" button.

Review the settings before click on Finish button.
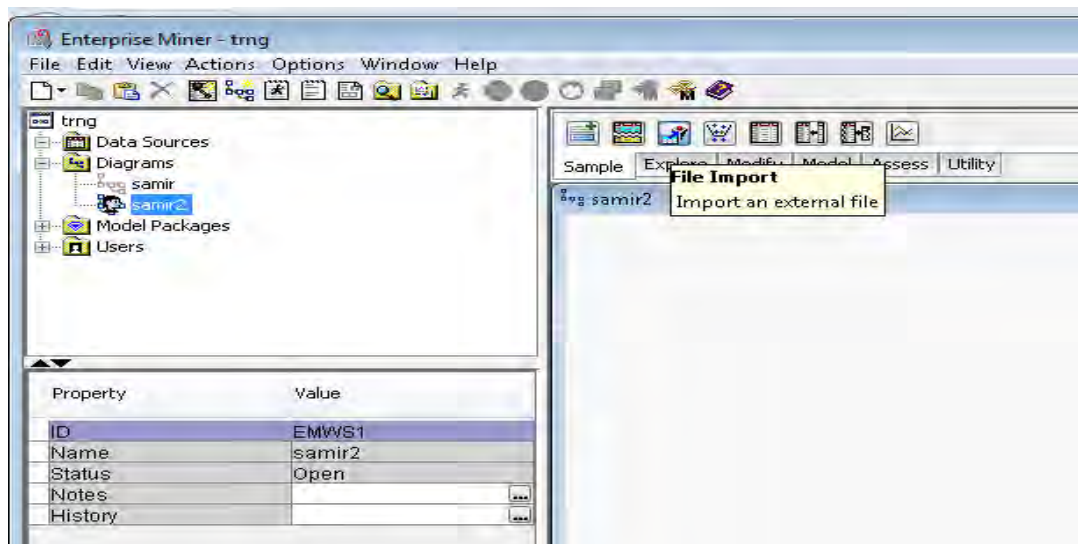

Click on Finish button to finish the creation of the project.
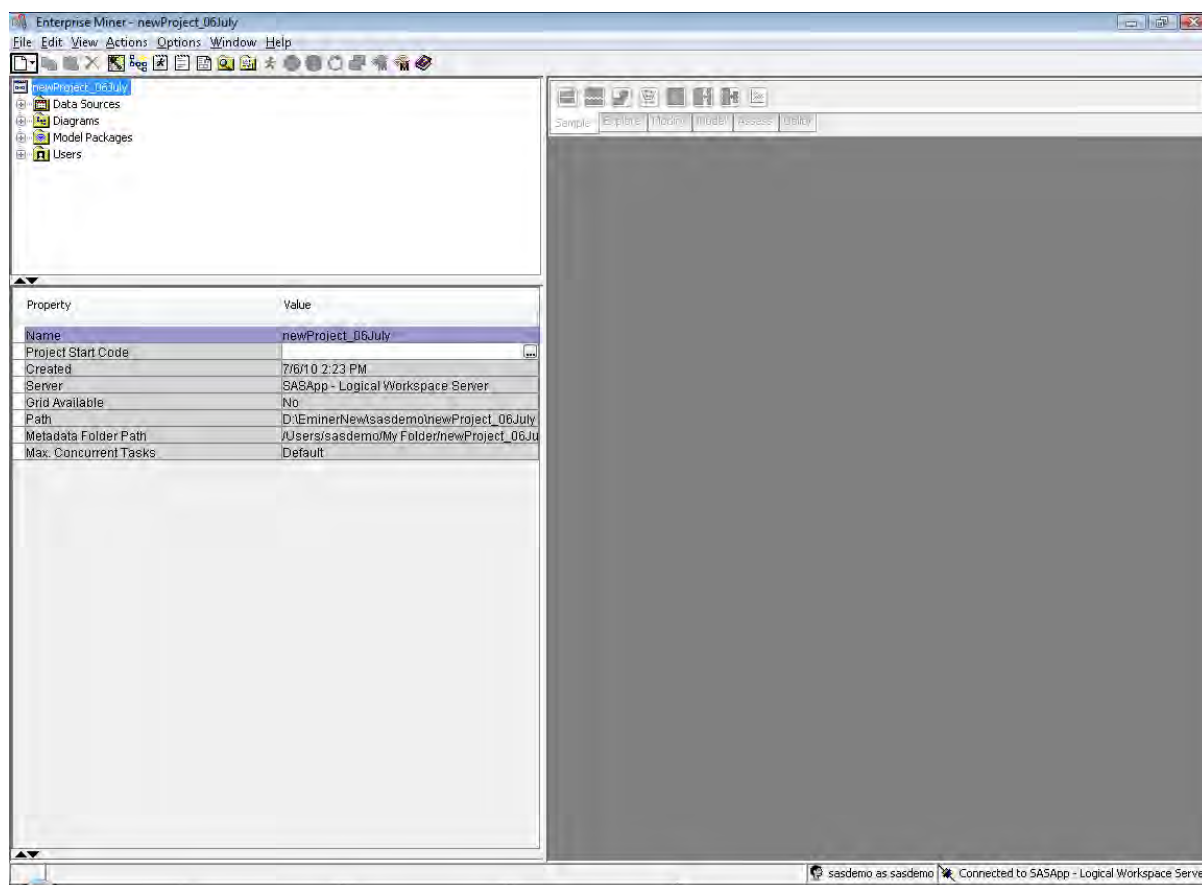
Now we can create a Enterprise Miner model in SAS Enterprise Miner.

Create Diagram

Create a new Diagram: File→ New→ Diagram or click on the icon just below the file which white in colour.
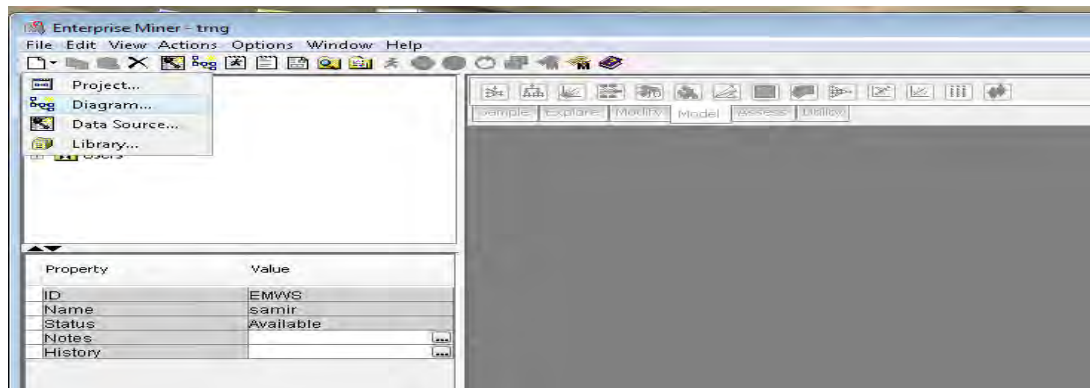
In diagram editor window Click on **Sample** → Select **File Import** and drag drop icon into process window
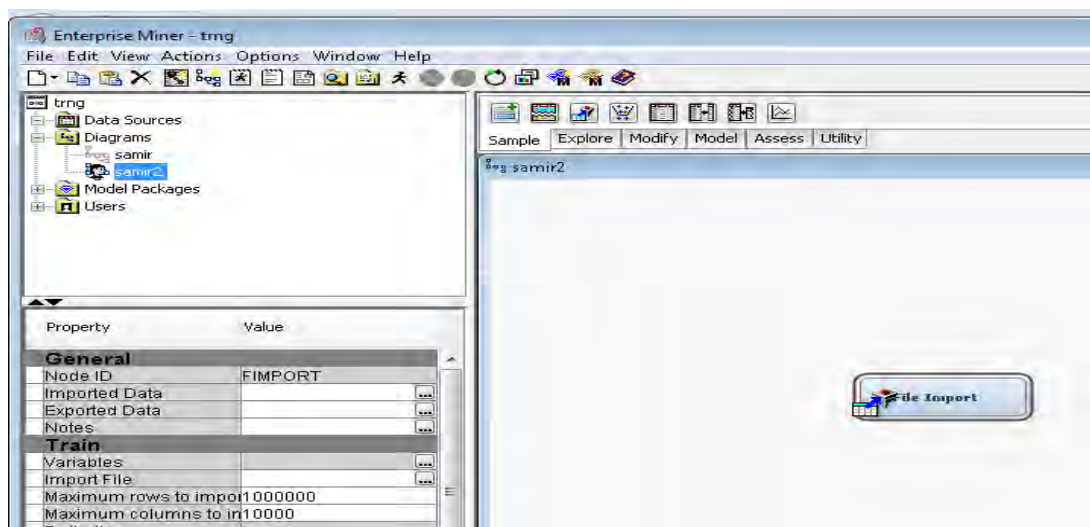
Now we select the file import icon from process window and then   from the property window Click on **Import File→My Computer→Browse** and **select location** and then the required **file →OK.**

 If we wish to see the data file click on **Preview**
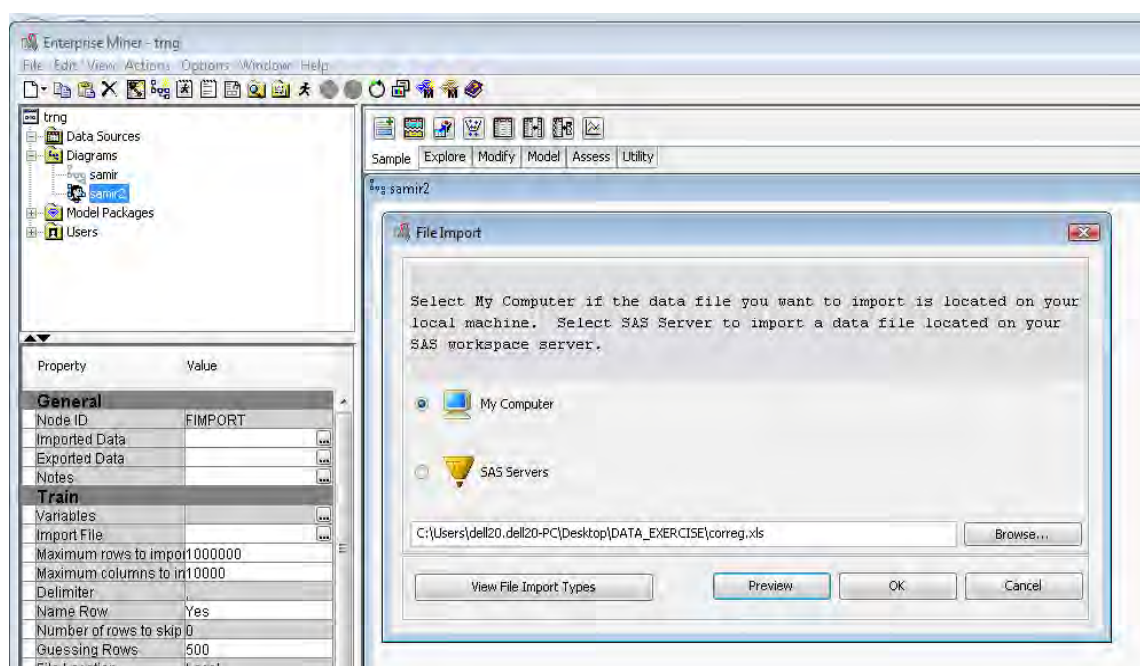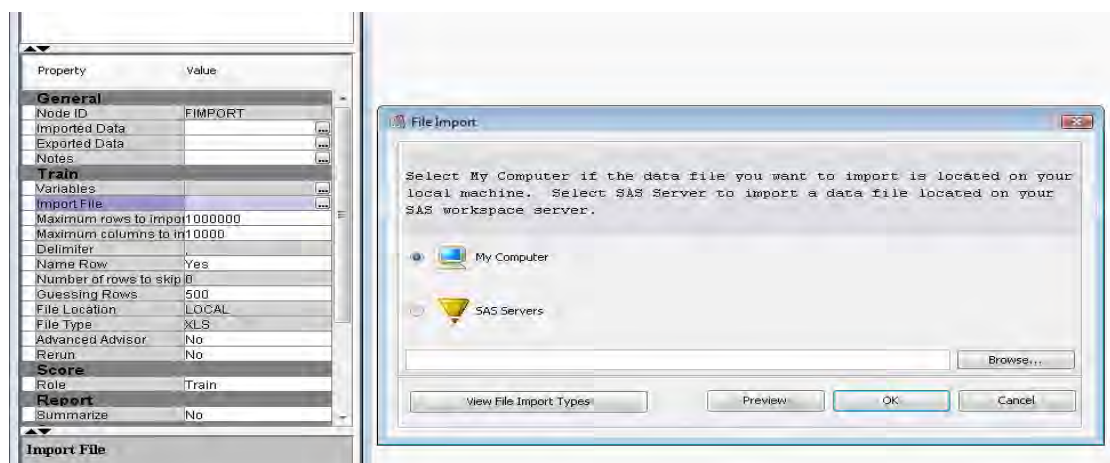
Now Click on **Model tab→Drag and Drop Regression icon** in the diagram panel and provide connection by just dragging the mouse from import file to regression diagram. Now from property window Click open Regression type and select Linear Regression.
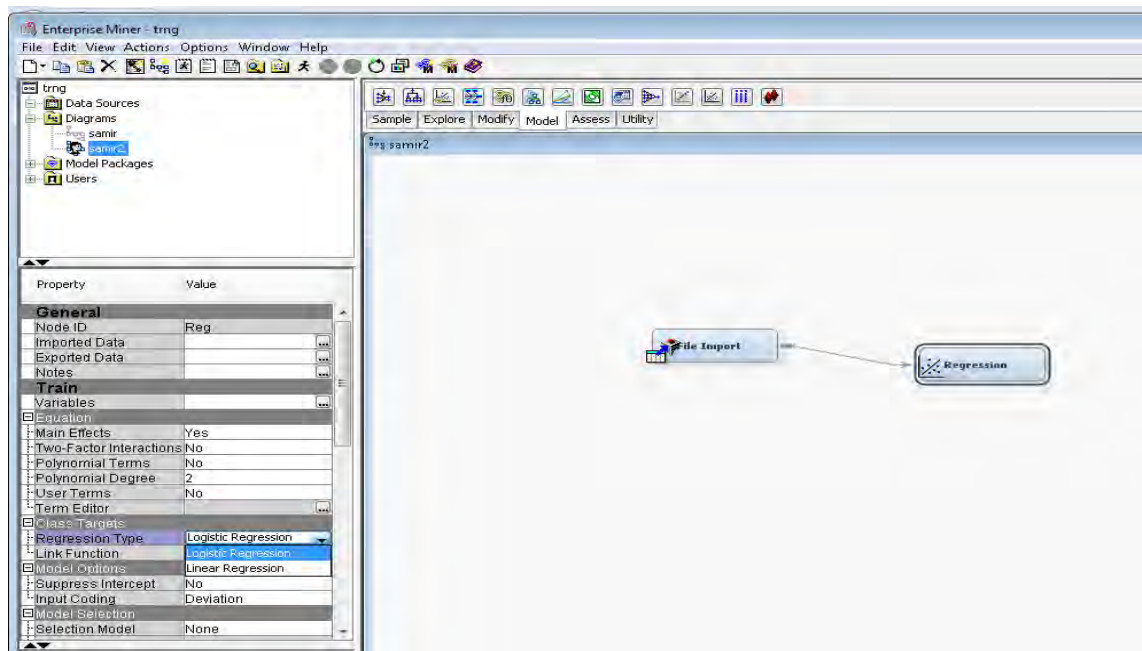
Right Click on **File Import→Edit Variable →Select the Target** (Dependent Variable) from the Role tab.
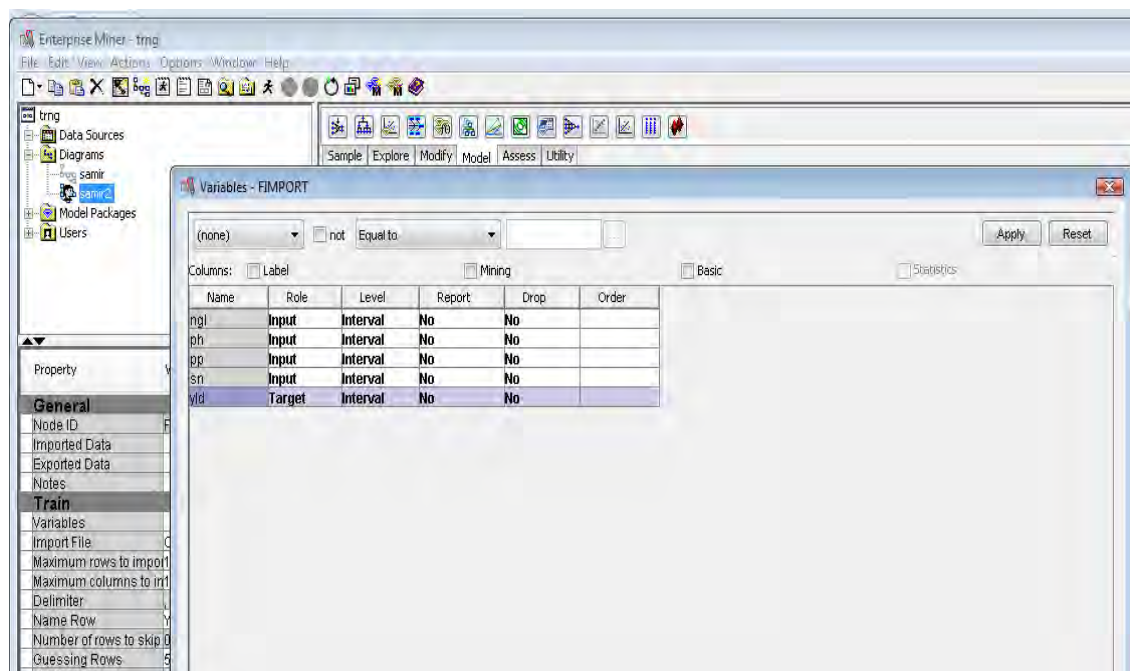

Now Right click on Regression tab and select **Run .** To view the results select result from dialogue box . The result obtained are shown below
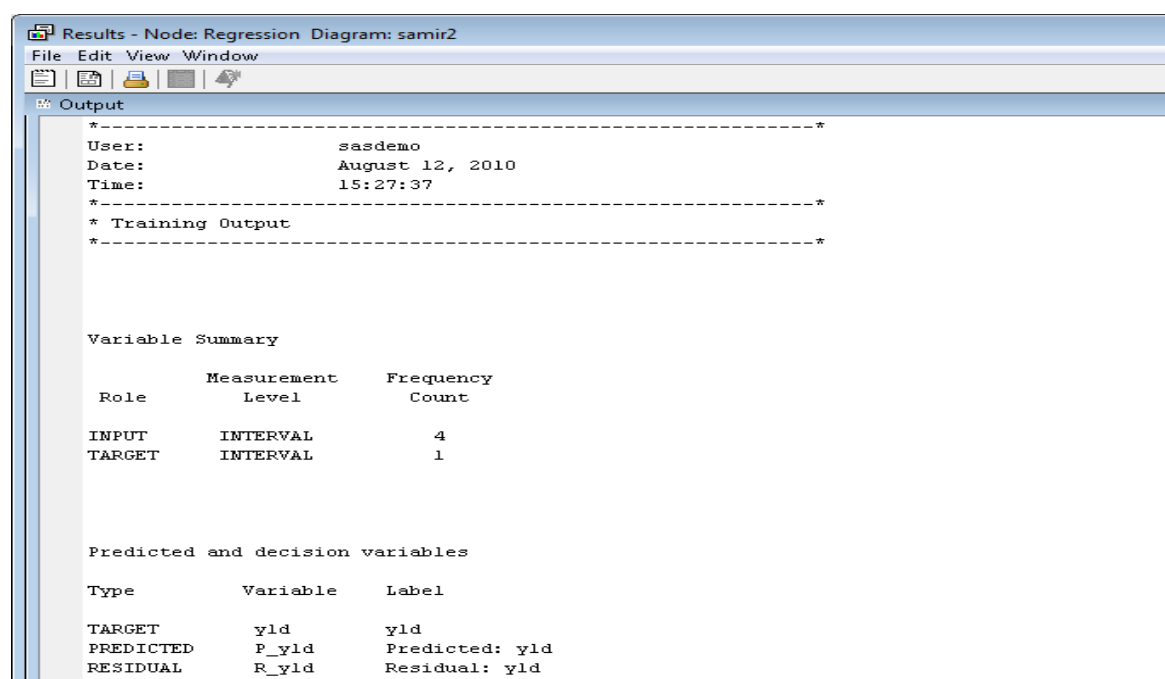
In the current project "sasdemo" user id is a public id which will be shared by users across the institutes in India. There is most likely chance that the projects created by the user "sasdemo" would be deleted automatically after certain period using a batch process. Hence we have to save the project's diagram on our local system that can be accessed later to continue the work. From this saved diagram project can be recreated at the lead centre server within few minute with out change in the work already done.

Saving the Diagram

The created diagram can be saved by Using the command File →Save Diagram(name) as
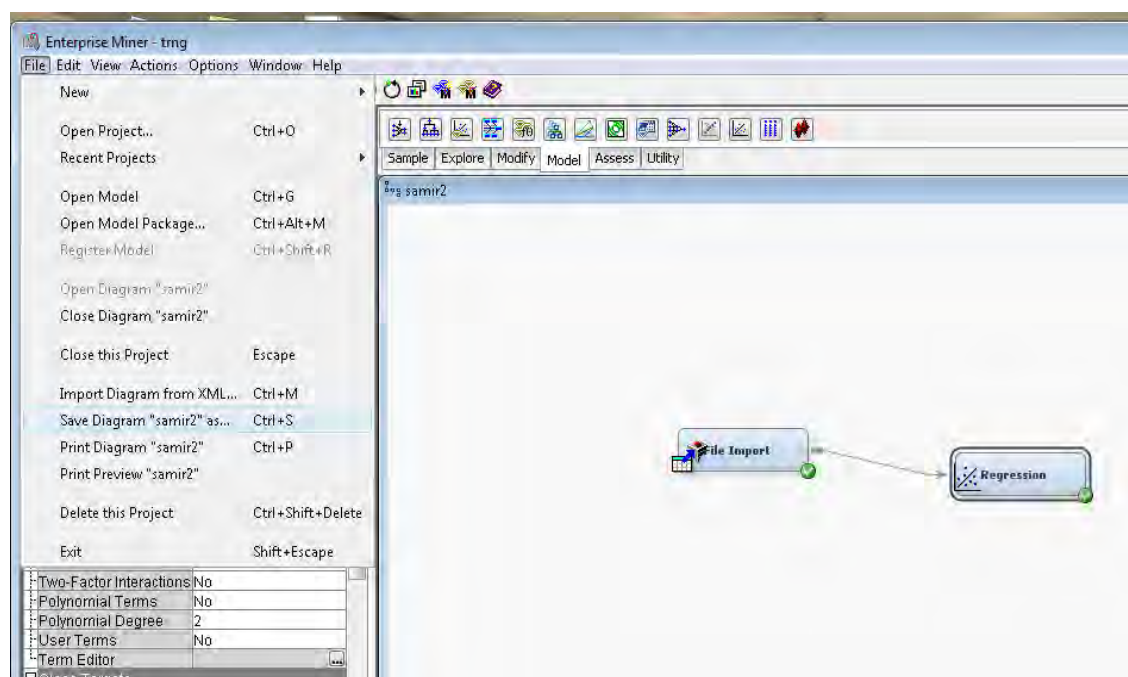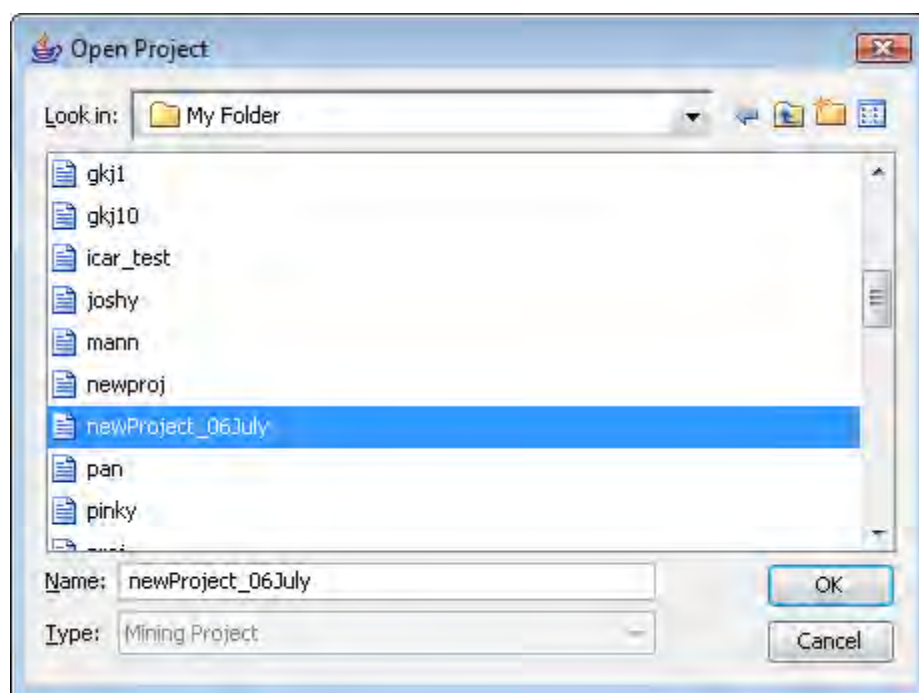
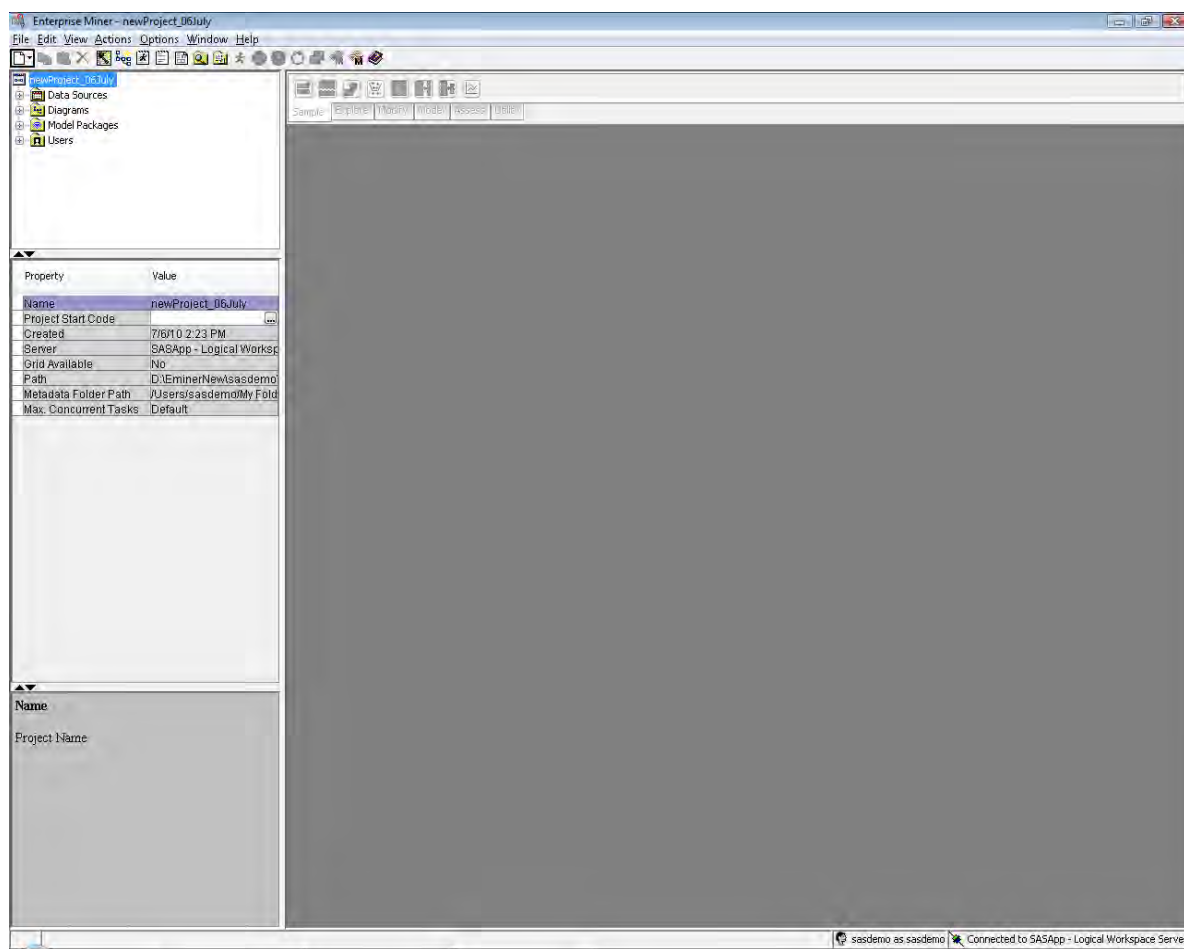Browse to the desired location and save with the some file name. The file will be saved with extension .XML
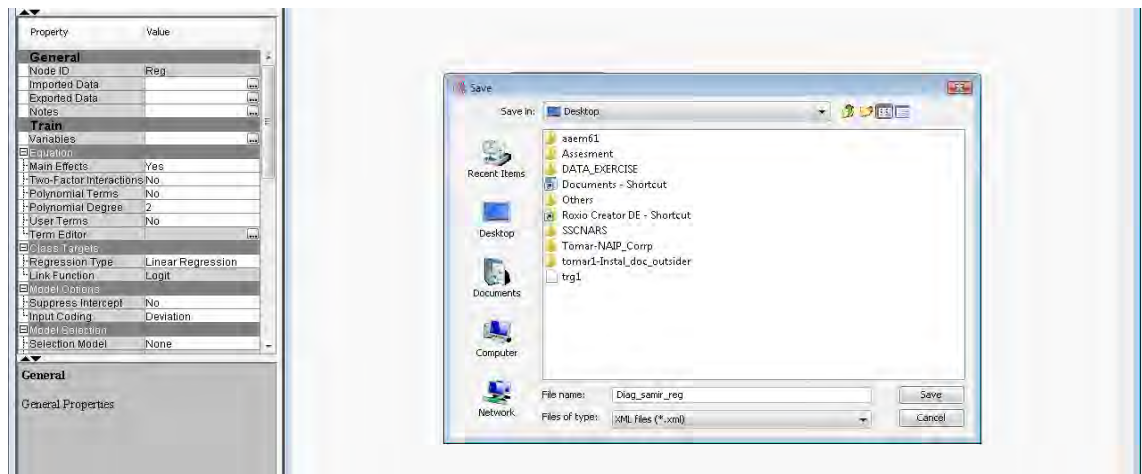
Open an existing project.

Browse to the location: SAS Folders\Users\sasdemo\My Folder

We select the project "newProject_06July Click on "OK" to open the project.

The project opens with the following screen:

To retrieve the saved diagram

Launch Enterprise Miner 6.1 and login using user id sasdemo. Click on " New Project" and create a new project. Now go the file import Diagram and browse the location where the diagram is saved.
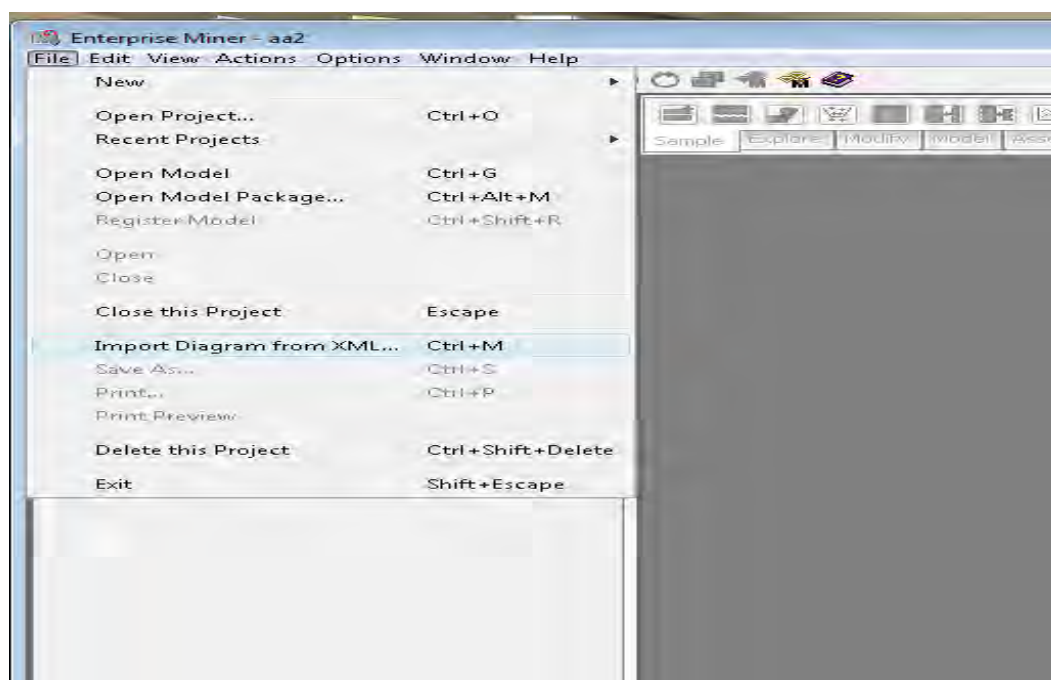
Select the diagram file and then click on open.The diagram will be displayed on the process flow.