Summary: Lesson 2 - Working with Data in a Project

Topic Summaries

Understanding SAS Data Structure

- The terms table, column, and row are commonly used across different types of databases. In SAS, a table is usually called a data set. You will also see the term **variable** for a column, and **observation** for a row.
- The descriptor portion of a SAS data set contains the data set attributes, such as the name, storage location, modified date, column attributes, and the number of rows and columns. You can view the properties of a data set by opening the Properties window from the data pop-up menu.
- All columns in a SAS data set must have three attributes: a name, a type, and a length. In addition, they can have a format, informat, and label.
    - According to standard SAS naming rules, column names can be 1 to 32 characters long. The name must start with a letter or underscore, and the remaining characters can be any combination of letters, underscores, or numbers.
    - A column's type is either character or numeric. Character variables can store any values, such as letters, numbers, special characters, and blanks. Numeric variables can store only numeric values, which can include the digits 0 through 9, a minus sign (-), a single decimal point (.), and E for scientific notation. Date and time variables are numeric variables with a format applied. A SAS date value is the number of days since 01JAN1960. A SAS time value is the number of seconds since midnight.
    - A column's length indicates the number of bytes used to store it. Character values are stored with a length of 1 to 32,767 bytes. One byte equals one character. All numeric variables have a default length of 8 bytes. Numeric values, no matter how many digits they contain, are stored as floating-point numbers in 8 bytes of storage. This provides space for 16 or 17 significant digits.
    - A format can be applied to data values to control how they are displayed in reports and in the data grid. An informat is an instruction that SAS Enterprise Guide uses to read data values into a column.
    - Labels for columns can be up to 256 characters in length and can include spaces and special symbols.

Accessing SAS and DBMS Data

- There are many methods for adding data to a project. In the Open Data window, you can add data from your local computer or network by using the Desktop, My Documents, My Computer, and My Network Places icons. You can add data from a server that is configured to work with SAS Enterprise, and you can add data from a SAS Folder in the SAS Platform for Business Analytics. when you add data to a project, a shortcut, or pointer, is added to the project.
- You can also access SAS and DBMS data by using a SAS library. A SAS library is a collection of files that are defined, recognized, and accessible by SAS. To define a library, you specify a library name, an engine specific to the type of files you want to access, and the location of the files. In a SAS program, you access a data set in a library with the following notation: libraryName.datasetName.
- Some libraries are automatically assigned by SAS and SAS Enterprise Guide, some might be assigned for your site by an administrator, and some you could assign yourself. The **Work** library is automatically created by SAS. The **Work** library points to a temporary folder, so the library is typically used as a location to create temporary data sets to use for reports and analyses. When you exit SAS Enterprise Guide, the data sets in the Work library are deleted. If you need to permanently store a data set, you can specify a different library to use. Custom libraries are typically set up by your SAS administrator. However, but you can use the Assign Project Library task to define libraries to SAS or DBMS data souces.
- After a library is assigned, you can access the data in the library by using the Open Data window. You can also browse libraries in the Server List of the Resources pane. When you add data from a library, a shortcut to the data is added to the Project Tree and Process Flow windows.
- You can explore data in a SAS library by opening a Data Exploration. A Data Exploration enables you to interactively select columns, filter rows, sort, and examine quick statisitics without add the data to your project.

Importing Microsoft Excel and Text Files

The Import Data Wizard enables you to read text or Microsoft Excel files. As you go through the steps in the wizard, you specify and verify information about the input and output data. Based on information you provide, the Import Data Wizard creates a SAS program that reads data from the data source and creates a SAS data set that you can use in analyses and queries.