

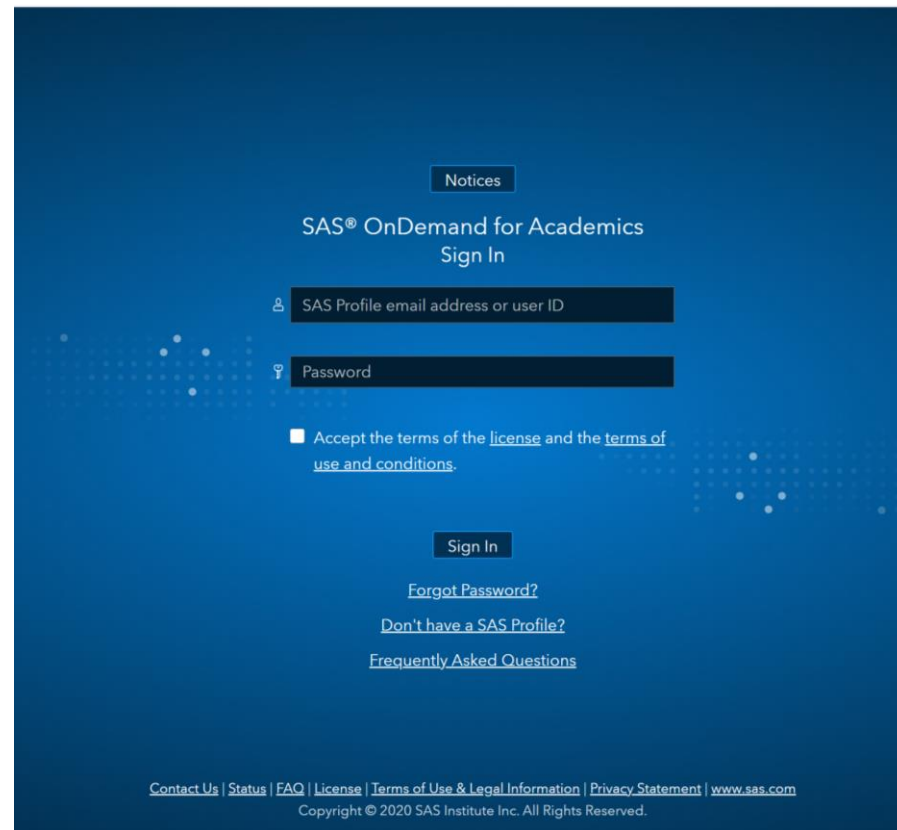
ANLT5050

Unit 5 Assignment 1 Tutorial



Access the SAS OnDemand for Academics Control Center

<https://odamid.oda.sas.com/SASODAControlCenter>



The screenshot shows the SAS OnDemand for Academics Sign In page. The background is a dark blue gradient with a subtle pattern of white dots. At the top, there is a 'Notices' button. Below it, the text 'SAS® OnDemand for Academics' and 'Sign In' are displayed. The sign-in form includes two input fields: 'SAS Profile email address or user ID' and 'Password'. Below the password field, there is a checkbox labeled 'Accept the terms of the [license](#) and the [terms of use and conditions](#)'. A 'Sign In' button is positioned below the checkbox. At the bottom of the form, there are three links: 'Forgot Password?', 'Don't have a SAS Profile?', and 'Frequently Asked Questions'. The footer contains a row of links: 'Contact Us', 'Status', 'FAQ', 'License', 'Terms of Use & Legal Information', 'Privacy Statement', and 'www.sas.com', followed by the copyright notice 'Copyright © 2020 SAS Institute Inc. All Rights Reserved.'



SAS OnDemand for Academics (SODA) Control Center

The screenshot displays the SAS OnDemand for Academics (SODA) Control Center dashboard. At the top, the SAS logo is on the left, and the user's location (United States) and name (Stefanie Reay) are on the right. The main heading is "SAS® OnDemand for Academics Dashboard". Below this, there are tabs for "Planned Events" and "Notices". A navigation bar includes "Applications", "Enrollments", and "Courses". The "Applications" tab is active, showing a list of SAS products:

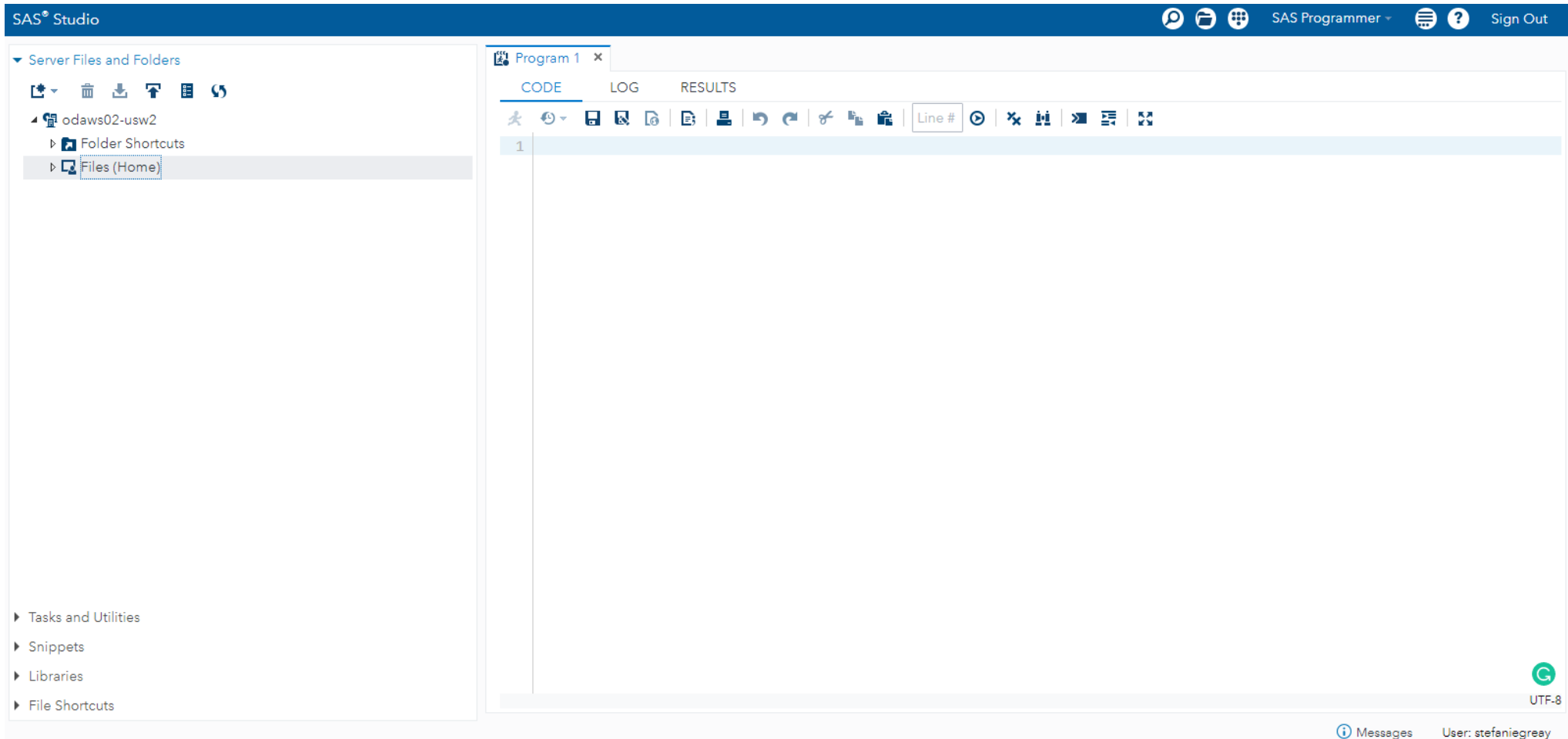
- SAS® Studio**: Write and run SAS code with a Web-based SAS development environment. *Actions: [Clear my saved tabs.](#)*
- SAS® Enterprise Guide®**: Deliver the power of SAS from an easy-to-use, point-and-click interface. ([Download Required](#))
- SAS® Enterprise Miner™**: Reveal valuable insights with powerful data mining software. ([Configuration Steps Required](#)) *Actions: [Clear my project locks.](#)*
- SAS® Forecast Studio**: Generate large numbers of high-quality forecasts automatically. ([Configuration Steps Required](#)) *Actions: [Manage your personal environment.](#)*
- JMP® Software access to SAS® hosted servers**: Statistical discovery software. Users must have a copy of JMP® software. ([Configuration Steps Required](#))

On the right side, there is a "Reference" section with links to the [Support Site](#), [Step-by-Step Reference Guides](#), and [Frequently Asked Questions](#). Below this is a "Quotas (learn more)" section showing progress bars for "Home Directory (46.5MB/5120MB)" at 1% and "Course Directory (207.0MB/3072MB)" at 7%.

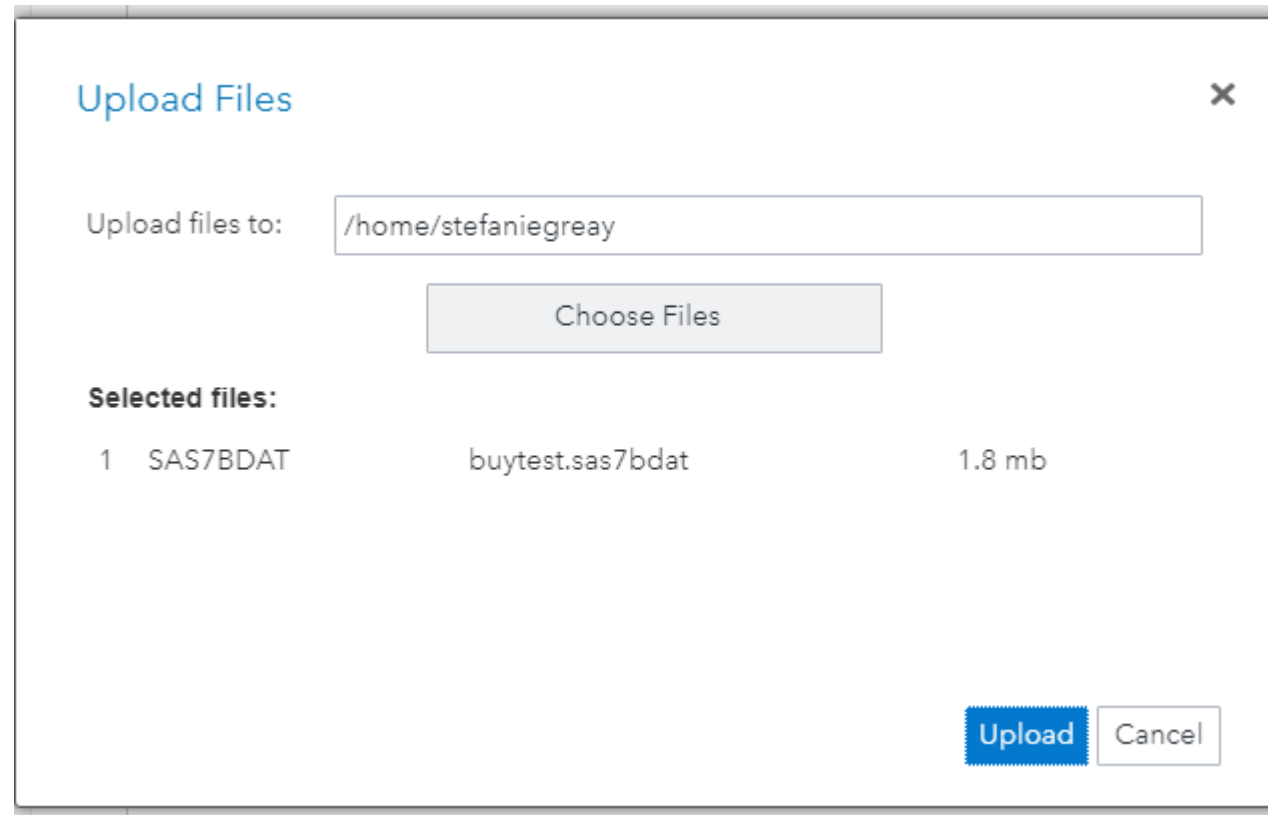
At the bottom, there is a link for "Other Ways to Access SAS® OnDemand for Academics Resources".



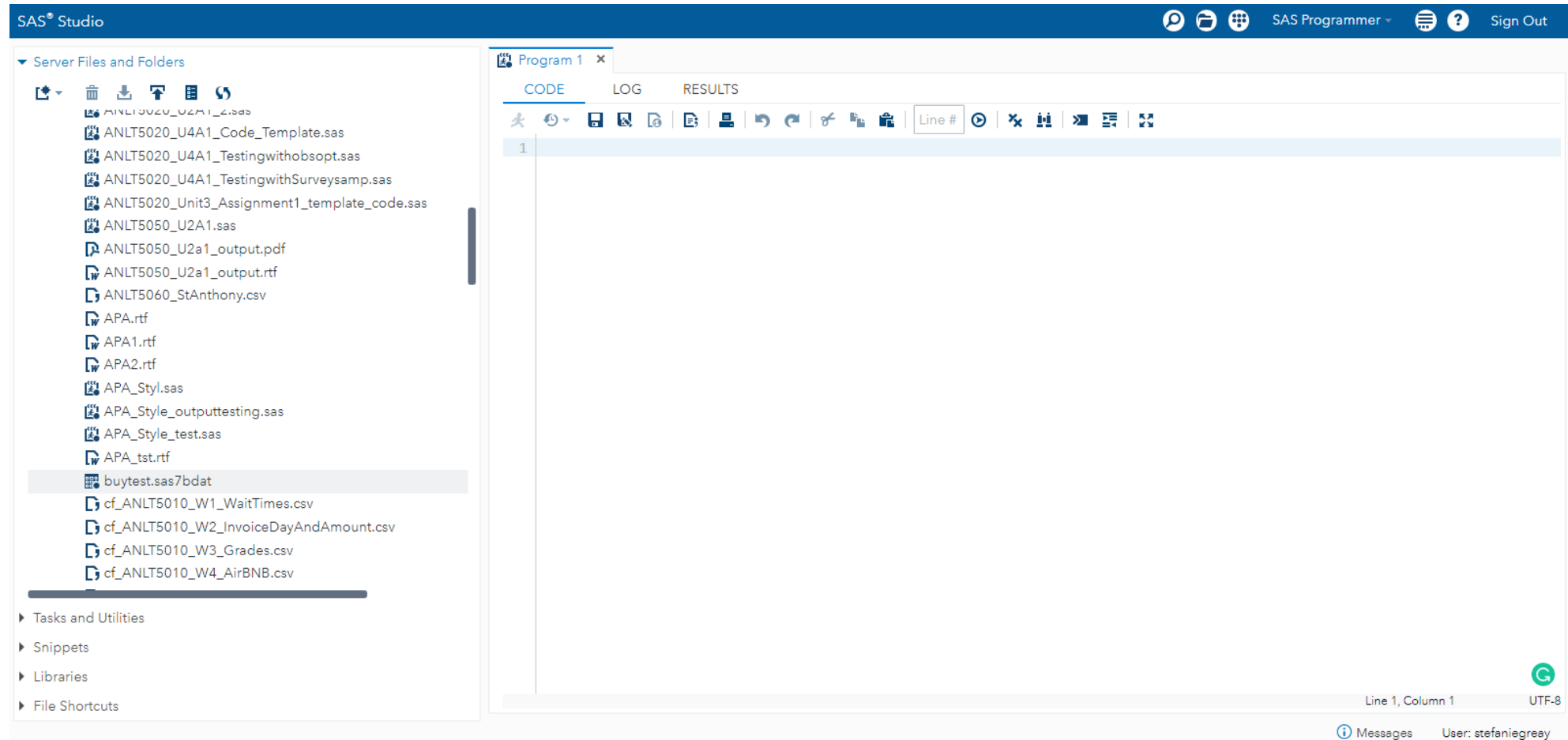
To upload the dataset to the SAS server, open SAS Studio, then click on “Files (Home)” and click the upload button.



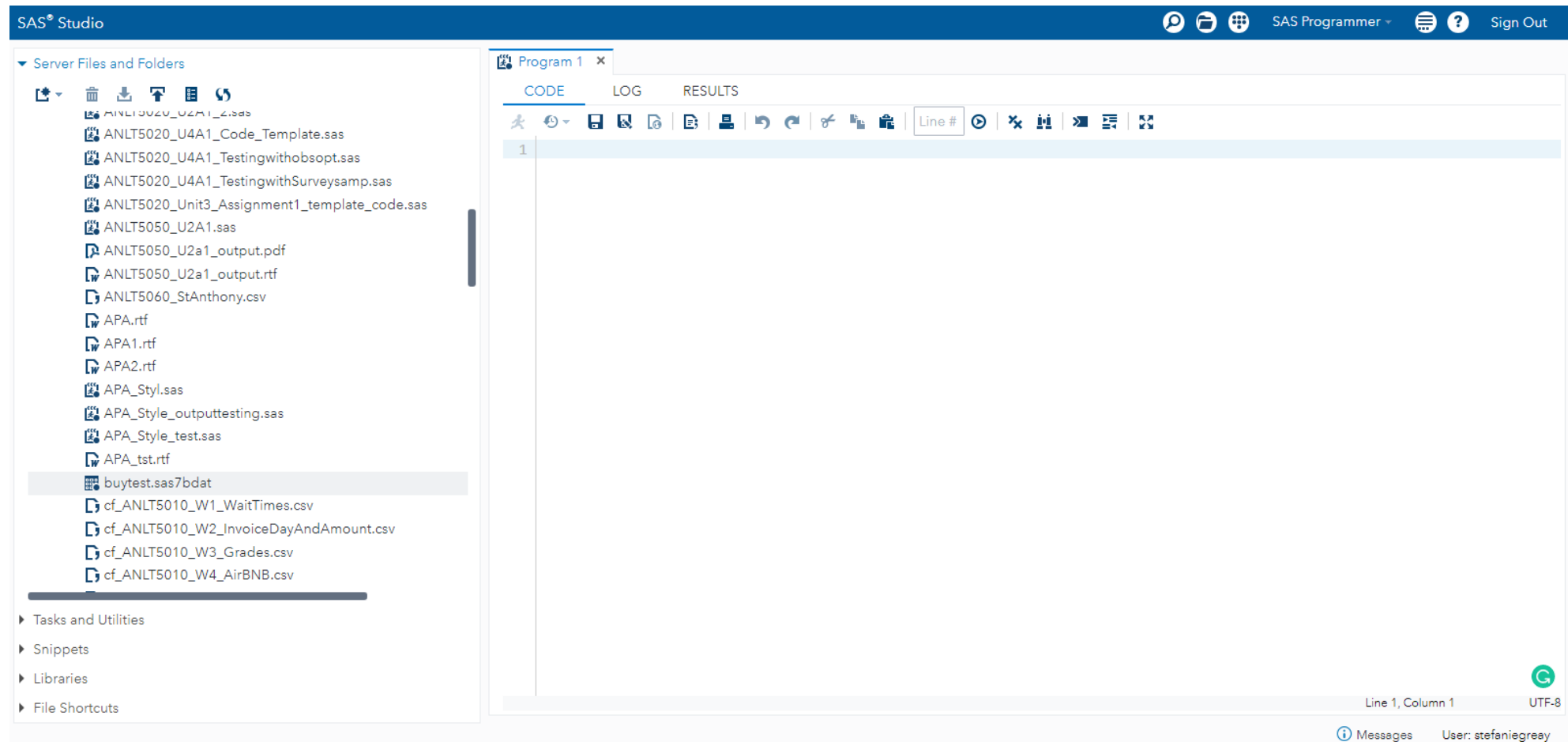
Click on “Choose Files” to browse to the file you want to upload, then click “Upload.”



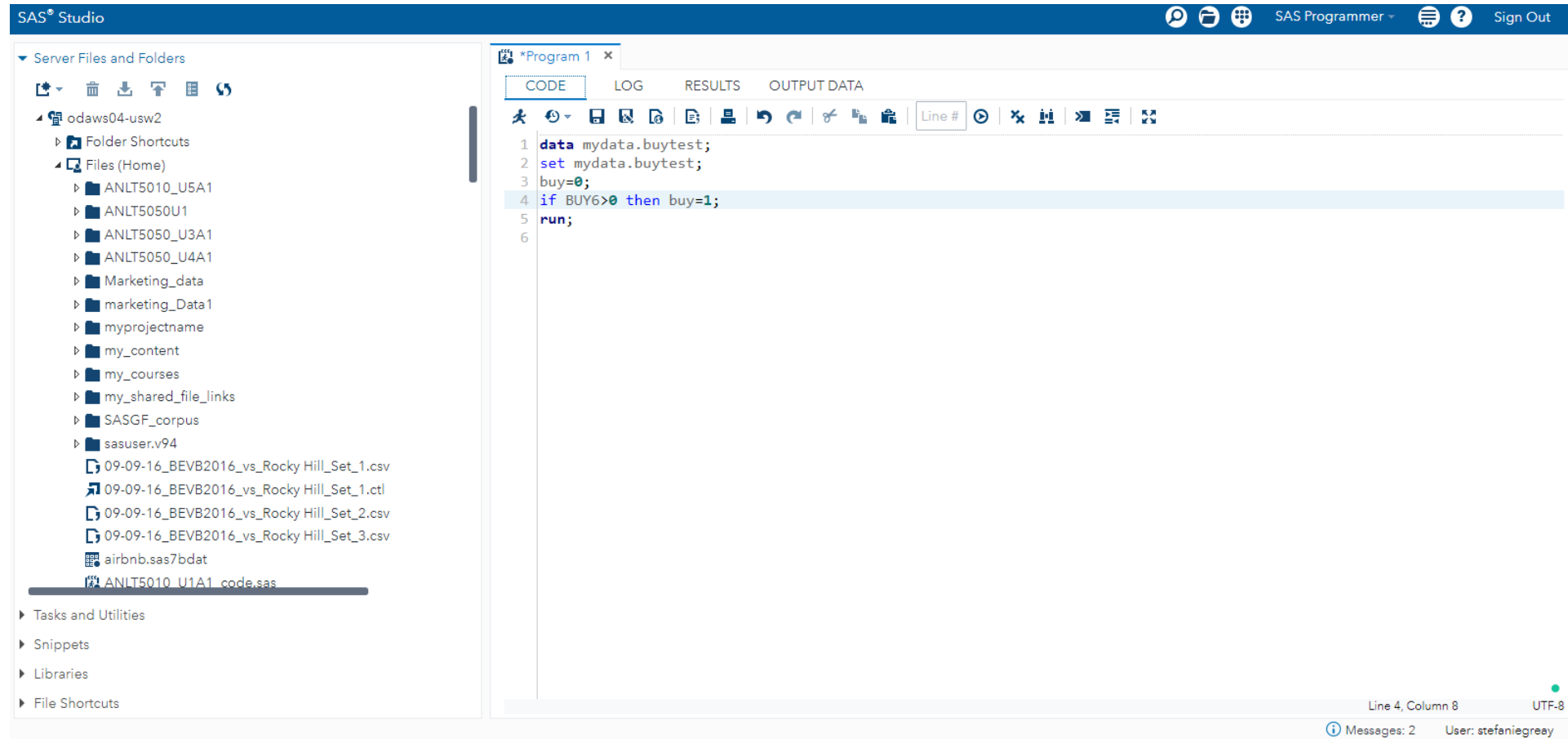
Verify that the upload was successful by scrolling down in your Files(Home) area.



Open a Program in SAS Studio



Use the following code to create a binary variable of “buy” to indicate if the person purchased within the last 6 months, using the existing variable of BUY6. Click “Run”



The screenshot displays the SAS Studio web interface. On the left, the 'Server Files and Folders' pane shows a directory structure under 'odaws04-usw2'. The main area is divided into tabs: 'CODE', 'LOG', 'RESULTS', and 'OUTPUT DATA'. The 'CODE' tab is active, showing a SAS program with the following lines:

```
1 data mydata.buytest;  
2 set mydata.buytest;  
3 buy=0;  
4 if BUY6>0 then buy=1;  
5 run;  
6
```

The status bar at the bottom indicates 'Line 4, Column 8' and 'UTF-8'. The bottom right corner shows 'Messages: 2' and 'User: stefaniegreay'.



Code

```
data mydata.buytest;  
set mydata.buytest;  
buy=0;  
if BUY6>0 then buy=1;  
run;
```



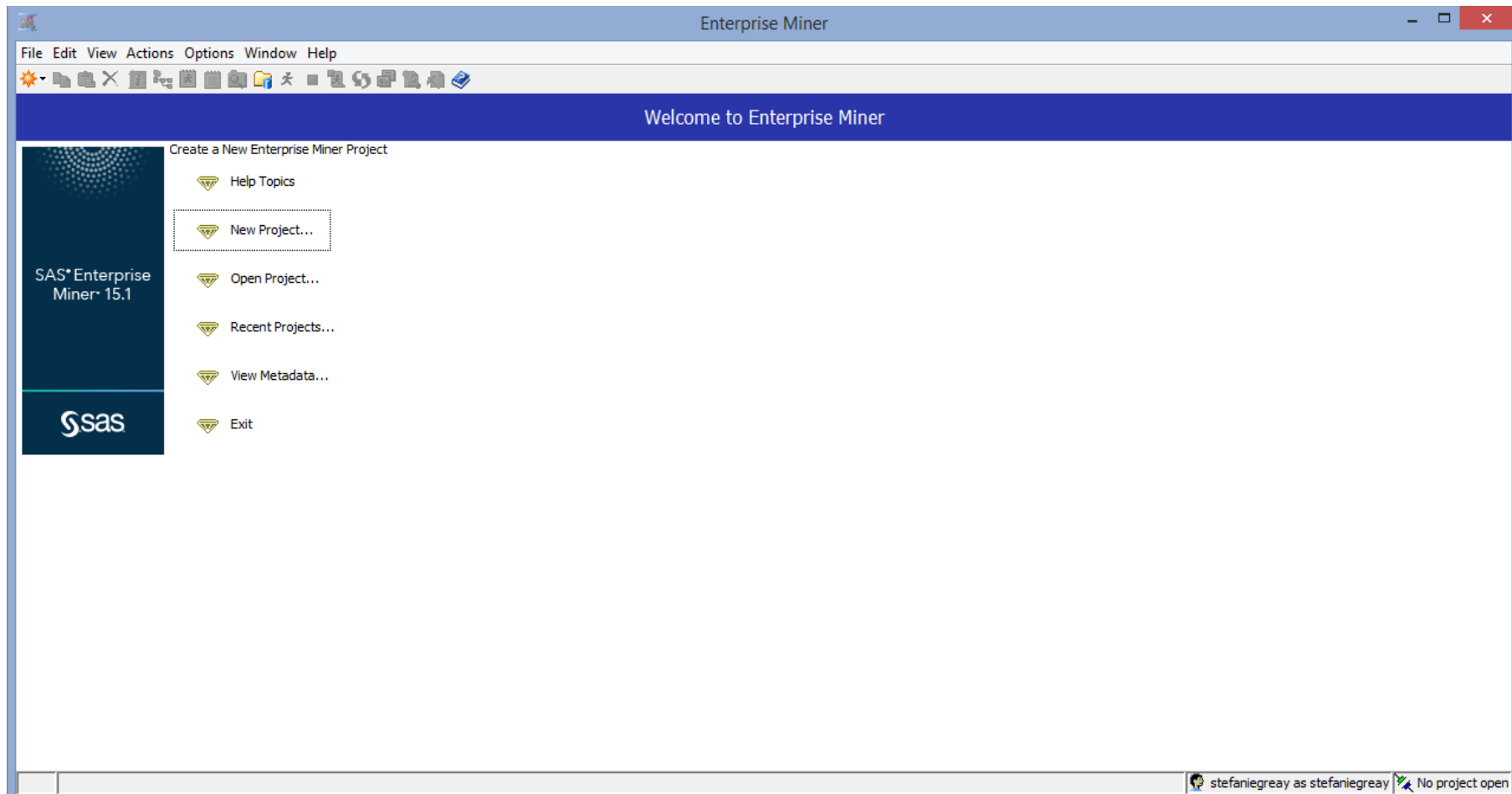
SAS Enterprise Miner Instructions

The following slides provide instructions on how to complete this task in SAS Enterprise Miner.

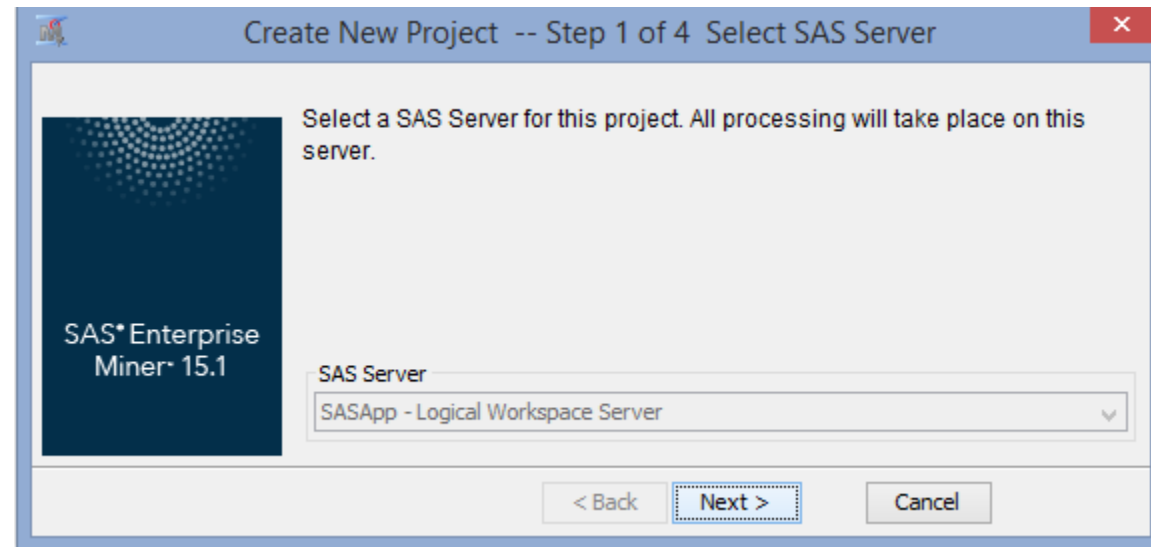
Once you have uploaded the dataset for this unit onto the SAS servers using SAS Studio, you may proceed from here using SAS Enterprise Miner.



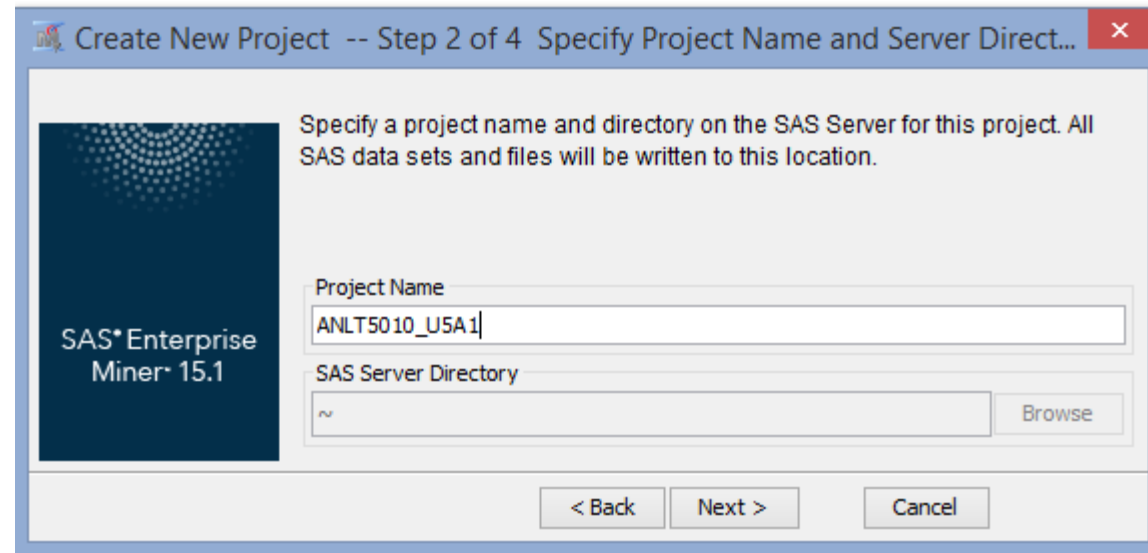
Once you download and start SAS Enterprise Miner, open a new project by clicking on “New Project.”



Click “Next>” to use the default SAS Server



Enter a project name and click “Next>”



Create New Project -- Step 2 of 4 Specify Project Name and Server Direct...

Specify a project name and directory on the SAS Server for this project. All SAS data sets and files will be written to this location.

SAS*Enterprise Miner 15.1

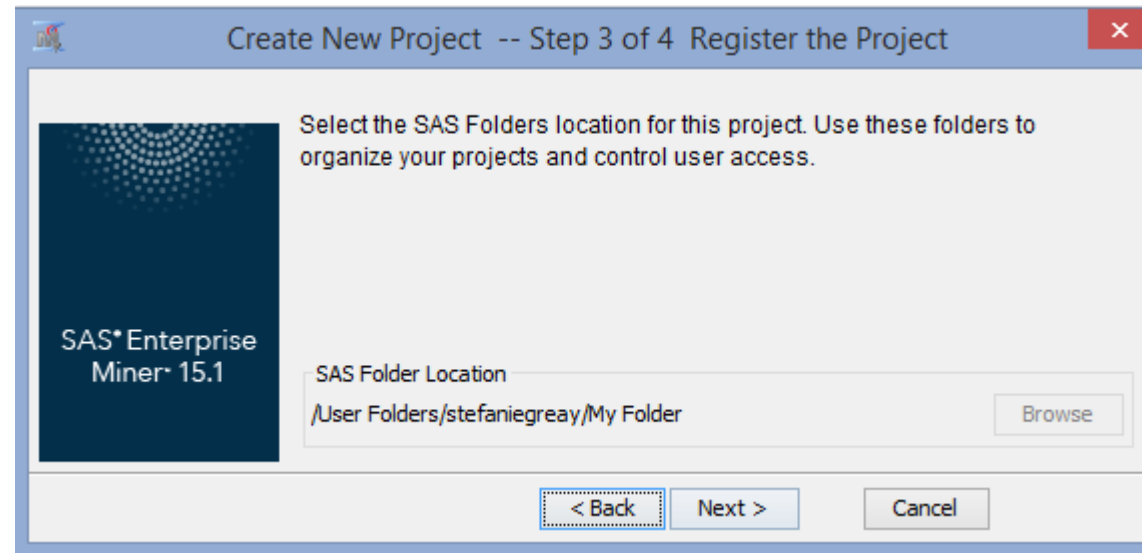
Project Name
ANLT5010_U5A1

SAS Server Directory
~ Browse

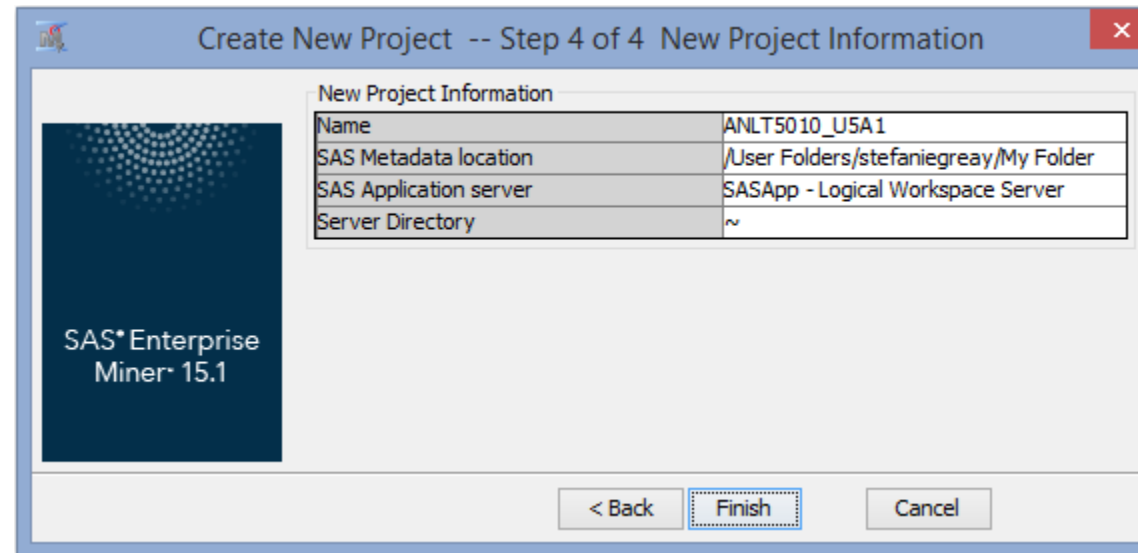
< Back Next > Cancel



Click “Next>”



Verify your entries and click “Finish”



Create New Project -- Step 4 of 4 New Project Information

New Project Information

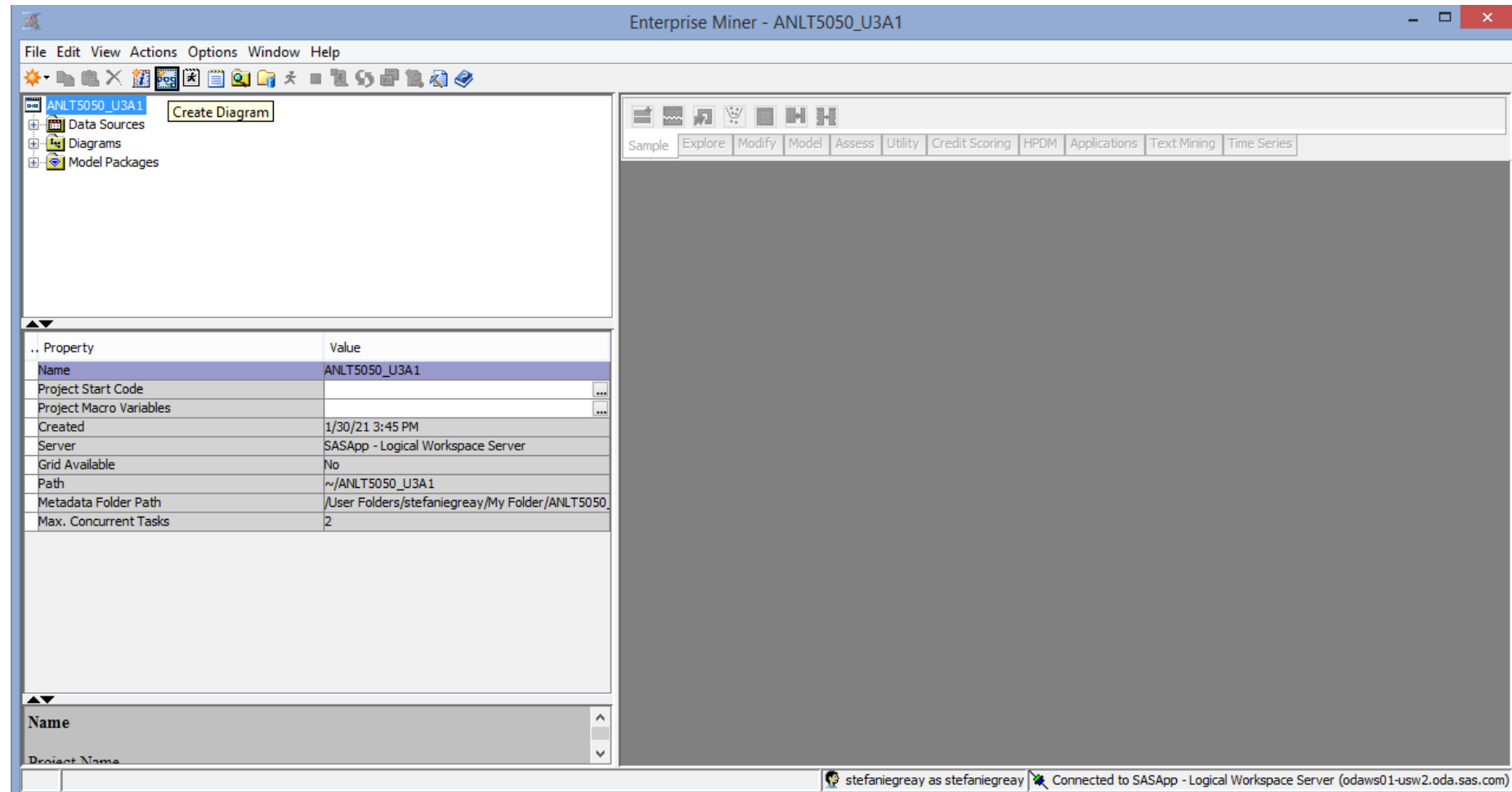
Name	ANLT5010_U5A1
SAS Metadata location	/User Folders/stefaniegreay/My Folder
SAS Application server	SASApp - Logical Workspace Server
Server Directory	~

SAS® Enterprise Miner® 15.1

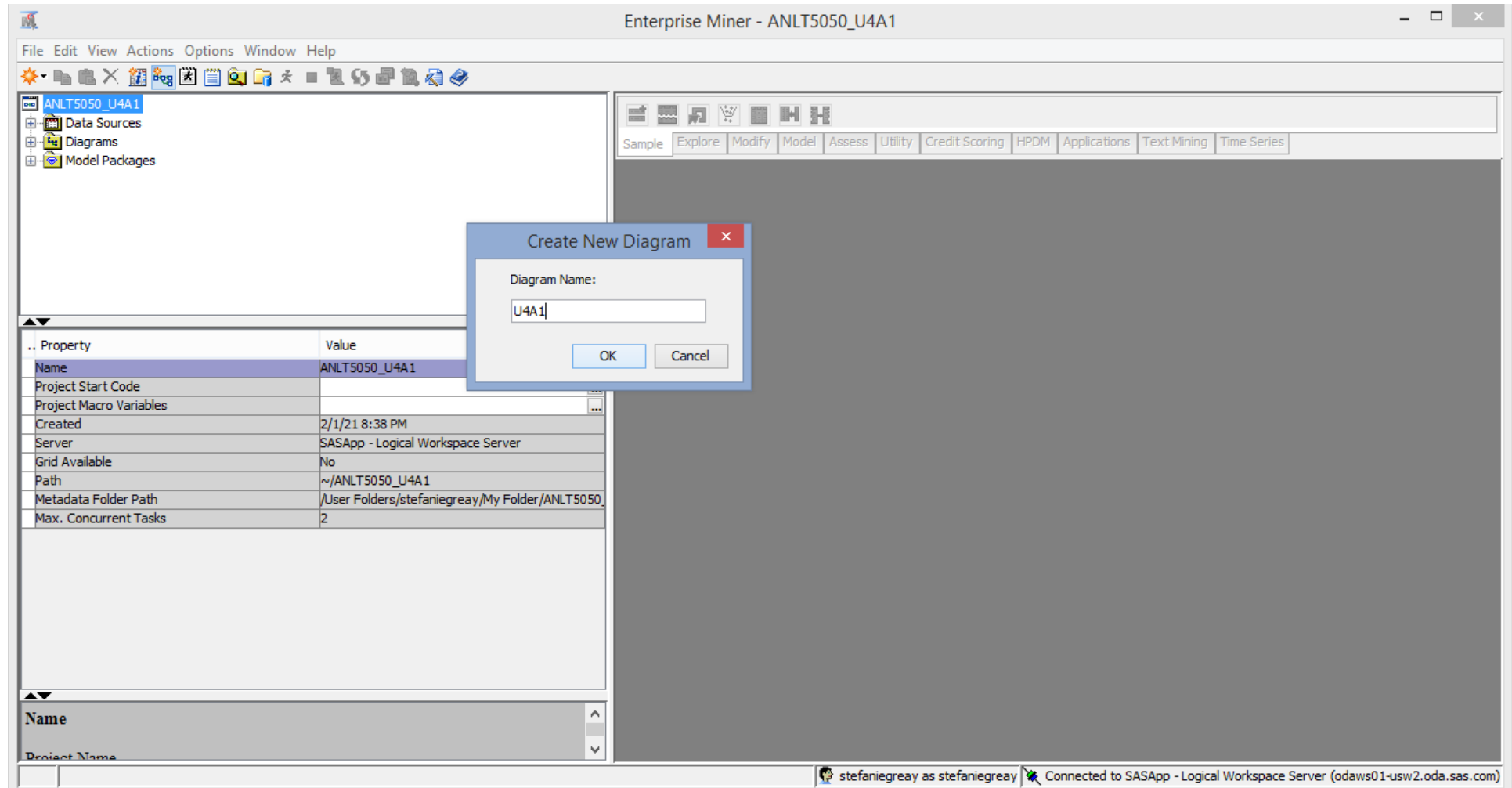
< Back Finish Cancel



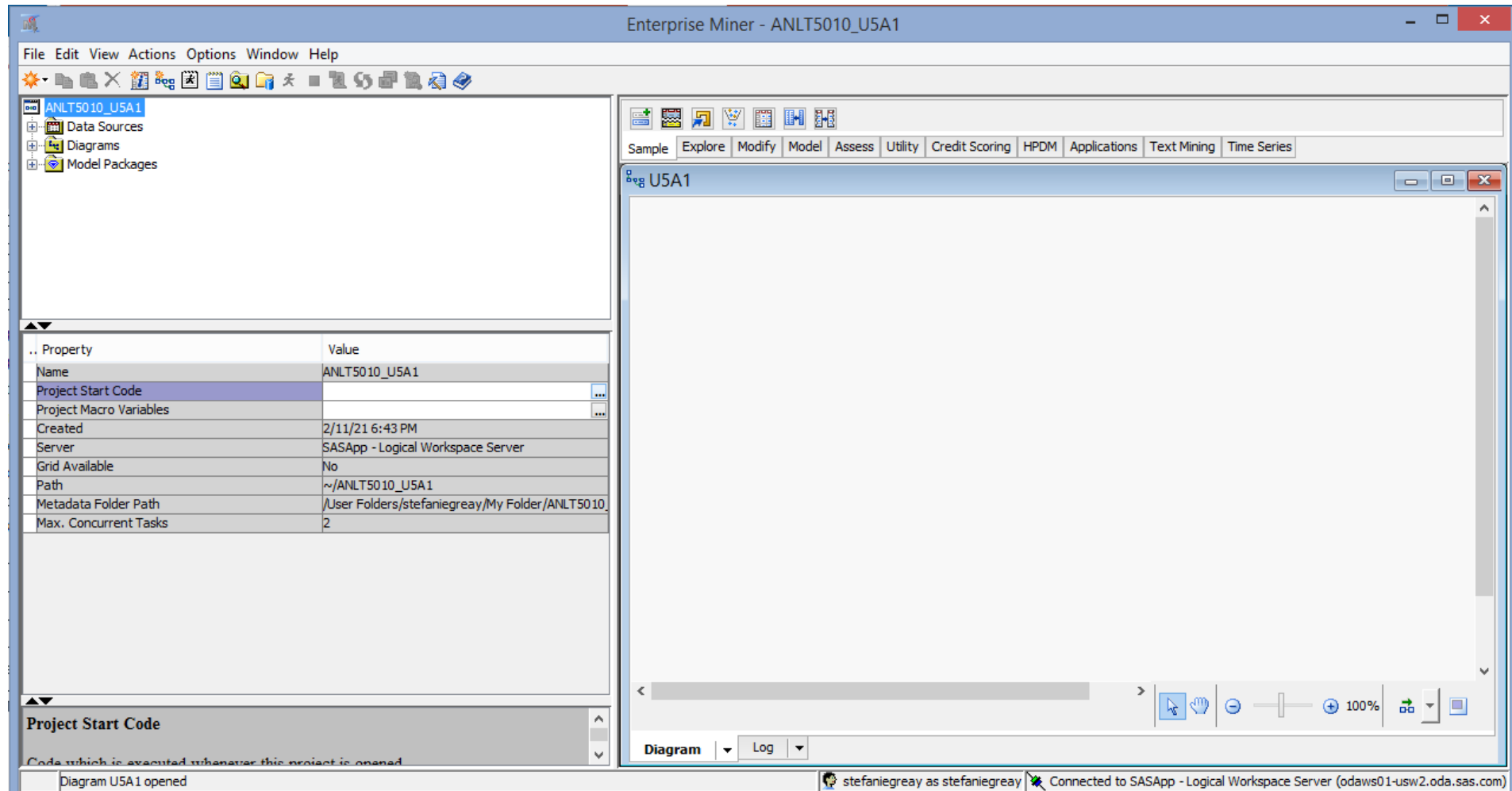
Click on the “Create Diagram” icon.



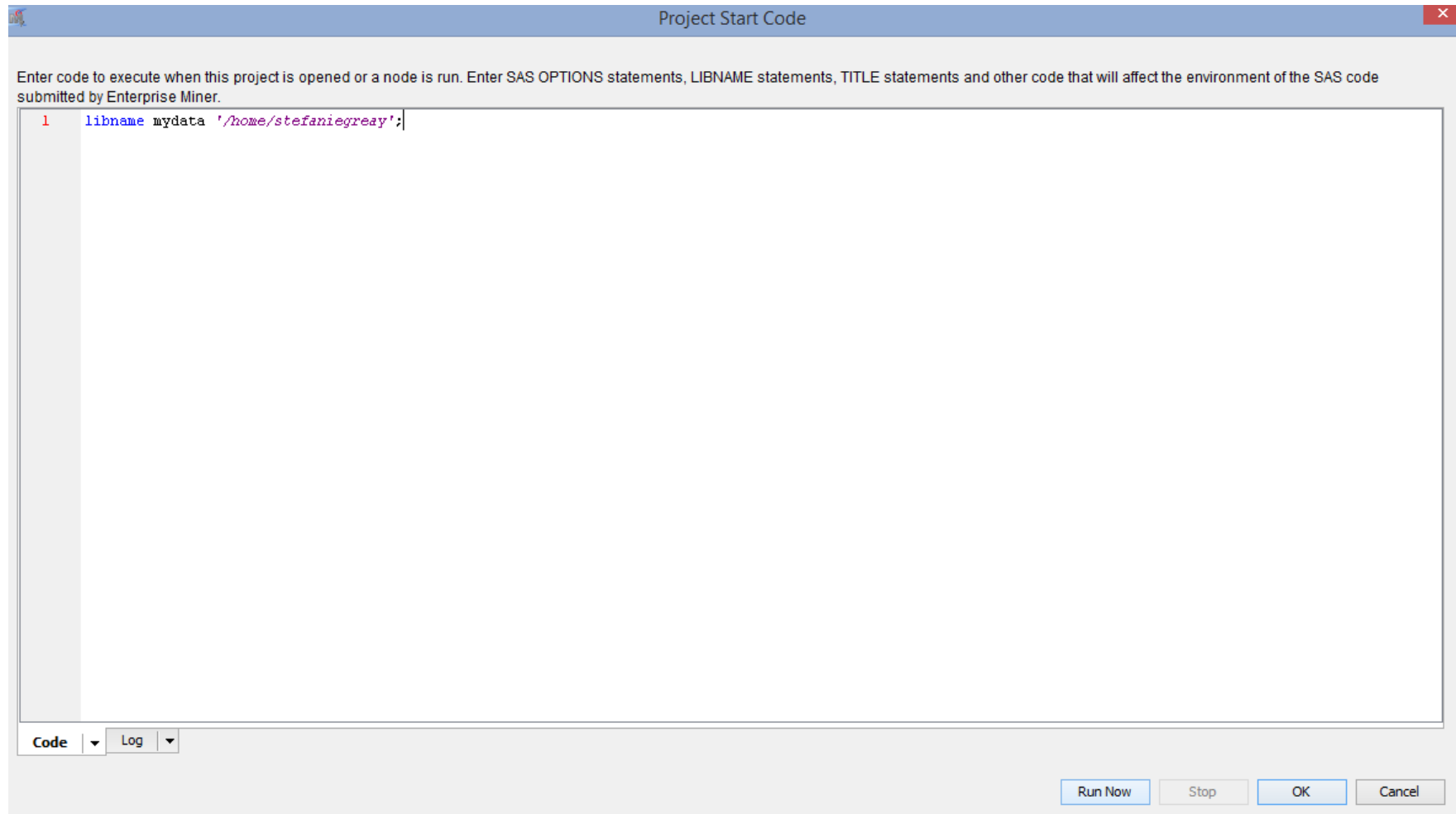
Name your diagram and click “OK.”



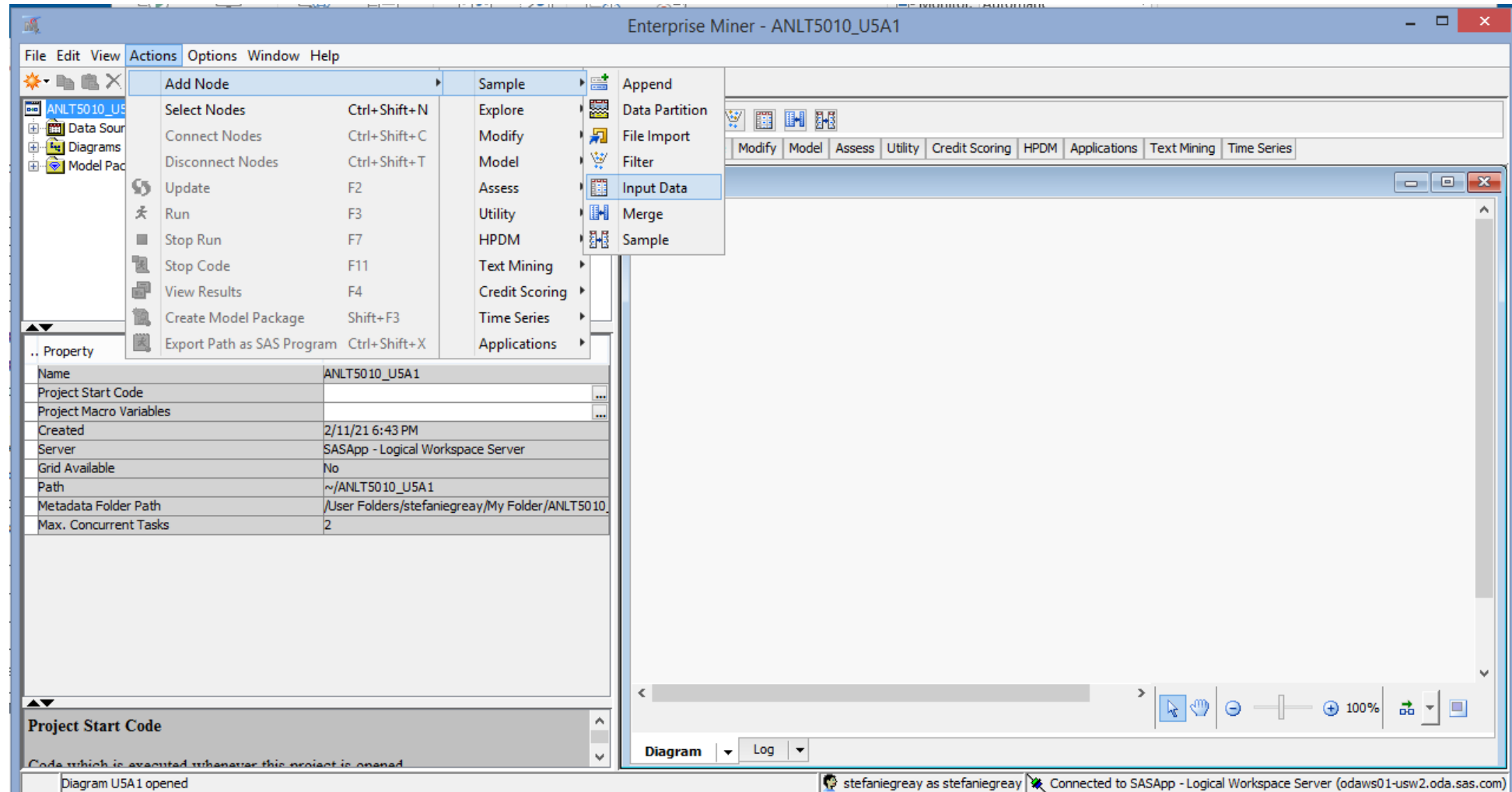
Click on the project, then click on the ellipses next to “Project Start Code.”



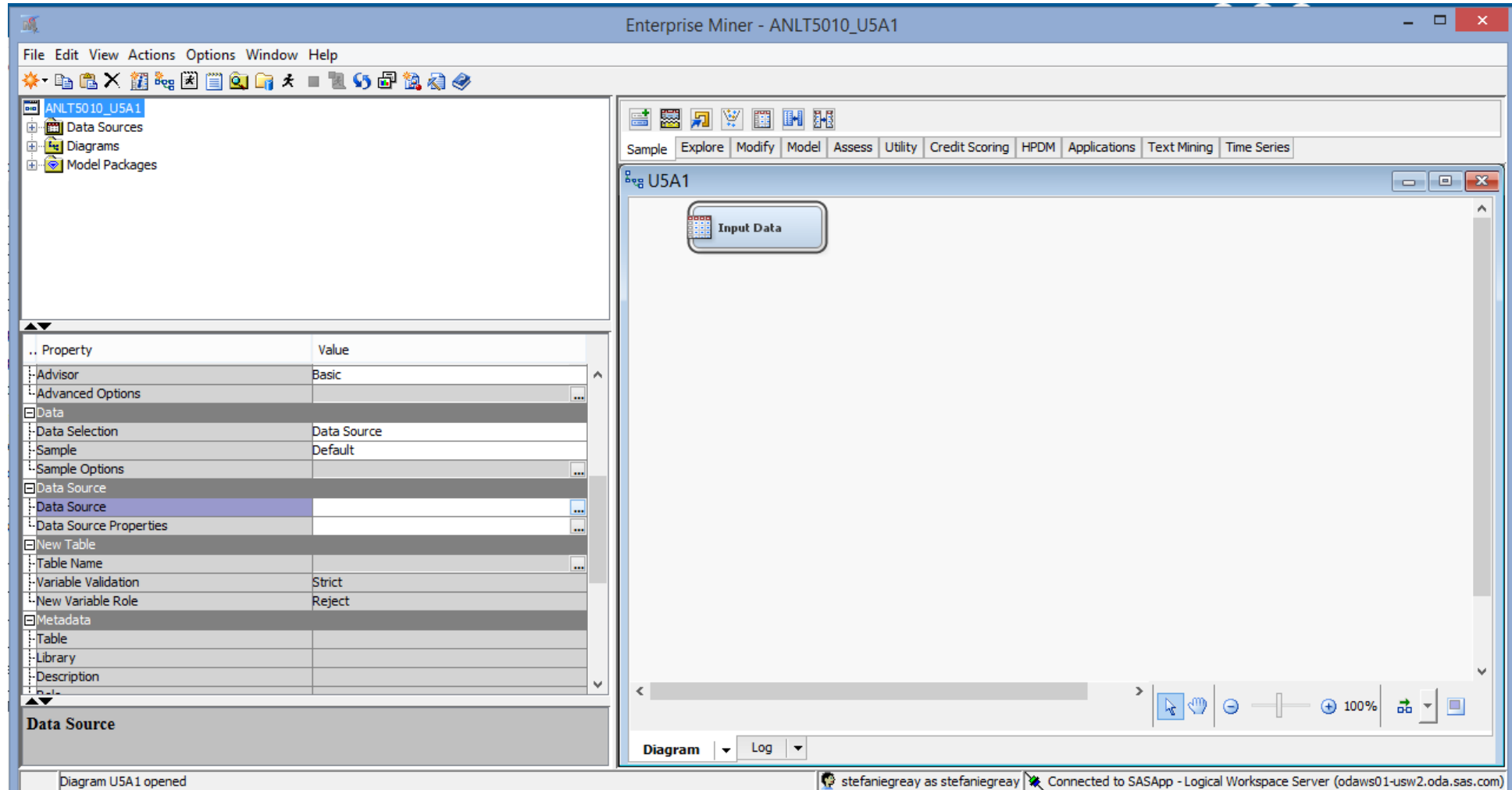
Add the library reference for where you uploaded the dataset in SAS studio, and click “Run Now.” Once it completes, click “OK.”



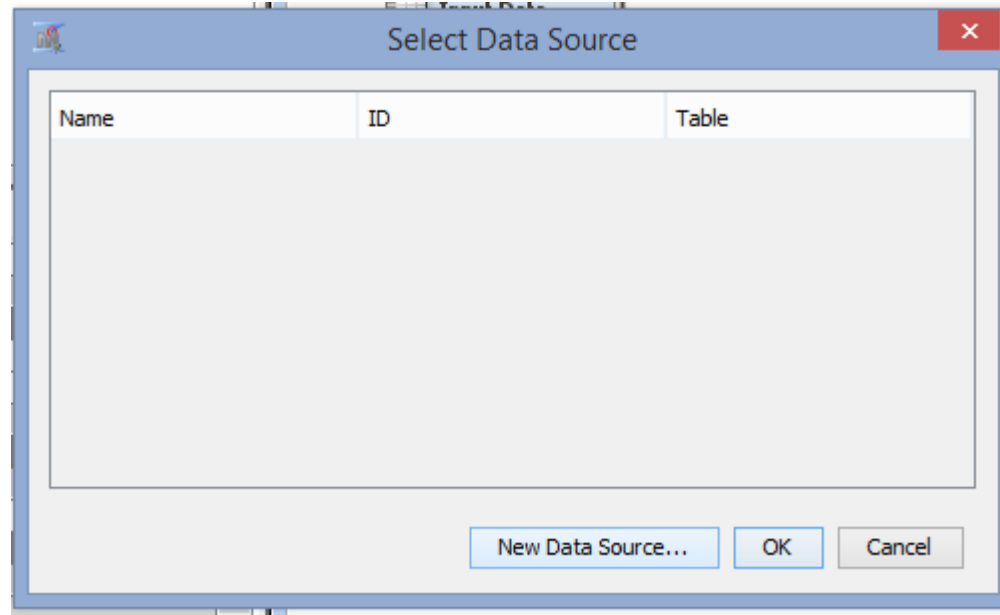
Click on Actions>Add Node>Sample>Input Data



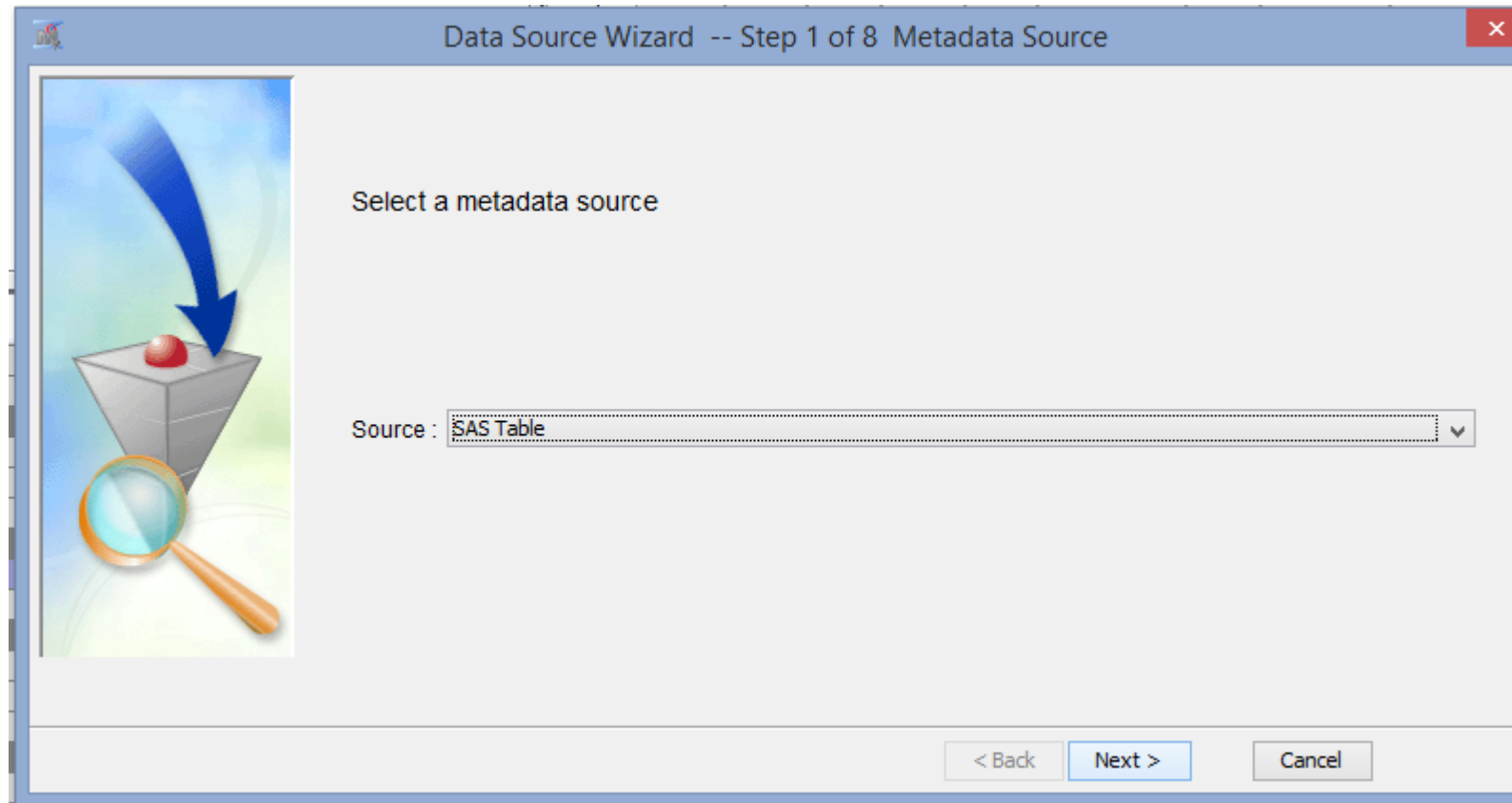
Click the ellipses (3 dots) next to “Data Source.”



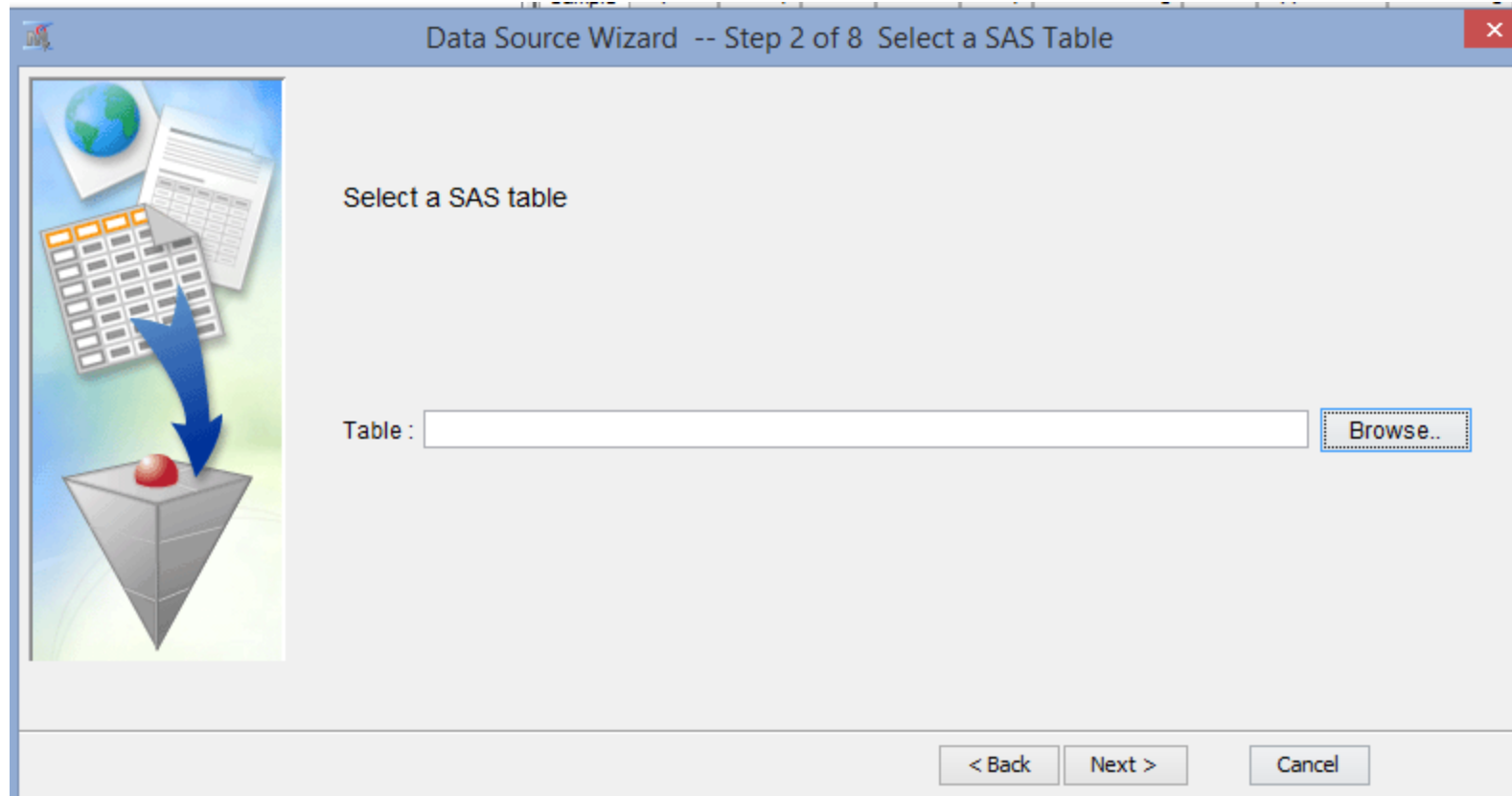
Click on “New Data Source”



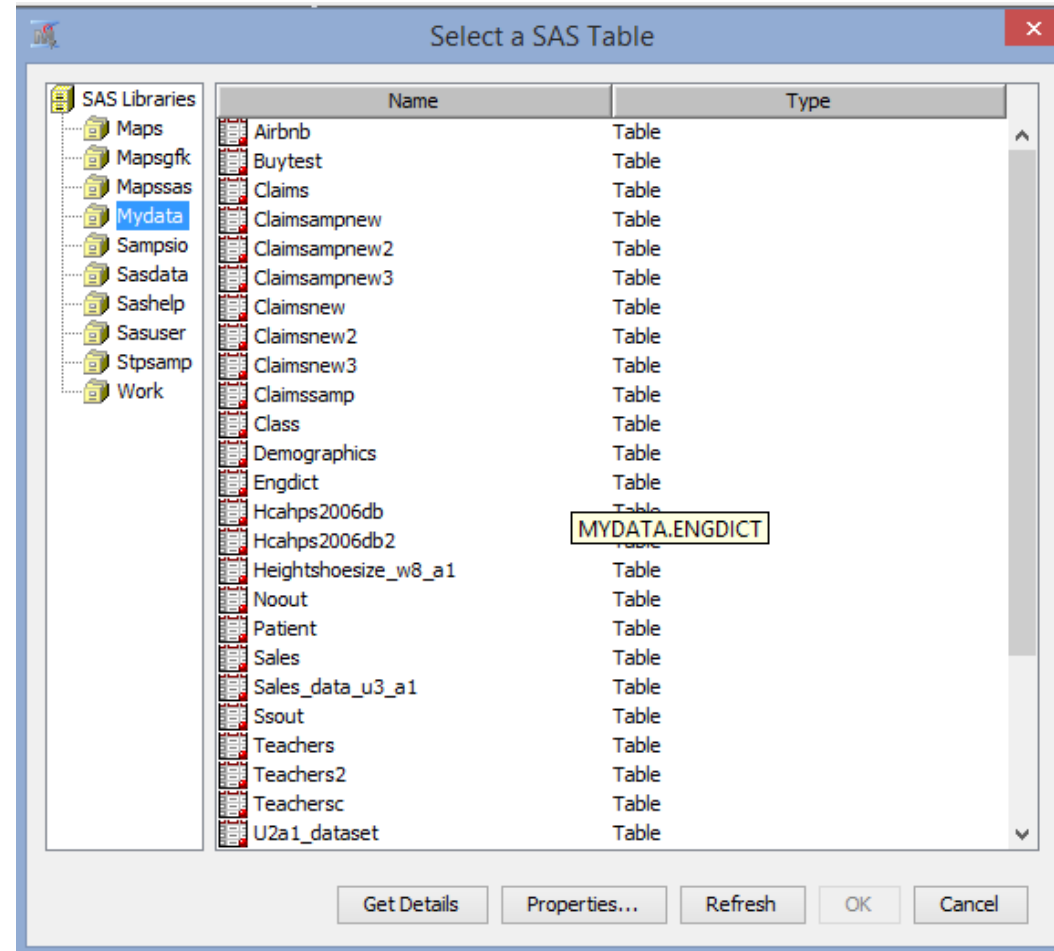
Leave it as “SAS Table” and click “Next >”



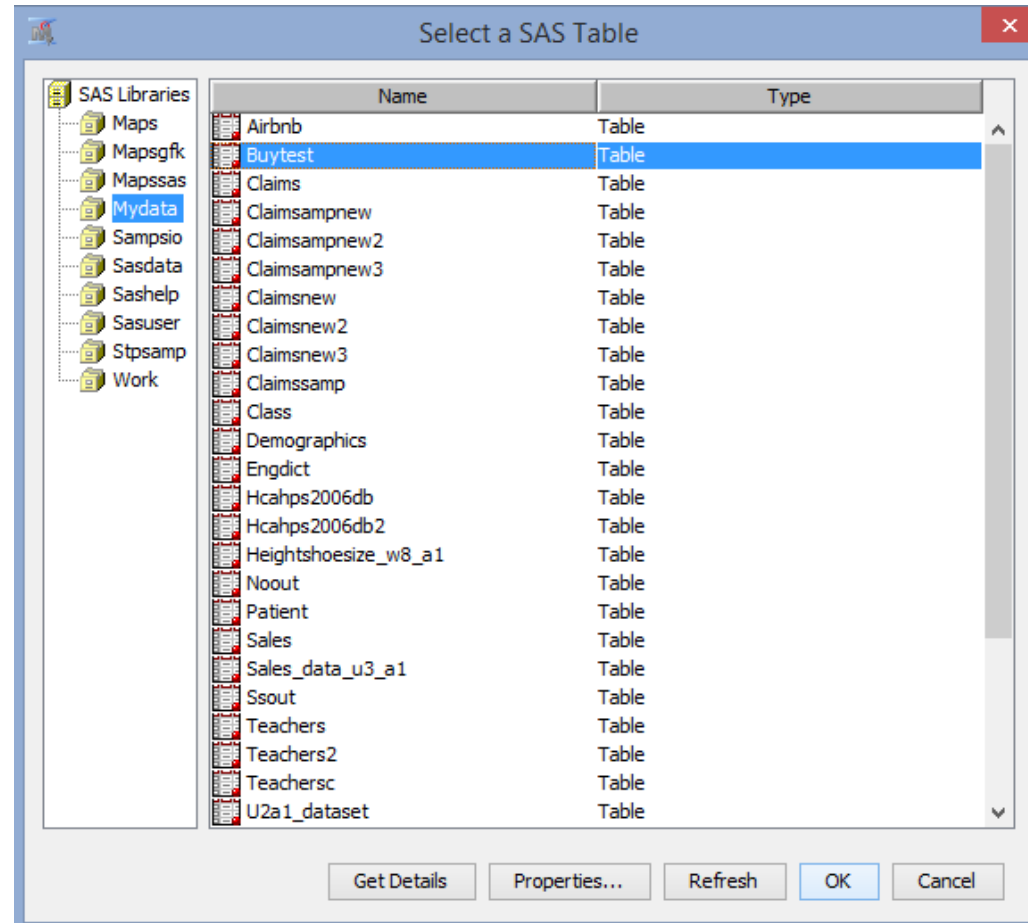
Click on “Browse”



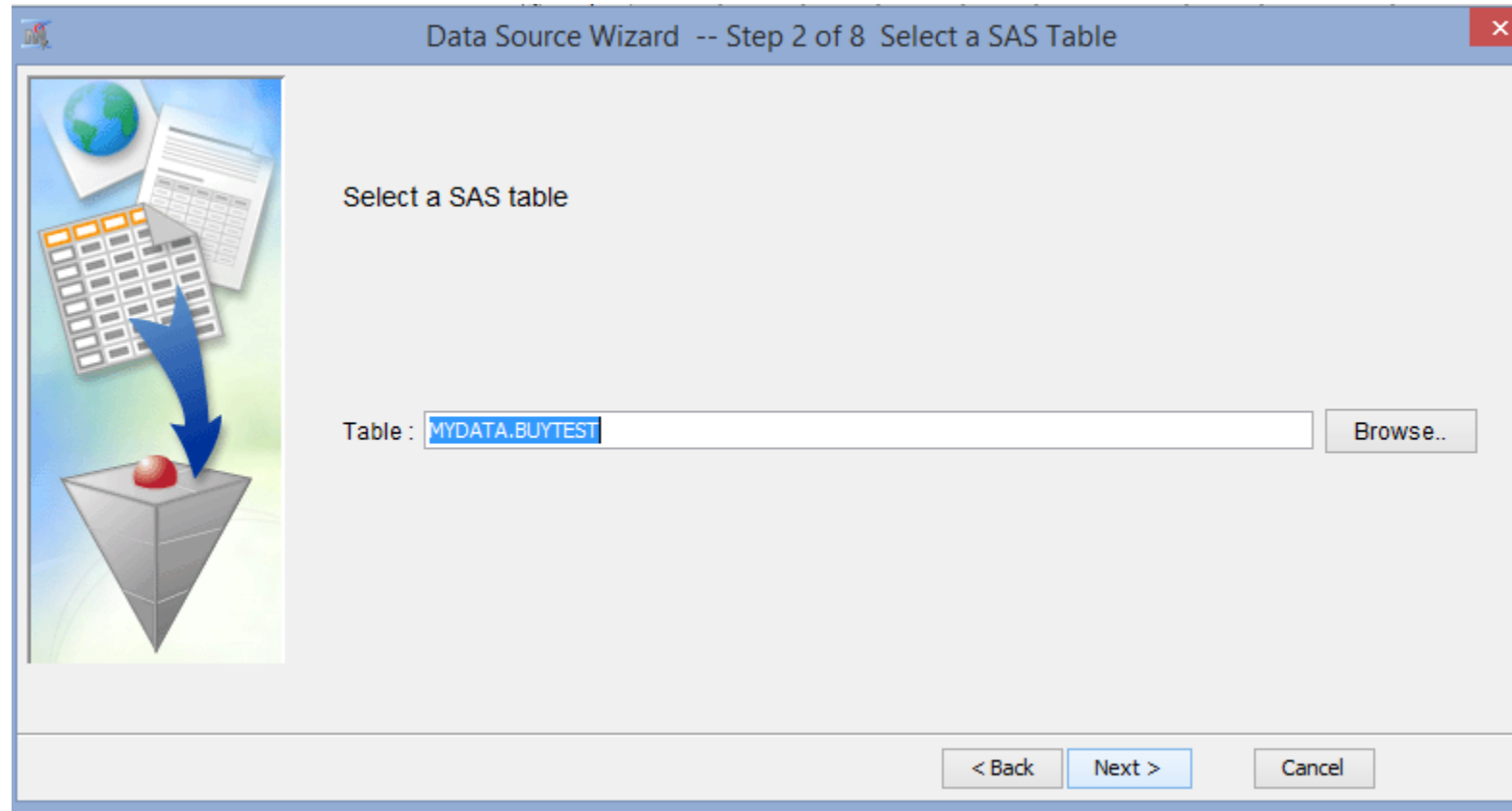
Double click on the libname you just set up in the project startup code.



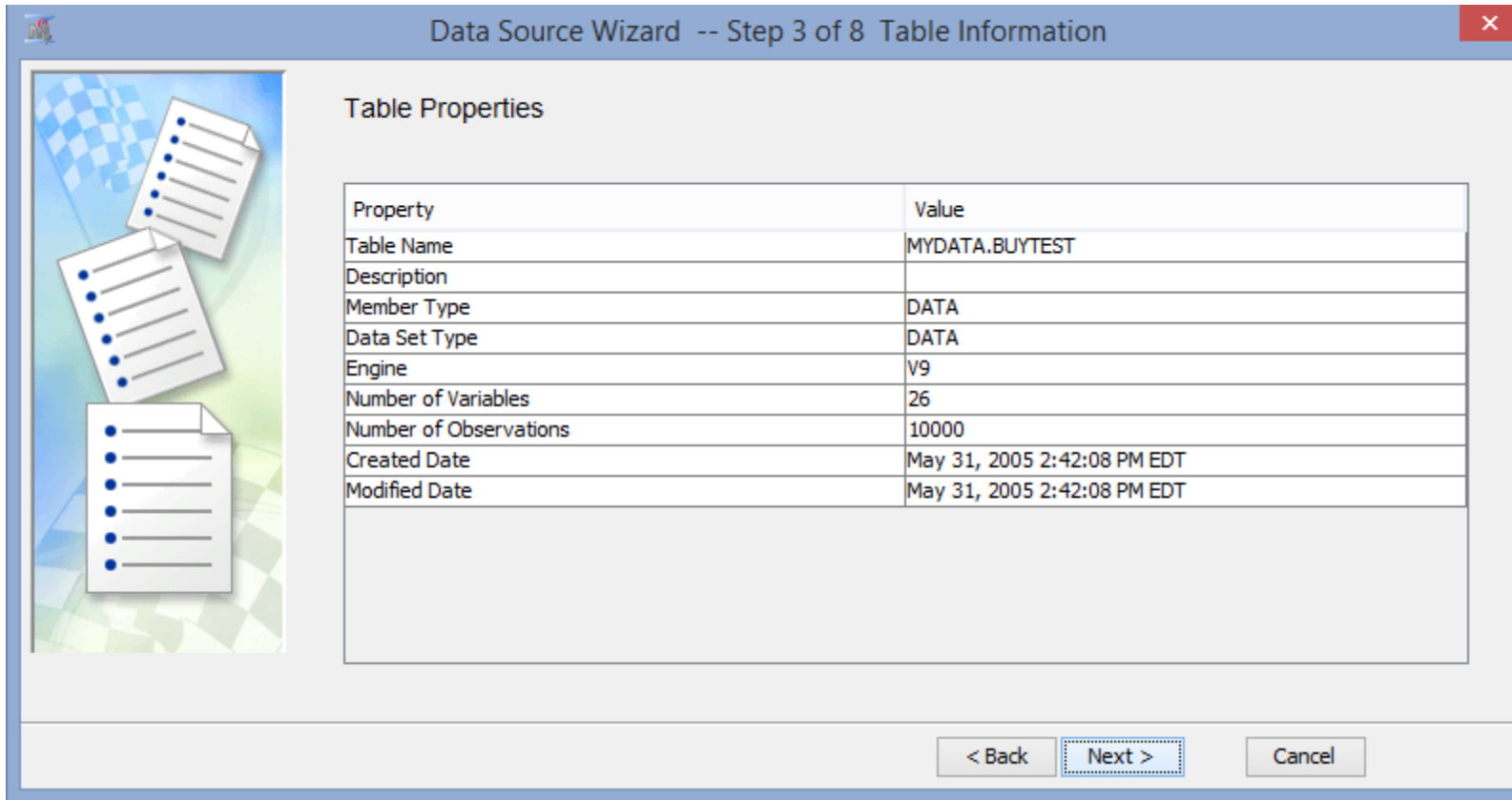
Double click to select the dataset for this unit, and click “OK”



Click “Next>”



Verify the options and click “Next>”

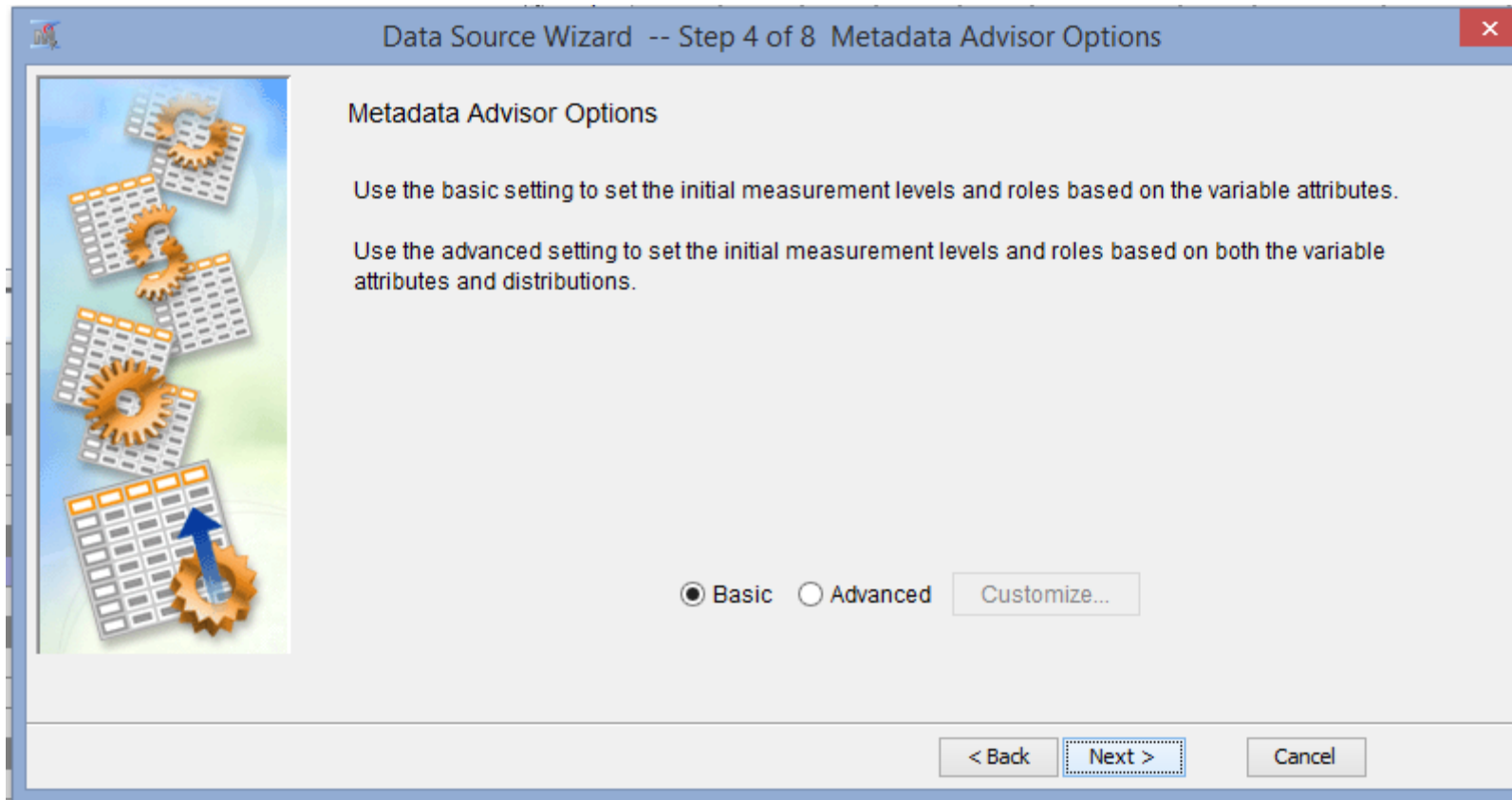


The screenshot shows a window titled "Data Source Wizard -- Step 3 of 8 Table Information". On the left is a decorative graphic of three overlapping document icons. The main area is titled "Table Properties" and contains a table with two columns: "Property" and "Value". The table lists various properties of the data source, including the table name, description, member type, data set type, engine, number of variables, number of observations, and creation/modification dates. At the bottom right, there are three buttons: "< Back", "Next >" (which is highlighted with a blue border), and "Cancel".

Property	Value
Table Name	MYDATA.BUYTEST
Description	
Member Type	DATA
Data Set Type	DATA
Engine	V9
Number of Variables	26
Number of Observations	10000
Created Date	May 31, 2005 2:42:08 PM EDT
Modified Date	May 31, 2005 2:42:08 PM EDT



Click “Next>”



Verify the variables and settings, adjust if necessary, and then click “Next>”

Data Source Wizard -- Step 5 of 8 Column Metadata

(none) ☐ not Equal to

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upp
AGE	Input	Interval	No		No	.	
BUY12	Input	Interval	No		No	.	
BUY18	Input	Interval	No		No	.	
BUY6	Input	Interval	No		No	.	
C1	Input	Interval	No		No	.	
C2	Input	Interval	No		No	.	
C3	Input	Interval	No		No	.	
C4	Input	Interval	No		No	.	
C5	Input	Interval	No		No	.	
C6	Input	Interval	No		No	.	
C7	Input	Interval	No		No	.	
CLIMATE	Input	Nominal	No		No	.	
COA6	Input	Interval	No		No	.	
DISCBUY	Input	Interval	No		No	.	
ETCO	Input	Interval	No		No	.	



You may choose to sample the dataset here, or just keep the full dataset, then click “Next>”

Data Source Wizard -- Step 6 of 8 Create Sample

Do you wish to create a sample data set?

☒ No ☐ Yes

Table Info

Columns 26

Rows 10000

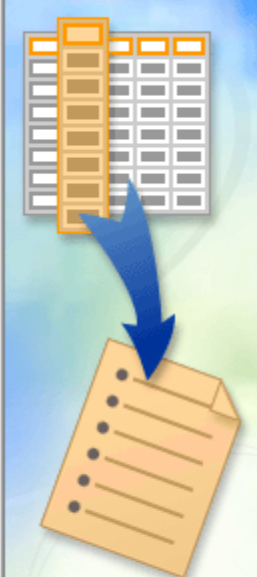
Sample Size

Type

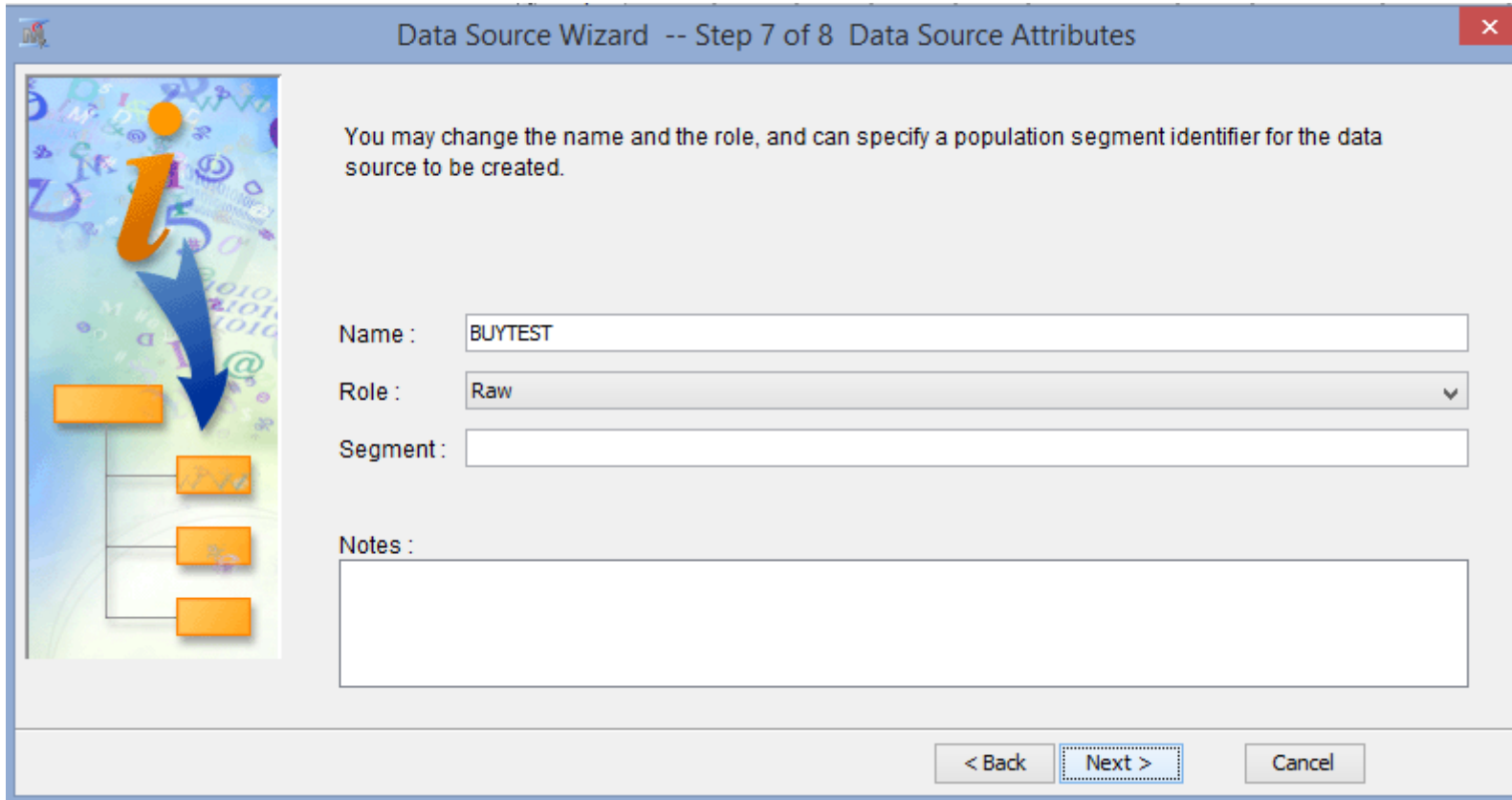
Percent

Rows

< Back Next > Cancel

An illustration on the left side of the wizard shows a grid representing a data table with several rows and columns. A blue arrow points from this grid down to a document icon, which is a yellow sheet of paper with several lines of text, representing a sampled dataset.

You may choose to adjust the role of the dataset, or leave it as the default, then click “Next>”



The image shows a screenshot of the 'Data Source Wizard -- Step 7 of 8: Data Source Attributes' window. The window has a blue title bar with a close button (X) in the top right corner. On the left side, there is a vertical sidebar with a blue background featuring a large orange 'i' icon and a blue arrow pointing downwards. Below the arrow are four orange rectangular buttons. The main area of the window is white and contains the following text and form elements:

You may change the name and the role, and can specify a population segment identifier for the data source to be created.

Name :

Role :

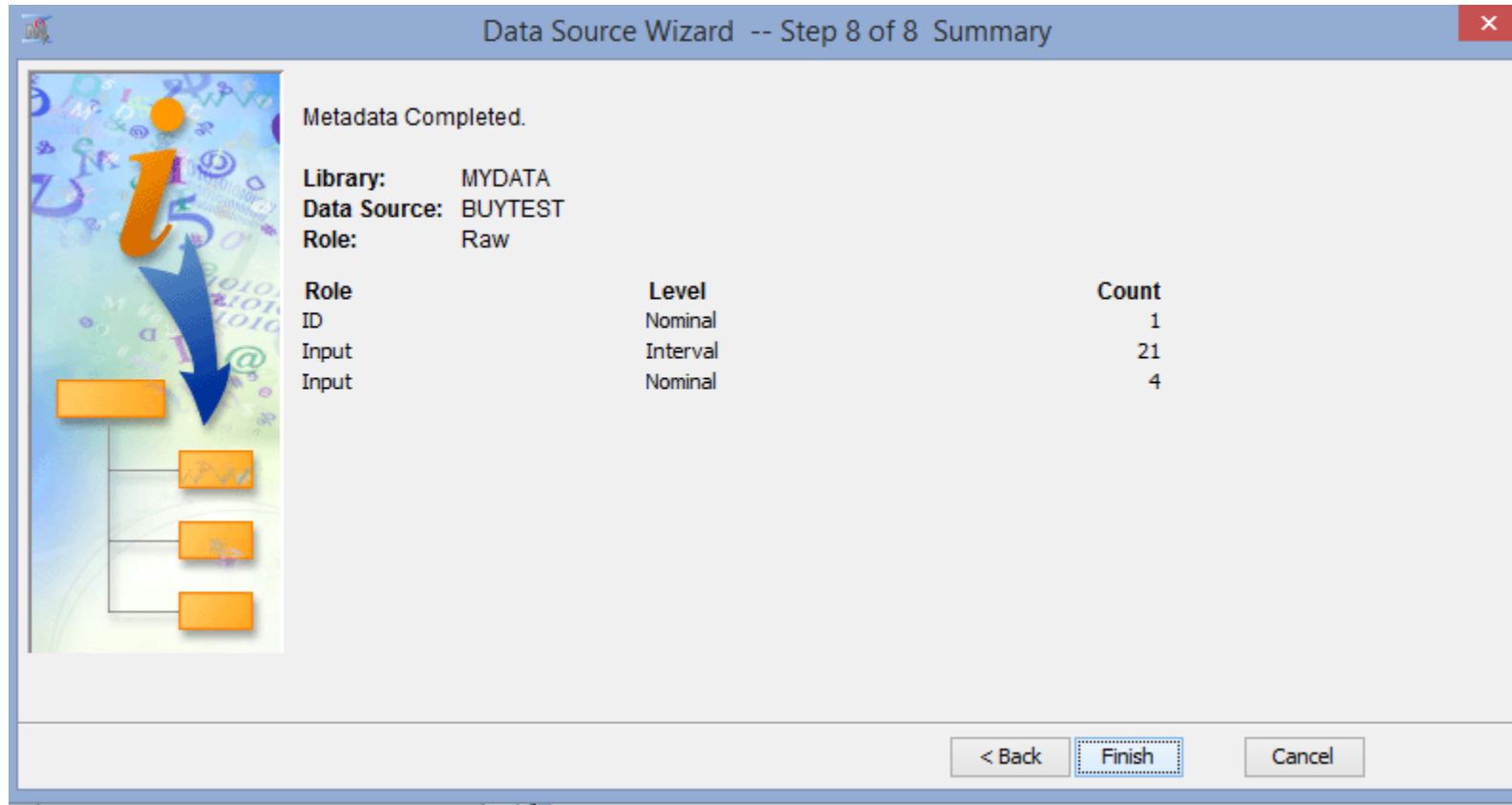
Segment :

Notes :

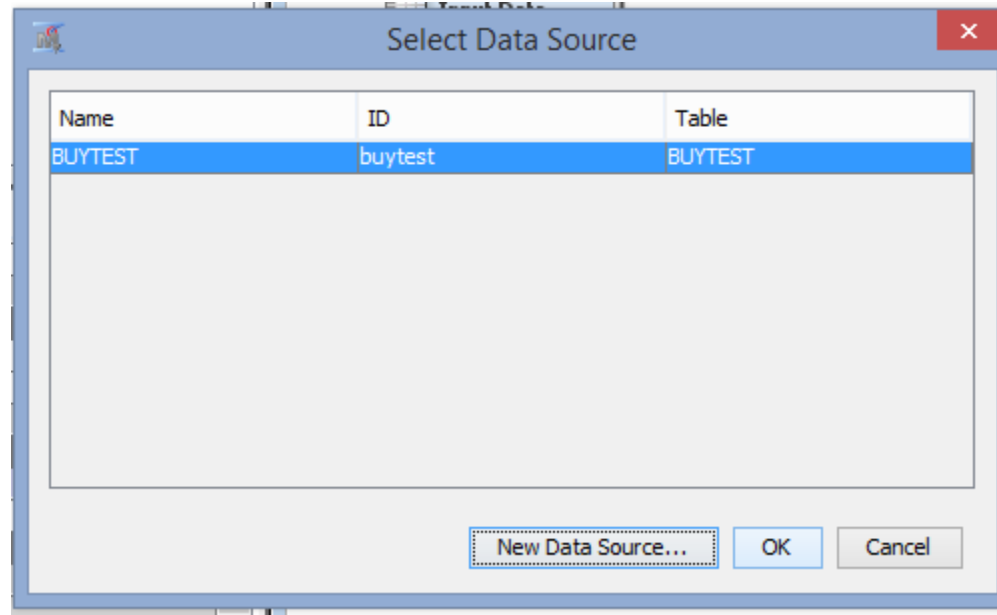
At the bottom of the window, there are three buttons: '< Back', 'Next >', and 'Cancel'. The 'Next >' button is highlighted with a dotted border.



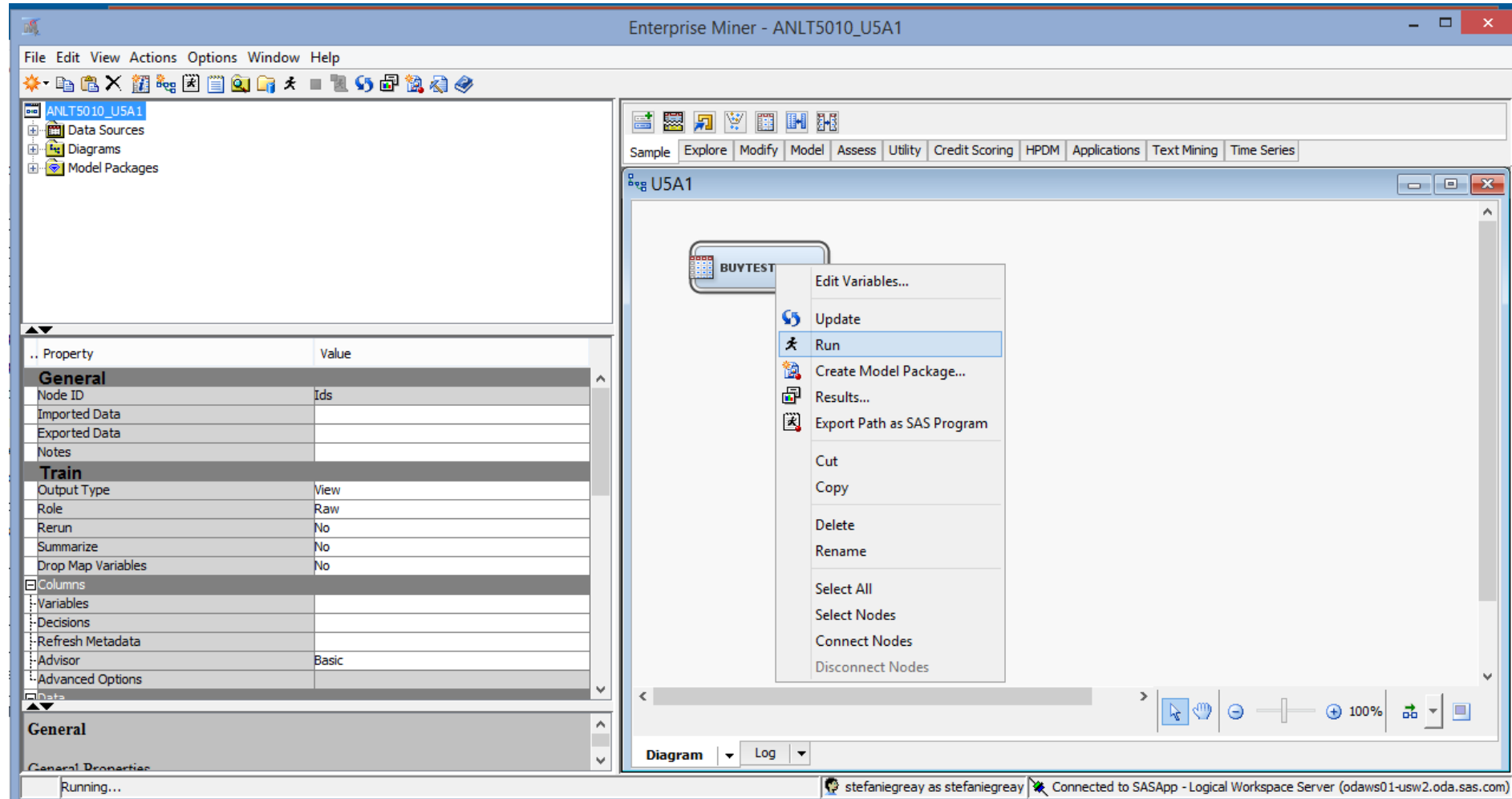
Click “Finish” to finish the data source registration within EG.



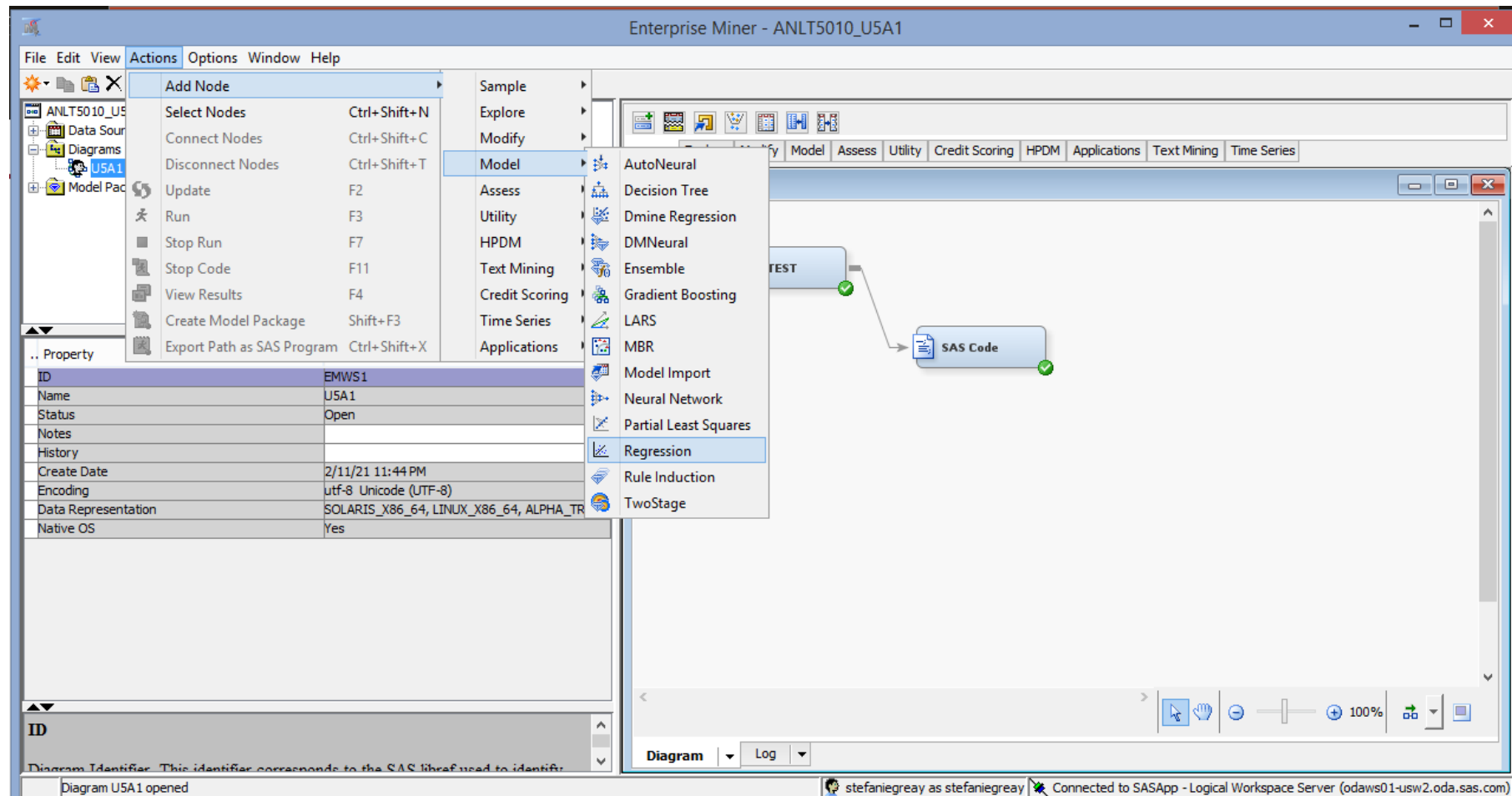
Click “OK” to complete the process. The name of the node should then change to the name of the dataset.



Right click on the dataset node and click “Run.”



Click on “Actions” > “Add Node” > “Model” > “Regression”



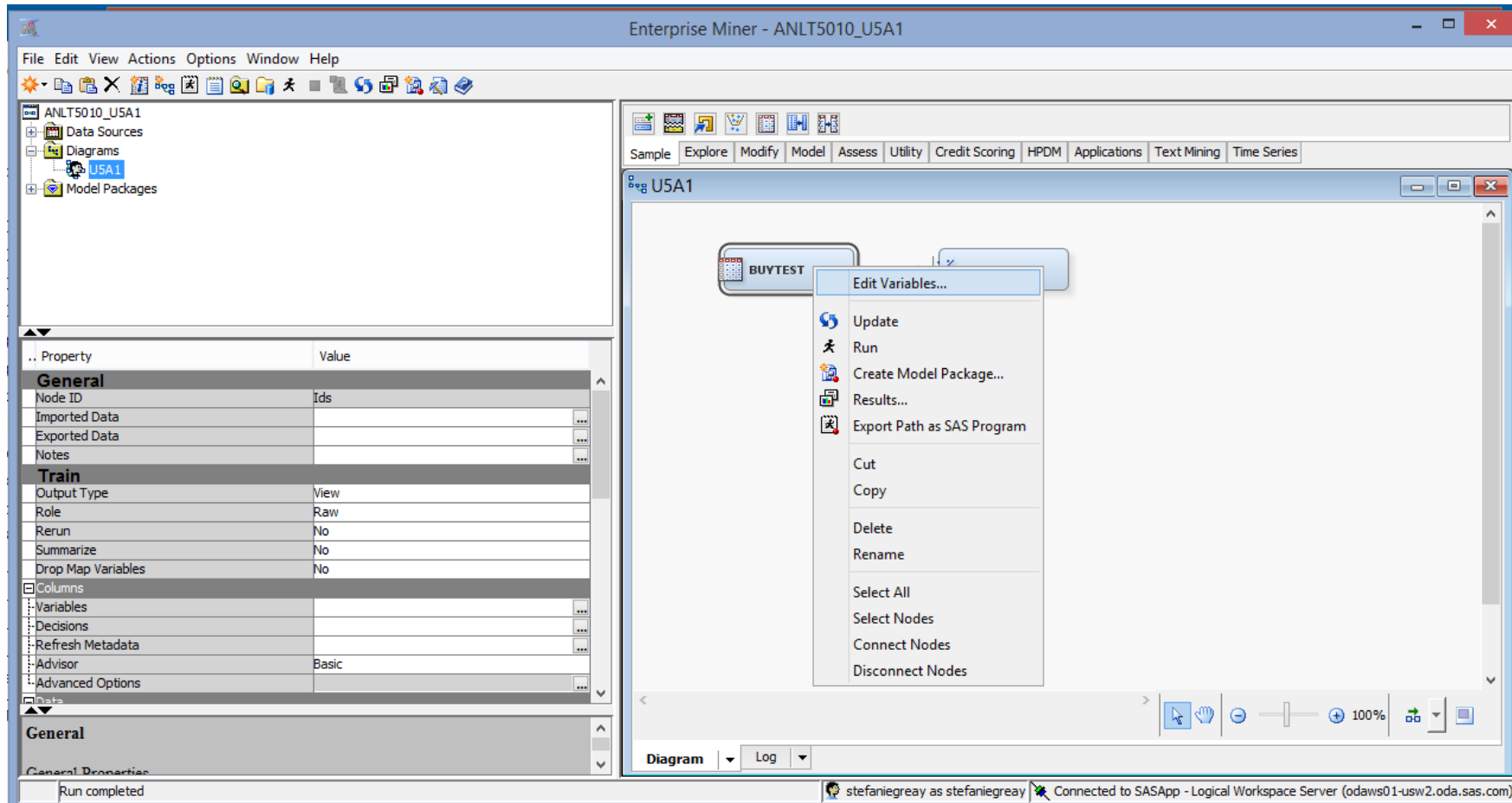
Connect the nodes

The screenshot displays the SAS Enterprise Miner interface. On the left, a project tree shows the hierarchy: ANLT5010_USA1 > Data Sources > Diagrams > U5A1. Below this, a properties window for the 'U5A1' diagram is visible, showing various settings under 'General', 'Train', 'Equation', 'Class Targets', and 'Model Options'. The main workspace on the right shows a diagram titled 'U5A1' with two nodes: 'BUYTEST' (a data source icon) and 'Regression' (a model icon). An arrow connects 'BUYTEST' to 'Regression', indicating the data flow. The status bar at the bottom indicates the user is 'stefaniegreay' and is connected to the SASApp - Logical Workspace Server.

Property	Value
General	
Node ID	Reg
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	...
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
General	
General Properties	
Run completed	



Right click on the dataset node and choose “edit variables.”



Change the “buy” variable we just made to “Target” and the “Buy6” variable to Drop=Yes

Variables - Ids

(none) ☐ not Equal to ☐

Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

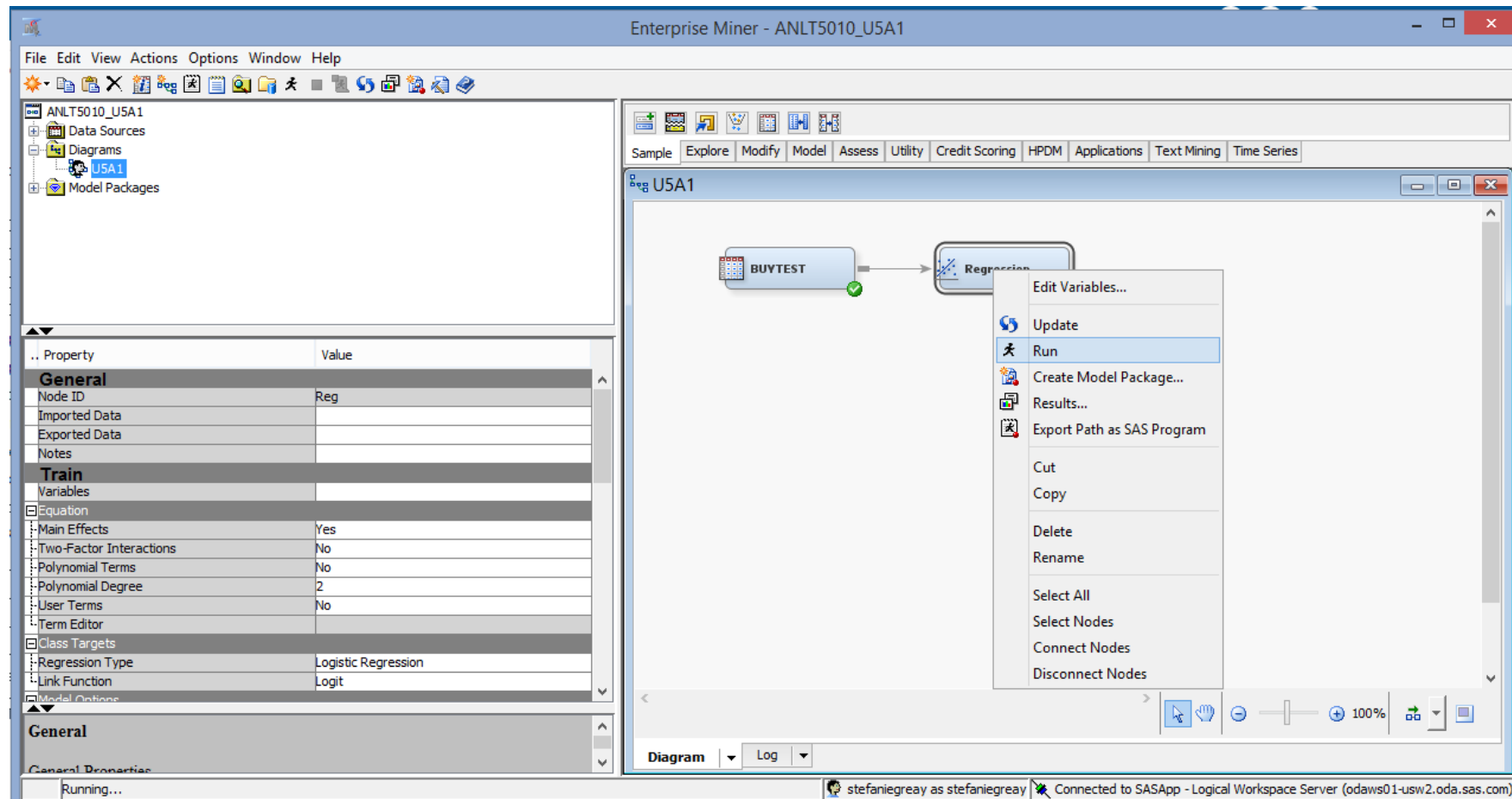
Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
AGE	Input	Interval	No		No	.	.
BUY12	Input	Interval	No		No	.	.
BUY18	Input	Interval	No		No	.	.
BUY6	Input	Interval	No		Yes	.	.
C1	Input	Interval	No		No	.	.
C2	Input	Interval	No		No	.	.
C3	Input	Interval	No		No	.	.
C4	Input	Interval	No		No	.	.
C5	Input	Interval	No		No	.	.
C6	Input	Interval	No		No	.	.
C7	Input	Interval	No		No	.	.
CLIMATE	Input	Nominal	No		No	.	.
COA6	Input	Interval	No		No	.	.
DISCBUY	Input	Interval	No		No	.	.
FICO	Input	Interval	No		No	.	.
ID	ID	Nominal	No		No	.	.
INCOME	Input	Interval	No		No	.	.
LOC	Input	Nominal	No		No	.	.
MARRIED	Input	Interval	No		No	.	.
ORGSRC	Input	Nominal	No		No	.	.
OWNHOME	Input	Interval	No		No	.	.
PURCHTOT	Input	Interval	No		No	.	.
RESPOND	Input	Interval	No		No	.	.
RETURN24	Input	Interval	No		No	.	.
SEX	Input	Nominal	No		No	.	.
VALUE24	Input	Interval	No		No	.	.
buy	Target	Interval	No		No	.	.

Apply Reset

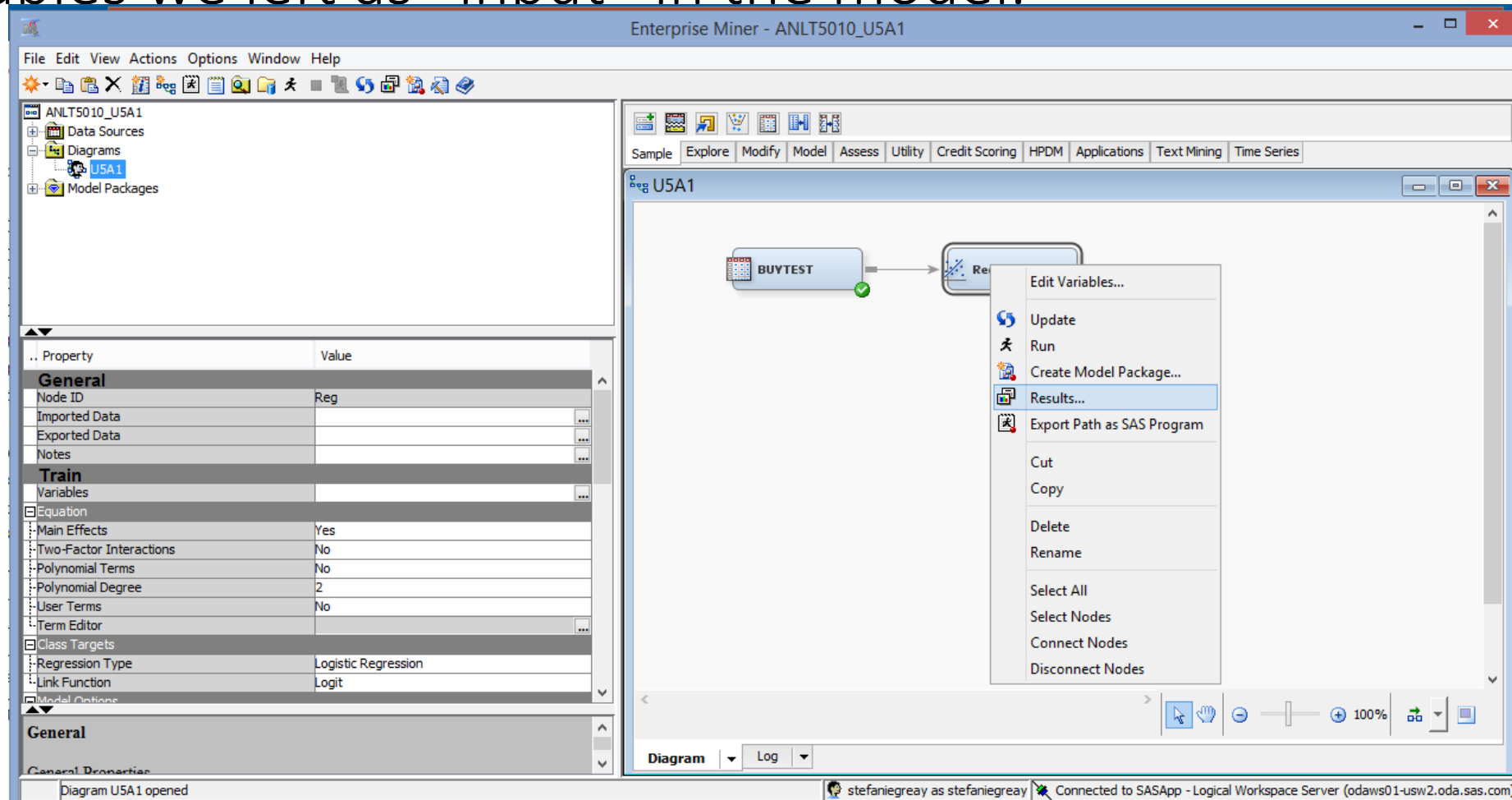
Explore... OK Cancel



Right click on “Regression” model and click “Run” (note that the default regression type is set to logistic regression, so no change is needed there.)



Right click on “Regression” model and click “Results” to view the results. Note that this is using the defaults and includes all variables we left as “input” in the model.



If you want to use one of the variable selection methods (forward, backward, stepwise, etc.), you can click on the “Regression” node and scroll down to “Selection Model” and then rerun the regression node.

The screenshot displays the SAS Enterprise Miner interface. The main window is titled "Enterprise Miner - ANLT5010_USA1". On the left, a project tree shows "Data Sources", "Diagrams", "USA1", and "Model Packages". The "USA1" diagram is selected, showing a workflow with two nodes: "BUYTEST" and "Regression". The "Regression" node is highlighted, and its properties are displayed in the lower-left pane. The "Selection Model" property is set to "None", and the "Selection Options" are set to "Stepwise". The "Regression Type" is "Logistic Regression", and the "Link Function" is "Logit". The "Suppression Intercept" is "No", and the "Input Coding" is "Deviation". The "Model Selection" section is expanded, showing "Selection Model" as "None", "Selection Criterion" as "Backward", "Use Selection Defaults" as "Forward", "Selection Options" as "Stepwise", "Optimization Options" as "None", "Technique" as "Default", "Default Optimization" as "Yes", "Max Iterations" as "0", "Max Function Calls" as "0", and "Maximum Time" as "1 Hour". The "Convergence Criteria" section is also visible. The bottom status bar shows "Diagram USA1 opened" and "Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)".

Property	Value
Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
Model Options	
Suppression Intercept	No
Input Coding	Deviation
Model Selection	
Selection Model	None
Selection Criterion	Backward
Use Selection Defaults	Forward
Selection Options	Stepwise
Optimization Options	None
Technique	Default
Default Optimization	Yes
Max Iterations	0
Max Function Calls	0
Maximum Time	1 Hour



Prediction of buying when >50 yrs old

One portion of the assignment asks that you predict the probability of an individual who is over the age of 50 buying the item, but this question is irrelevant considering that age is not a significant variable in the logistic regression model. As such, you may ignore this question/portion of the assignment for my class.



Assumptions for logistic regression

Be sure that you discuss and check (when possible) all of the assumptions of logistic regression, which include:

- 1) the response/dependent/target variable is binary (for binary logistic regression) or ordinal (for ordinal logistic regression)
- 2) no multicollinearity exists between the independent variables
- 3) the observations are independent of each other
- 4) there is a linear relationship between the independent variables and the log odds
- 5) there is a significantly large sample



SAS Documentation Reference

The link below brings you to the SAS Documentation on the Regression Node, which has an example, including interpretation of the output.

<https://documentation.sas.com/?docsetId=emref&docsetTarget=n1jqz z8cssr9m2n1ktx2iyv87q56.htm&docsetVersion=14.3&locale=en>

