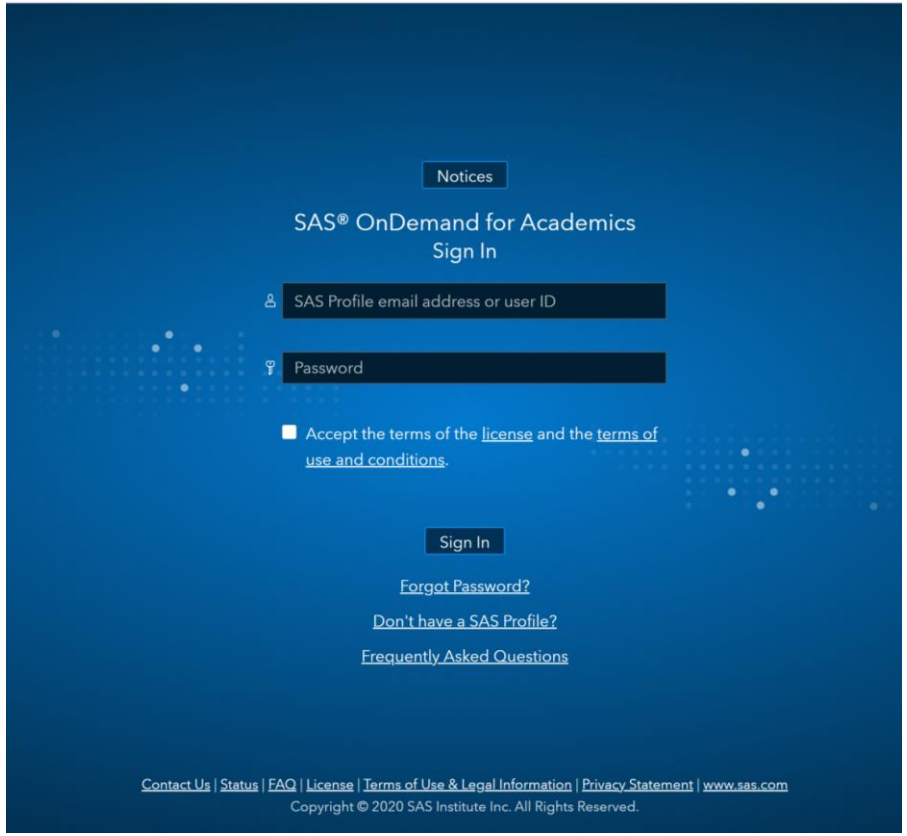
ANLT5050

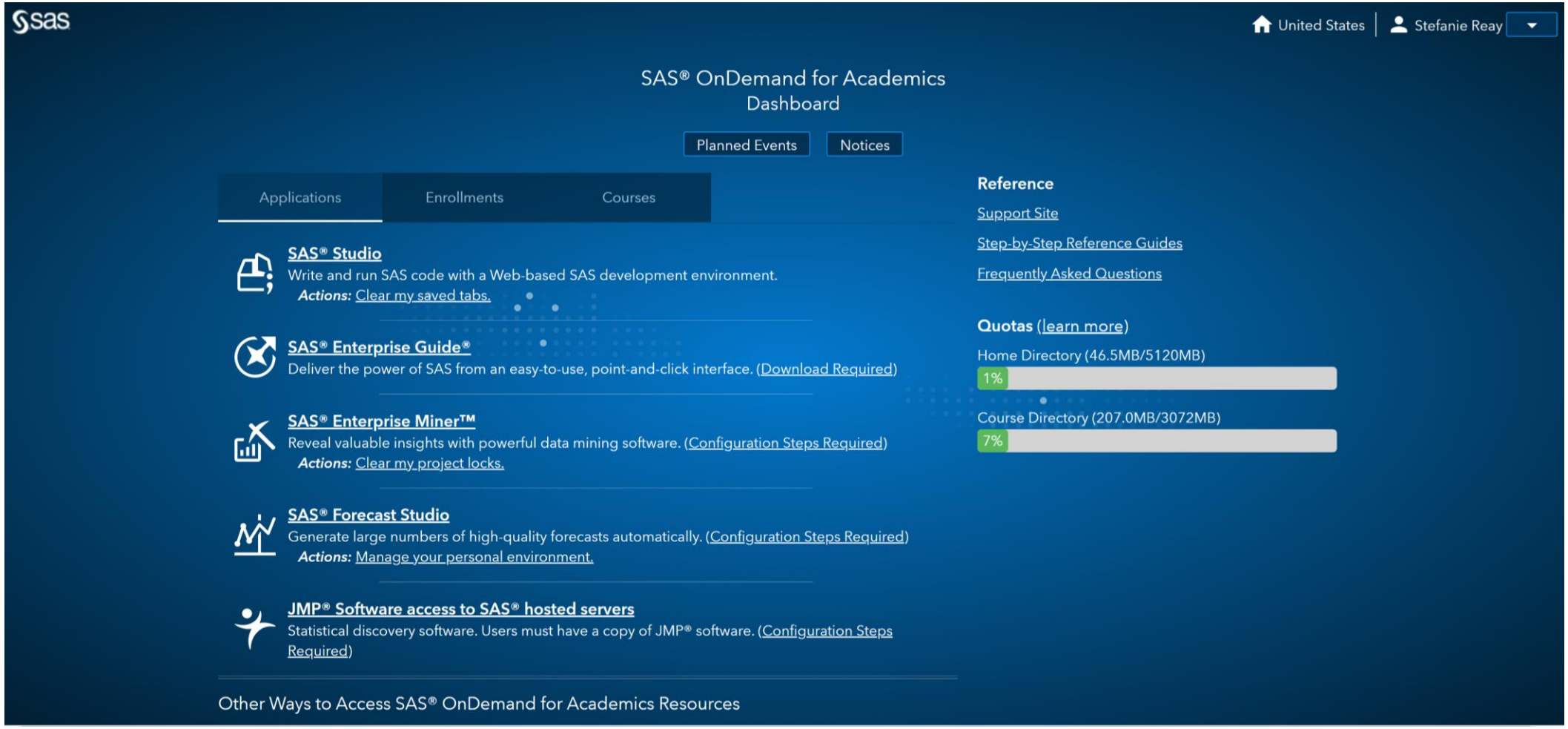
Unit 5 Assignment 1 Tutorial

Access the SAS OnDemand for Academics Control Center

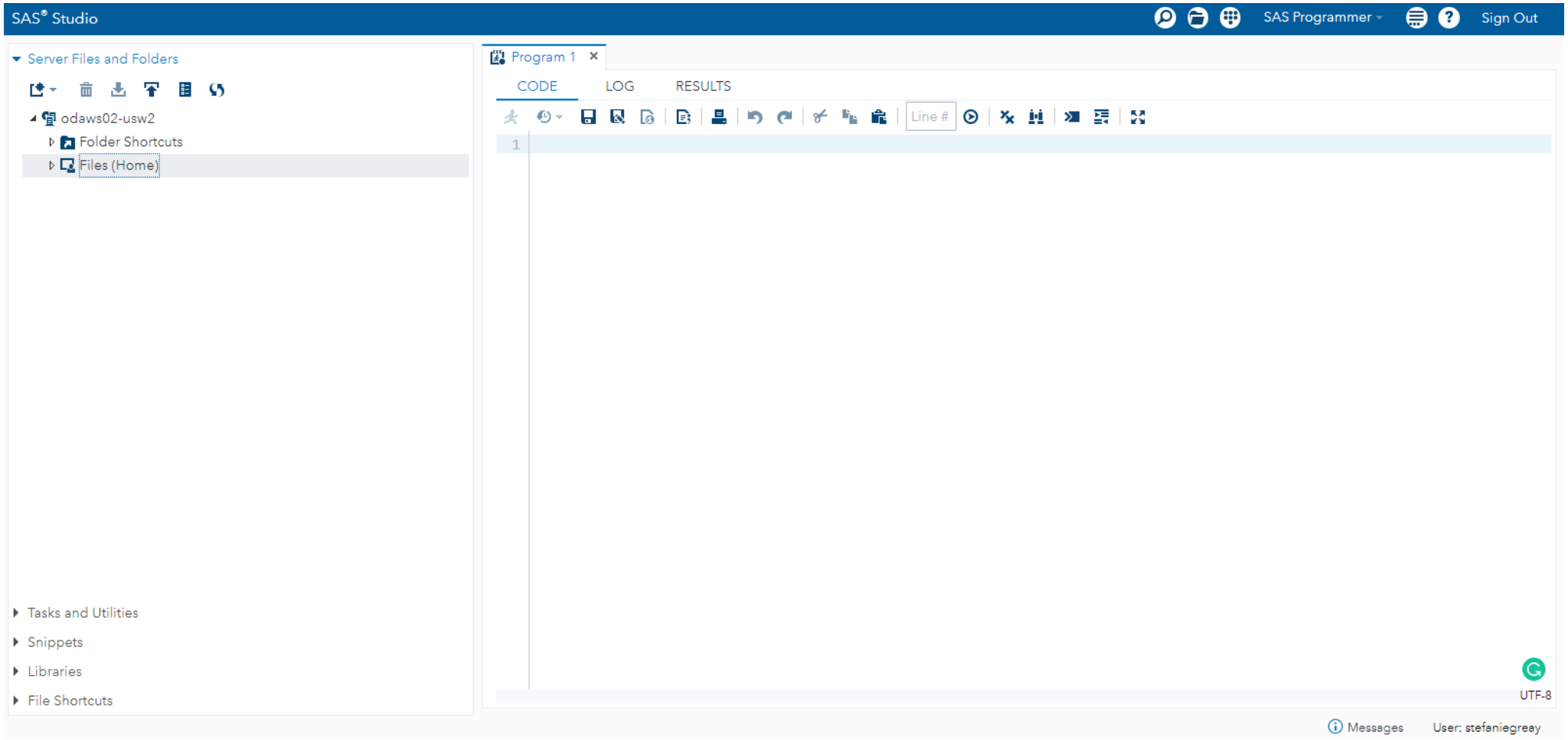
https://odamid.oda.sas.com/SASODAControlCenter



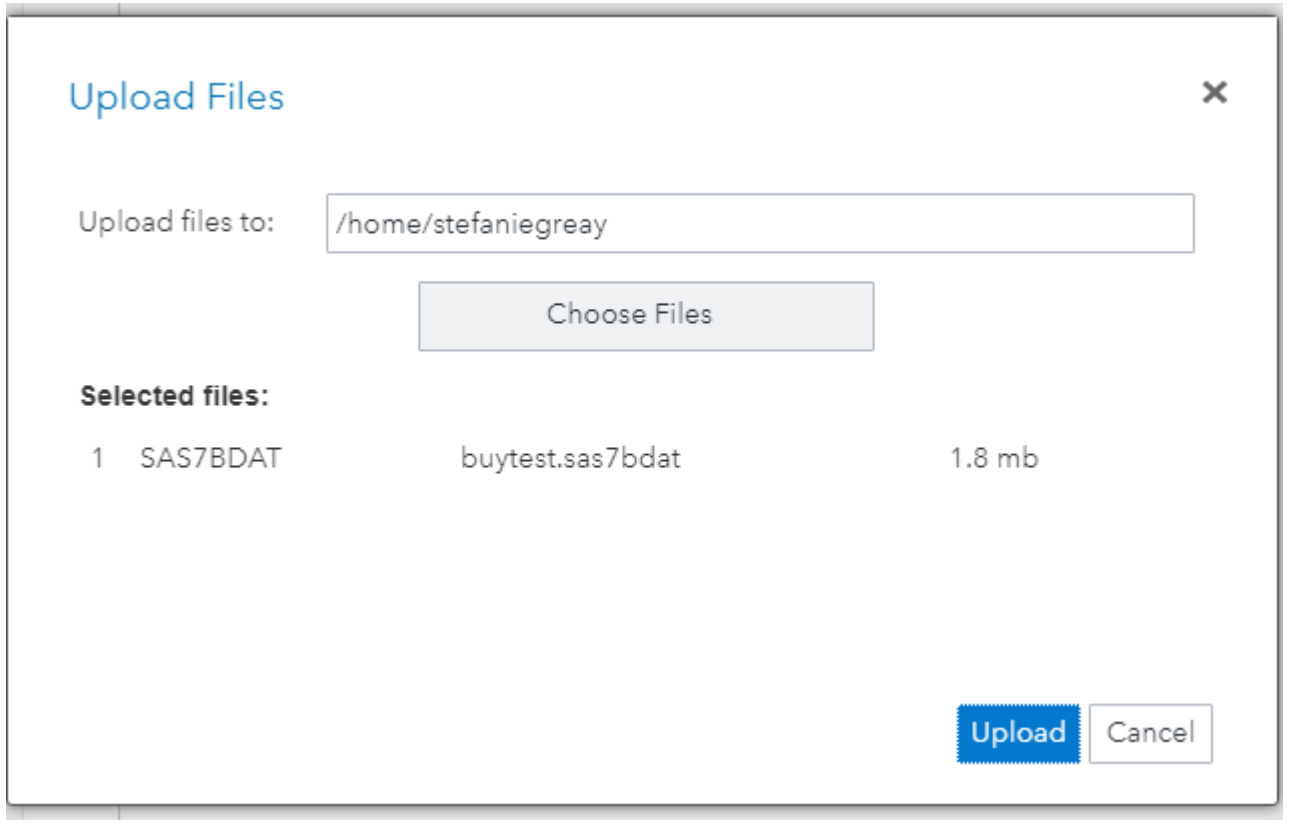
SAS OnDemand for Academics (SODA) Control Center



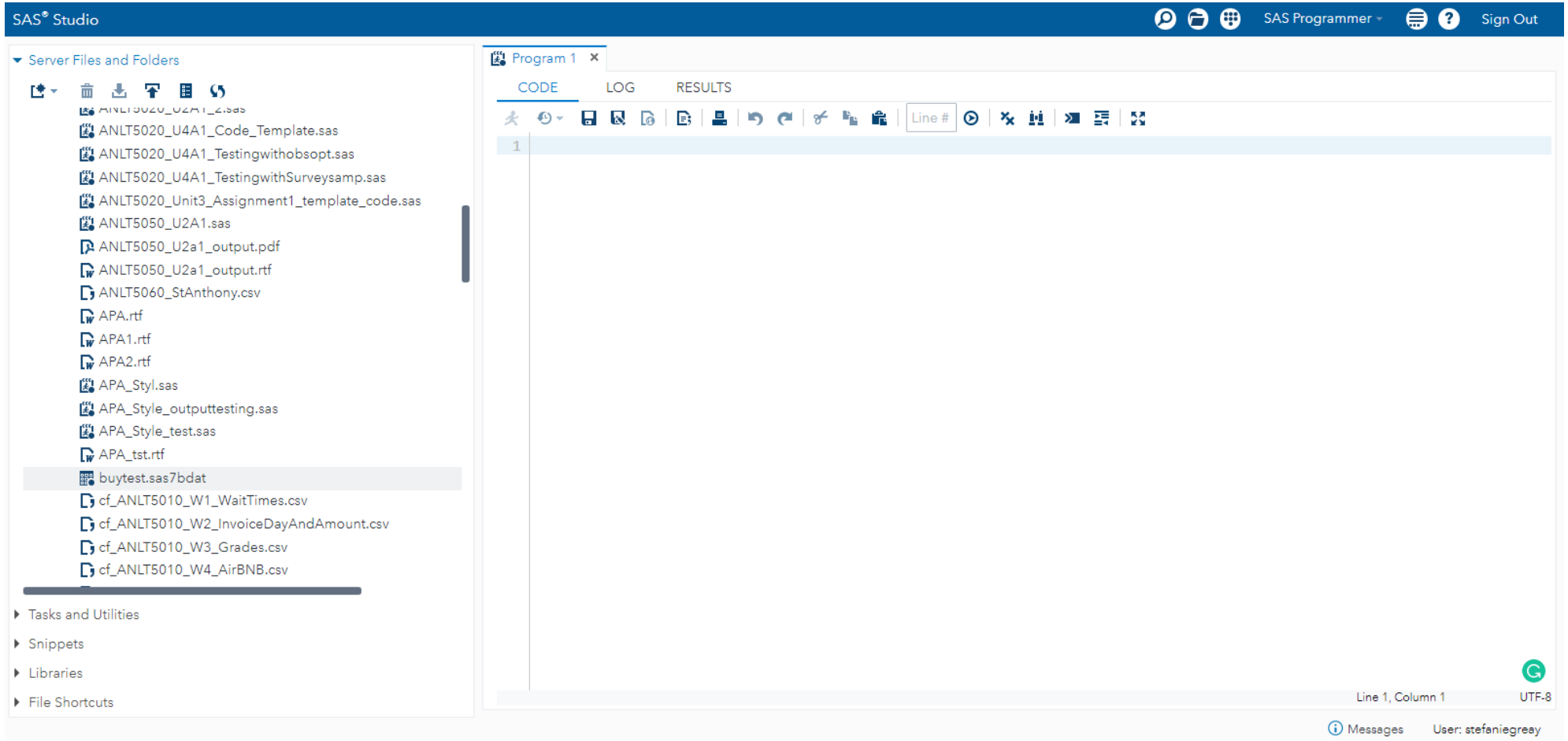
To upload the dataset to the SAS server, open SAS Studio, then click on “Files (Home)” and click the upload button.



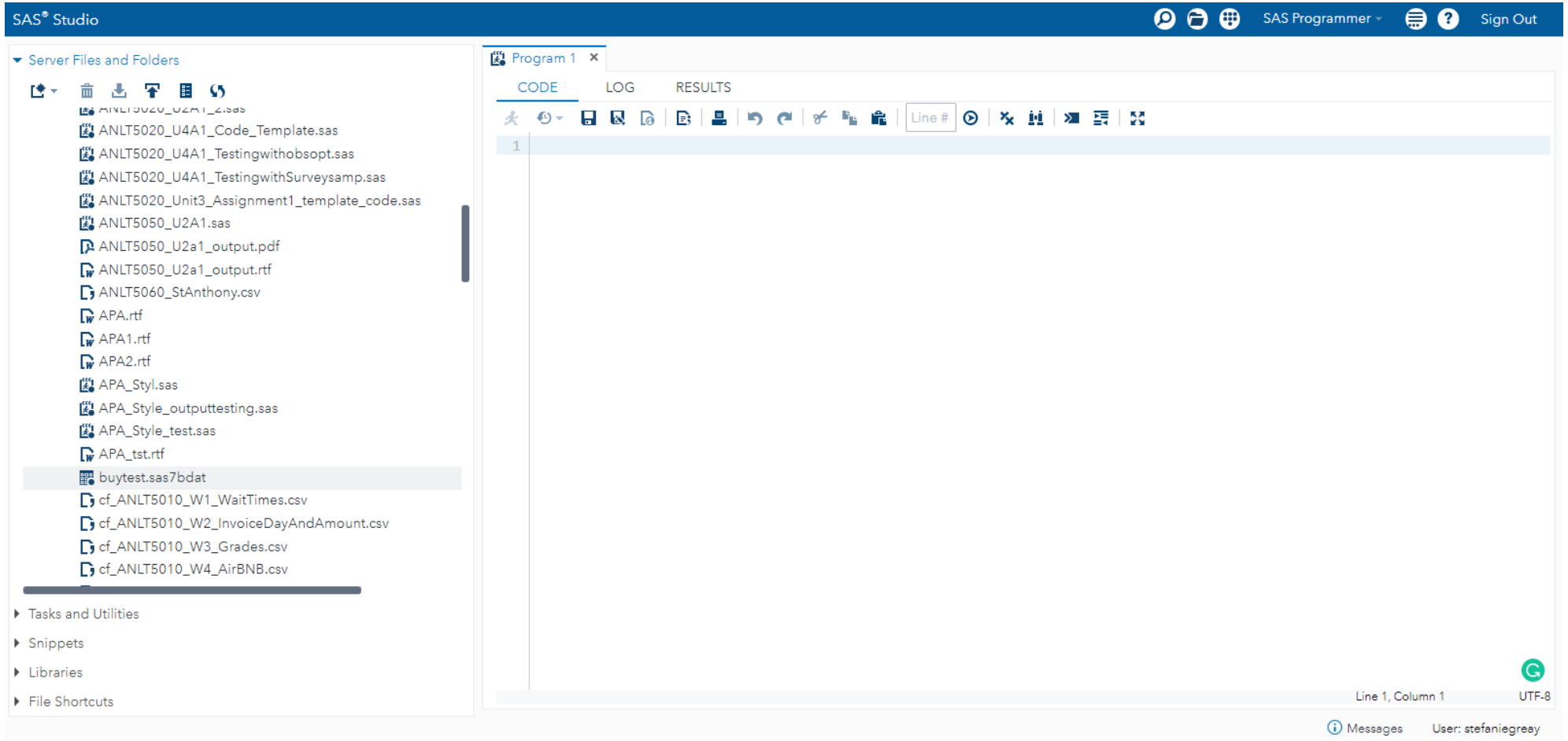
Click on “Choose Files” to browse to the file you want to upload, then click “Upload.”



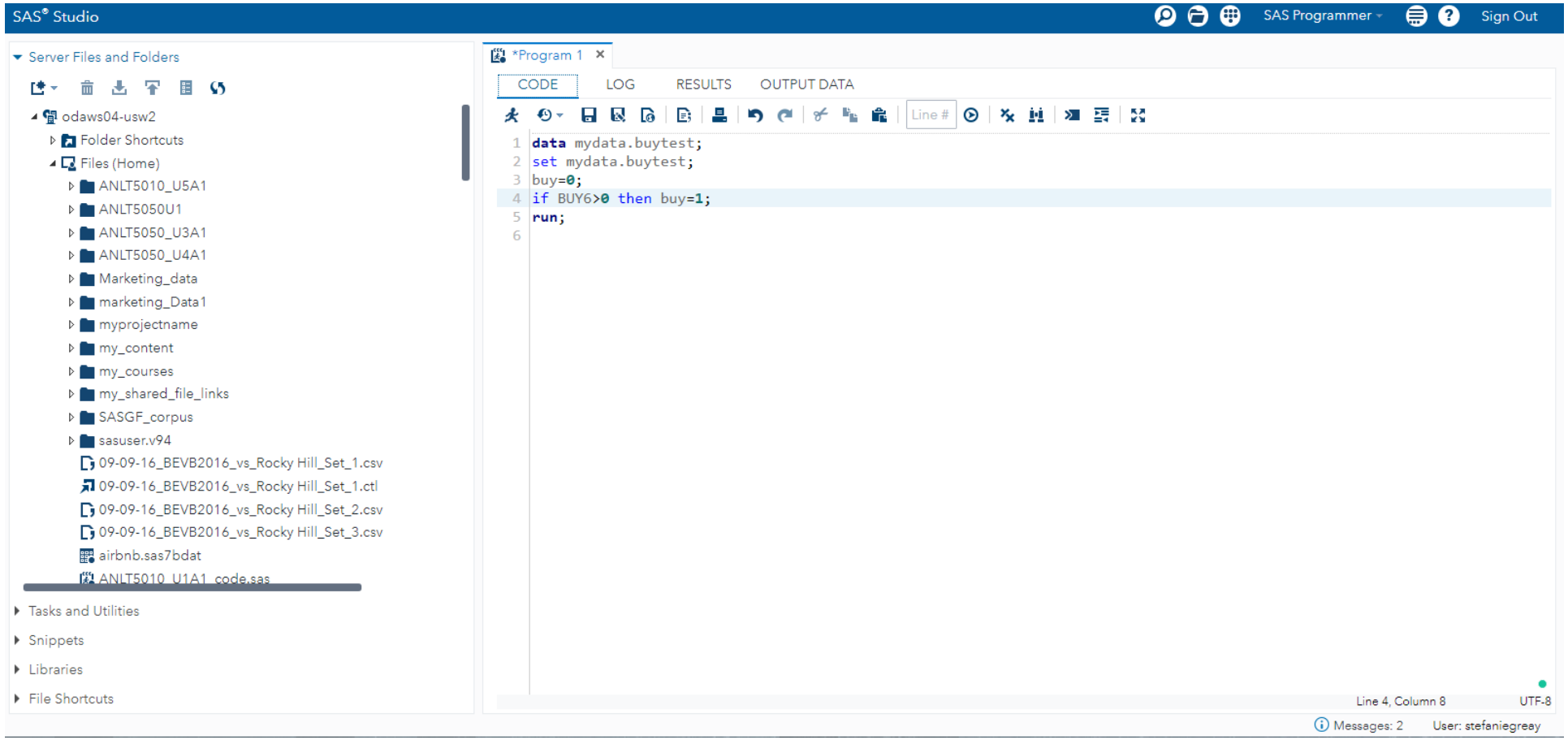
Verify that the upload was successful by scrolling down in your Files(Home) area.



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”



Code

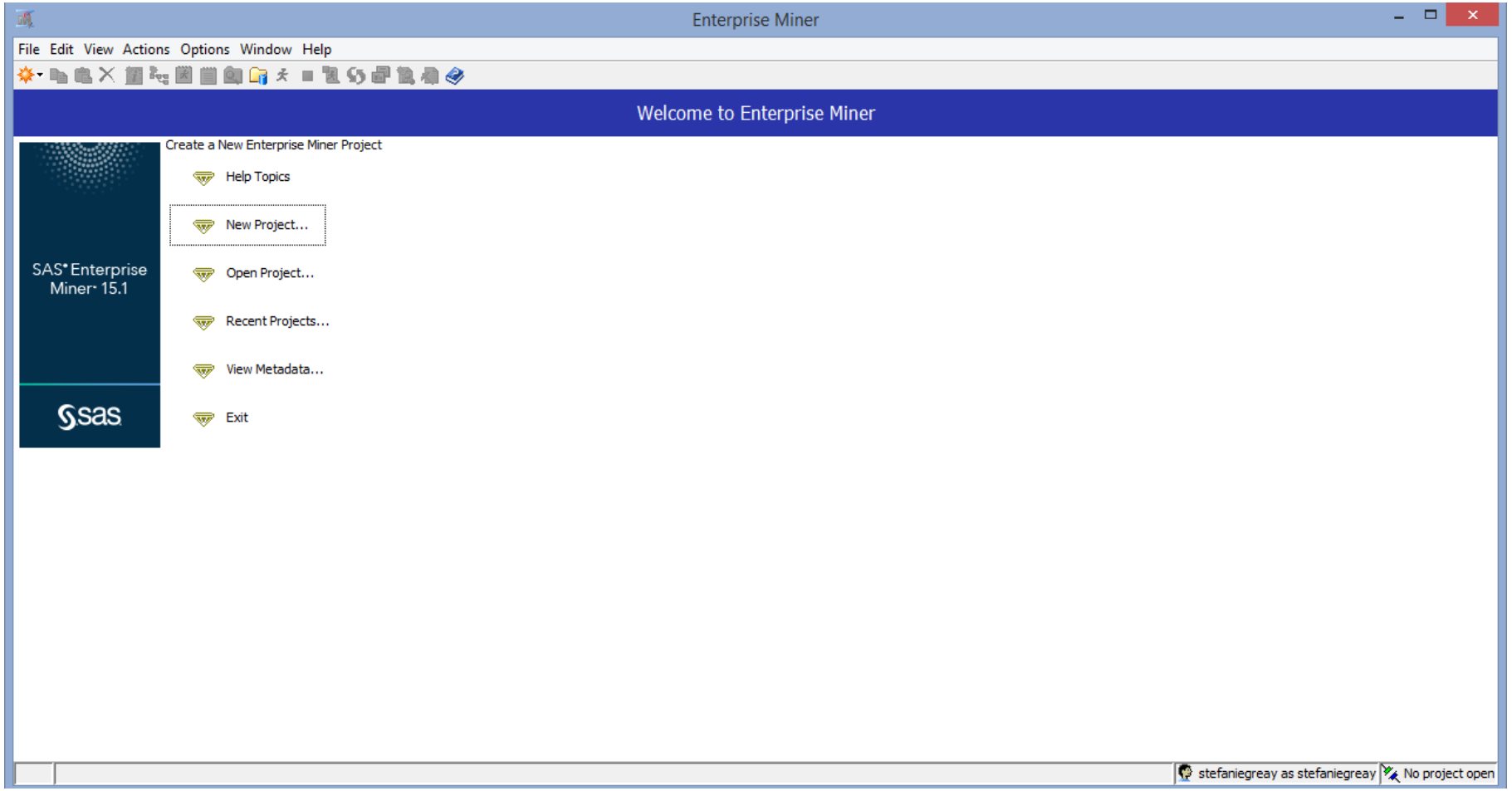
```
data mydata.buytest;
set mydata.buytest;
buy=0;
if BUY6>0 then buy=1;
run;
```

SAS Enterprise Miner Instructions

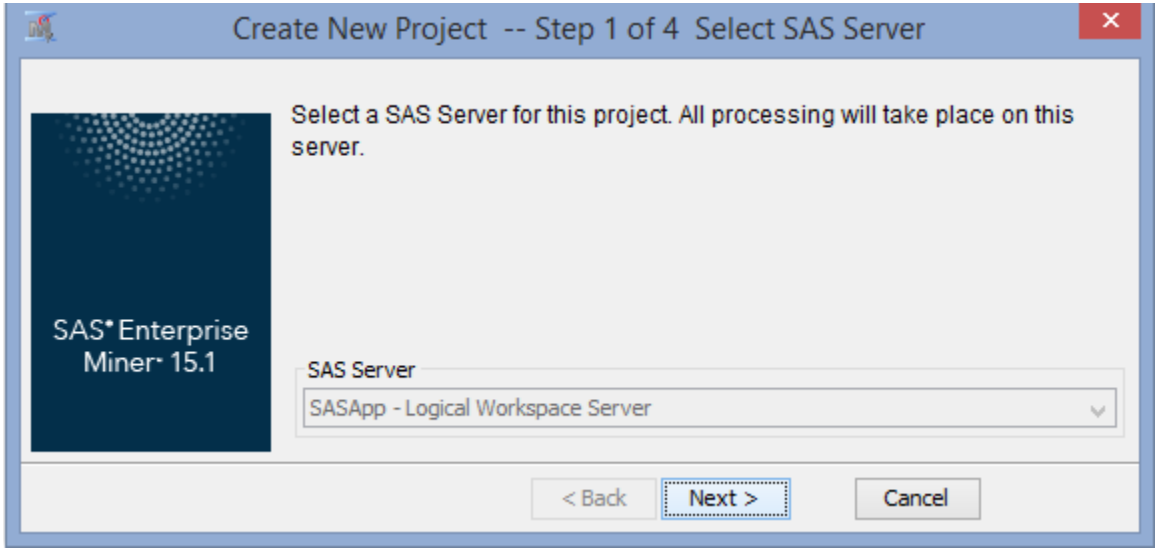
The following slides provide instructions on how to complete this task in SAS Enterprise Miner.

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.

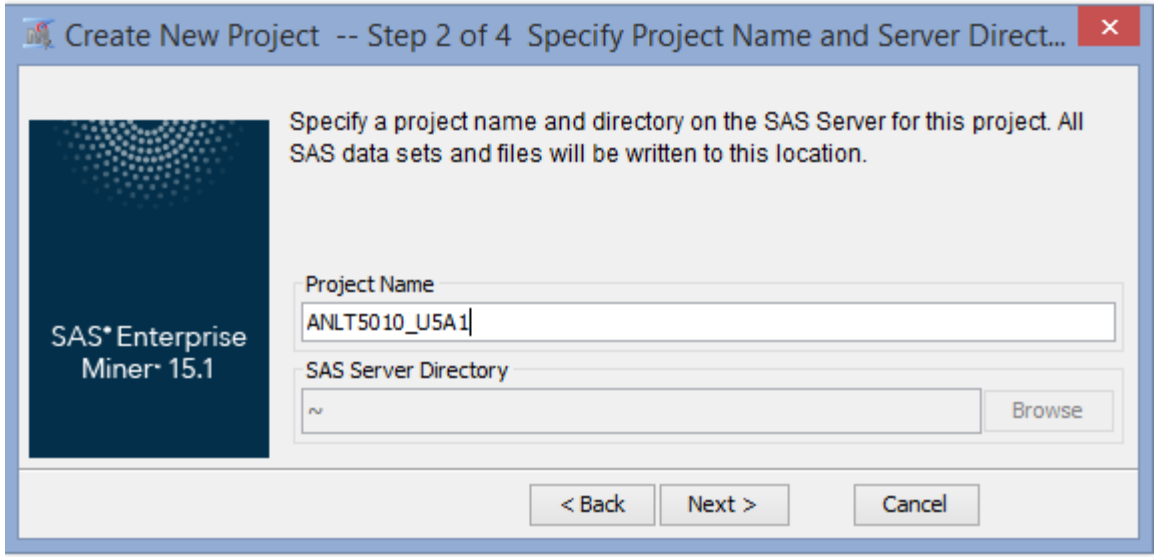
Once you download and start SAS Enterprise Miner, open a new project by clicking on “New Project.”



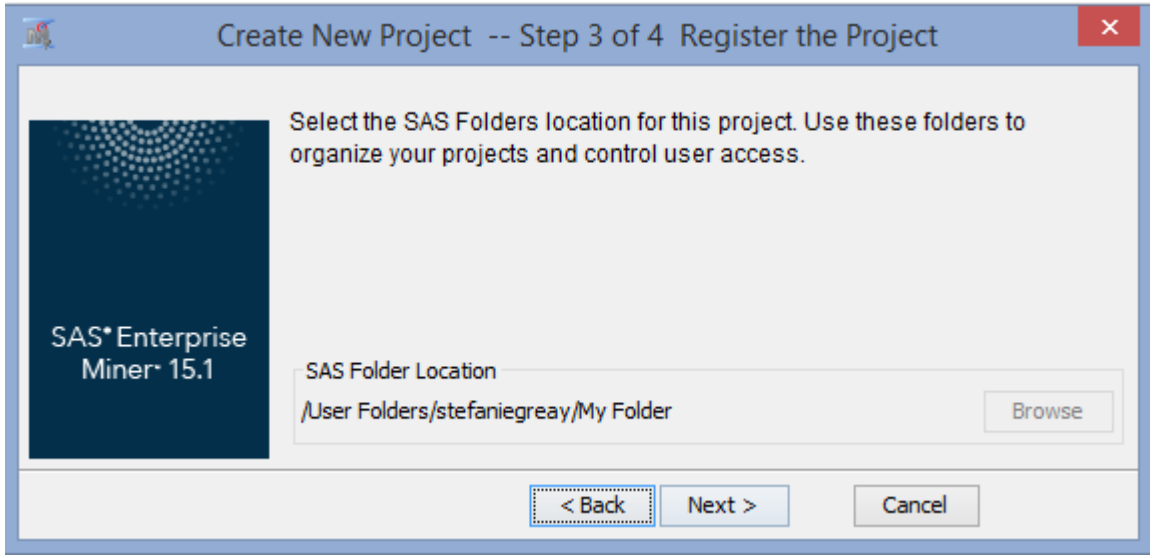
Click “Next>” to use the default SAS Server



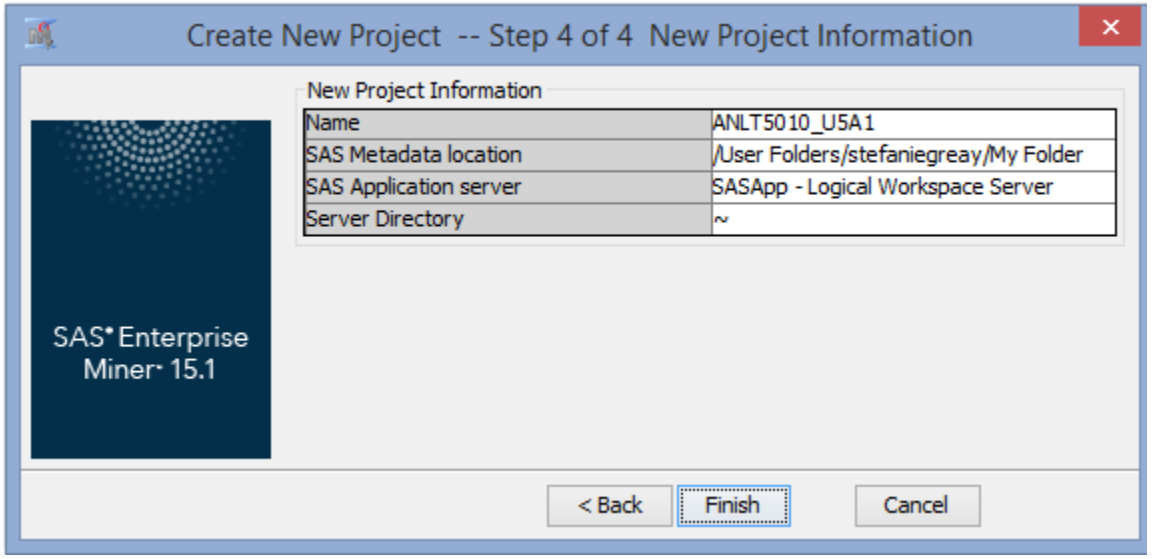
Enter a project name and click “Next>”



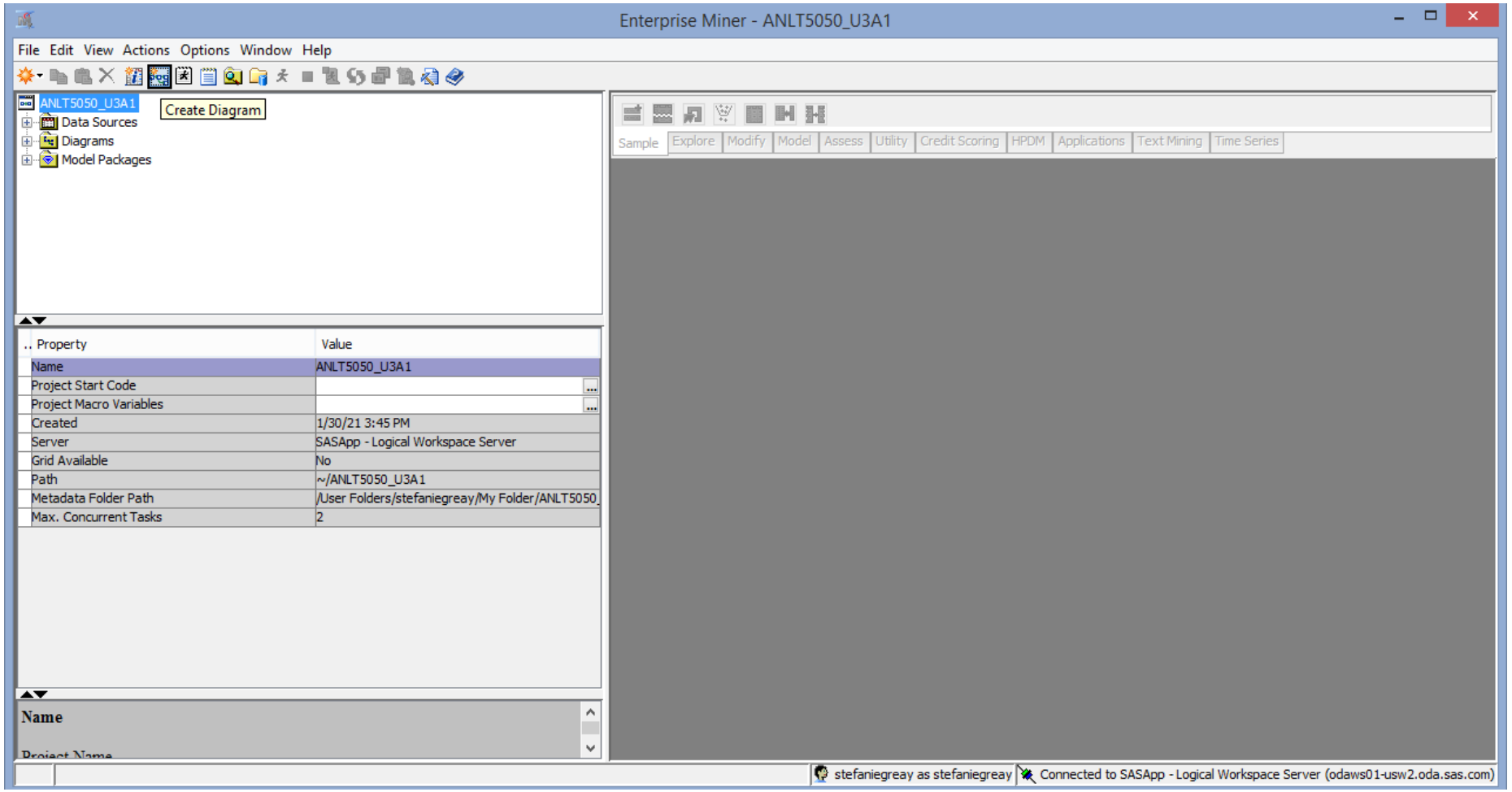
Click “Next>”



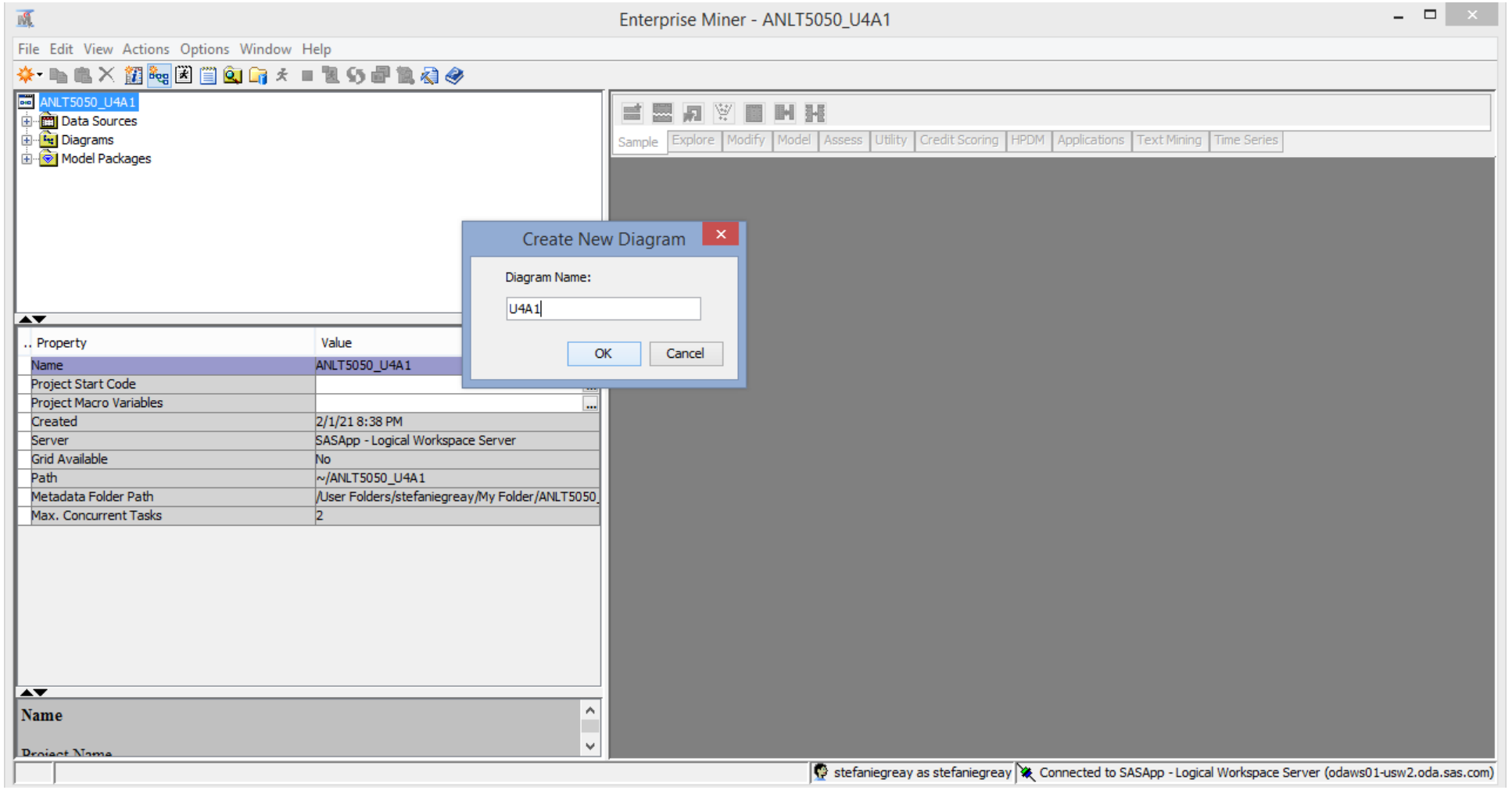
Verify your entries and click “Finish”



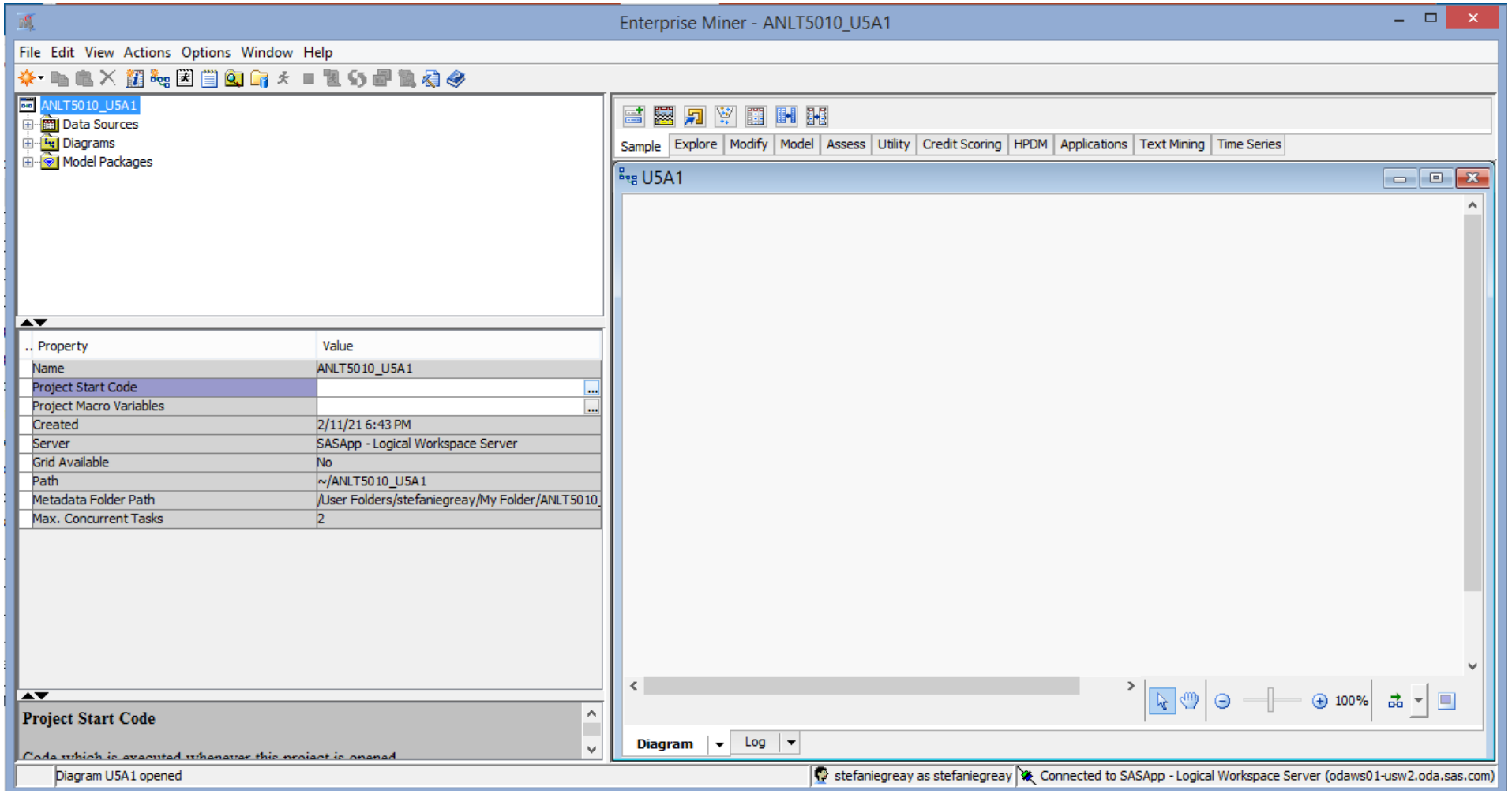
Click on the “Create Diagram” icon.



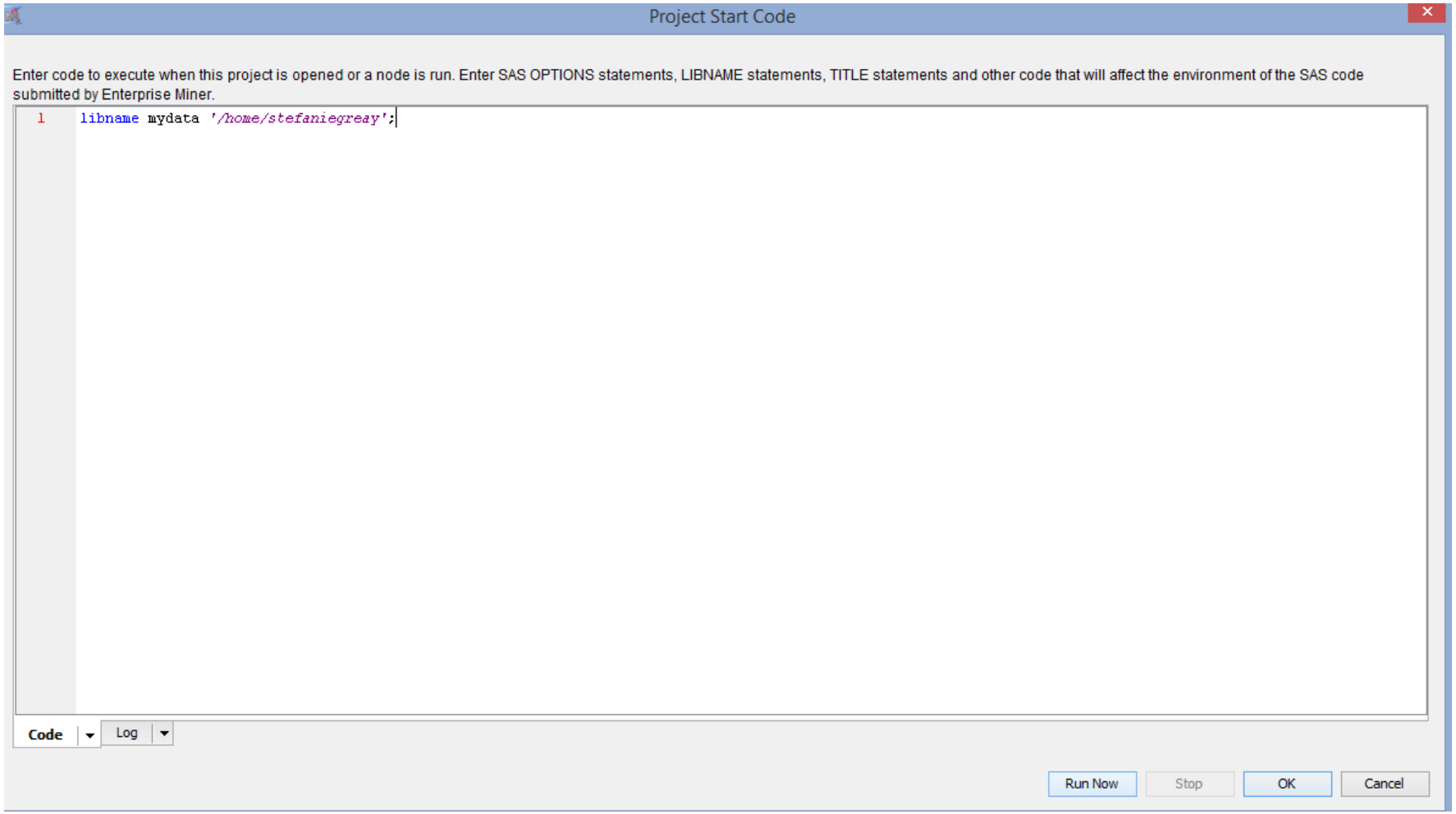
Name your diagram and click “OK.”



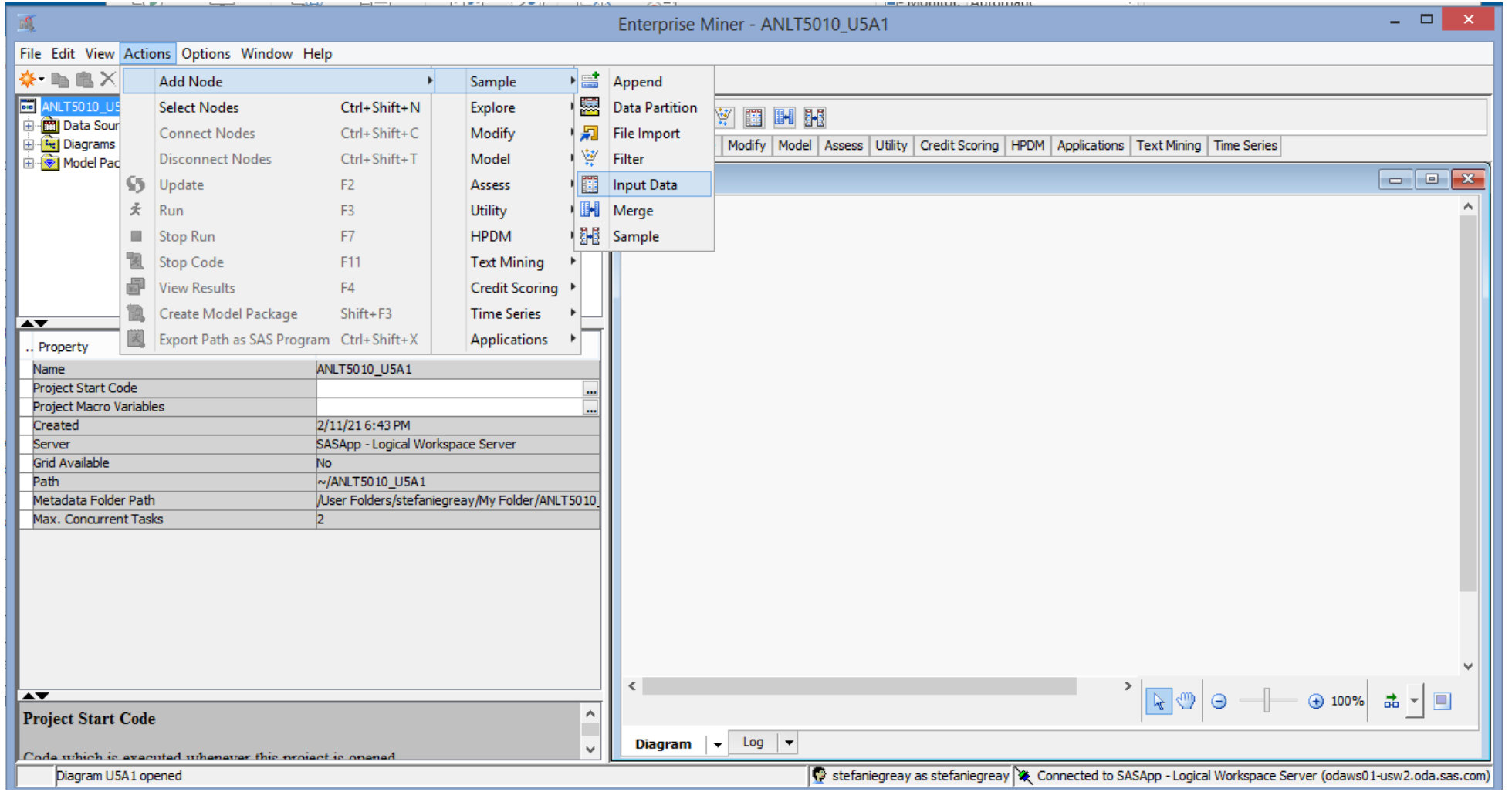
Click on the project, then click on the ellipses next to “Project Start Code.”



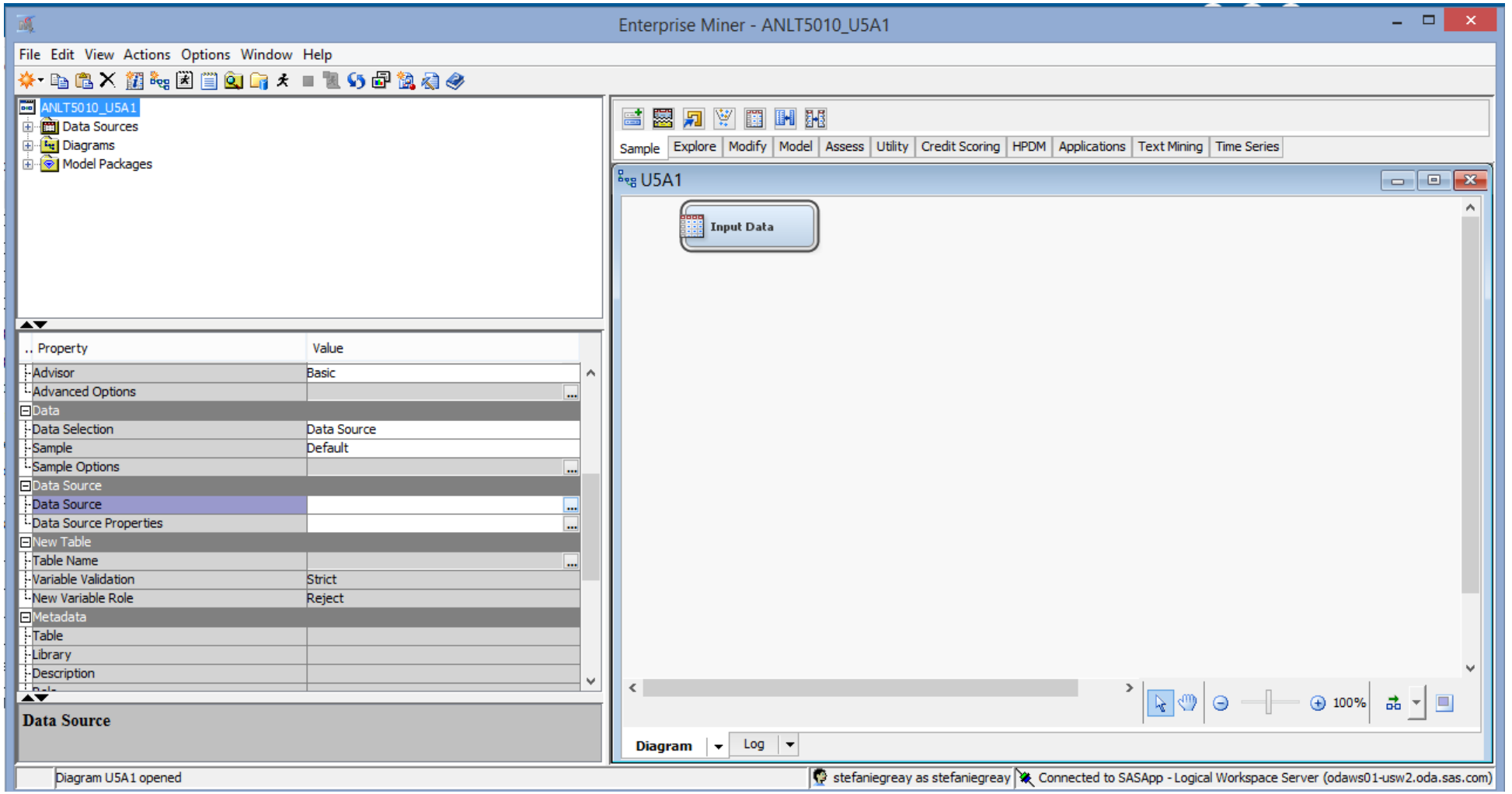
Add the library reference for where you uploaded the dataset in SAS studio, and click “Run Now.” Once it completes, click “OK.”



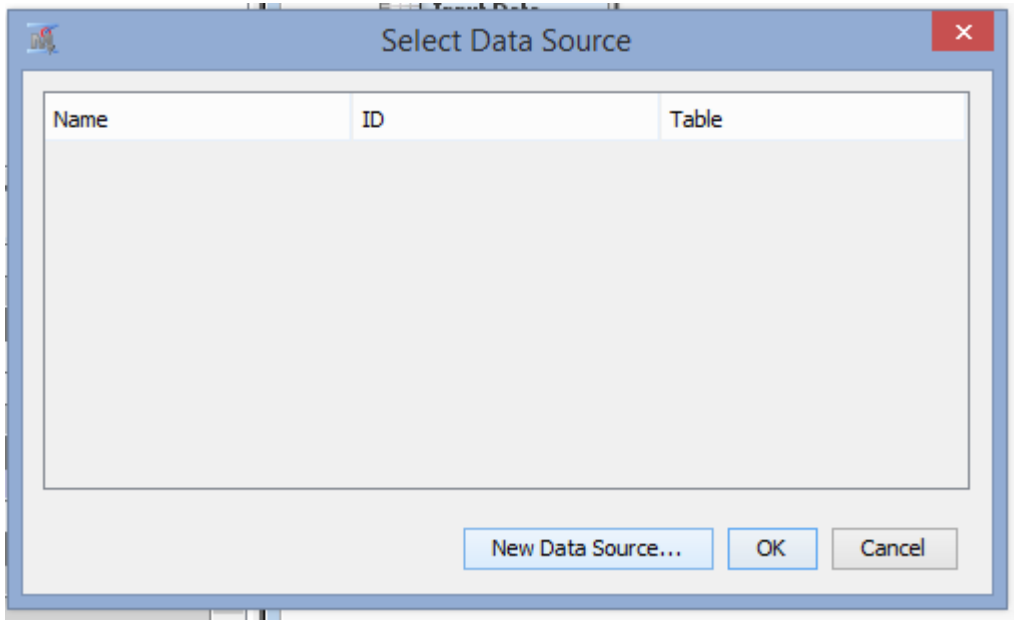
Click on Actions>Add Node>Sample>Input Data



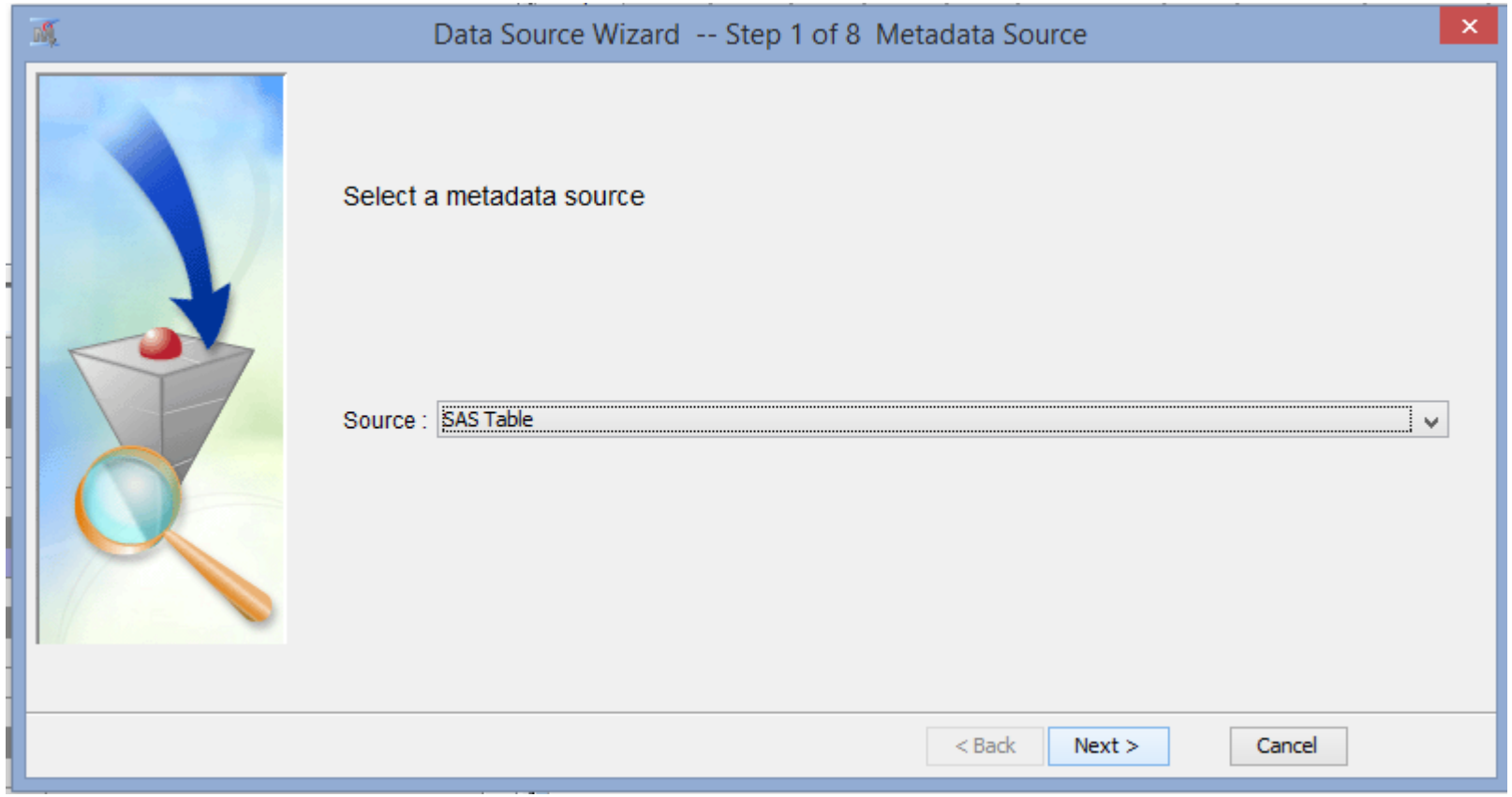
Click the ellipses (3 dots) next to “Data Source.”



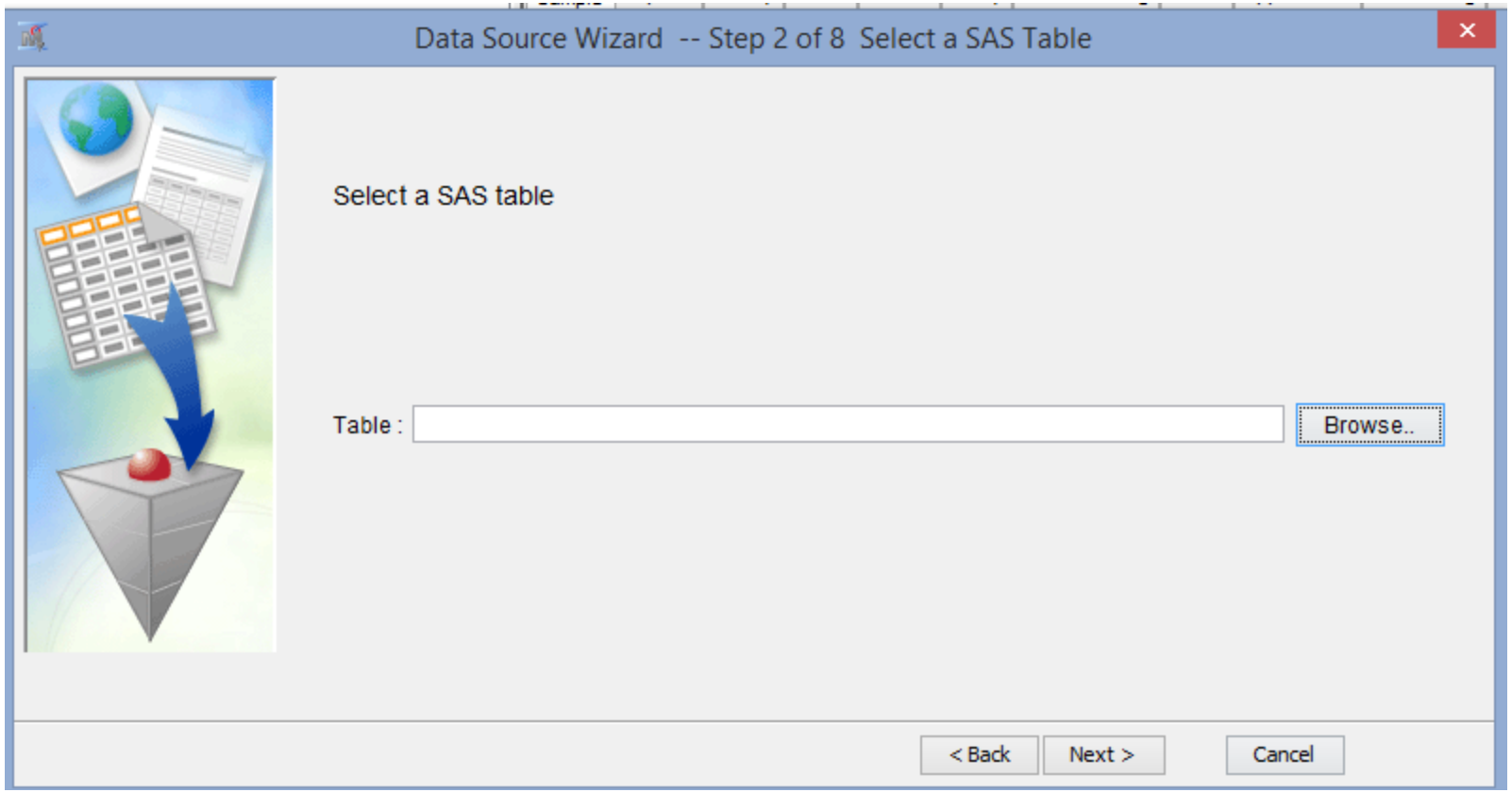
Click on “New Data Source”



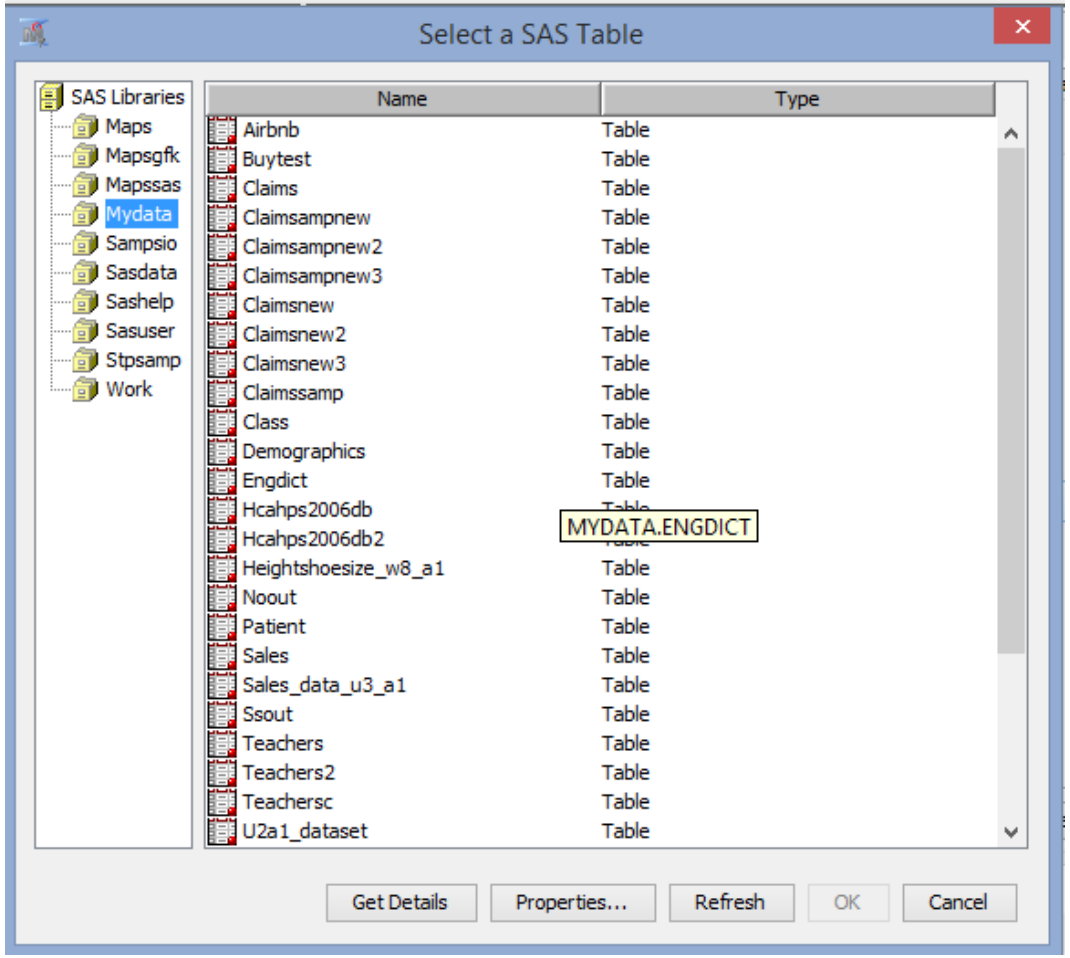
Leave it as “SAS Table” and click “Next >”



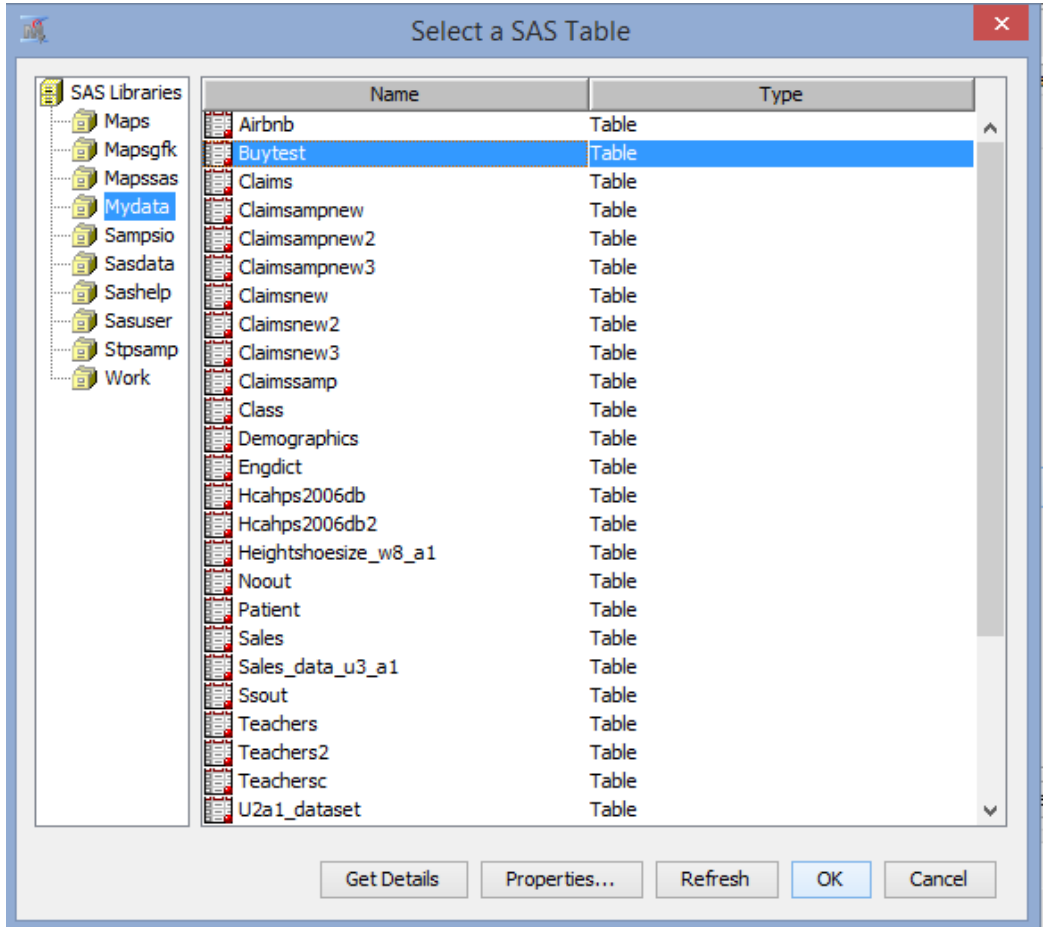
Click on “Browse”



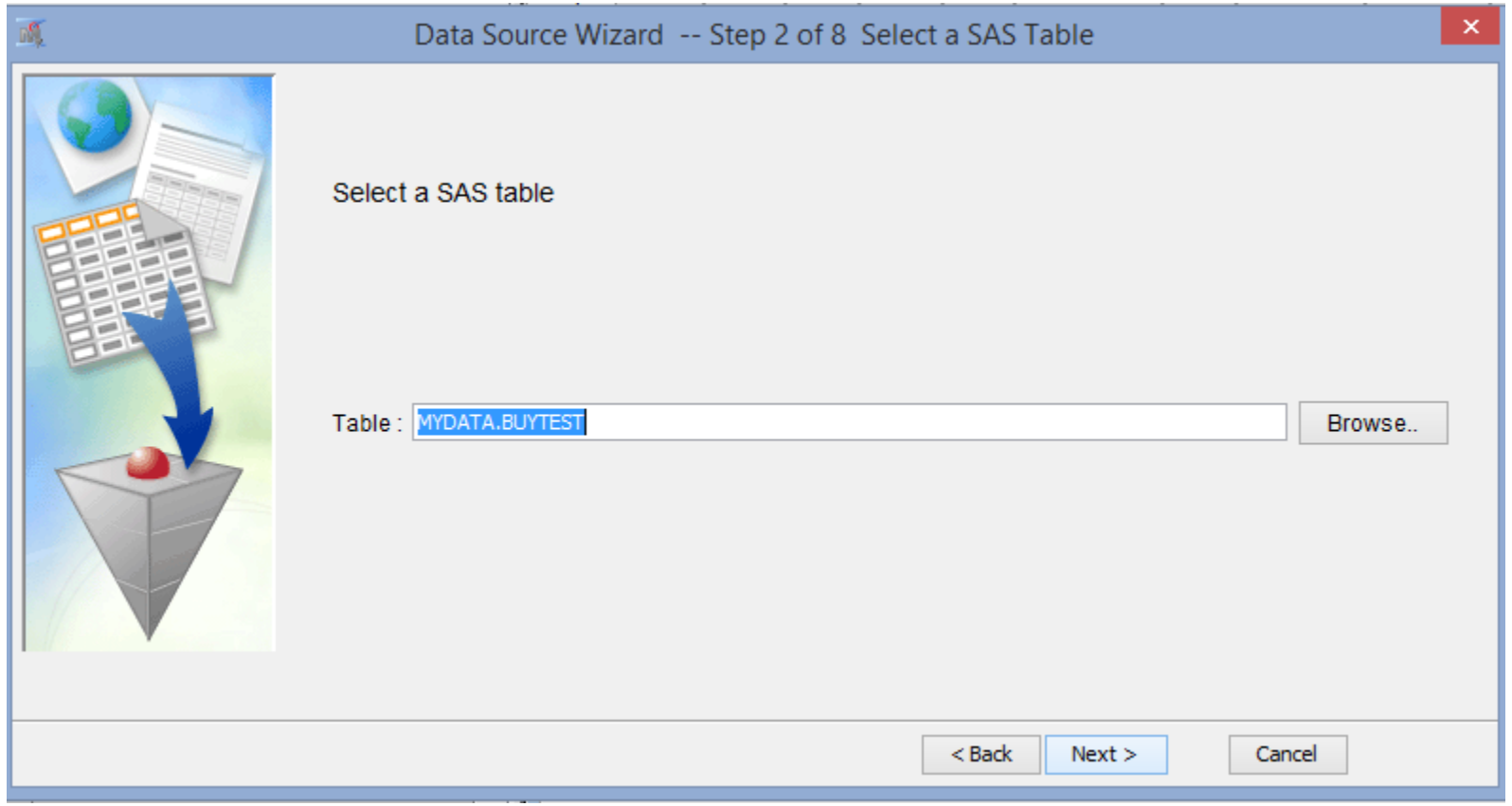
Double click on the libname you just set up in the project startup code.



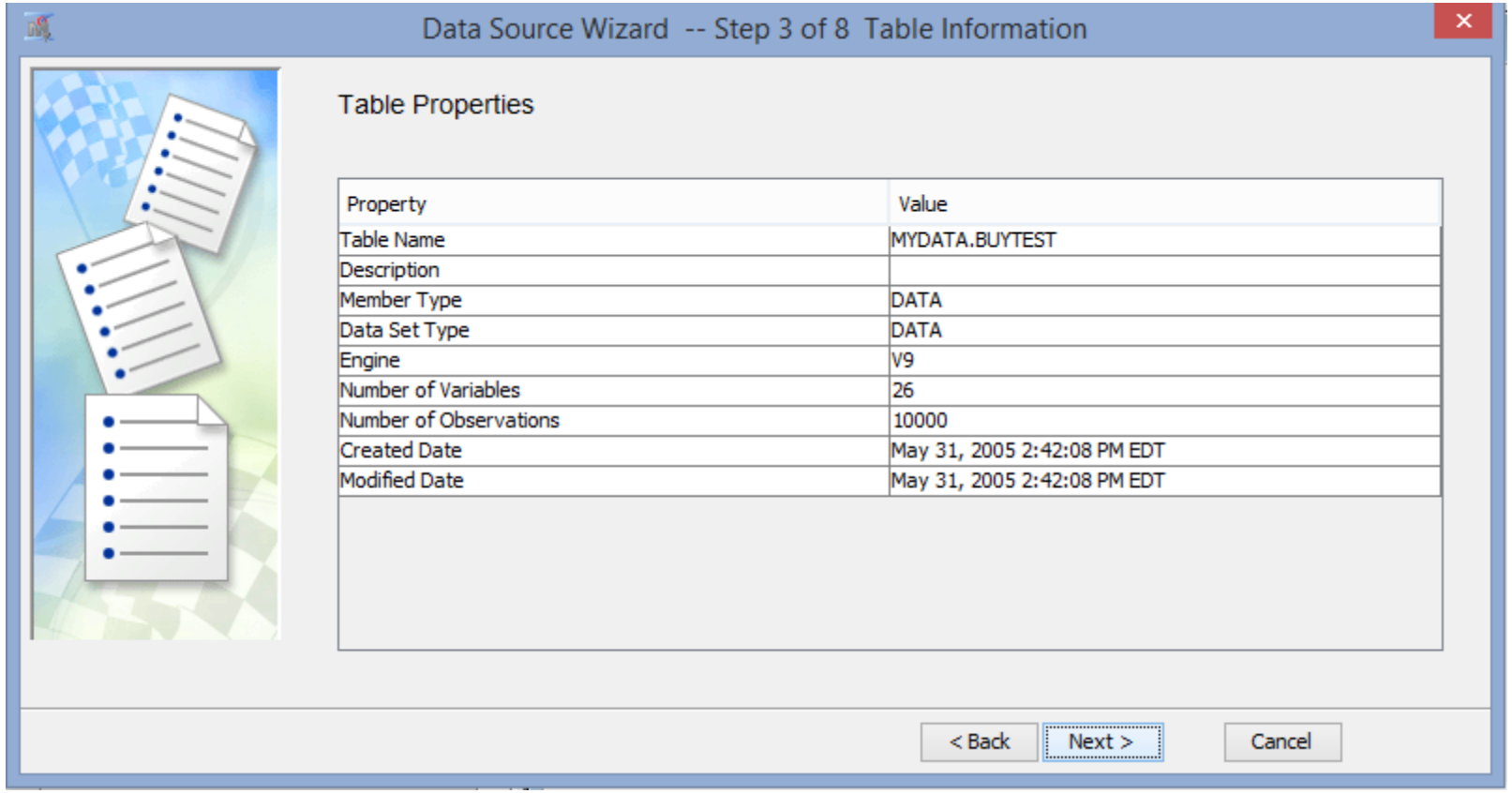
Double click to select the dataset for this unit, and click “OK”



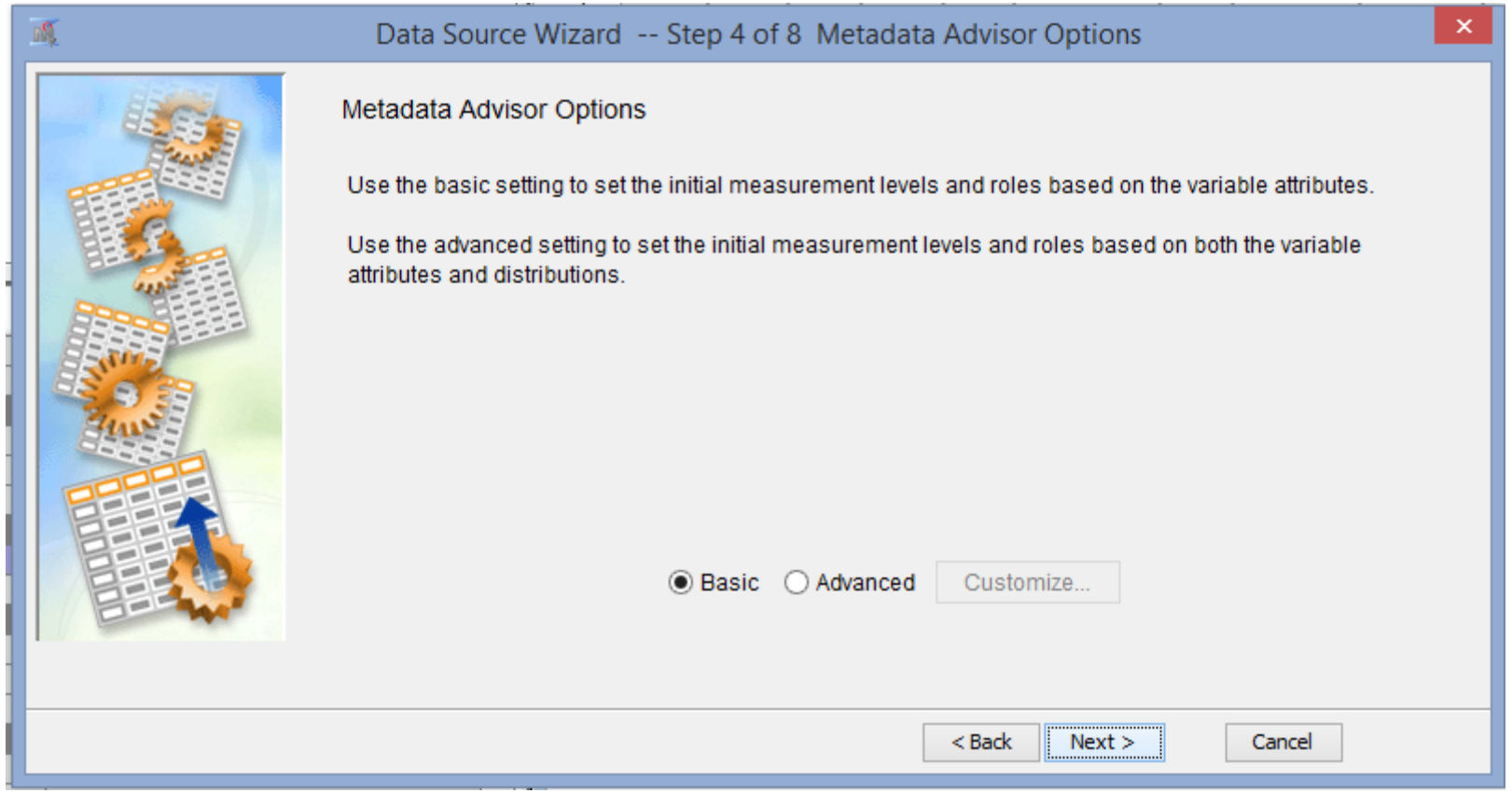
Click “Next>”



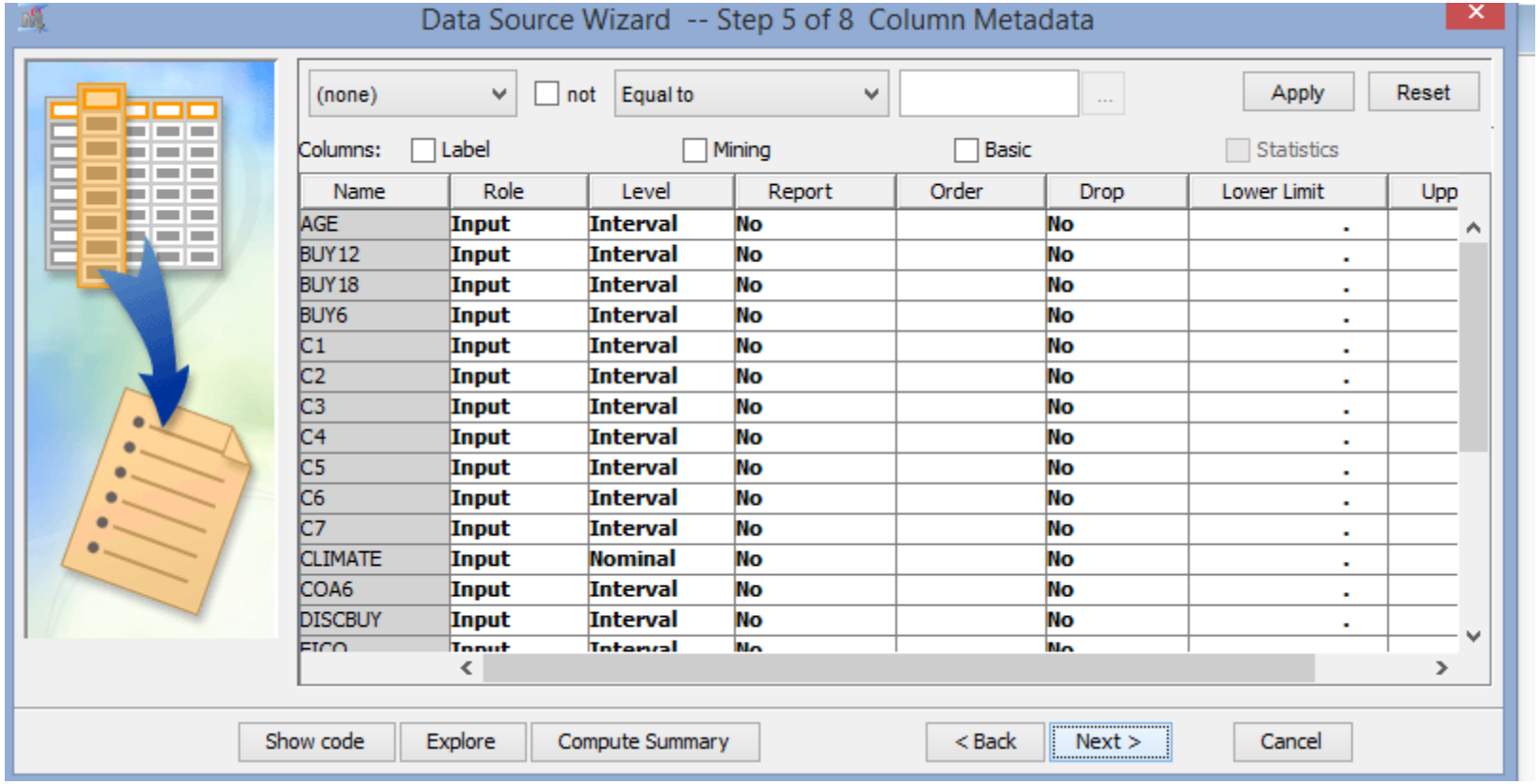
Verify the options and click “Next>”



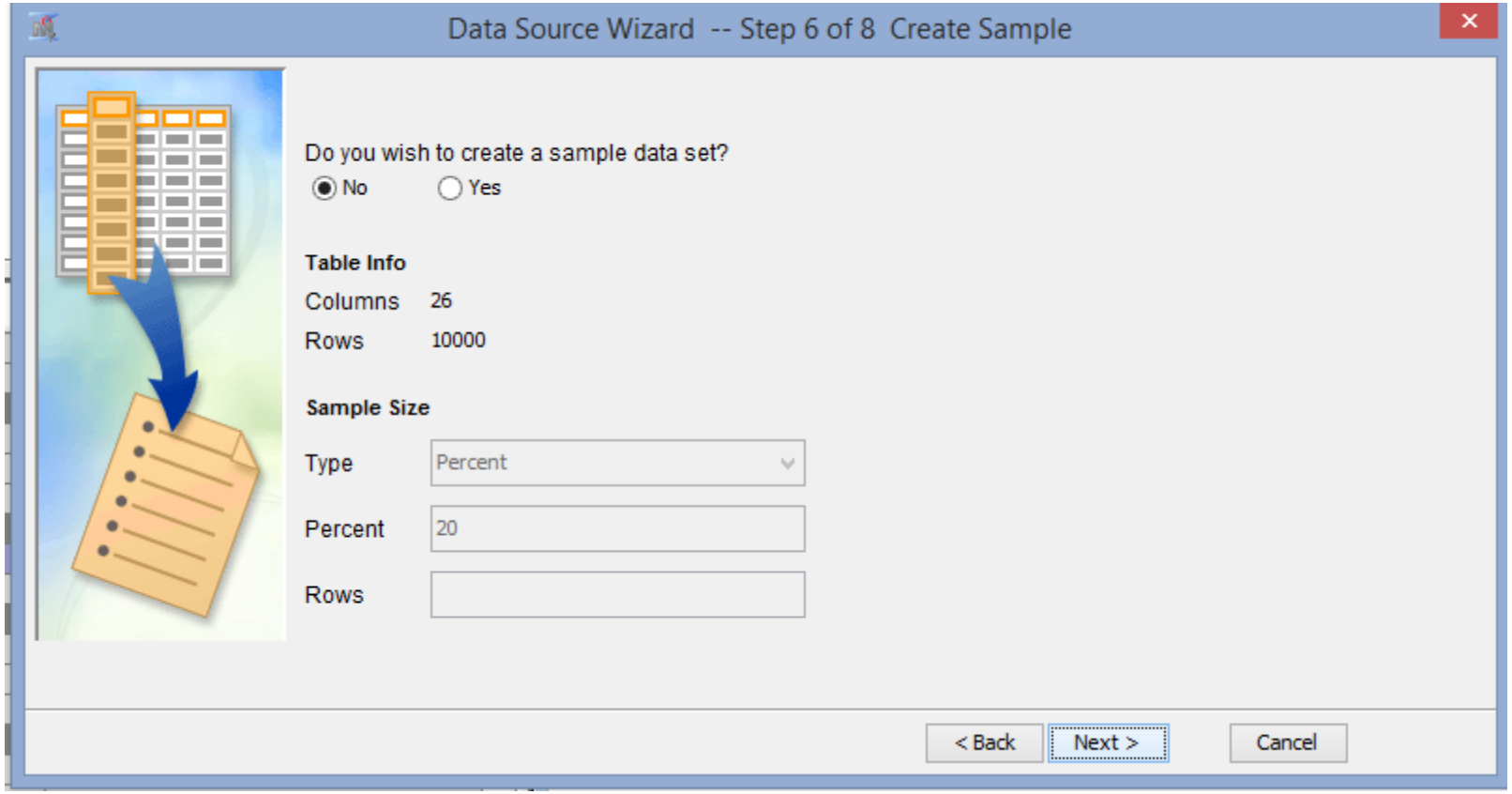
Click “Next>”



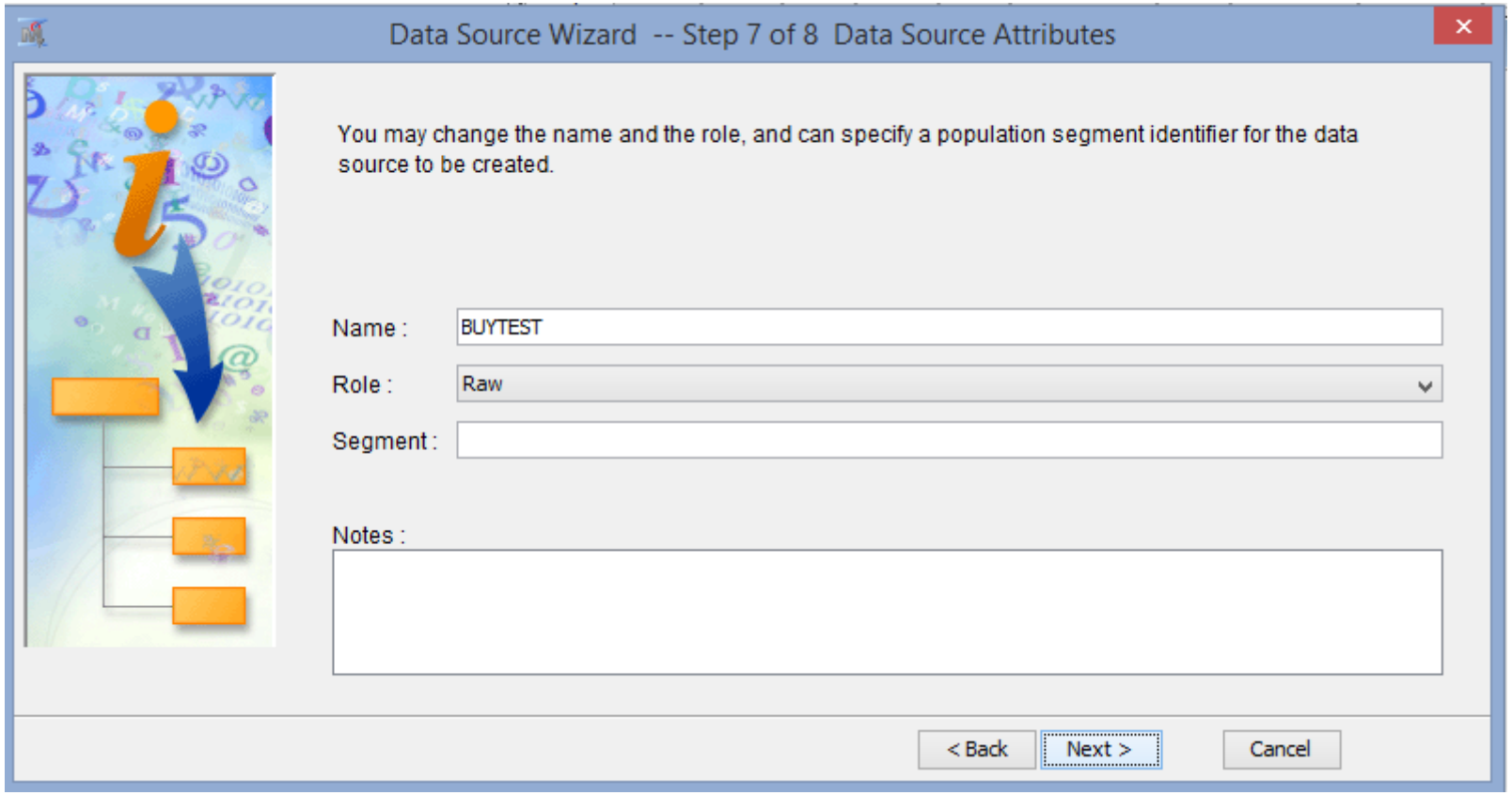
Verify the variables and settings, adjust if necessary, and then click “Next>”



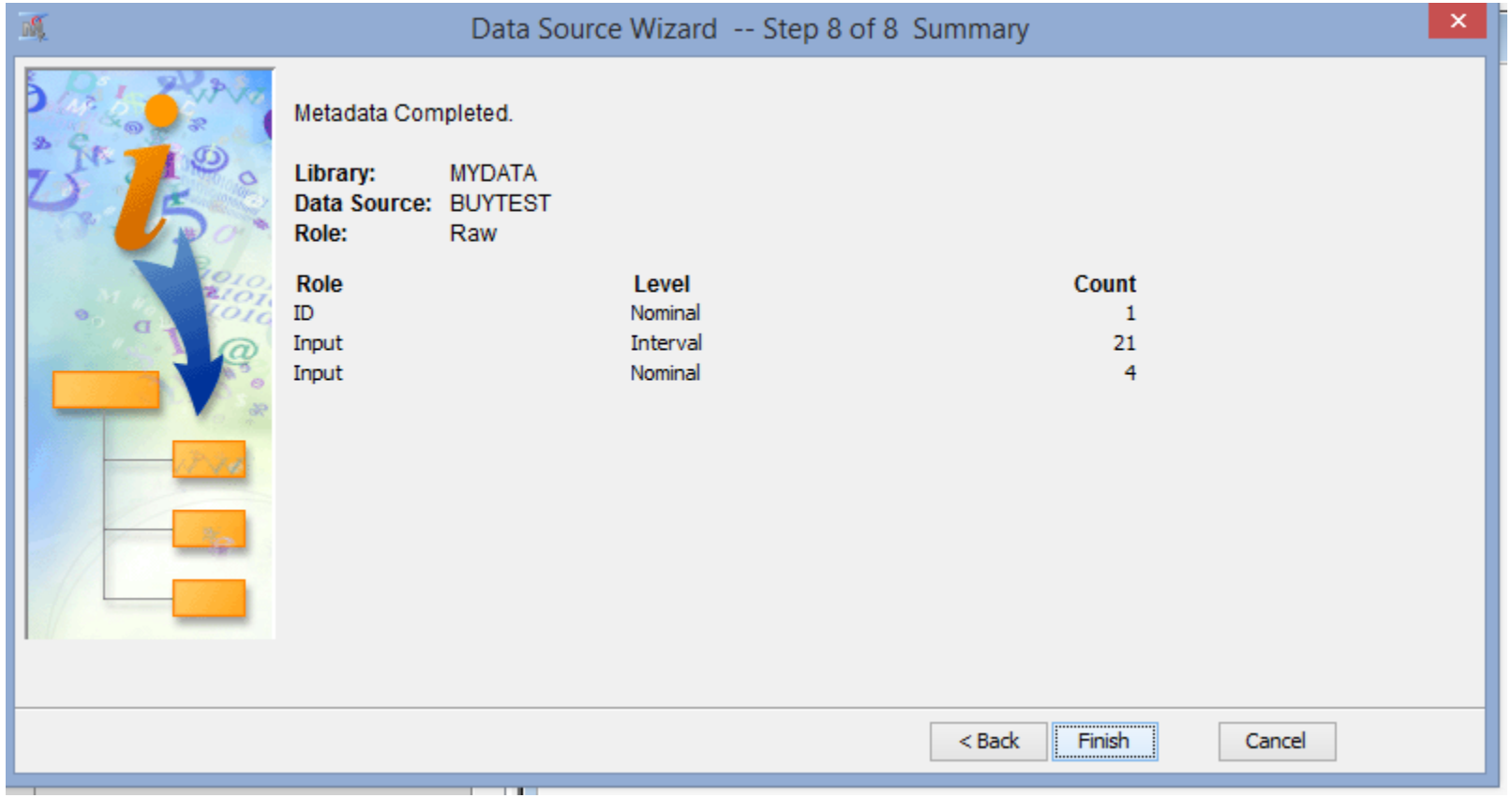
You may choose to sample the dataset here, or just keep the full dataset, then click “Next>”



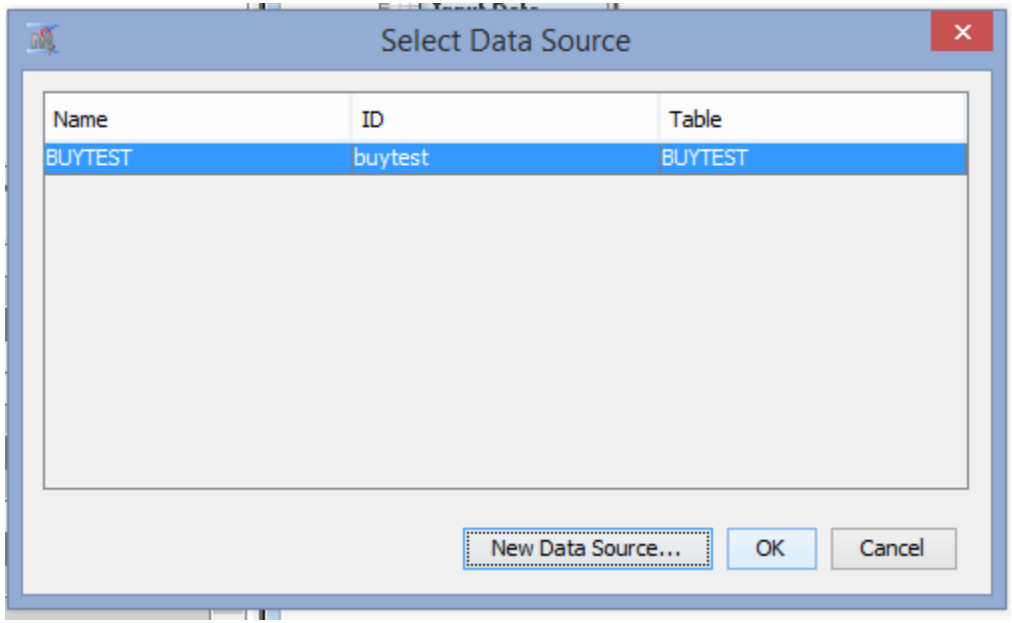
You may choose to adjust the role of the dataset, or leave it as the default, then click “Next>”



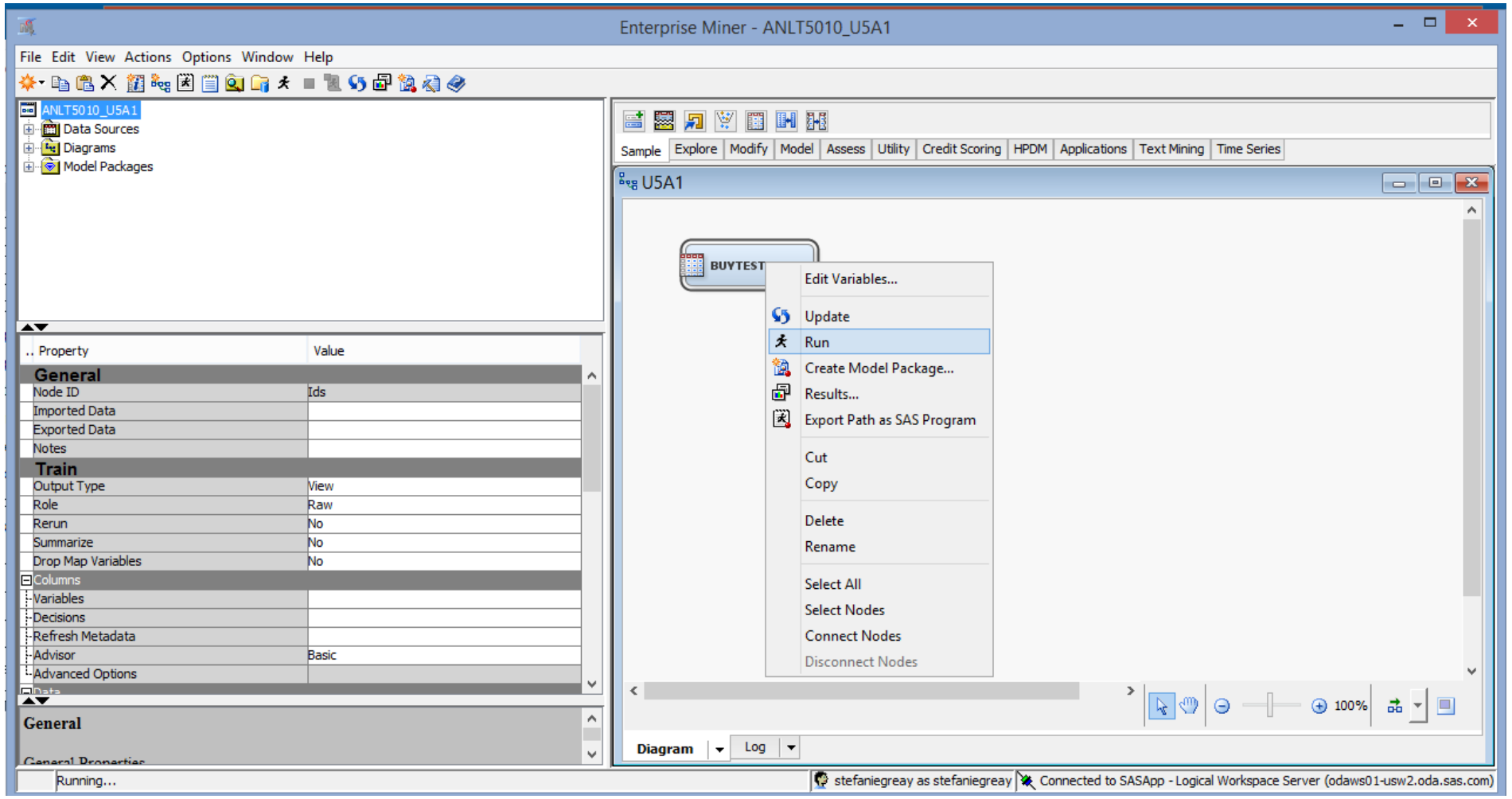
Click “Finish” to finish the data source registration within EG.



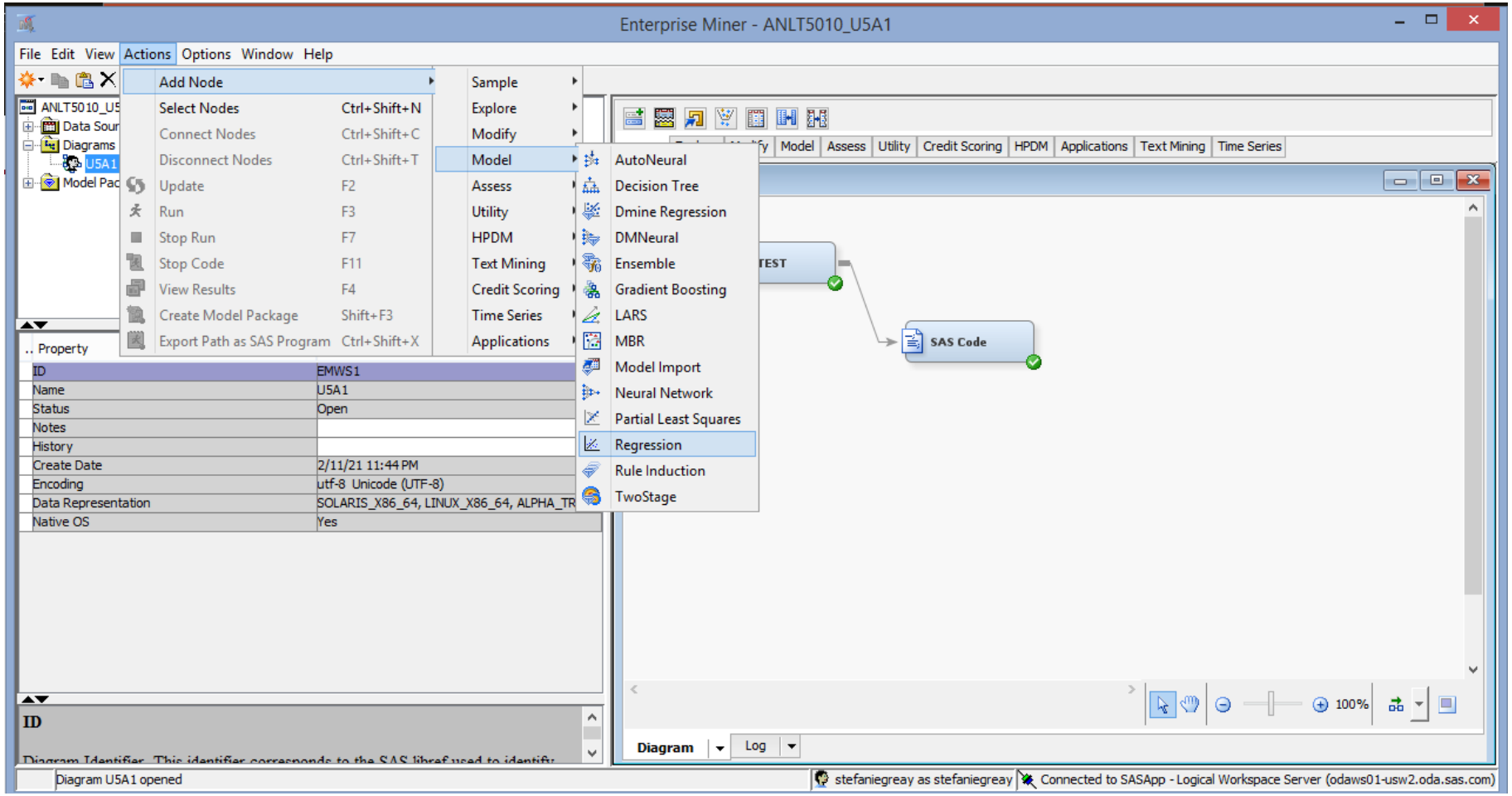
Click “OK” to complete the process. The name of the node should then change to the name of the dataset.



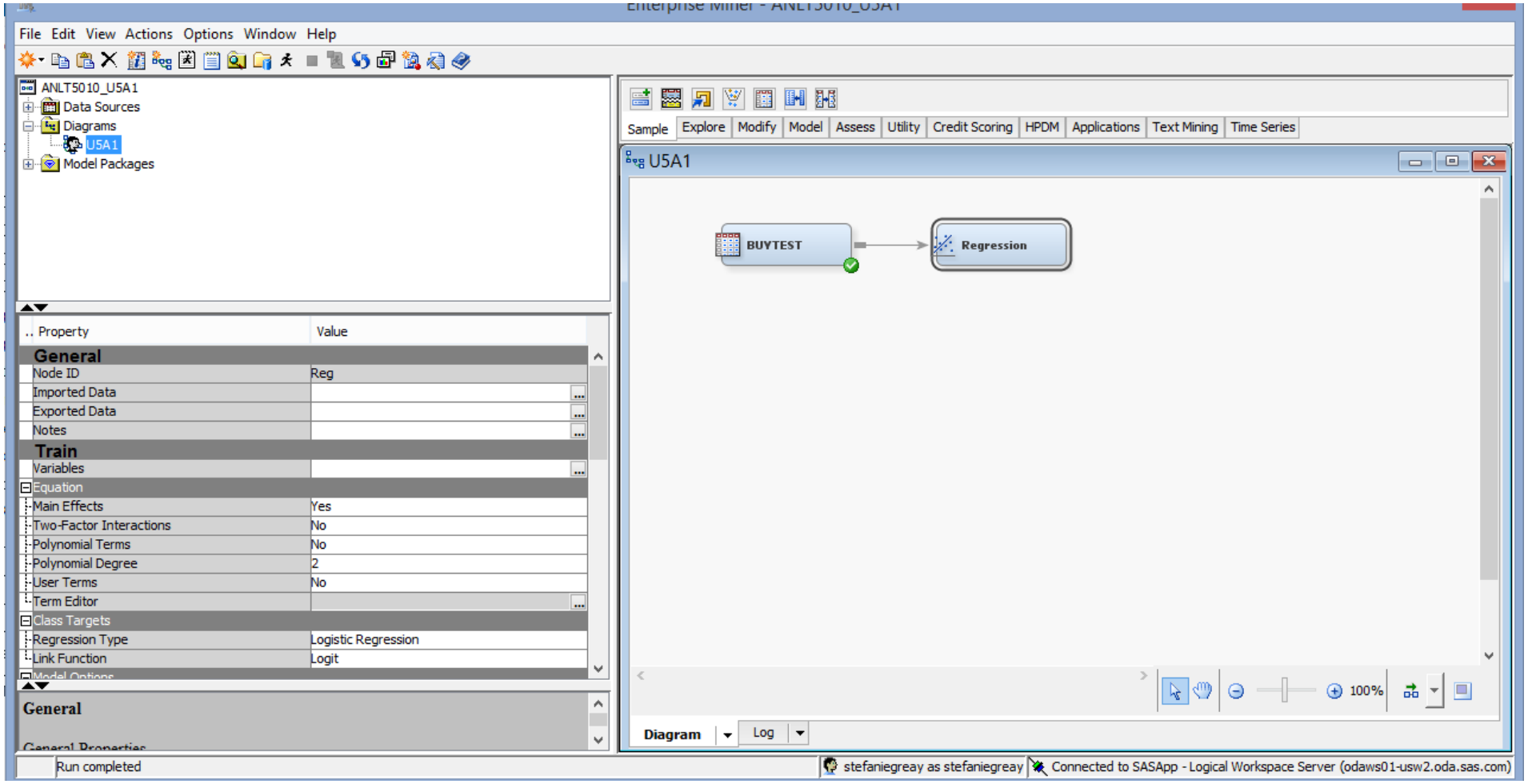
Right click on the dataset node and click “Run.”



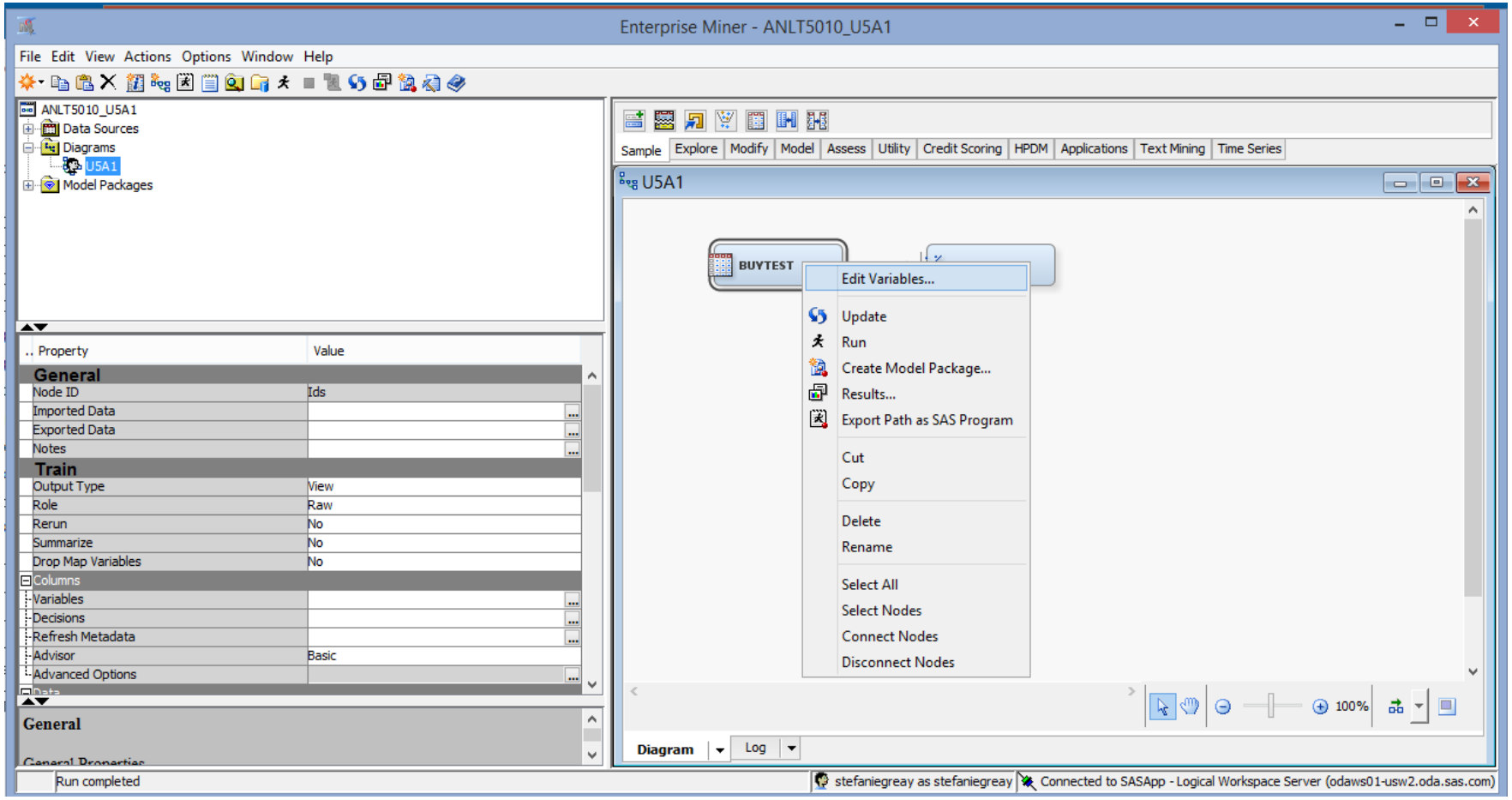
Click on “Actions” > “Add Node” > “Model” > “Regression”



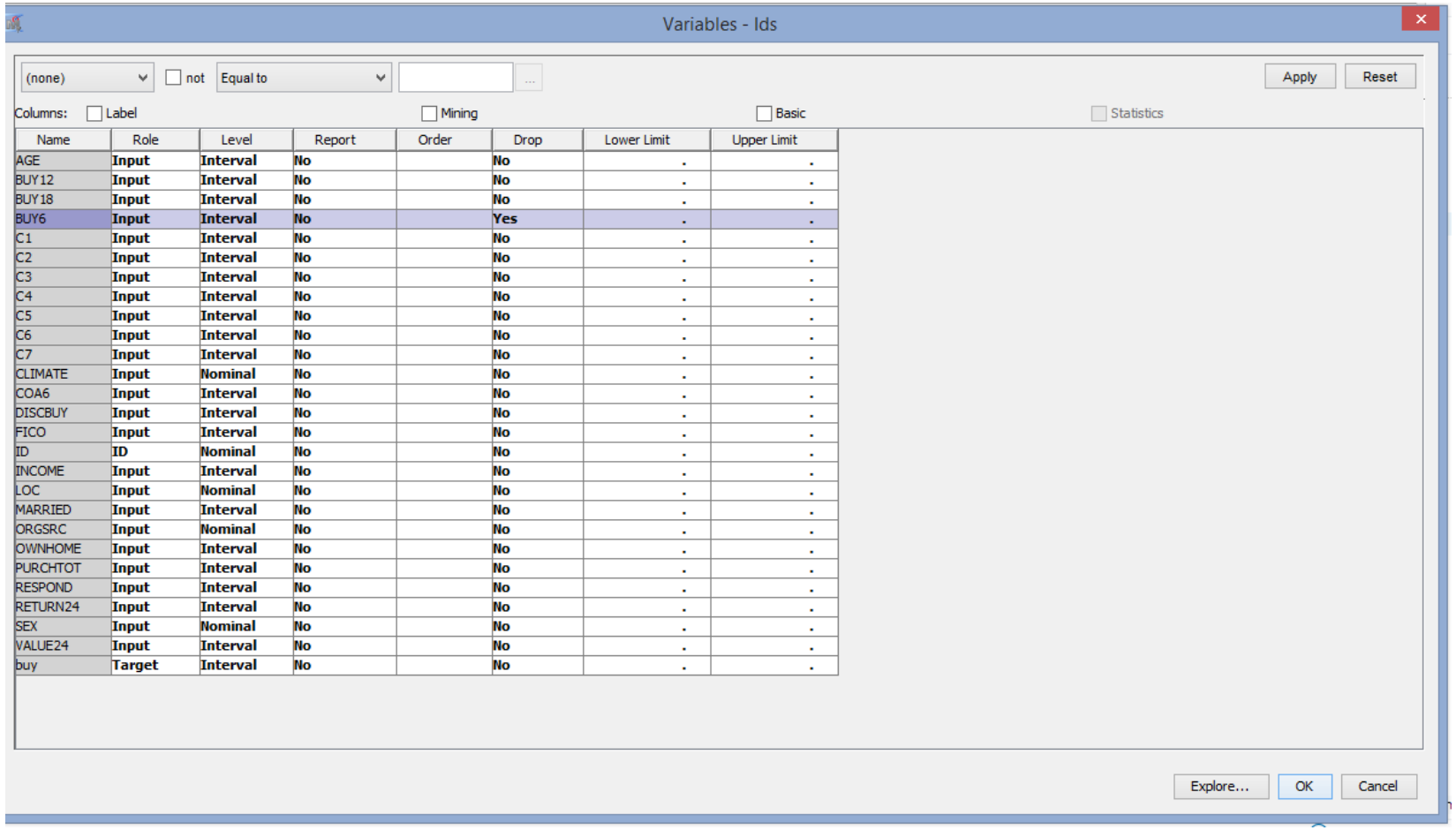
Connect the nodes



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

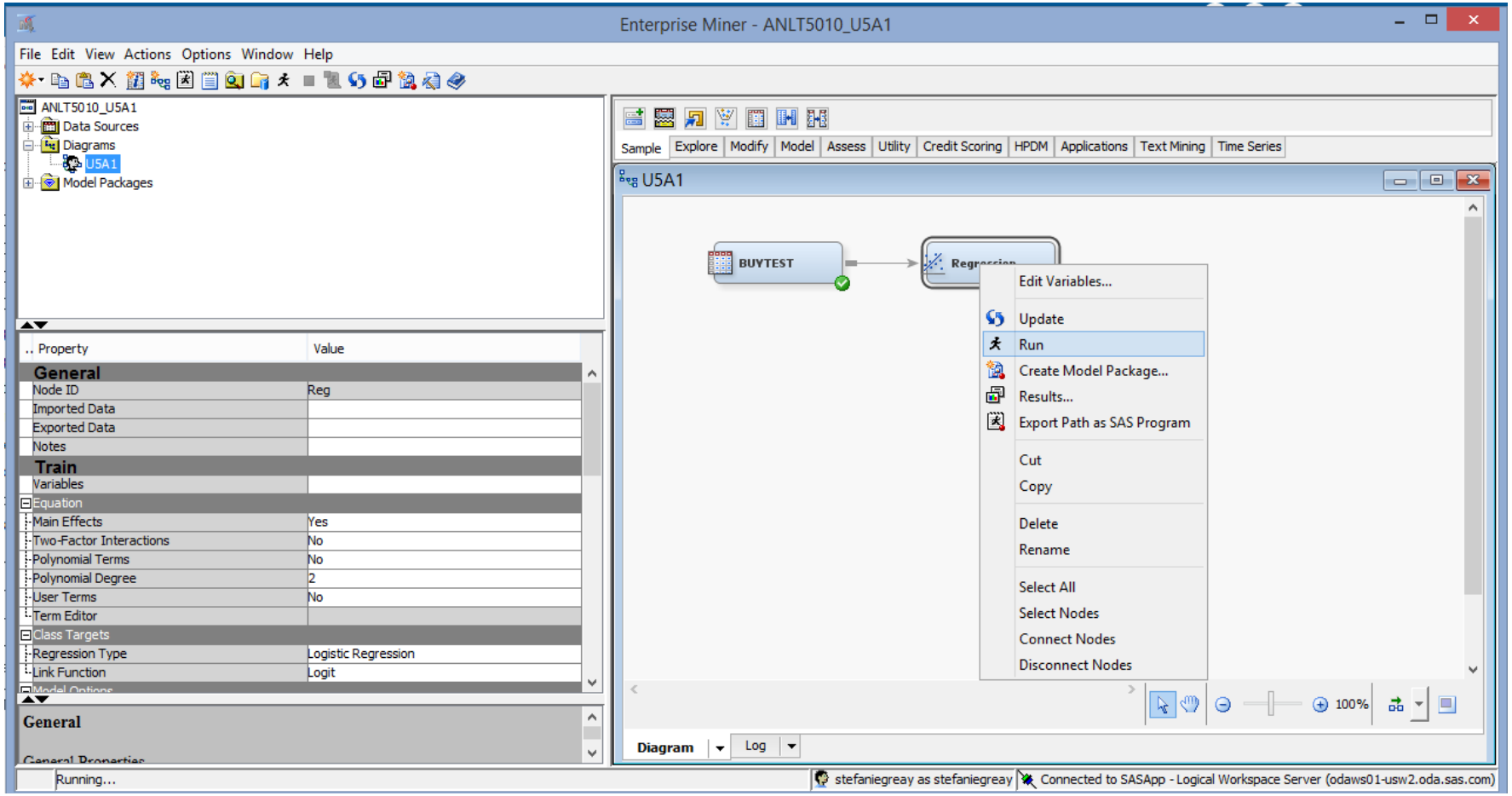


Change the “buy” variable we just made to “Target” and the “Buy6” variable to Drop=Yes

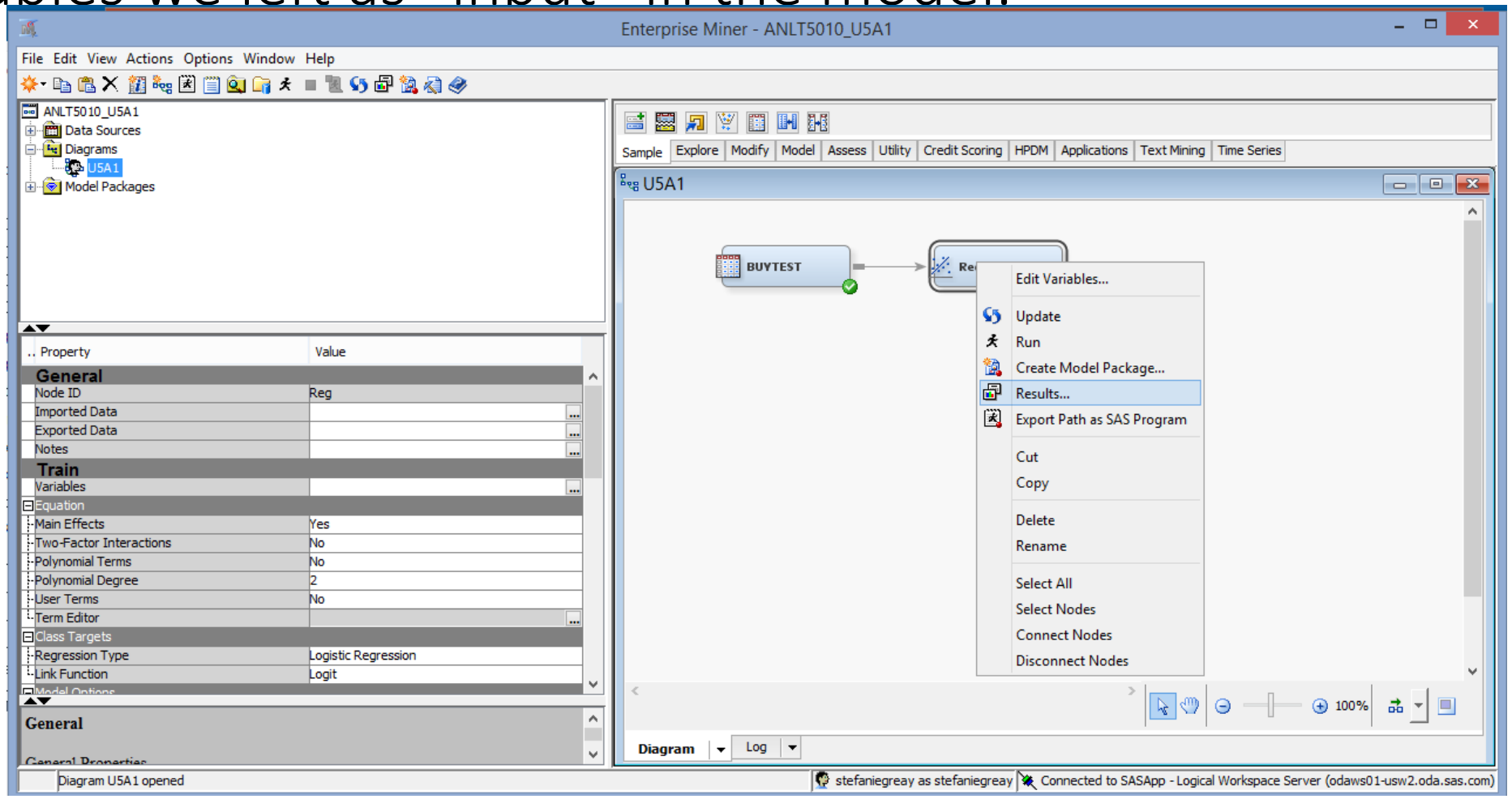


Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
AGE	Input	Interval	No		No	.	.
BUY12	Input	Interval	No		No	.	.
BUY18	Input	Interval	No		No	.	.
BUY6	Input	Interval	No		Yes	.	.
C1	Input	Interval	No		No	.	.
C2	Input	Interval	No		No	.	.
C3	Input	Interval	No		No	.	.
C4	Input	Interval	No		No	.	.
C5	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	.	.

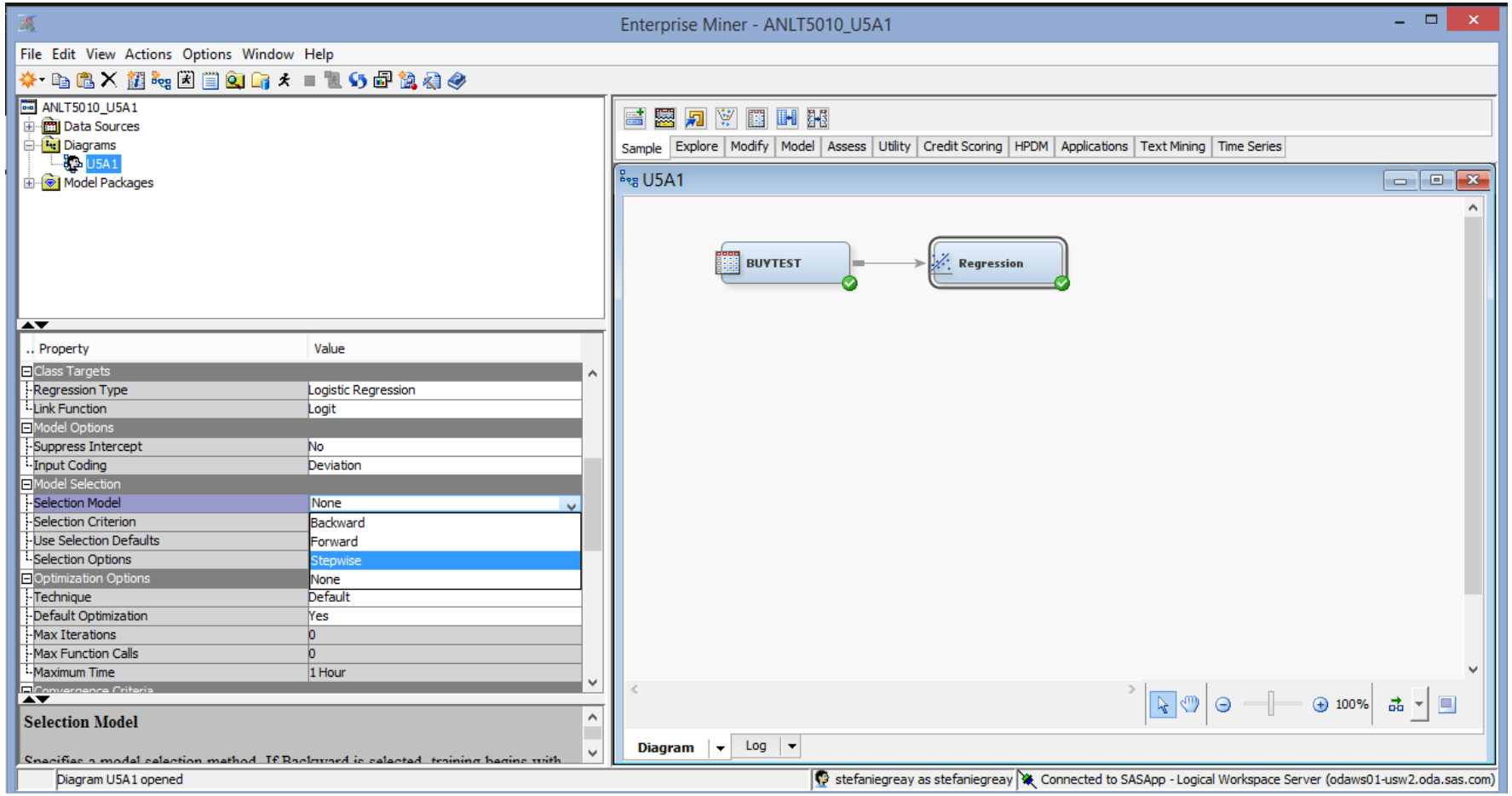
Right click on “Regression” model and click “Run” (note that the default regression type is set to logistic regression, so no change is needed there.)



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.



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=n1jqzz8cssr9m2n1ktx2iyv87q56.htm&docsetVersion=14.3&locale=en