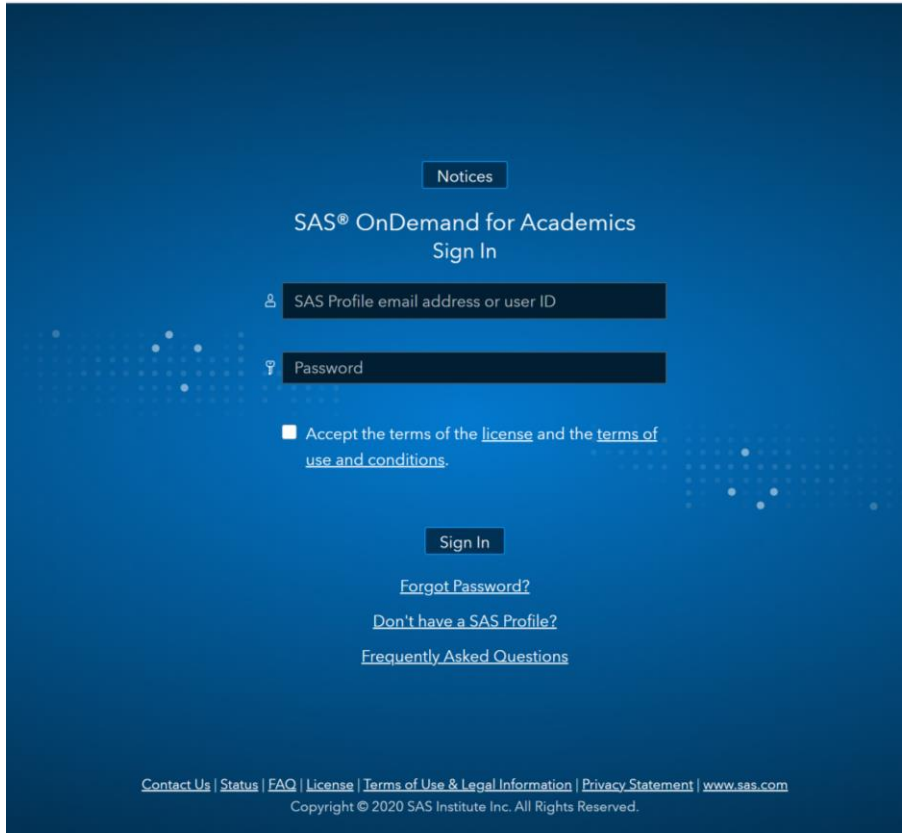
ANLT5050

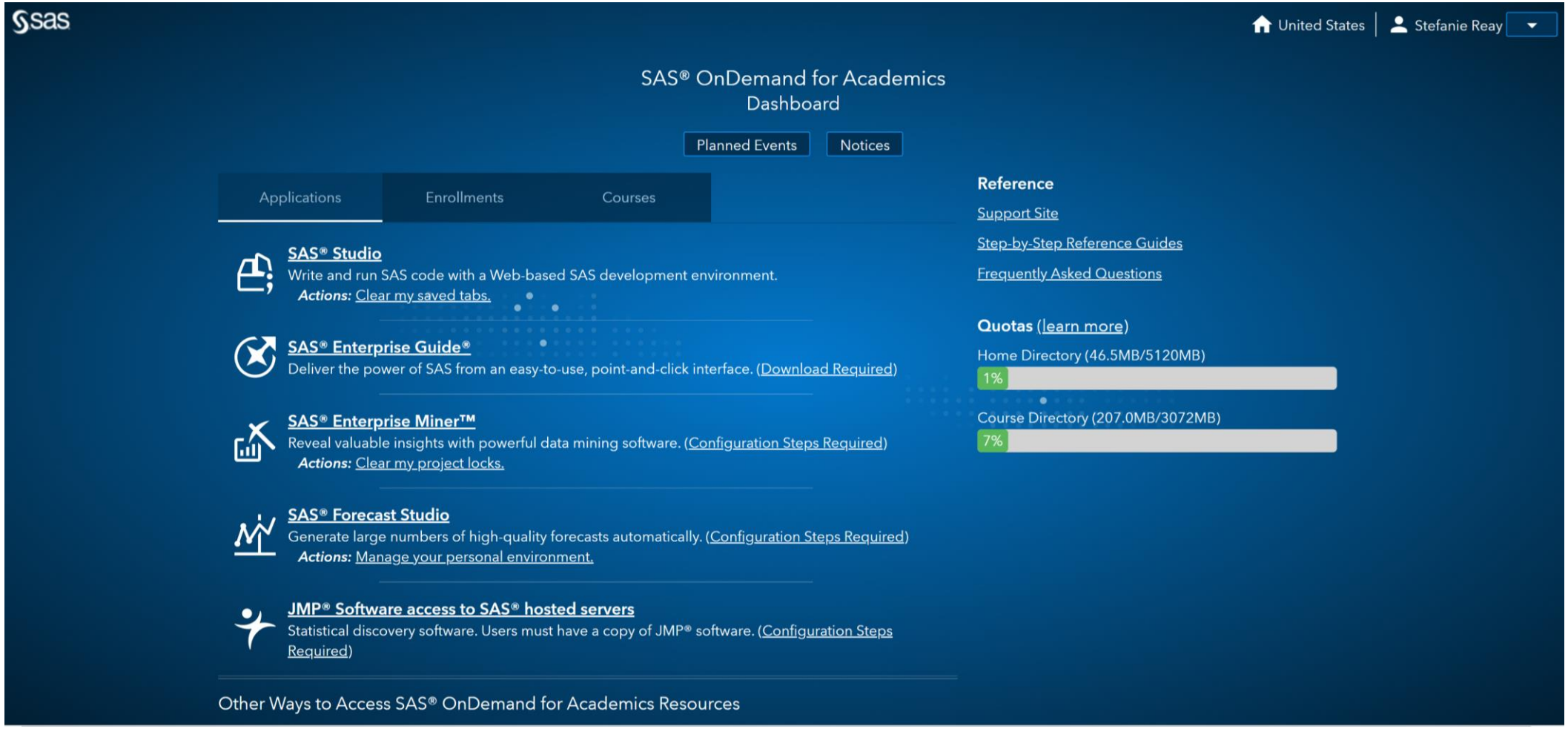
Unit 7 Assignment 2 Tutorial

Access the SAS OnDemand for Academics Control Center

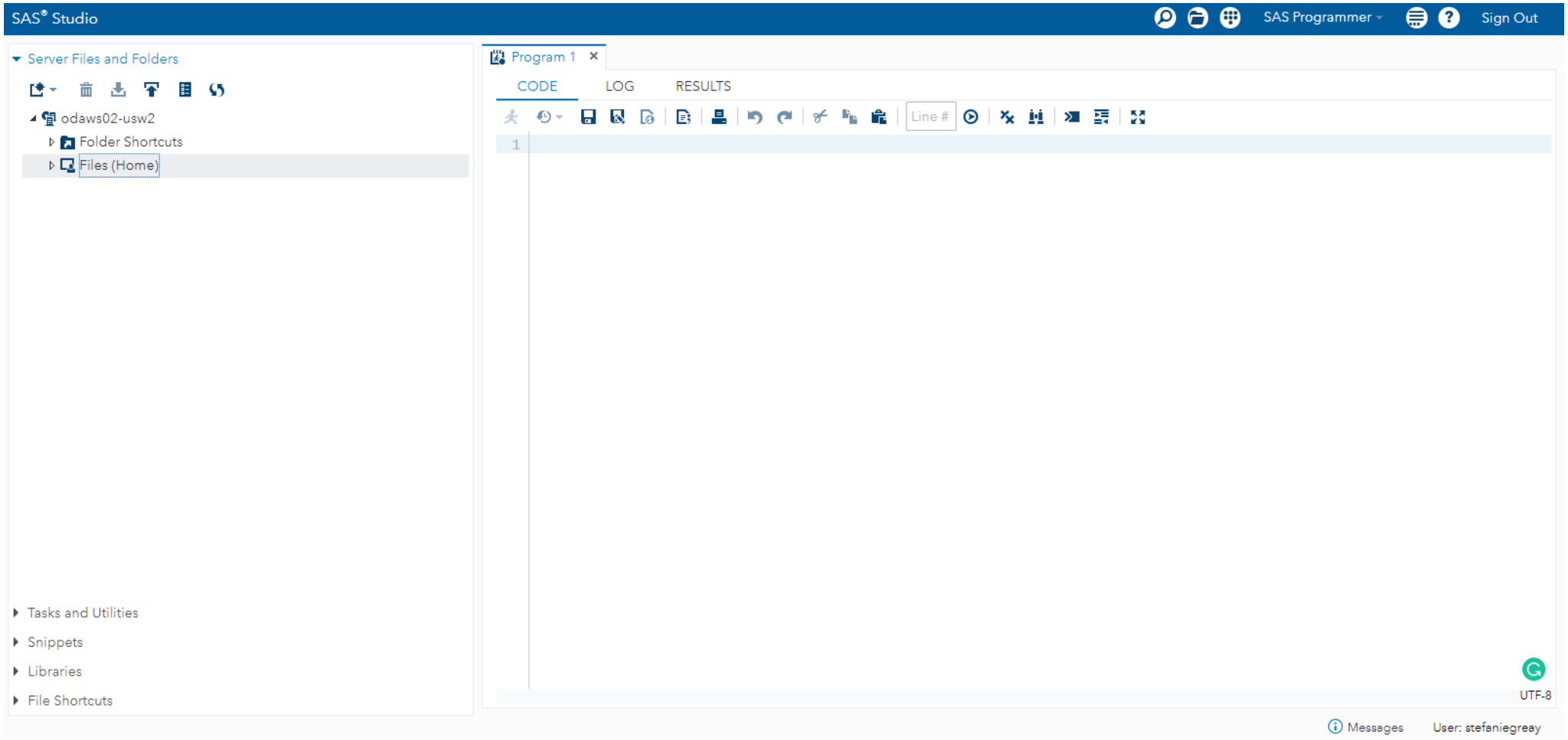
https://odamid.oda.sas.com/SASODAControlCenter



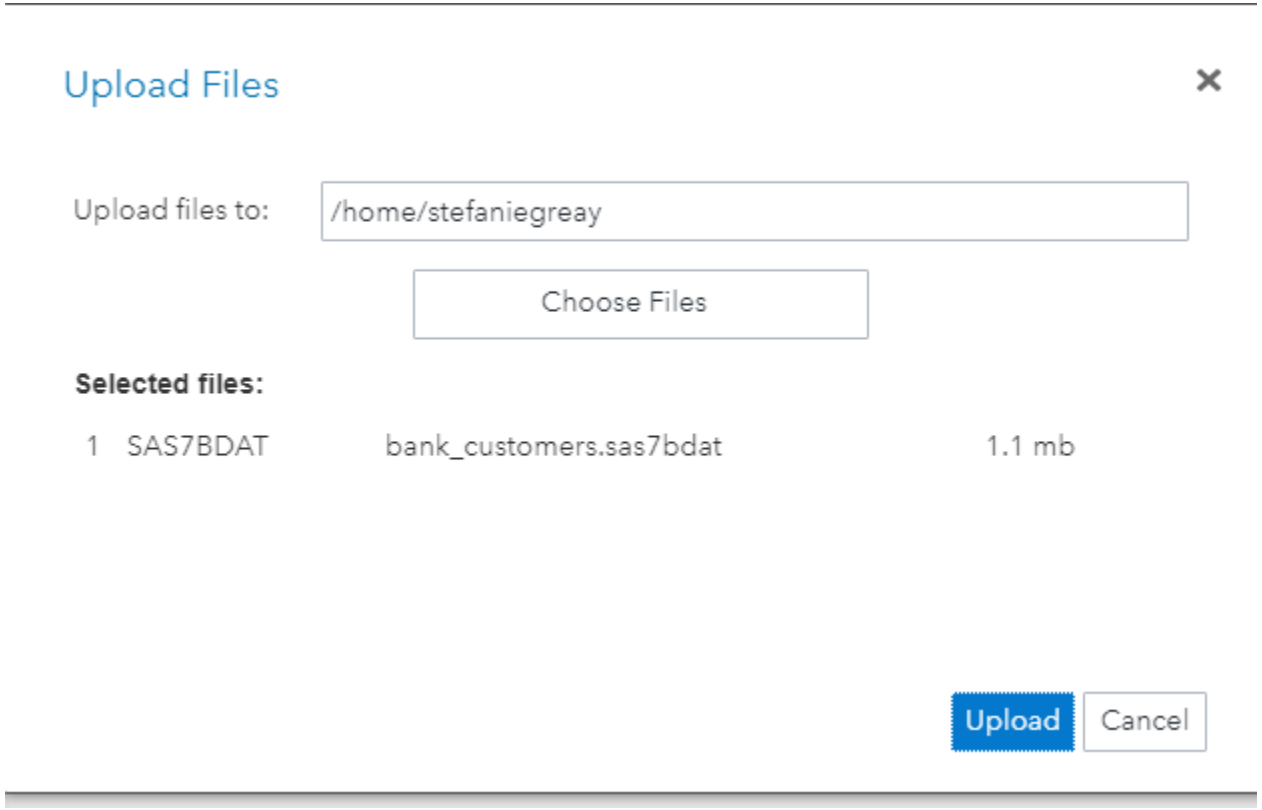
SAS OnDemand for Academics (SODA) Control Center



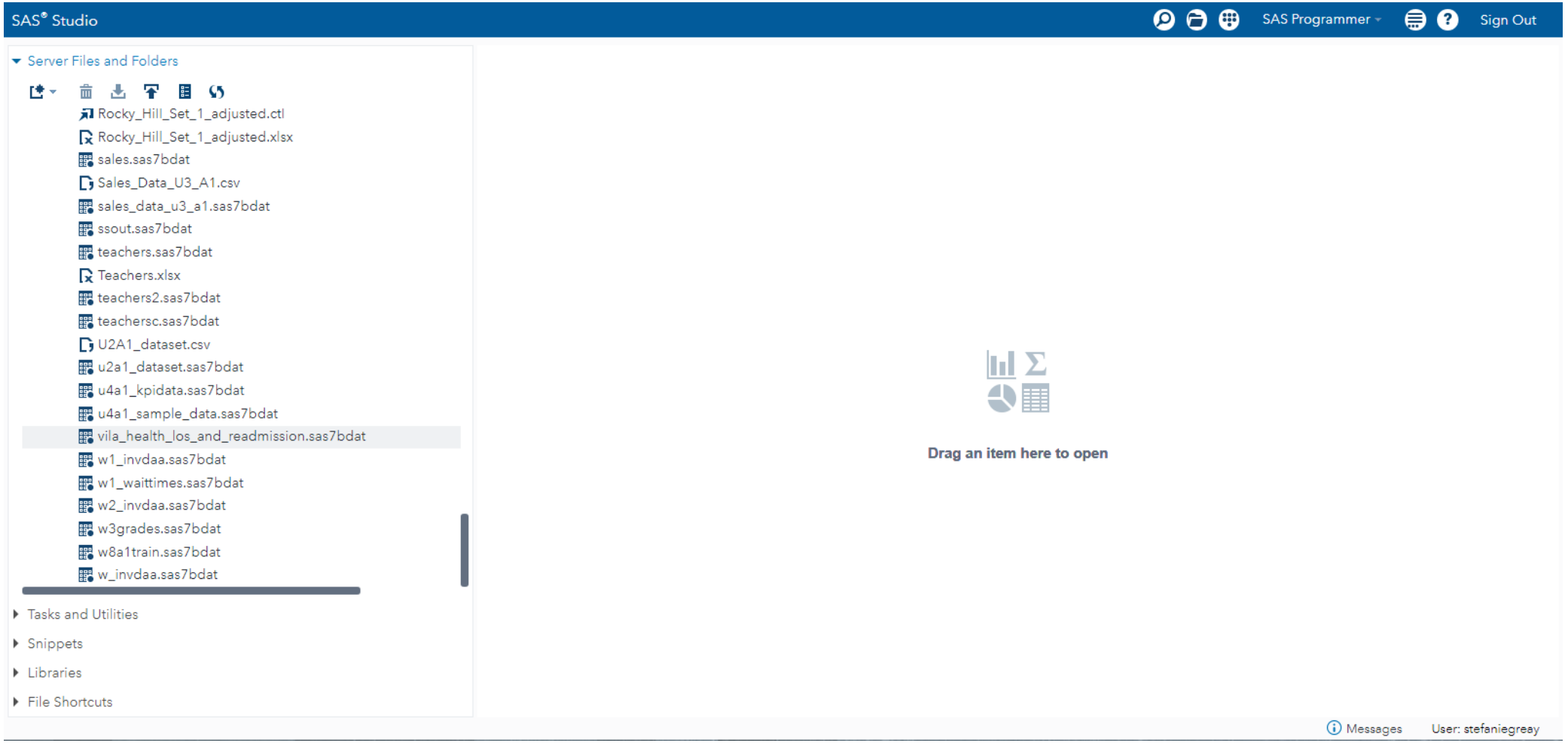
To upload the dataset to the SAS server, open SAS Studio, then click on “Files (Home)” and click the upload button.



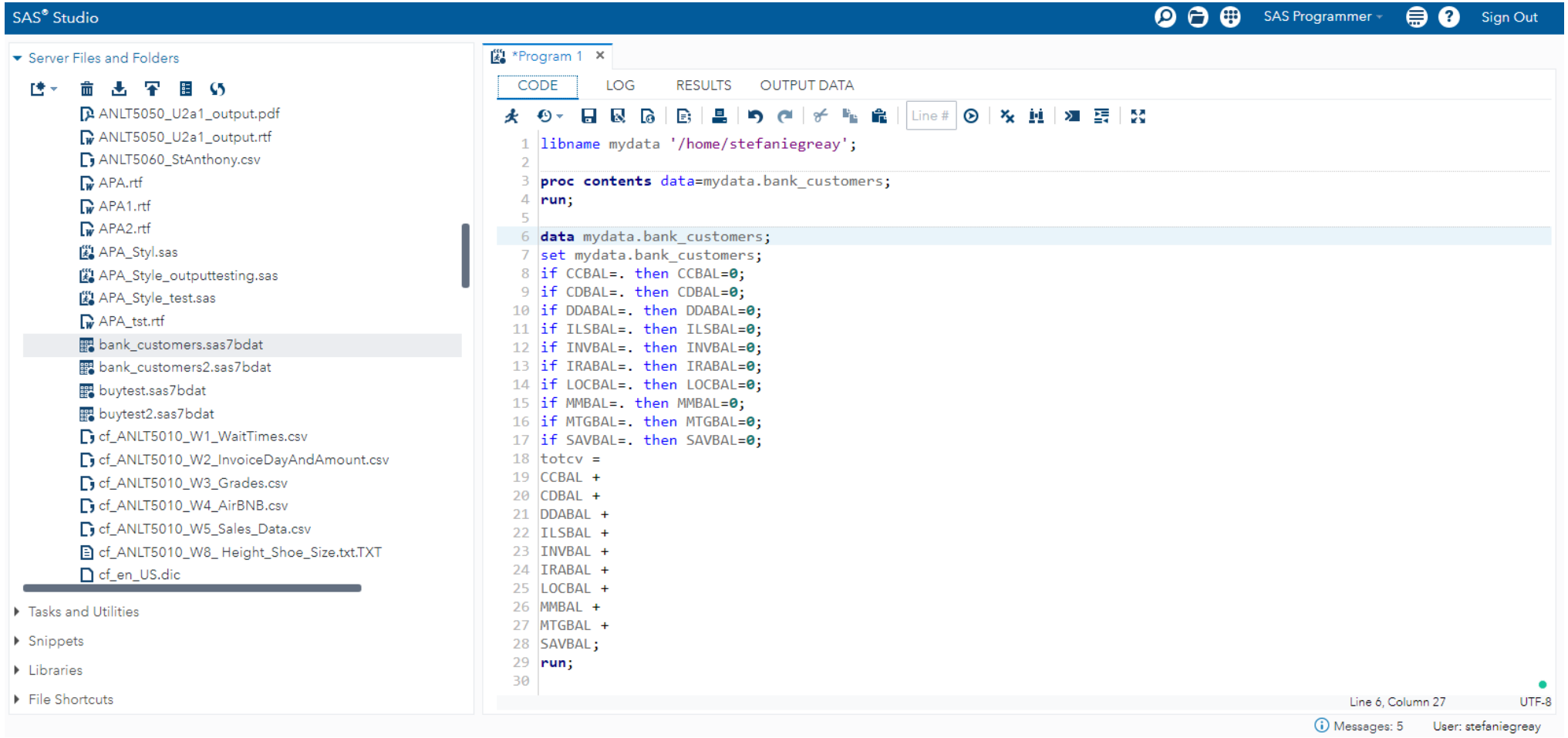
Click on “Choose Files” to browse to the file you want to upload, then click “Upload.”



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



Use the Program Editor in SAS Studio to create a total customer value variable



Sample Code

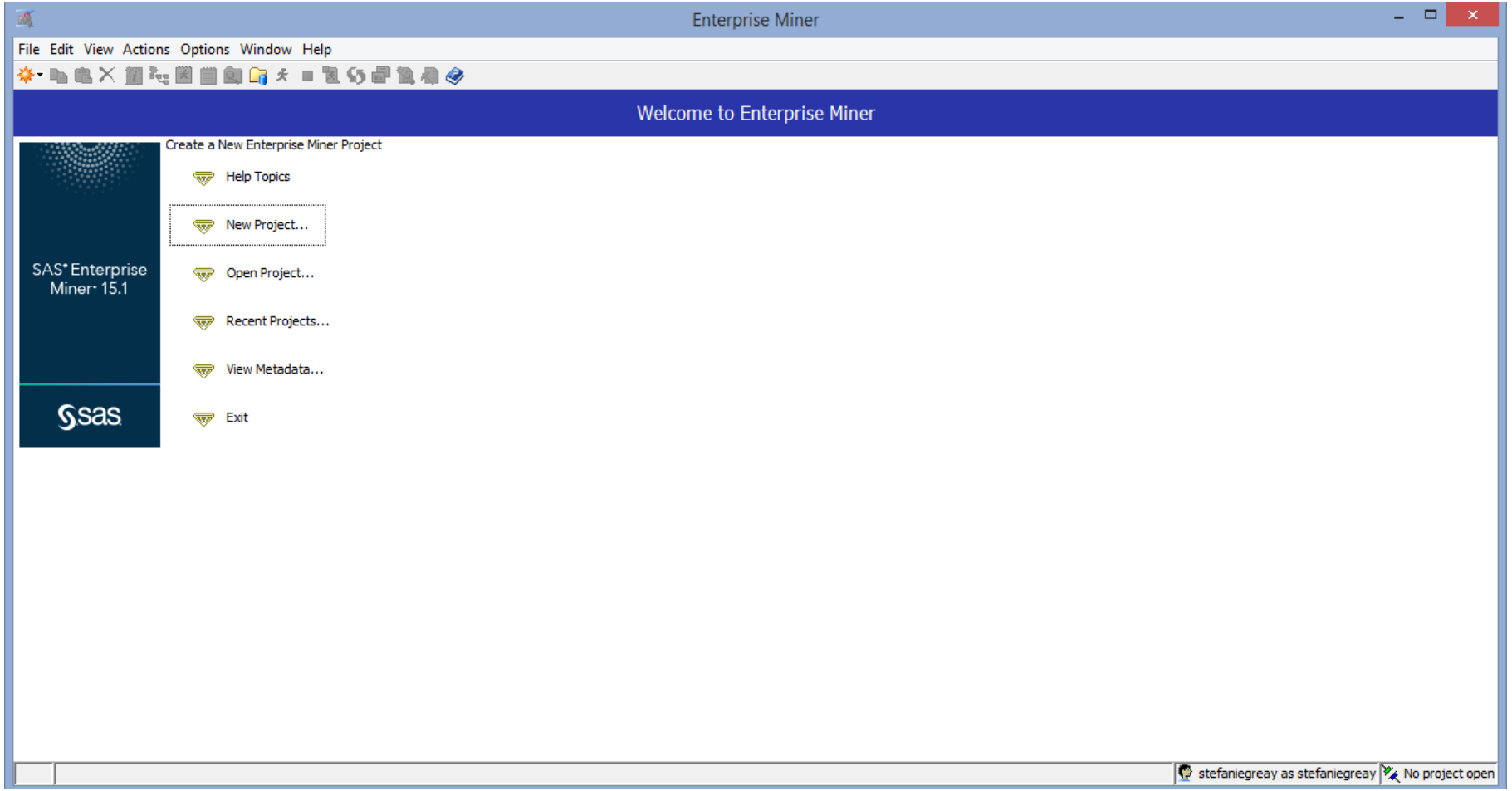
```
libname mydata '/home/stefaniegreay';

proc contents data=mydata.bank_customers;
run;

data mydata.bank_customers;
set mydata.bank_customers;
if CCBAL=. then CCBAL=0;
if CDBAL=. then CDBAL=0;
if DDABAL=. then DDABAL=0;
if ILSBAL=. then ILSBAL=0;
if INVBAL=. then INVBAL=0;
if IRABAL=. then IRABAL=0;
if LOCBAL=. then LOCBAL=0;
if MMBAL=. then MMBAL=0;
if MTGBAL=. then MTGBAL=0;
if SAVBAL=. then SAVBAL=0;
totcv =  CCBAL + CDBAL + DDABAL + ILSBAL + INVBAL + IRABAL+ LOCBAL + MMBAL + MTGBAL + SAVBAL;
run;
```

SAS Enterprise Miner Instructions

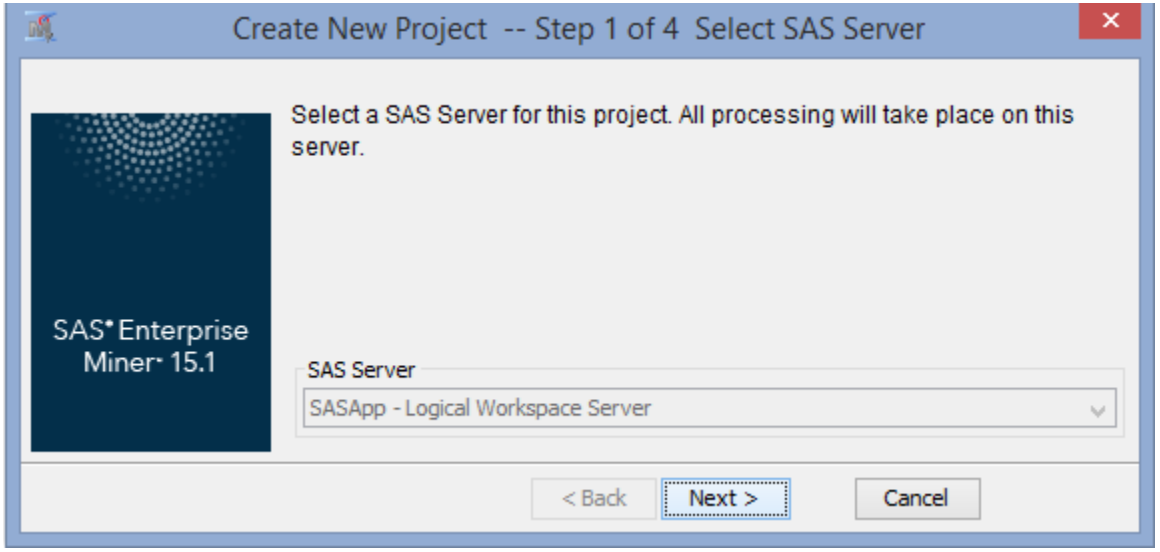
The following slides provide instructions on how to complete this task in SAS Enterprise Miner.

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.

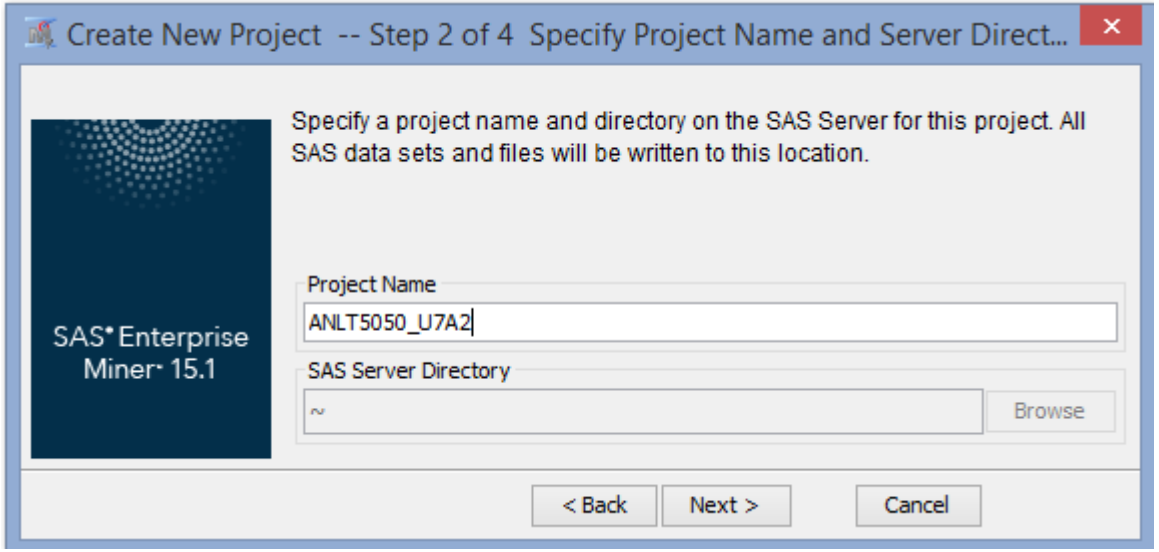
Once you download and start SAS Enterprise Miner, open a new project by clicking on “New Project.”



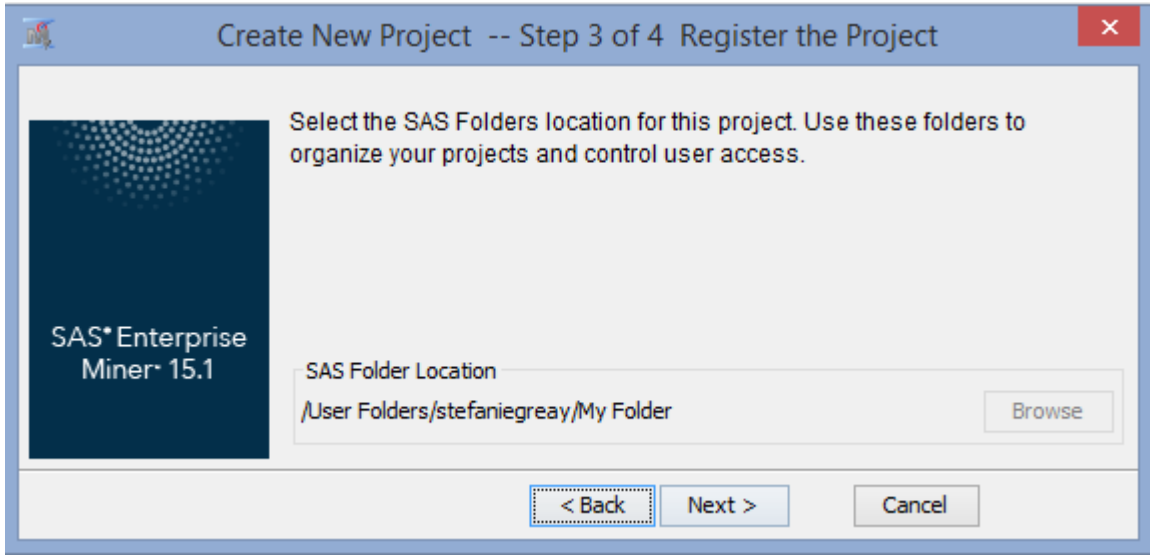
Click “Next>” to use the default SAS Server



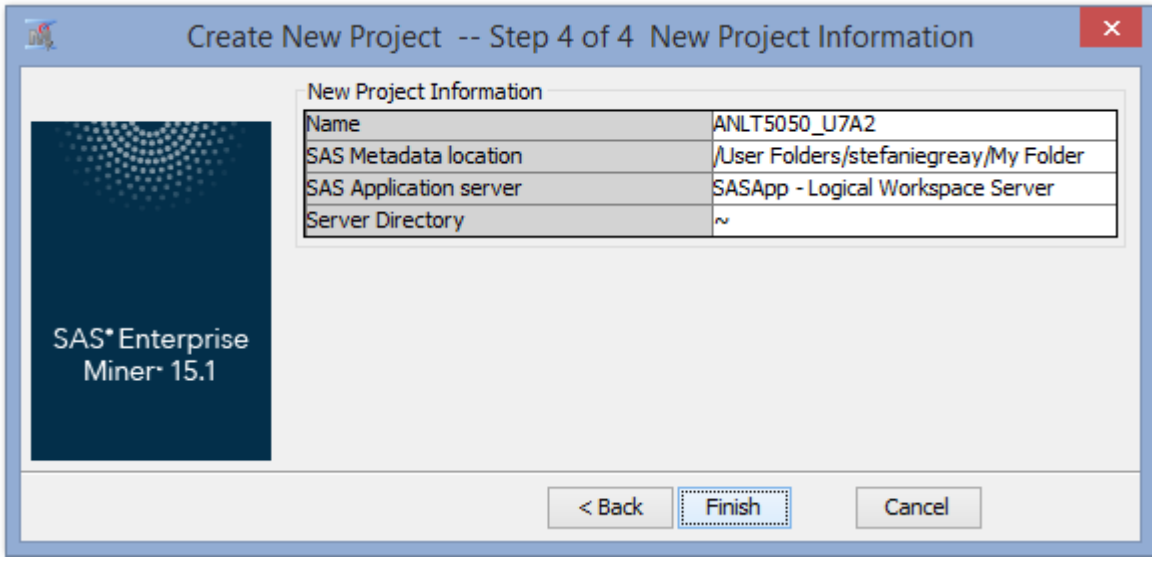
Enter a project name and click “Next>”



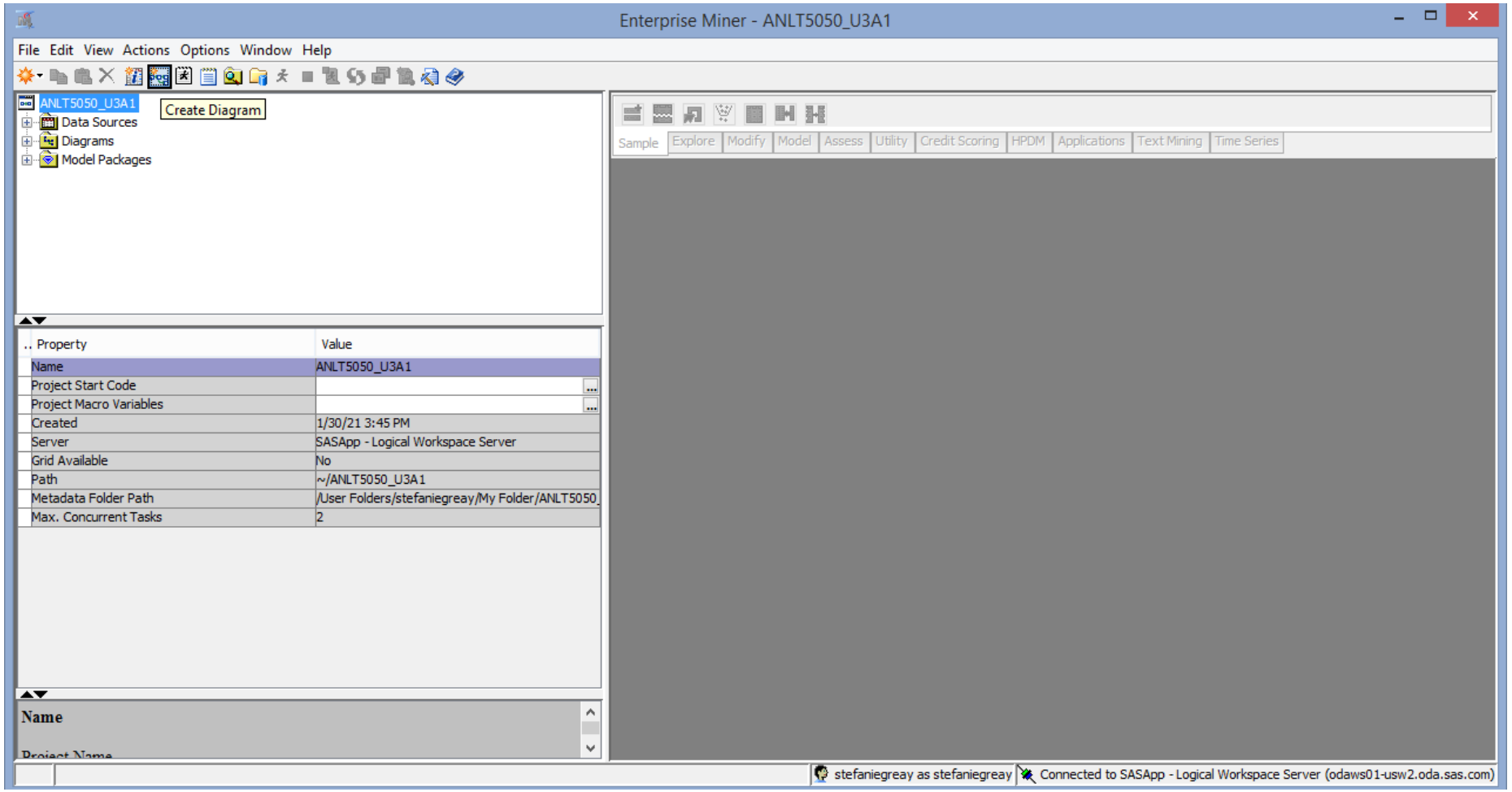
Click “Next>”



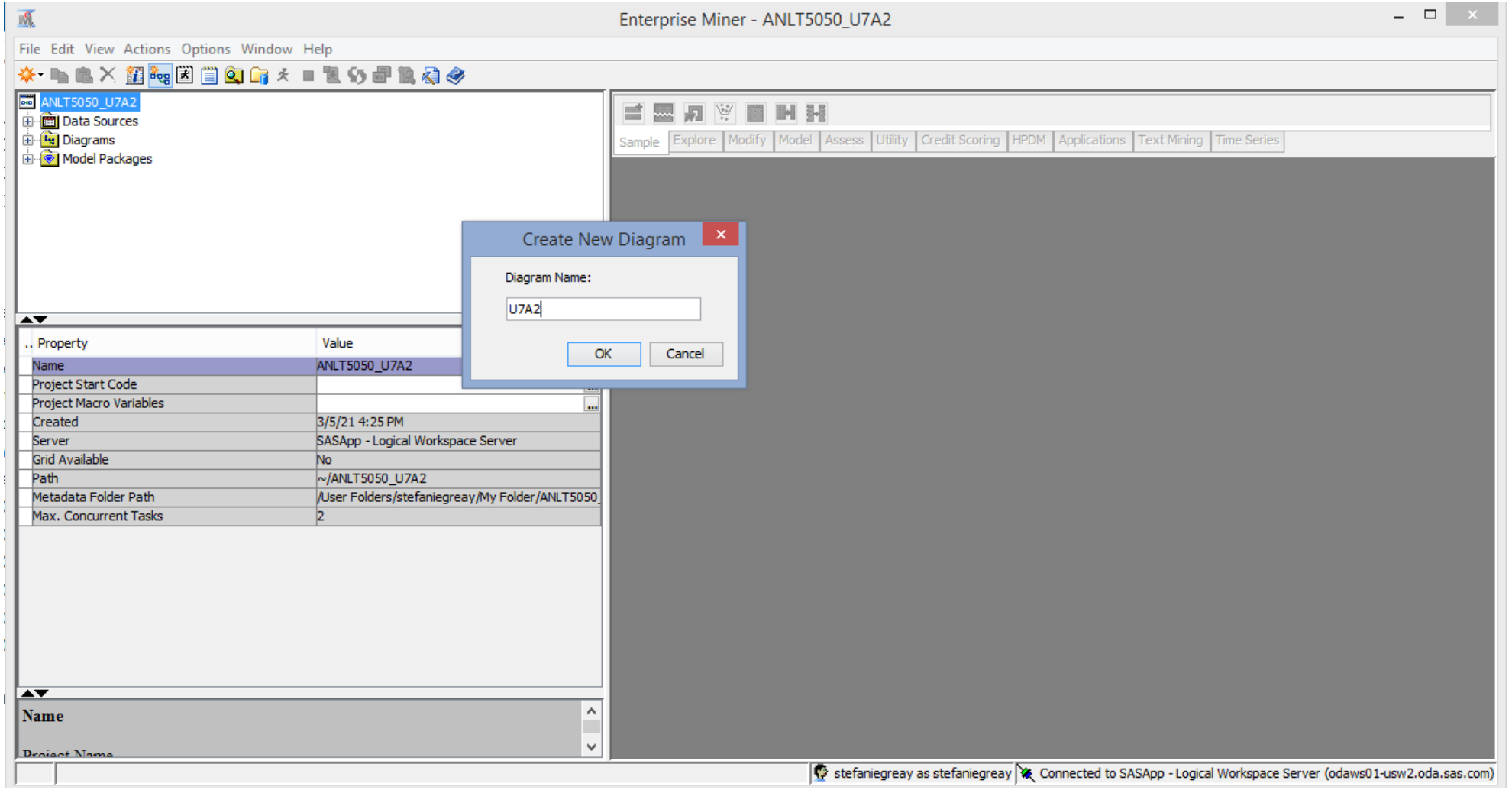
Verify your entries and click “Finish”



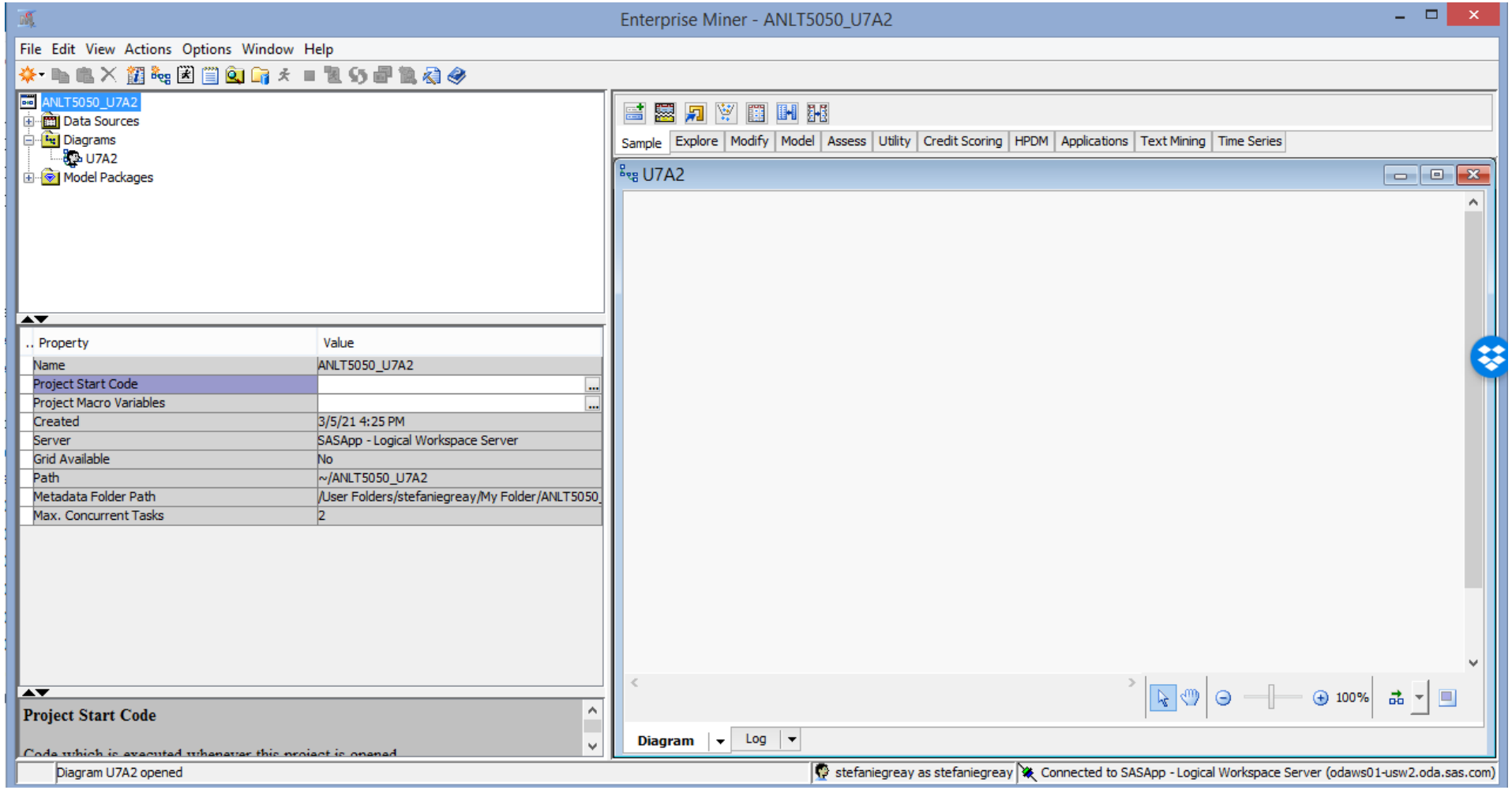
Click on the “Create Diagram” icon.



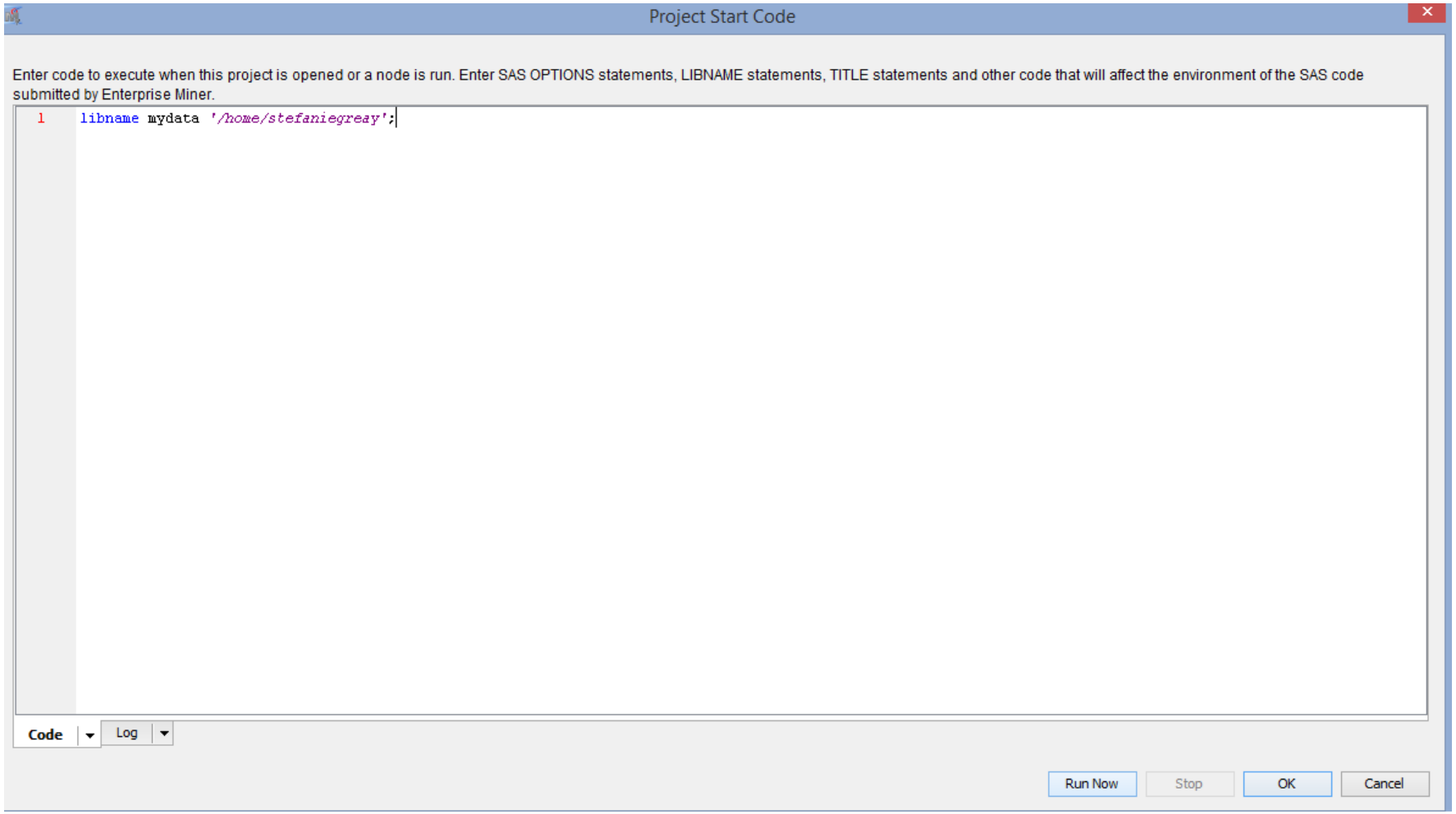
Name your diagram and click “OK.”



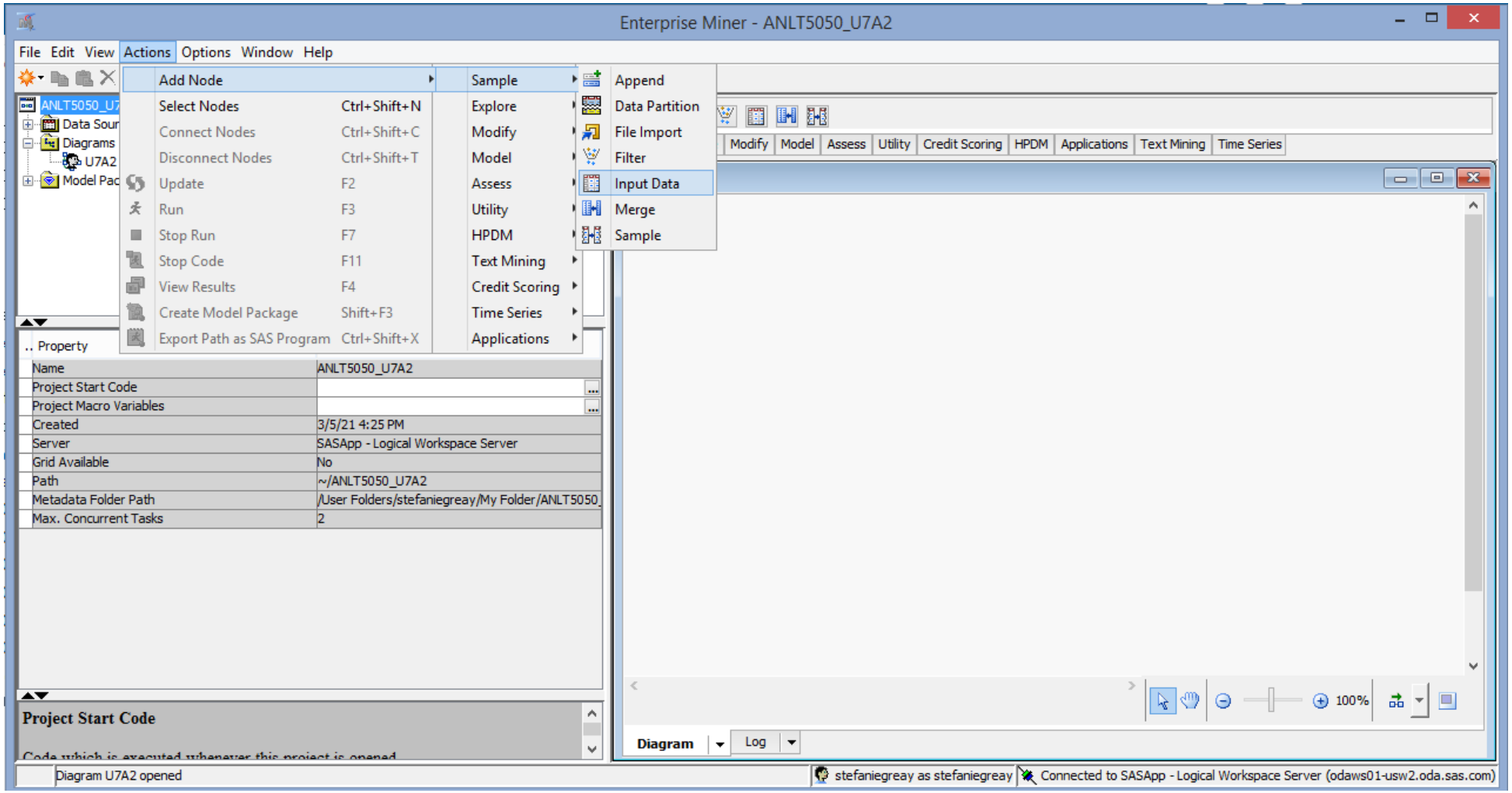
Click on the project, then click on the ellipses next to “Project Start Code.”



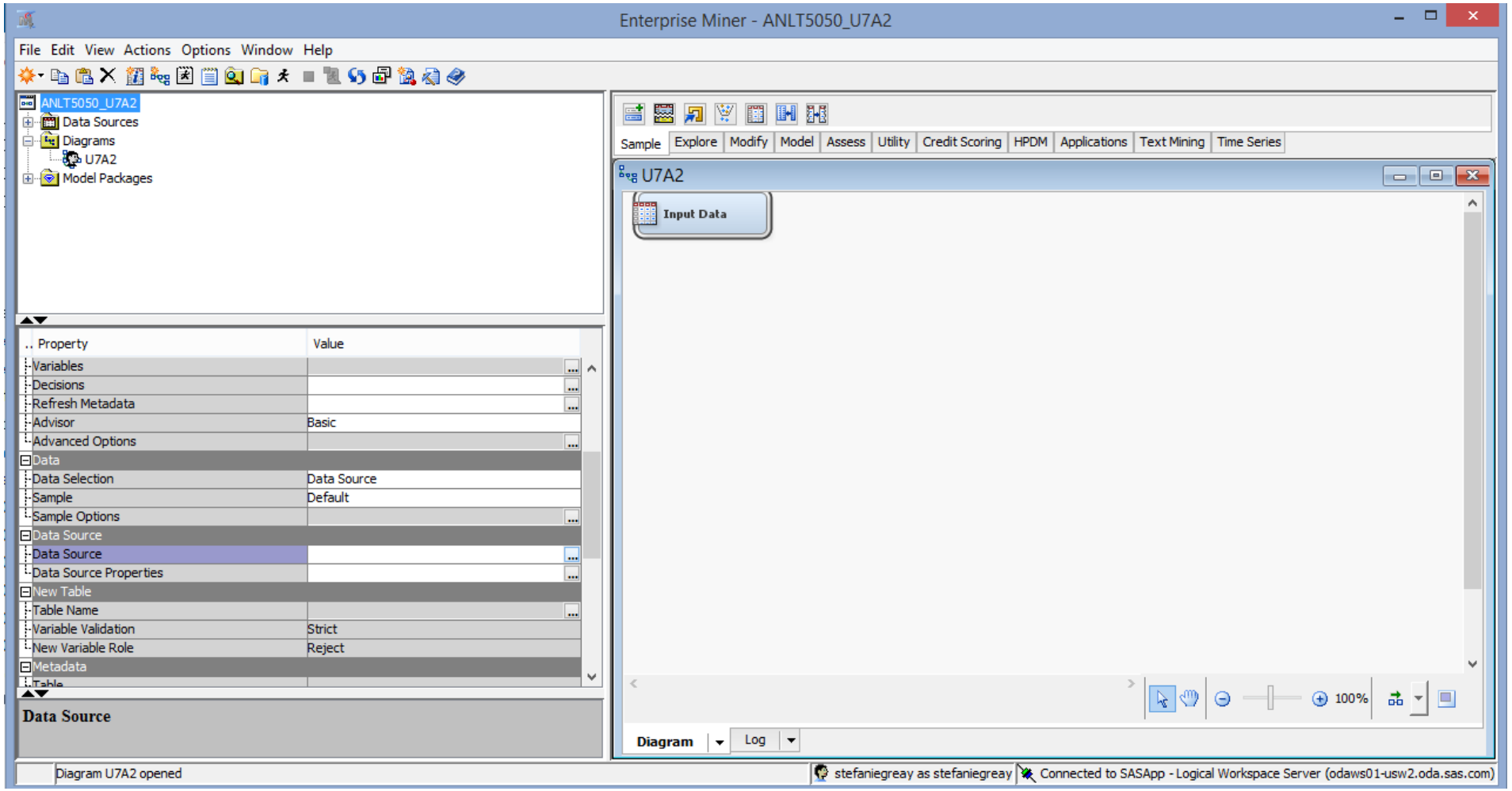
Add the library reference for where you uploaded the dataset in SAS studio, and click “Run Now.” Once it completes, click “OK.”



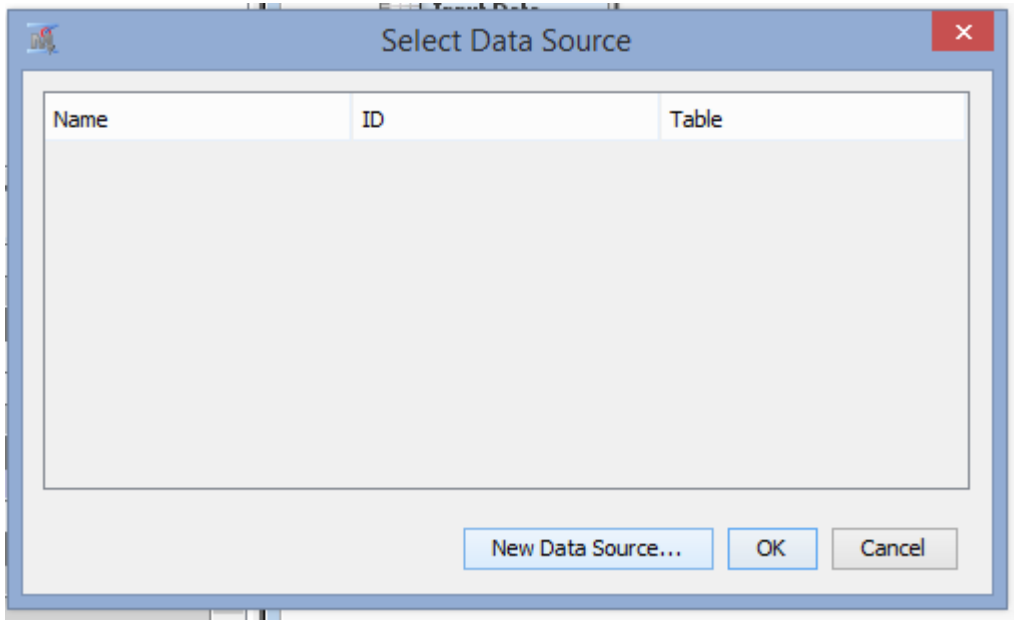
Click on Actions>Add Node>Sample>Input Data



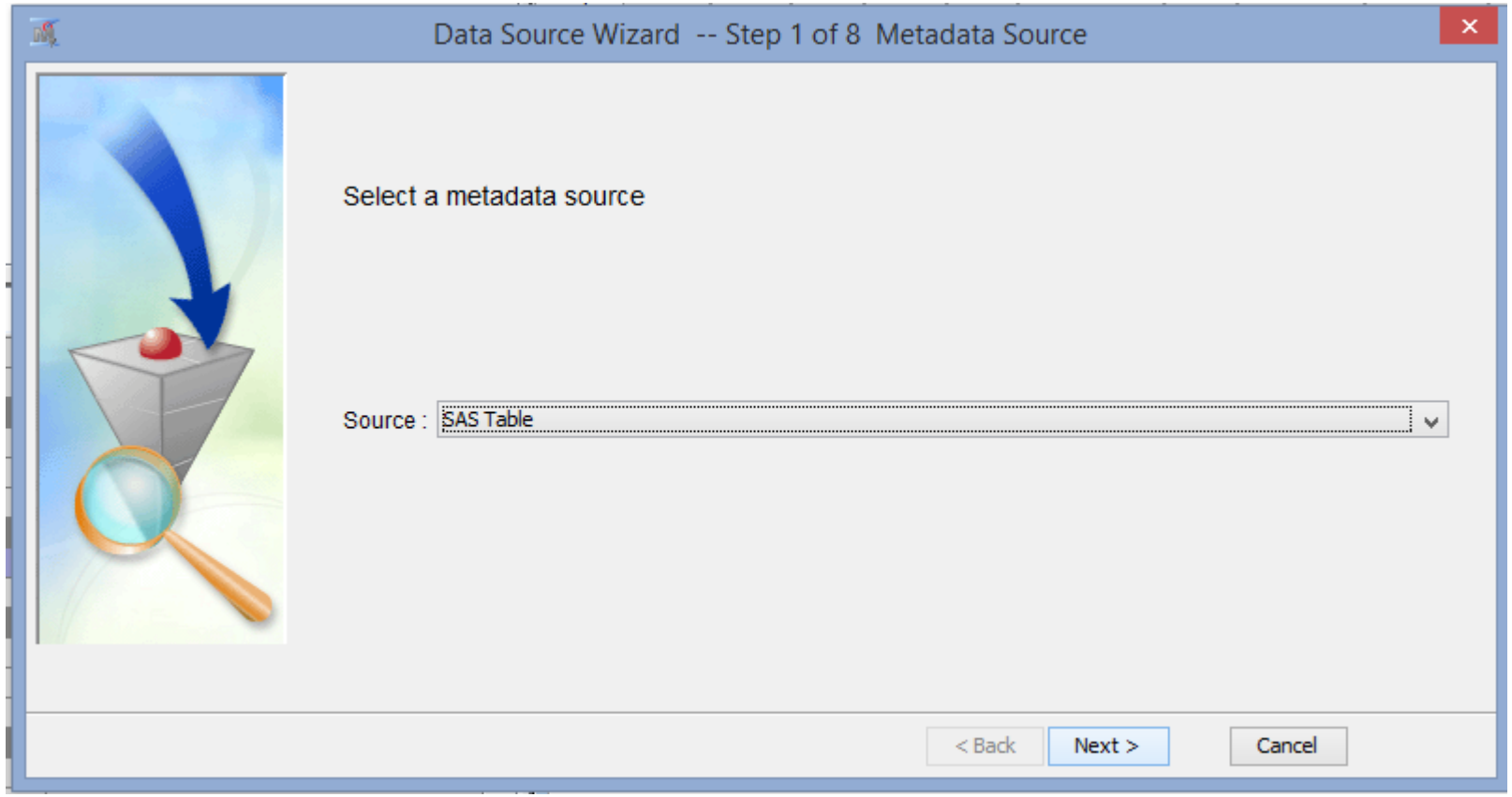
Click the ellipses (3 dots) next to “Data Source.”



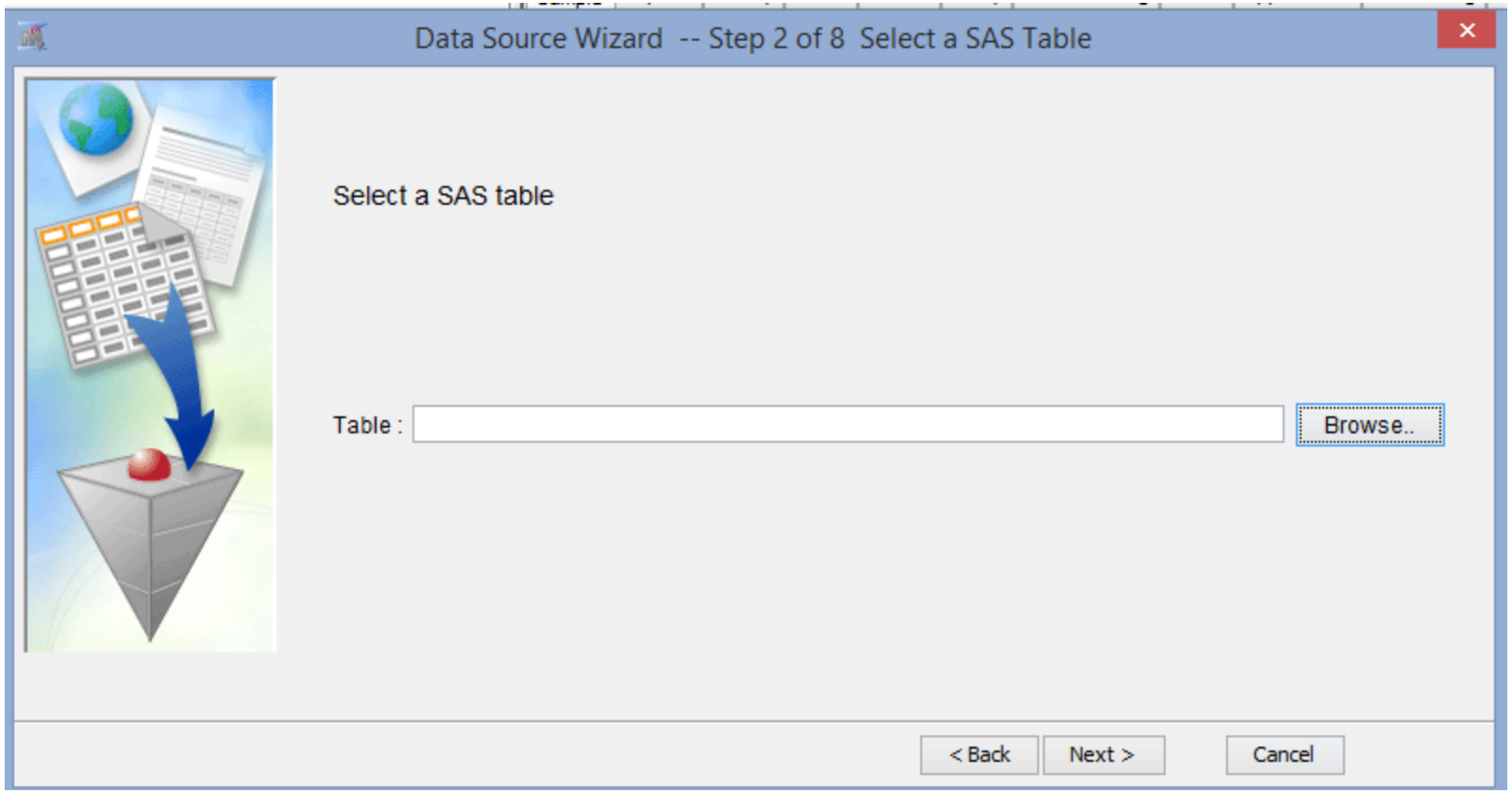
Click on “New Data Source”



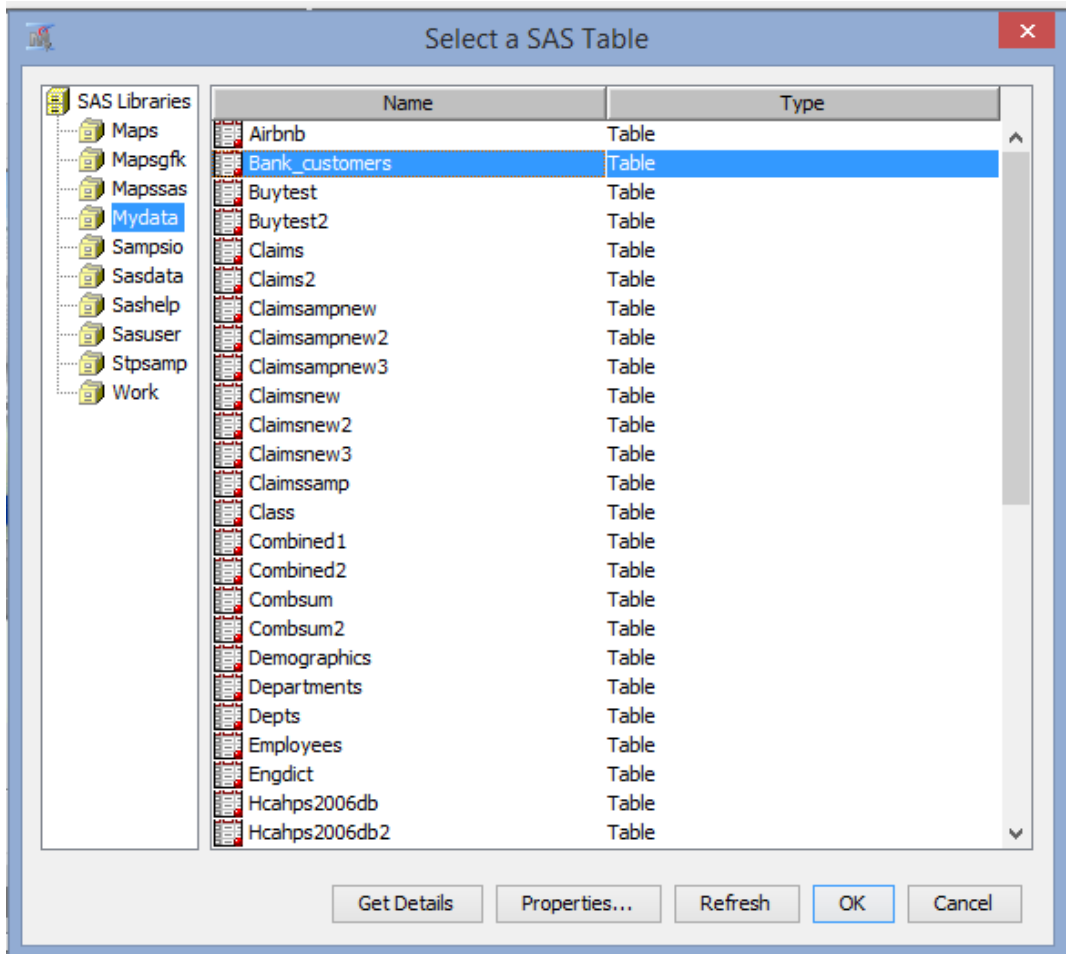
Leave it as “SAS Table” and click “Next >”



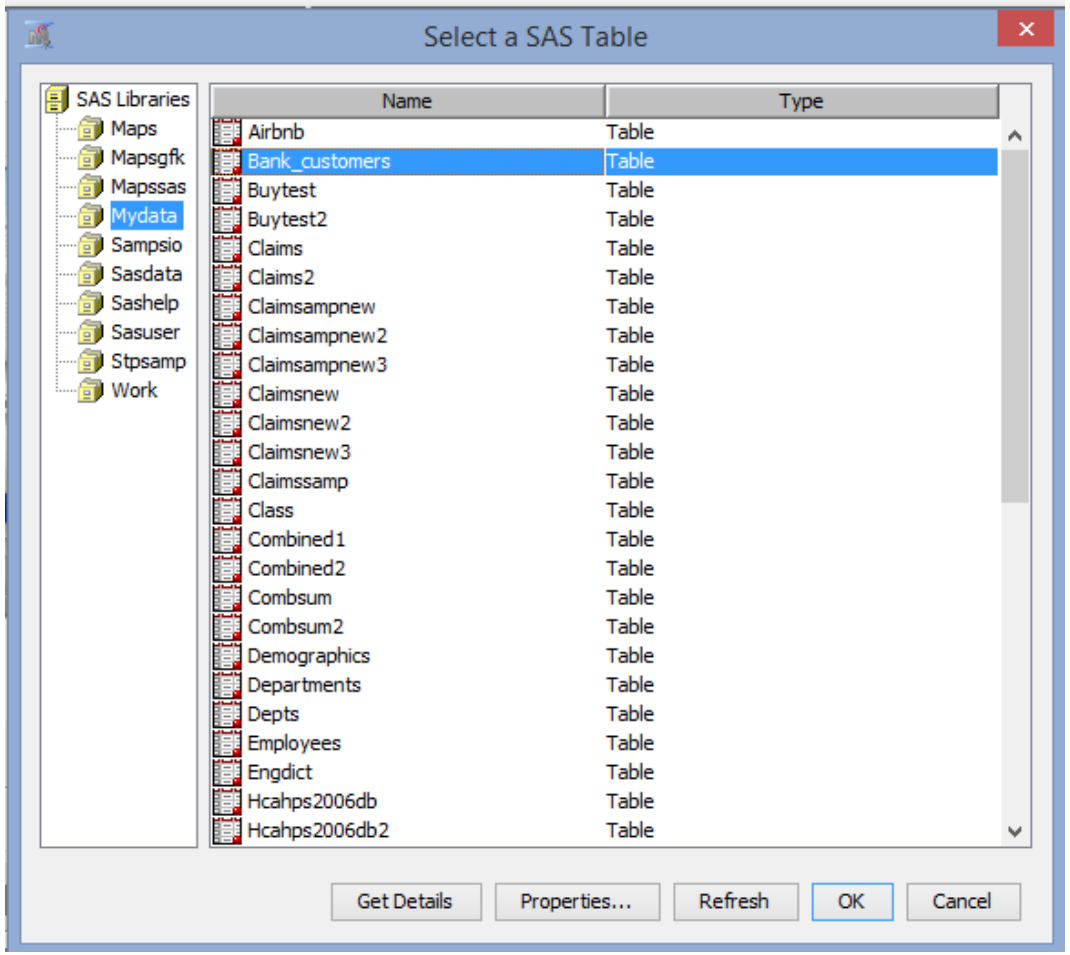
Click on “Browse”



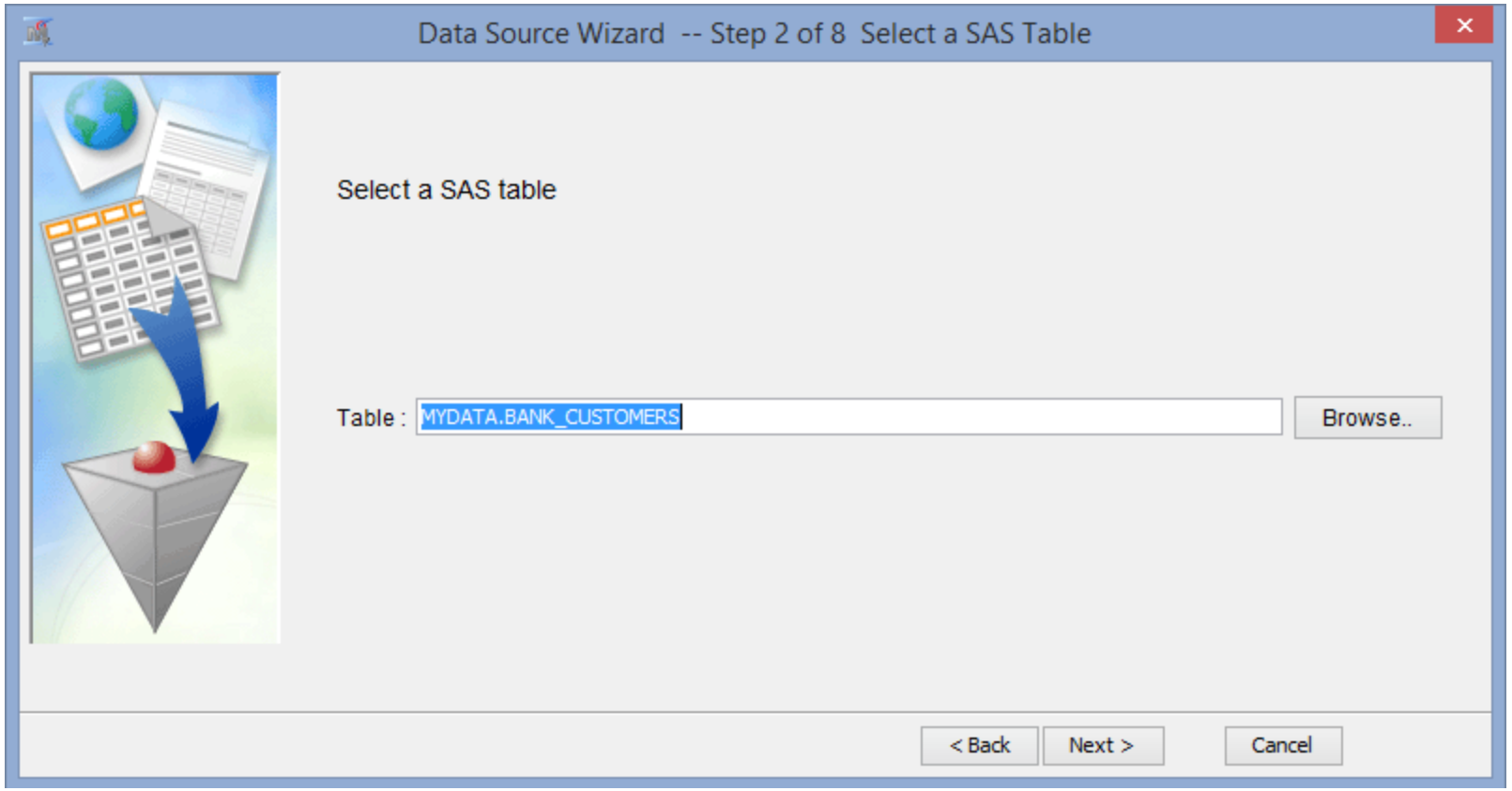
Double click on the libname you just set up in the project startup code.



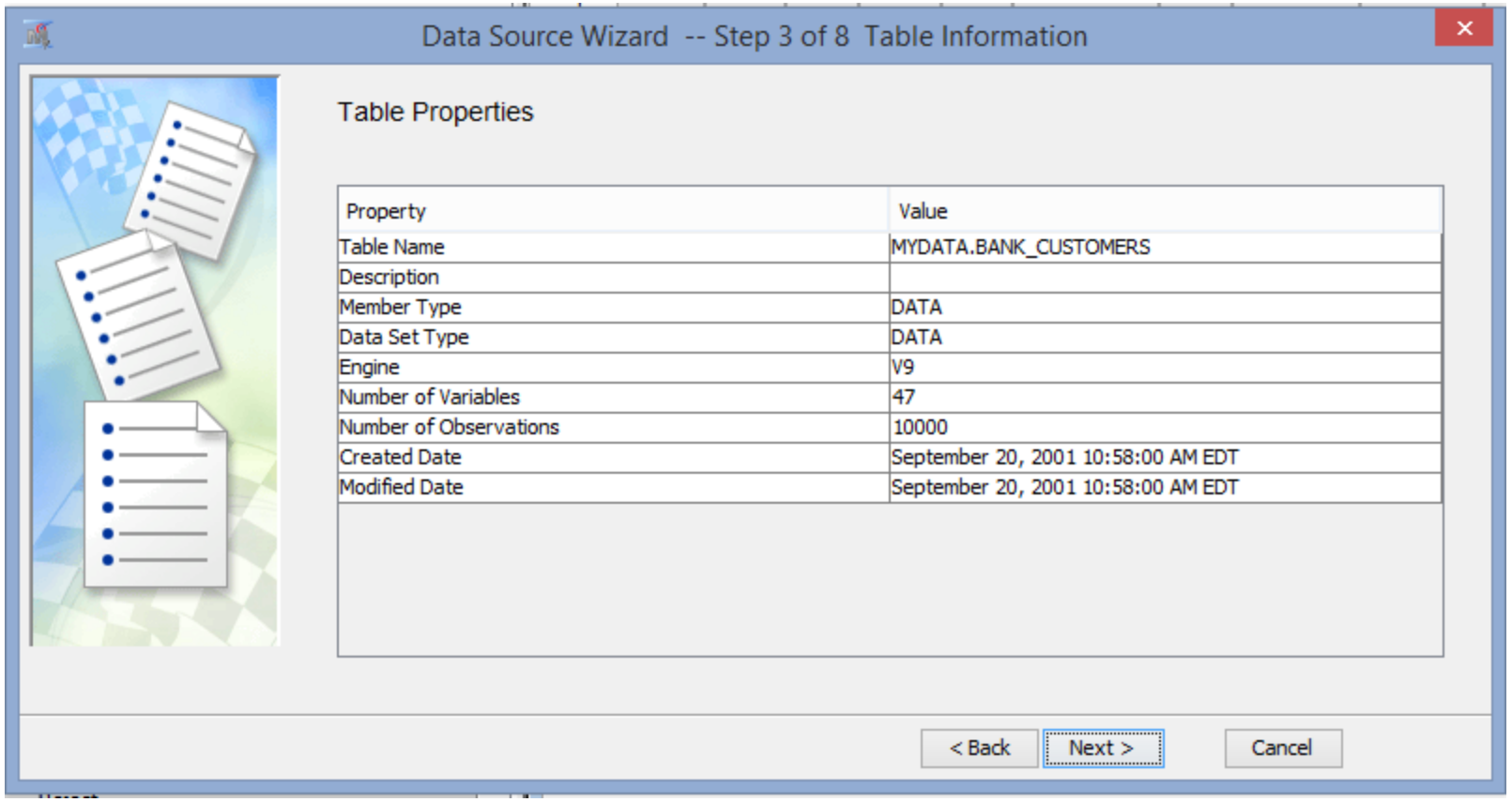
Double click to select the dataset for this unit, and click “OK”



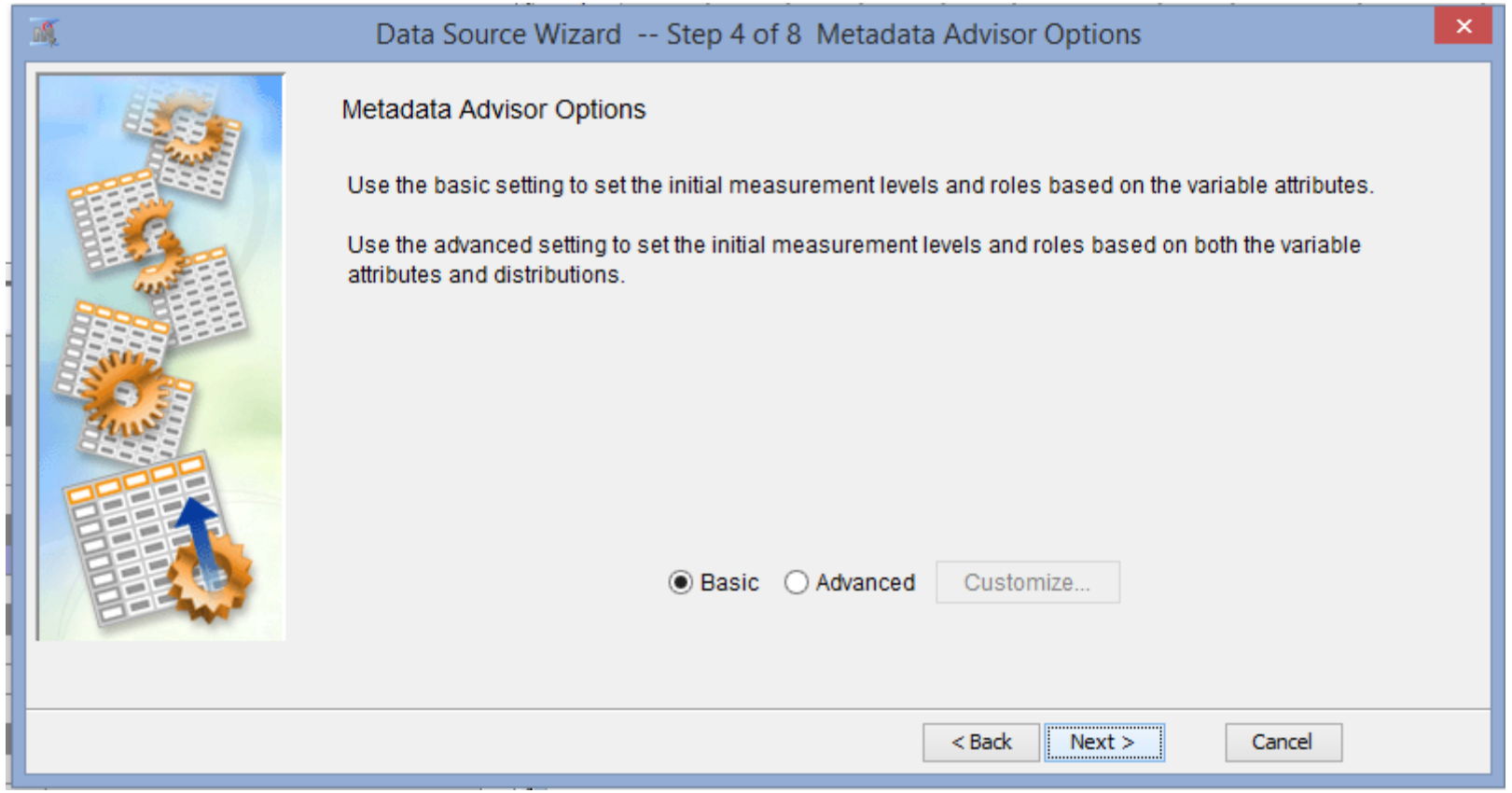
Click “Next>”



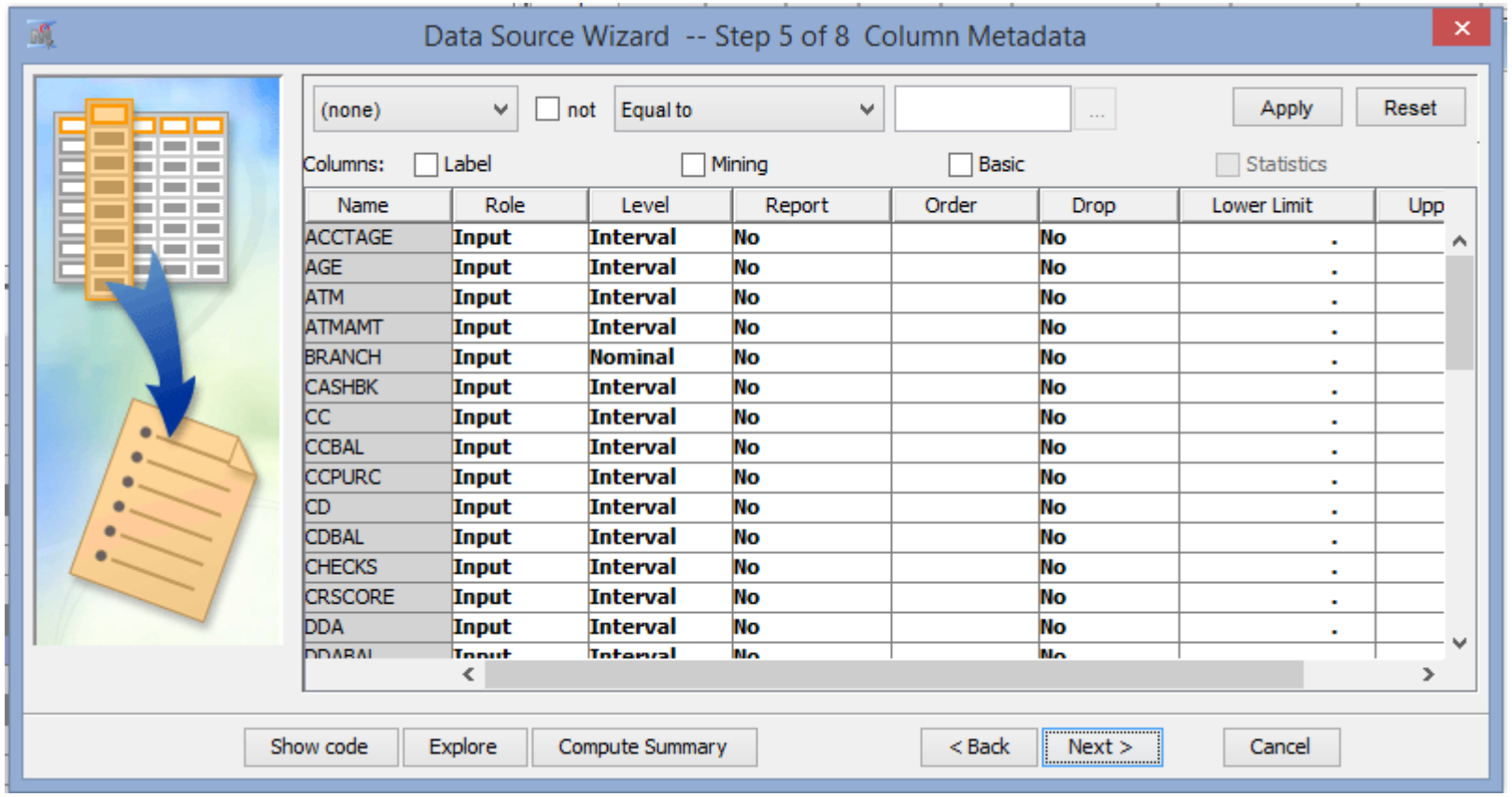
Verify the options and click “Next>”



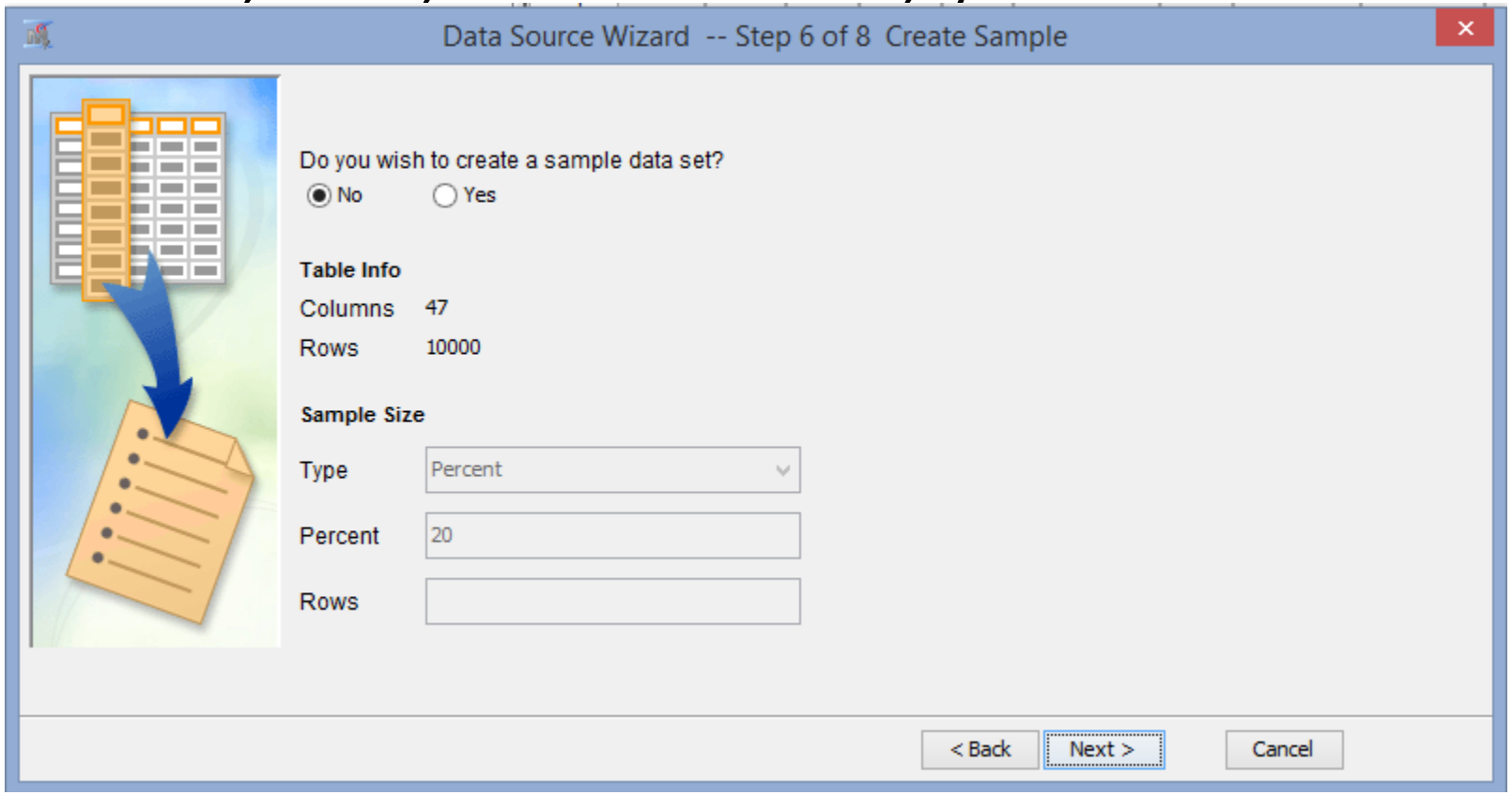
Click “Next>”



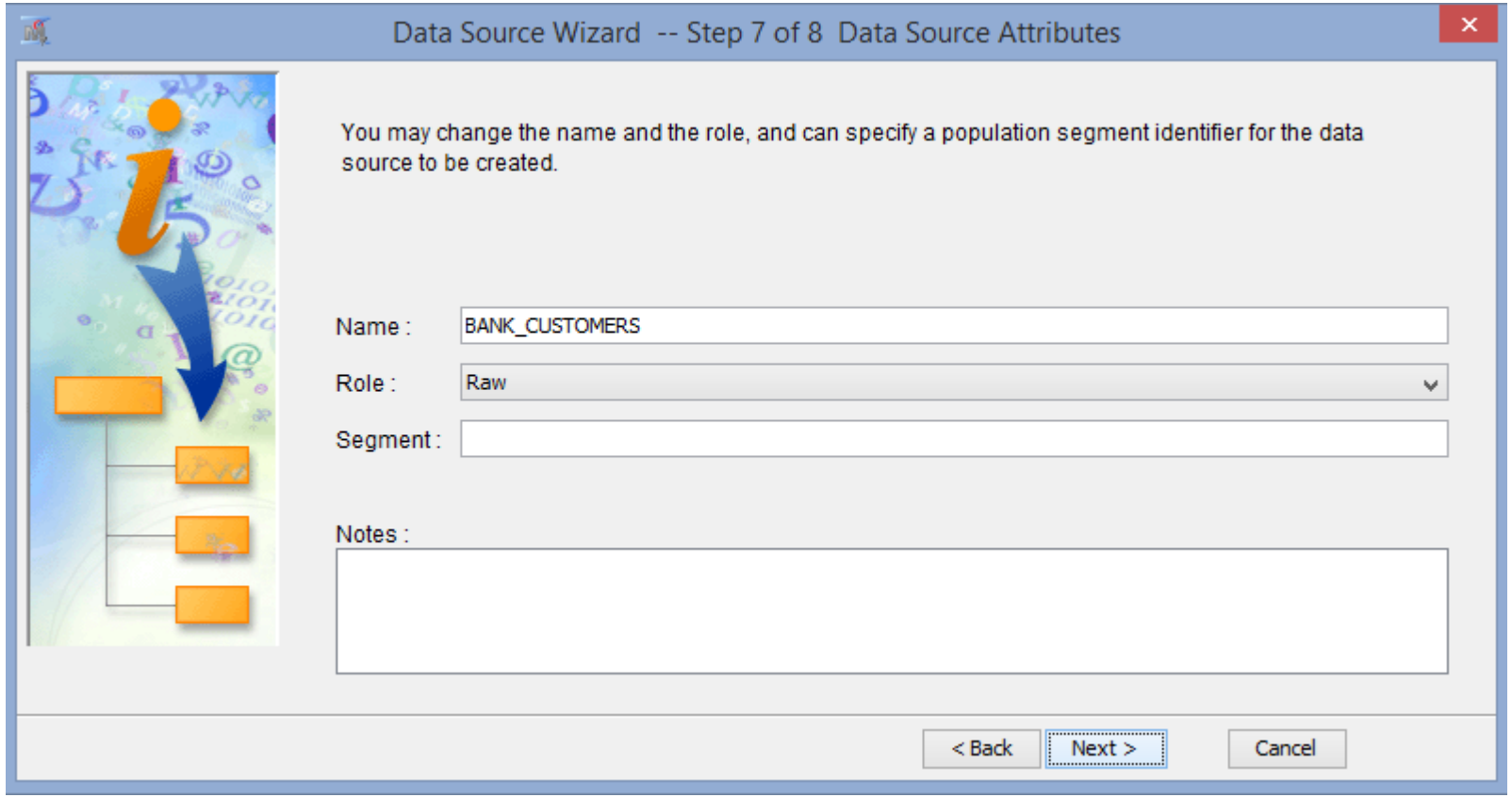
Verify the variables and settings, adjust if necessary, and then click “Next>”



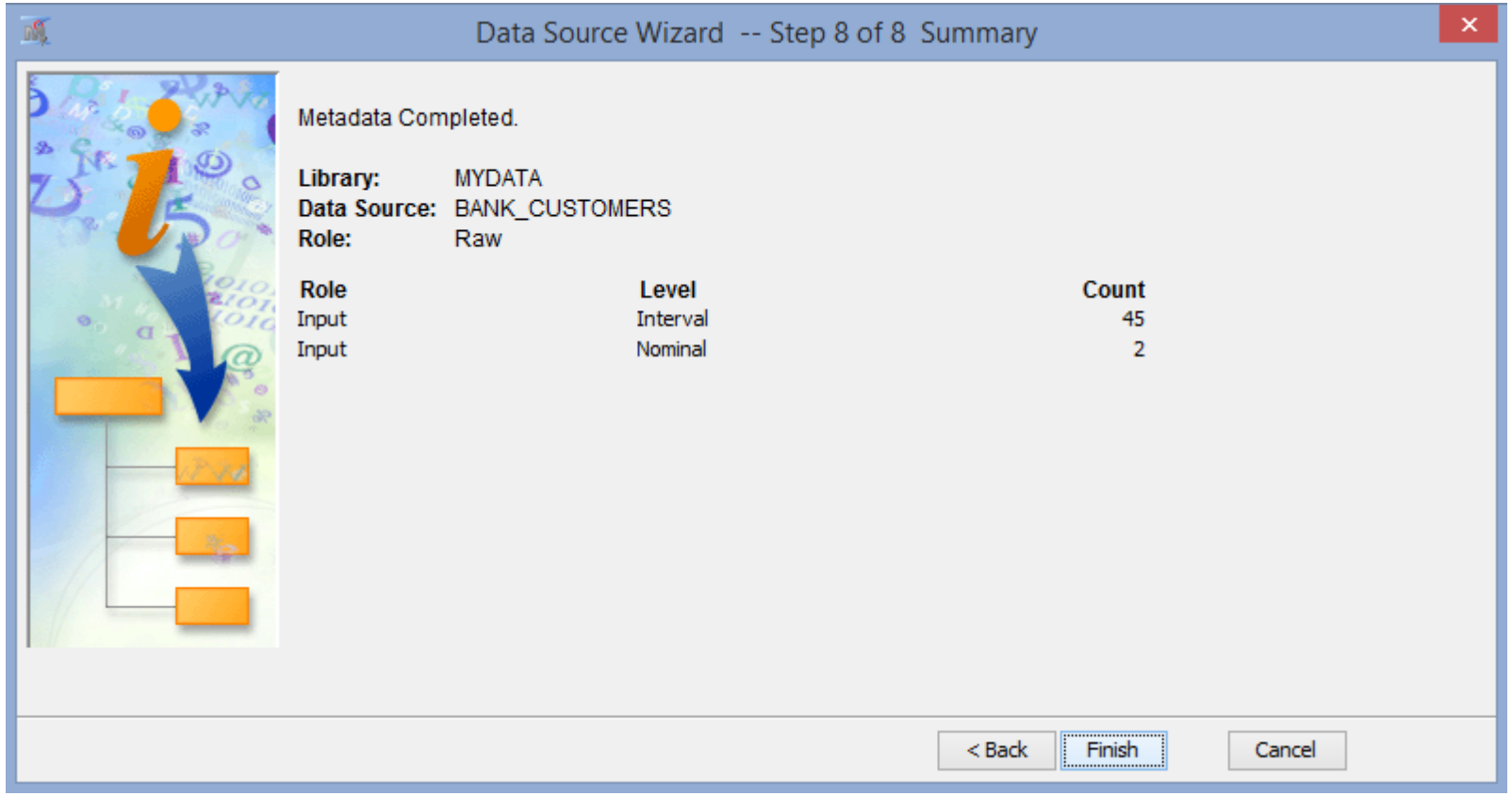
You may choose to sample the dataset here, or just keep the full dataset, then click “Next>.” If you want to split into train, test, and validate, you could do this here.



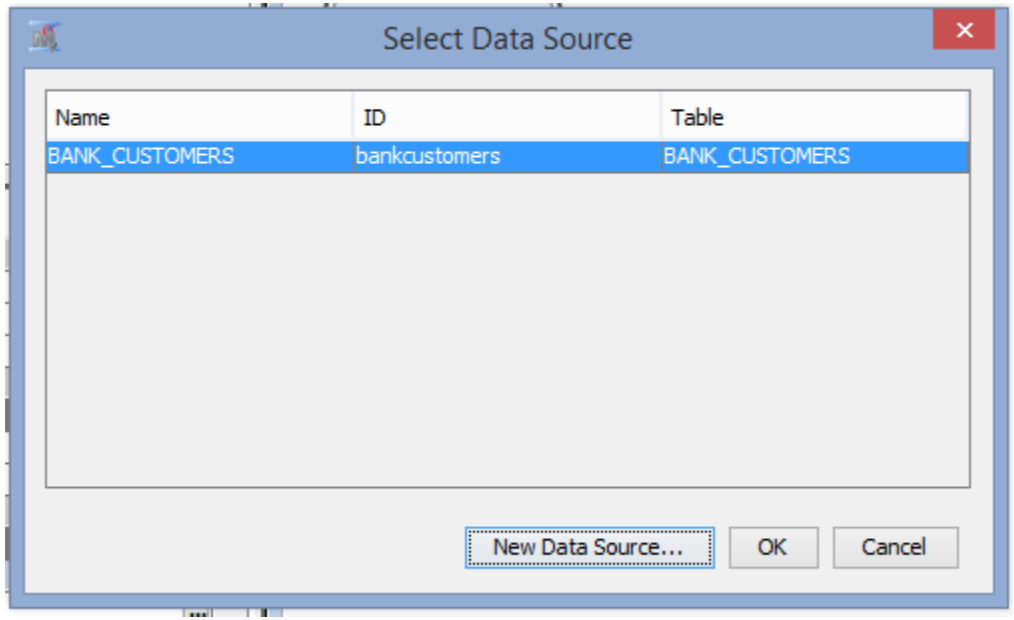
You may choose to adjust the role of the dataset, or leave it as the default, then click “Next>”



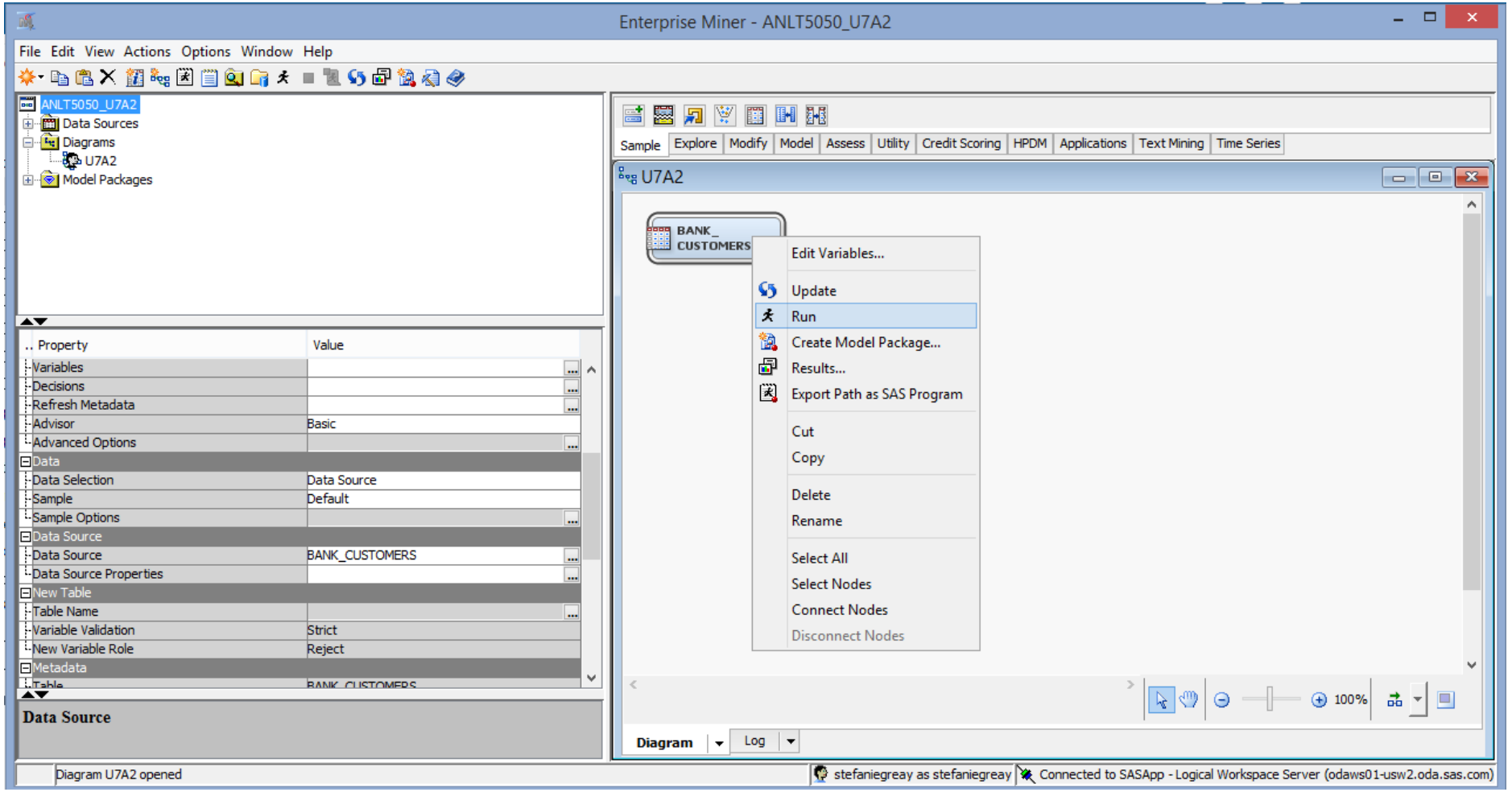
Click “Finish” to finish the data source registration within EM.



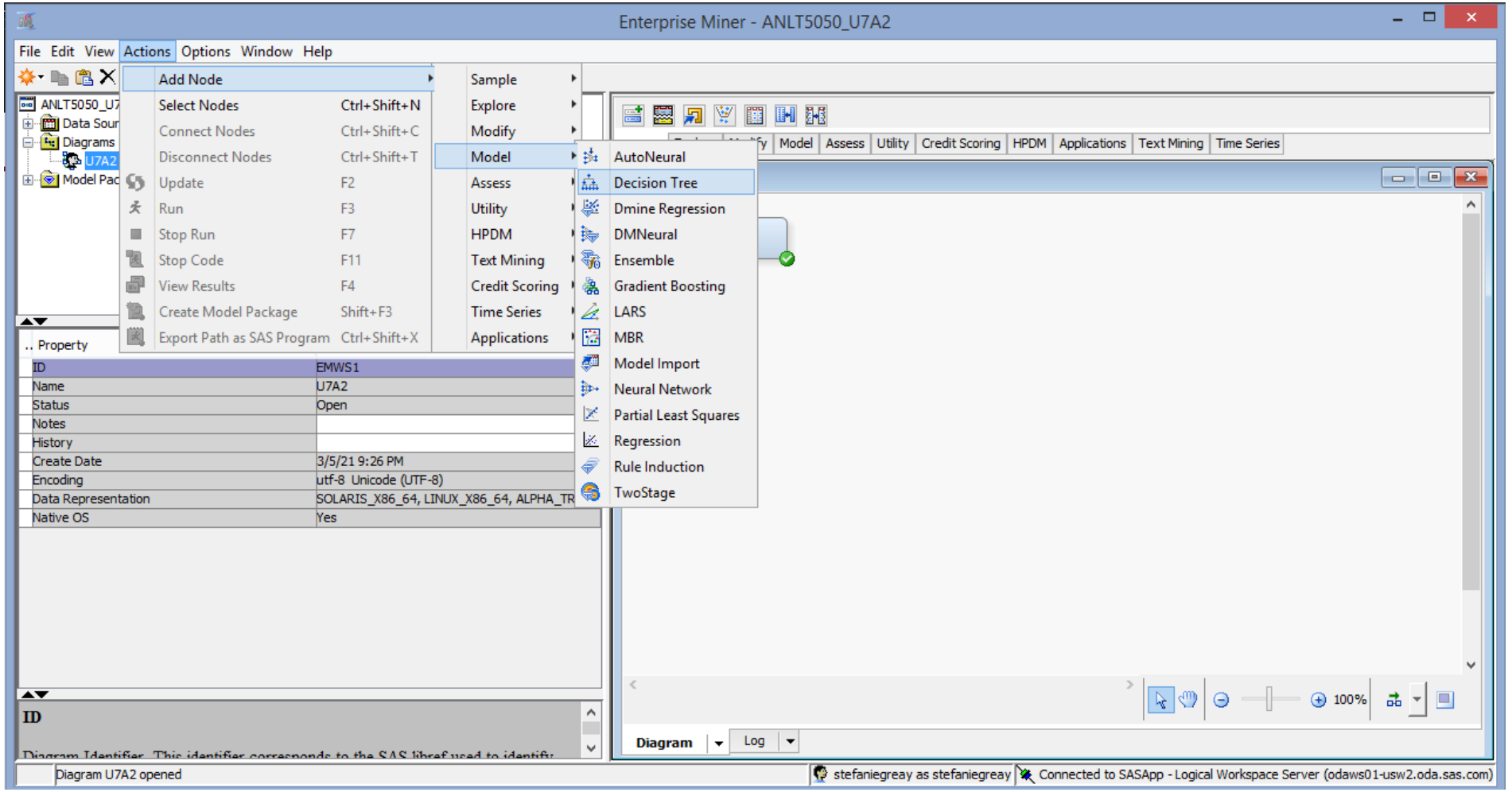
Click “OK” to complete the process. The name of the node should then change to the name of the dataset.



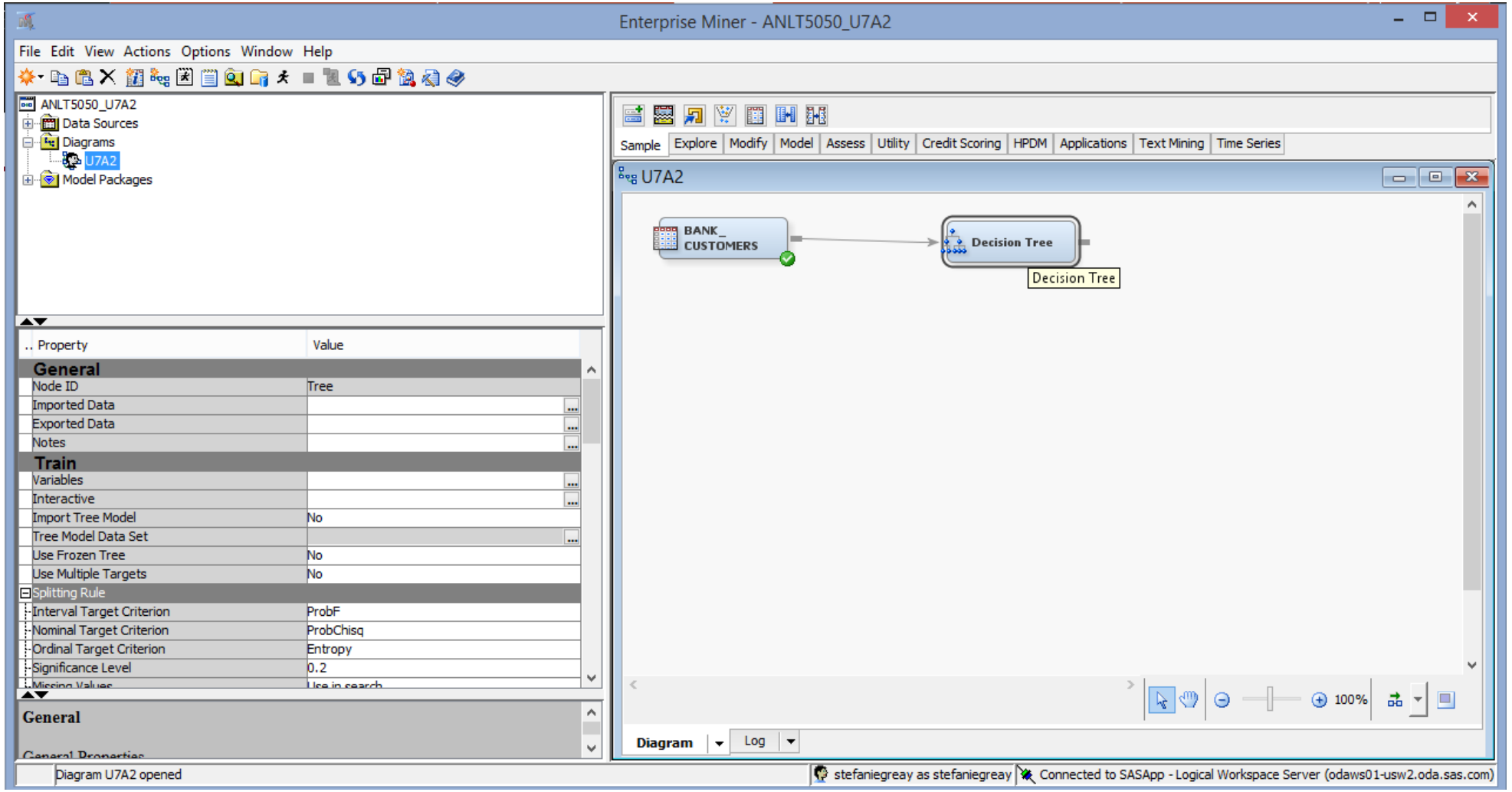
Right click on the dataset node and click “Run.”



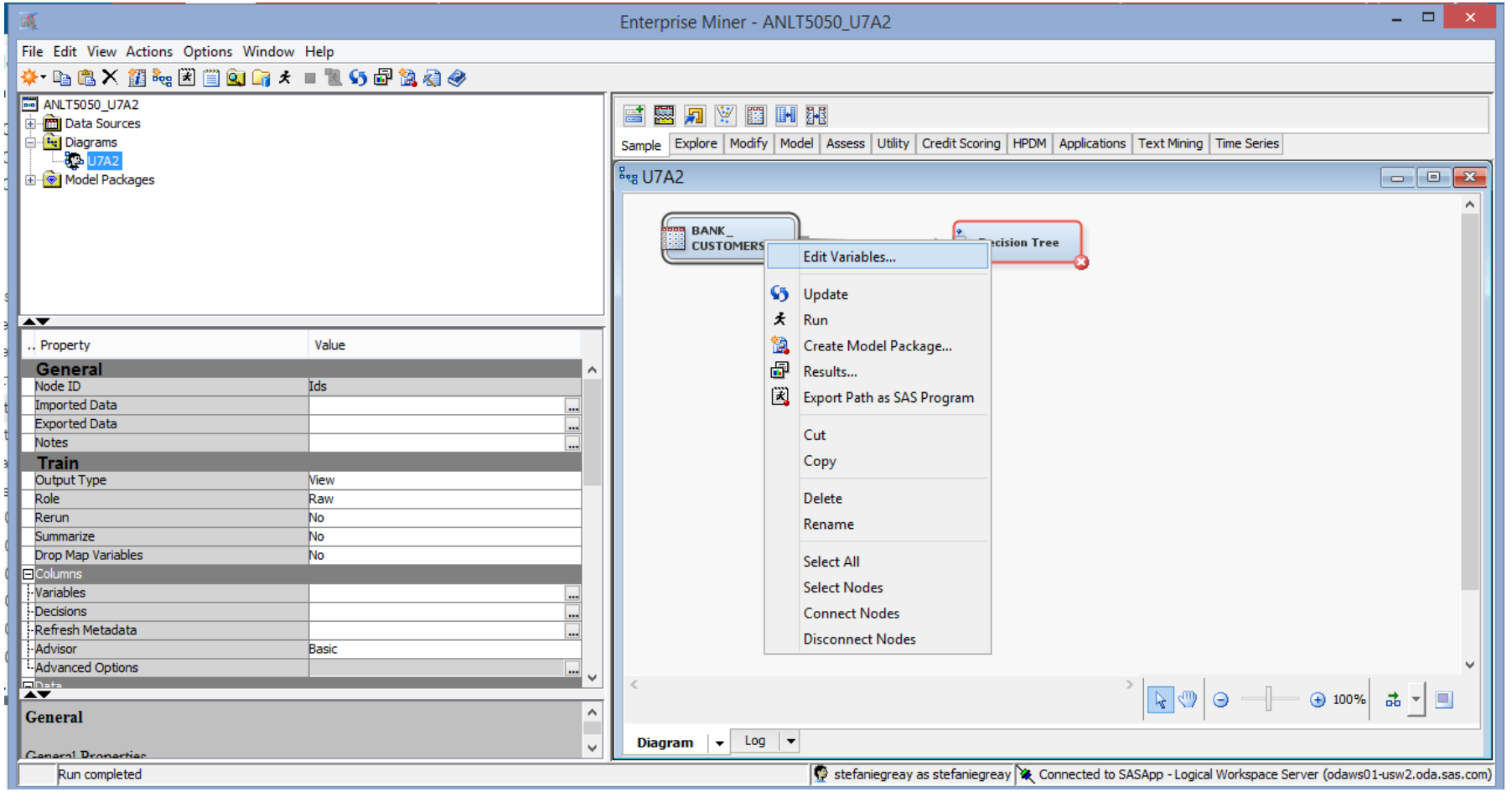
Click on “Actions” > “Add Node” > “Model” > “Decision Tree”



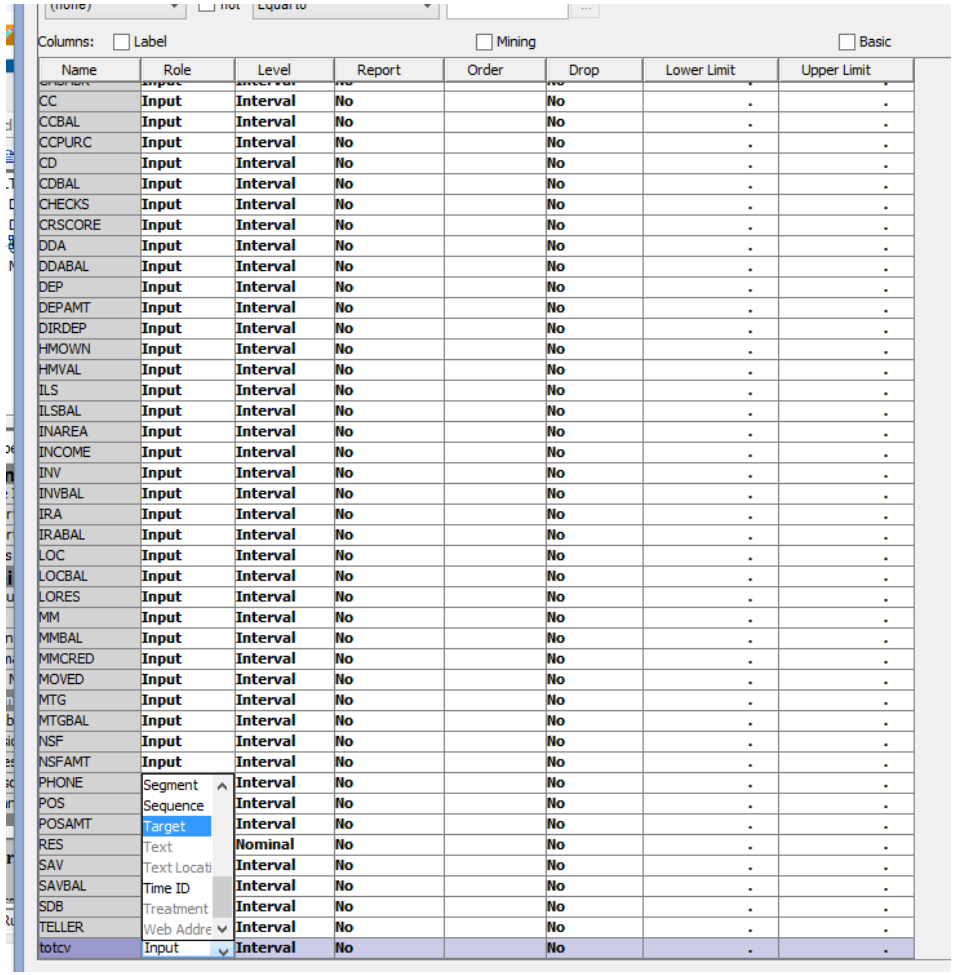
Connect the nodes



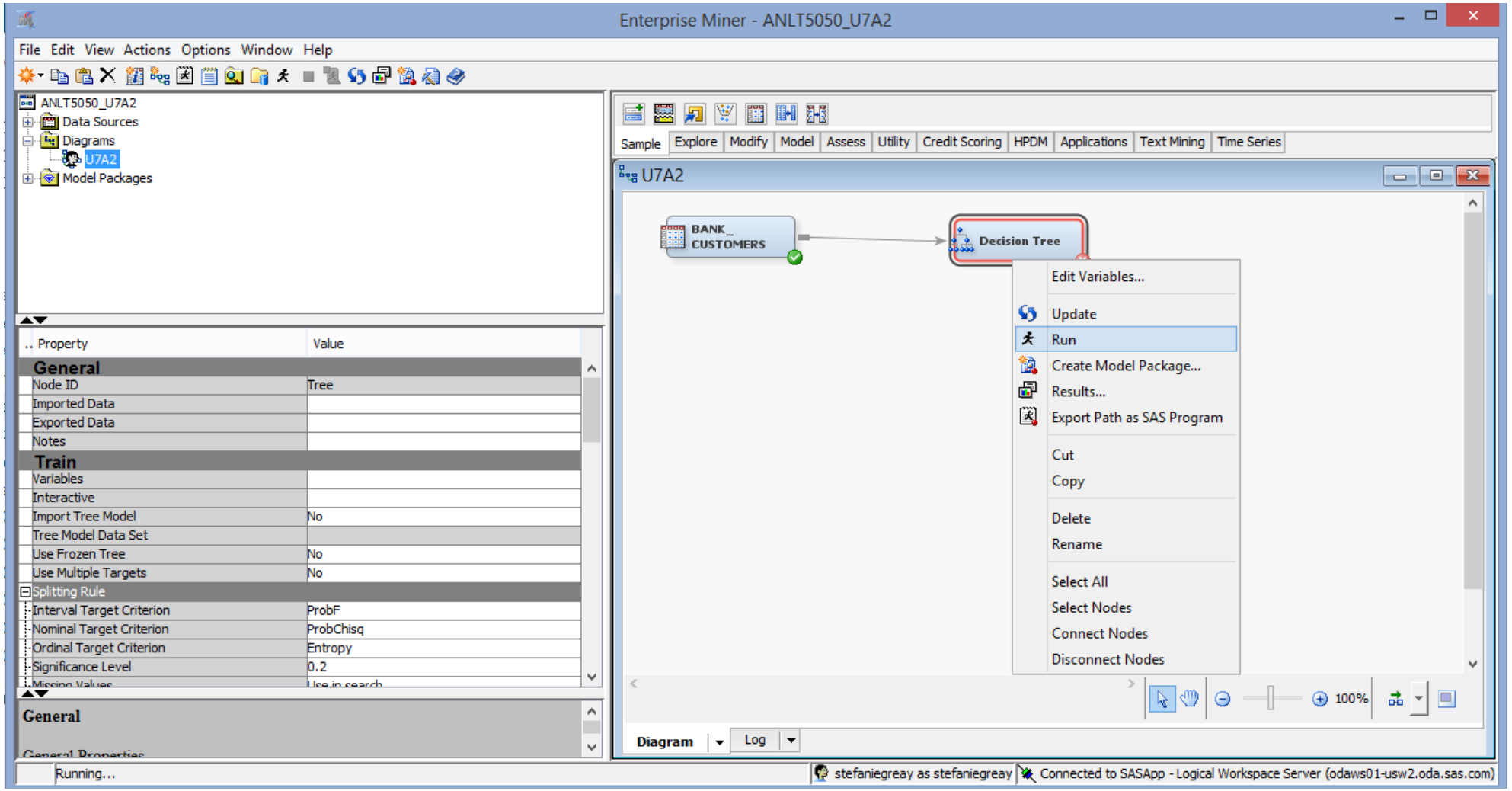
Right click on the dataset node and choose “edit variables.”



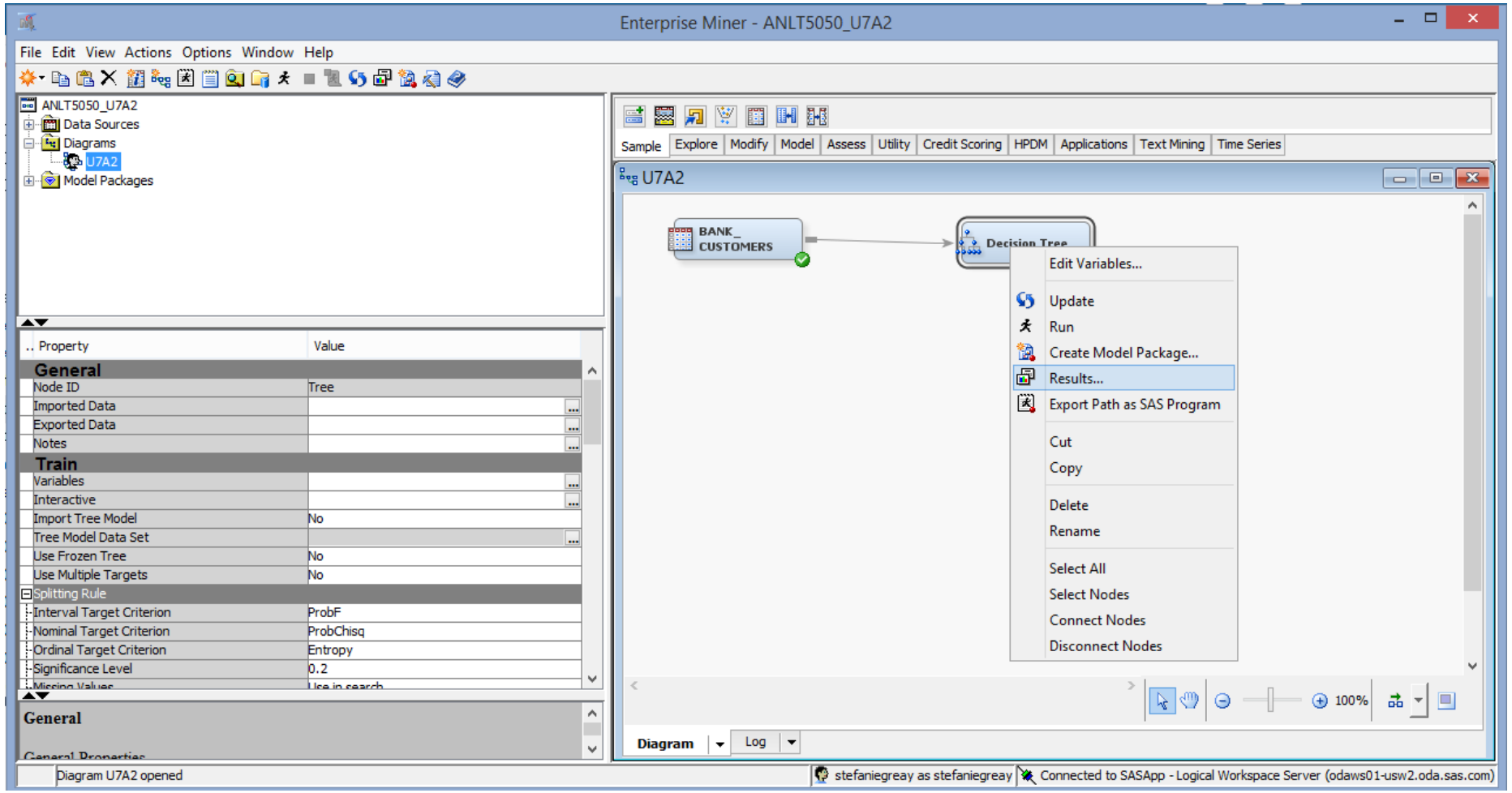
Change the totcv variable to “target” and leave all others as “Input,” then click “OK.”



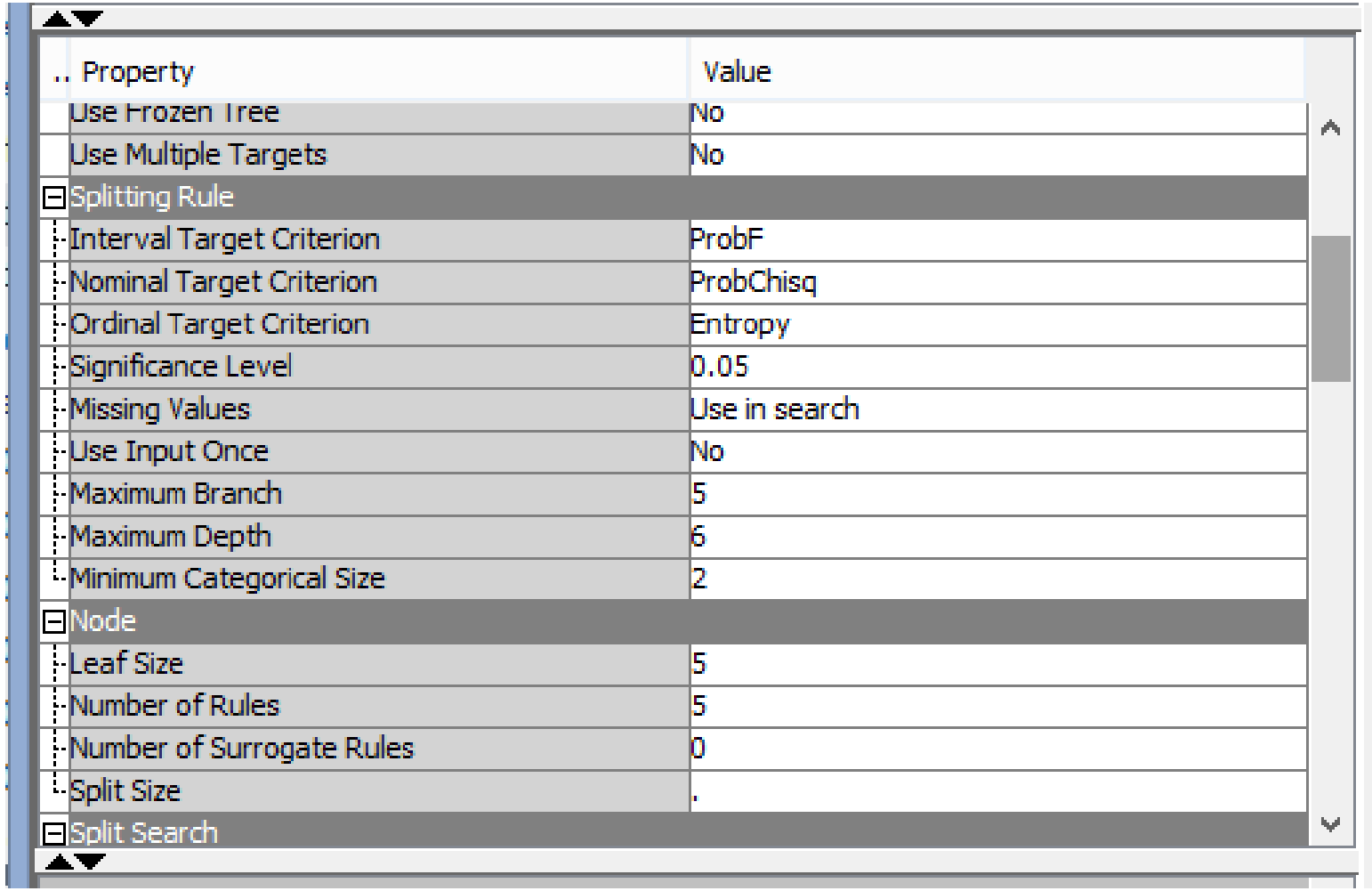
Right click on “Decision Tree” node and click “Run”



Right click on the “Decision Tree” node and click “Results” to view the results.



Options for splitting criteria (adjust the significance level, maximum branch, maximum depth and minimum categorical size parameters here).

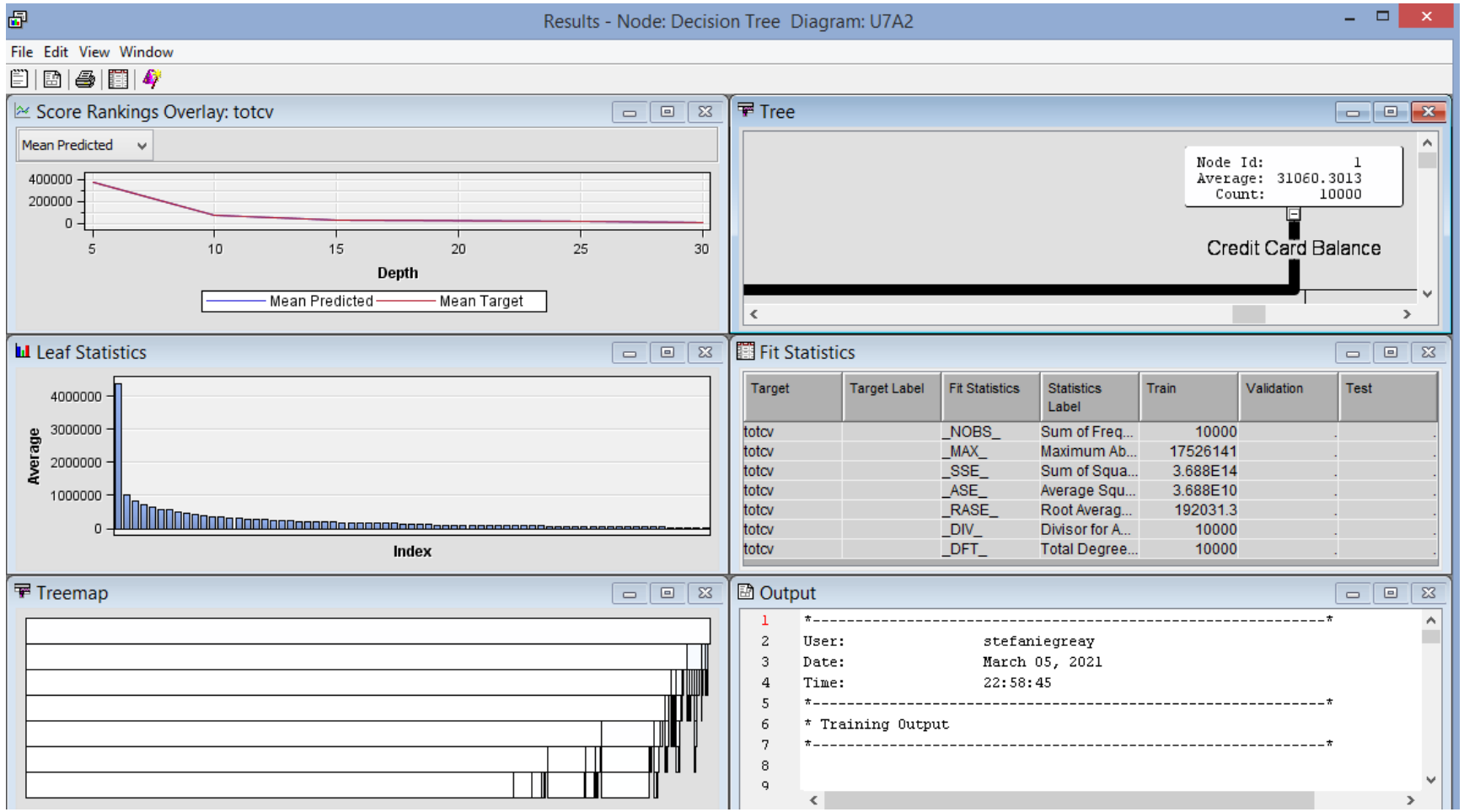


.. Property	Value
Use Frozen Tree	No
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.05
Missing Values	Use in search
Use Input Once	No
Maximum Branch	5
Maximum Depth	6
Minimum Categorical Size	2
Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
Split Search	

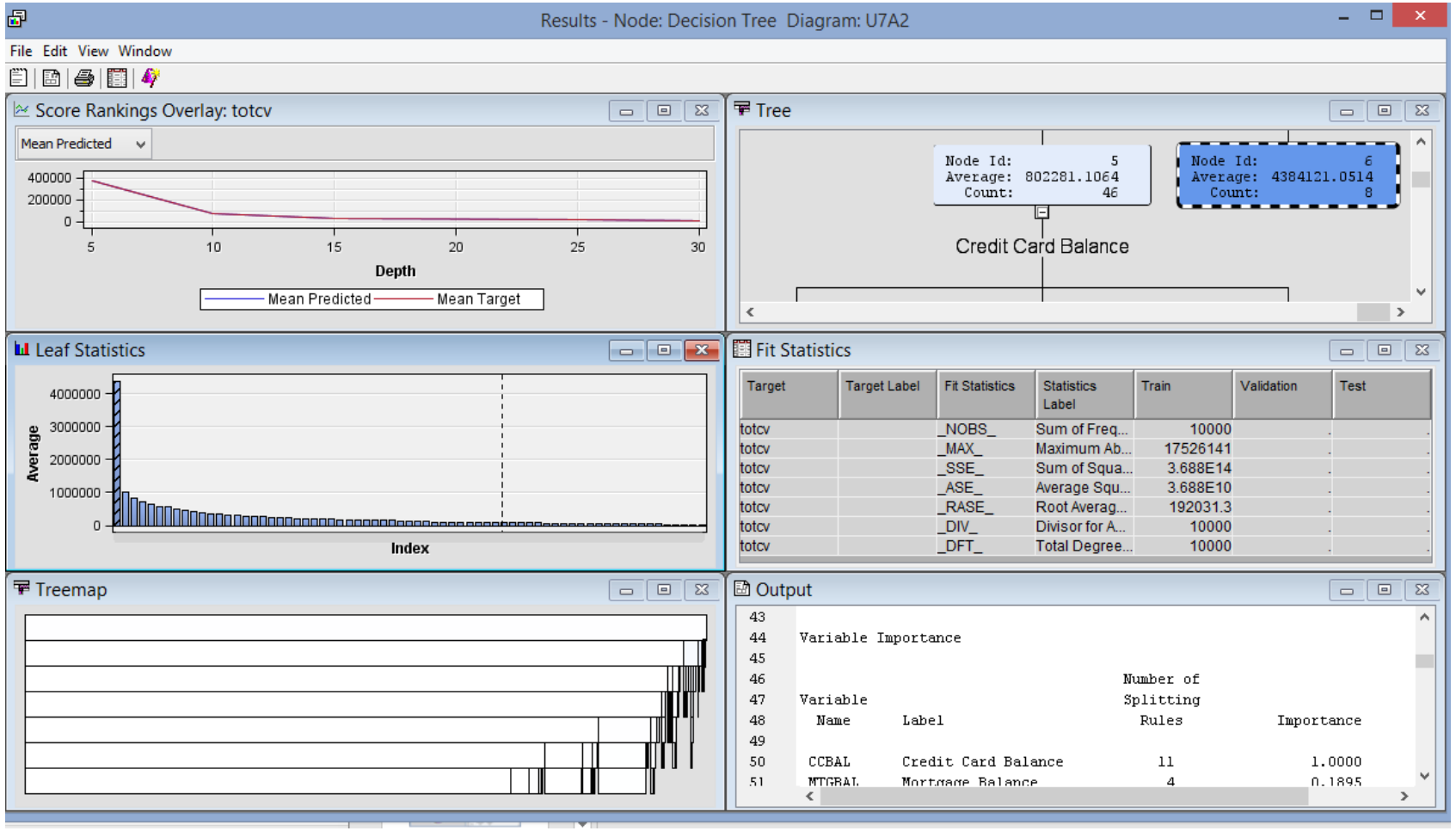
Considerations for splitting options

- Be sure that you understand the system constraints of whatever system you plan to implement the decision tree into, and make the splitting options fit those constraints.
- If you have categorical variables, like binary variables, be sure that you set the minimum categories setting accordingly so you do not automatically exclude these variables.
- The maximum branch setting is the maximum number of splits allowed in a branch (horizontal).
- The maximum depth setting is the maximum number of levels or layers of splits (vertical).

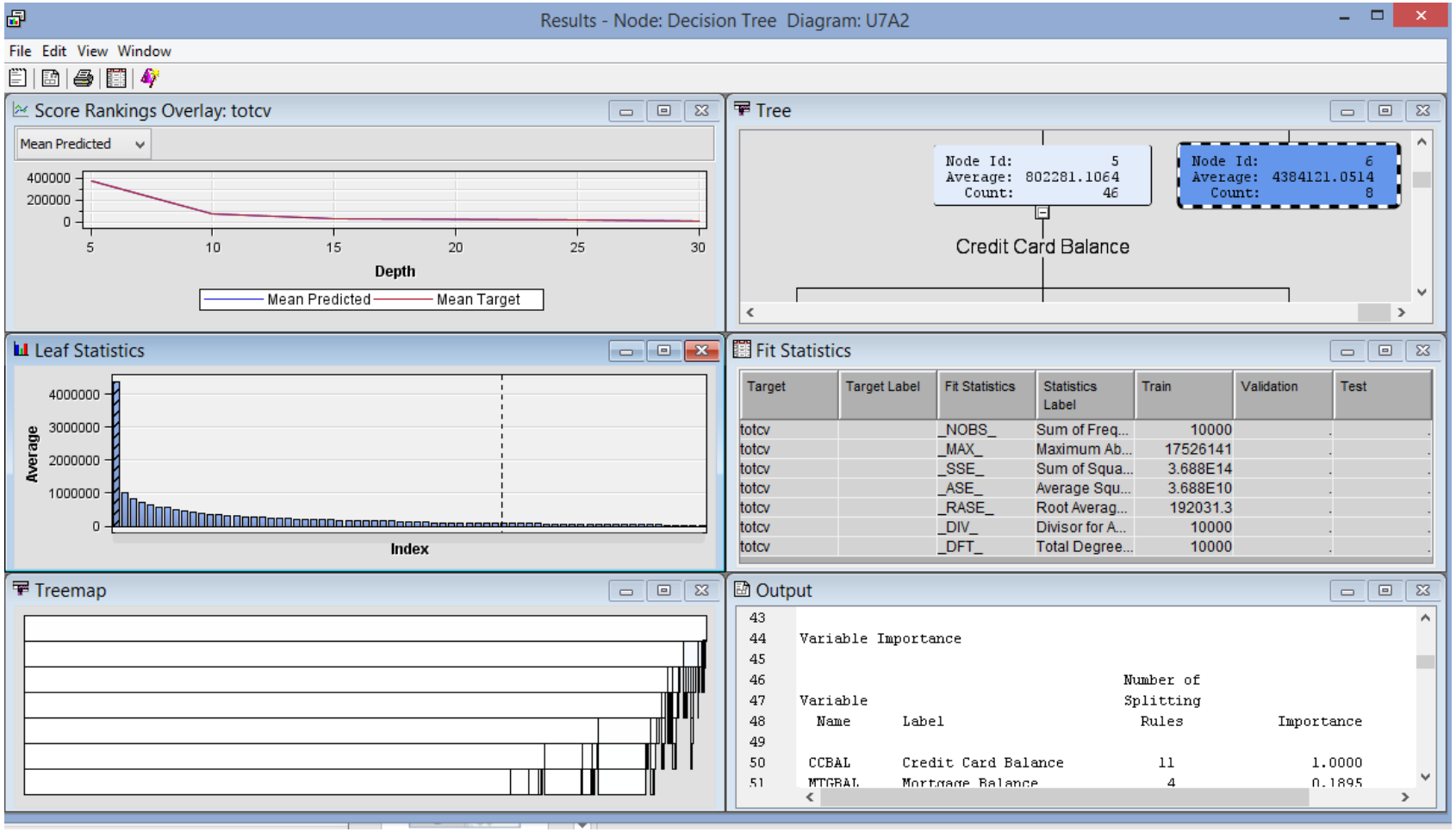
After you re-run the decision tree node (after making any adjustments to the split settings), right click on the node and select “Results.”



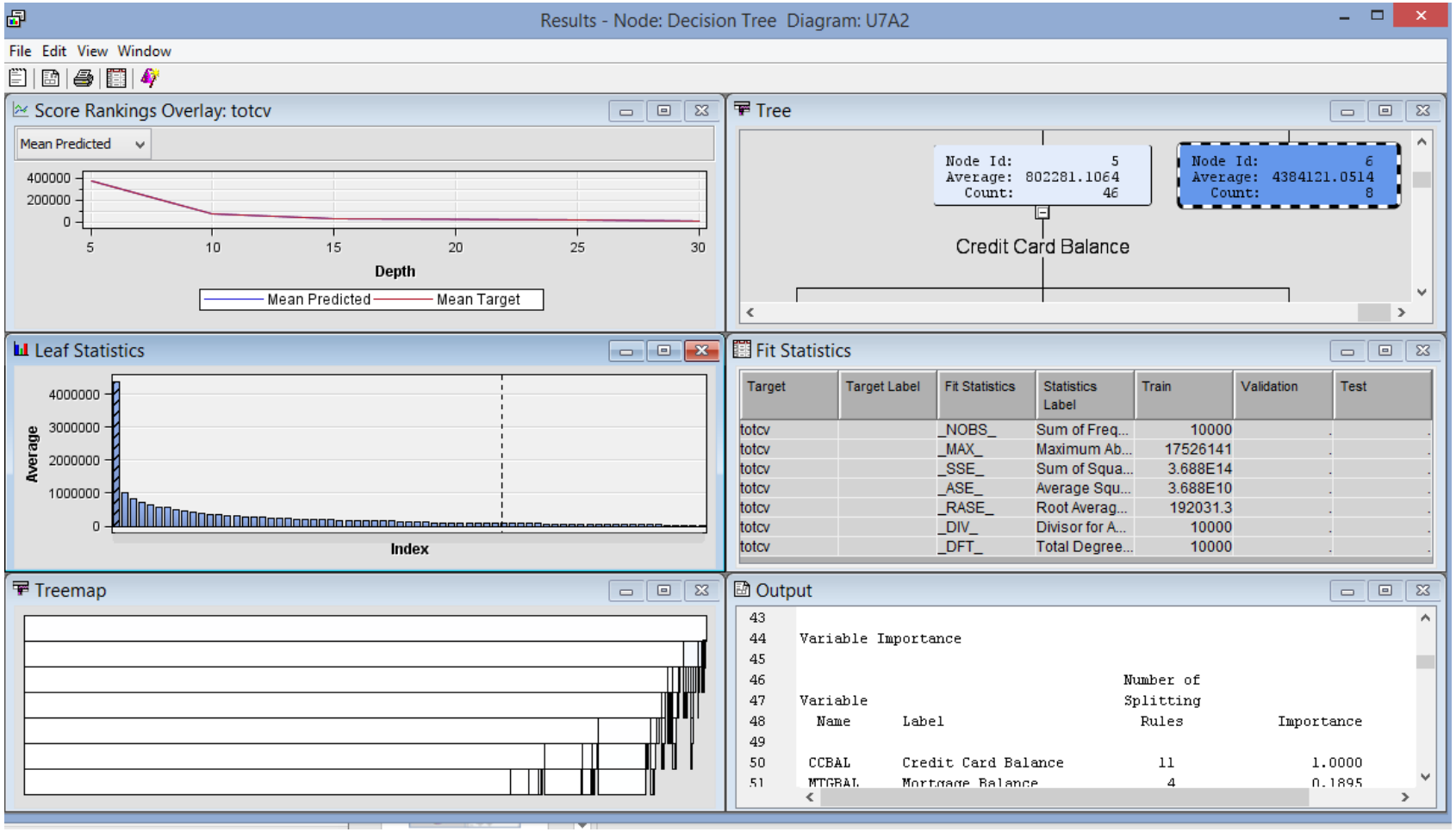
If you click on the bars in the “Leaf Statistics” window, it finds the appropriate leaf in the tree window and highlights it. You can then follow the logic that goes into that leaf by following it up the tree.



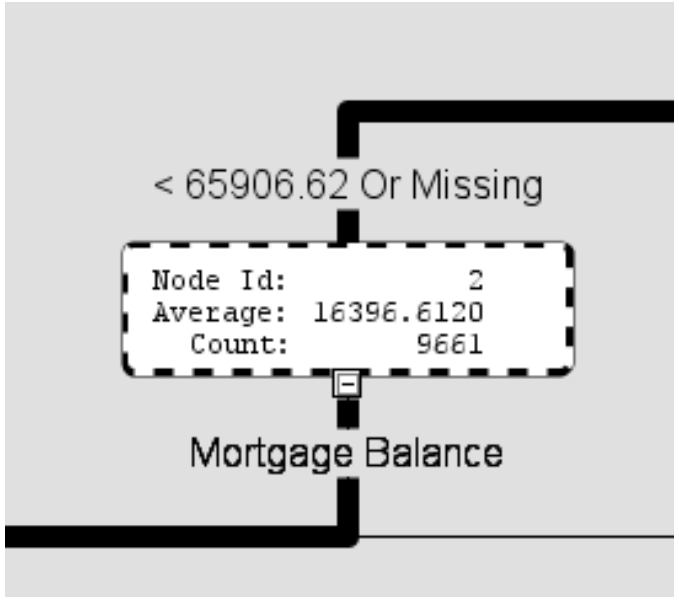
The Output window has the variable importance table that shows what variables contributed to the splitting/tree creation, as well as the leaf report.



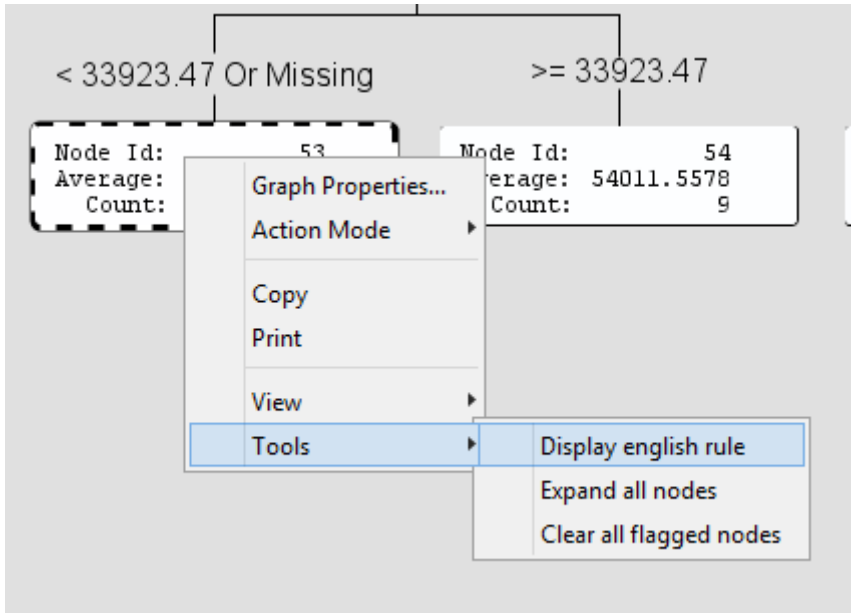
The Tree pane shows a graphical display of the tree, including the splitting criteria, count of observations, leaf identifier and the average of the target variable. This can be used to summarize the rules.



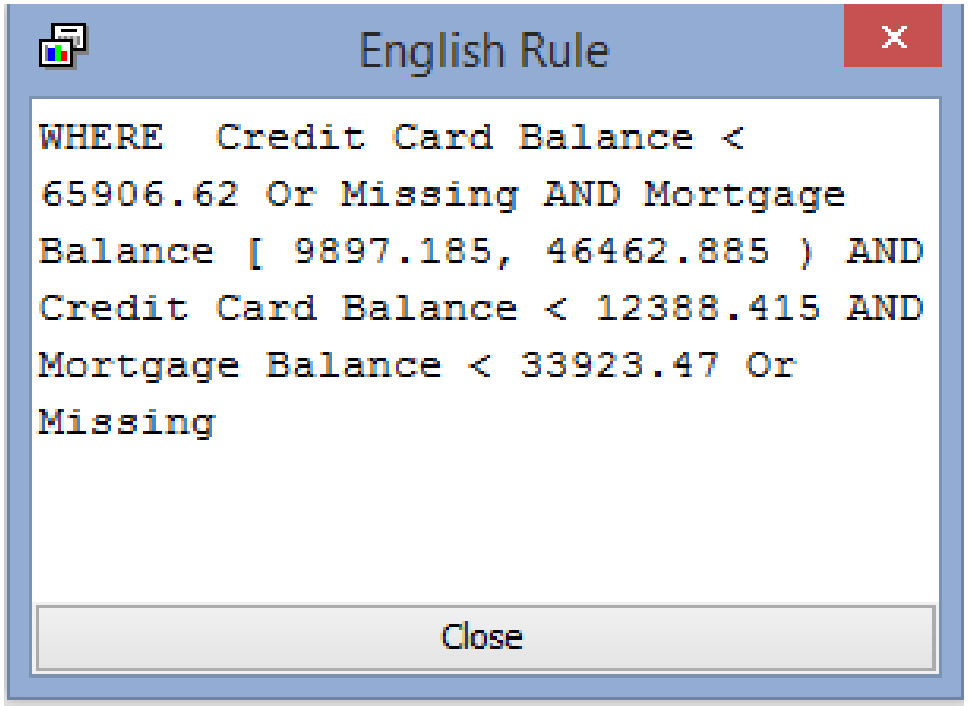
An example of interpreting one of the intermediate nodes. The Node Id is a reference for this node in the other output and results, the average is the average of the totcv (total customer value) variable we are using as the target variable. The count is the number of records or observations that fall into the criteria in this node. The logic above indicates the split criteria that created this node. The label below indicates the variable used in the next split.



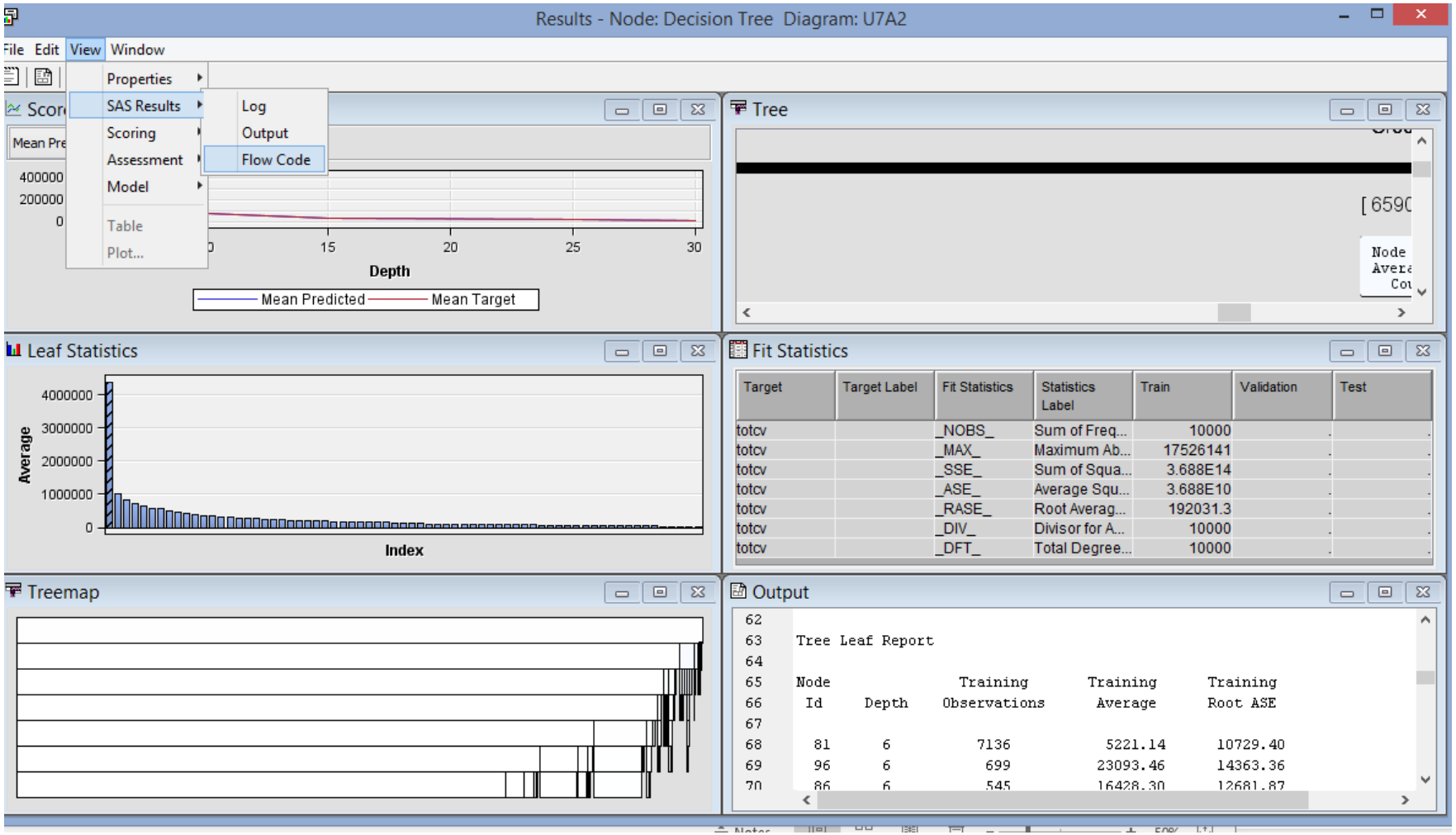
An example of interpreting one of the end nodes. The Node Id is a reference for this node in the other output and results, the average is the average of the totcv (total customer value) variable we are using as the target variable. The count is the number of records or observations that fall into the criteria in this node. The logic above indicates the split criteria that created this node. You can also display the rule that gets us to this node by right clicking on the node, then choosing “Tools” then “Display English rule.”



The English Rule will look something like this, and will contain all logic that is required to filter all the way down to that node/leaf (from the top of the tree).



To output the SAS code that includes the logic created by the decision tree, click on View, SAS Results, then Flow Code.



Best practice is to create a table that includes the splitting logic from the decision tree, instead of (or in addition to) the tree diagram itself, especially for executive summaries, as they are clearer and more easily reviewed and interpreted. It could look something like this, and would include all of the end nodes in the tree, at a minimum:

Node ID	Average Customer Total Value	Count	Logic
53	$34,303	19	WHERE Credit Card Balance < 65906.62 Or Missing AND Mortgage Balance [9897.185, 46462.885) AND Credit Card Balance < 12388.415 AND Mortgage Balance < 33923.47 Or Missing

The Assessment Score Distribution in the output will give you a good idea of what logical cutoffs for total customer value (tcv) you can use to identify “high value customers” or “top customers.” You can then focus the interpretation of the decision tree nodes/leaves on the leaves that contain those high value or top customers.

SAS Documentation Reference

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

https://documentation.sas.com/?docsetId=emref&docsetTarget=n0cx4ud03paymdn1kargegaduеml.htm&docsetVersion=14.3&locale=en