

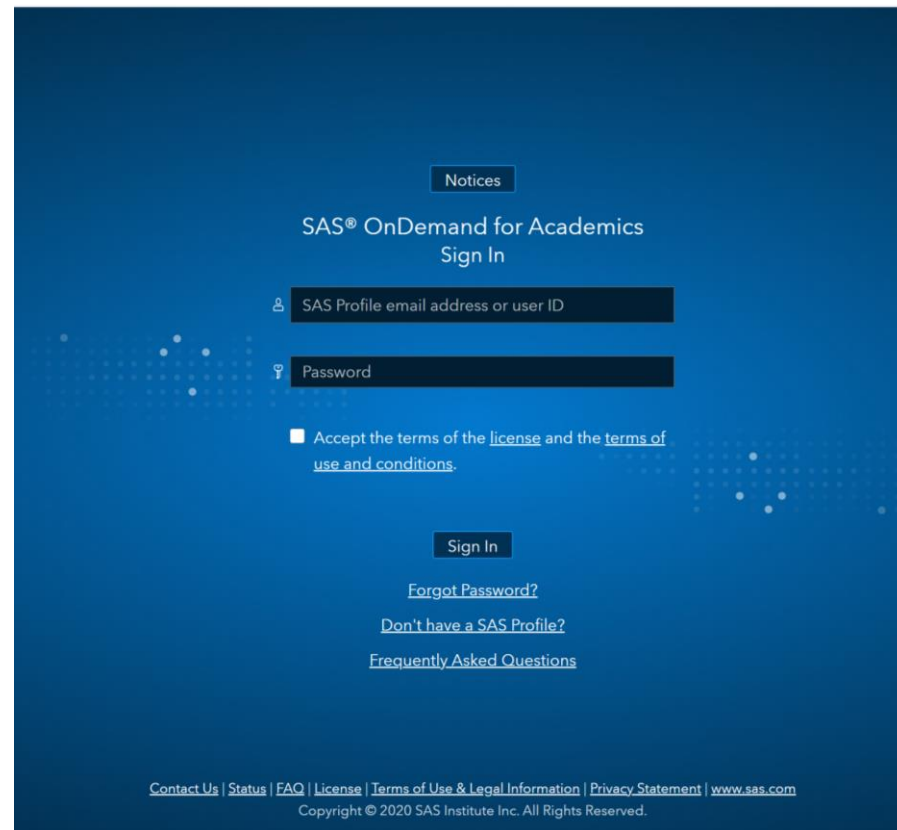
ANLT5050

Unit 7 Assignment 2 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 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, SAS® Enterprise Guide®, SAS® Enterprise Miner™, SAS® Forecast Studio, and JMP® Software access to SAS® hosted servers. Each product entry includes an icon, the product name, a brief description, and a link to actions. 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" section with a link to learn more, showing progress bars for the Home Directory (1% of 46.5MB/5120MB) and Course Directory (7% of 207.0MB/3072MB). At the bottom, there is a link to "Other Ways to Access SAS® OnDemand for Academics Resources".

SAS

United States | Stefanie Reay

SAS® OnDemand for Academics Dashboard

Planned Events Notices

Applications Enrollments Courses

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](#))

Reference

[Support Site](#)
[Step-by-Step Reference Guides](#)
[Frequently Asked Questions](#)

Quotas ([learn more](#))

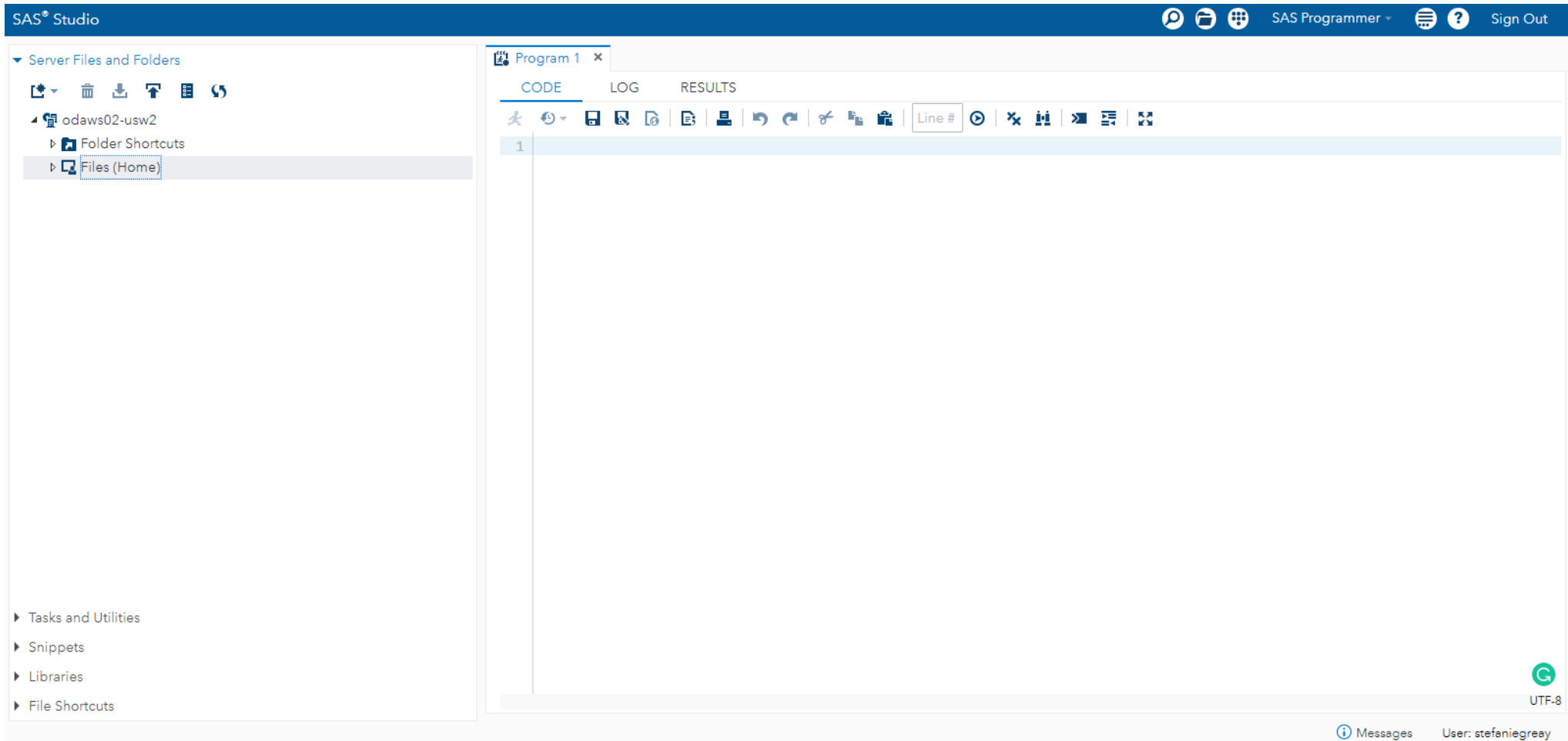
Home Directory (46.5MB/5120MB)
1%

Course Directory (207.0MB/3072MB)
7%

[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.”

Upload Files ×

Upload files to:

Selected files:

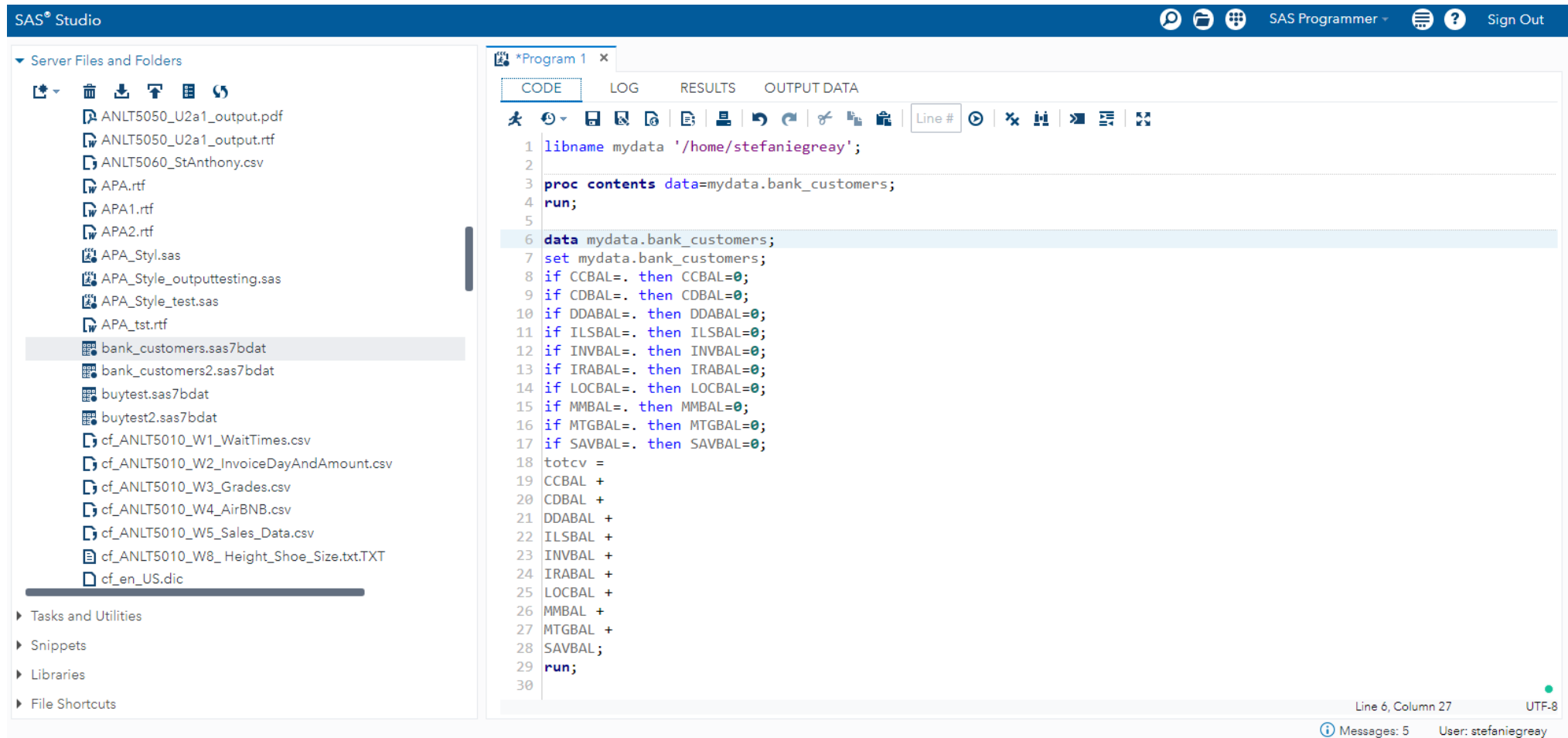
1	SAS7BDAT	bank_customers.sas7bdat	1.1 mb
---	----------	-------------------------	--------



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;
```



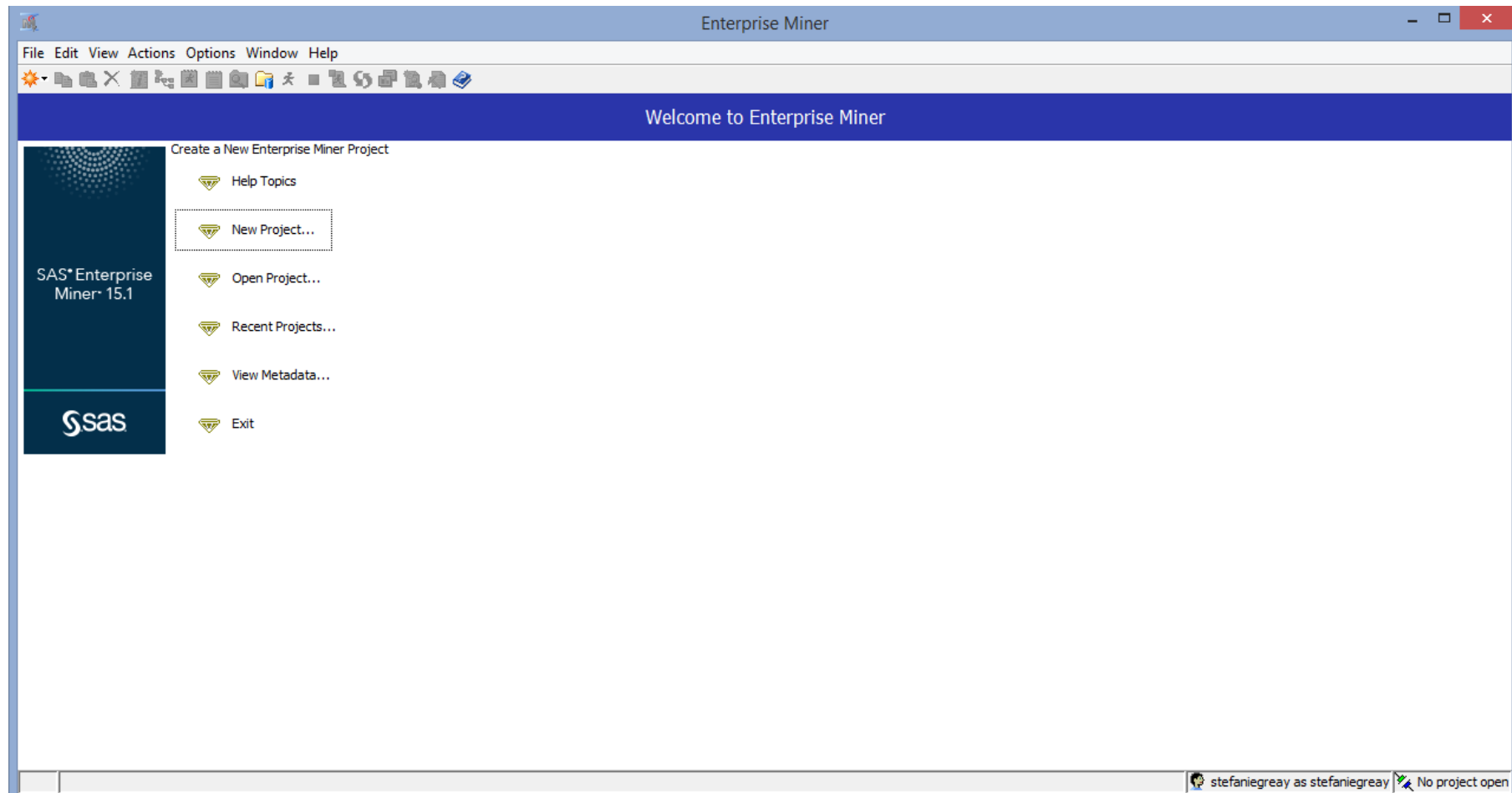
SAS Enterprise Miner Instructions

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

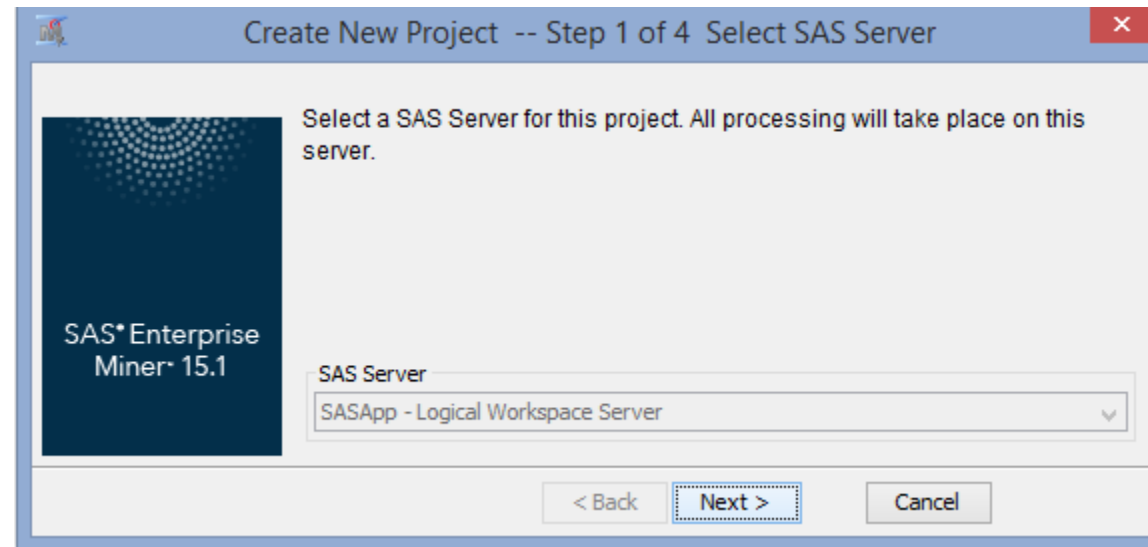
Once you have uploaded the dataset for this unit onto the SAS servers using SAS Studio, you may proceed from here using SAS Enterprise Miner.



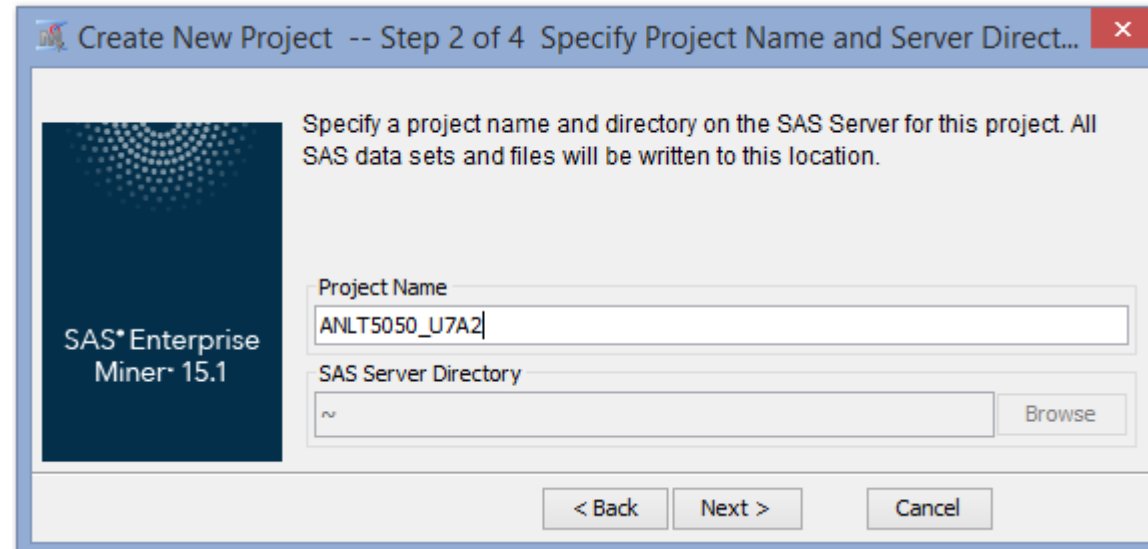
Once you download and start SAS Enterprise Miner, open a new project by clicking on “New Project.”



Click “Next>” to use the default SAS Server



Enter a project name and click “Next>”



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

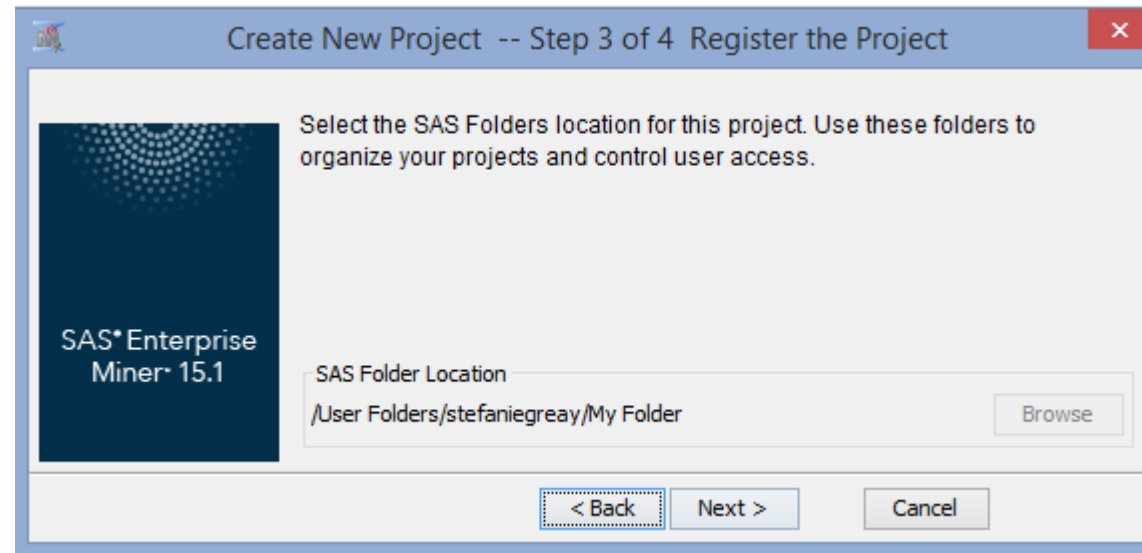
Project Name
ANLT5050_U7A2

SAS Server Directory
~ Browse

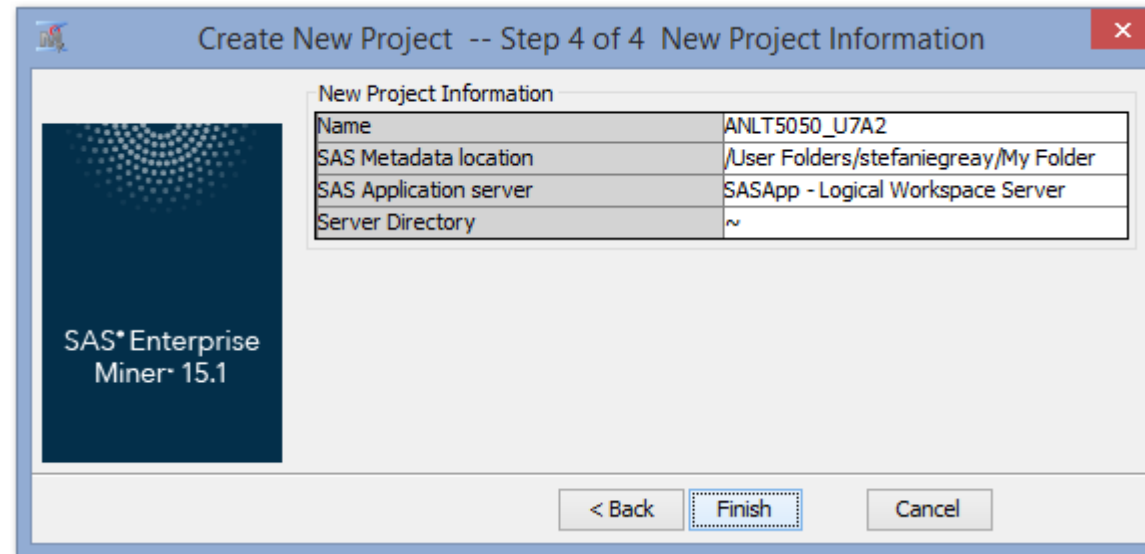
< Back Next > Cancel



Click “Next>”



Verify your entries and click “Finish”



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

New Project Information

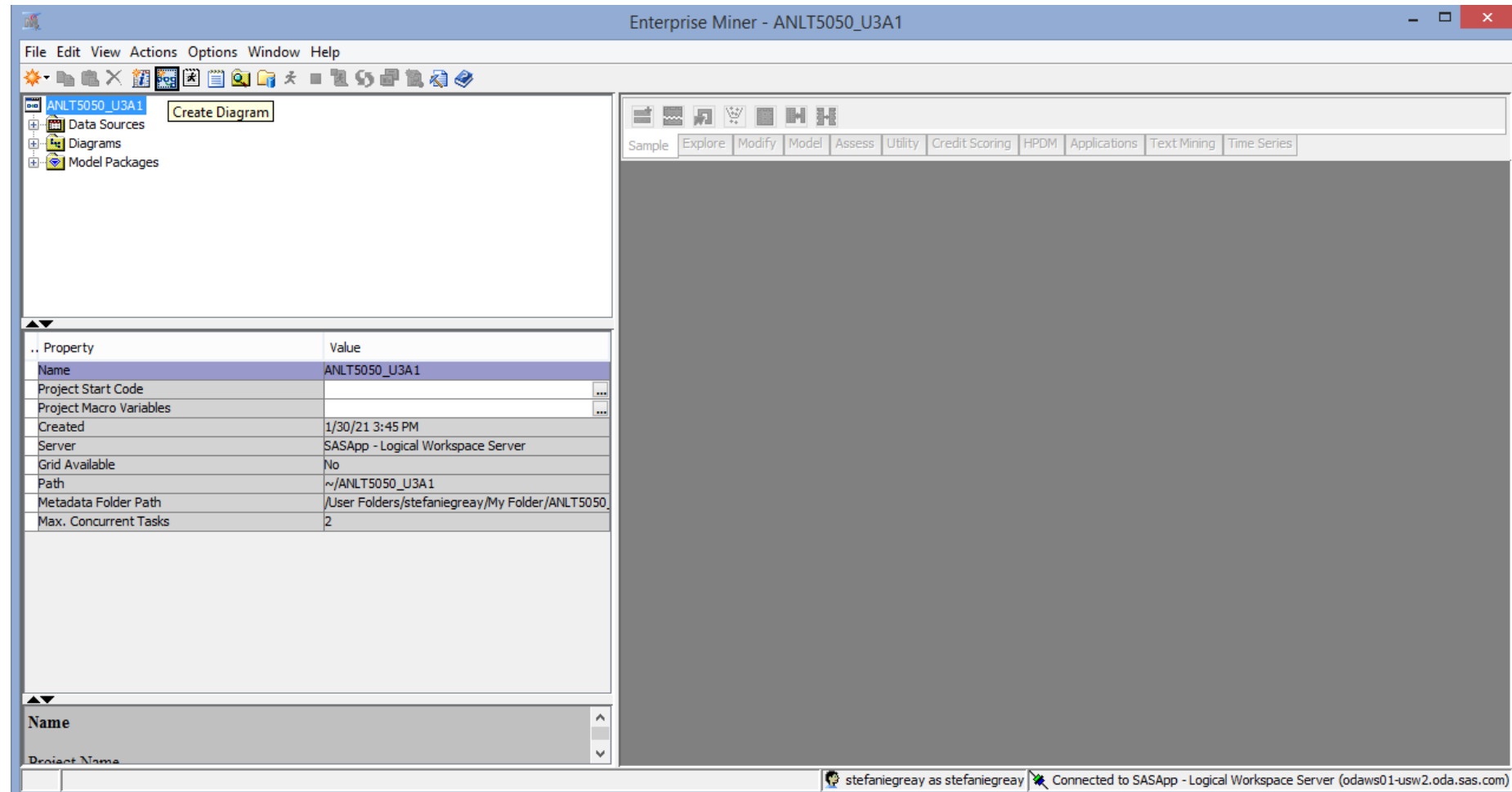
Name	ANLT5050_U7A2
SAS Metadata location	/User Folders/stefaniegreay/My Folder
SAS Application server	SASApp - Logical Workspace Server
Server Directory	~

SAS® Enterprise Miner 15.1

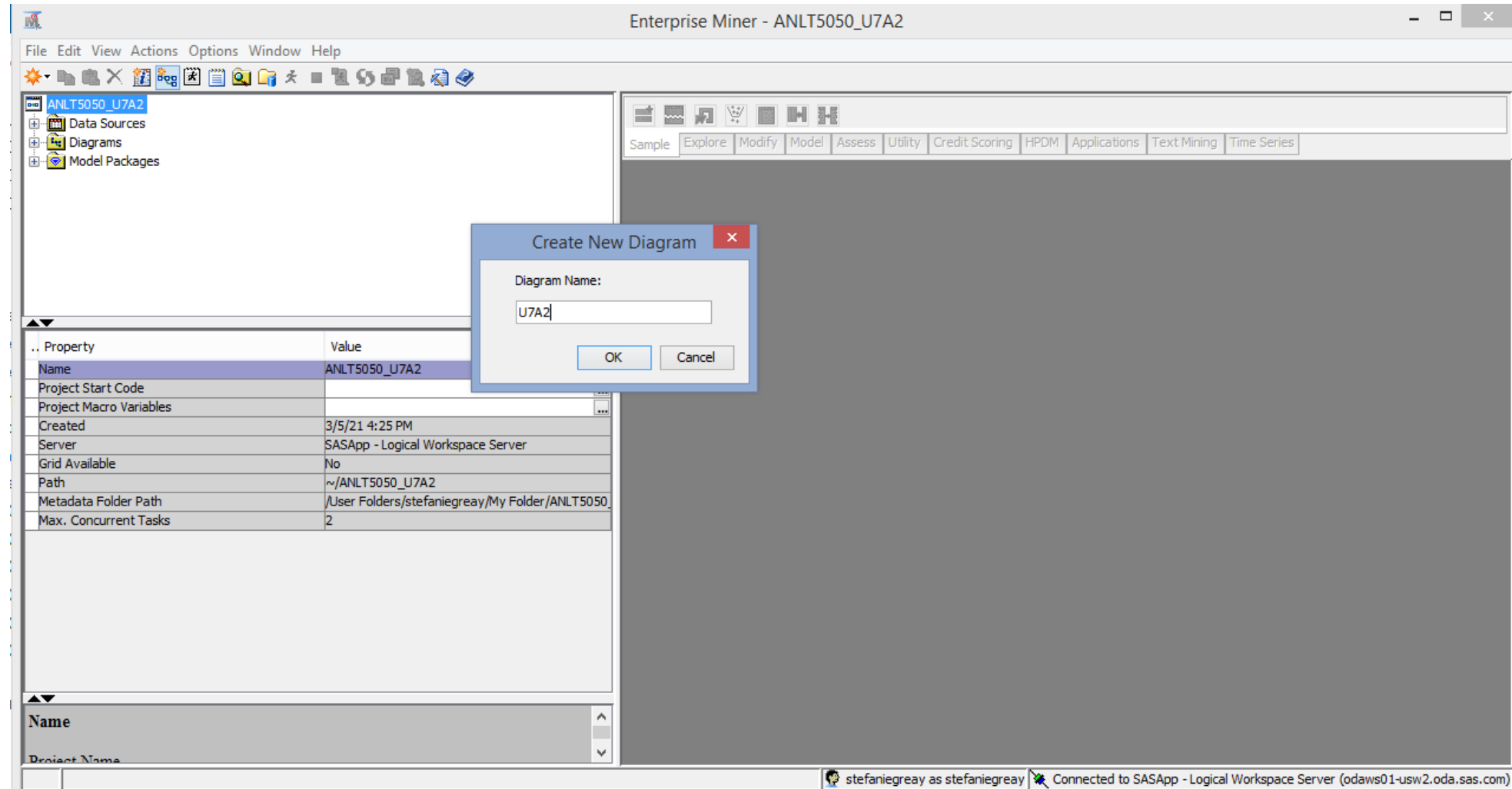
< Back Finish Cancel



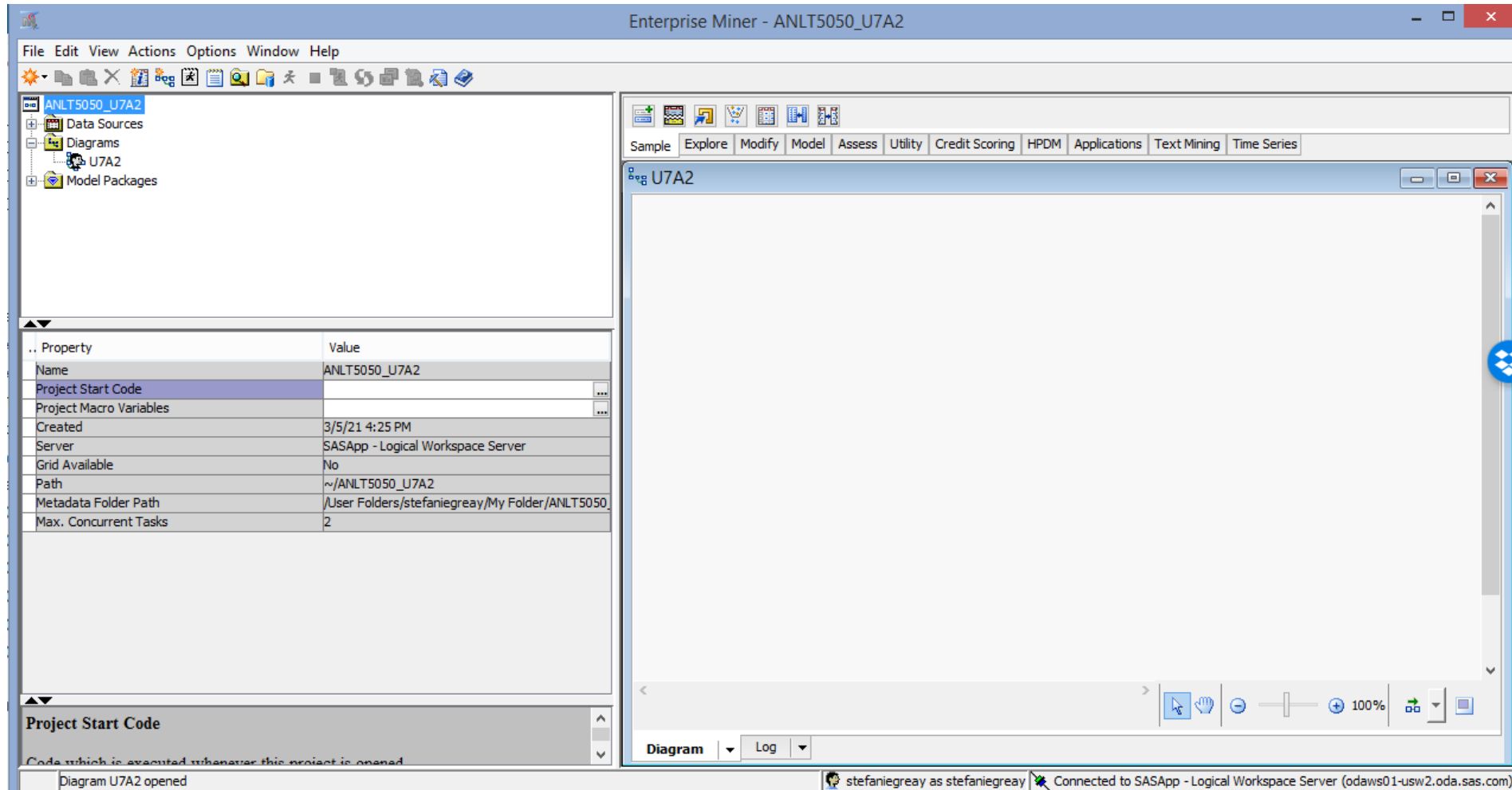
Click on the “Create Diagram” icon.



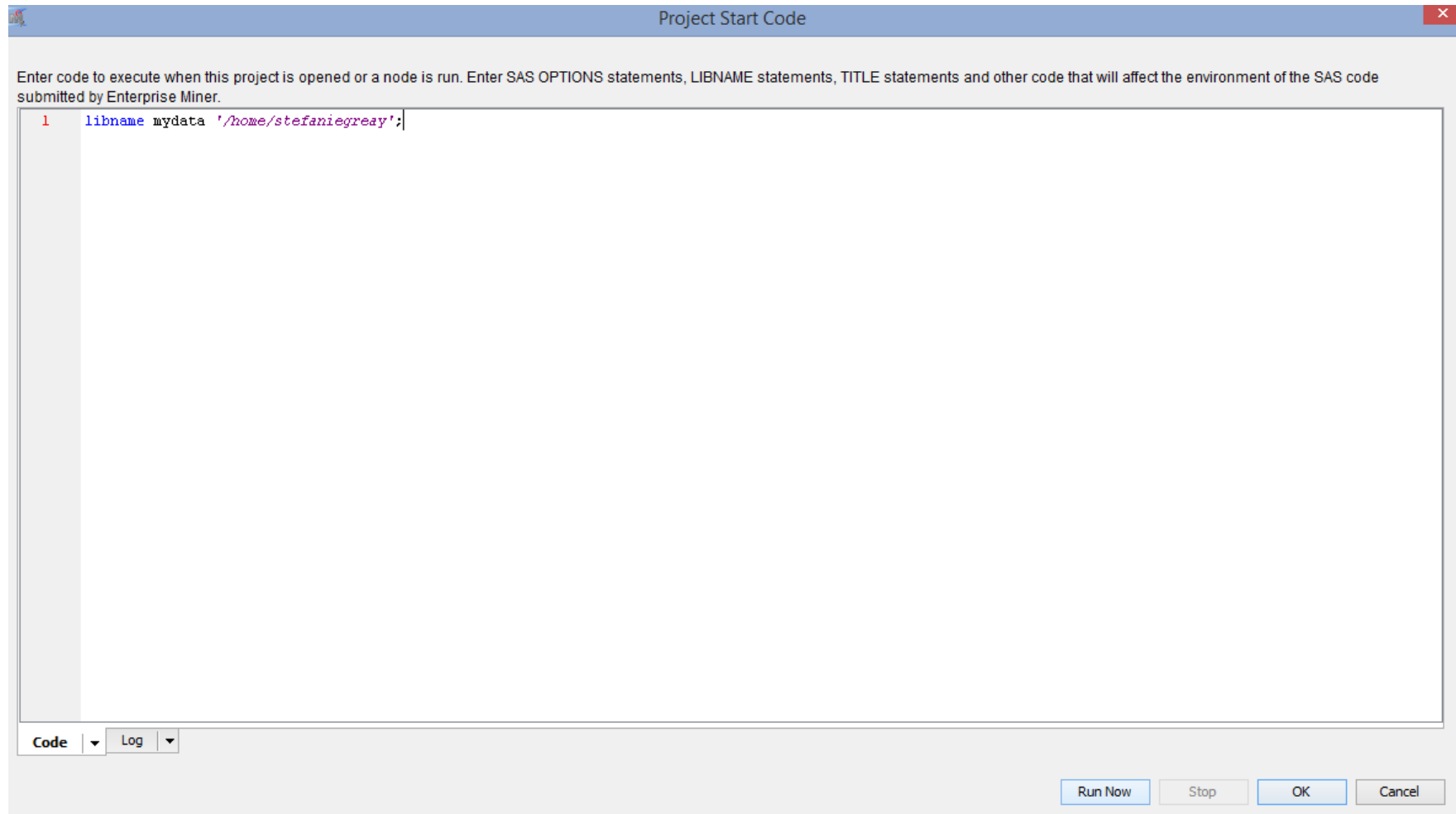
Name your diagram and click “OK.”



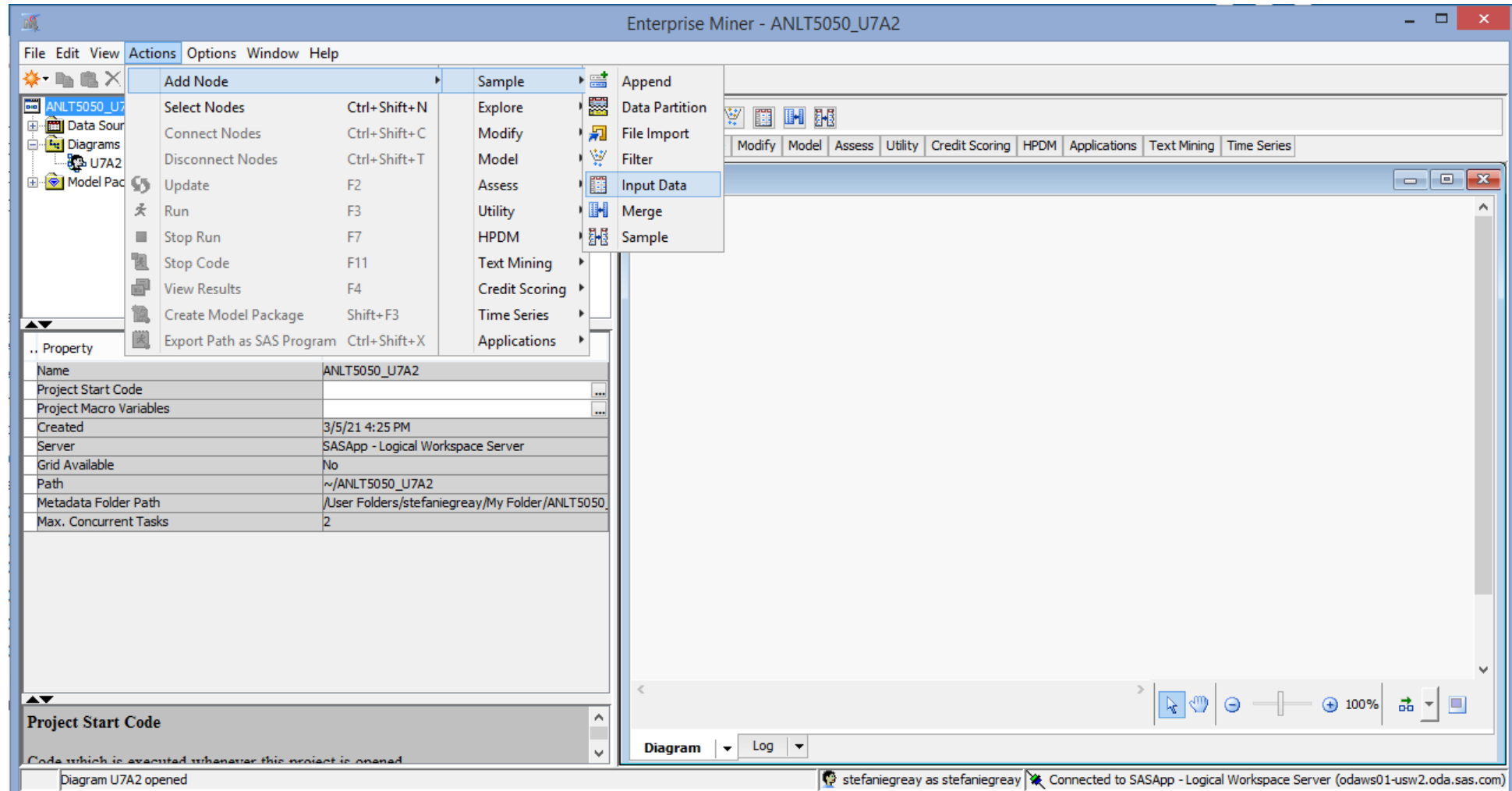
Click on the project, then click on the ellipses next to “Project Start Code.”



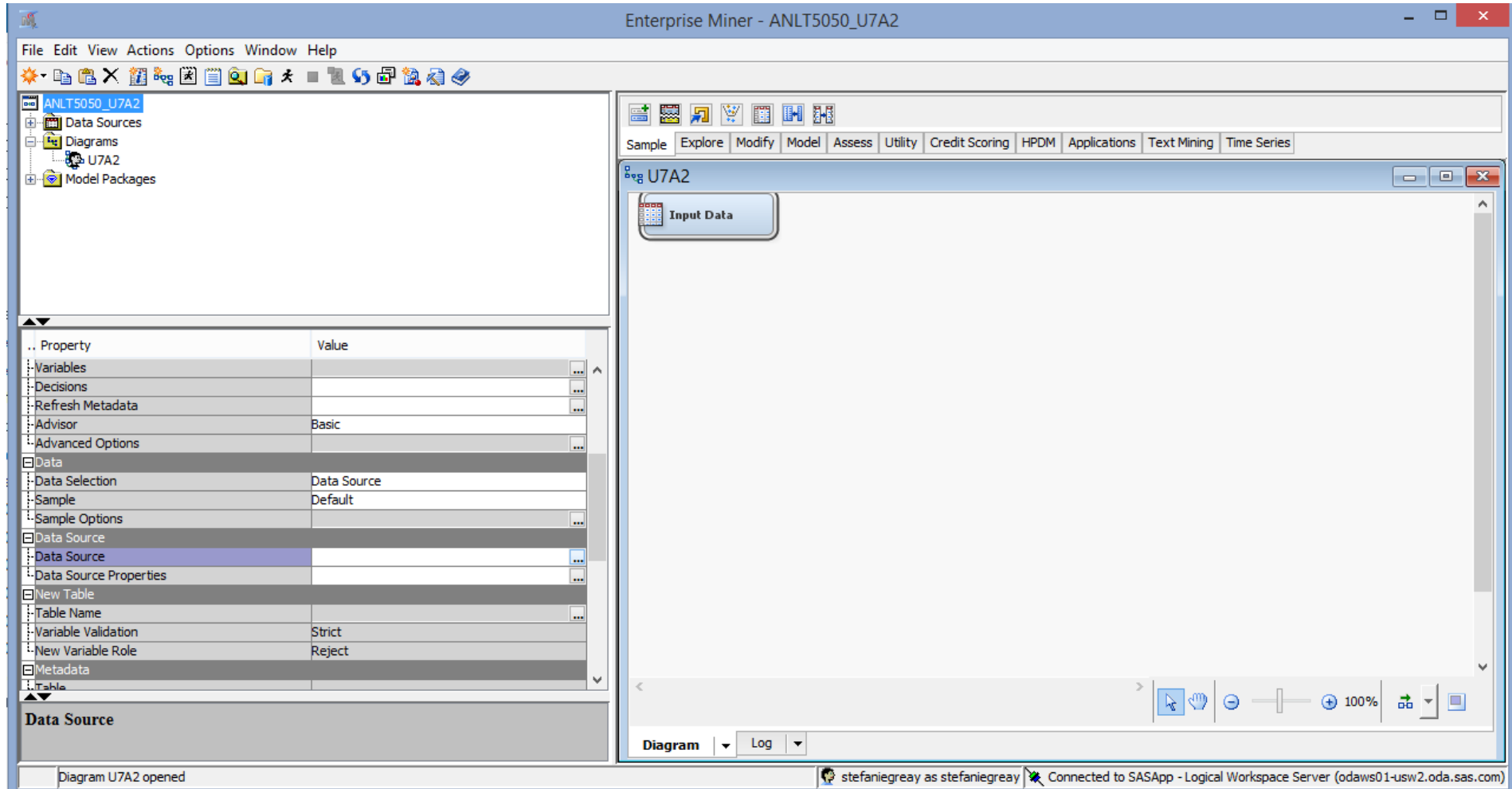
Add the library reference for where you uploaded the dataset in SAS studio, and click “Run Now.” Once it completes, click “OK.”



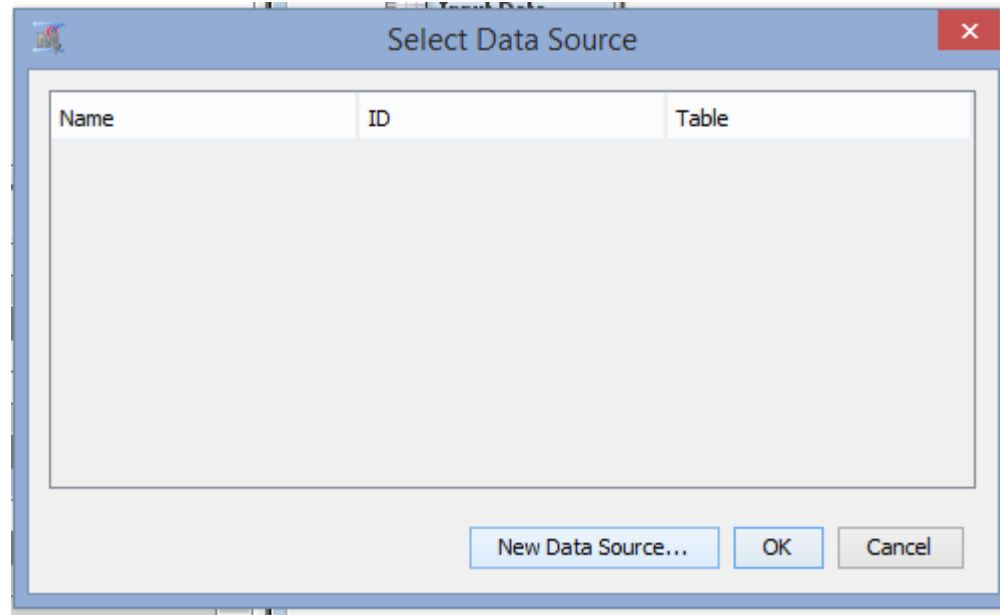
Click on Actions>Add Node>Sample>Input Data



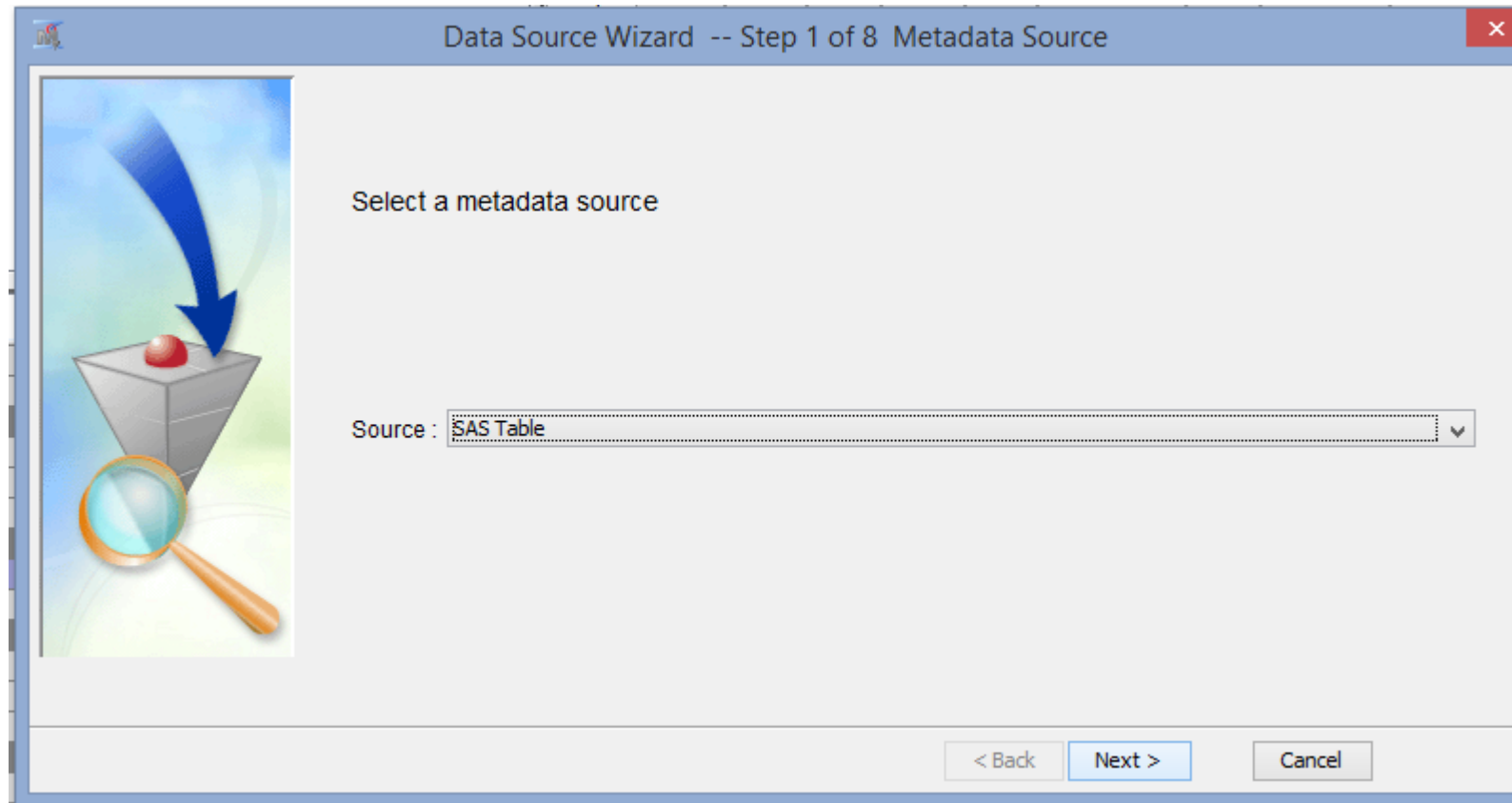
Click the ellipses (3 dots) next to “Data Source.”



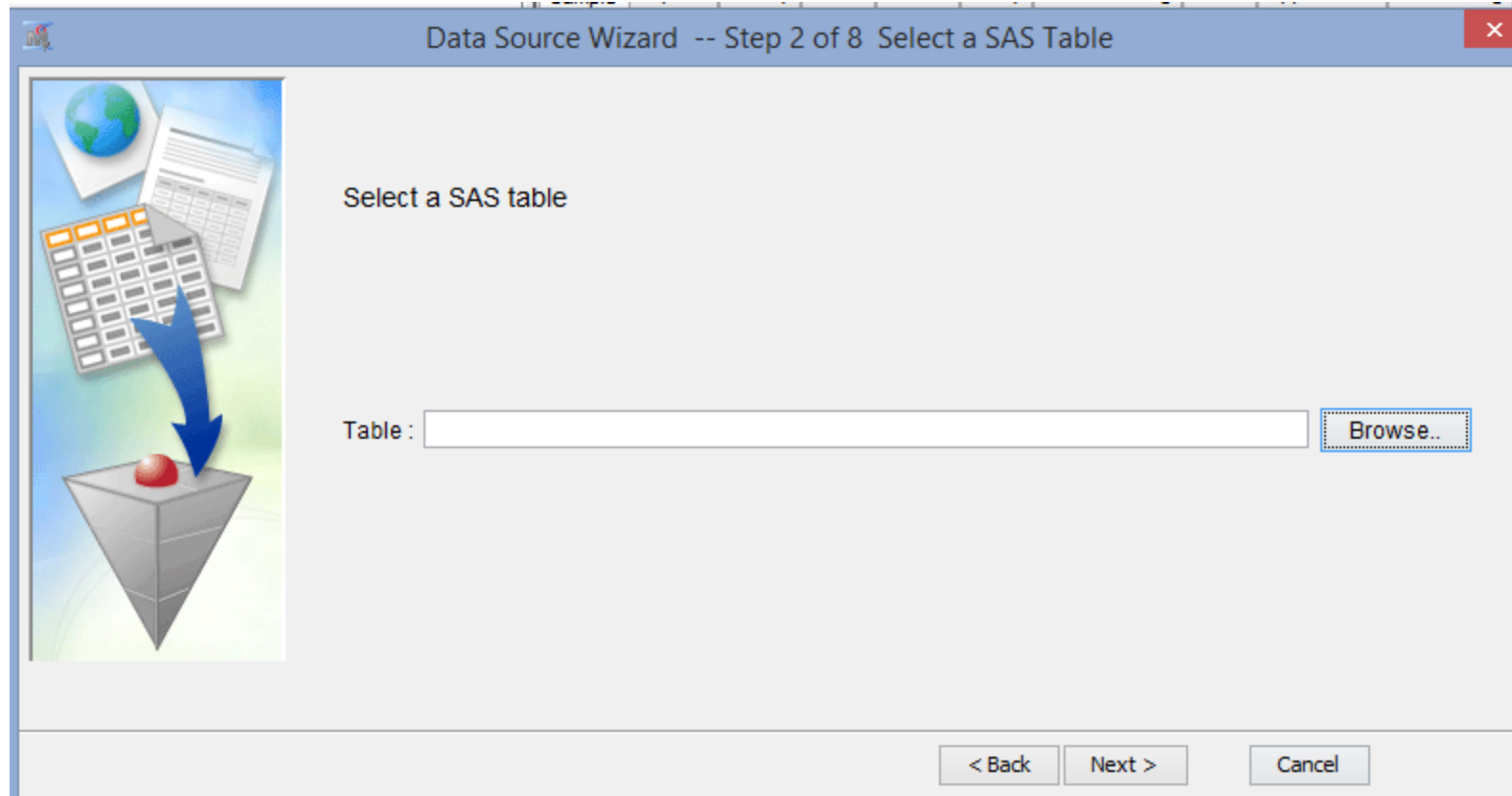
Click on “New Data Source”



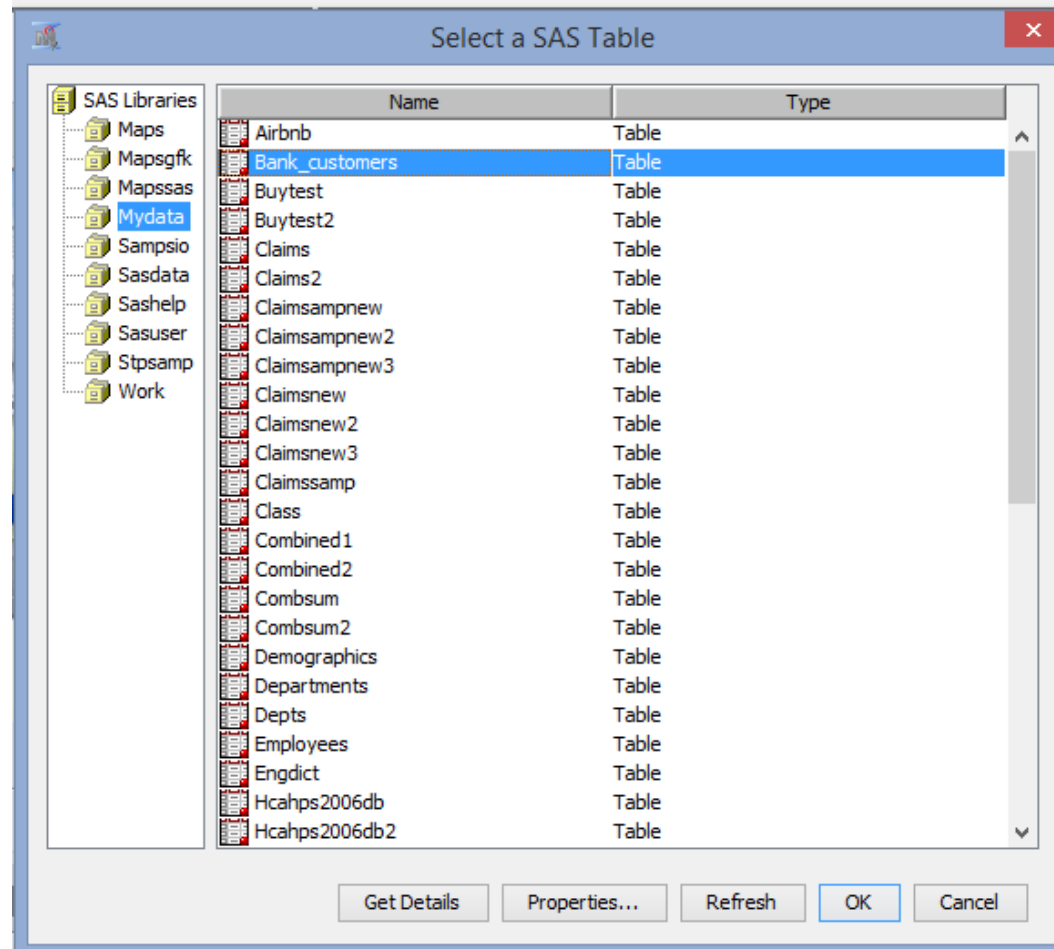
Leave it as “SAS Table” and click “Next >”



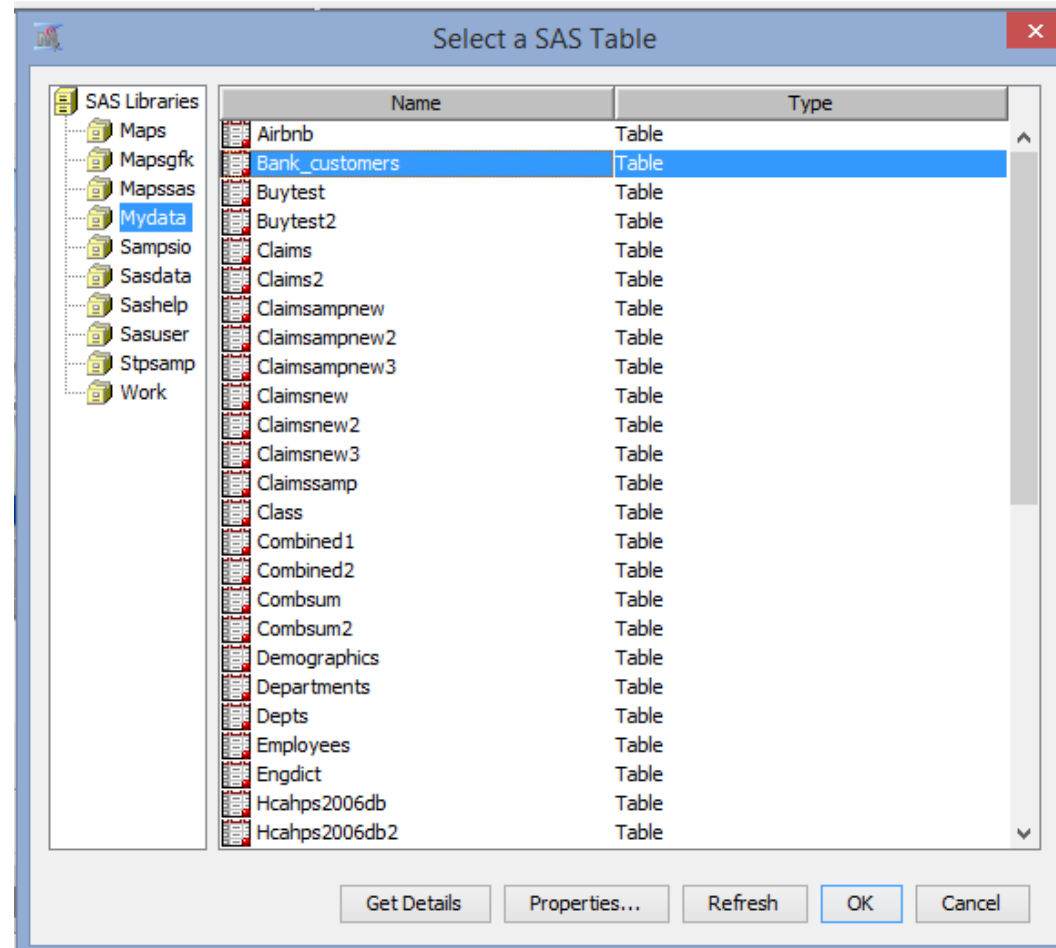
Click on “Browse”



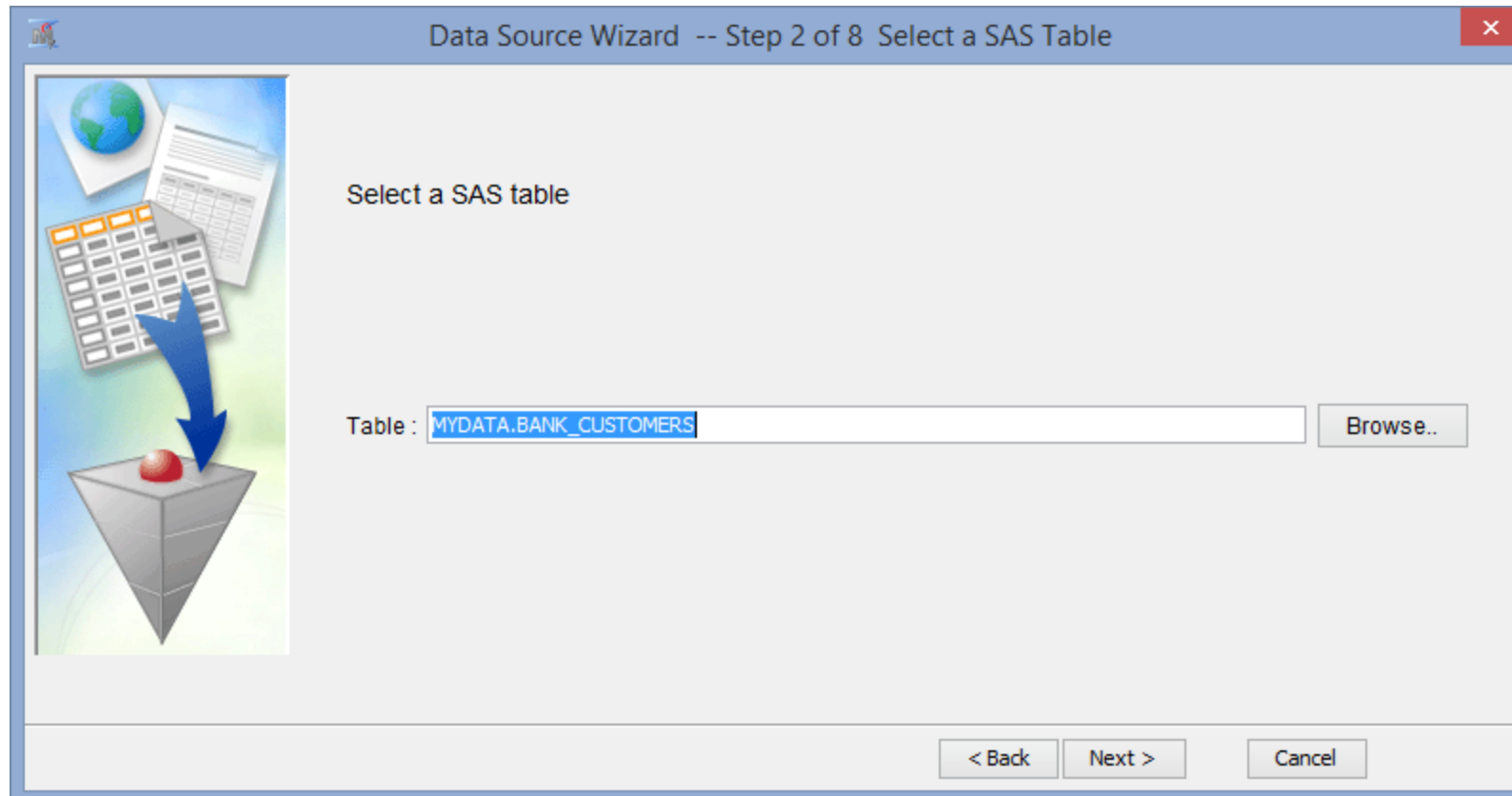
Double click on the libname you just set up in the project startup code.



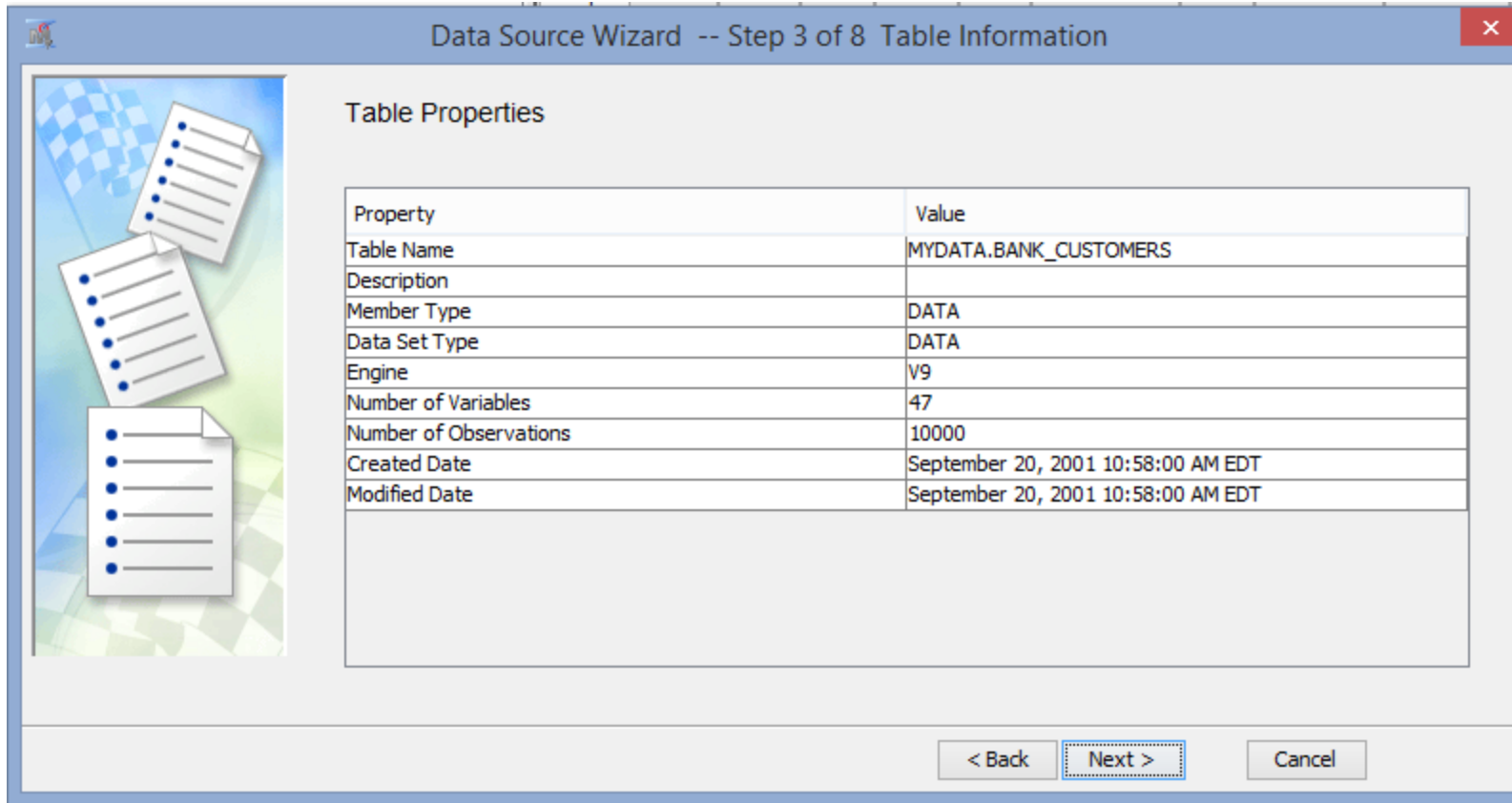
Double click to select the dataset for this unit, and click “OK”



Click “Next>”



Verify the options and click “Next>”



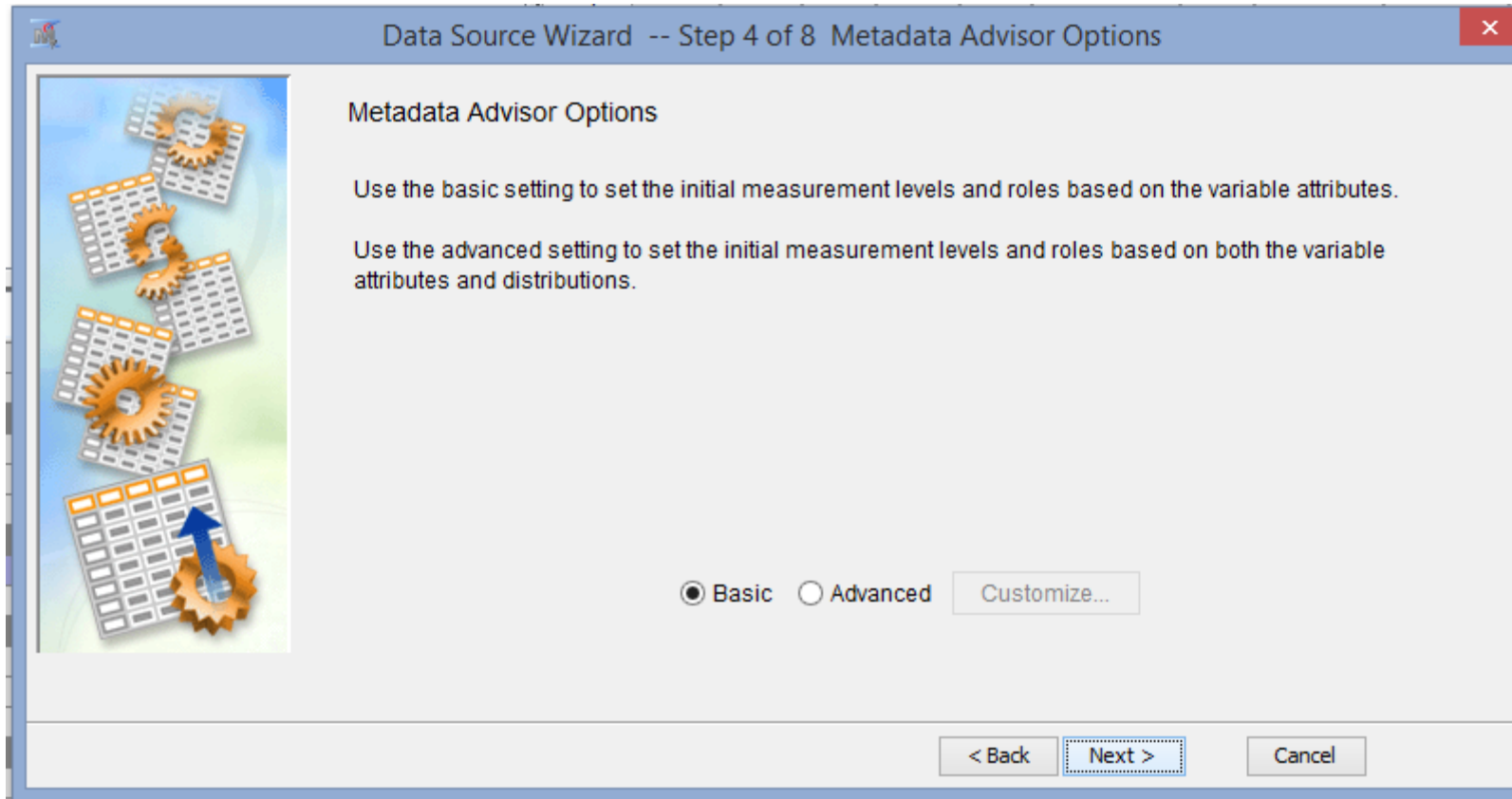
The screenshot shows a Windows-style dialog box titled "Data Source Wizard -- Step 3 of 8 Table Information". On the left is a decorative graphic of three overlapping document icons. The main area is titled "Table Properties" and contains a table with two columns: "Property" and "Value". The table lists various metadata for a 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 >", and "Cancel". The "Next >" button is highlighted with a dashed blue border, indicating it is the recommended action.

Property	Value
Table Name	MYDATA.BANK_CUSTOMERS
Description	
Member Type	DATA
Data Set Type	DATA
Engine	V9
Number of Variables	47
Number of Observations	10000
Created Date	September 20, 2001 10:58:00 AM EDT
Modified Date	September 20, 2001 10:58:00 AM EDT

< Back **Next >** Cancel



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 ☐ Apply

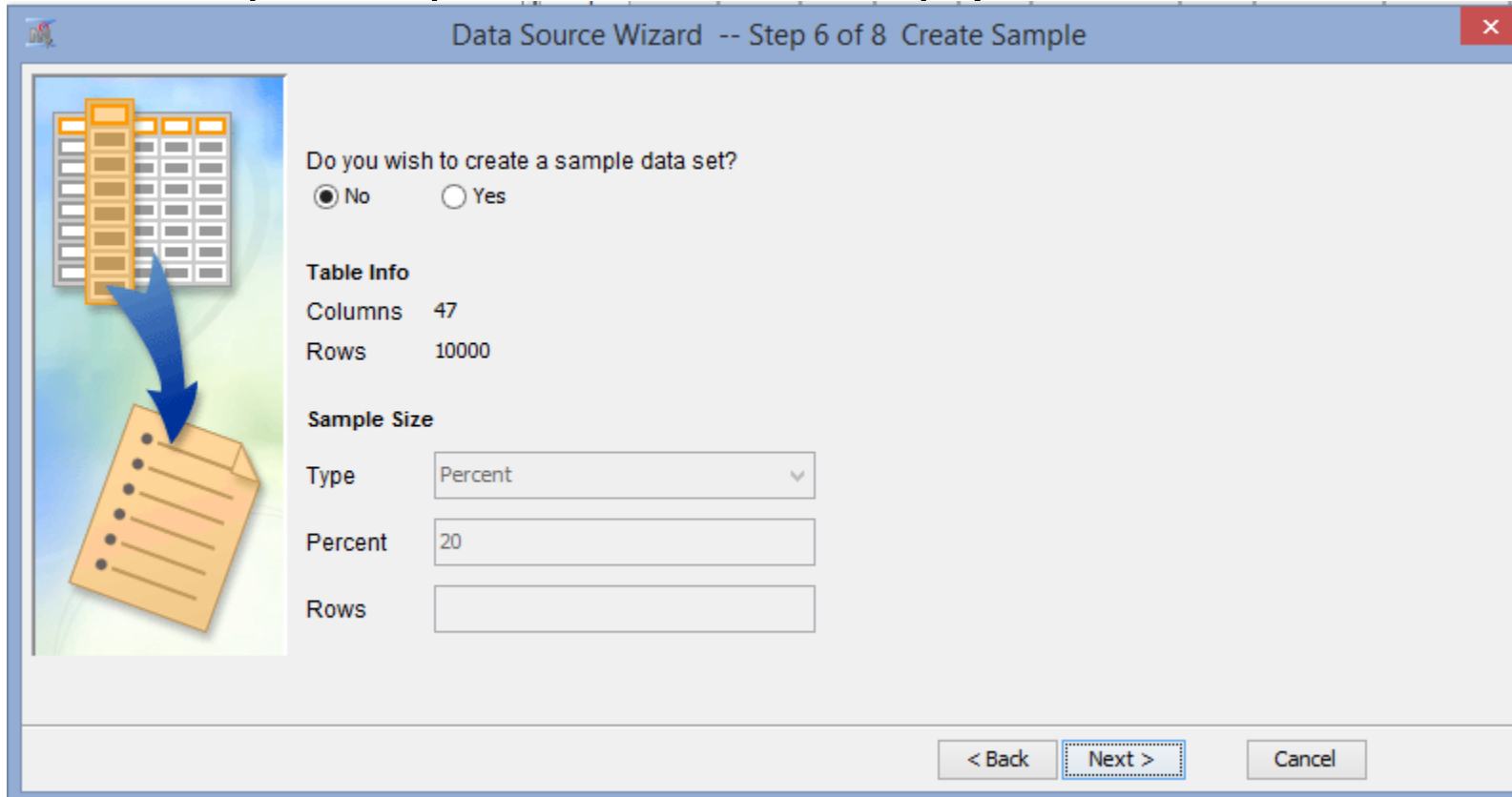
Columns: ☐ Label ☐ Mining ☐ Basic ☐ Statistics

Name	Role	Level	Report	Order	Drop	Lower Limit	Upp
ACCTAGE	Input	Interval	No		No	.	
AGE	Input	Interval	No		No	.	
ATM	Input	Interval	No		No	.	
ATMAMT	Input	Interval	No		No	.	
BRANCH	Input	Nominal	No		No	.	
CASHBK	Input	Interval	No		No	.	
CC	Input	Interval	No		No	.	
CCBAL	Input	Interval	No		No	.	
CCPURC	Input	Interval	No		No	.	
CD	Input	Interval	No		No	.	
CDBAL	Input	Interval	No		No	.	
CHECKS	Input	Interval	No		No	.	
CRSCORE	Input	Interval	No		No	.	
DDA	Input	Interval	No		No	.	
DDARAI	Input	Interval	No		No	.	

Show code Explore Compute Summary < Back Next > Cancel



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.



Data Source Wizard -- Step 6 of 8 Create Sample

Do you wish to create a sample data set?

☒ No ☐ Yes

Table Info

Columns 47

Rows 10000

Sample Size

Type

Percent

Percent

20

Rows

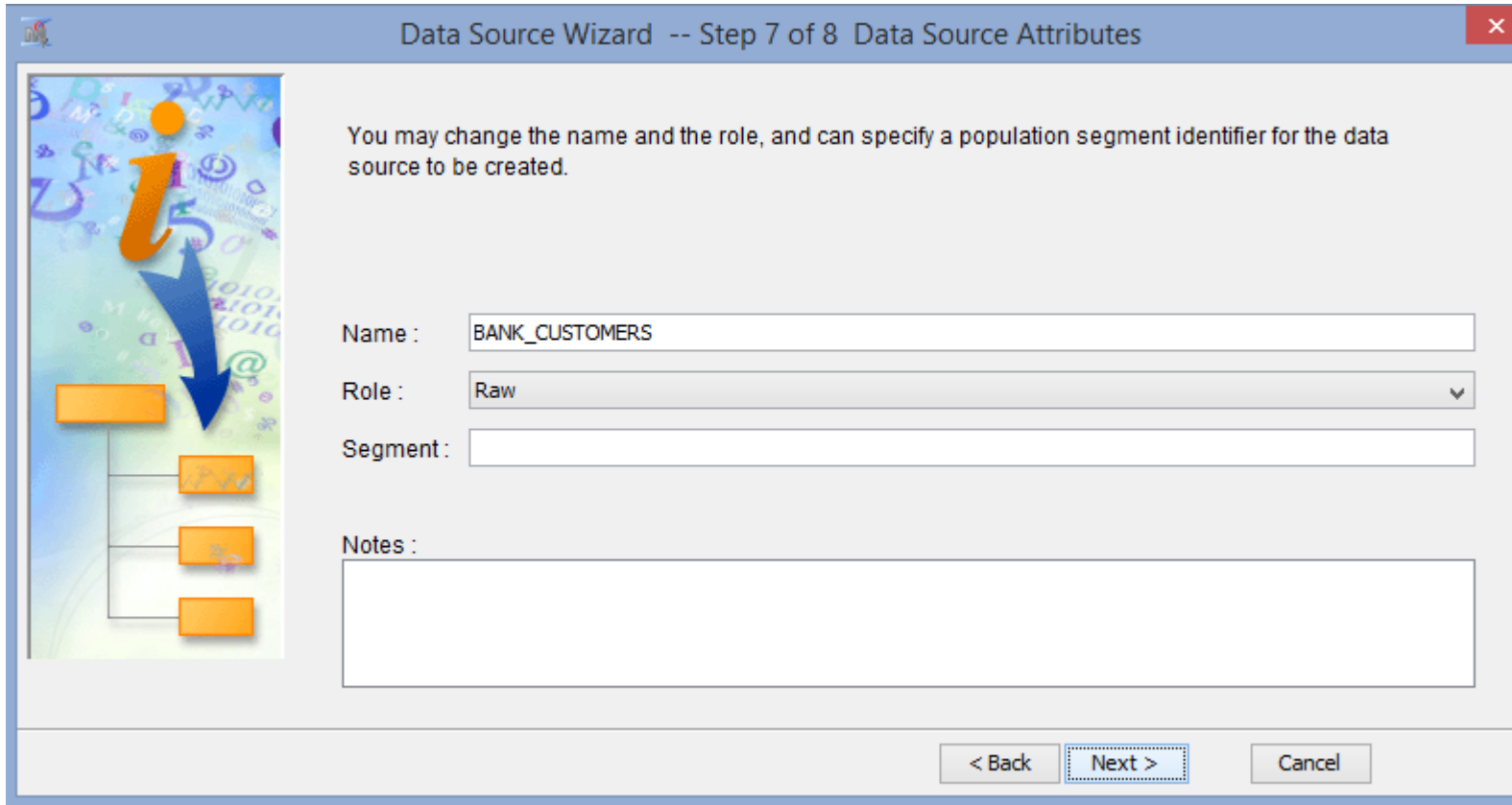
< Back

Next >

Cancel



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' window, specifically Step 7 of 8, titled 'Data Source Attributes'. The window has a blue title bar with a close button in the top right corner. On the left side, there is a vertical panel with a blue background featuring a large orange 'i' icon and a blue arrow pointing downwards. Below the arrow are three orange rectangular boxes connected by lines. The main area of the window is light gray and contains the following text and controls:

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

Name :

Role :

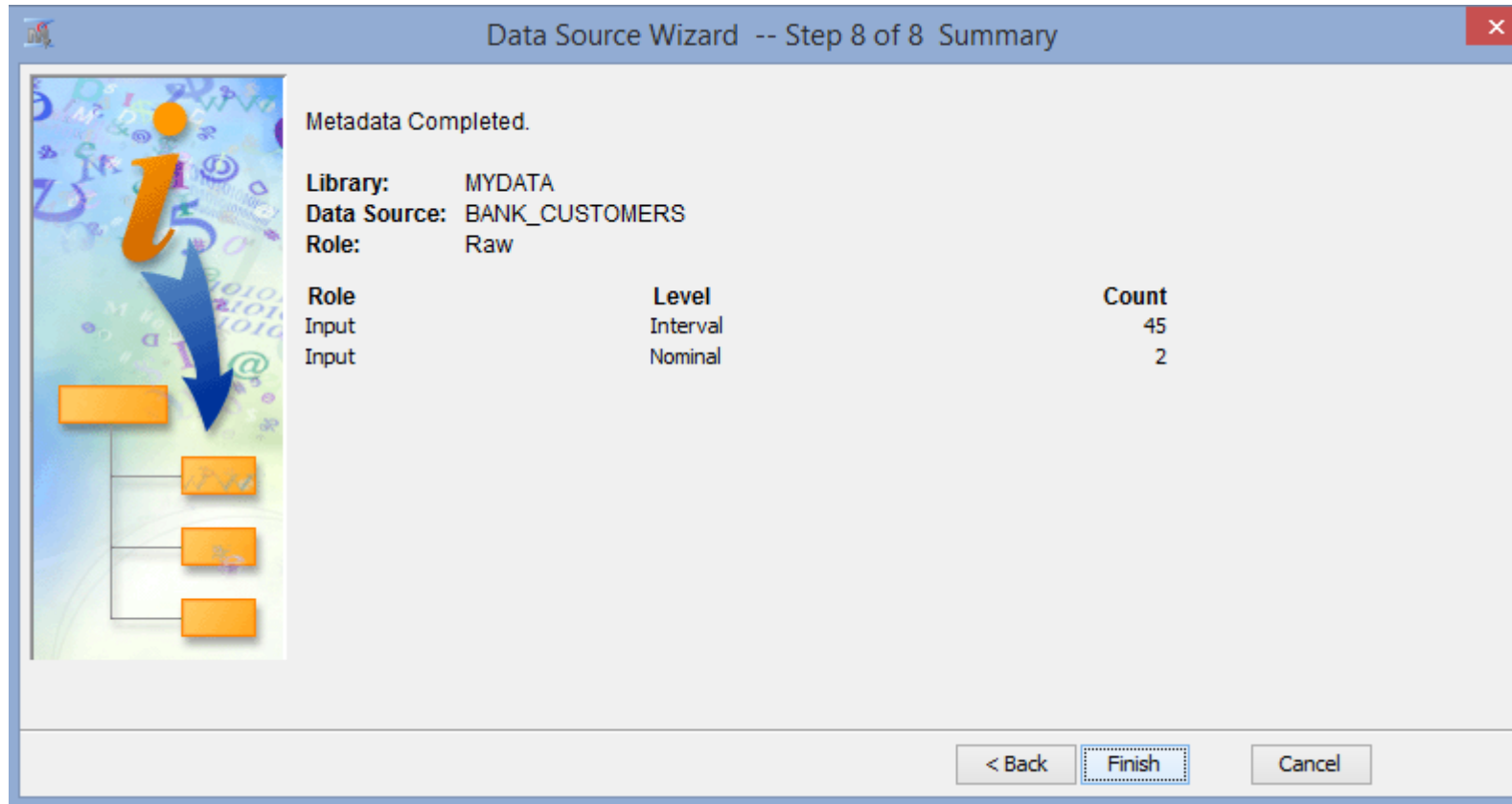
Segment :

Notes :

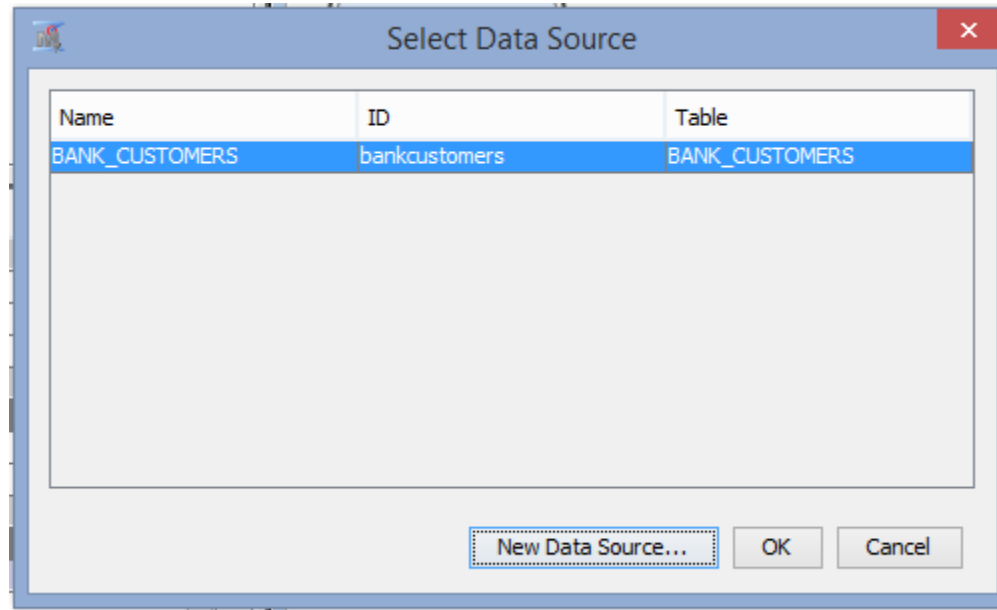
At the bottom right, there are three buttons: '< Back', 'Next >' (which is highlighted with a blue border), and 'Cancel'.



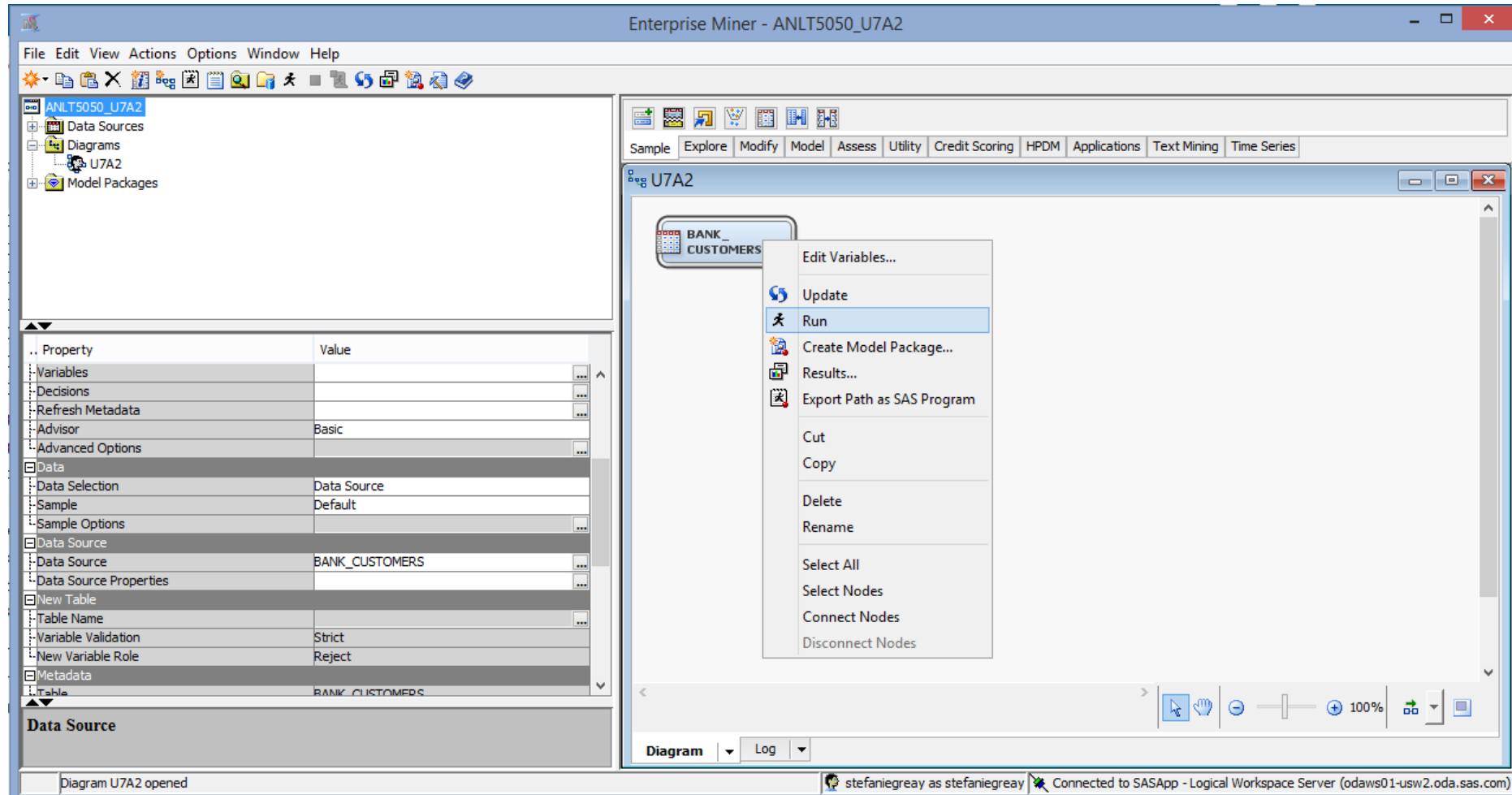
Click “Finish” to finish the data source registration within EM.



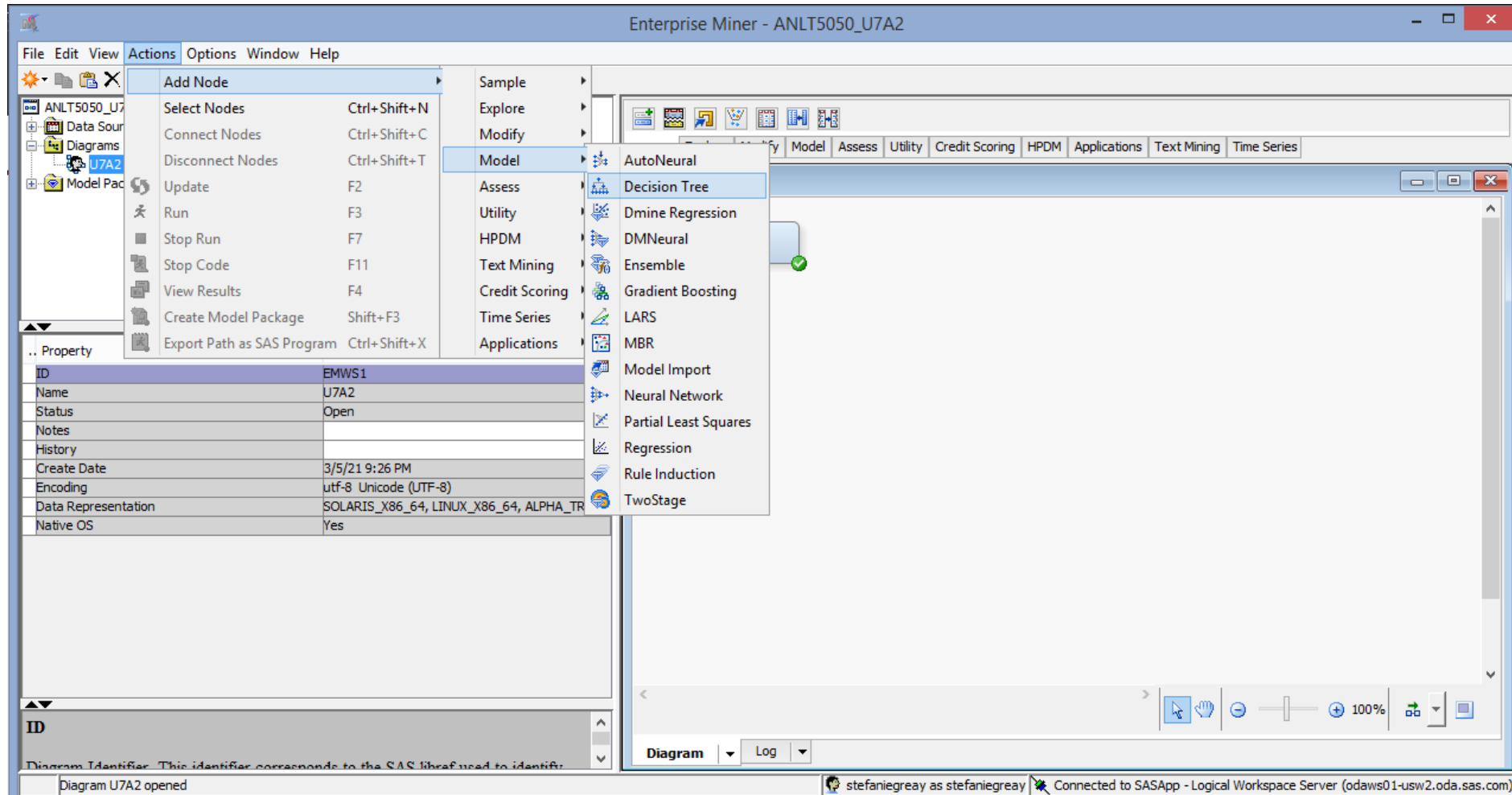
Click “OK” to complete the process. The name of the node should then change to the name of the dataset.



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



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



Connect the nodes

The screenshot displays the SAS Enterprise Miner interface. The main workspace shows a workflow diagram with two nodes: 'BANK_CUSTOMERS' (a data source icon) and 'Decision Tree' (a model icon). An arrow connects the output of 'BANK_CUSTOMERS' to the input of 'Decision Tree'. The 'Decision Tree' node is labeled 'Decision Tree' in a yellow box.

The left sidebar shows the project structure for 'ANLT5050_U7A2', including 'Data Sources', 'Diagrams', and 'Model Packages'. The 'U7A2' diagram is selected.

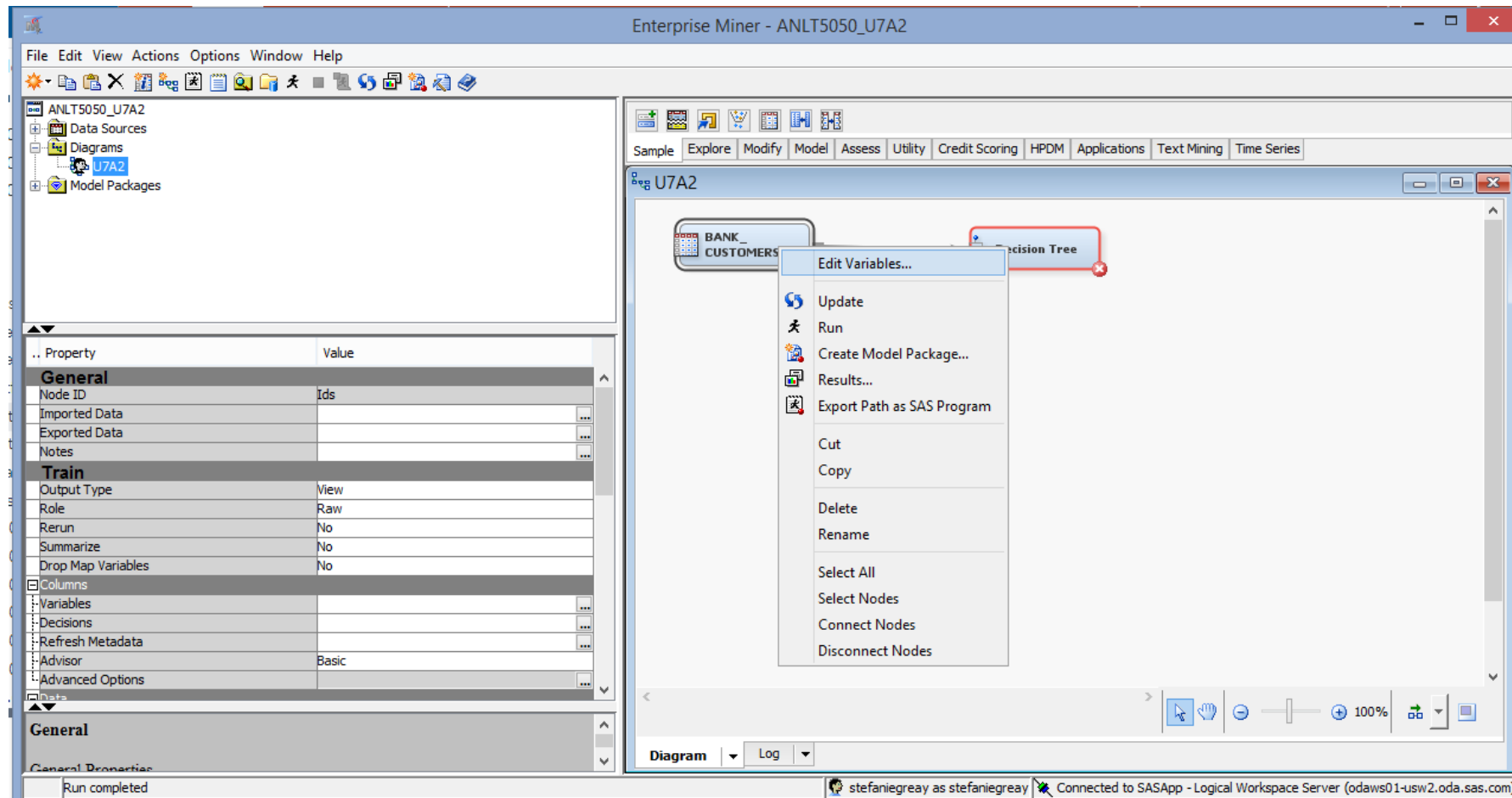
The bottom-left pane shows the properties for the selected 'Decision Tree' node. The 'General' tab is active, displaying the following properties:

Property	Value
General	
Node ID	Tree
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Interactive	
Import Tree Model	No
Tree Model Data Set	
Use Frozen Tree	No
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search

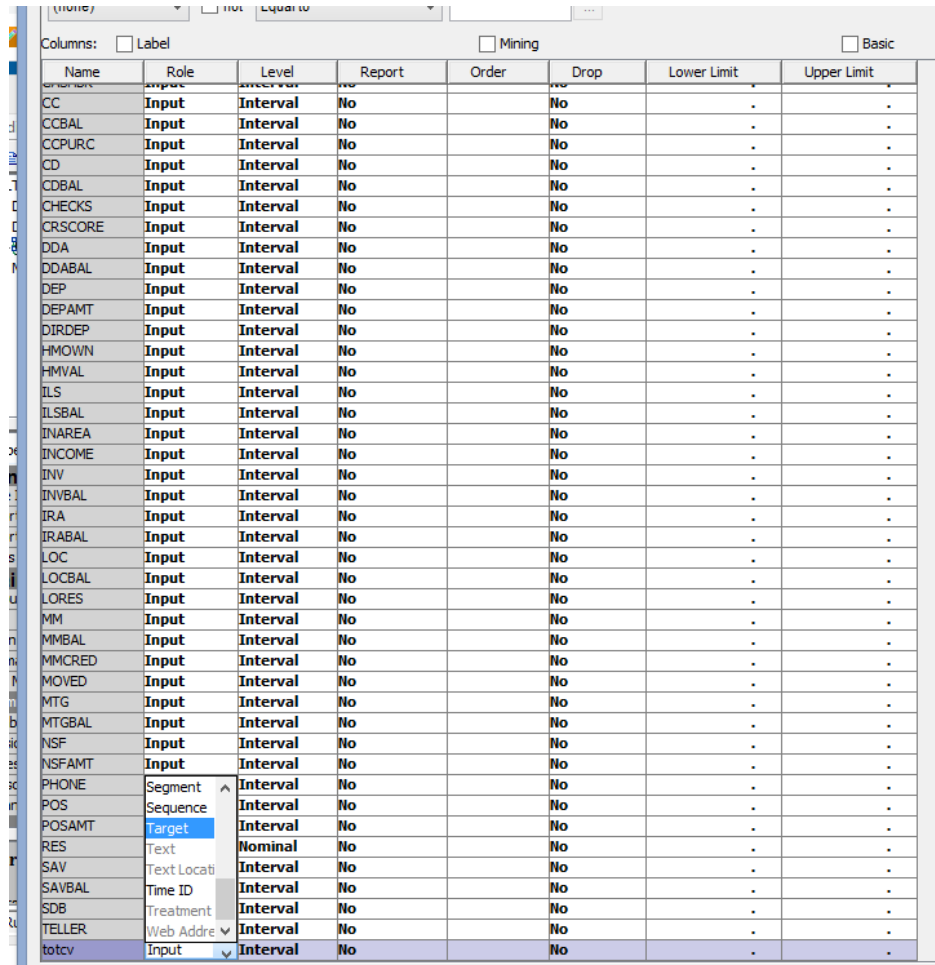
The bottom status bar indicates the user is 'stefaniegreay as stefaniegreay' and is 'Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)'.



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.”



Columns: ☐ Label ☐ Mining ☐ Basic

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
CC	Input	Interval	No		No	.	.
CCBAL	Input	Interval	No		No	.	.
CCPURC	Input	Interval	No		No	.	.
CD	Input	Interval	No		No	.	.
CDBAL	Input	Interval	No		No	.	.
CHECKS	Input	Interval	No		No	.	.
CRSCORE	Input	Interval	No		No	.	.
DDA	Input	Interval	No		No	.	.
DDABAL	Input	Interval	No		No	.	.
DEP	Input	Interval	No		No	.	.
DEPAMT	Input	Interval	No		No	.	.
DIRDEP	Input	Interval	No		No	.	.
HMOOWN	Input	Interval	No		No	.	.
HMVAL	Input	Interval	No		No	.	.
ILS	Input	Interval	No		No	.	.
ILSBAL	Input	Interval	No		No	.	.
INAREA	Input	Interval	No		No	.	.
INCOME	Input	Interval	No		No	.	.
INV	Input	Interval	No		No	.	.
INVBAL	Input	Interval	No		No	.	.
IRA	Input	Interval	No		No	.	.
IRABAL	Input	Interval	No		No	.	.
LOC	Input	Interval	No		No	.	.
LOCBAL	Input	Interval	No		No	.	.
LORES	Input	Interval	No		No	.	.
MM	Input	Interval	No		No	.	.
MMBAL	Input	Interval	No		No	.	.
MMCREC	Input	Interval	No		No	.	.
MOVED	Input	Interval	No		No	.	.
MTG	Input	Interval	No		No	.	.
MTGBAL	Input	Interval	No		No	.	.
NSF	Input	Interval	No		No	.	.
NSFAMT	Input	Interval	No		No	.	.
PHONE	Segment	Interval	No		No	.	.
POS	Sequence	Interval	No		No	.	.
POSAMT	Target	Interval	No		No	.	.
RES	Text	Nominal	No		No	.	.
SAV	Text Locati	Interval	No		No	.	.
SAVBAL	Time ID	Interval	No		No	.	.
SDB	Treatment	Interval	No		No	.	.
TELLER	Web Addre	Interval	No		No	.	.
totcv	Input	Interval	No		No	.	.



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

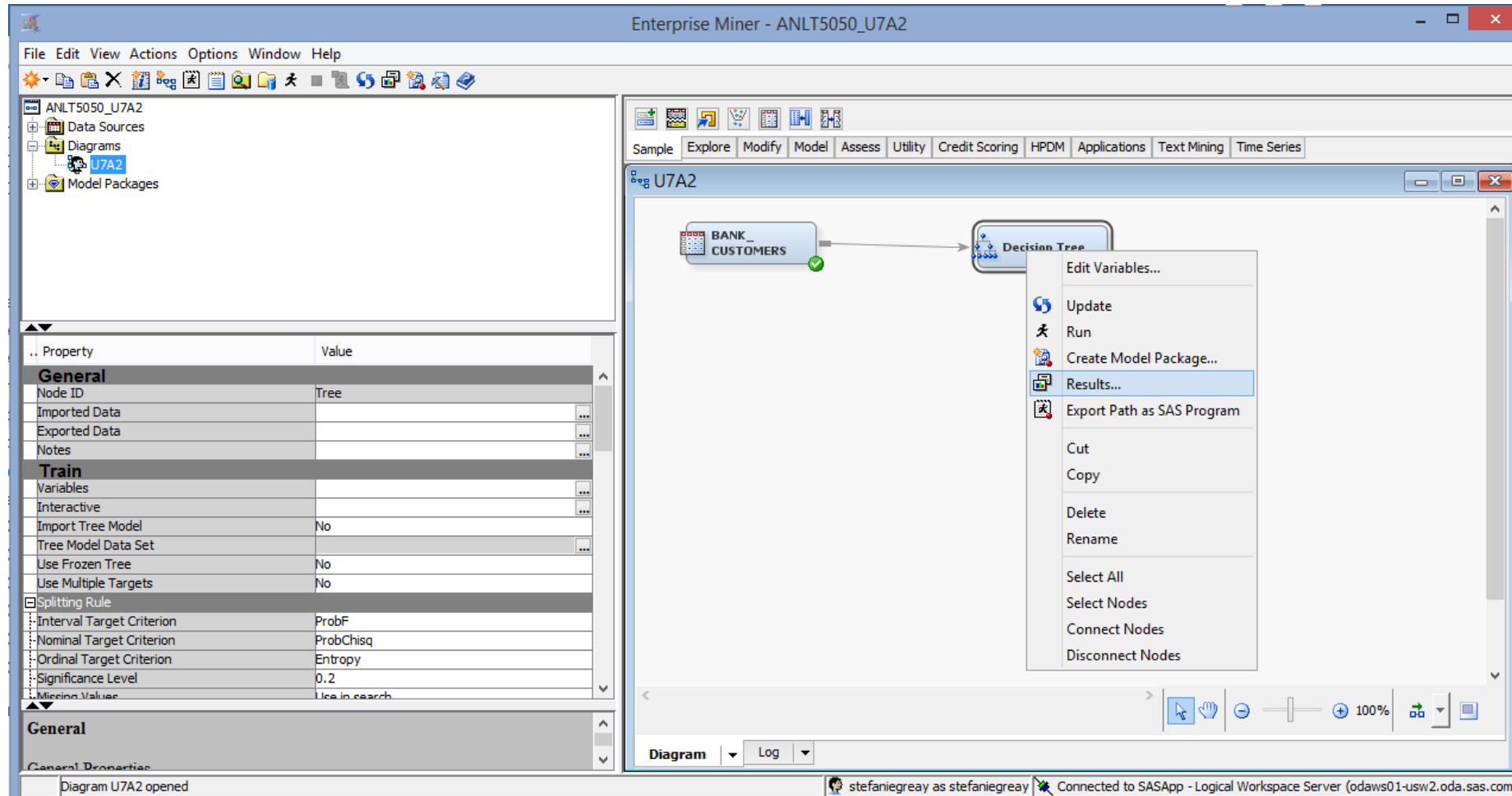
The screenshot displays the SAS Enterprise Miner interface. On the left, a project tree shows 'ANLT5050_U7A2' with sub-items 'Data Sources', 'Diagrams', 'U7A2', and 'Model Packages'. Below this is a properties table for the selected 'Decision Tree' node.

Property	Value
General	
Node ID	Tree
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Interactive	
Import Tree Model	No
Tree Model Data Set	
Use Frozen Tree	No
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search

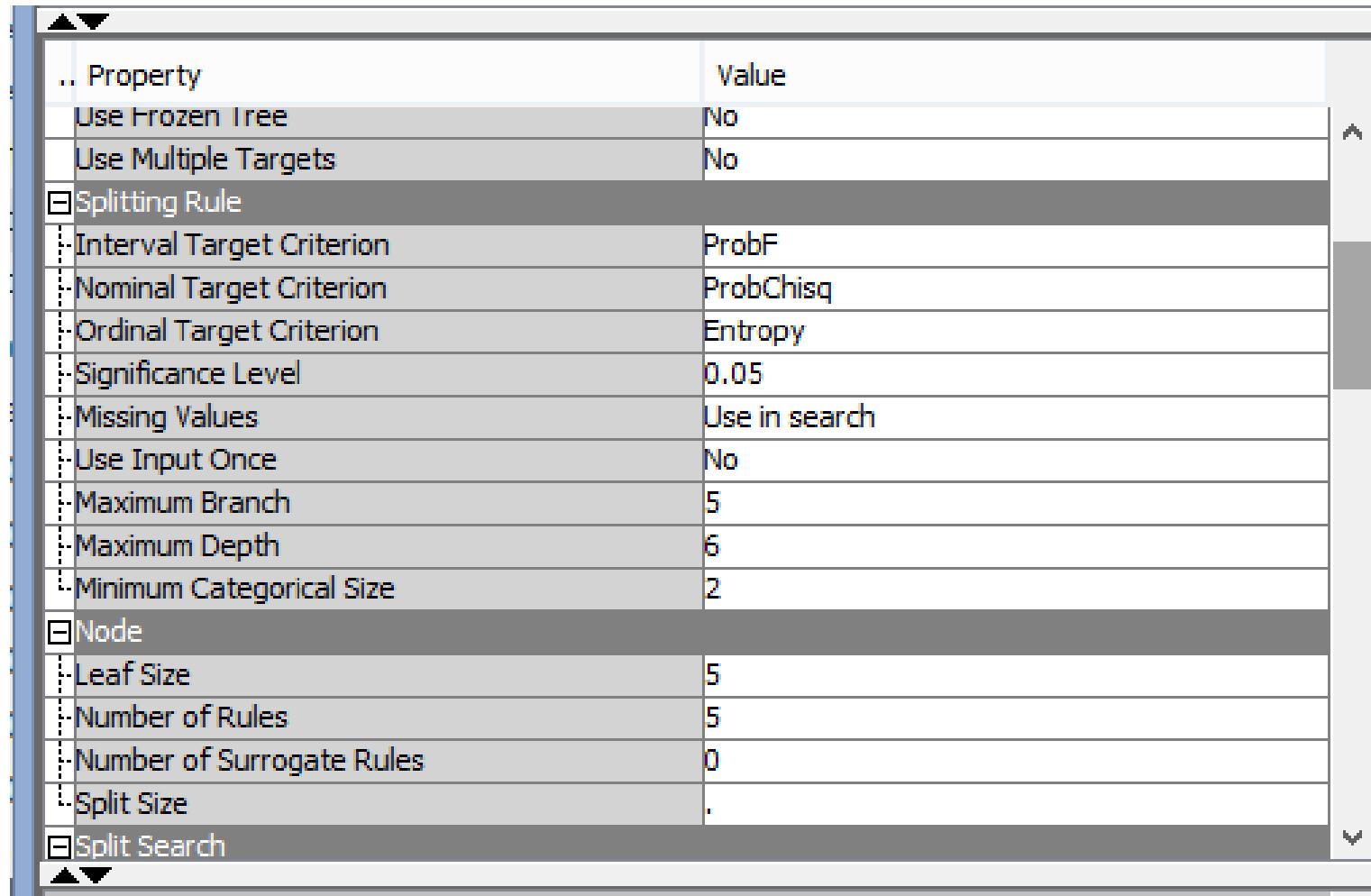
The main workspace shows a diagram with a 'BANK_CUSTOMERS' data source node connected to a 'Decision Tree' model node. A right-click context menu is open over the 'Decision Tree' node, with the 'Run' option highlighted. Other menu options include 'Edit Variables...', 'Update', 'Create Model Package...', 'Results...', 'Export Path as SAS Program', 'Cut', 'Copy', 'Delete', 'Rename', 'Select All', 'Select Nodes', 'Connect Nodes', and 'Disconnect Nodes'. The status bar at the bottom indicates the user is 'stefaniegreay' and is connected to the 'SASApp - Logical Workspace Server'.



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).



The screenshot shows a software window with a table of options for splitting criteria. The table has two columns: 'Property' and 'Value'. The options are grouped into sections: 'Splitting Rule', 'Node', and 'Split Search'. The 'Splitting Rule' section includes '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), and 'Minimum Categorical Size' (2). The 'Node' section includes 'Leaf Size' (5), 'Number of Rules' (5), 'Number of Surrogate Rules' (0), and 'Split Size' (.). The 'Split Search' section is currently empty.

Property	Value
Use Frozen Tree	No
Use Multiple Targets	No
<input checked="" type="checkbox"/> 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
<input checked="" type="checkbox"/> Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rules	0
Split Size	.
<input checked="" type="checkbox"/> Split Search	

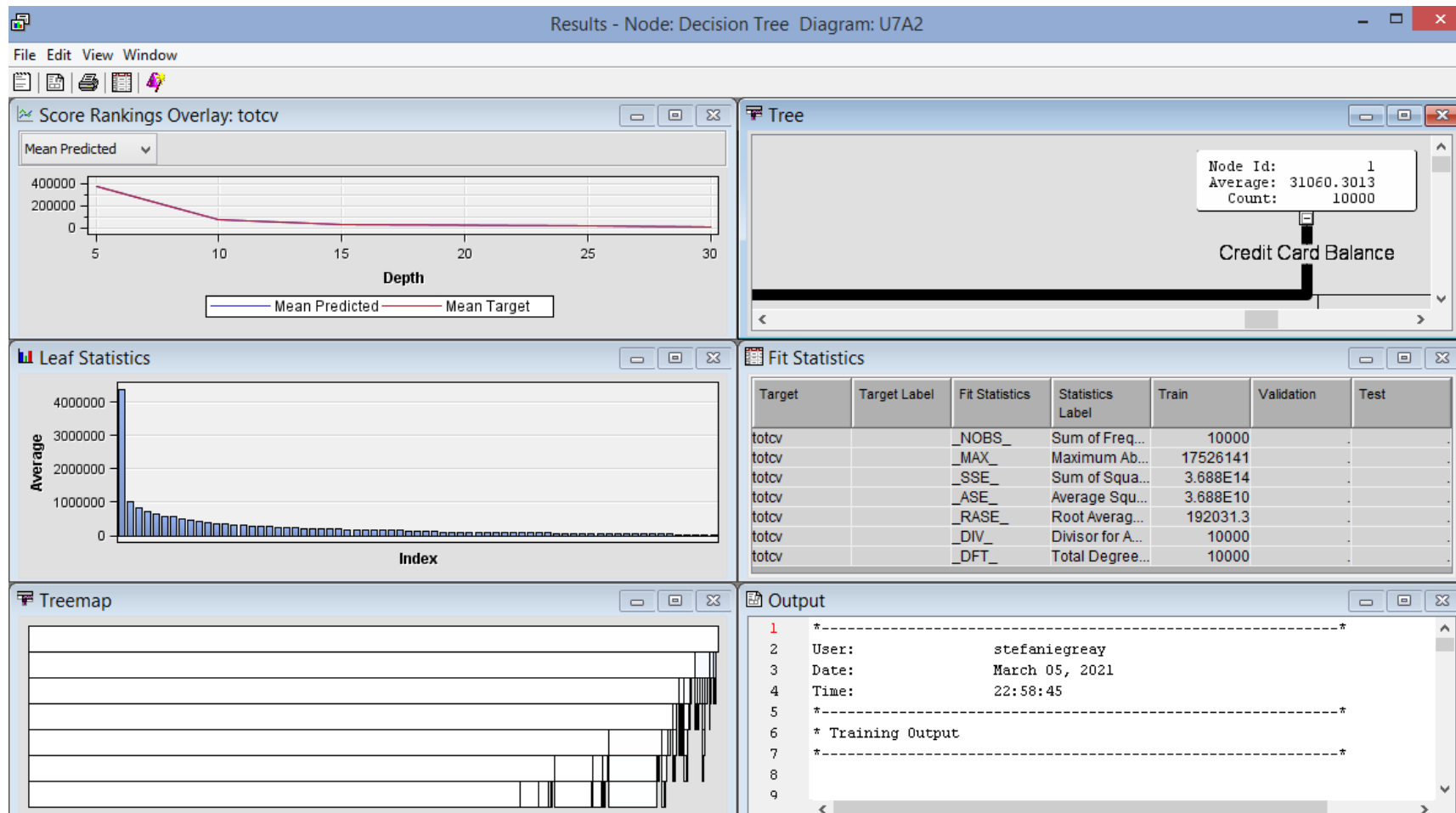


Considerations for splitting options

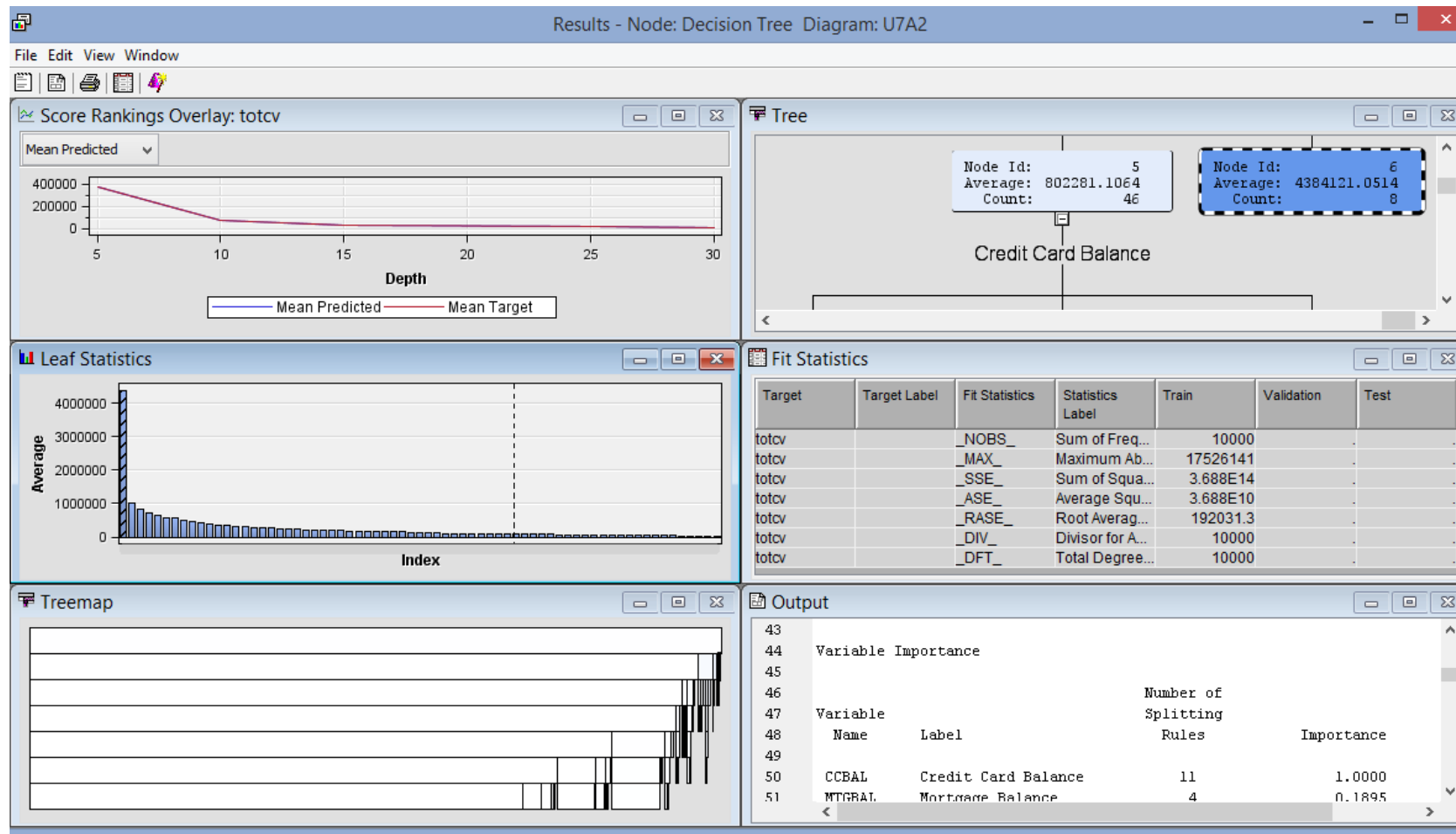
- Be sure that you understand the system constraints of whatever system you plan to implement the decision tree into, and make the splitting options fit those constraints.
- If you have categorical variables, like binary variables, be sure that you set the minimum categories setting accordingly so you 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).



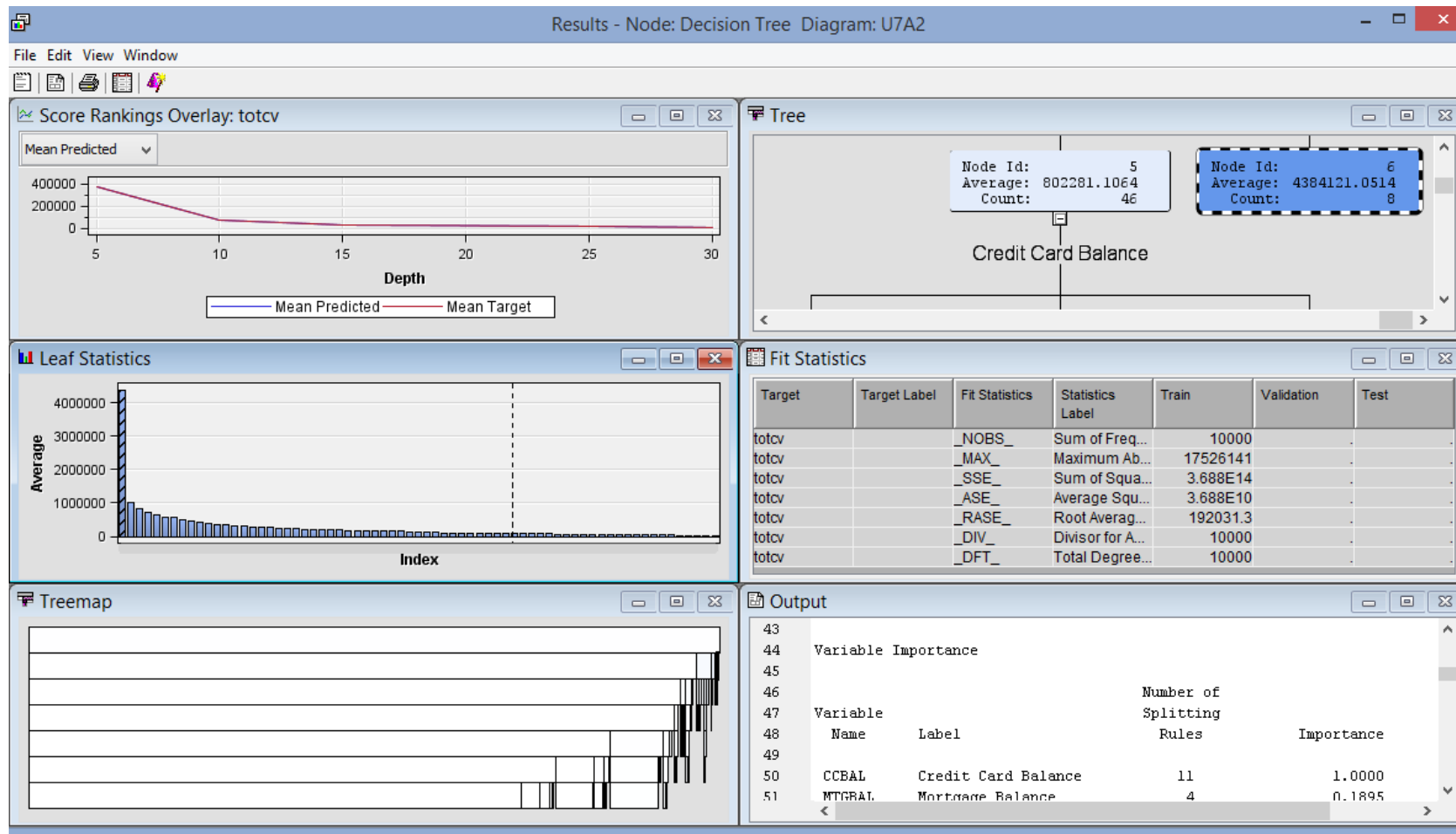
After you re-run the decision tree node (after making any adjustments to the split settings), right click on the node and select “Results.”



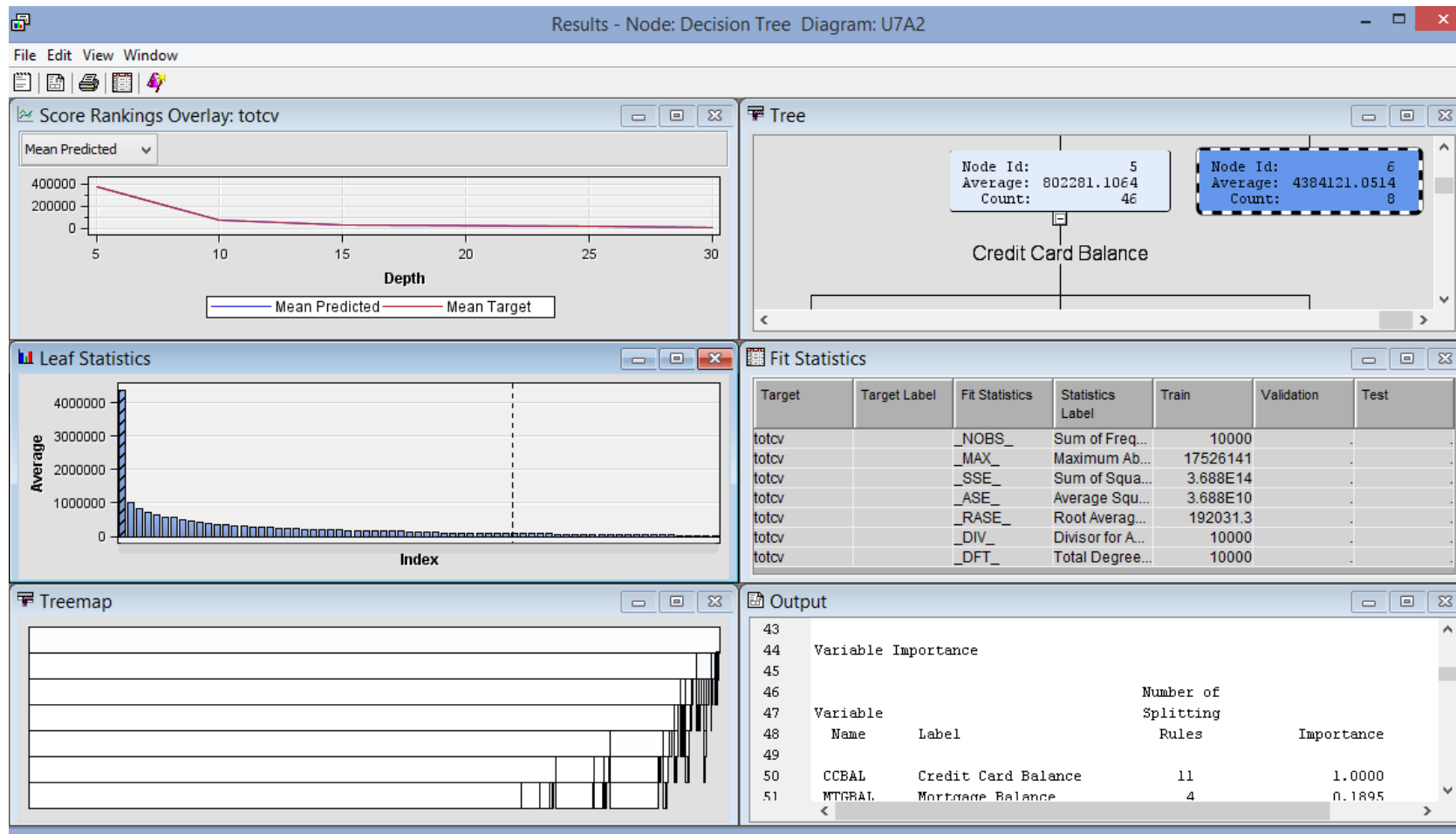
If you click on the bars in the “Leaf Statistics” window, it finds the appropriate leaf in the tree window and highlights it. You can then follow the logic that goes into that leaf by following it up the tree.



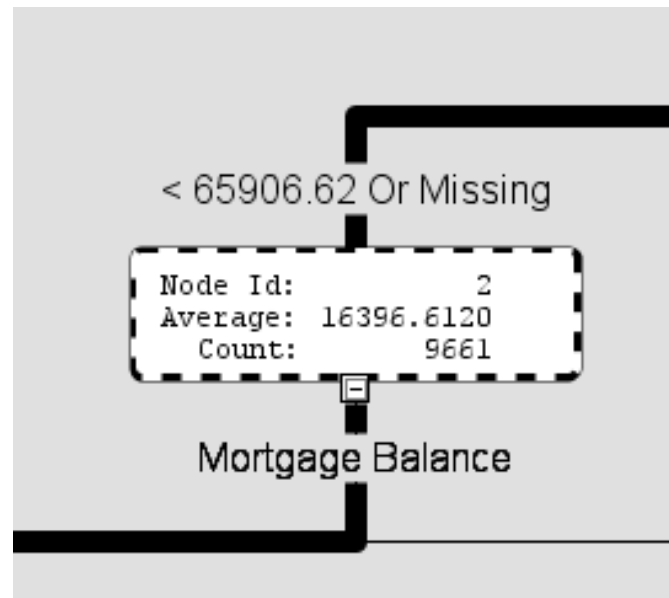
The Output window has the variable importance table that shows what variables contributed to the splitting/tree creation, as well as the leaf report.



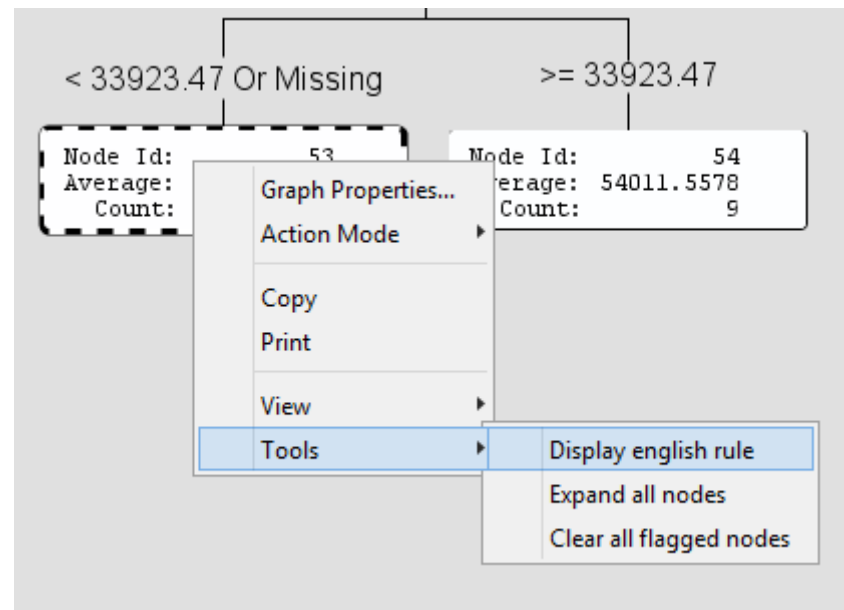
The Tree pane shows a graphical display of the tree, including the splitting criteria, count of observations, leaf identifier and the average of the target variable. This can be used to summarize the rules.



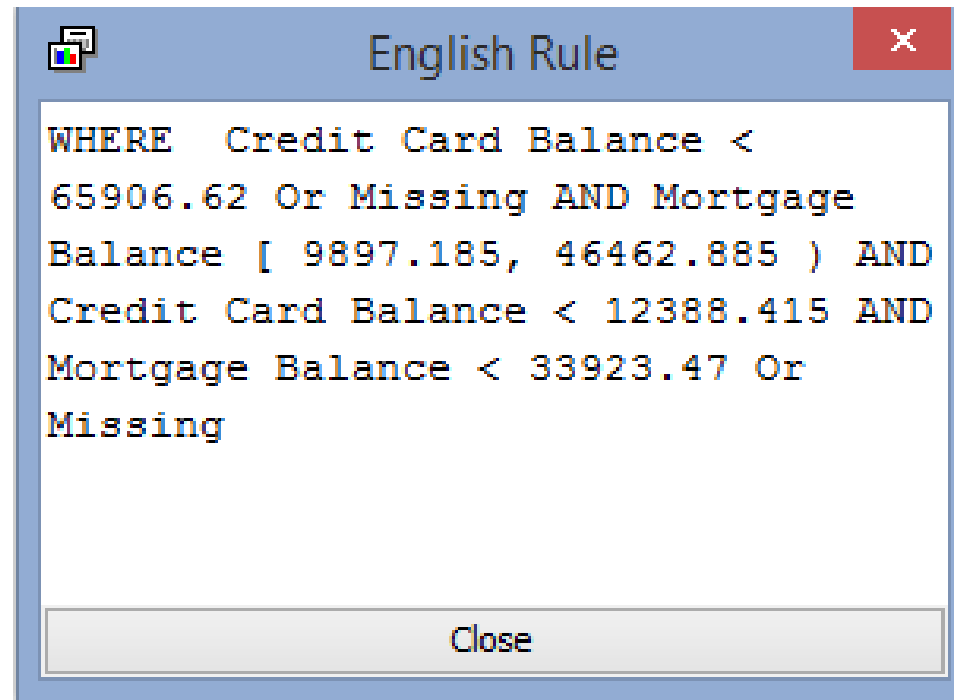
An example of interpreting one of the intermediate nodes. The Node Id is a reference for this node in the other output and results, the average is the average of the totcv (total customer value) variable we are using as the target variable. The count is the number of records or observations that fall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



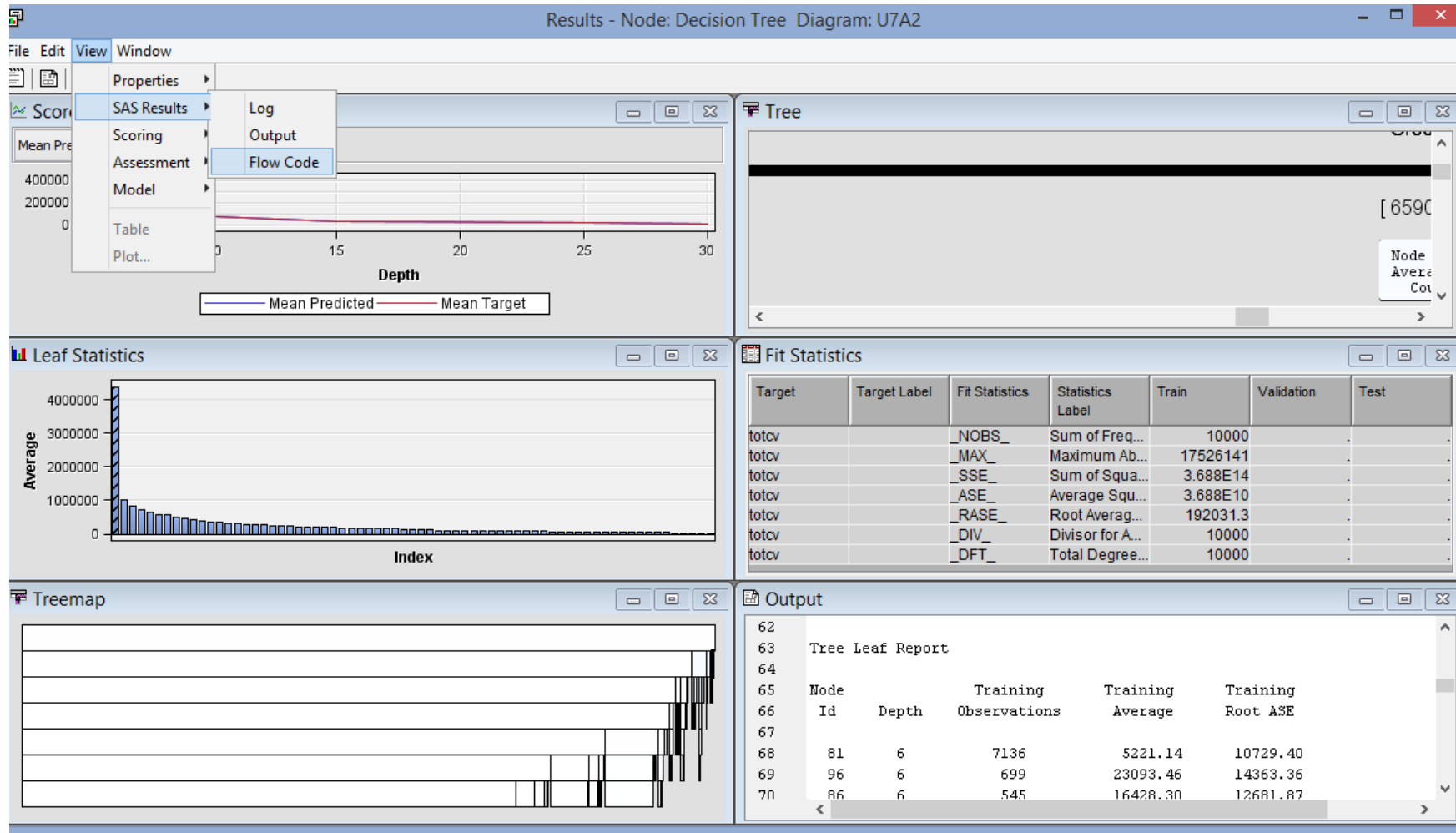
An example of interpreting one of the end nodes. The Node Id is a reference for this node in the other output and results, the average is the average of the totcv (total customer value) variable we are using as the target variable. The count is the number of records or observations that fall into the 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=n0cx4ud03paymdn1kargegadueml.htm&docsetVersion=14.3&locale=en>

