

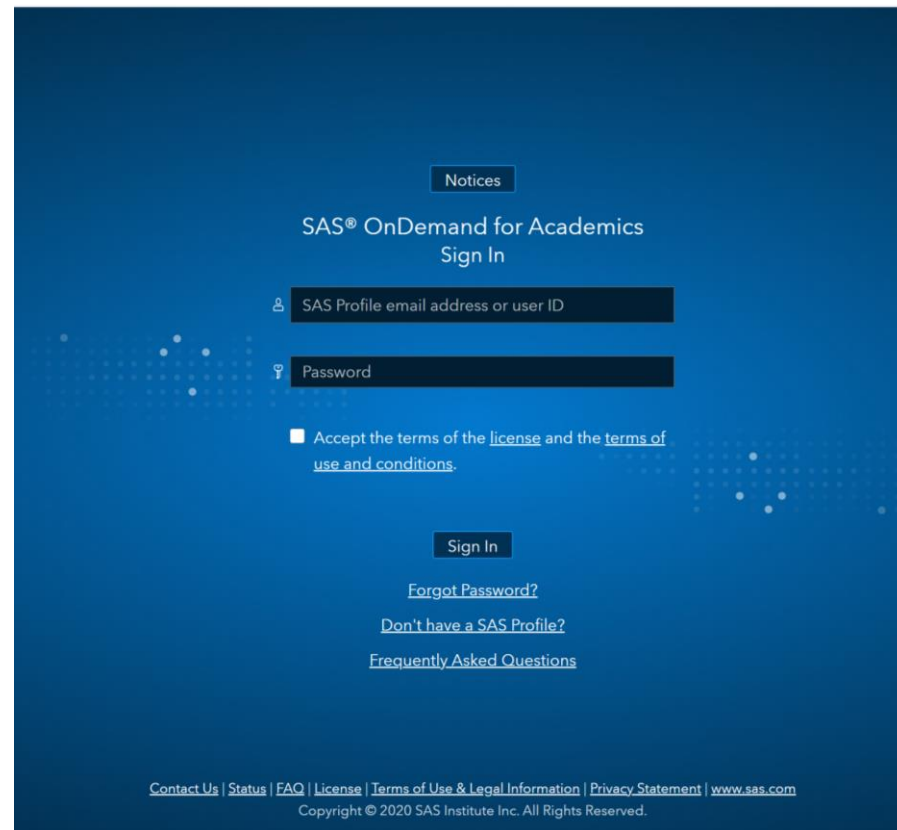
ANLT5070

Unit 8 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 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".



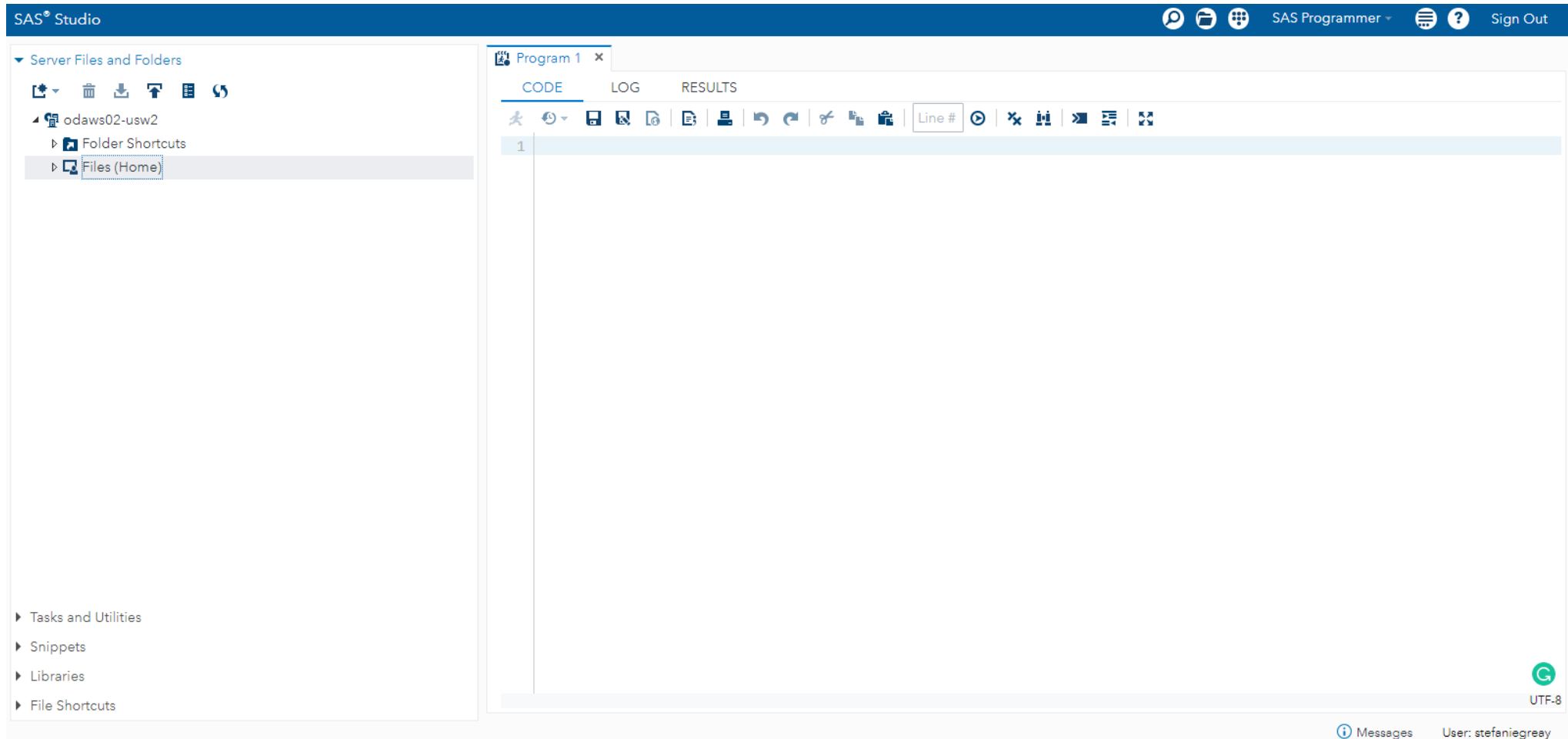
Open the csv file and save it as xlsx format.

If you upload and import the csv file, you will notice that the columns do not all align and that it creates extra rows. This is because the text in some of the fields contains commas, and these are treated as additional delimiters in the import process.

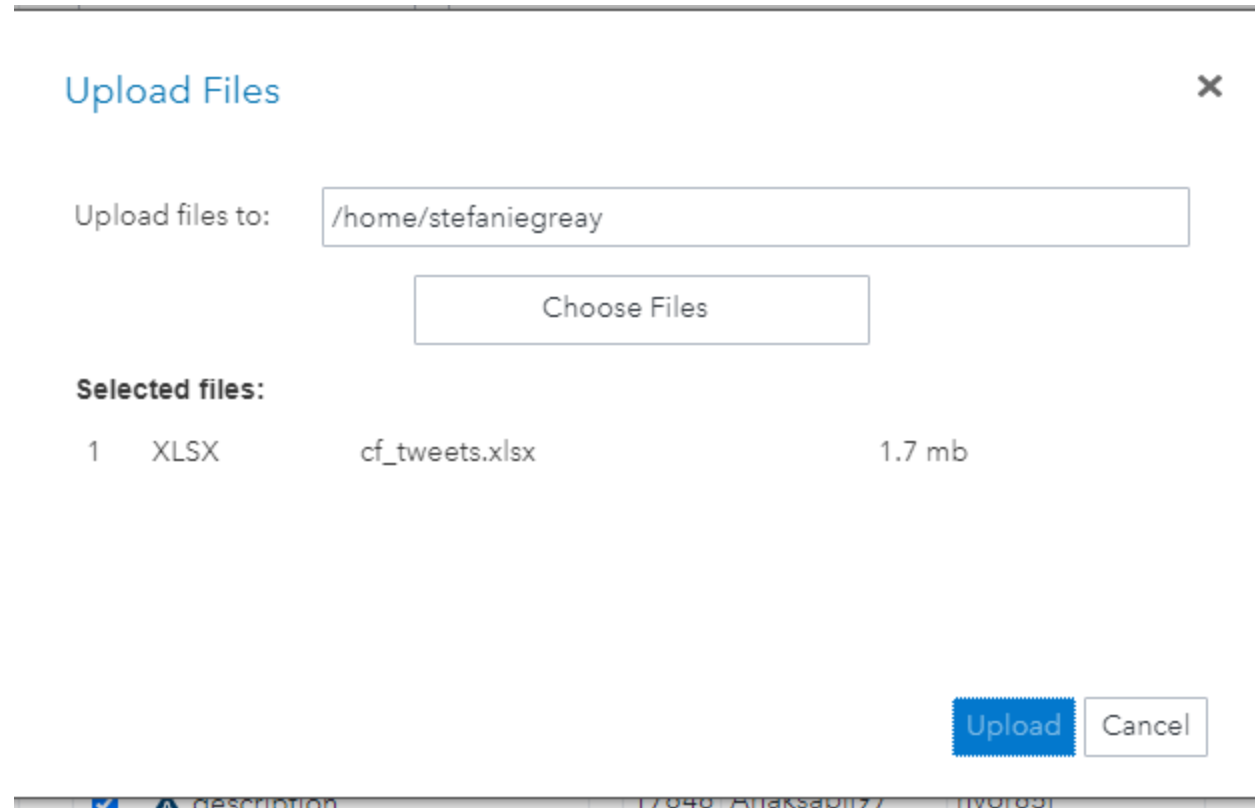
The easiest way to resolve this is to open the csv file in MS Excel (since it is so small, this is perfectly fine), and save it as MS Excel format (i.e. .xlsx extension or format), then upload and import it.



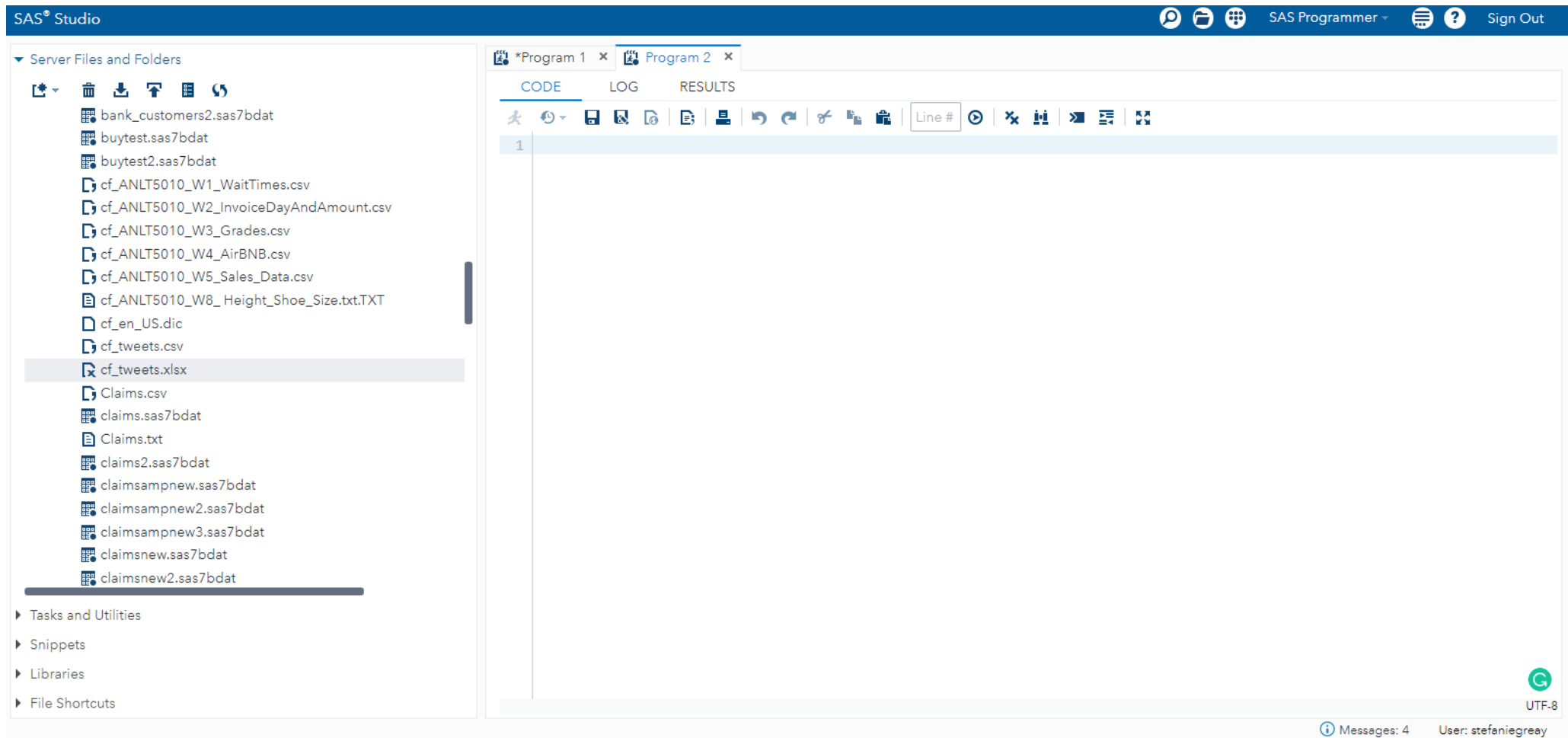
To upload the dataset to the SAS server, open SAS Studio, then click on “Files (Home)” and click the upload button.



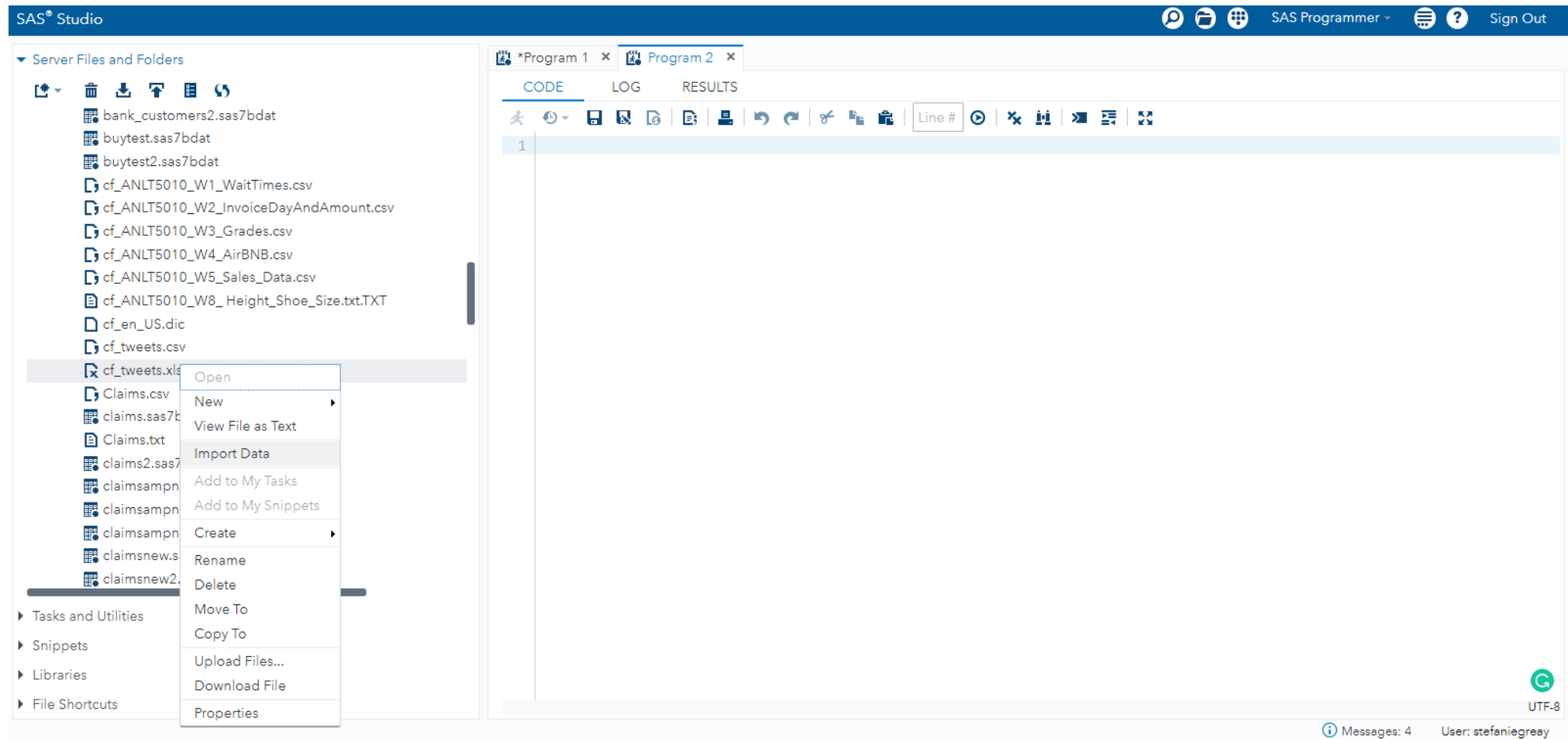
Click on “Choose Files” to browse to the file you want to upload, then click “Upload.”



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



Right click on the cf_tweets.xlsx file and choose “Import Data.”



Click the run guy to run the import process.

The screenshot displays the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'cf_tweets.xlsx' selected. The main editor area shows the 'FILE INFORMATION' tab for 'cf_tweets.xlsx', indicating the source location is '/home/stefaniegreay'. Below this, the 'OUTPUT DATA' section shows the SAS server is 'SASApp', the data set name is 'IMPORT1', and the library is 'WORK'. The 'CODE' tab is active, displaying the following SAS code:

```
1 /* Generated Code (IMPORT) */  
2 /* Source File: cf_tweets.xlsx */  
3 /* Source Path: /home/stefaniegreay */  
4 /* Code generated on: 3/7/21, 11:08 AM */  
5  
6 %web_drop_table(WORK.IMPORT1);  
7  
8  
9 FILENAME REFFILE '/home/stefaniegreay/cf_tweets.xlsx';  
10  
11 PROC IMPORT DATAFILE=REFFILE  
12
```

At the top of the editor, there is a 'Run (F3)' button. The bottom status bar indicates 'Messages: 4' and 'User: stefaniegreay'.



Open a new SAS program.

The screenshot displays the SAS Studio web interface. On the left, the 'Server Files and Folders' pane lists various files, with 'cf_tweets.csv' selected. The main workspace shows the 'FILE INFORMATION' for 'cf_tweets.csv', including its source location and output data settings. A context menu is open over the workspace, offering options like 'New SAS Program (F4)', 'New Import Data', and 'New Query'. The 'RESULTS' tab is active, displaying a 'Table of Contents' for 'The CONTENTS Procedure'.

FILE INFORMATION

SOURCE FILE

File name: **cf_tweets.csv**
Source location: **/home/stefaniegreay**
End of line delimiter: **Default**

OUTPUT DATA

SAS server: **SASApp**
Data set name: **IMPORT**
Library: **WORK**

RESULTS

Table of Contents

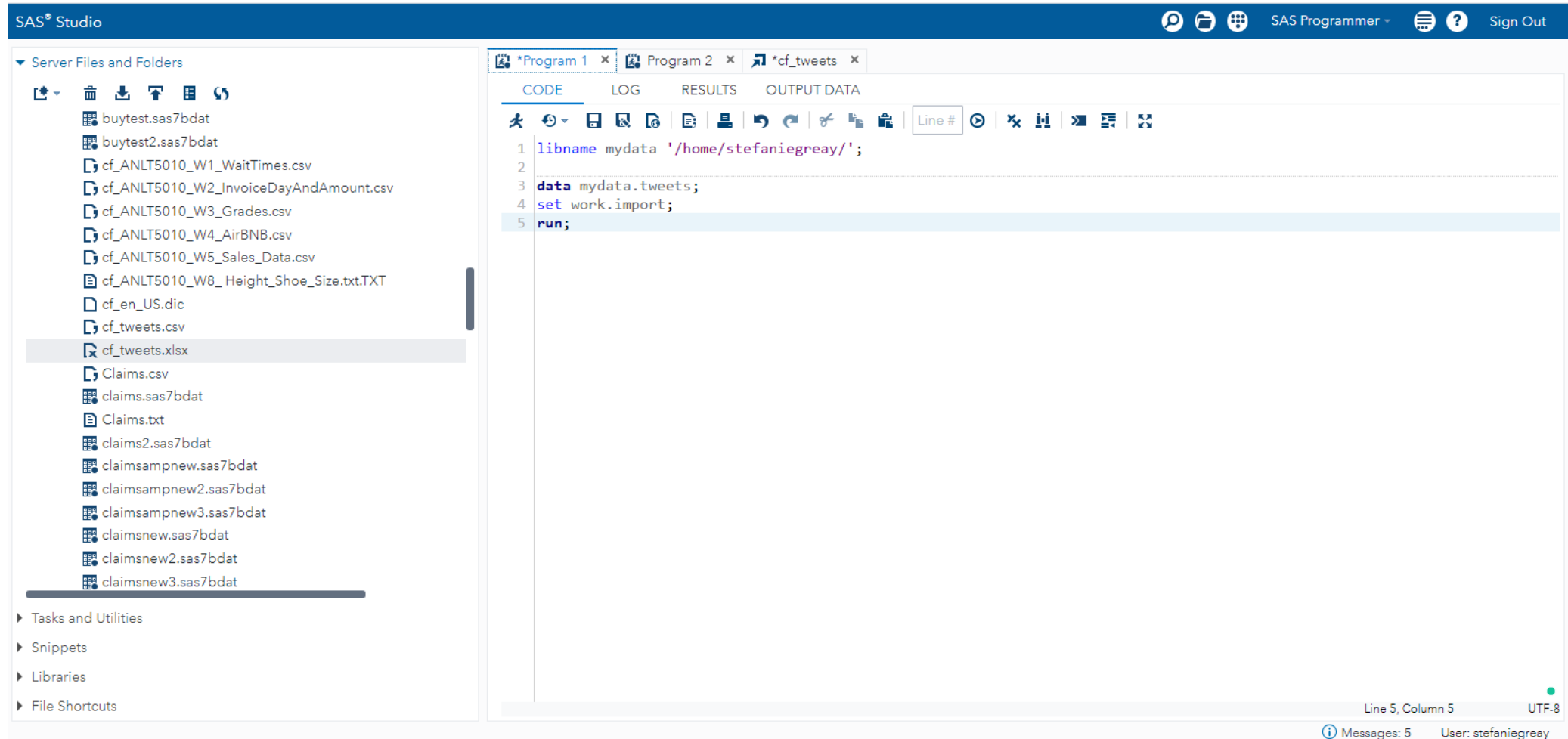
The CONTENTS Procedure

Data Set Name	WORK.IMPORT	Observations	28955
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	03/07/2021 10:58:24	Observation Length	272
Last Modified	03/07/2021 10:58:24	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			

Messages: 1 User: stefaniegreay



Write a libname statement and data step to save the imported data into a permanent dataset (and run it).



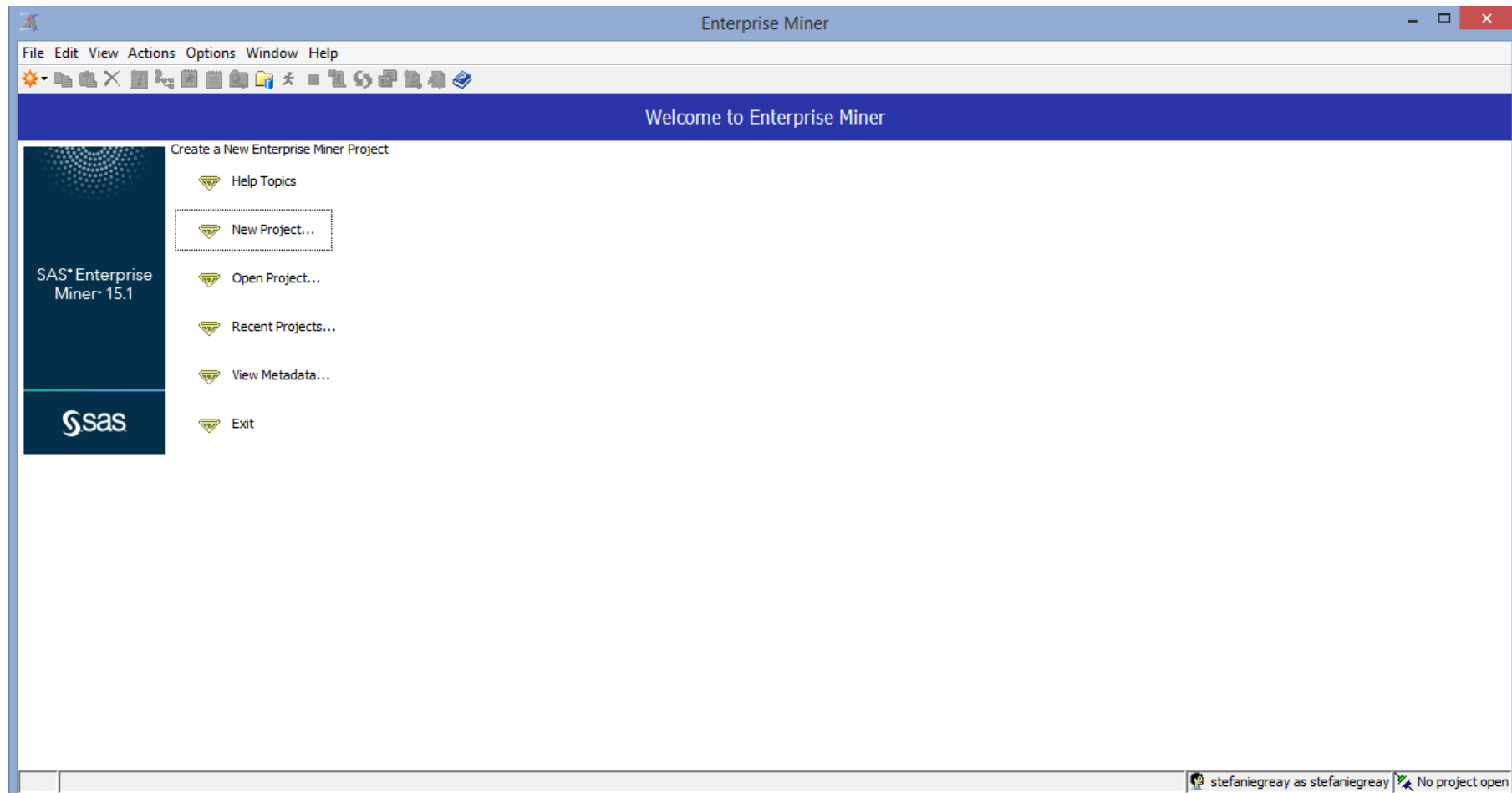
SAS Enterprise Miner Instructions

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

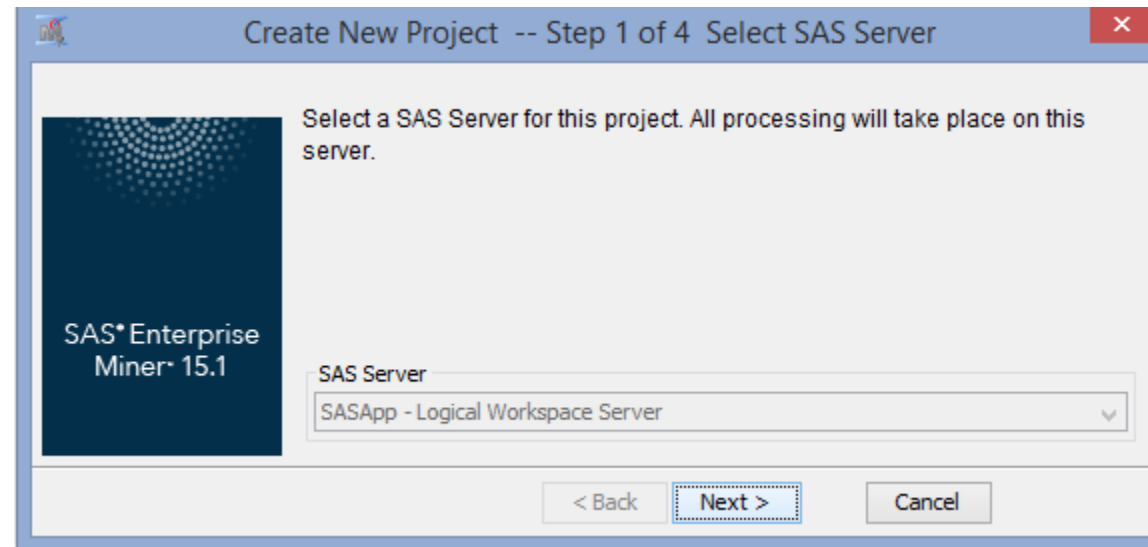
Once you have uploaded the dataset for this unit onto the SAS servers using SAS Studio, you may proceed from here using SAS Enterprise Miner.



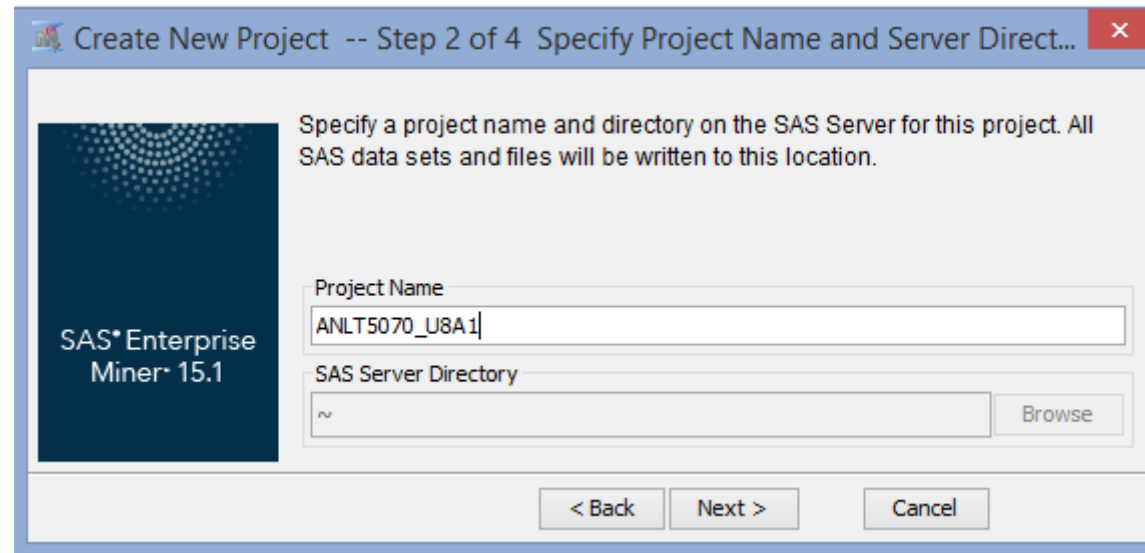
Once you download and start SAS Enterprise Miner, open a new project by clicking on “New Project.”



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

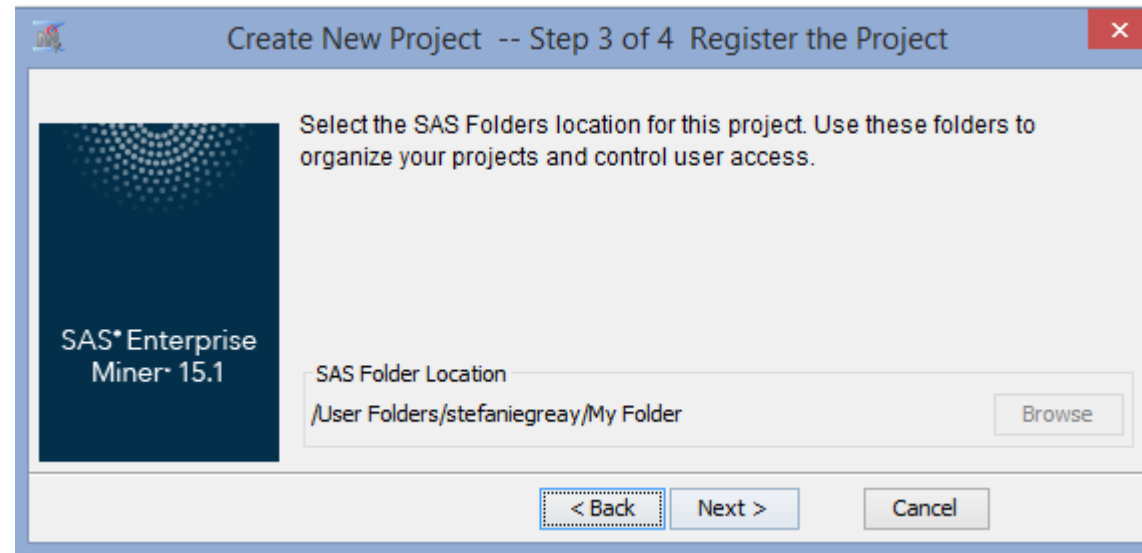
Project Name
ANLT5070_U8A1

SAS Server Directory
~ Browse

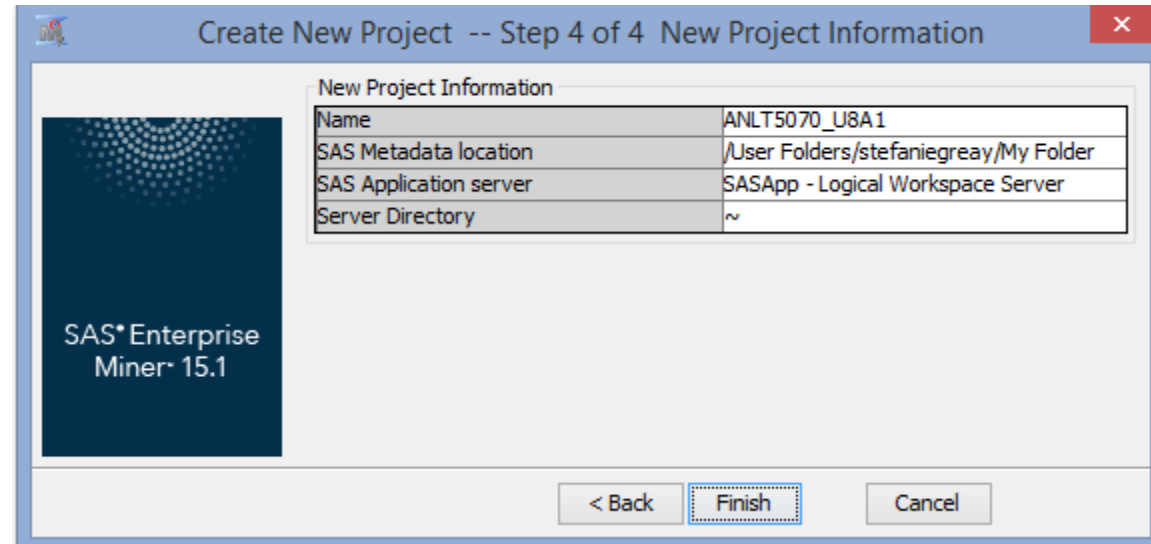
< Back Next > Cancel



Click “Next>”



Verify your entries and click “Finish”



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

New Project Information

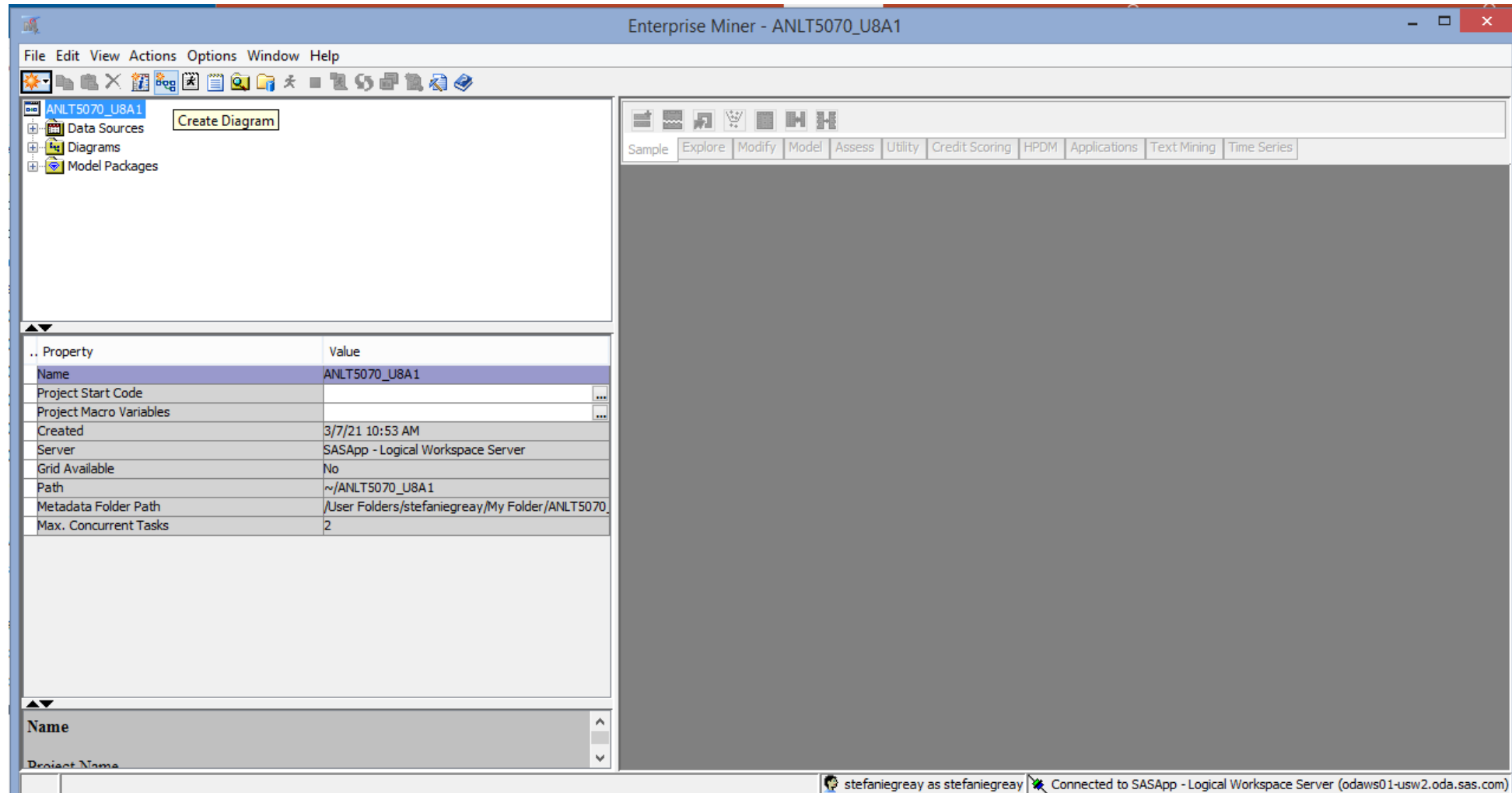
Name	ANLT5070_U8A1
SAS Metadata location	/User Folders/stefaniegreay/My Folder
SAS Application server	SASApp - Logical Workspace Server
Server Directory	~

SAS® Enterprise Miner® 15.1

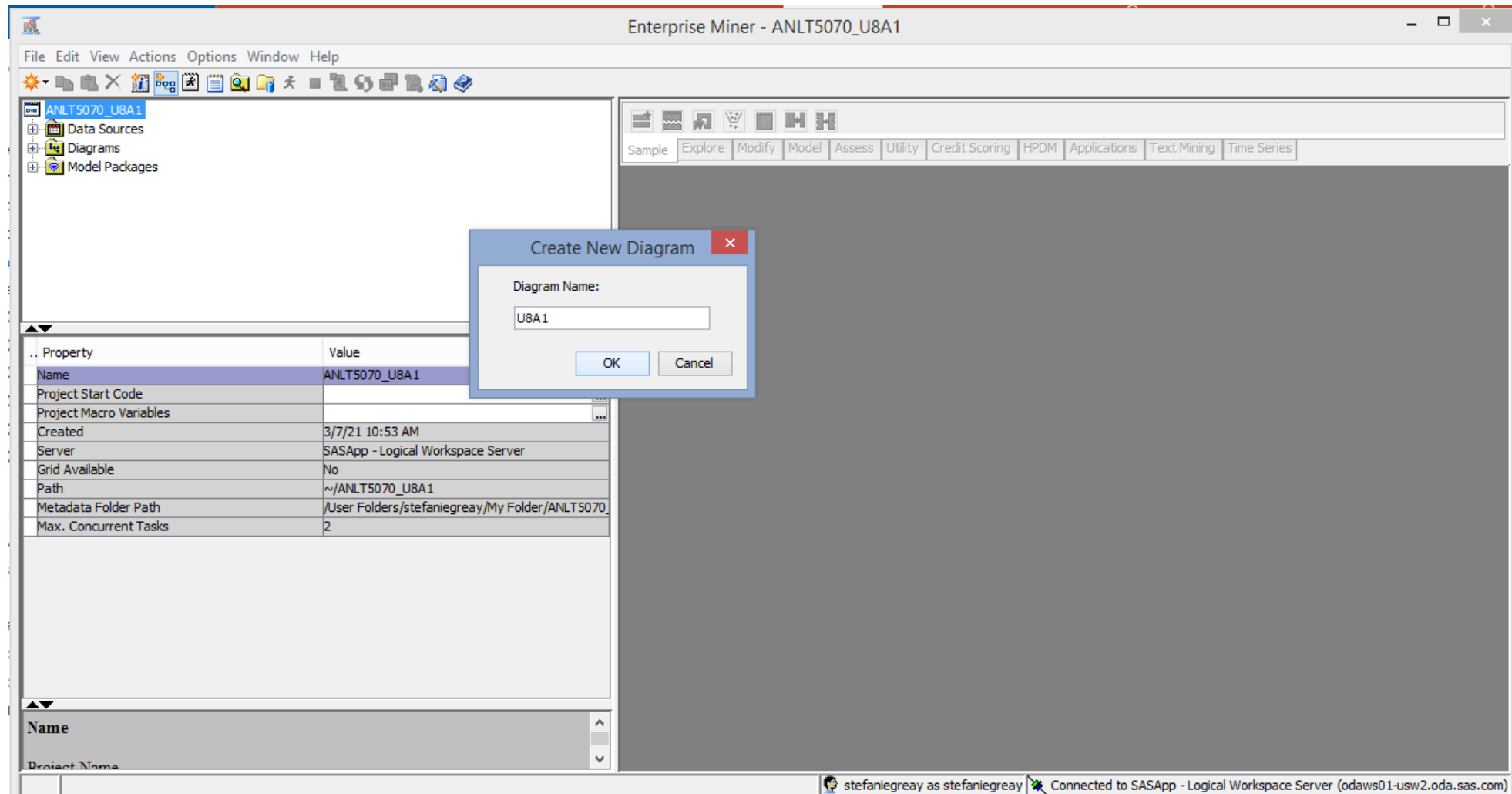
< Back Finish Cancel



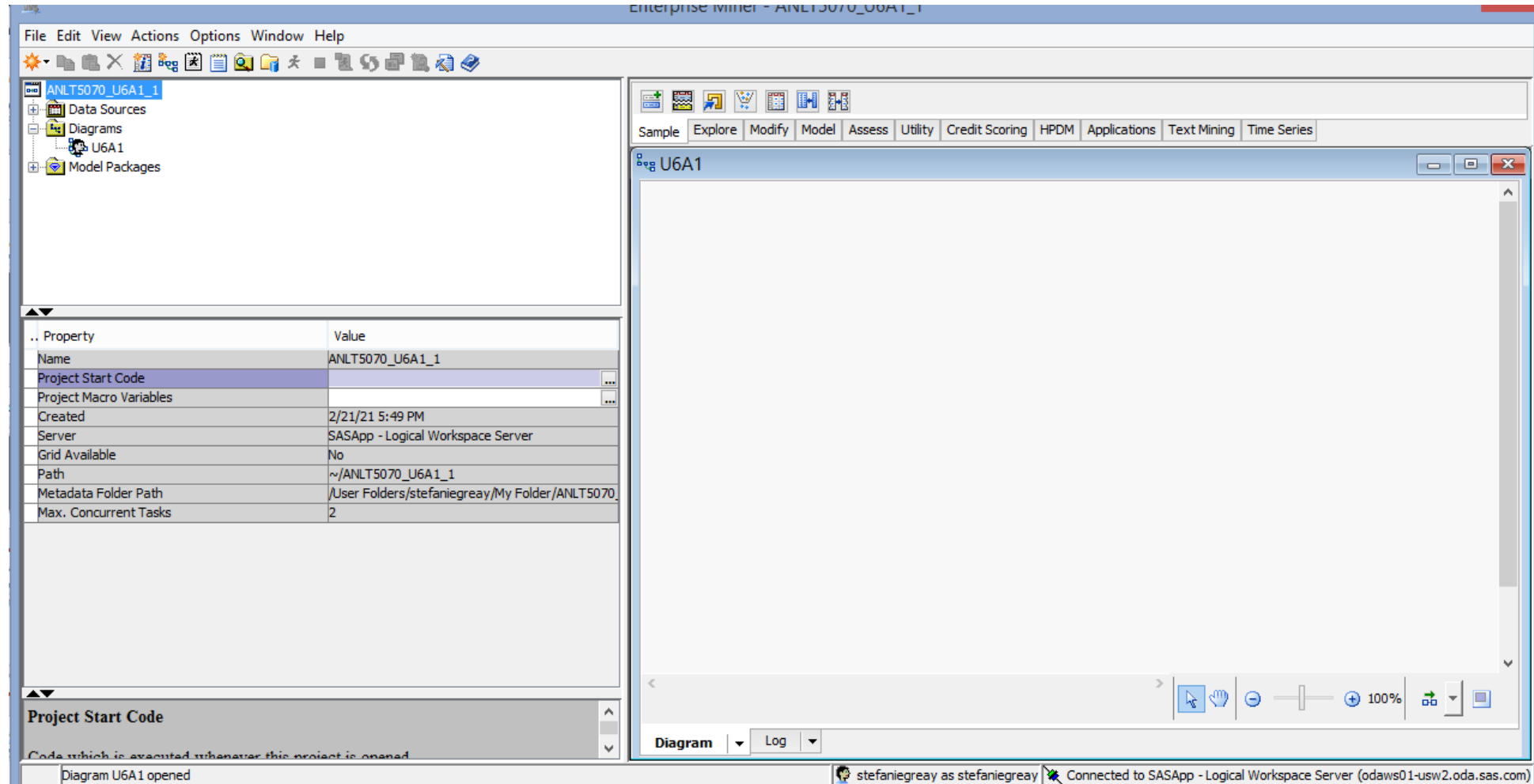
Click on the “Create Diagram” icon.



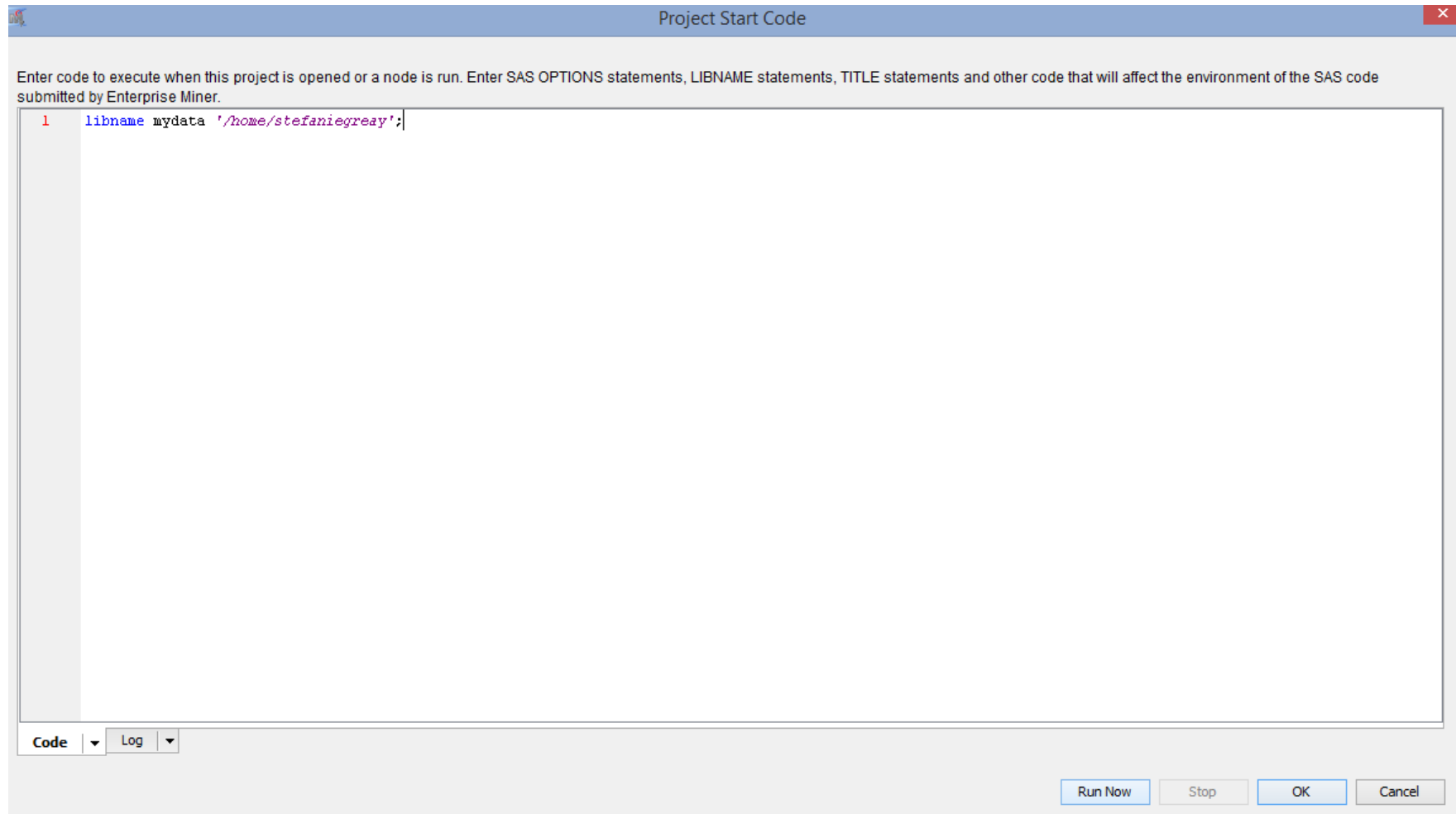
Name your diagram and click “OK.”



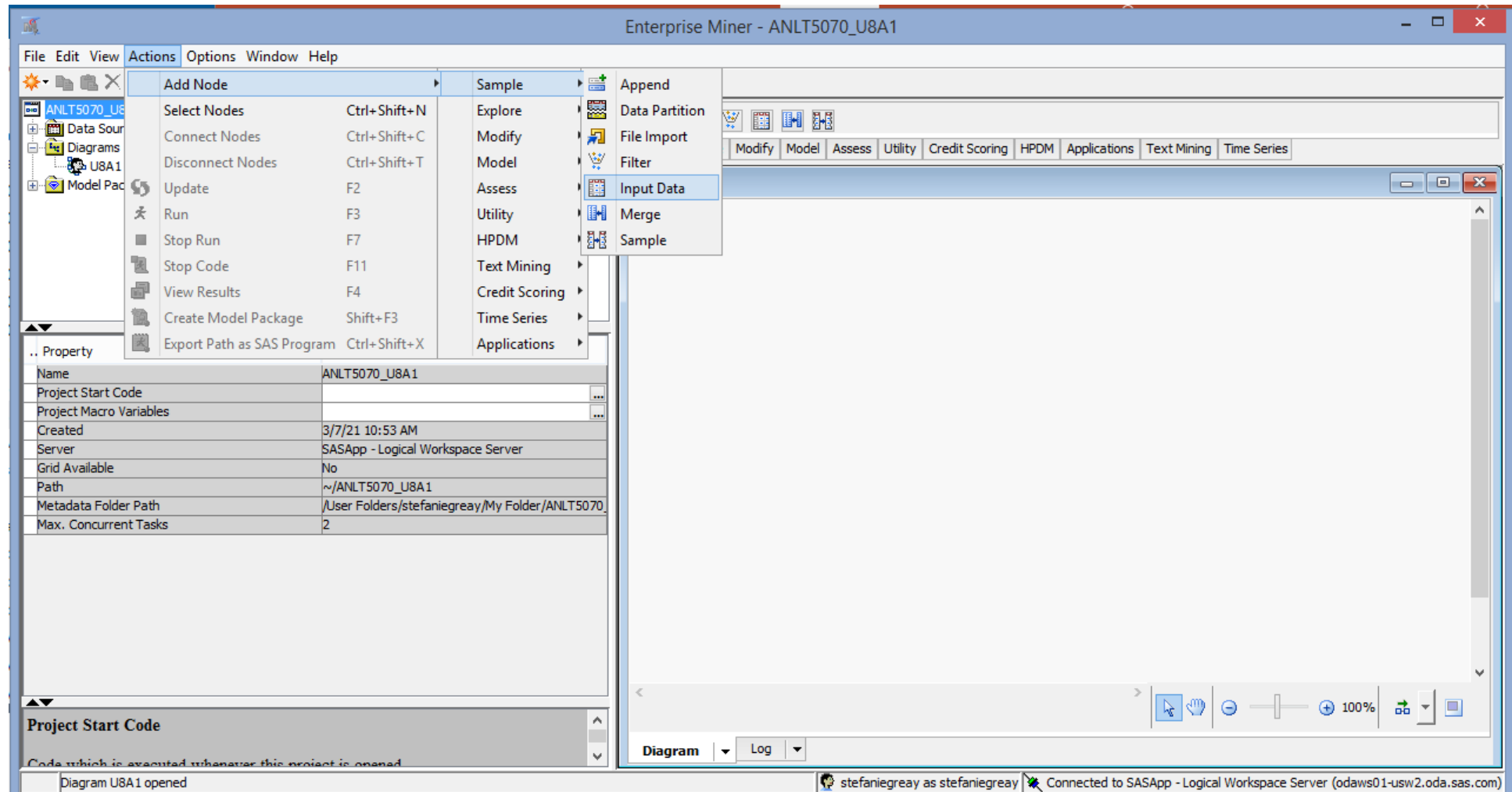
Click on the project, then click on the ellipses next to “Project Start Code.”



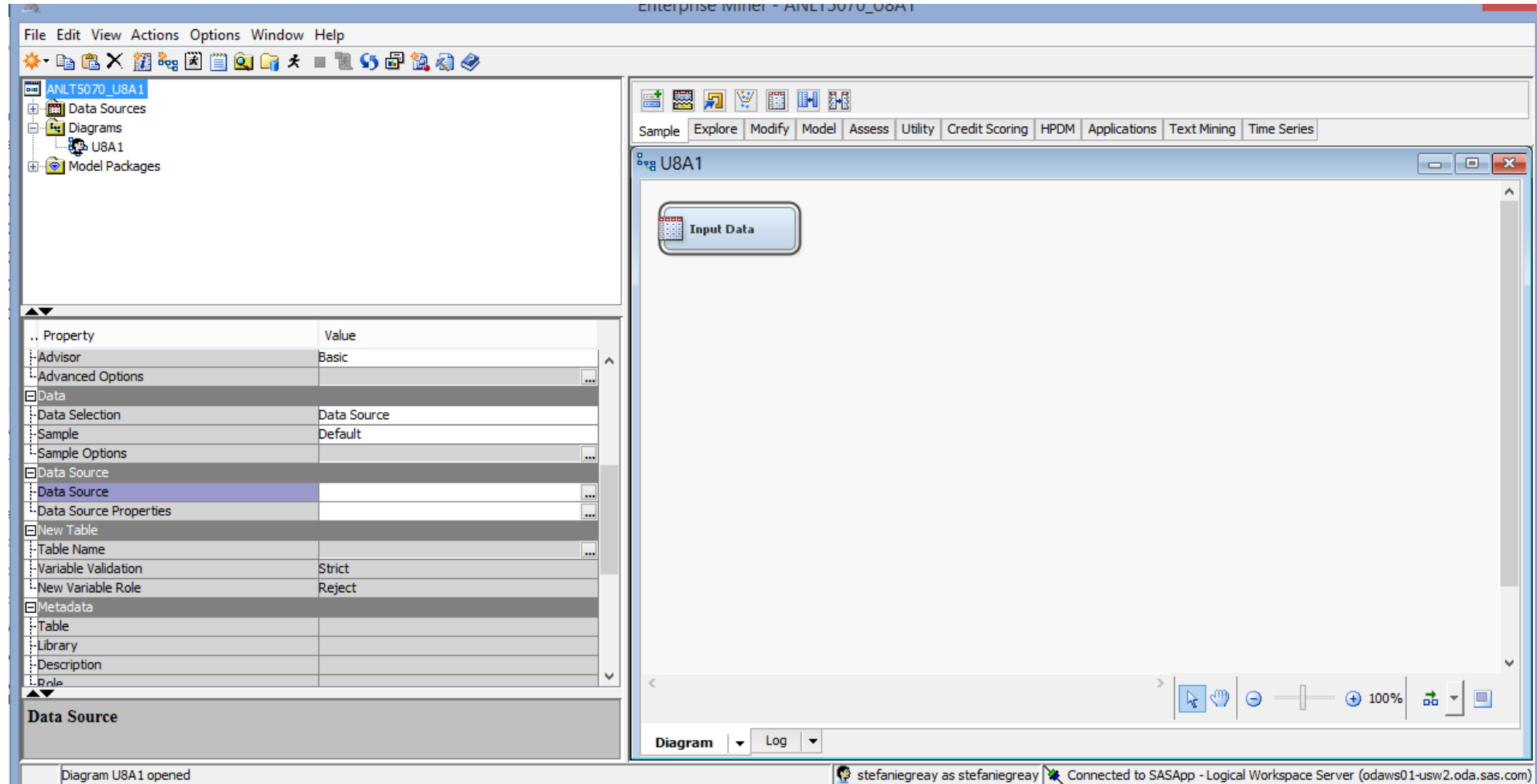
Add the library reference for where you uploaded the dataset in SAS studio, and click “Run Now.” Once it completes, click “OK.”



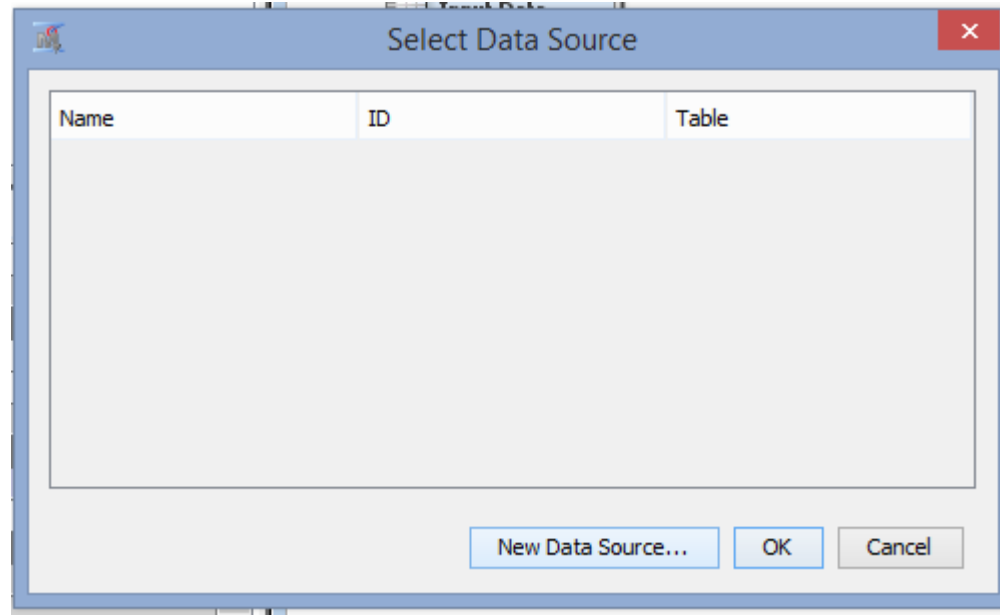
Click on Actions>Add Node>Sample>Input Data



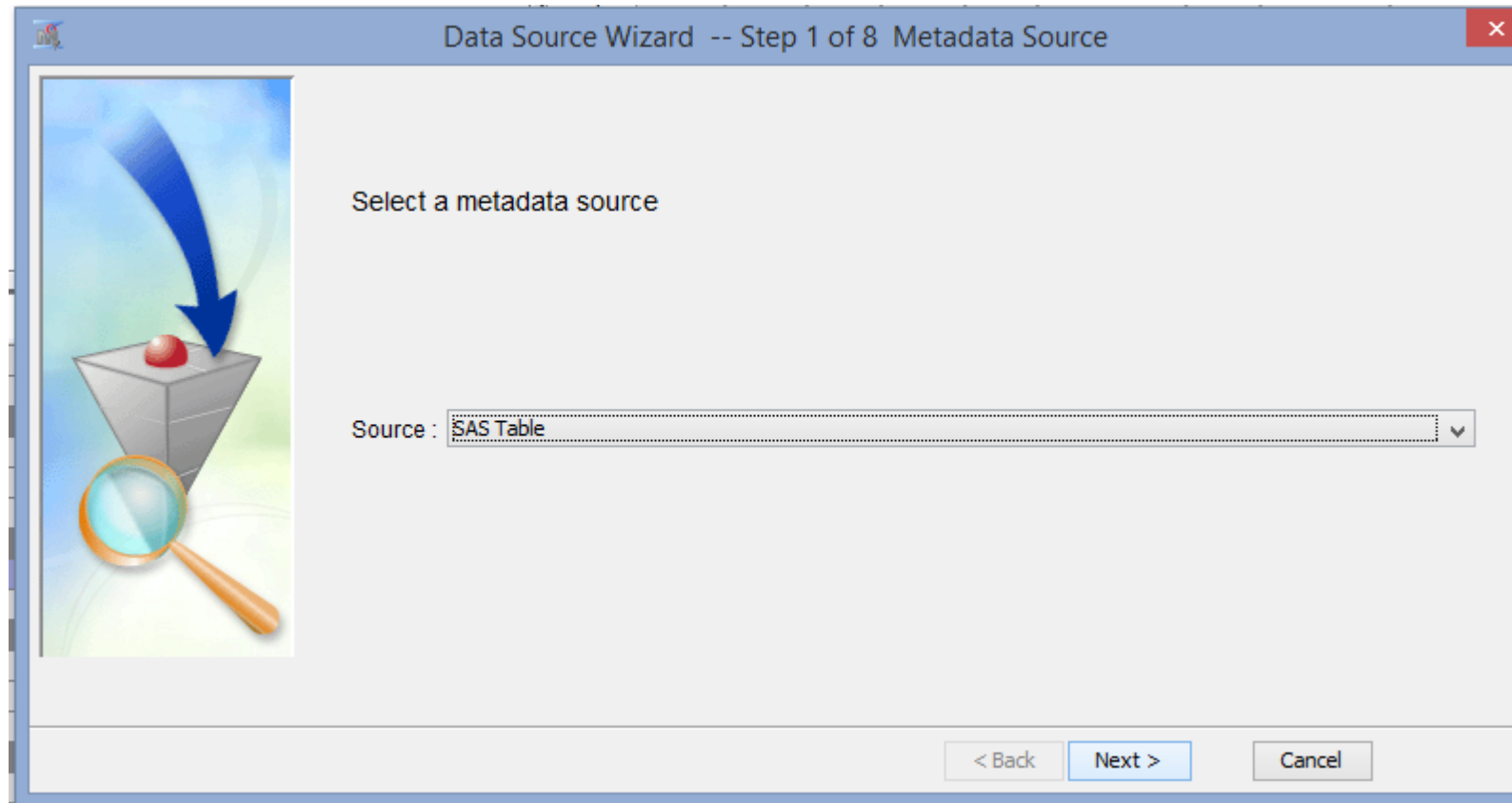
Click the ellipses (3 dots) next to “Data Source.”



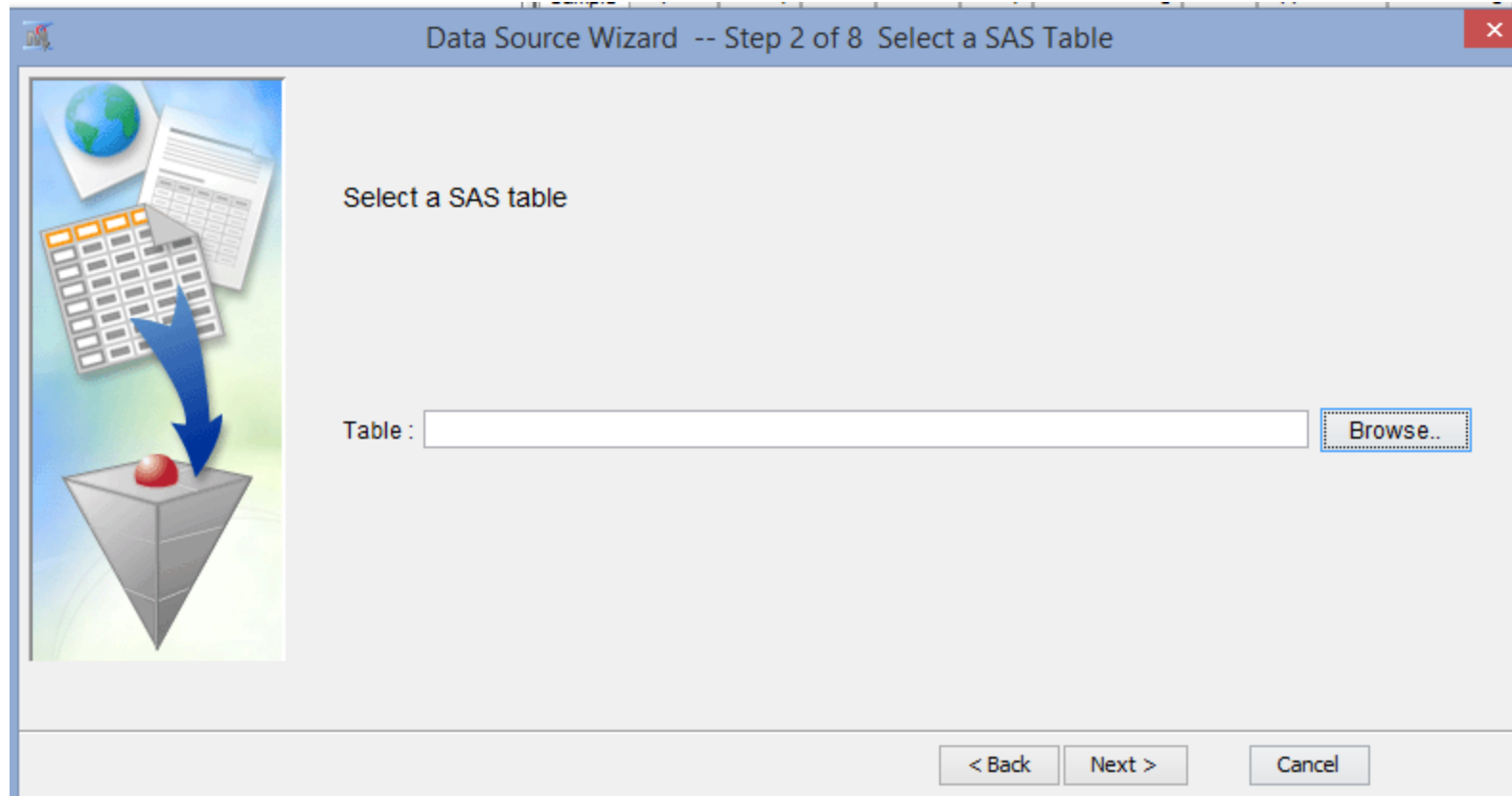
Click on “New Data Source”



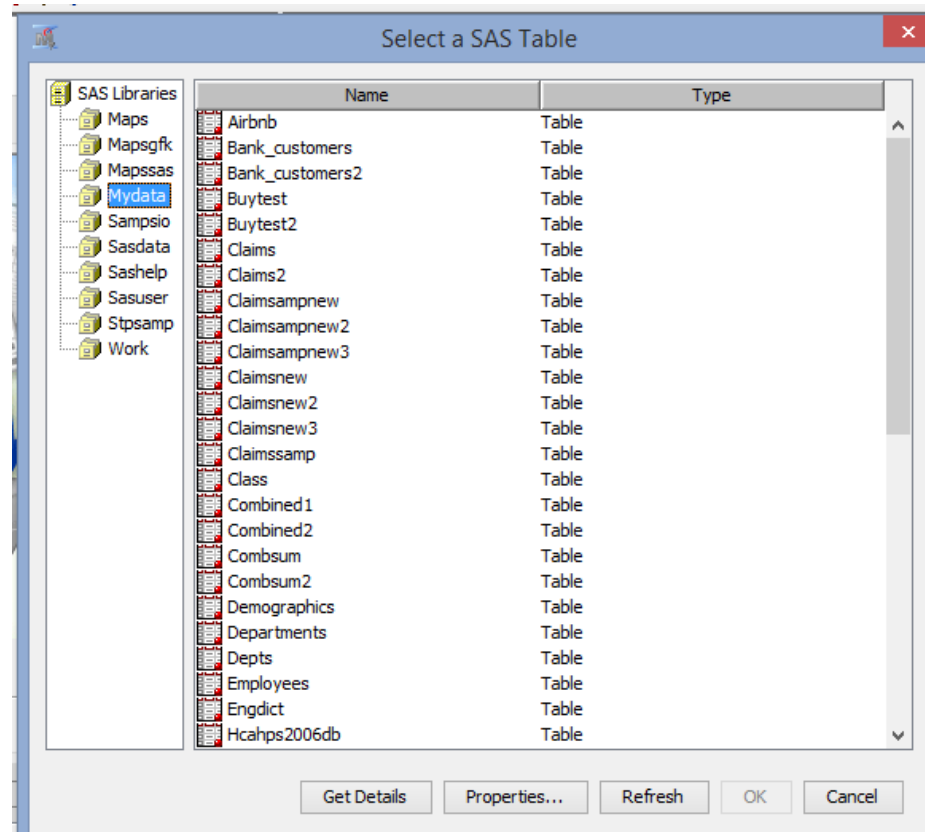
Leave it as “SAS Table” and click “Next >”



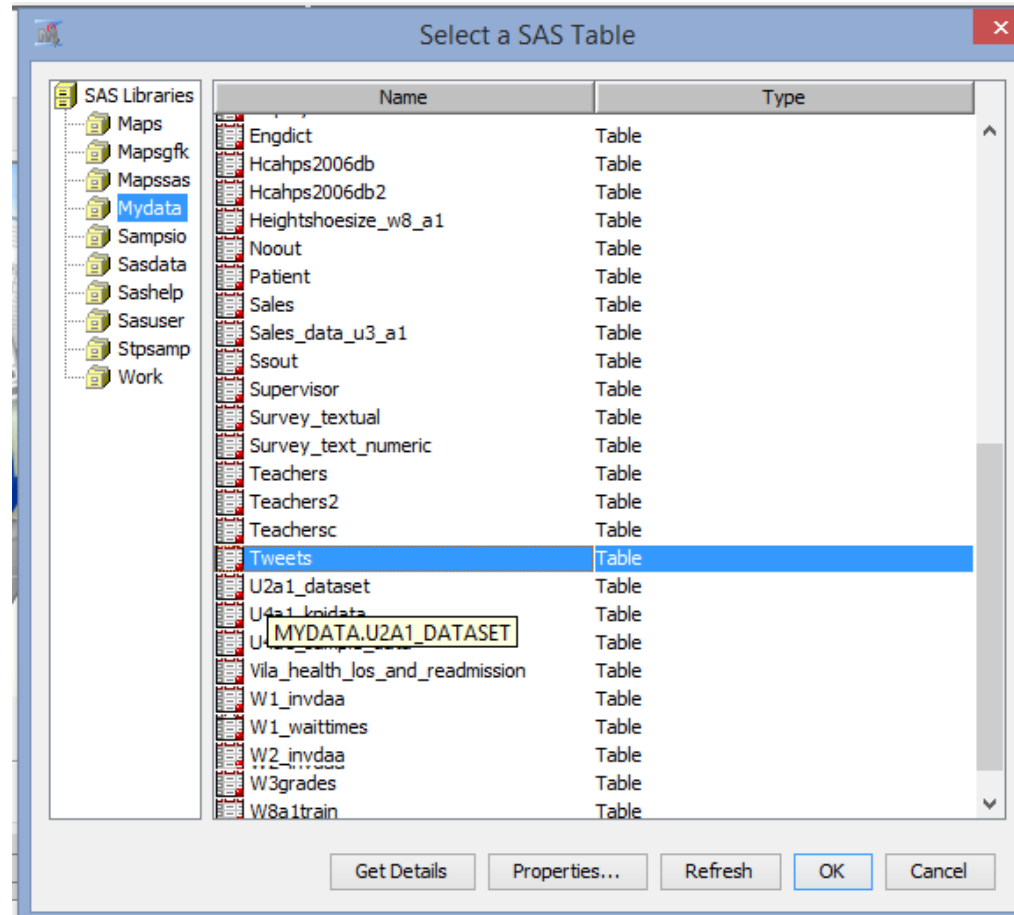
Click on “Browse”



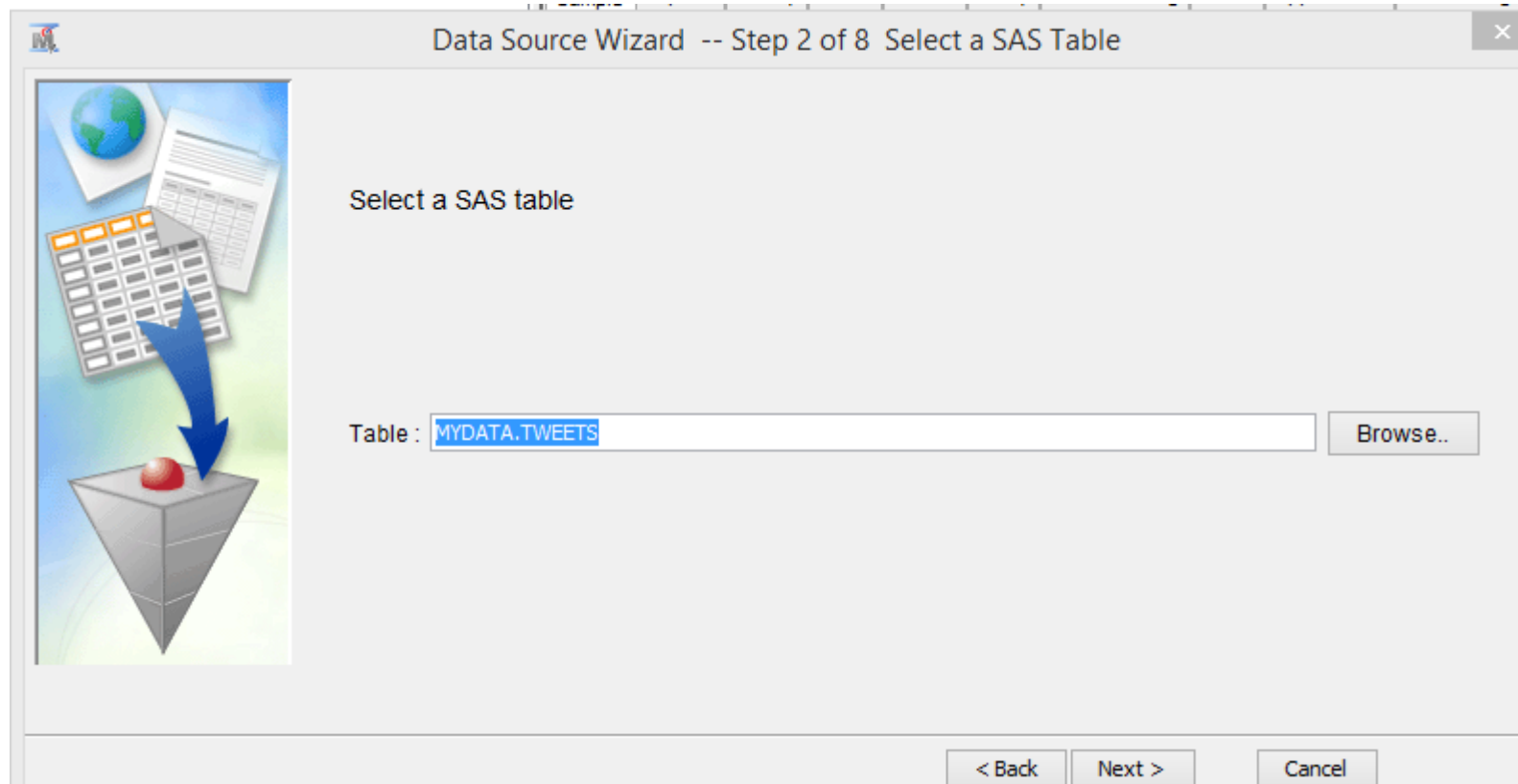
Double click on the libname you just set up in the project startup code.



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



Click “Next>”



Verify the options and click “Next>”

Data Source Wizard -- Step 3 of 8 Table Information

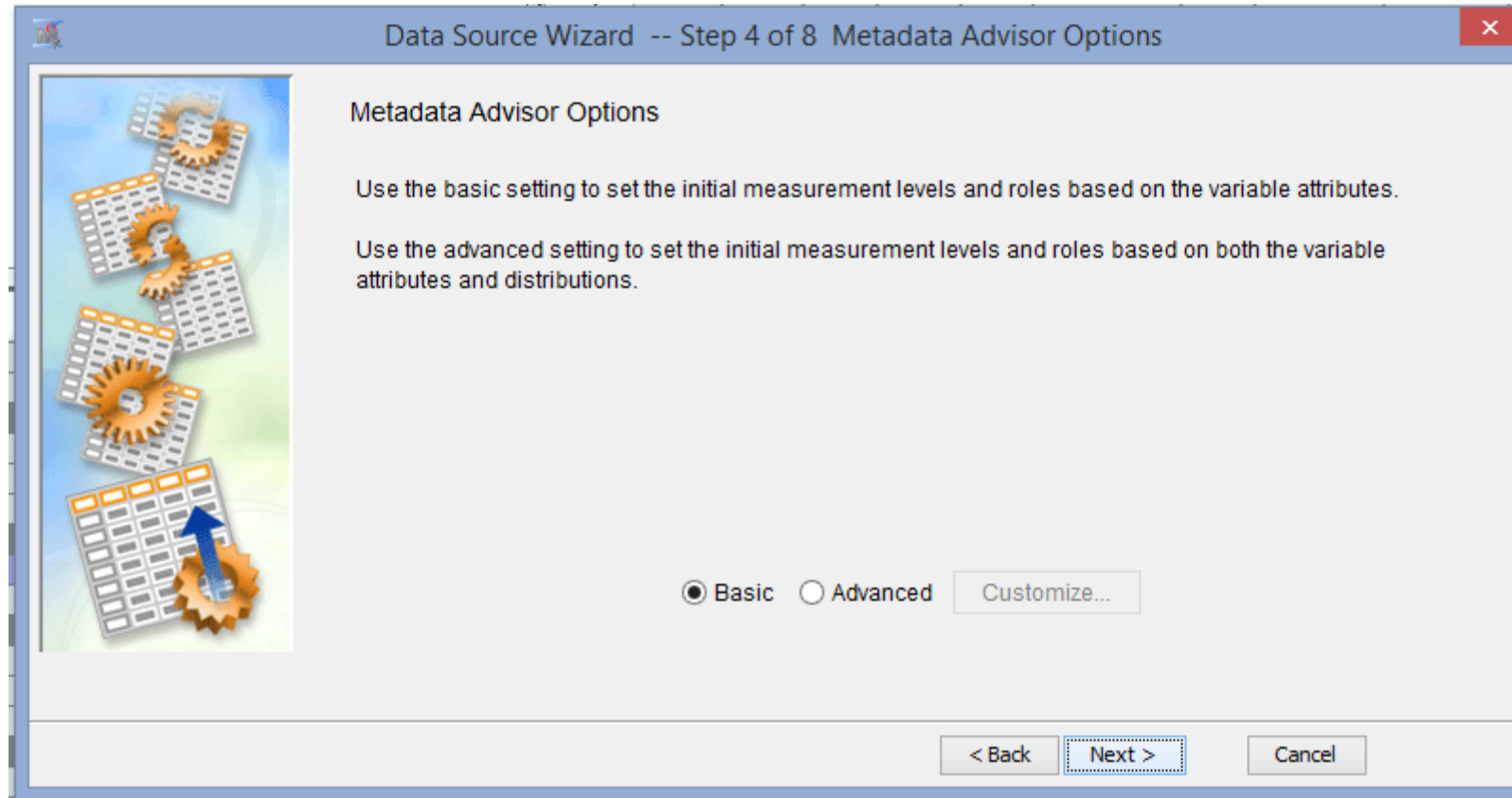
Table Properties

Property	Value
Table Name	MYDATA.TWEETS
Description	
Member Type	DATA
Data Set Type	DATA
Engine	V9
Number of Variables	8
Number of Observations	17410
Created Date	March 7, 2021 4:11:16 PM EST
Modified Date	March 7, 2021 4:11:16 PM EST

< 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	Upper L
description	Text	Nominal	No		No	.	
followers	Input	Interval	No		No	.	
location	Text	Nominal	No		No	.	
name	Input	Nominal	No		No	.	
numberstatuses	Input	Interval	No		No	.	
time	Time ID	Interval	No		No	.	
tweets	Text	Nominal	No		No	.	
username	Text	Nominal	No		No	.	

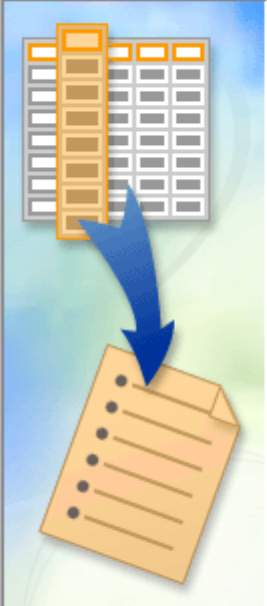
< >

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>”

Data Source Wizard -- Step 6 of 8 Create Sample



Do you wish to create a sample data set?

☒ No ☐ Yes

Table Info

Columns 8
Rows 17410

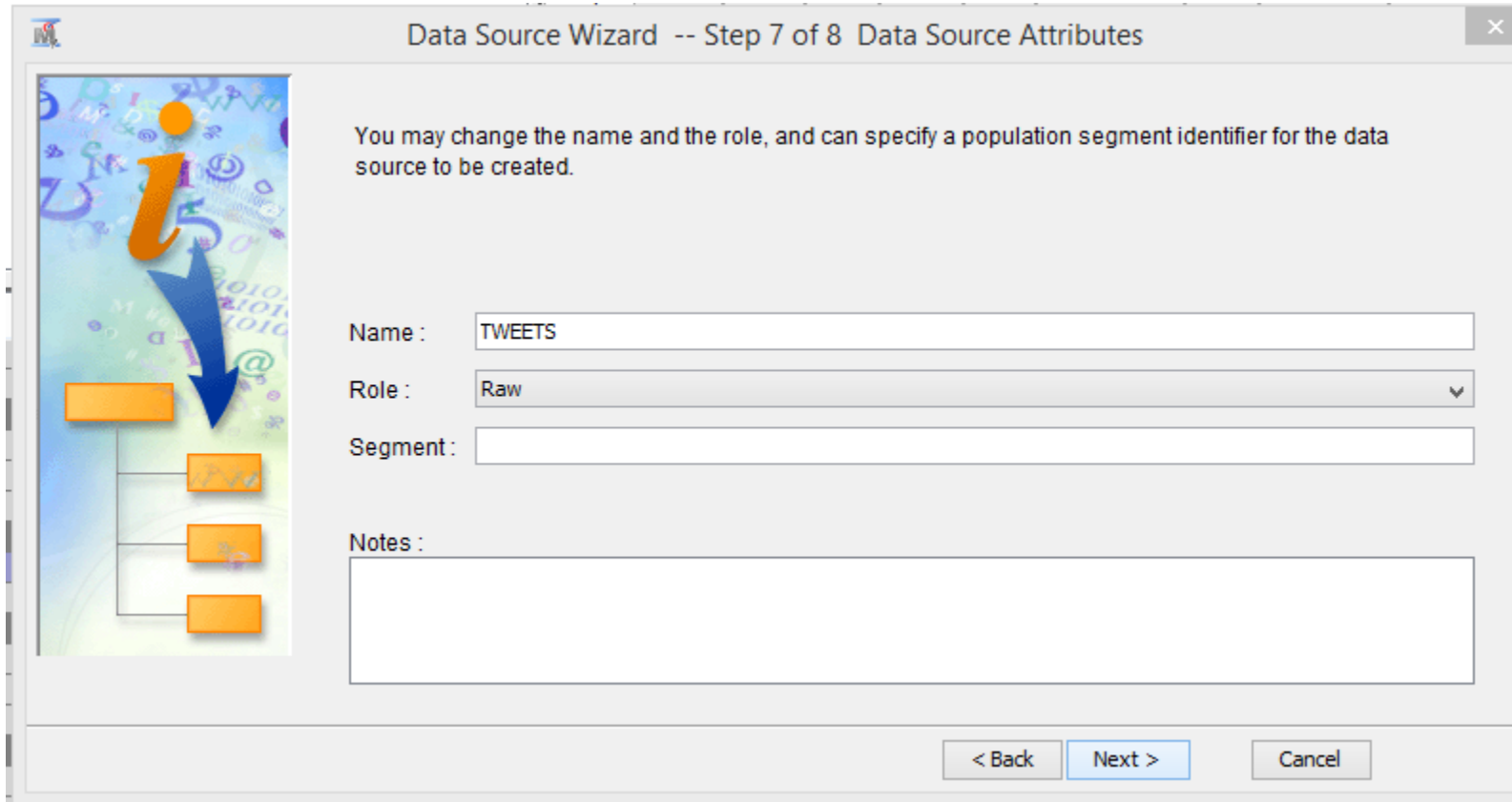
Sample Size

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 Data Source Attributes'. The window has a title bar with a close button. On the left side, there is a vertical panel with a blue background featuring a large orange 'i' icon and a blue arrow pointing down towards a series of four orange rectangular boxes connected by lines. The main area of the window 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 :

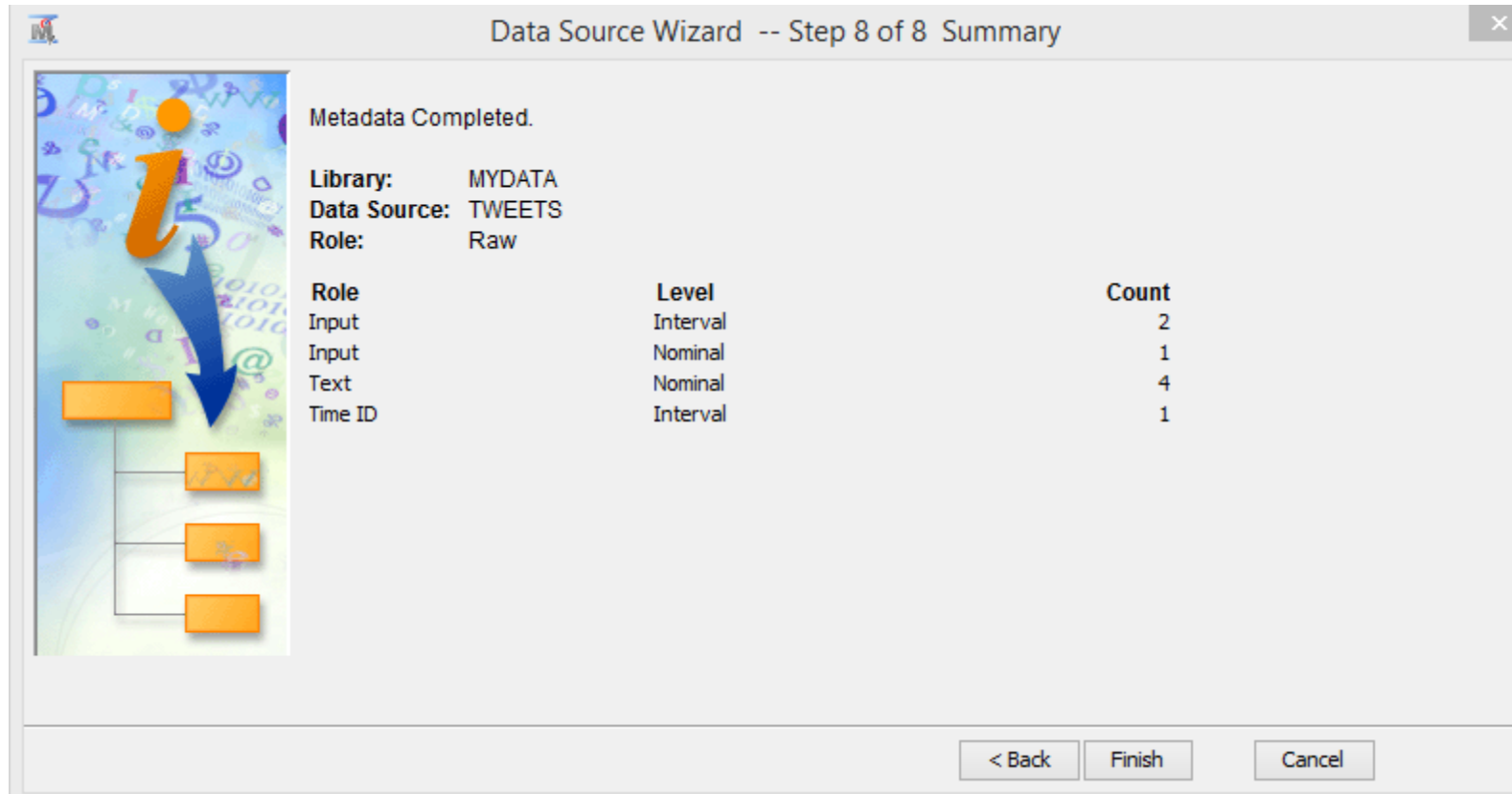
Segment :

Notes :

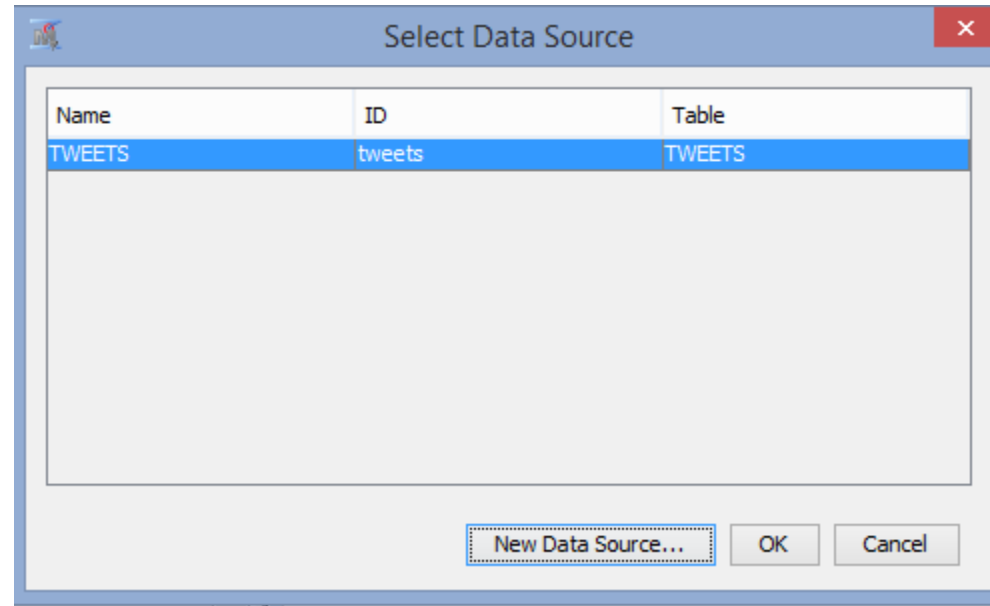
At the bottom right, there are three buttons: '< Back', 'Next >', and 'Cancel'.



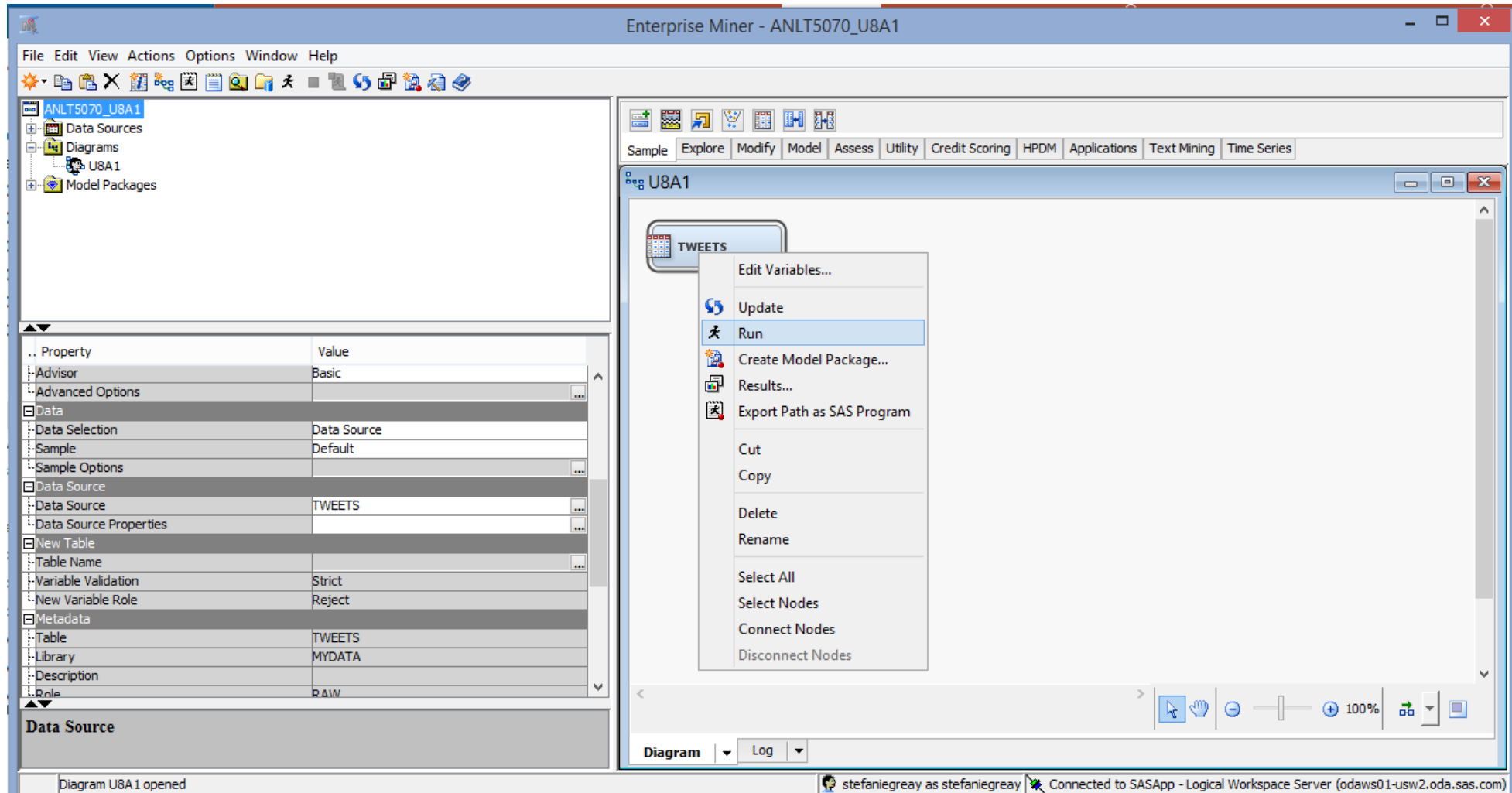
Click “Finish” to finish the data source registration within EM.



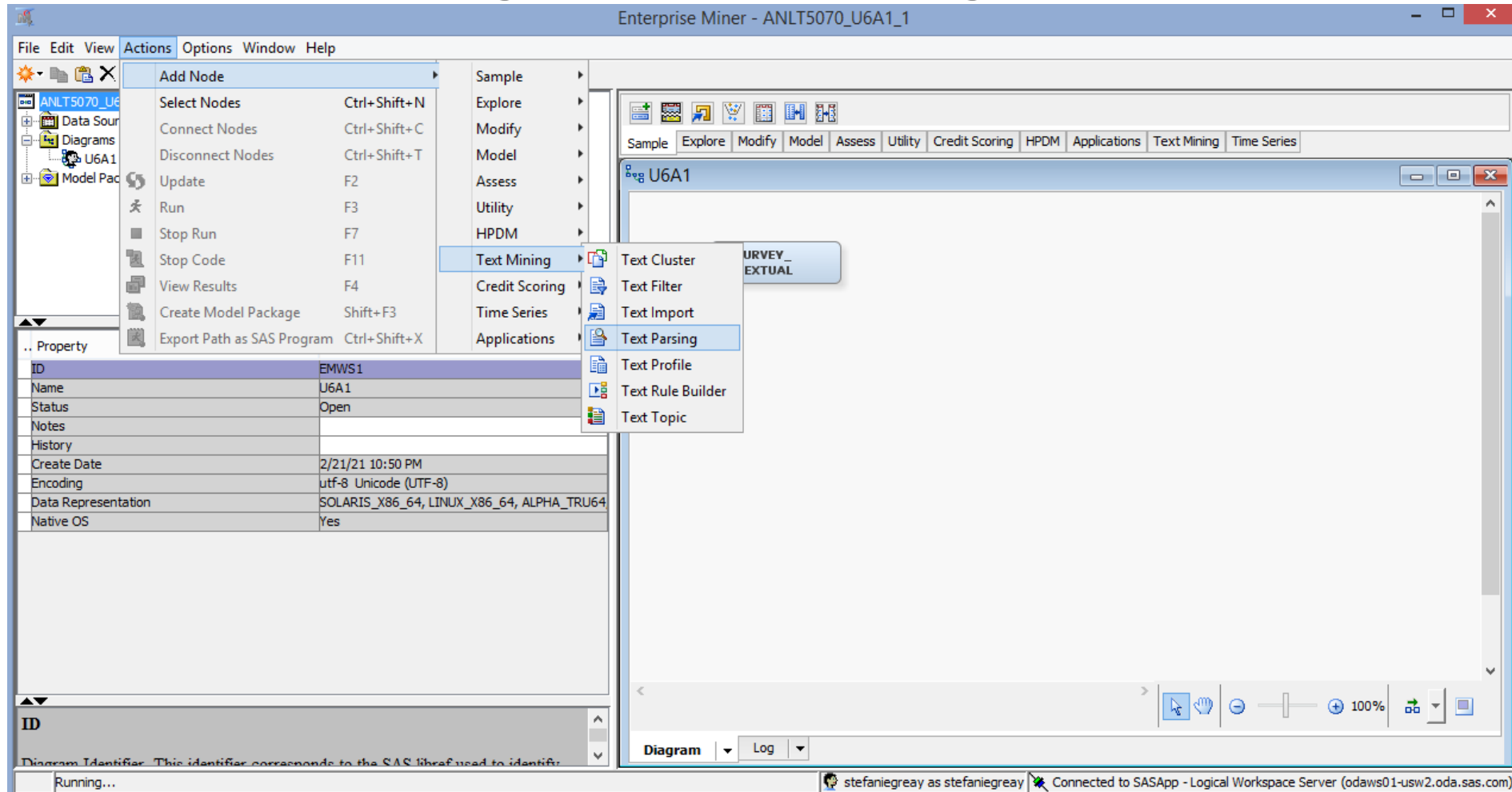
Click “OK” to complete the process. The name of the node should then change to the name of the dataset.



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



To add the Text Parsing node, click on “Actions”>”Add Node”>”Text Mining”>”Text Parsing”



Connect the nodes

The screenshot displays the SAS Enterprise Miner software interface. The main window is titled "Enterprise Miner - ANLT5070_U8A1". The left sidebar shows a project tree with "ANLT5070_U8A1" selected, containing "Data Sources", "Diagrams", "U8A1", and "Model Packages". Below the tree is a table of properties and values.

Property	Value
Custom Entities	
Ignore	
Ignore Parts of Speech	'Aux' 'Conj' 'Det' 'Interj' 'Part' 'Prep' 'Pron'
Ignore Types of Entities	
Ignore Types of Attributes	'Num' 'Punct'
Synonyms	
Stem Terms	Yes
Synonyms	SASHELP.ENGSYNMS
Filter	
Start List	
Stop List	SASHELP.ENGSTOP
Select Languages	
Report	
Number of Terms to Display	20000
Status	
Create Time	3/7/21 4:23 PM
Run ID	
Last Error	
General	
General Properties	

The main workspace shows a diagram titled "U8A1" with two nodes: "TWEETS" and "Text Parsing". An arrow connects "TWEETS" to "Text Parsing". The bottom status bar indicates the user is "stefaniegreay as stefaniegreay" and is "Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)".



Run the Text Parsing node using the default values by right clicking and selecting “Run.”

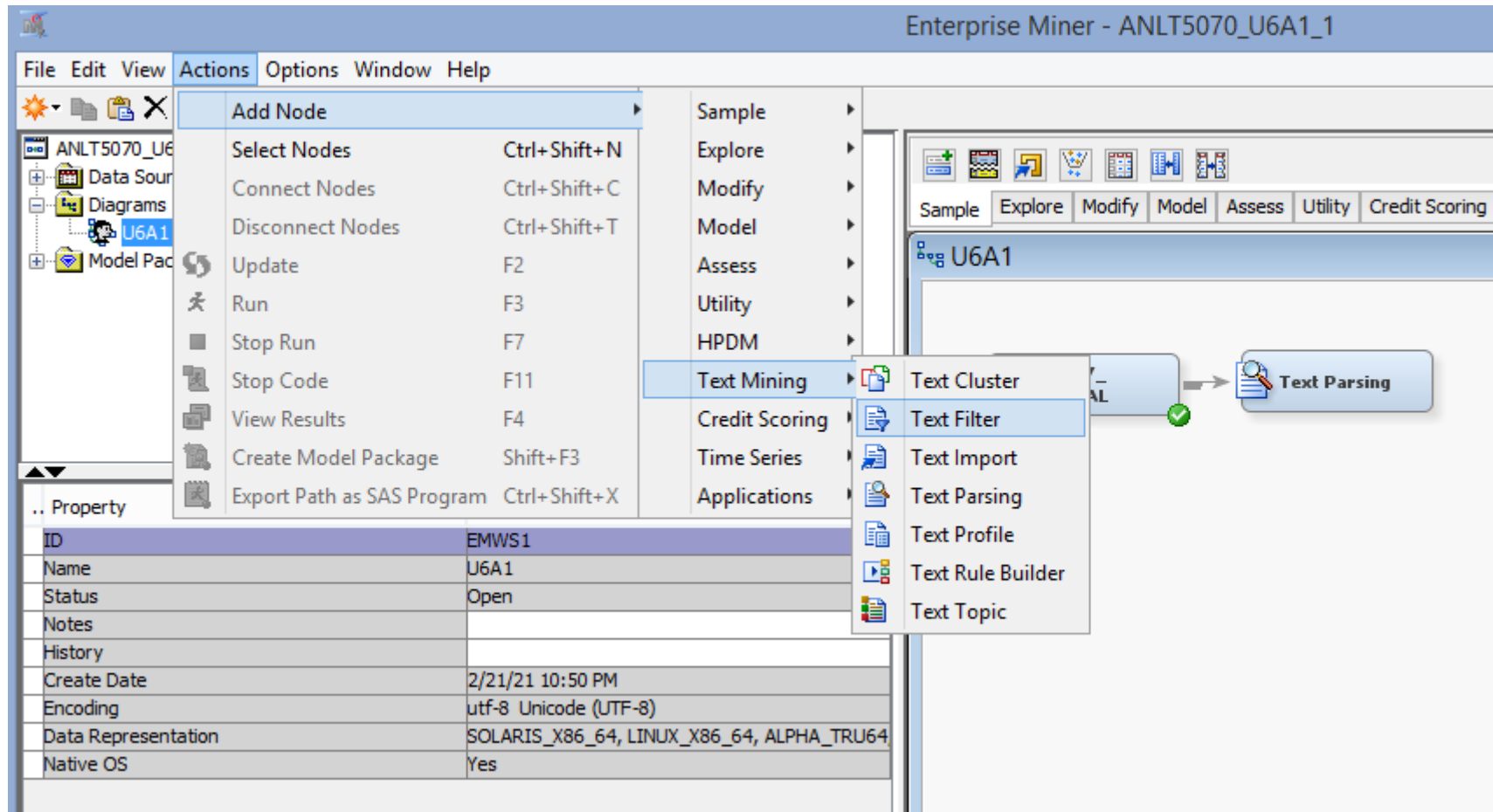
The screenshot displays the SAS Enterprise Miner interface. On the left, a tree view shows the project structure: ANLT5070_U8A1, Data Sources, Diagrams, U8A1, and Model Packages. Below this, a table lists various properties and their values:

Property	Value
Custom Entities	
Ignore	
Ignore Parts of Speech	Aux' 'Conj' 'Det' 'Interj' 'Part' 'Prep' 'Pron'
Ignore Types of Entities	
Ignore Types of Attributes	Num' 'Punct'
Synonyms	
Stem Terms	Yes
Synonyms	SASHELP.ENGSYNMS
Filter	
Start List	
Stop List	SASHELP.ENGSTOP
Select Languages	
Report	
Number of Terms to Display	20000
Status	
Create Time	3/7/21 4:23 PM
Run ID	
Last Error	
General	
General Properties	
Running...	

The main workspace shows a diagram with a 'TWEETS' data source node connected to a 'Text Parsing' node. A right-click context menu is open over the 'Text Parsing' node, with the 'Run' option highlighted. The menu includes options like 'Edit Variables...', 'Update', 'Run', 'Create Model Package...', 'Results...', 'Export Path as SAS Program', 'Cut', 'Copy', 'Delete', 'Rename', 'Select All', 'Select Nodes', 'Connect Nodes', and 'Disconnect Nodes'. The bottom status bar indicates the user is 'stefaniegreay as stefaniegreay' and is 'Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)'.



To add the Text Filter node, click on
“Actions”>“Add Node”>“Text Mining”>“Text Filter”



Connect the nodes

The screenshot displays the SAS Enterprise Miner interface. The main workspace shows a workflow diagram with three nodes: 'TWEETS' (data source), 'Text Parsing' (text mining process), and 'Text Filter' (text mining process), connected by arrows indicating a sequential flow. The left pane shows the project structure with 'ANLT5070_U8A1' selected. The bottom pane shows the 'General' properties for the selected node, including a table of properties and values.

Property	Value
Maximum Number of Terms	.
Import Synonyms	
Document Filters	
Search Expression	
Subset Documents	
Results	
Filter Viewer	
Spell-Checking Results	
Exported Synonyms	
Report	
Terms to View	All
Number of Terms to Display	20000
Status	
Create Time	3/7/21 4:25 PM
Run ID	
Last Error	
Last Status	
Last Run Time	

General

Running...



Run the Text Filter node using the default values by right clicking and selecting “Run.”

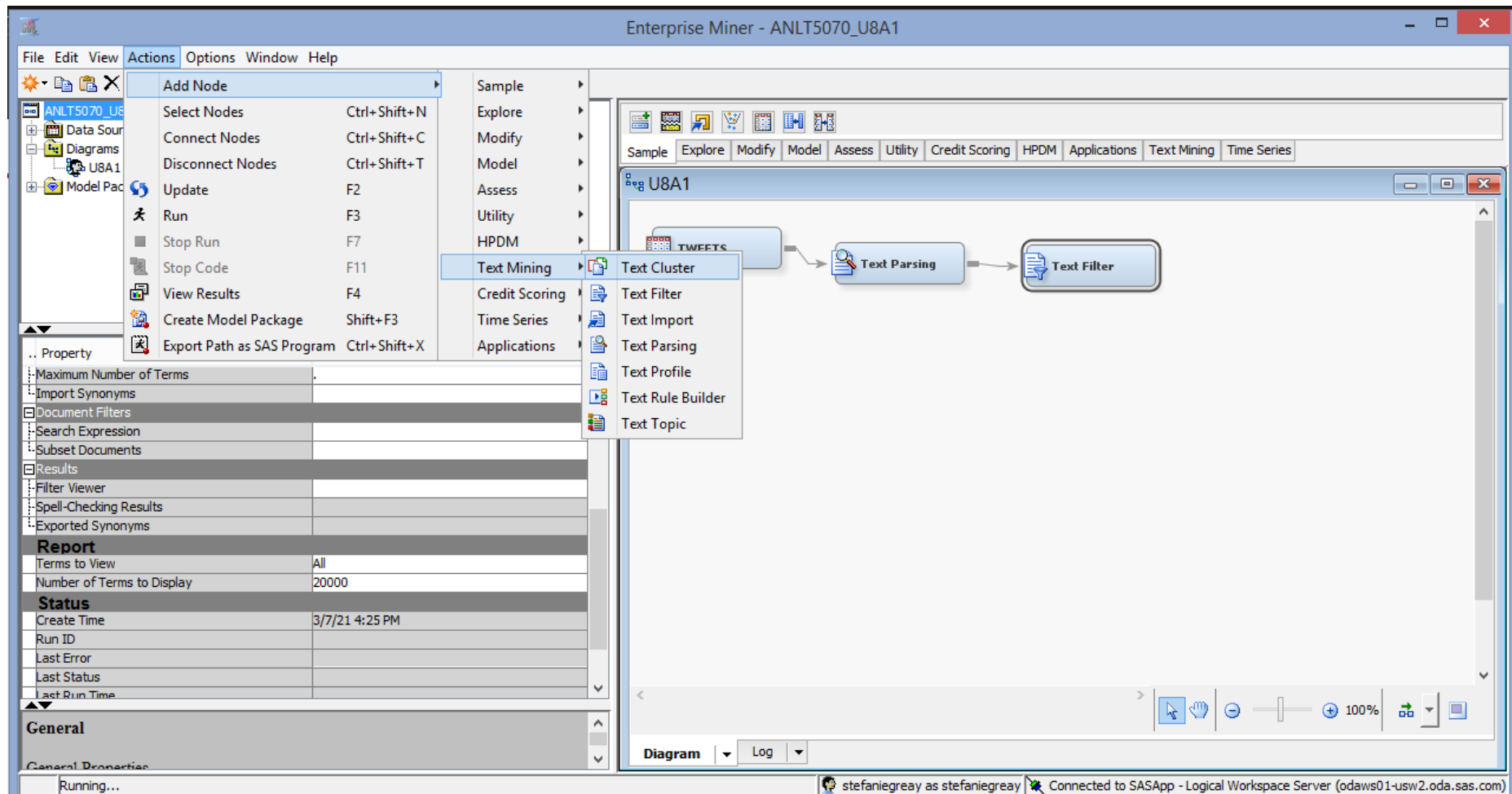
The screenshot displays the SAS Enterprise Miner interface. On the left, a project tree shows 'ANLT5070_U8A1' containing 'Data Sources', 'Diagrams', 'U8A1', and 'Model Packages'. Below this, a table lists properties and their values:

Property	Value
Maximum Number of Terms	
Import Synonyms	
Document Filters	
Search Expression	
Subset Documents	
Results	
Filter Viewer	
Spell-Checking Results	
Exported Synonyms	
Report	
Terms to View	All
Number of Terms to Display	20000
Status	
Create Time	3/7/21 4:25 PM
Run ID	
Last Error	
Last Status	
Last Run Time	

The main workspace shows a workflow diagram with nodes 'TWEETS', 'Text Parsing', and 'Text'. A right-click context menu is open over the 'Text' node, with the 'Run' option selected. The menu includes options like '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 bottom status bar indicates the user is 'stefaniegreay' and is connected to 'SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)'.



Add the Text Cluster node by clicking on “Actions”>”Add Node”>”Text Mining”>”Text Cluster”



Connect the nodes, then run the cluster node.

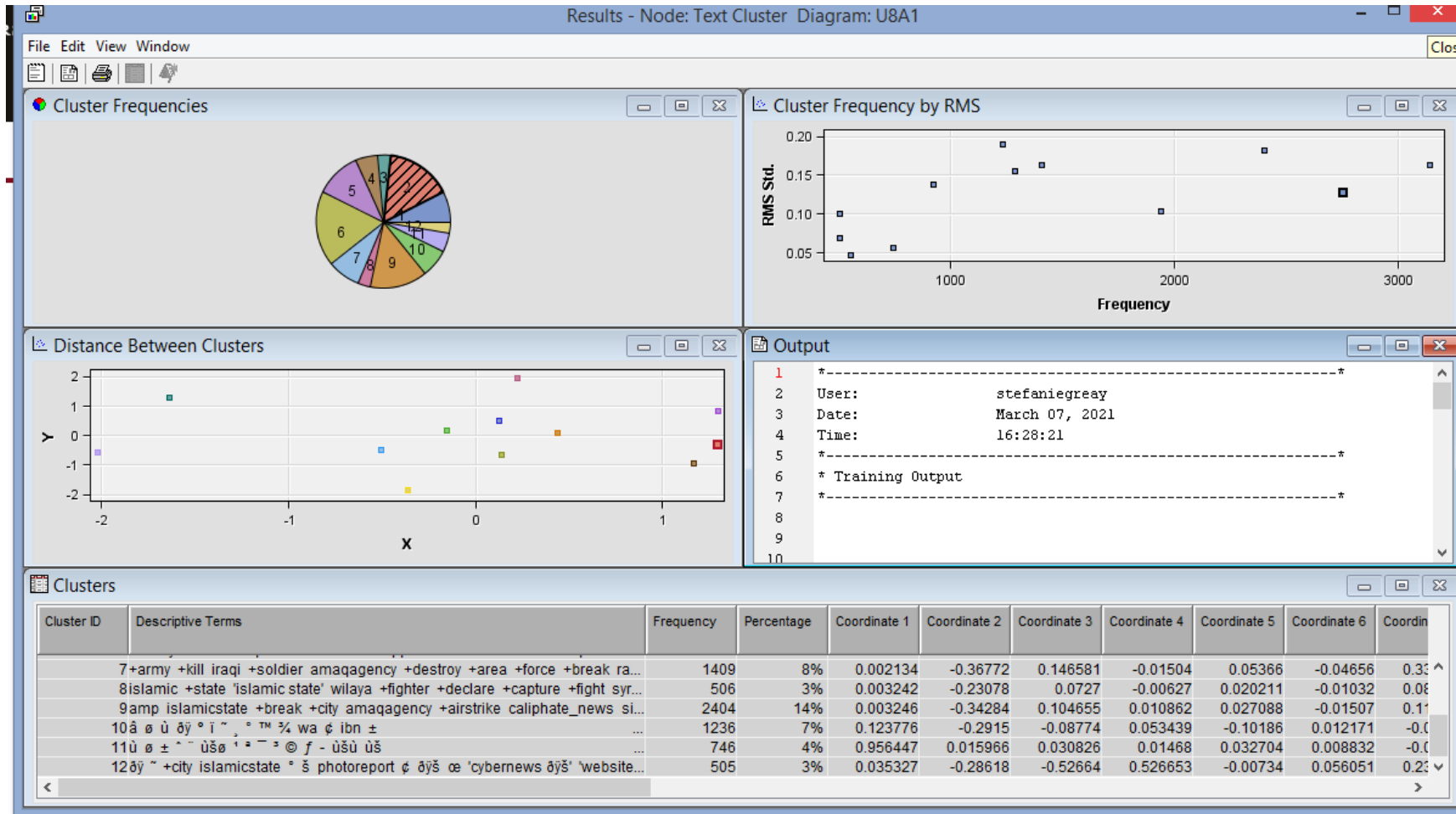
The screenshot displays the SAS Enterprise Miner software interface. The main window, titled "Enterprise Miner - ANLT5070_U8A1", shows a workflow diagram with four nodes: "TWEETS", "Text Parsing", "Text Filter", and "Text Cluster". The nodes are connected sequentially, and each has a green checkmark indicating it is ready to run. A context menu is open over the "Text Cluster" node, listing various actions such as "Edit Variables...", "Update", "Run", "Create Model Package...", "Results...", "Export Path as SAS Program", "Cut", "Copy", "Delete", "Rename", "Select All", "Select Nodes", "Connect Nodes", and "Disconnect Nodes". The "Run" option is highlighted. On the left side, a tree view shows the project structure, including "Data Sources", "Diagrams", "U8A1", and "Model Packages". Below the tree view, a table displays the properties of the selected node, "TextCluster".

Property	Value
General	
Node ID	TextCluster
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Transform	
SVD Resolution	Low
Max SVD Dimensions	100
Cluster	
Exact or Maximum Number	Maximum
Number of Clusters	40
Cluster Algorithm	Expectation-Maximization
Descriptive Terms	15
Status	
Create Time	3/7/21 4:27 PM
Run ID	
General	

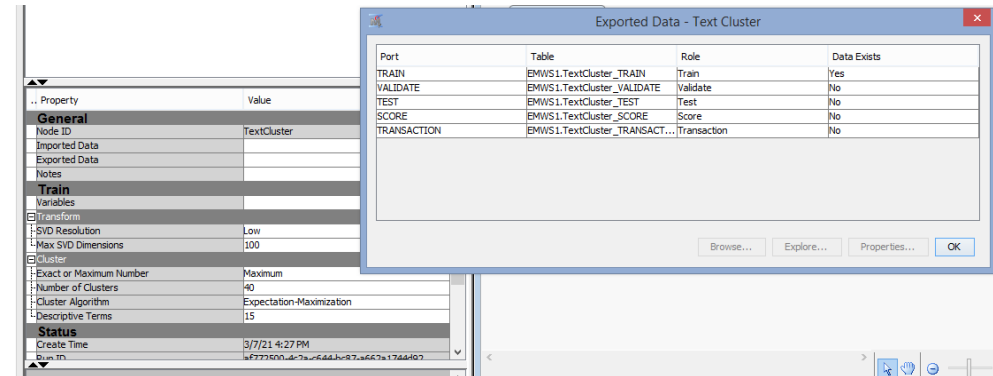
The status bar at the bottom indicates the user is "stefaniegreay as stefaniegreay" and is "Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)".



Right click on the “Text Cluster” node and select “Results” to view the default cluster results.



The results can be reviewed by right clicking on each of the text cluster nodes and exploring the resulting summaries or exploring the output dataset (by clicking on the ellipses next to “export datasets” from this node that contains the original dataset as well as the cluster assignment for each observation.

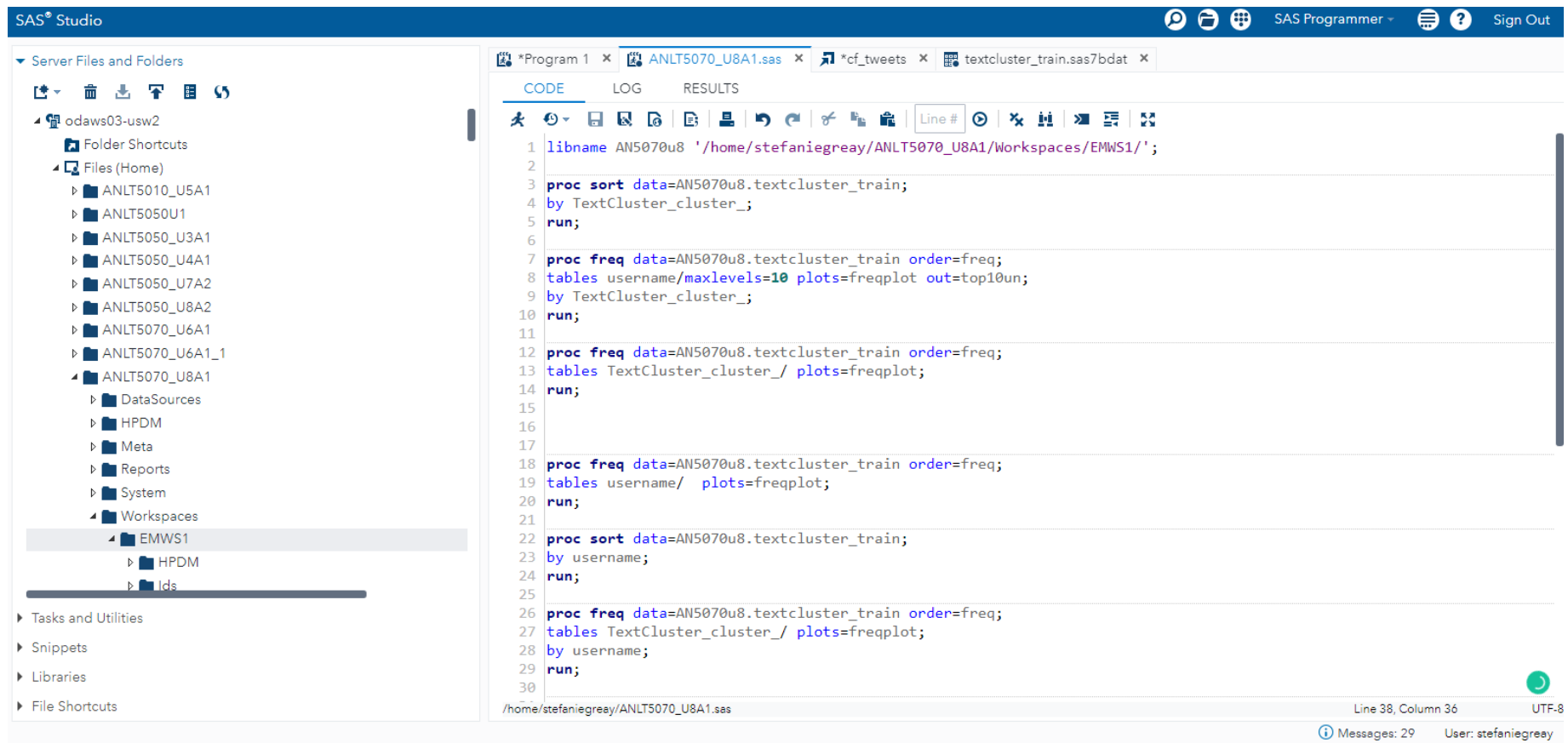


To review the output/export dataset from the text cluster node, you can navigate to that project in SAS Studio, and explore it using the functions or code there.

The screenshot displays the SAS Studio interface. On the left, the 'Server Files and Folders' pane shows a project structure under 'ANLT5070_U8A1'. The 'Workspaces' section is expanded, showing a folder named 'TextCluster'. Inside this folder, several files are listed, including 'EMWS1', 'HPDM', 'Ids', 'System', 'ODS', 'BATCH.sas', 'CDELTA_TRAIN.sas', 'EMCUSTOMGRAPH.xml', 'EMLOG.log', 'EMNOTES.txt', 'EMOUTPUT.out', 'EMPUBLISHSCORE.sas', and 'EMREPORT.log'. The main workspace area shows a table with columns 'tweets', 'TextCluster_cluster_', and 'name'. The table contains 17,410 rows and 44 columns. The 'tweets' column contains text data, and the 'TextCluster_cluster_' column contains cluster identifiers. The 'name' column contains user names. The bottom status bar shows 'Messages: 6' and 'User: stefaniegreay'.



Sample code is provided to explore the export dataset from the Text Cluster node.



The screenshot displays the SAS Studio web interface. On the left, the 'Server Files and Folders' pane shows a project structure with a workspace named 'EMWS1'. The main editor area is titled 'ANLT5070_U8A1.sas' and contains the following SAS code:

```
1 libname AN5070u8 '/home/stefaniegreay/ANLT5070_U8A1/Workspaces/EMWS1/';
2
3 proc sort data=AN5070u8.textcluster_train;
4 by TextCluster_cluster_;
5 run;
6
7 proc freq data=AN5070u8.textcluster_train order=freq;
8 tables username/maxlevels=10 plots=freqplot out=top10un;
9 by TextCluster_cluster_;
10 run;
11
12 proc freq data=AN5070u8.textcluster_train order=freq;
13 tables TextCluster_cluster_/ plots=freqplot;
14 run;
15
16
17
18 proc freq data=AN5070u8.textcluster_train order=freq;
19 tables username/ plots=freqplot;
20 run;
21
22 proc sort data=AN5070u8.textcluster_train;
23 by username;
24 run;
25
26 proc freq data=AN5070u8.textcluster_train order=freq;
27 tables TextCluster_cluster_/ plots=freqplot;
28 by username;
29 run;
30
```

The status bar at the bottom indicates the current position is Line 38, Column 36, and the user is stefaniegreay.

