

# ANLT5070

## Unit 2 Assignment 1 Tutorial



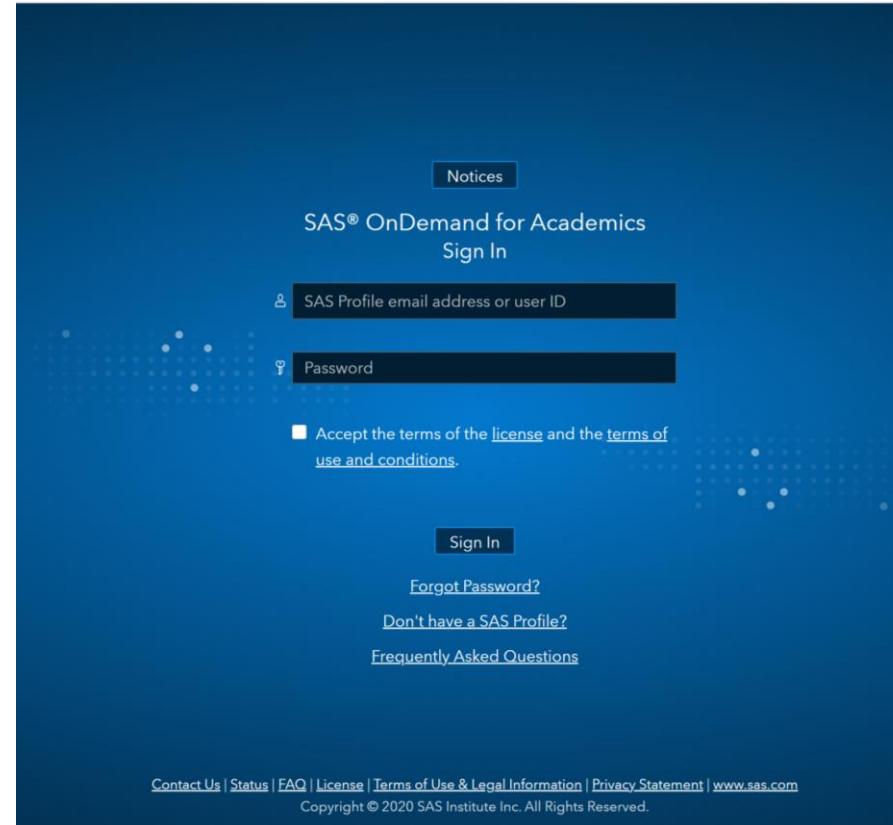
# Repeated Steps from U1A1

- The following steps are repeats from U1A1, in case those were not completed accurately or completely. The upload of the pdf files technically does not need to be completed again if it was successfully completed for last unit's assignment.
- If you successfully completed U1A1 and imported the 31 pdf files into SAS datasets using the Text Import node, and saved your project, you may skip to slide 32 to open the project and continue working on the next steps for this assignment.



# Access the SAS OnDemand for Academics Control Center

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



# SAS OnDemand for Academics (SODA) Control Center

The screenshot shows the SAS OnDemand for Academics (SODA) Control Center dashboard. At the top right, there are links for "United States" and a user profile for "Stefanie Reay". The main title is "SAS® OnDemand for Academics Dashboard". Below the title, there are two tabs: "Planned Events" and "Notices", with "Planned Events" being the active tab. A navigation bar at the top has three items: "Applications", "Enrollments", and "Courses", with "Applications" being the active tab. The main content area displays five software applications with their icons and descriptions:

- 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 "Support Site", "Step-by-Step Reference Guides", and "Frequently Asked Questions". Below that is a "Quotas" section with two progress bars:

- Home Directory (46.5MB/5120MB) at 1%
- Course Directory (207.0MB/3072MB) at 7%

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



# Data in the SAS OnDemand for Academics environment

- To upload data for use in the SAS OnDemand for Academics environment, you must upload it through SAS Studio
- Once you upload the files in SAS Studio, they will be accessible through SAS Studio, SAS Enterprise Guide, and SAS Enterprise Miner



# Click on “SAS Studio” to start SAS Studio

The screenshot shows the SAS OnDemand for Academics Dashboard. At the top, there is a navigation bar with a home icon, "United States", and a user profile for "Stefanie Reay". Below the navigation bar, the title "SAS® OnDemand for Academics Dashboard" is displayed. There are two tabs at the top right: "Planned Events" and "Notices", with "Planned Events" being the active tab. A horizontal menu bar below the title includes "Applications", "Enrollments", and "Courses", with "Applications" being the active tab. The main content area lists several SAS applications:

- 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 of the dashboard, there is a "Reference" section with links to "Support Site", "Step-by-Step Reference Guides", and "Frequently Asked Questions". Below that is a "Quotas" section with two progress bars: "Home Directory (46.5MB/5120MB)" at 1% and "Course Directory (207.0MB/3072MB)" at 7%.

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

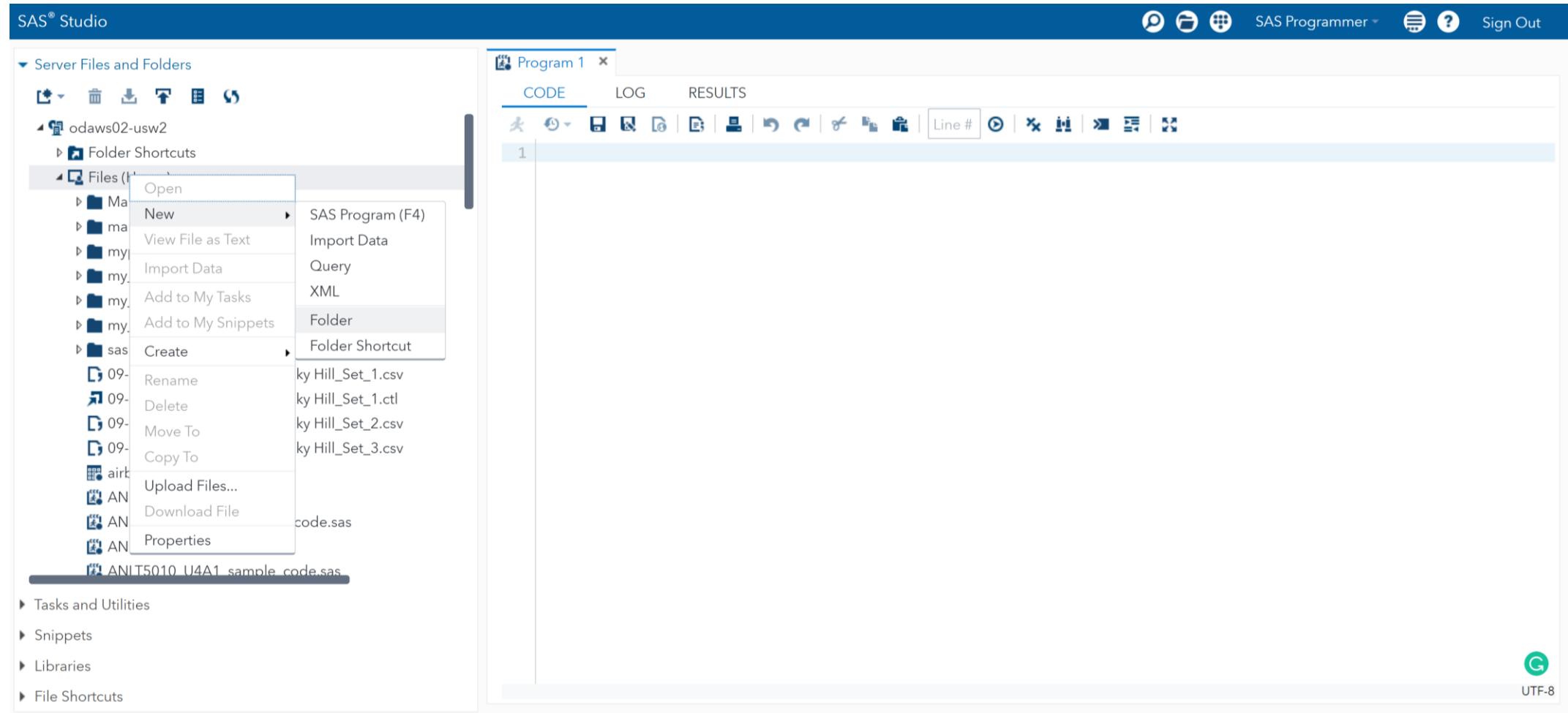


Click on “Files (Home)” to make the upload button appear in dark blue.

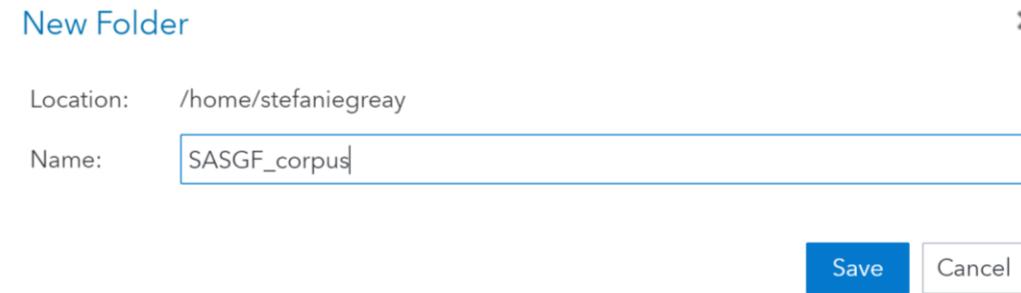
The screenshot shows the SAS Studio interface. On the left, there is a sidebar titled "Server Files and Folders" which includes options like "Tasks and Utilities", "Snippets", "Libraries", and "File Shortcuts". Below this, under "Server Files and Folders", there is a tree view showing a folder structure: "odaws02-usw2" has "Folder Shortcuts" and "Files (Home)". The "Files (Home)" item is highlighted with a blue selection bar. The main workspace is titled "Program 1" and contains tabs for "CODE", "LOG", and "RESULTS". The "CODE" tab is active, showing a single line of code "1". The "RESULTS" tab is visible but empty. At the bottom right of the workspace, there is a green circular icon with a white letter "G" and the text "UTF-8". The top navigation bar includes icons for help, search, and connectivity, along with user information "SAS Programmer" and "Sign Out".



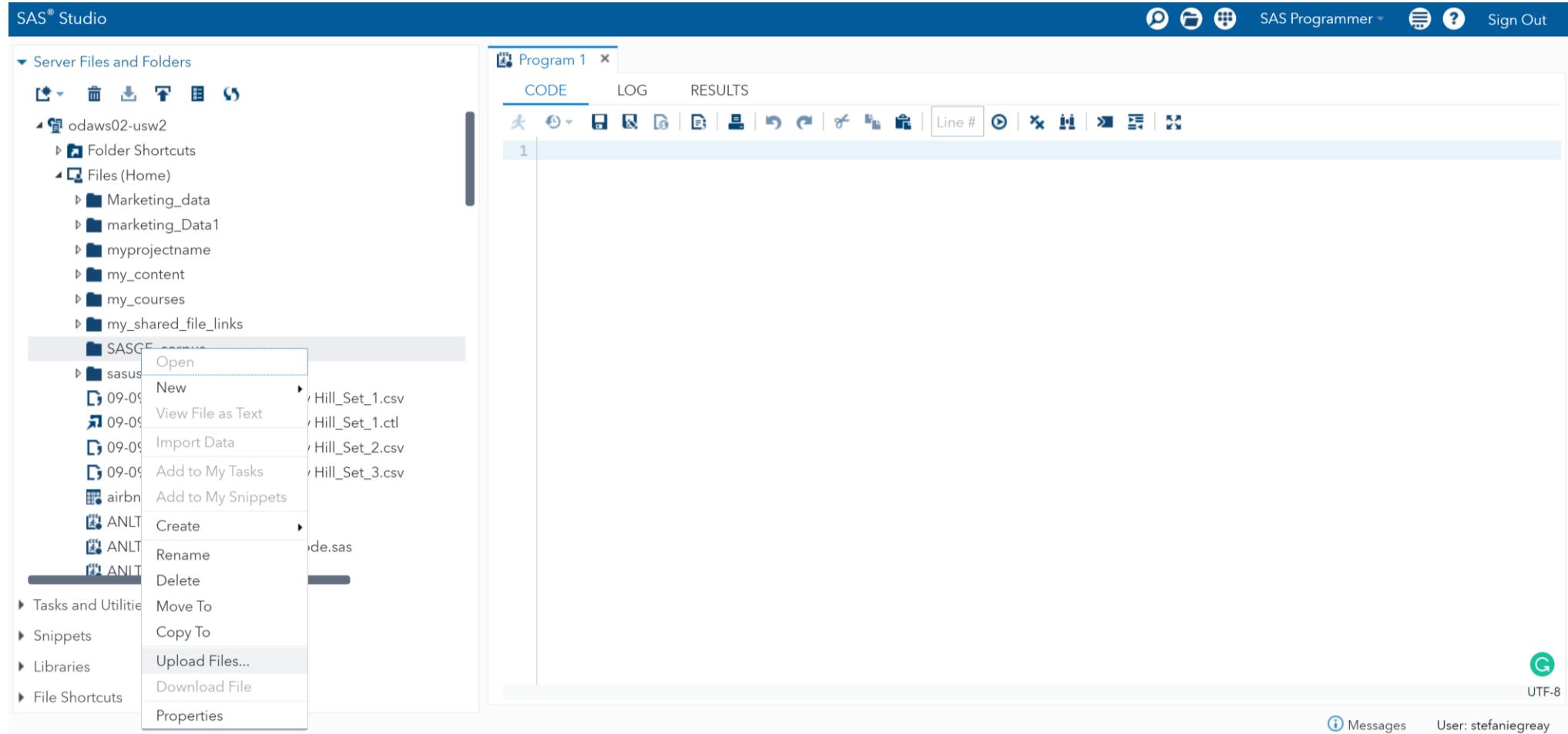
# Right click on “Files(Home)” and select “New” and then “Folder”



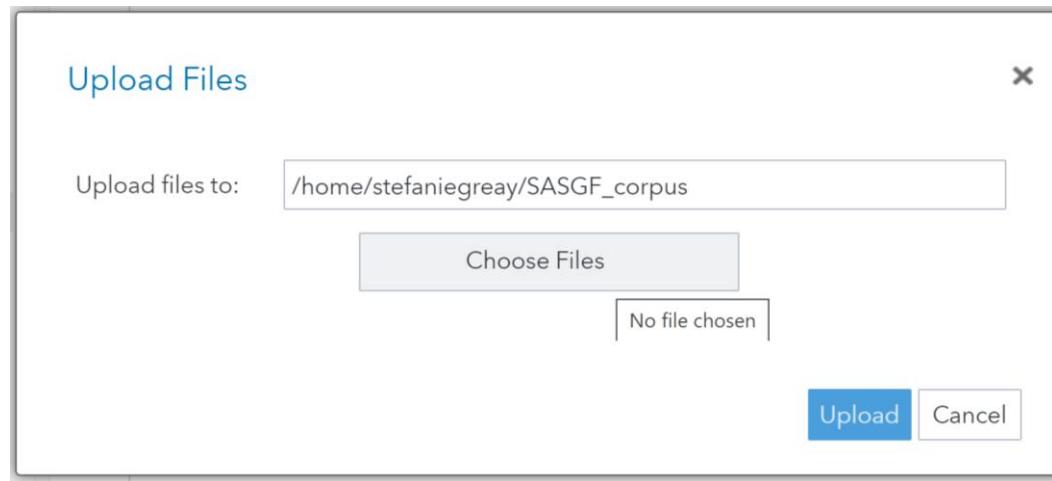
Type a name for your new folder, then click  
“Save.”



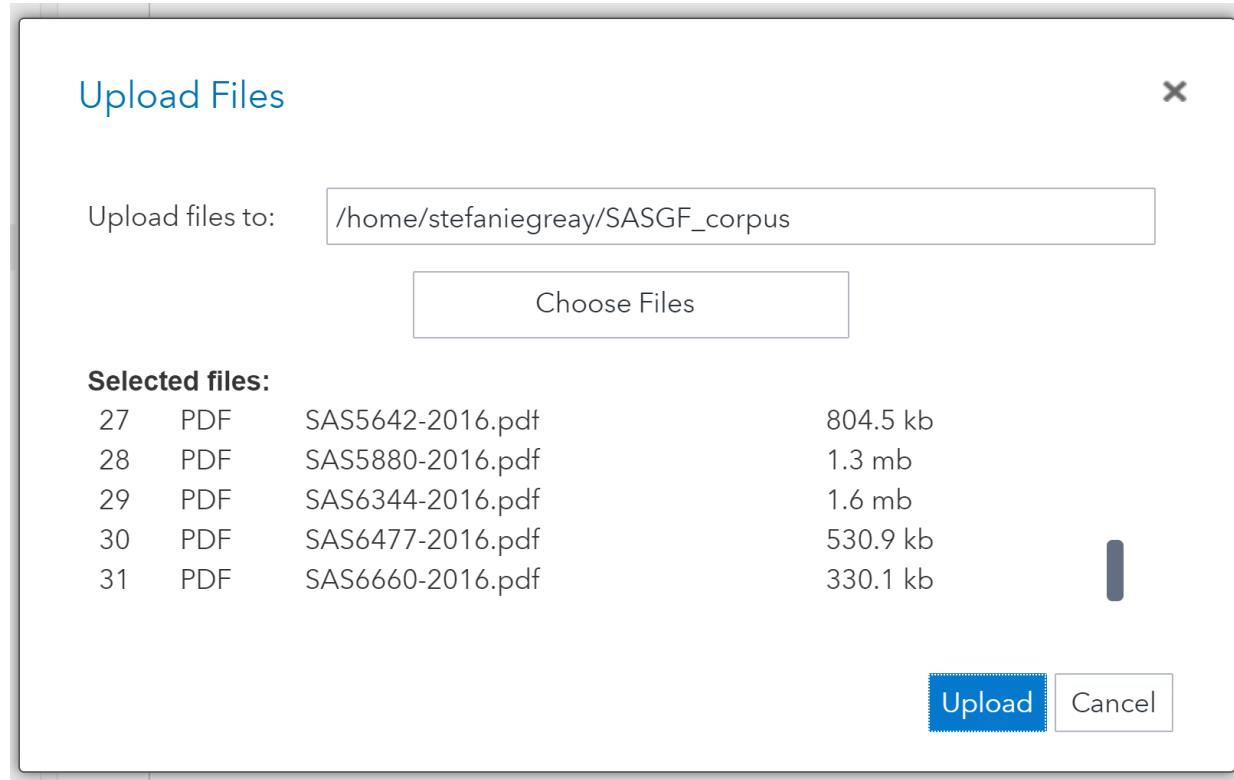
# Right click on your new folder, and choose “Upload Files.”



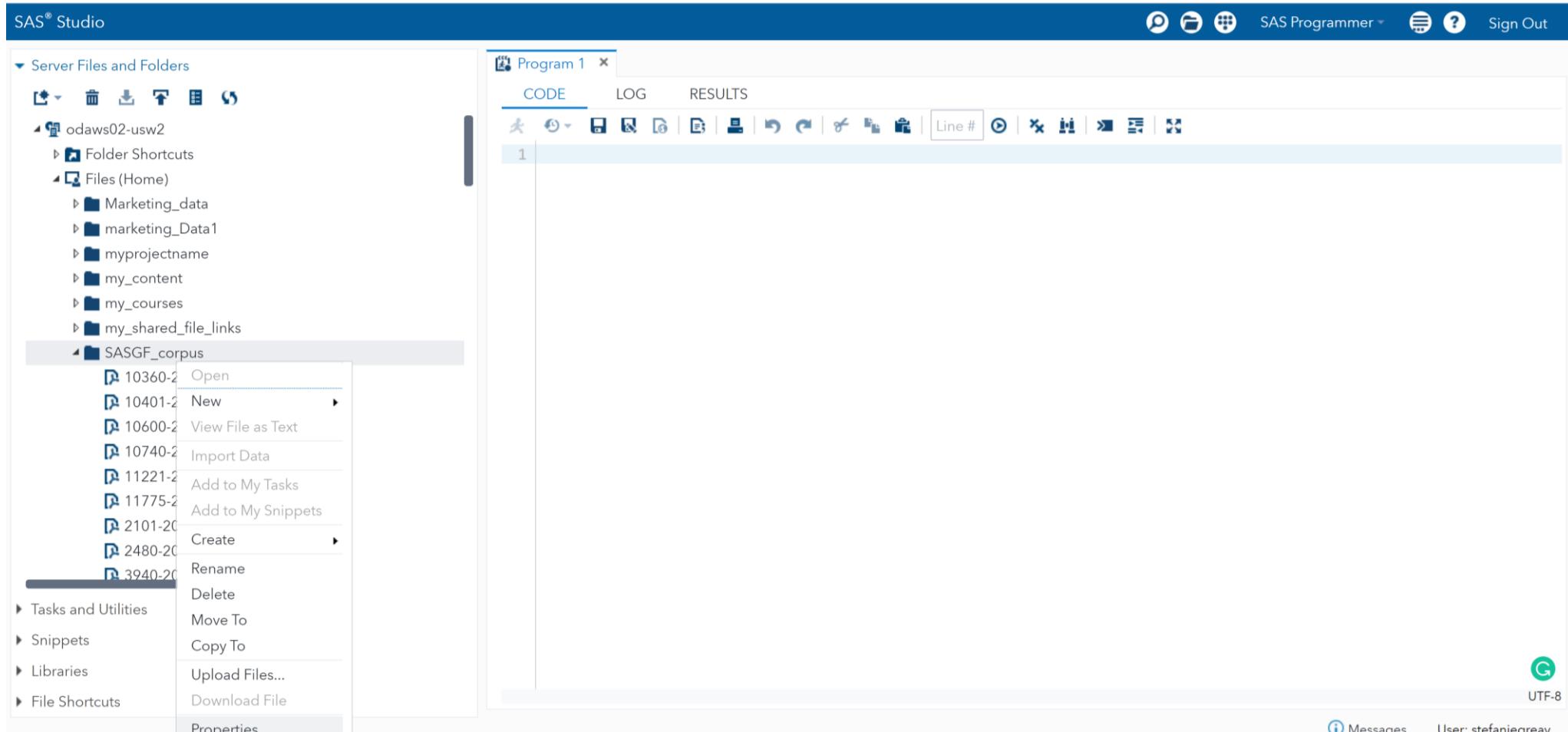
# Click “Choose Files.”



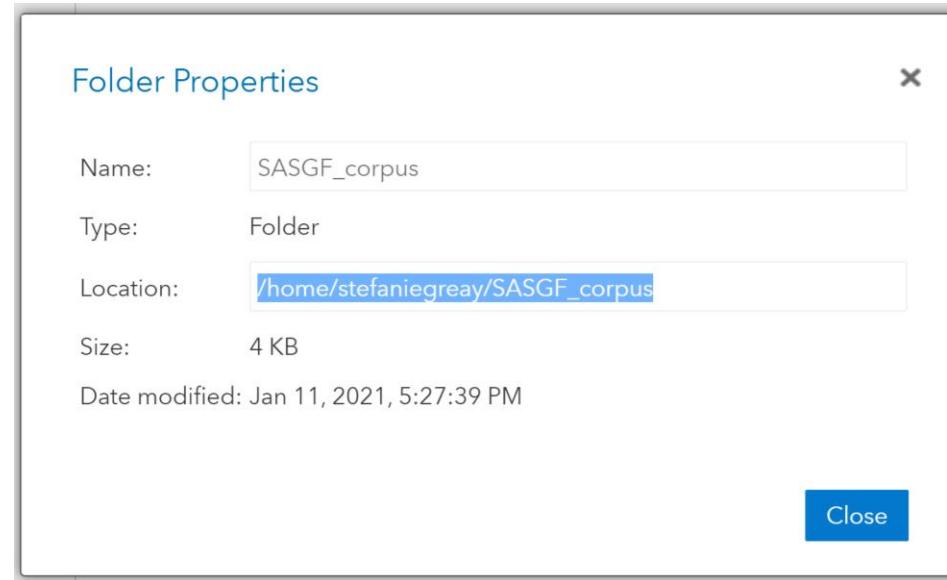
# Choose all 31 pdf files, and click on “Upload.”



# Right click on the new folder, and select “properties.”



Select and copy the folder location and save it to a text file (for use in the following steps), then click “Close.”



# Open SAS Enterprise Miner

- You have now uploaded the 31 pdf files to the SAS On Demand for Academics server and saved the folder location.
- The next step is to open SAS Enterprise Miner and access the data to start the text mining process.



# Click on SAS Enterprise Miner and then click “Keep”

The screenshot shows the SAS® OnDemand for Academics Dashboard. At the top, there are tabs for "Planned Events" and "Notices". Below this, a navigation bar has tabs for "Applications", "Enrollments", and "Courses", with "Applications" being the active tab. The main content area displays five application cards:

- 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 [Support Site](#), [Step-by-Step Reference Guides](#), and [Frequently Asked Questions](#). Below that, "Quotas" are shown for "Home Directory" (46.5MB/5120MB) at 1% and "Course Directory" (207.0MB/3072MB) at 7%. A warning message at the bottom left says: "This type of file can harm your computer. Do you want to keep main (3).jnlp anyway?" with buttons for "Keep" and "Discard".



# Click on “main.jnlp” to start validation and installation

The screenshot shows the SAS® OnDemand for Academics Dashboard. At the top, there's a navigation bar with the SAS logo, a dropdown for 'United States', a user profile for 'Stefanie Reay', and a dropdown menu. Below the header, the title 'SAS® OnDemand for Academics Dashboard' is displayed, along with 'Planned Events' and 'Notices' buttons.

The main content area features a navigation bar with tabs: 'Applications' (which is selected), 'Enrollments', and 'Courses'. The 'Applications' section lists several 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)

To the right, there's a 'Reference' section with links to 'Support Site', 'Step-by-Step Reference Guides', and 'Frequently Asked Questions'. Below that is a 'Quotas' section showing usage statistics:

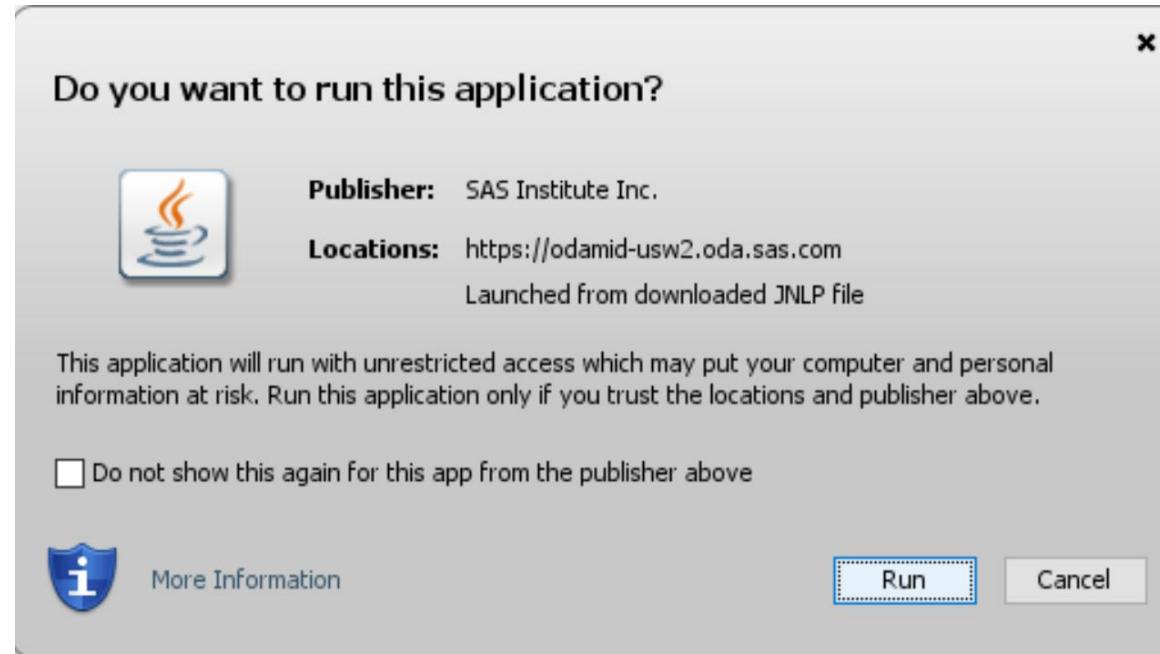
- Home Directory (46.5MB/5120MB) at 1%
- Course Directory (207.0MB/3072MB) at 7%

At the bottom, a taskbar shows a file named 'main (3).jnlp' and includes 'Show all' and 'X' buttons. A small speaker icon in the bottom right corner indicates audio is available.

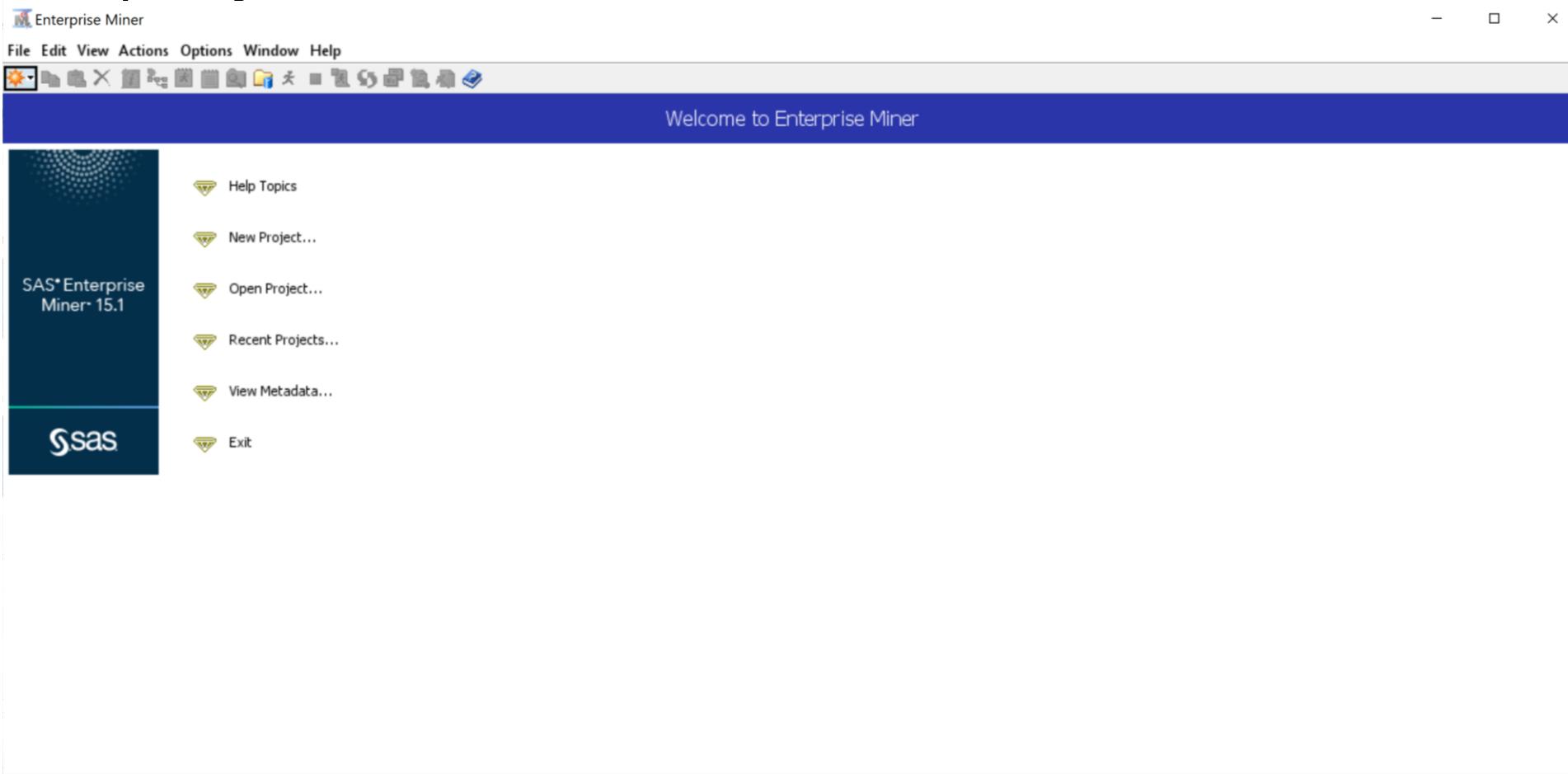
You will see a “Starting application...” window pop up.



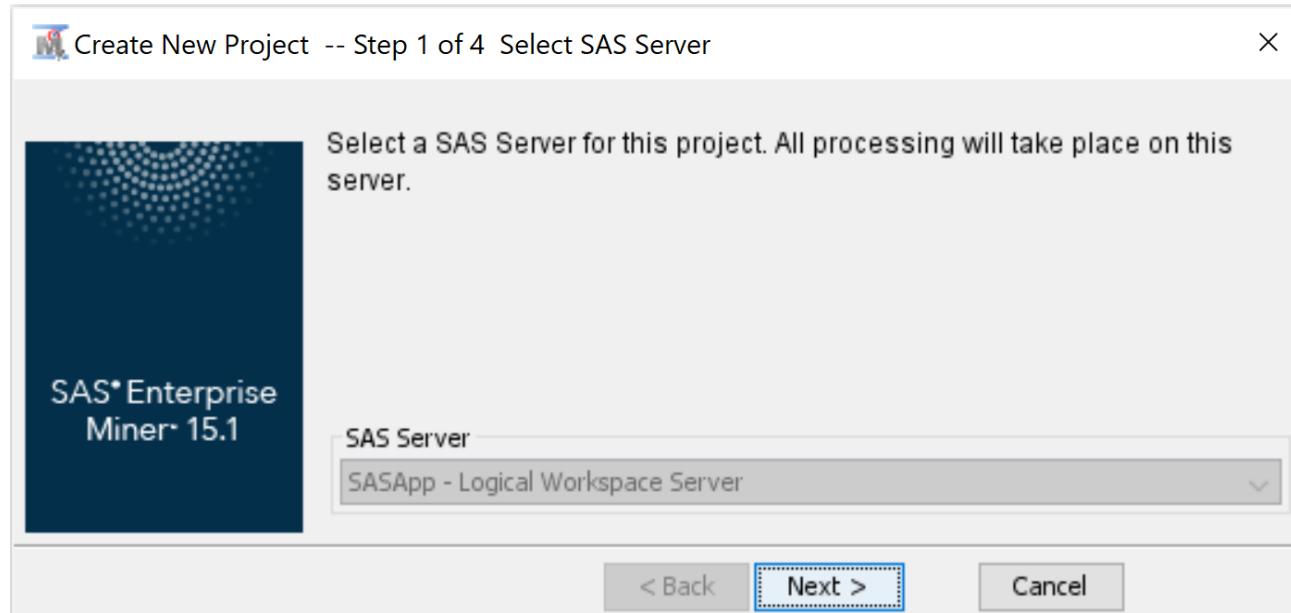
You will see a “Starting application...” window pop up. Click “Run.”



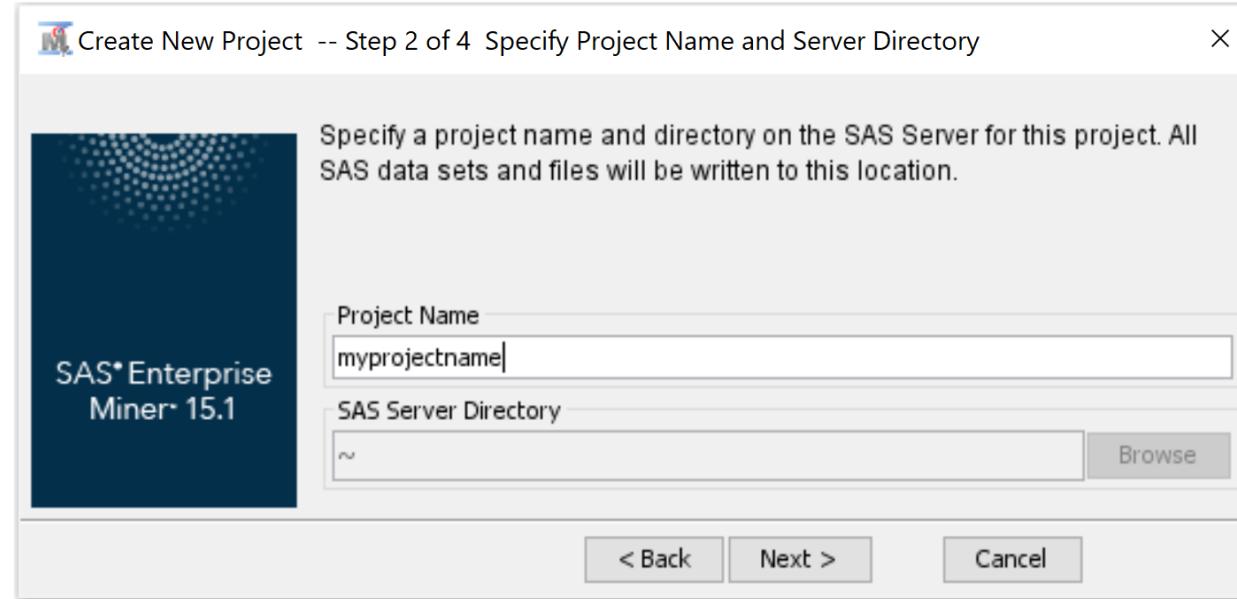
The Enterprise Miner Application will then pop up. Clicking on “New Project” will allow you to start a new project.



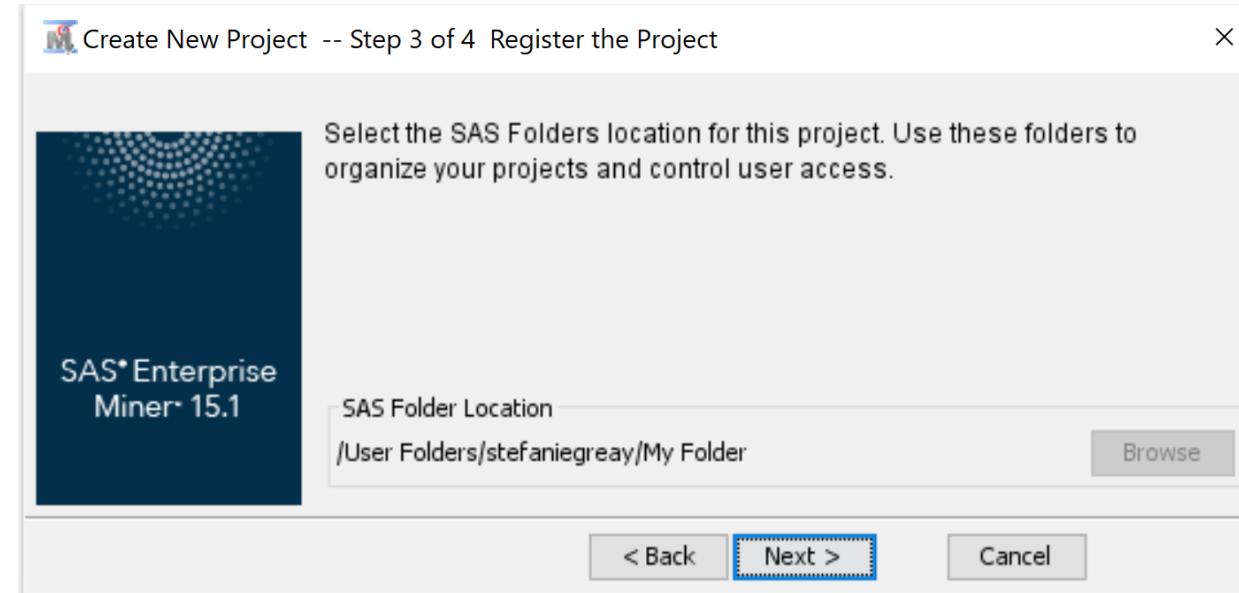
Click on “Next >” to continue setting up the project using the default SASApp server.



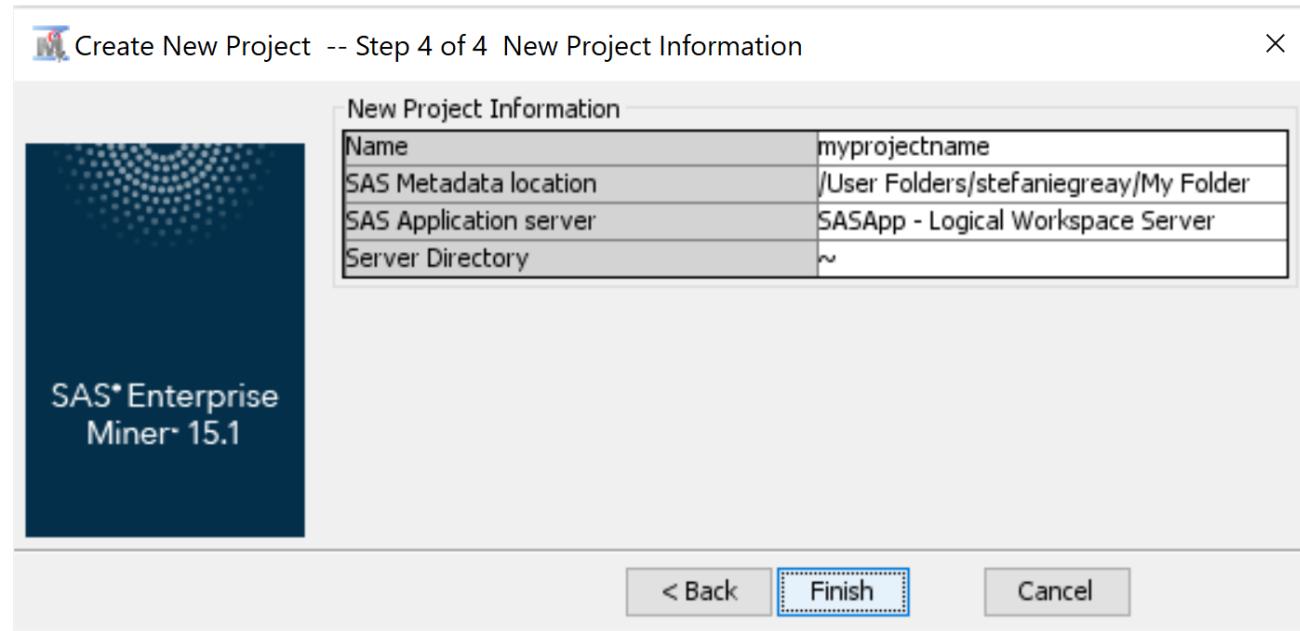
Enter a name for your project, then click “Next >.”



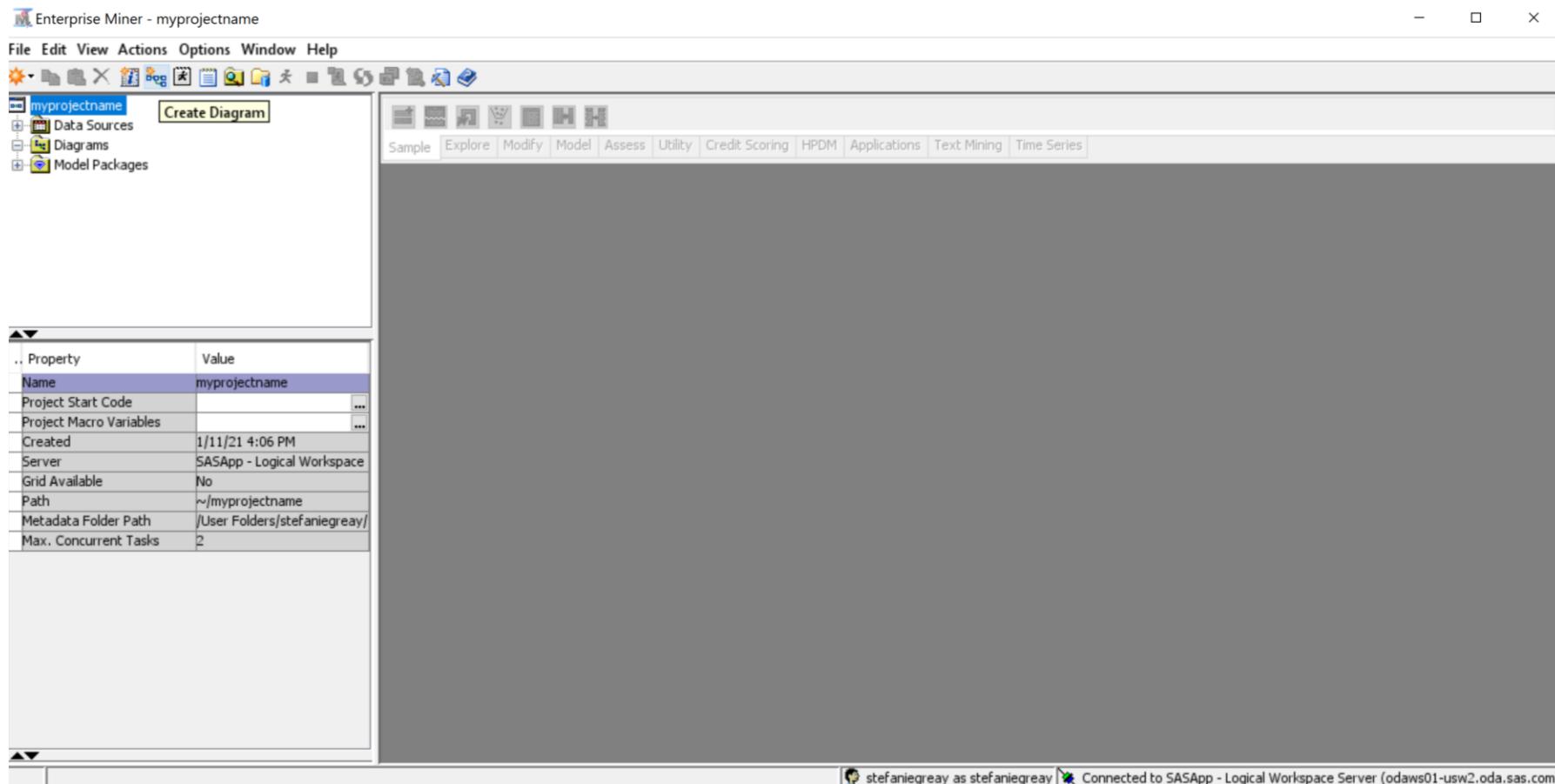
Click “Next >” to save the project in the default location.



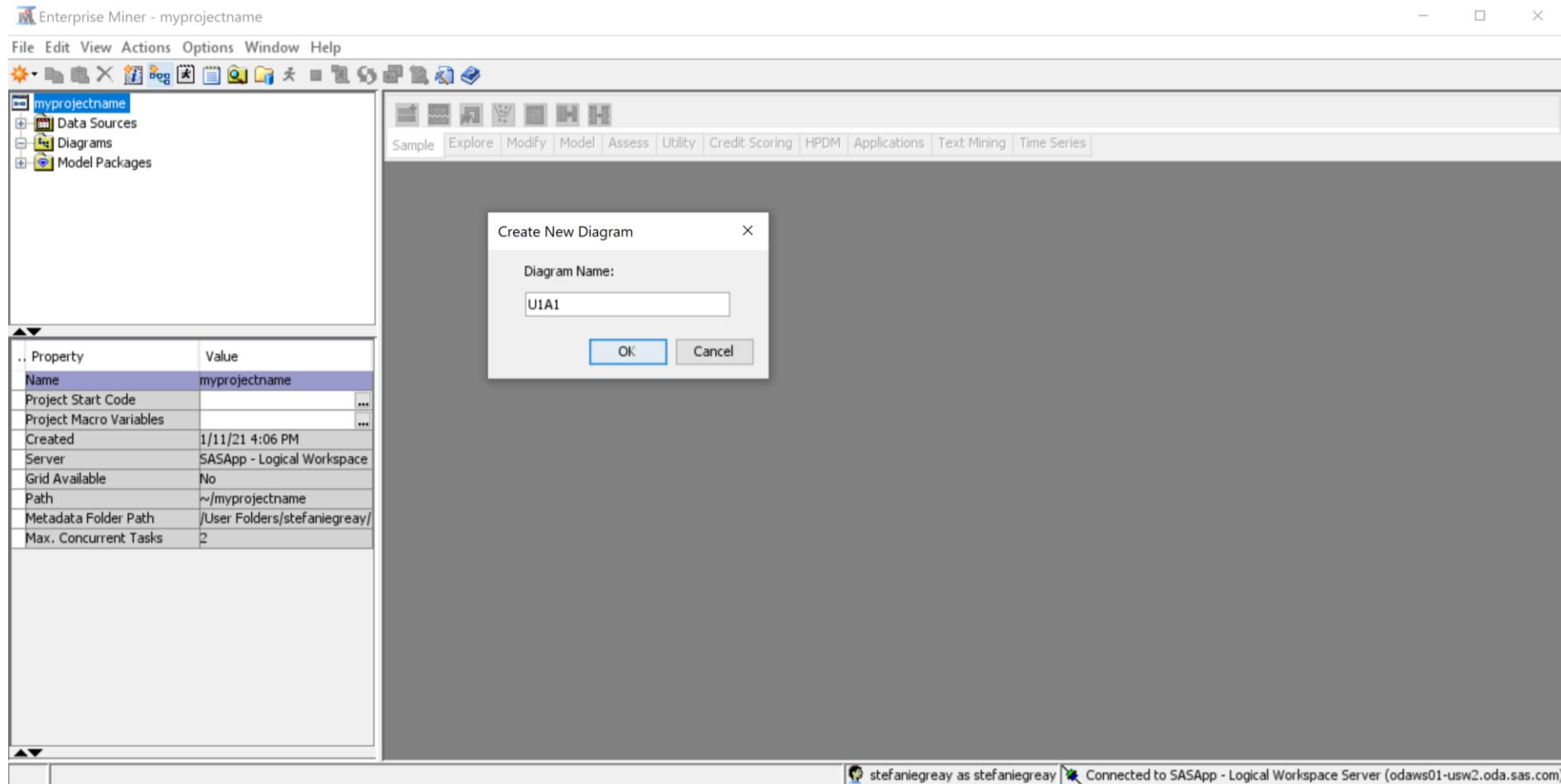
# Click “Finish” to complete the SAS Enterprise Miner project setup.



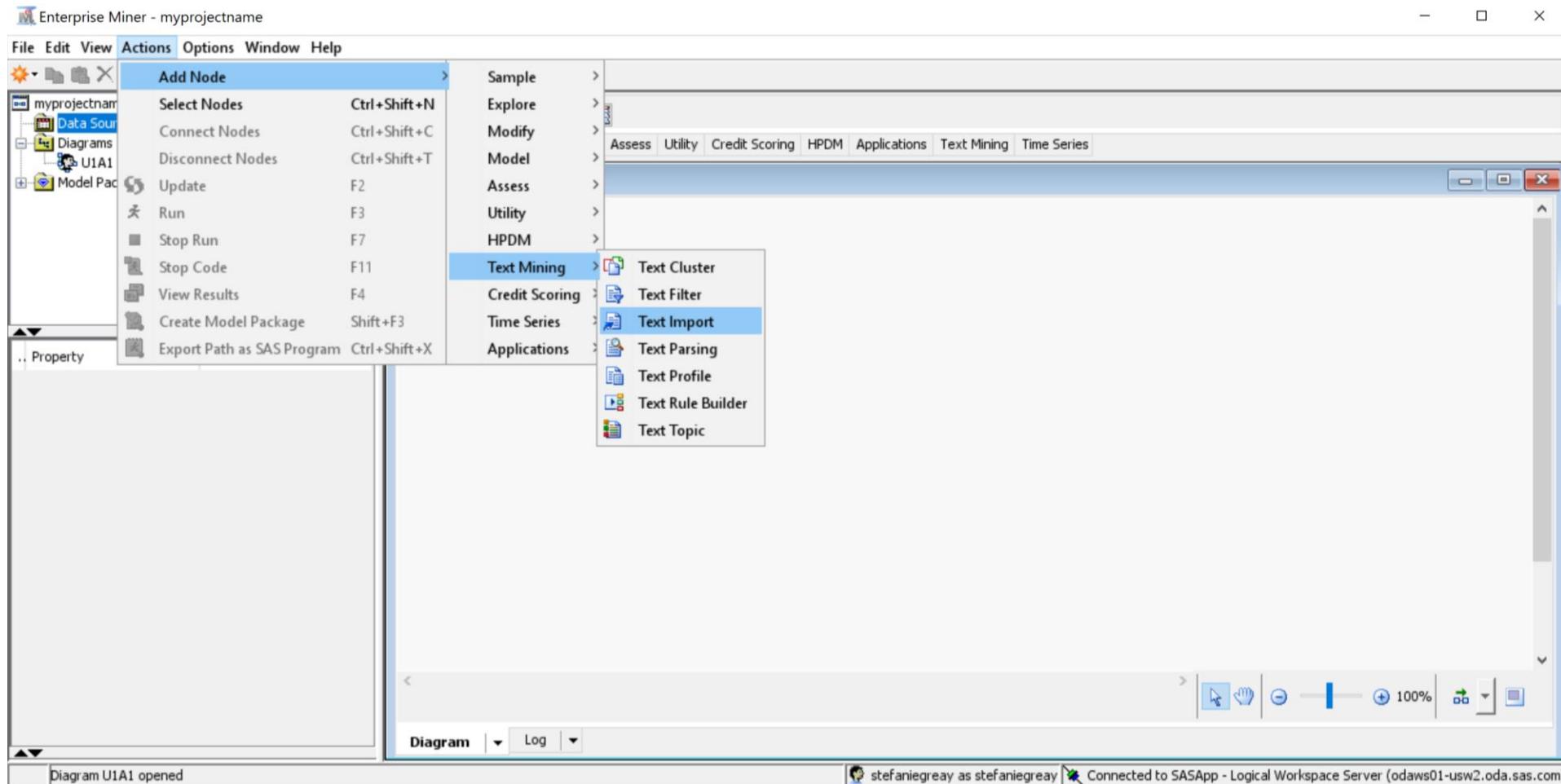
You can create a diagram using the icon directly to the left of the notebook with the running guy.



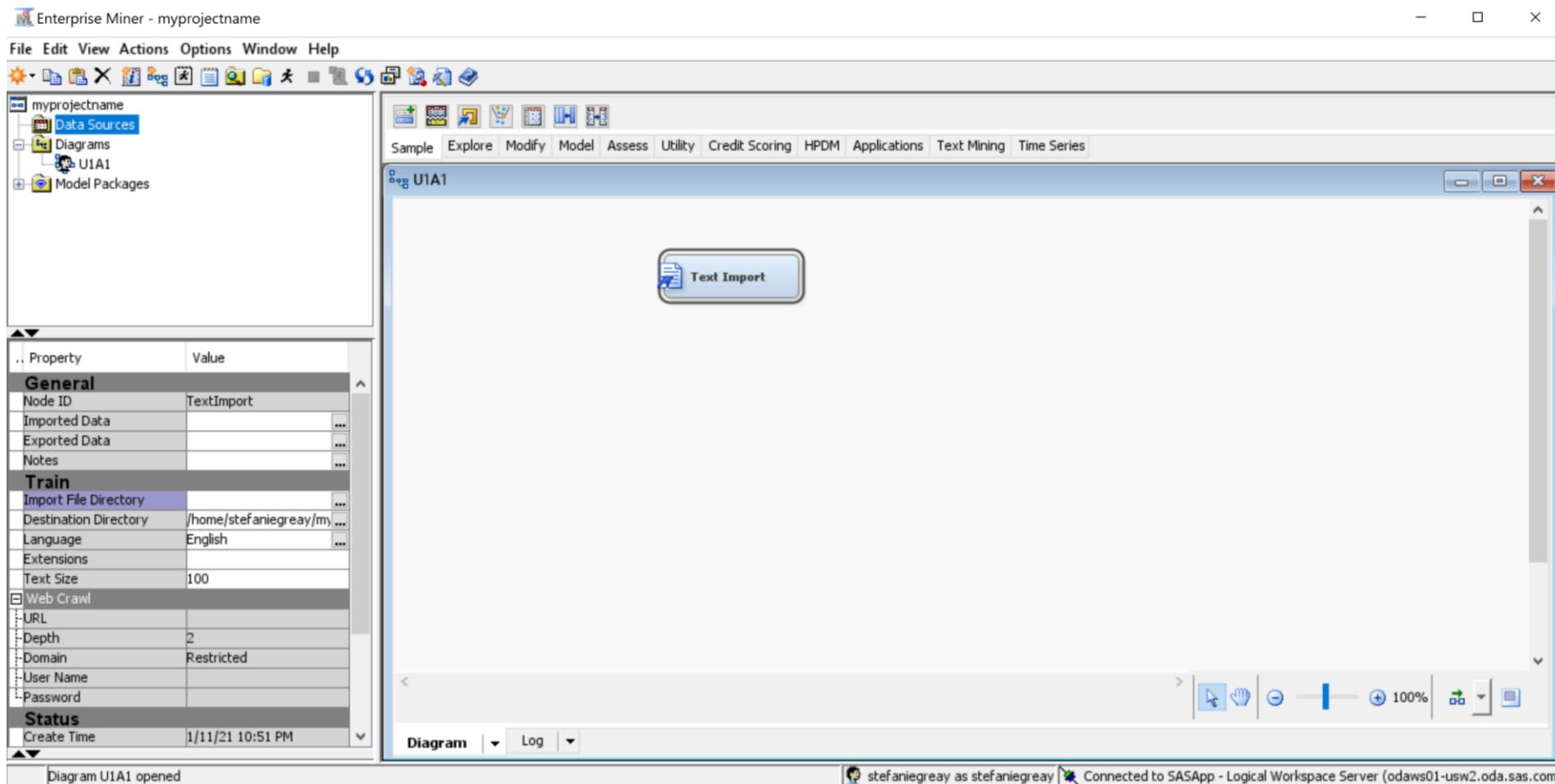
# Name the diagram in the dialog box that pops up, then click “OK.”



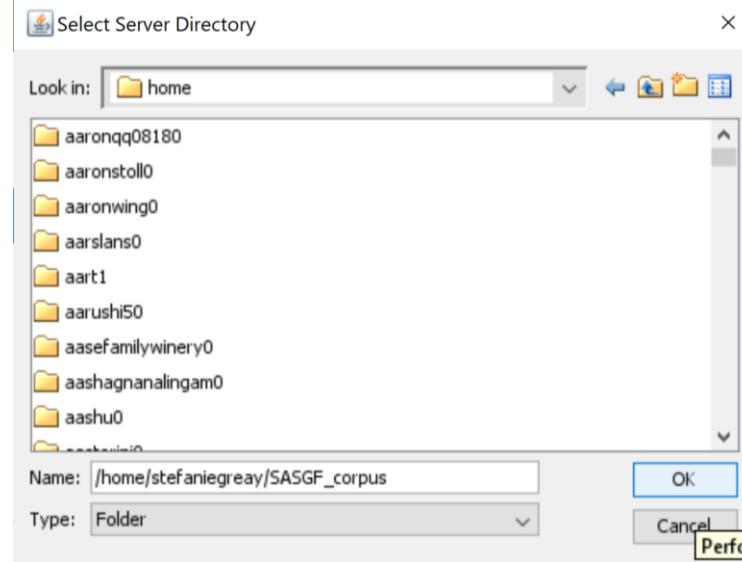
Click on “Action” > “Add Node” > “Text Mining” > “Text Import”



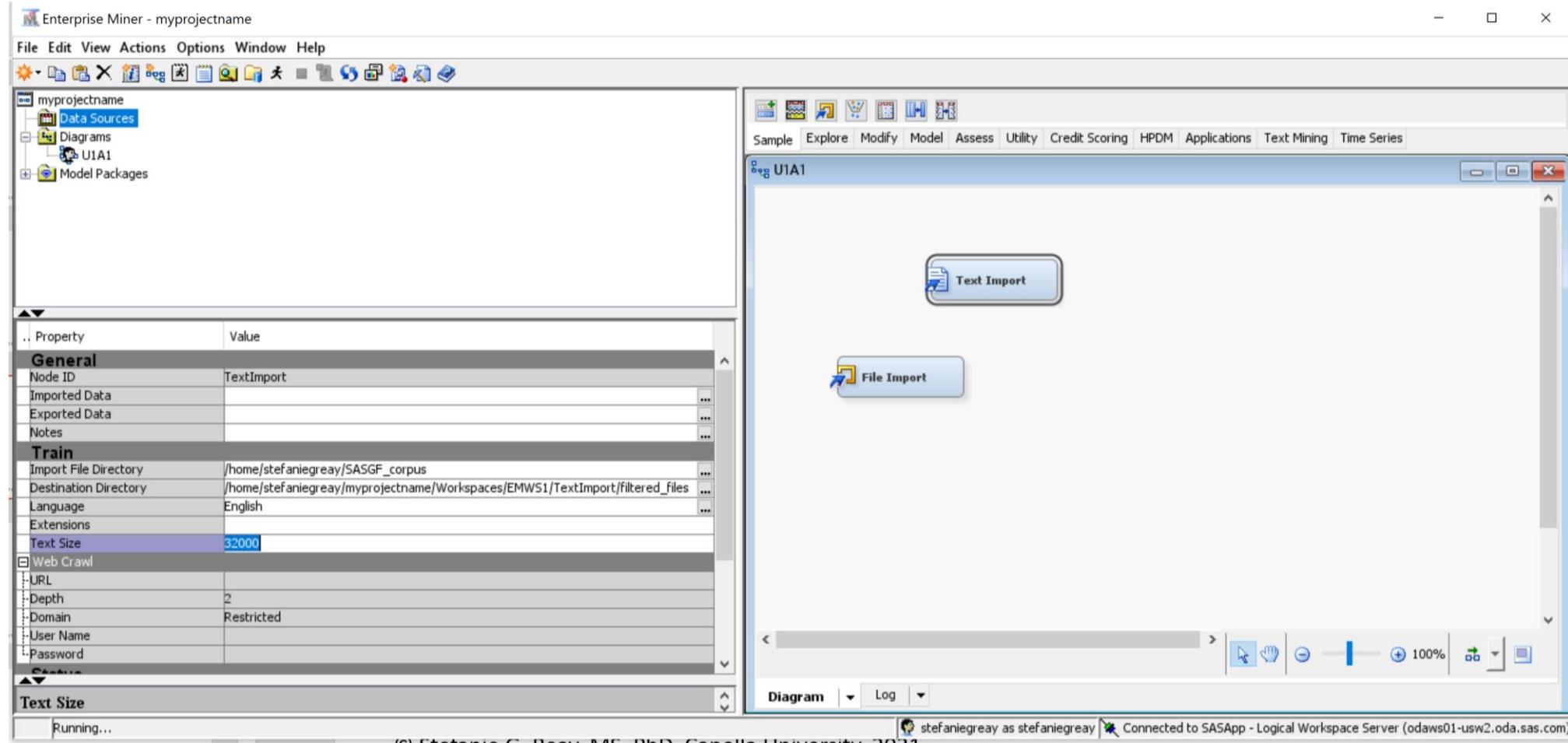
Click on the “Text Import” node you just added, and click the 3 dots next to “Import File Directory.”



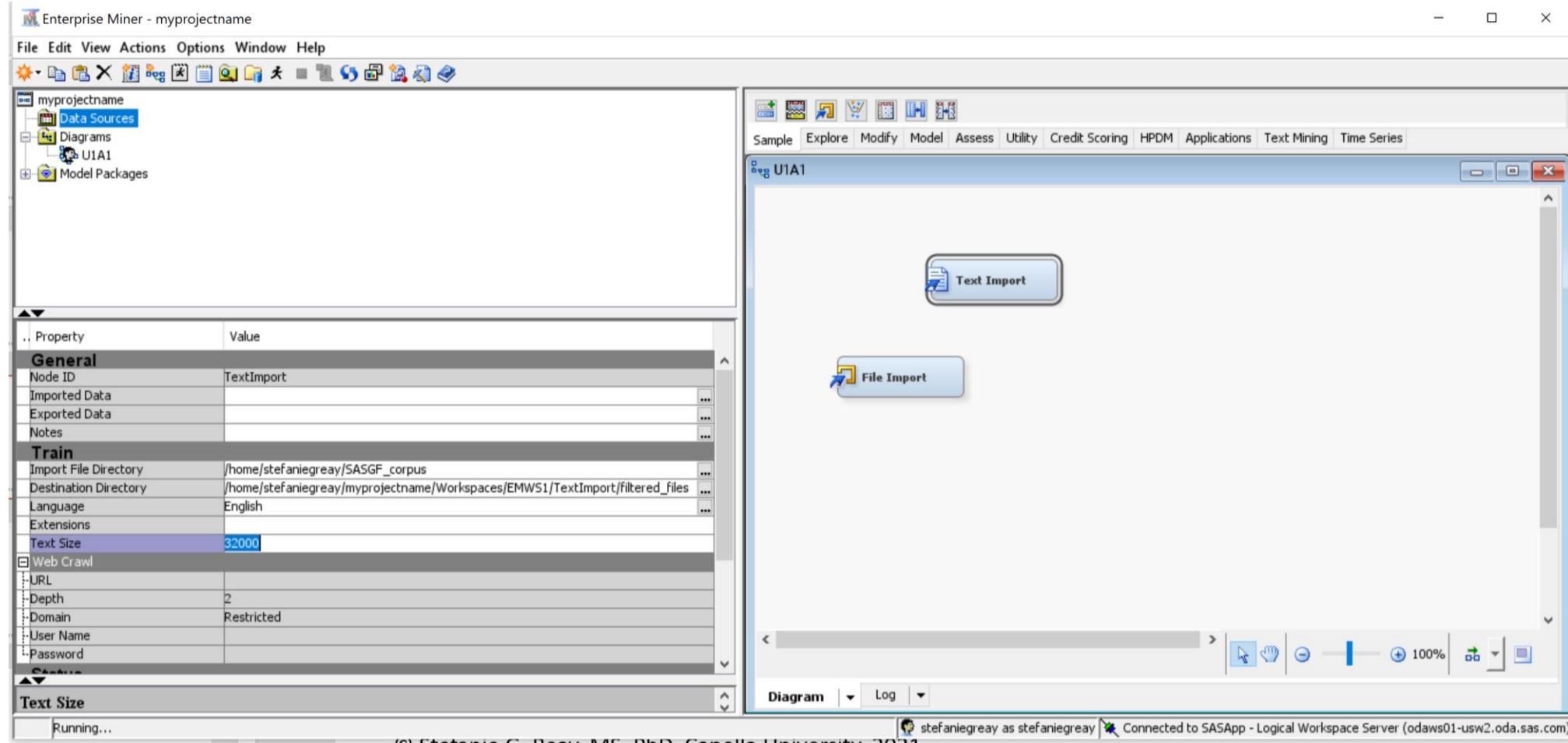
Paste the SAS OnDemand server directory you uploaded your files to in the box next to “Name:”



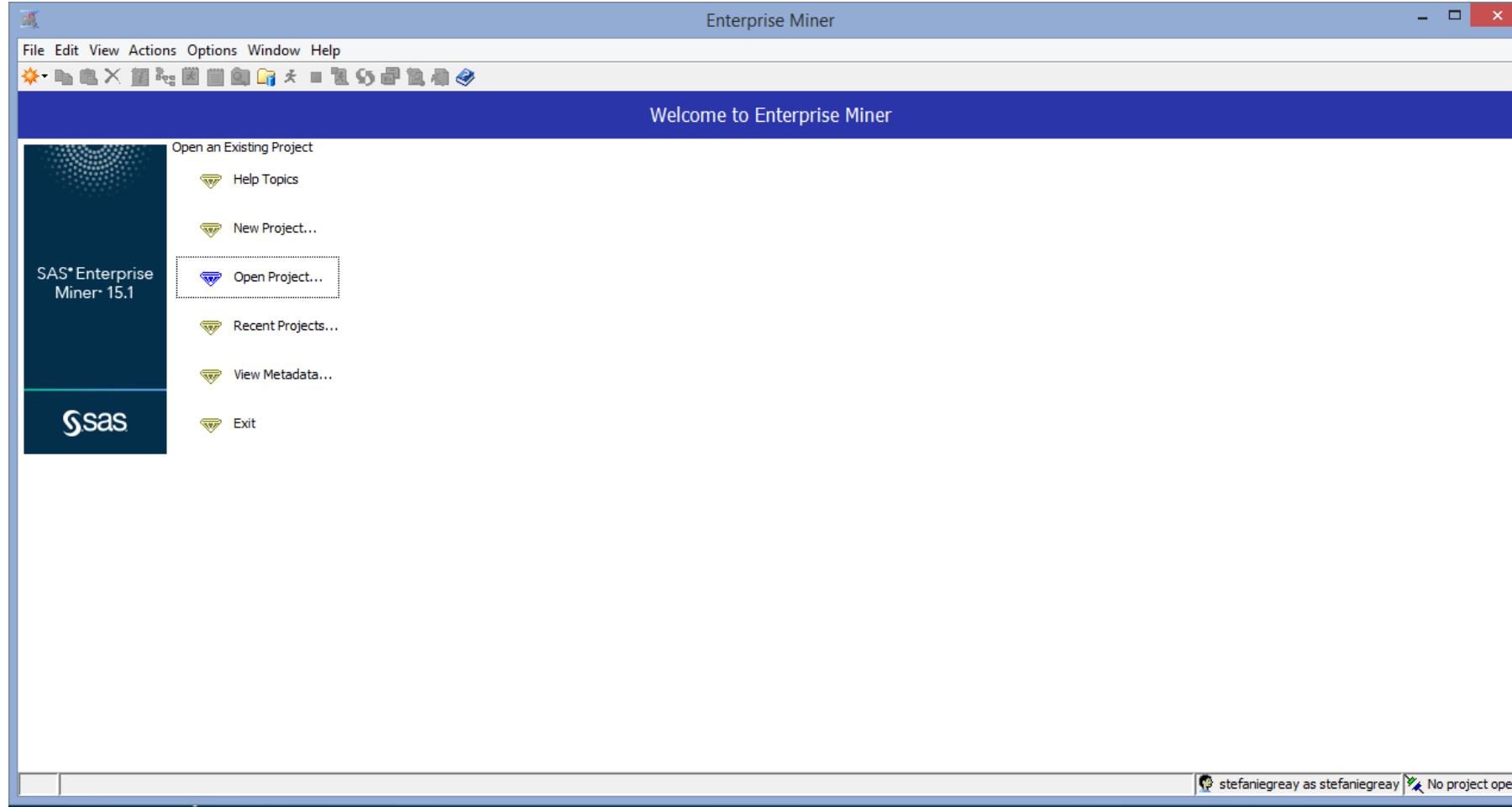
# Change the text size to 32,000, which is the maximum possible size.



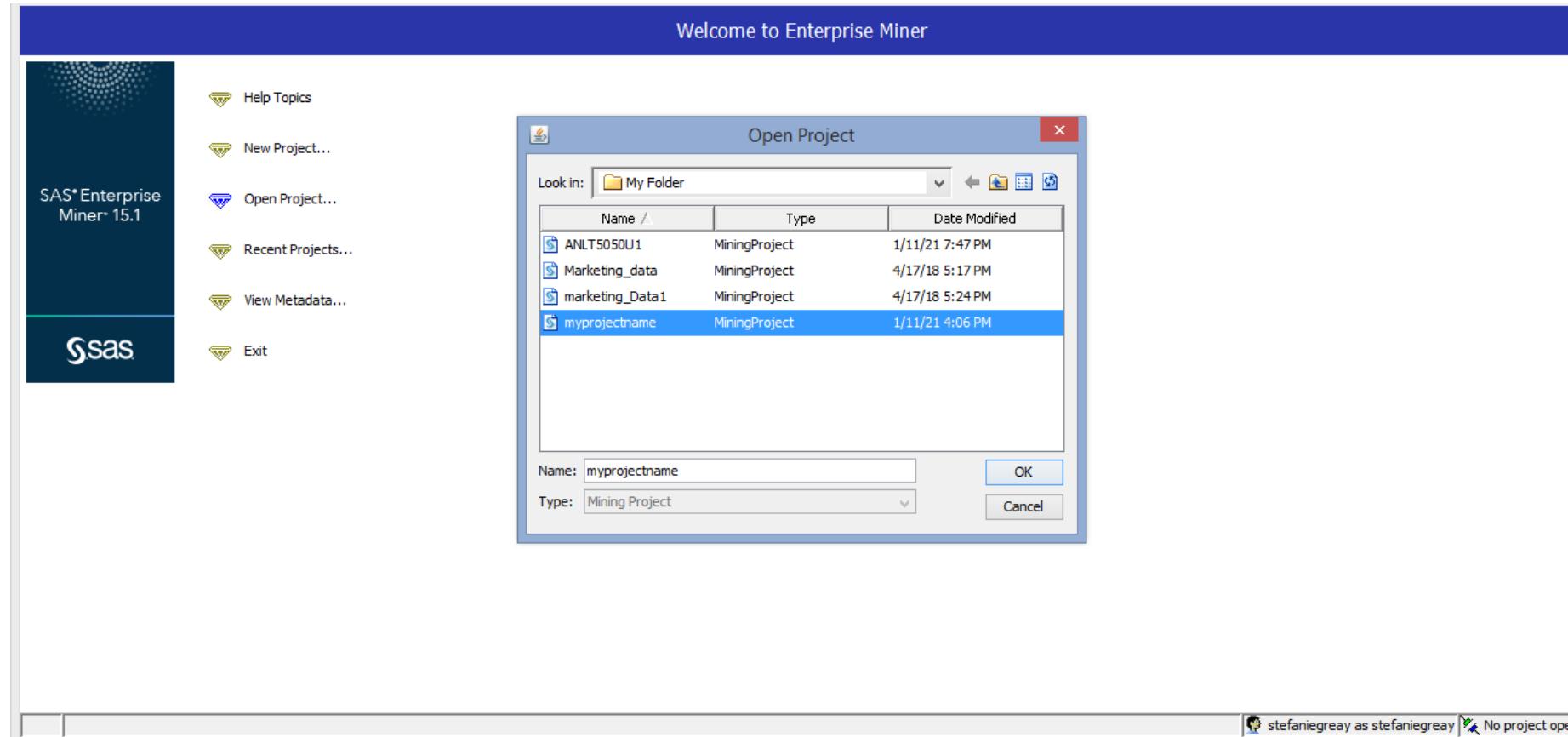
# Change the text size to 32,000, which is the maximum possible size.



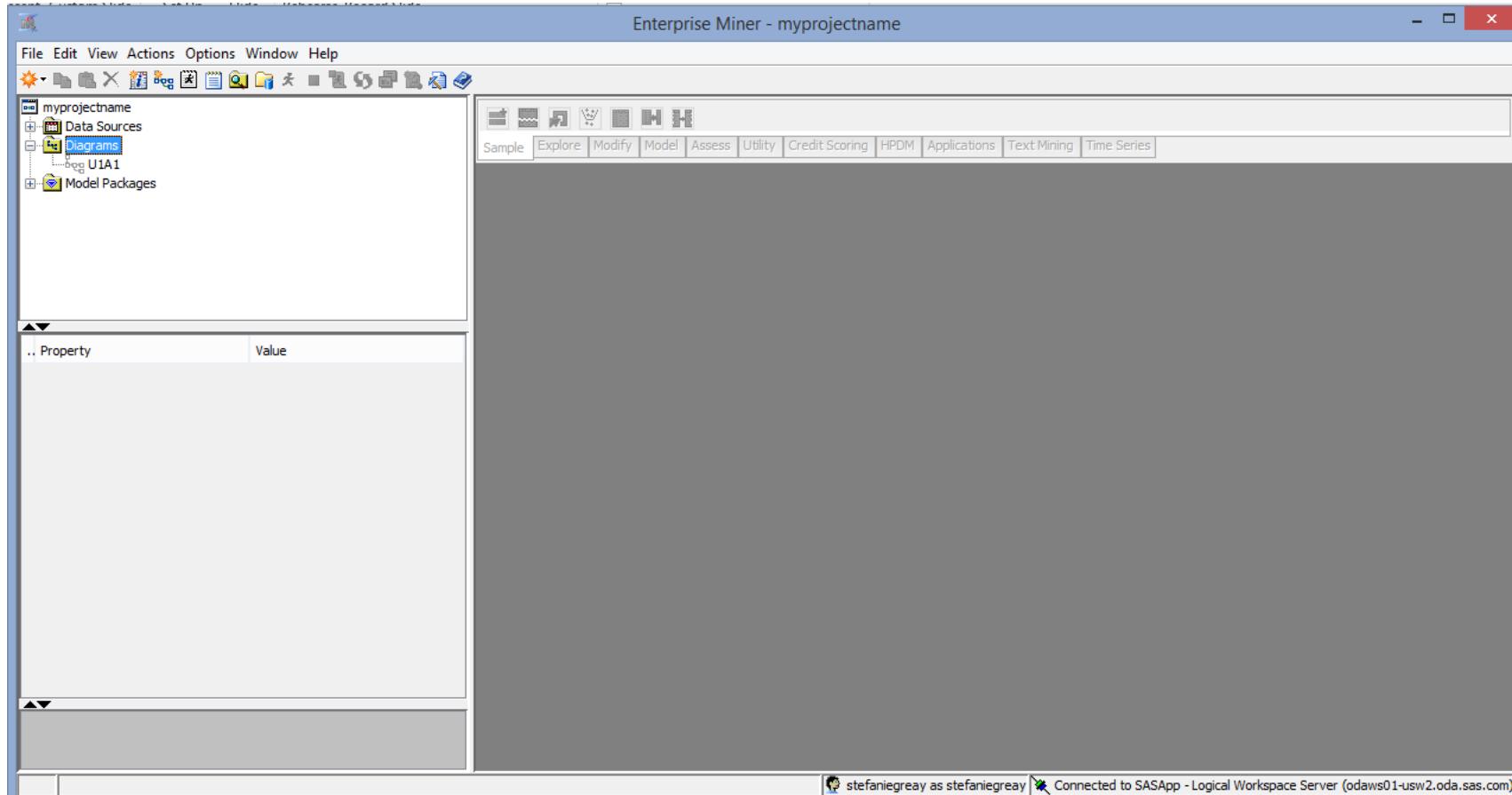
# START HERE if you successfully completed U1A1 – open project



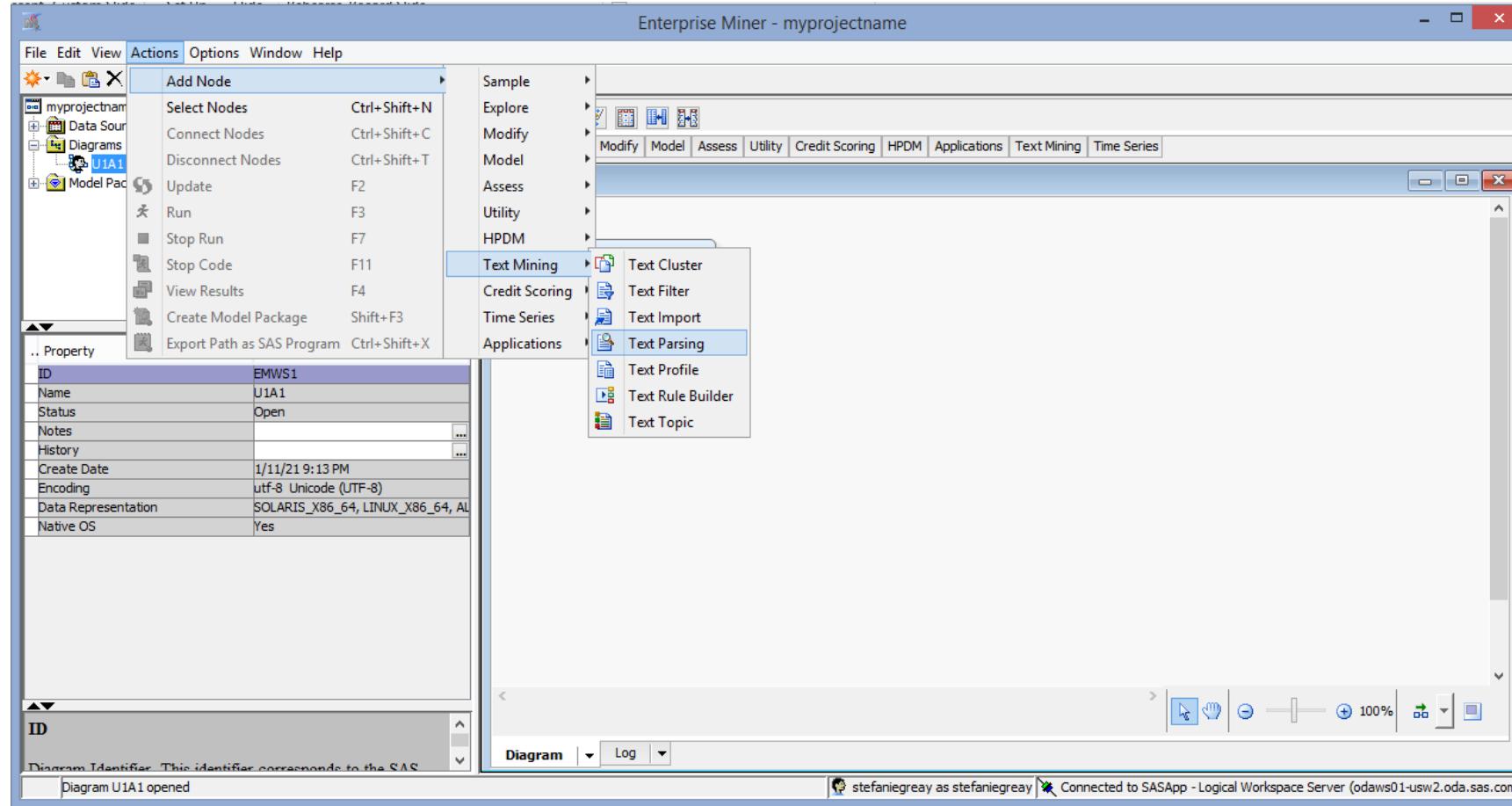
Select the project you created in U1A1 (or just now), and click “OK” to open the project.



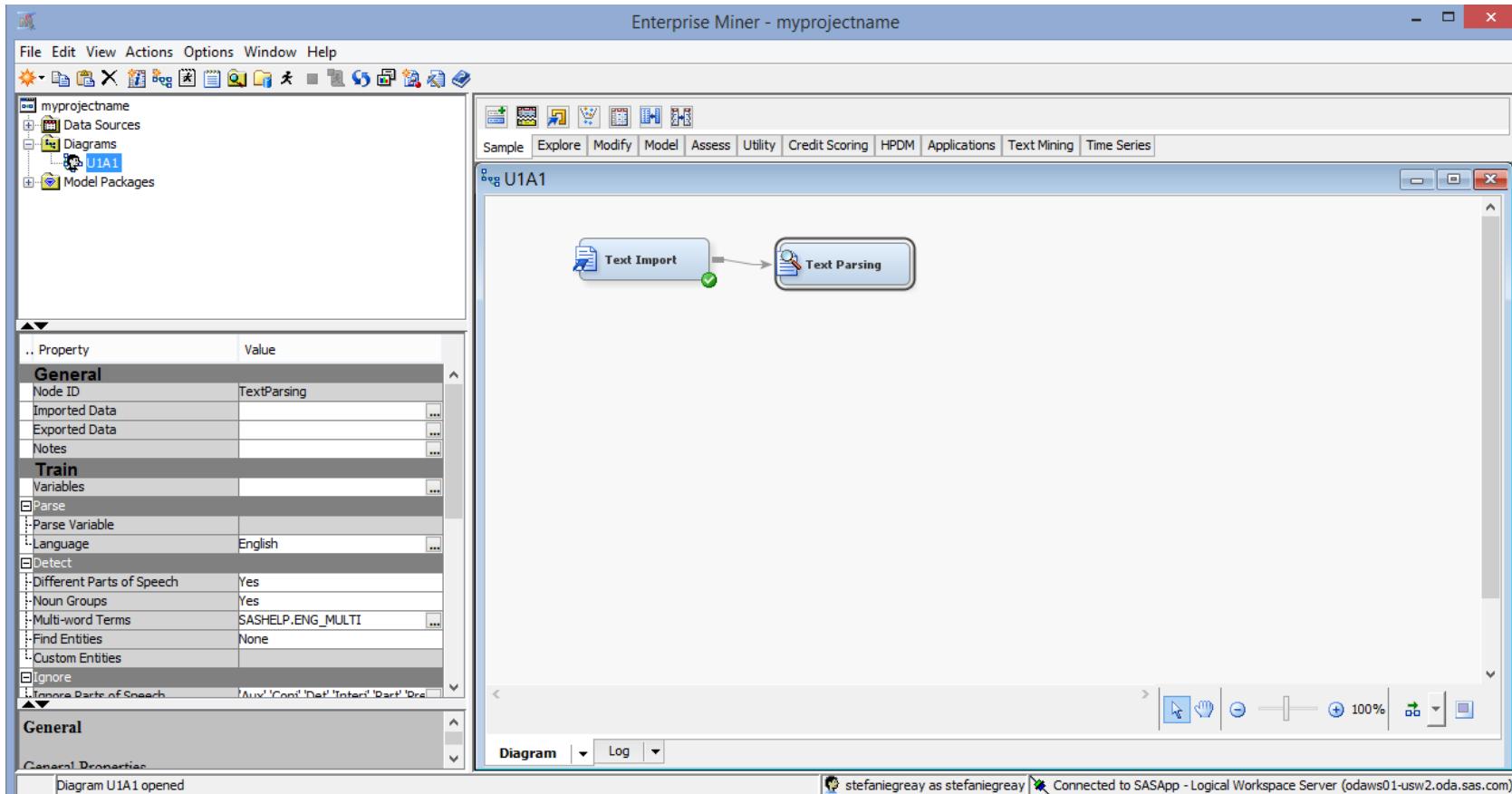
Click on “Diagrams,” navigate to the diagram you created, and double click on it to open it.



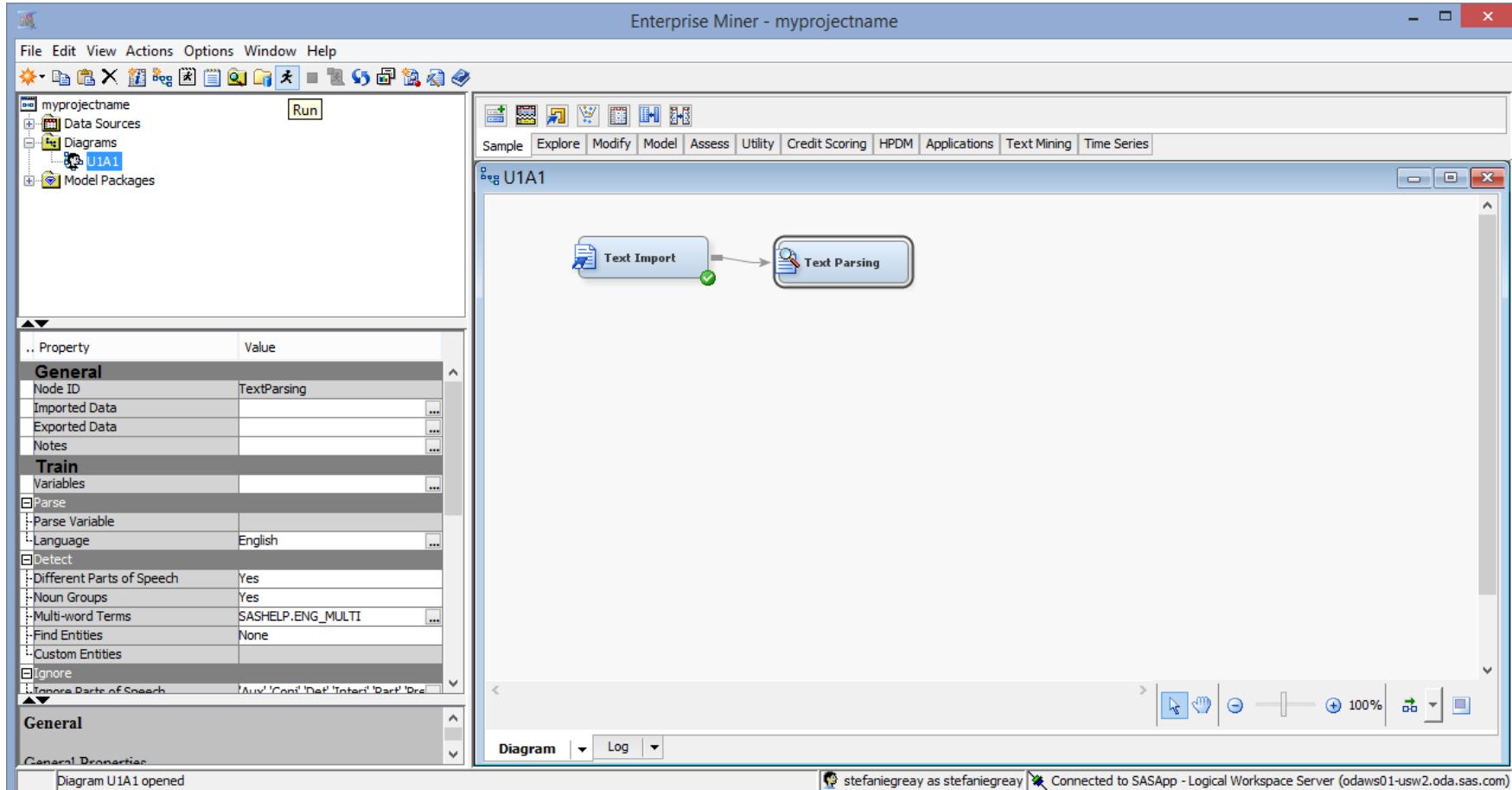
Add the text parsing node by clicking on  
Actions>Add Node>Text Mining>Text Parsing



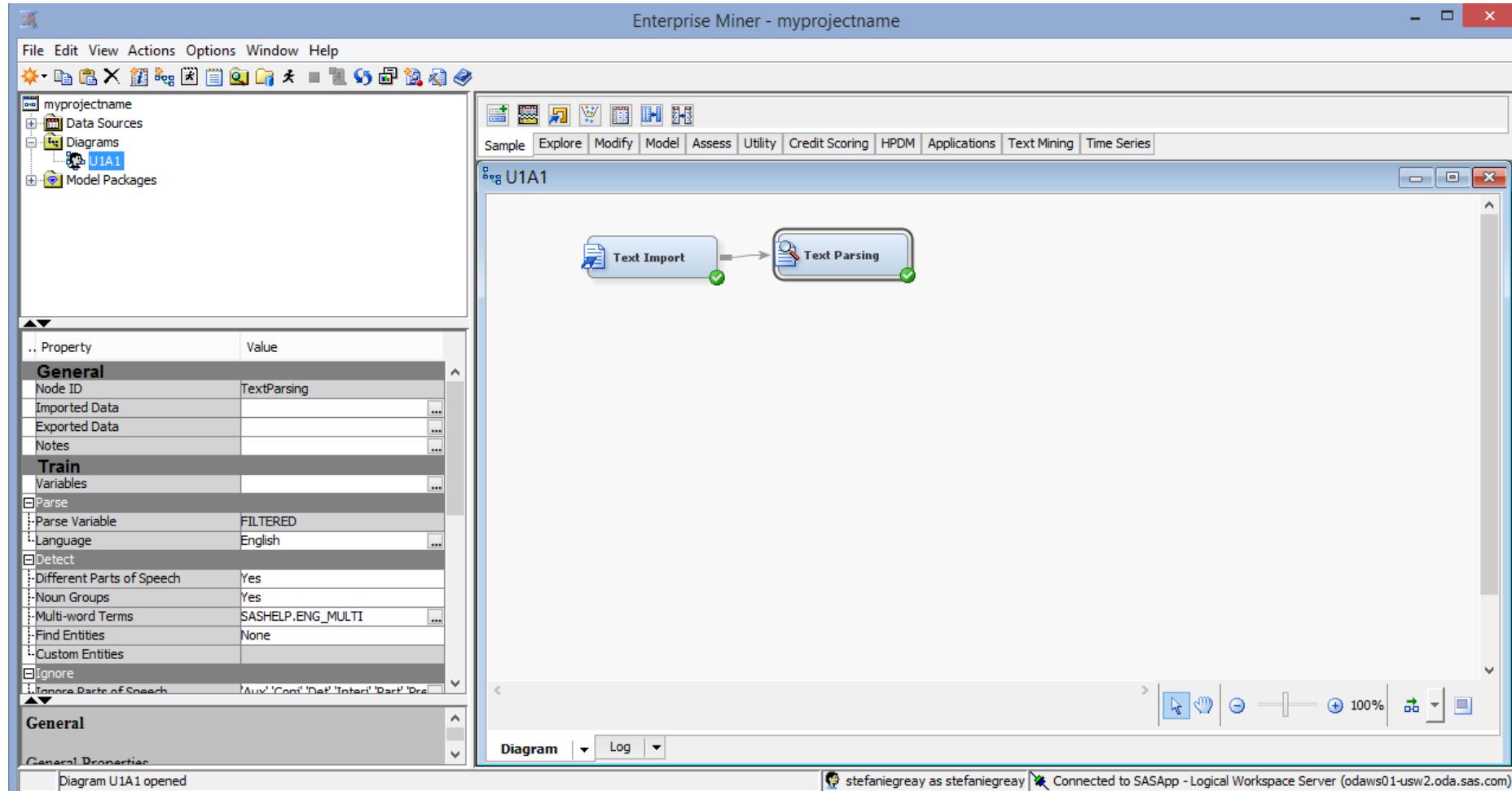
Drag the Text Parsing node to the right of the Text Import, and connect it using the connection box and arrow.



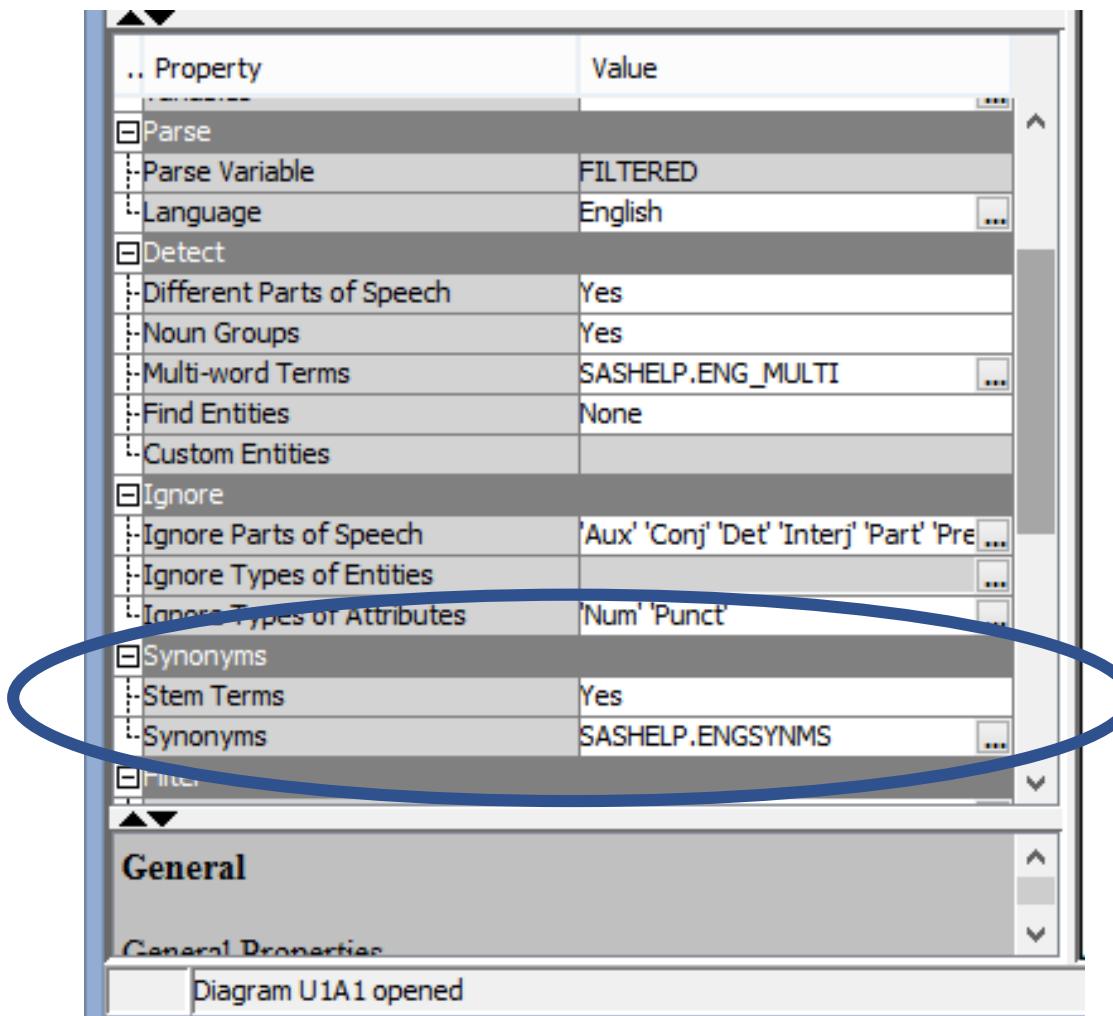
Run the Text Parsing node, so it recognizes the connection and the input from the Text import node, by clicking on “Text Parsing” and clicking the run guy.



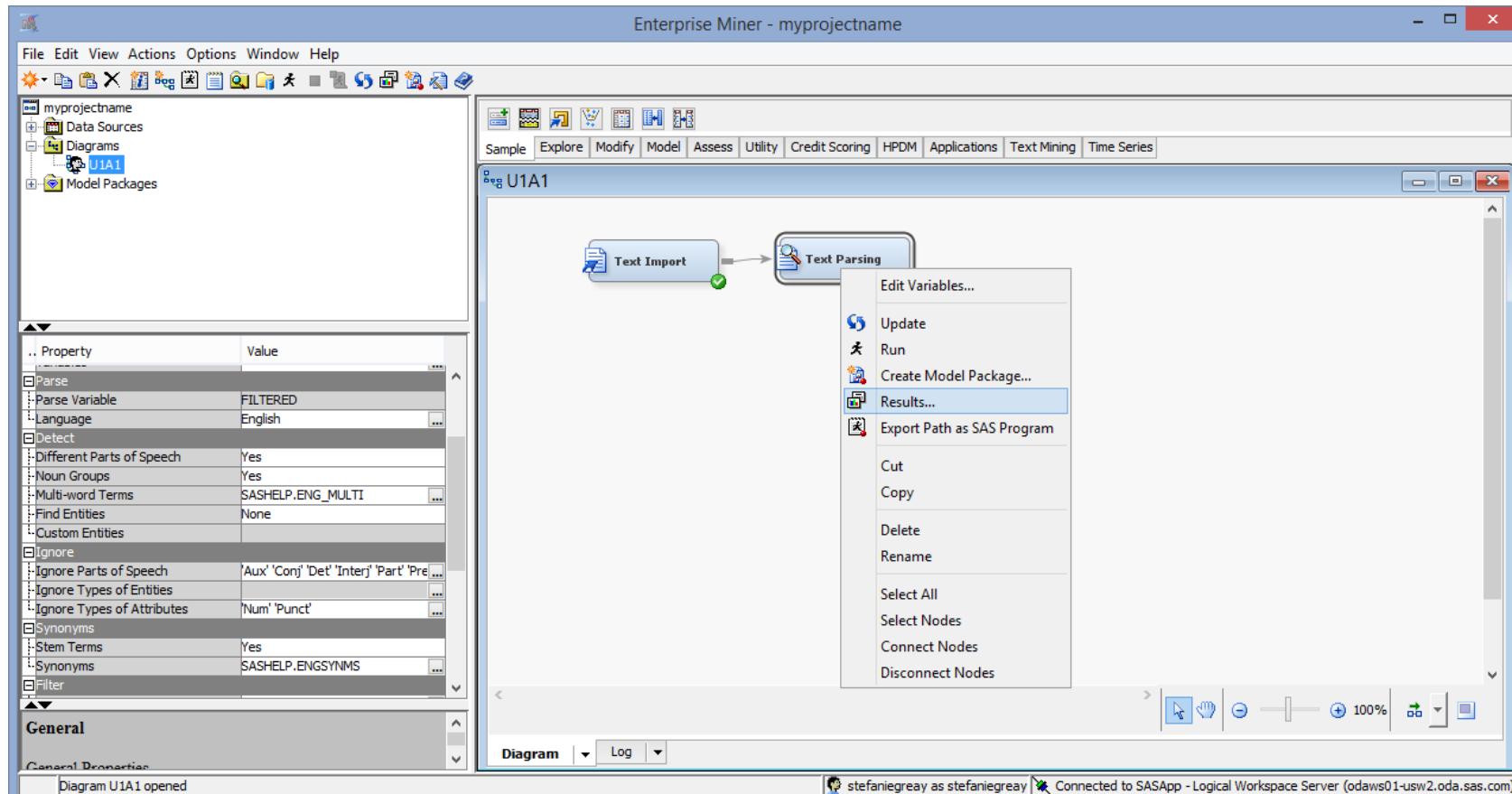
You will know that it was successful when you see the green circle with a check mark, and you see “FILTERED” next to the Parse Variable in the left menu.



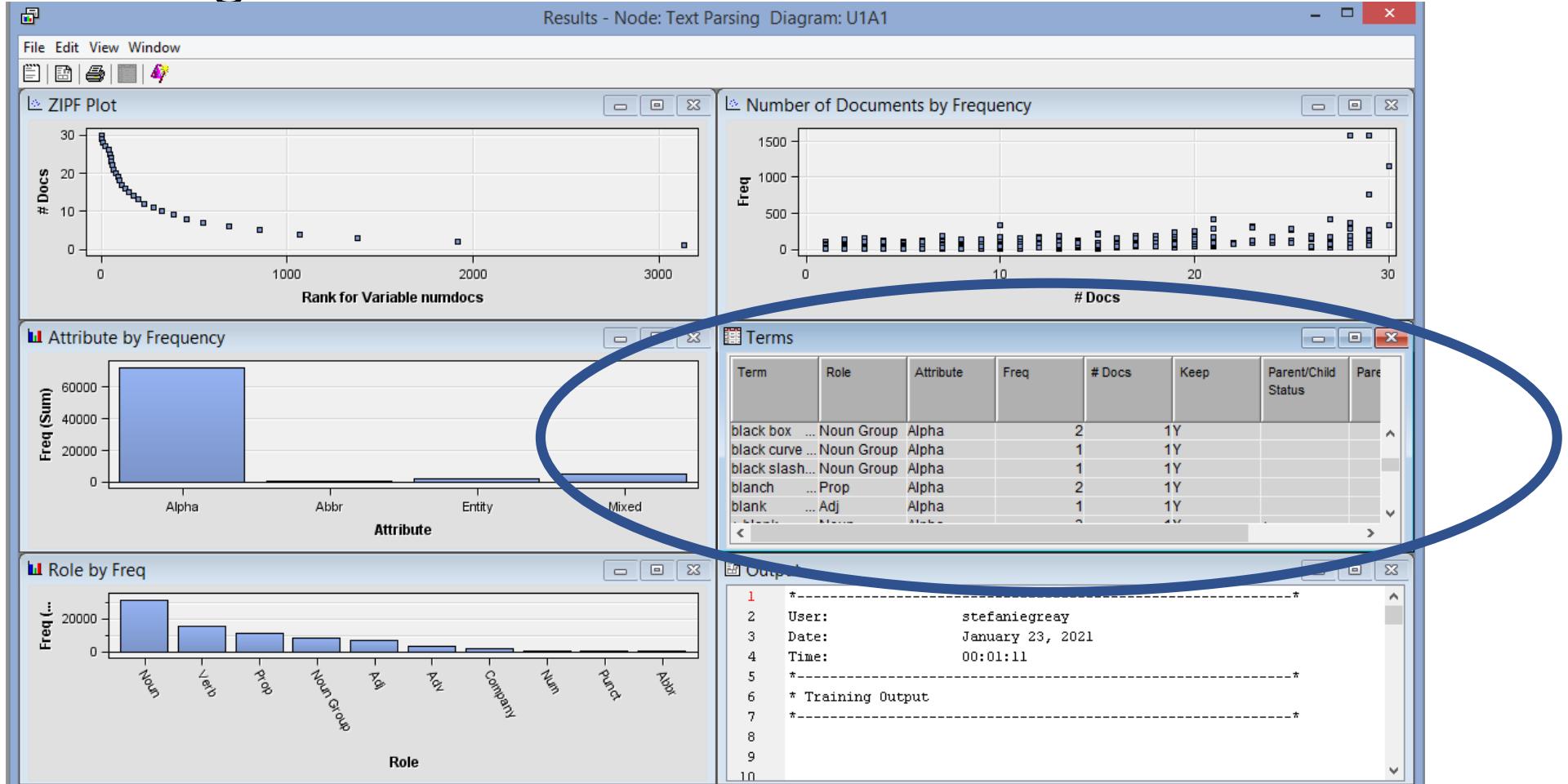
Use the default synonym list (no changes required). This is where you could edit it if you wish.



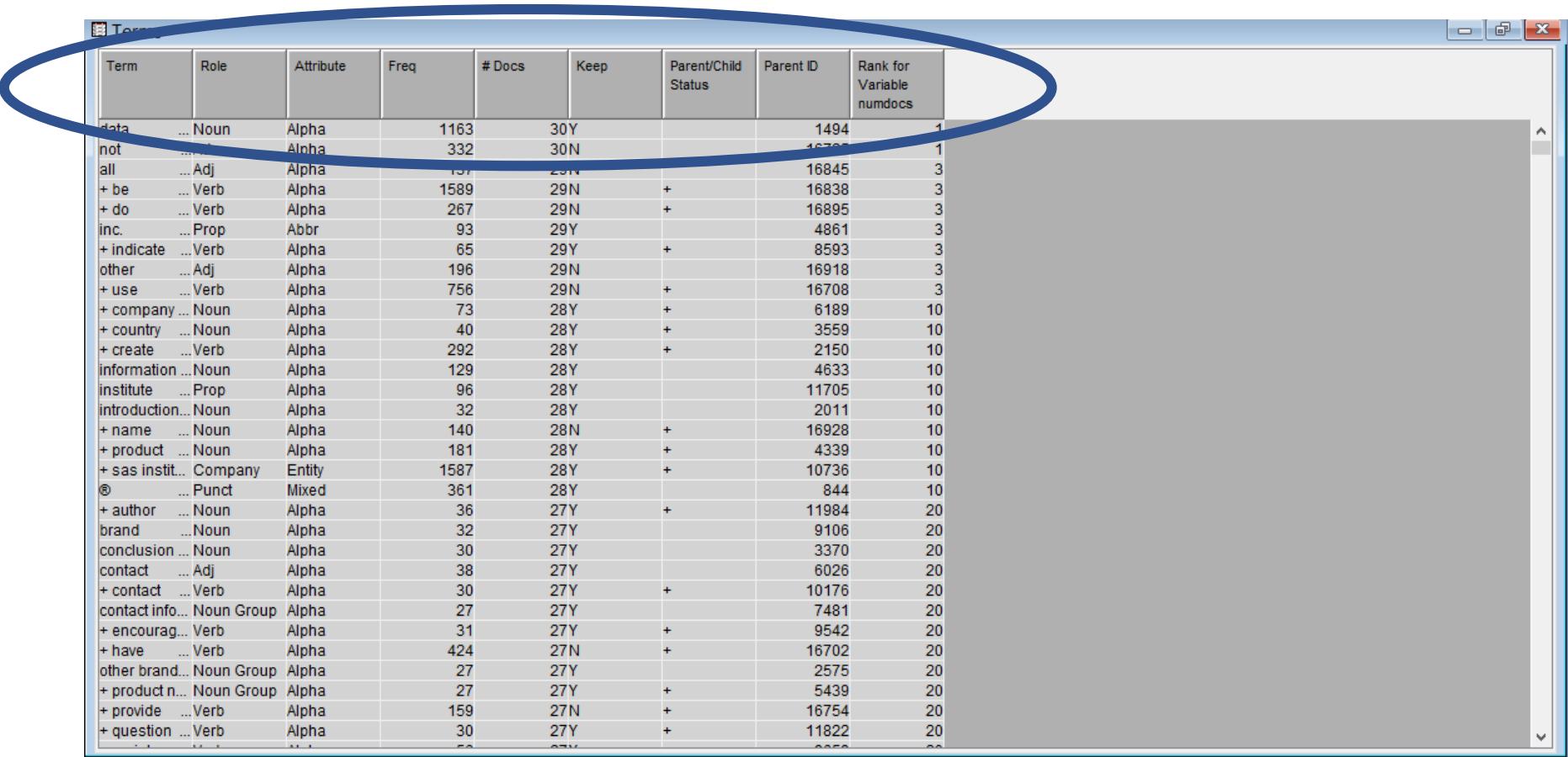
You can review the results by right clicking on the “Text Parsing” Node, and selecting “Results.”



The results pane will look like this. The area you want to review is the “Terms” window, which you can expand by double clicking on it.



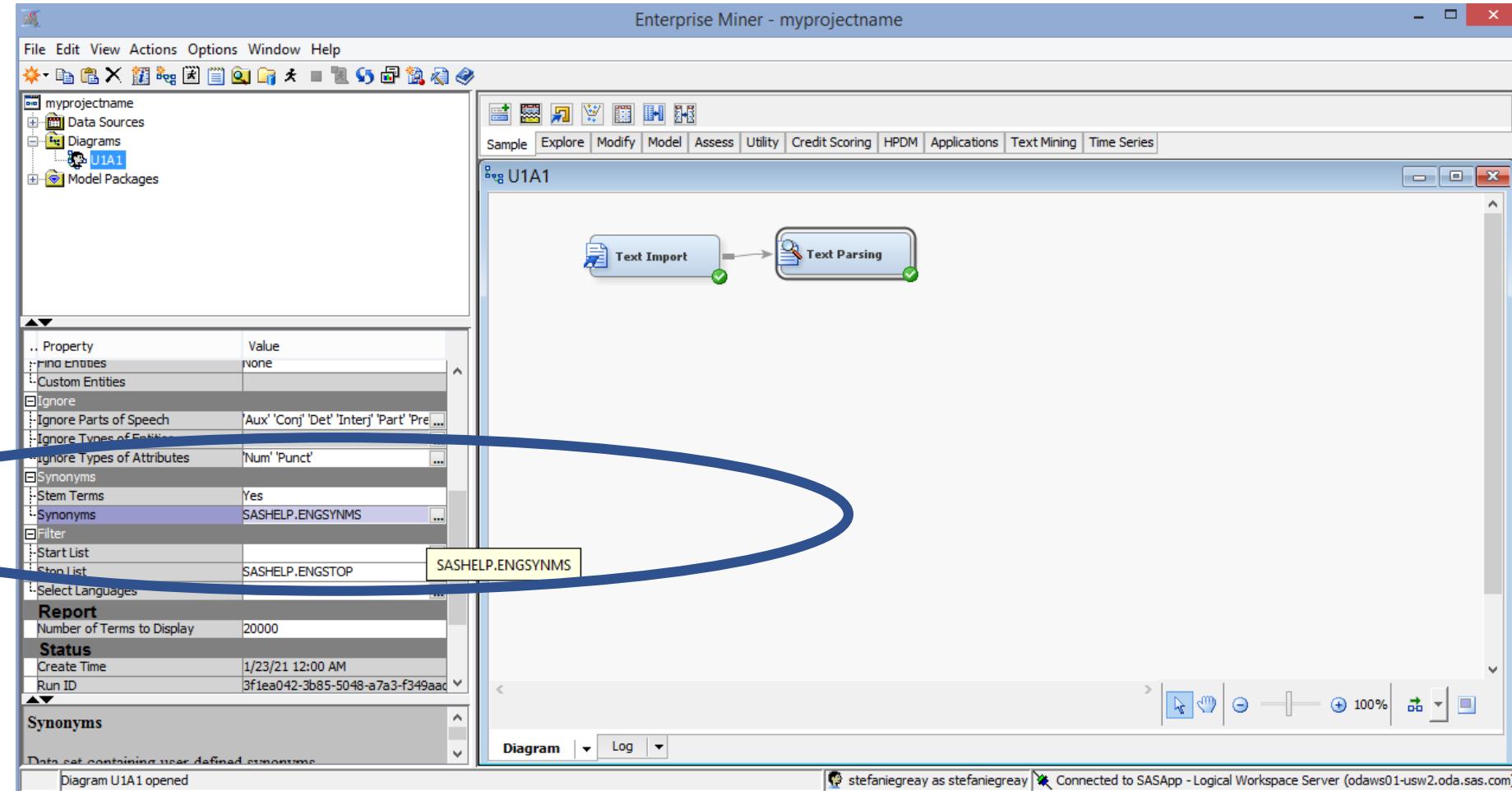
Scan through the terms to see if there are any custom synonyms you want to set in your parsing. You can sort by any of the variables (like frequency and number of docs the term shows up in) to review them more easily. If you determine none are needed, you can close this window and proceed.



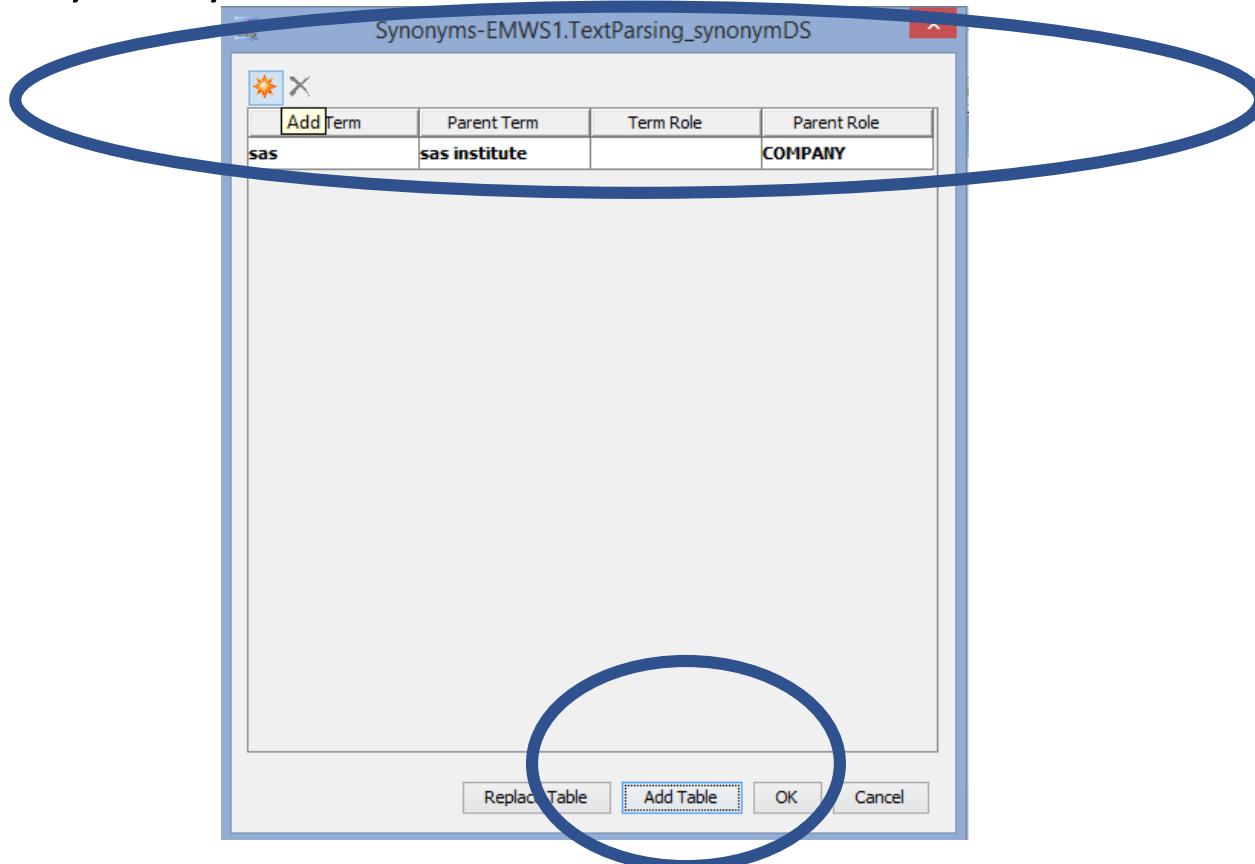
Term	Role	Attribute	Freq	# Docs	Keep	Parent/Child Status	Parent ID	Rank for Variable numdocs
data	... Noun	Alpha	1163	30Y			1494	1
not	... Noun	Alpha	332	30N			16702	1
all	... Adj	Alpha	157	29N			16845	3
+ be	... Verb	Alpha	1589	29N	+		16838	3
+ do	... Verb	Alpha	267	29N	+		16895	3
inc.	... Prop	Abbr	93	29Y			4861	3
+ indicate	... Verb	Alpha	65	29Y	+		8593	3
other	... Adj	Alpha	196	29N			16918	3
+ use	... Verb	Alpha	756	29N	+		16708	3
+ company	... Noun	Alpha	73	28Y	+		6189	10
+ country	... Noun	Alpha	40	28Y	+		3559	10
+ create	... Verb	Alpha	292	28Y	+		2150	10
information	... Noun	Alpha	129	28Y			4633	10
institute	... Prop	Alpha	96	28Y			11705	10
introduction	... Noun	Alpha	32	28Y			2011	10
+ name	... Noun	Alpha	140	28N	+		16928	10
+ product	... Noun	Alpha	181	28Y	+		4339	10
+ sas instit...	Company	Entity	1587	28Y	+		10736	10
@	... Punct	Mixed	361	28Y			844	10
+ author	... Noun	Alpha	36	27Y	+		11984	20
brand	... Noun	Alpha	32	27Y			9106	20
conclusion	... Noun	Alpha	30	27Y			3370	20
contact	... Adj	Alpha	38	27Y			6026	20
+ contact	... Verb	Alpha	30	27Y	+		10176	20
contact info...	Noun Group	Alpha	27	27Y			7481	20
+ encourag...	Verb	Alpha	31	27Y	+		9542	20
+ have	... Verb	Alpha	424	27N	+		16702	20
other brand...	Noun Group	Alpha	27	27Y			2575	20
+ product n...	Noun Group	Alpha	27	27Y	+		5439	20
+ provide	... Verb	Alpha	159	27N	+		16754	20
+ question	... Verb	Alpha	30	27Y	+		11822	20



If you choose to add synonyms, you can do so back in the left menu, by clicking on the elipses next to the synonym dataset.

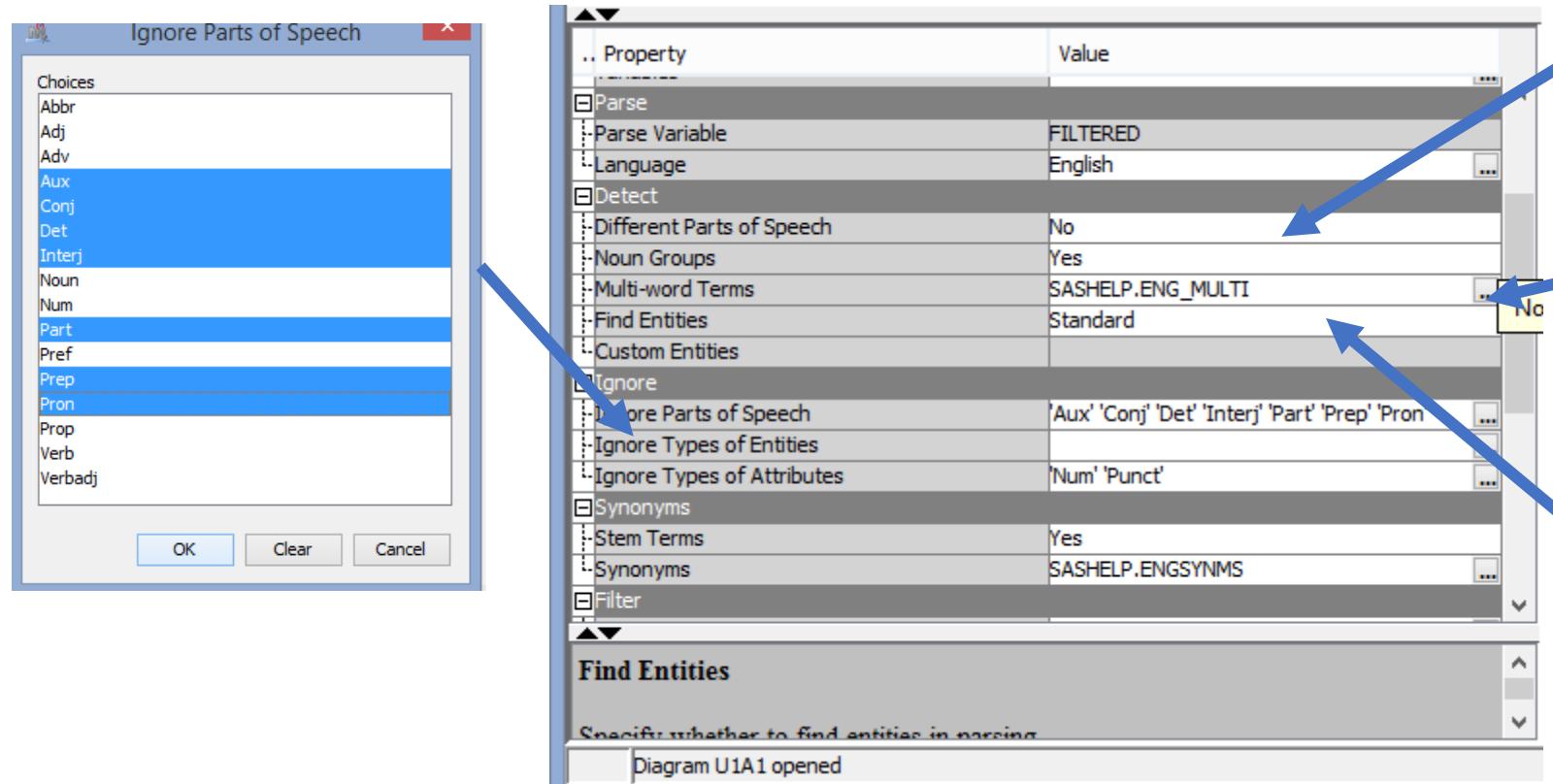


You can then add any synonyms by clicking on the star on the top left, and choosing to add new. Alternatively, if you had an existing table of synonyms, you could add them using the “add table” button below. For this assignment, neither of these are necessary, if you don’t find any specific synonyms you wish to add.



Back in the main menu, we have to change a few additional settings to meet the list of requirements in the assignment.

Default:



Change “Different parts of speech to “No”

Any multi-word terms can be added to the list by clicking here. (None required for this assignment.)

Choose “Standard” in the Find Entities option.



Download the dictionary from the resources of the assignment in the courseroom. NOTE: clicking on it opens it in a new window, then you have to right click and select “save as” to download it to your computer.

---

- Complete the steps required to create a dictionary of words from a document collection using NLP.
- Identify sentences and parts of speech from the text in a document collection to parse data sets.

Refer to the Text Parsing Scoring Guide for more details.

**Walkthroughs:** You may view the following walkthrough videos, linked in the Resources, to help you complete this assignment:

- Creating an English Dictionary Data Set Walkthrough.
- Text Parsing Node – Parts of Speech Walkthrough.

#### Additional Requirements

Please adhere to the following:

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA current edition style and formatting.
- **Number of resources:** Include a list of any articles or readings you reference or use to complete your assignment.
- **Length of paper:** 1–2 double-spaced pages, including the word dictionary.
- **Font:** Arial, 10 point.

#### Resources

[Text Parsing Scoring Guide.](#)

[APA Style and Format.](#)

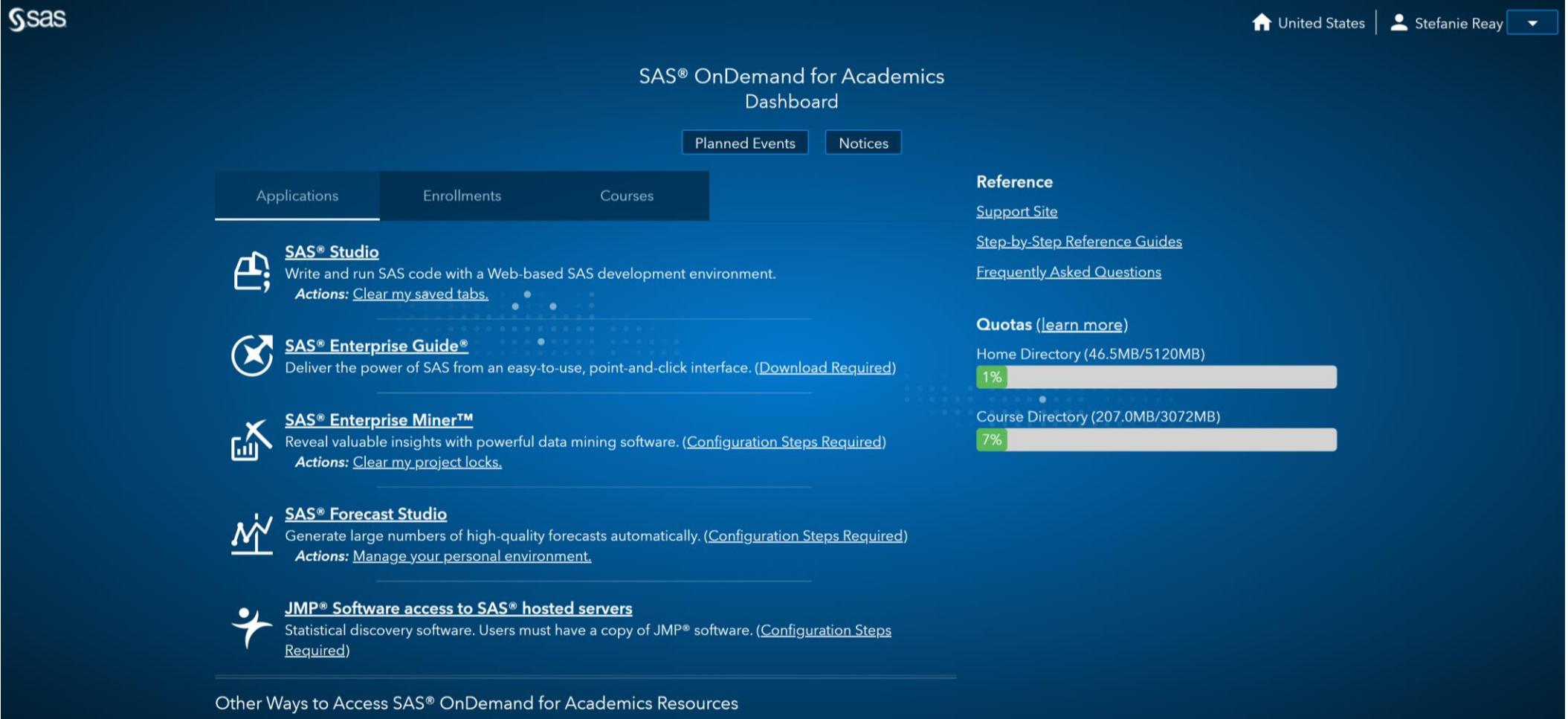
[Data Dictionary \[Data\].](#)

[Creating an English](#) Select this link to launch this material in a new window.

[Text Parsing Node – Parts of Speech Walkthrough.](#)



# Go to “SAS Studio” to upload the dictionary file.



The screenshot shows the SAS OnDemand for Academics Dashboard. At the top, there are links for "Planned Events" and "Notices". Below that, a navigation bar has tabs for "Applications", "Enrollments", and "Courses", with "Applications" being the active tab. The main content area displays five software applications:

- 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 "Support Site", "Step-by-Step Reference Guides", and "Frequently Asked Questions". Below that is a "Quotas" section showing usage statistics for Home and Course Directories.

Other Ways to Access SAS® OnDemand for Academics Resources

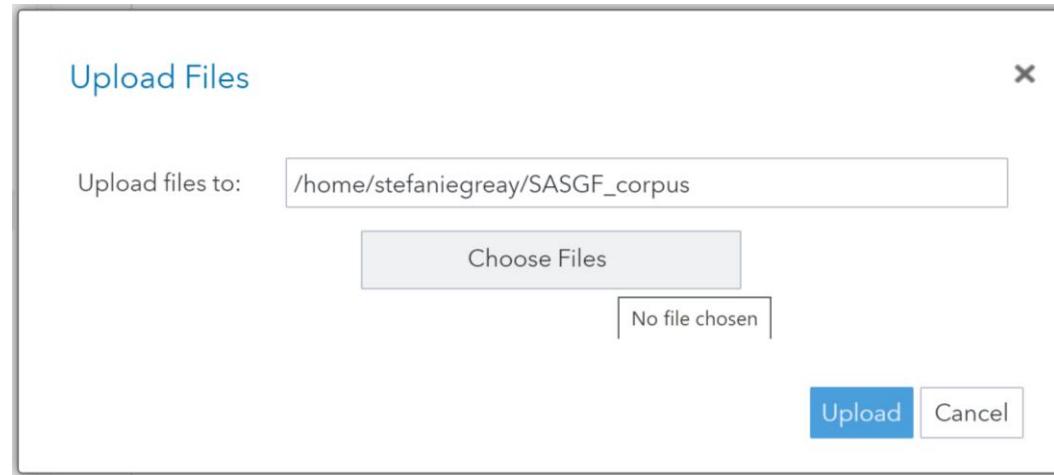


Click on “Files (Home)” to make the upload button appear in dark blue.

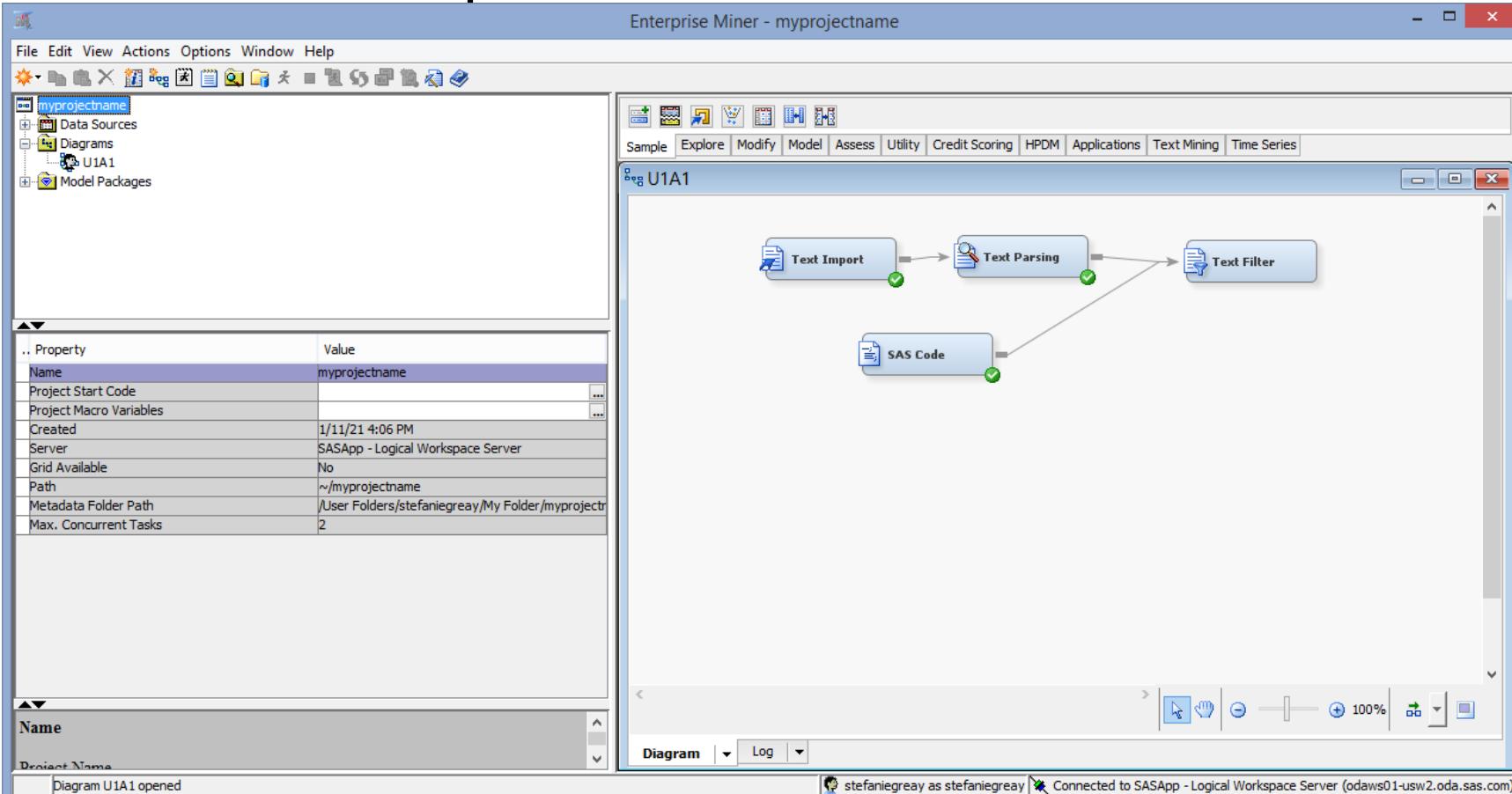
The screenshot shows the SAS Studio interface. On the left, there is a sidebar titled "Server Files and Folders" which includes options like "Tasks and Utilities", "Snippets", "Libraries", and "File Shortcuts". Below this, under "Server Files and Folders", there is a tree view showing a folder structure: "odaws02-usw2" has "Folder Shortcuts" and "Files (Home)". The "Files (Home)" item is highlighted with a blue selection bar. The main workspace is titled "Program 1" and contains tabs for "CODE", "LOG", and "RESULTS". The "CODE" tab is active, showing a single line of code "1". Below the tabs is a toolbar with various icons. The bottom right corner of the workspace shows "UTF-8", a green circular icon with a white letter 'G', and the text "User: stefaniegreay". The top right of the screen shows user information: "SAS Programmer", a help icon, and "Sign Out".



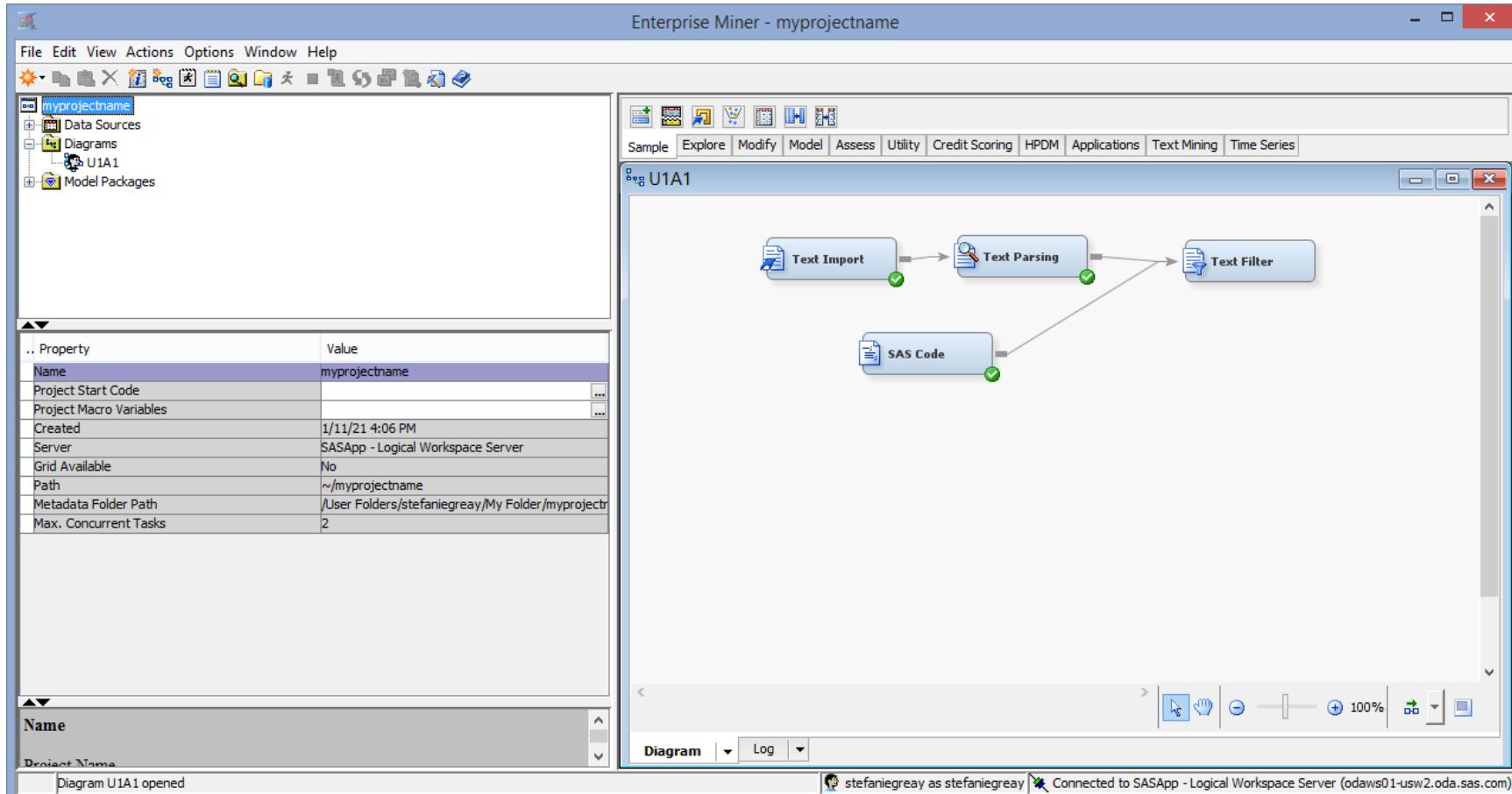
Click “Choose Files” after specifying whatever location you wish to upload the file to, then click “Upload” once you have chosen the file.



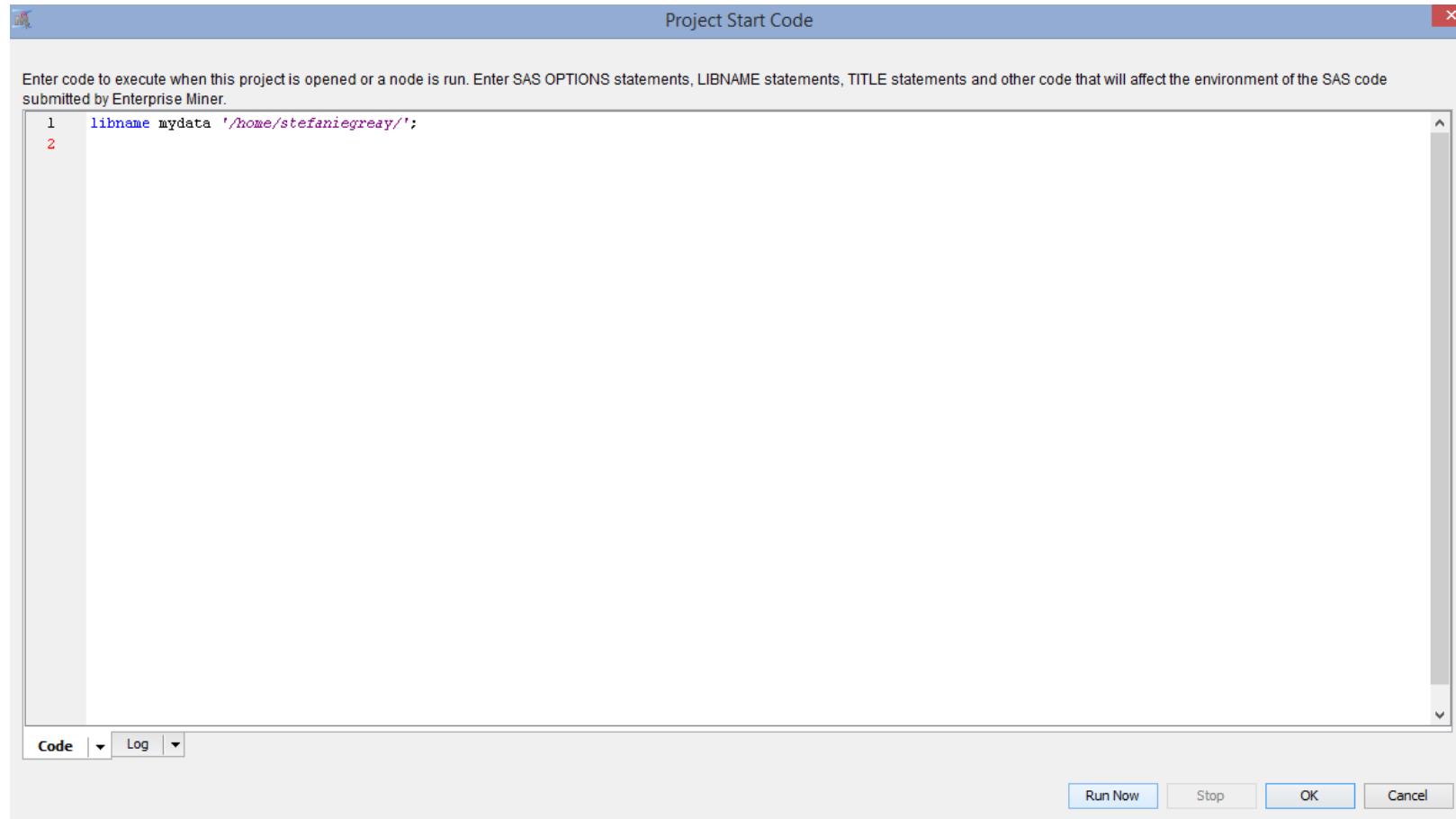
To be sure that this library will be searchable in SAS Enterprise Miner, double click on the project name in SAS Enterprise Miner.



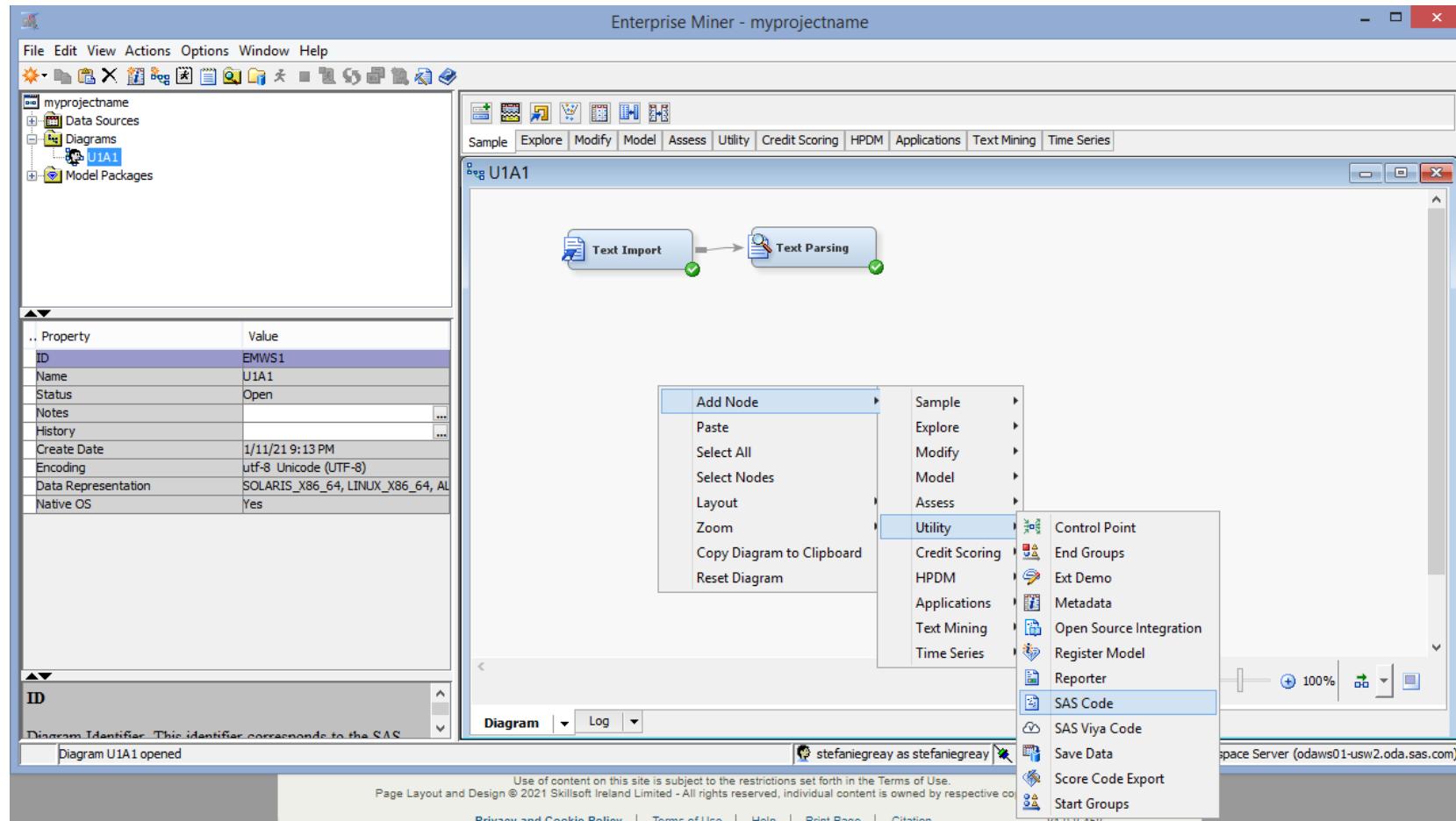
# Click on the elipses next to the “Project Start Code”



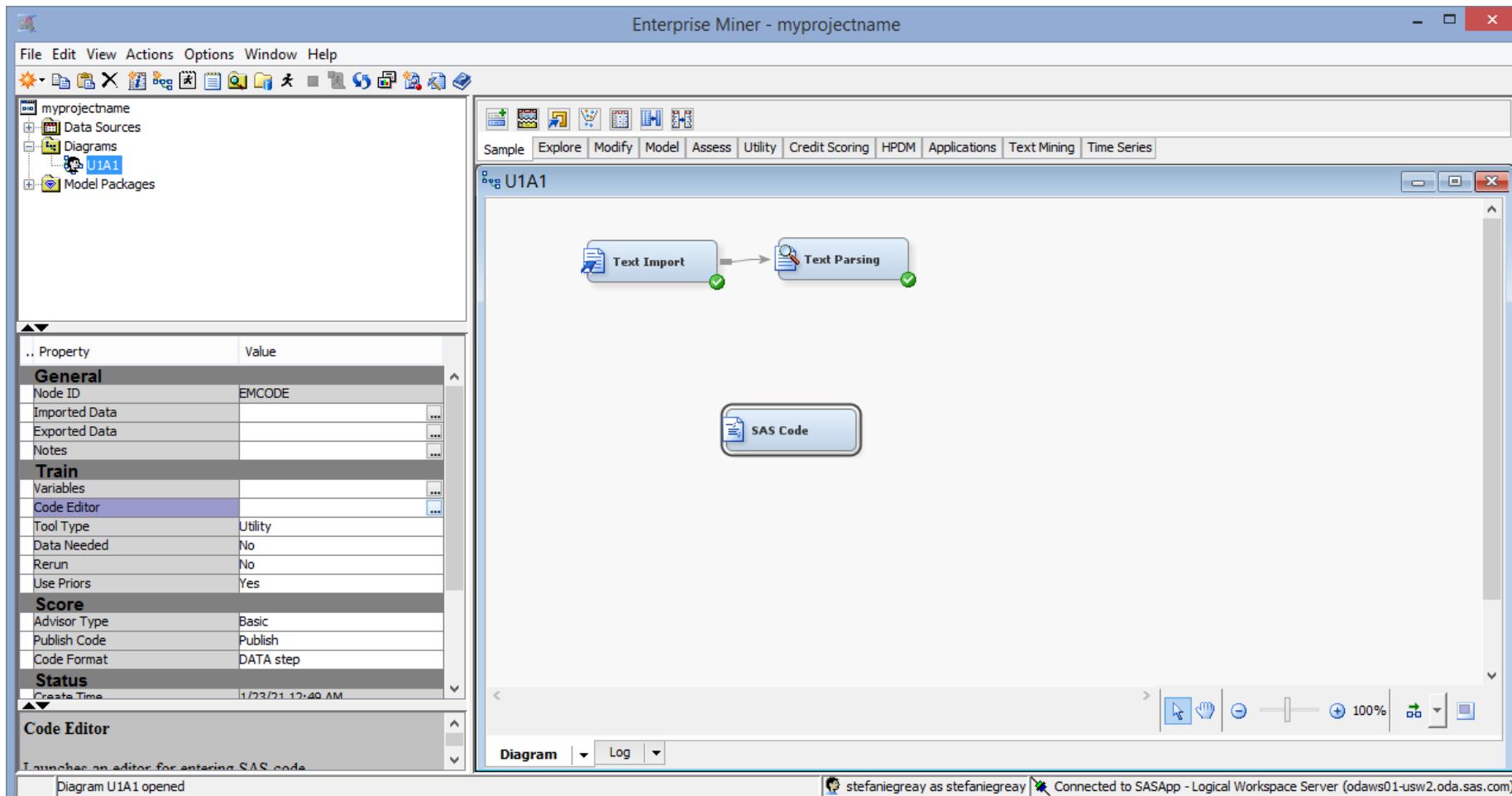
Add the libname statement for the library you will store your imported SAS dataset of the dictionary, click “Run Now” and then click “OK”



Add a code node to your diagram by right clicking, then choosing Add Node>Utility>SAS Code



Open the code for editing by clicking on the ellipses next to the “Code Editor.”



Add the code to import the dictionary file you uploaded through SAS Studio.

Training Code - Code Node

File Edit Run View

Macro

Train

Utility

EM\_REGISTER

EM\_REPORT

EM\_DATACODE

Macros Macro Variables

Training Code

```
libname mydata '/home/stefaniegreay/';  
data mydata.engdict (keep=term pos);  
    length inputterm term $32;  
    infile '/home/stefaniegreay/cf_en_US.dic'  
        truncover;  
    input linetxt $80.;  
    i=1;  
    do until (inputterm = ' ');  
        inputterm = scan(linetxt, i, ' ');  
        if inputterm ne ' ' then do;  
            location=index(inputterm,'/');  
            if location gt 0 then  
                term = substr(inputterm,1,location-1);
```

Output Log Result Log

1  
2

stefaniegreay as stefaniegreay - myprojectname - U1A1 - EMCODE - Complete

# Sample code for dictionary import

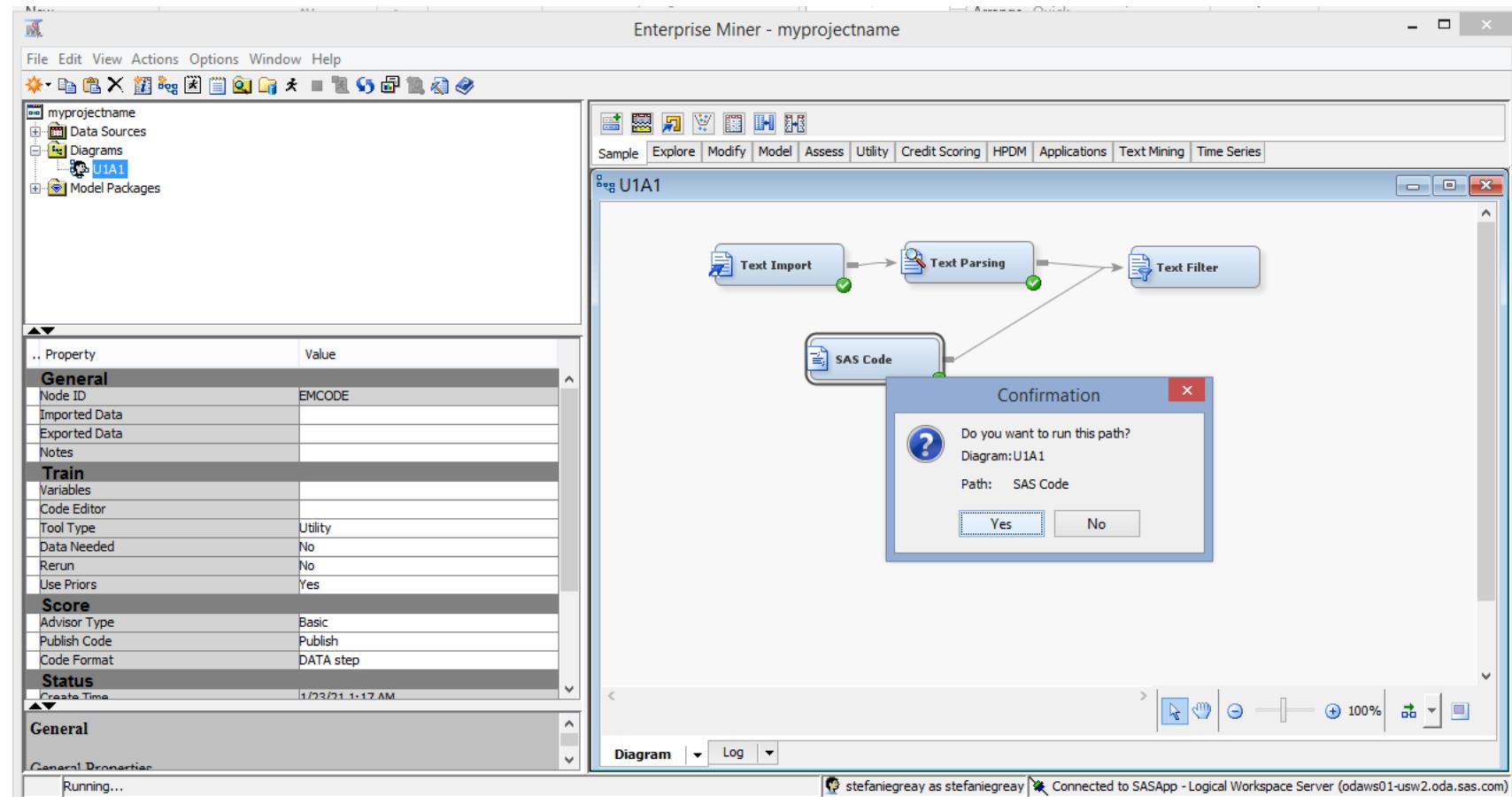
```
libname mydata '/home/stefaniegreay/';

data mydata.engdict (keep=term pos);
length inputterm term $32;
infile '/home/stefaniegreay/cf_en_US.dic'
truncover;
input linetxt $80.;

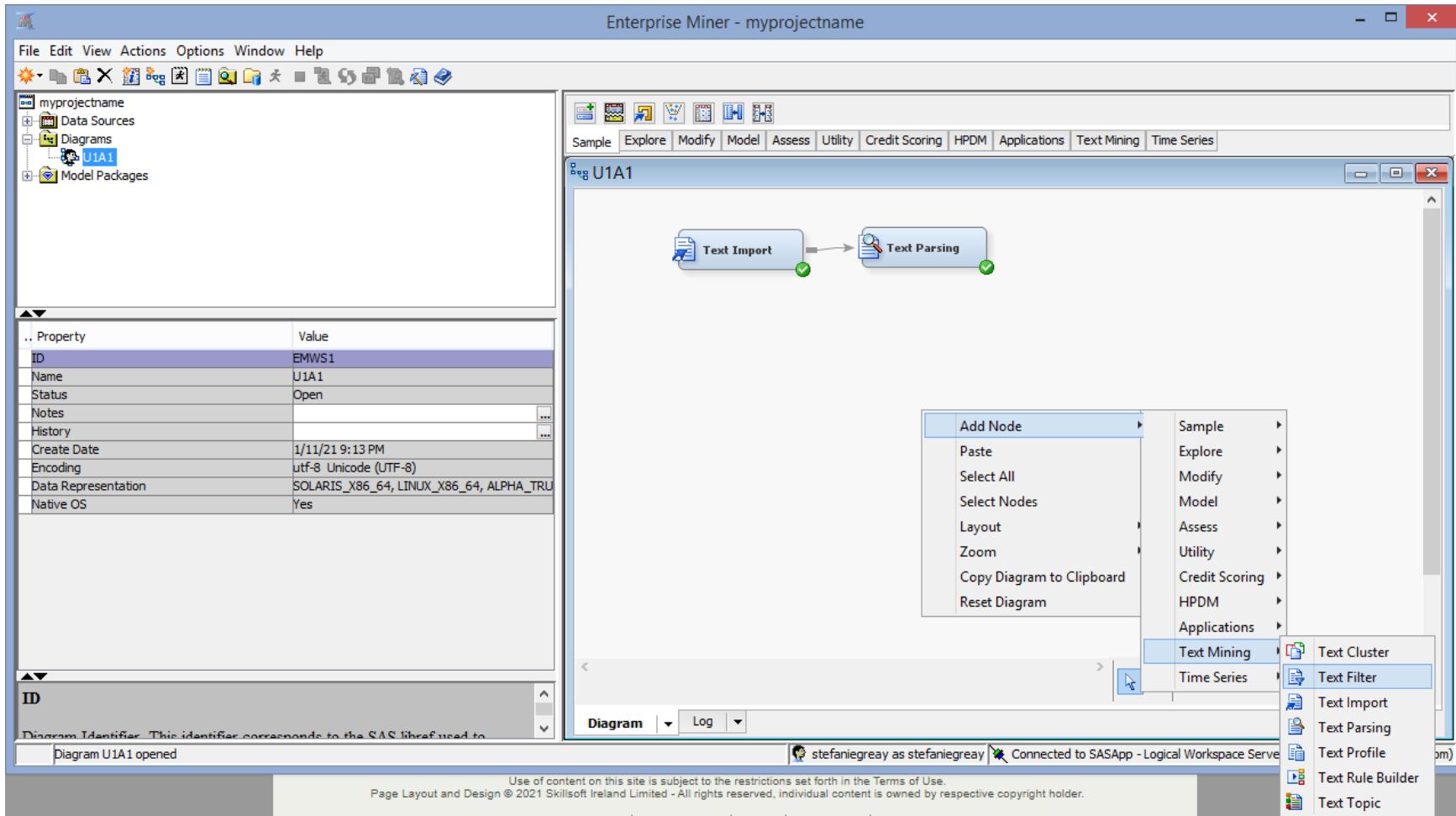
i=1;
do until (inputterm = ' ');
inputterm = scan(linetxt, i, ' ');
if inputterm ne '' then do;
location=index(inputterm,'/');
if location gt 0 then
term = substr(inputterm,1,location-1);
if location eq 0 then
term = inputterm;
if lowercase(term) ne term then pos = 'Prop';
term = lowercase(term);
output;
end;
i=i+1;
end;
run;
```



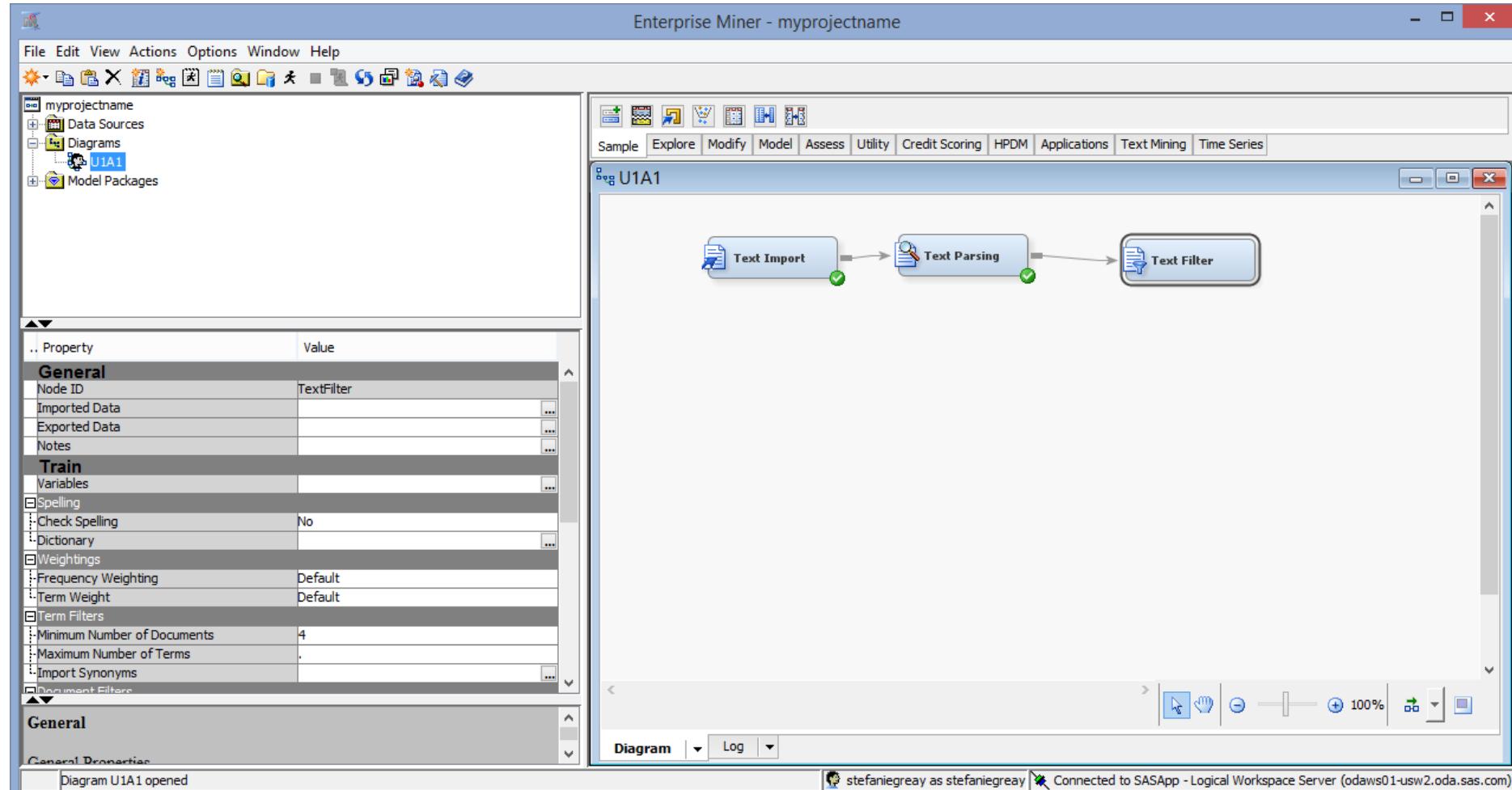
Run this code using the run guy or right clicking on the “SAS Code” node and clicking “run,” and confirm that you wish to run it by clicking “Yes.” (Note that I already have the Text Filter node in this screenshot. It is ok if you do not yet.)



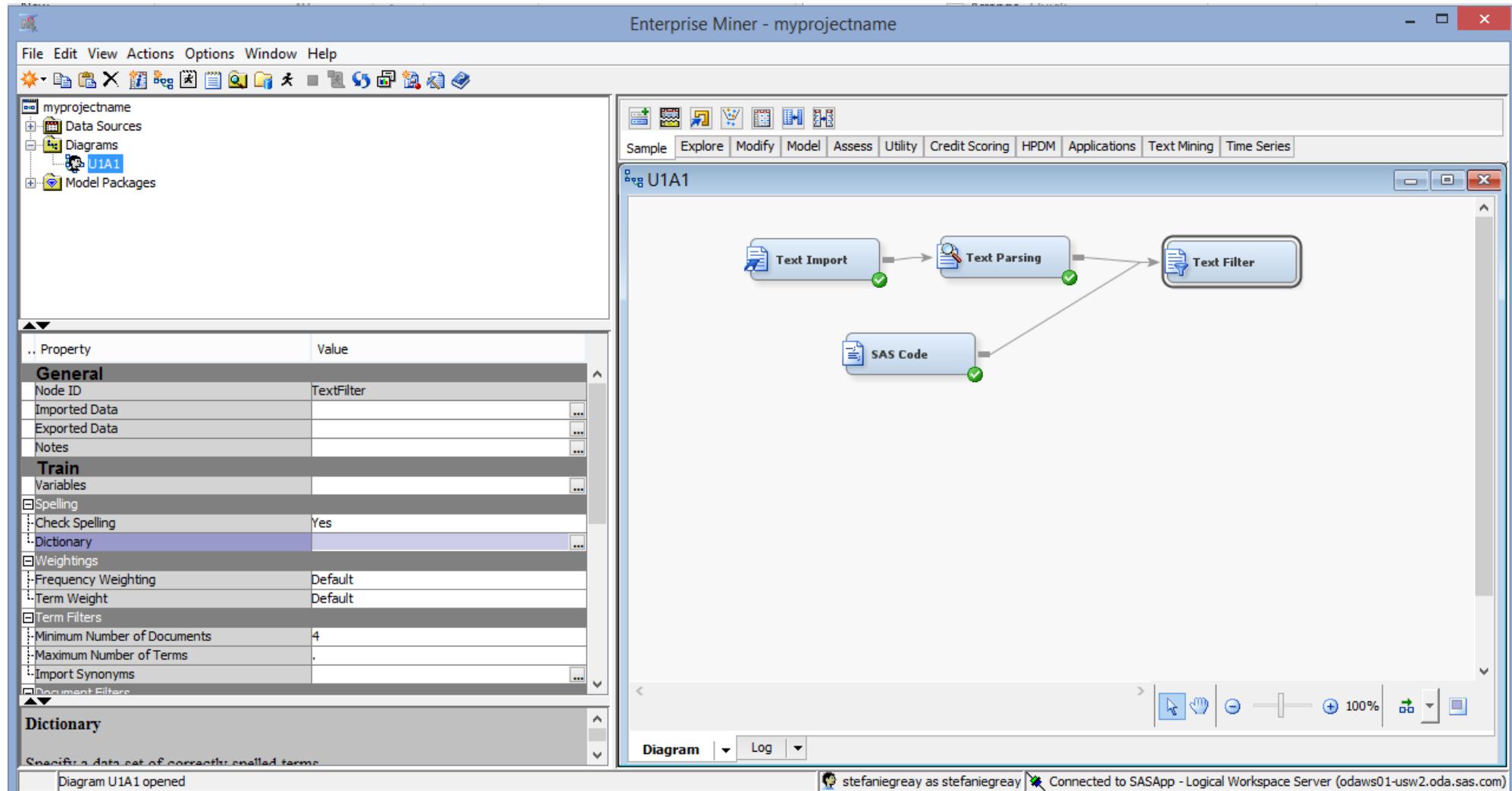
Add the text filter node to the diagram (in order to specify your dictionary that you just created) using right click, then Add Node>Text Mining>Text Filter.



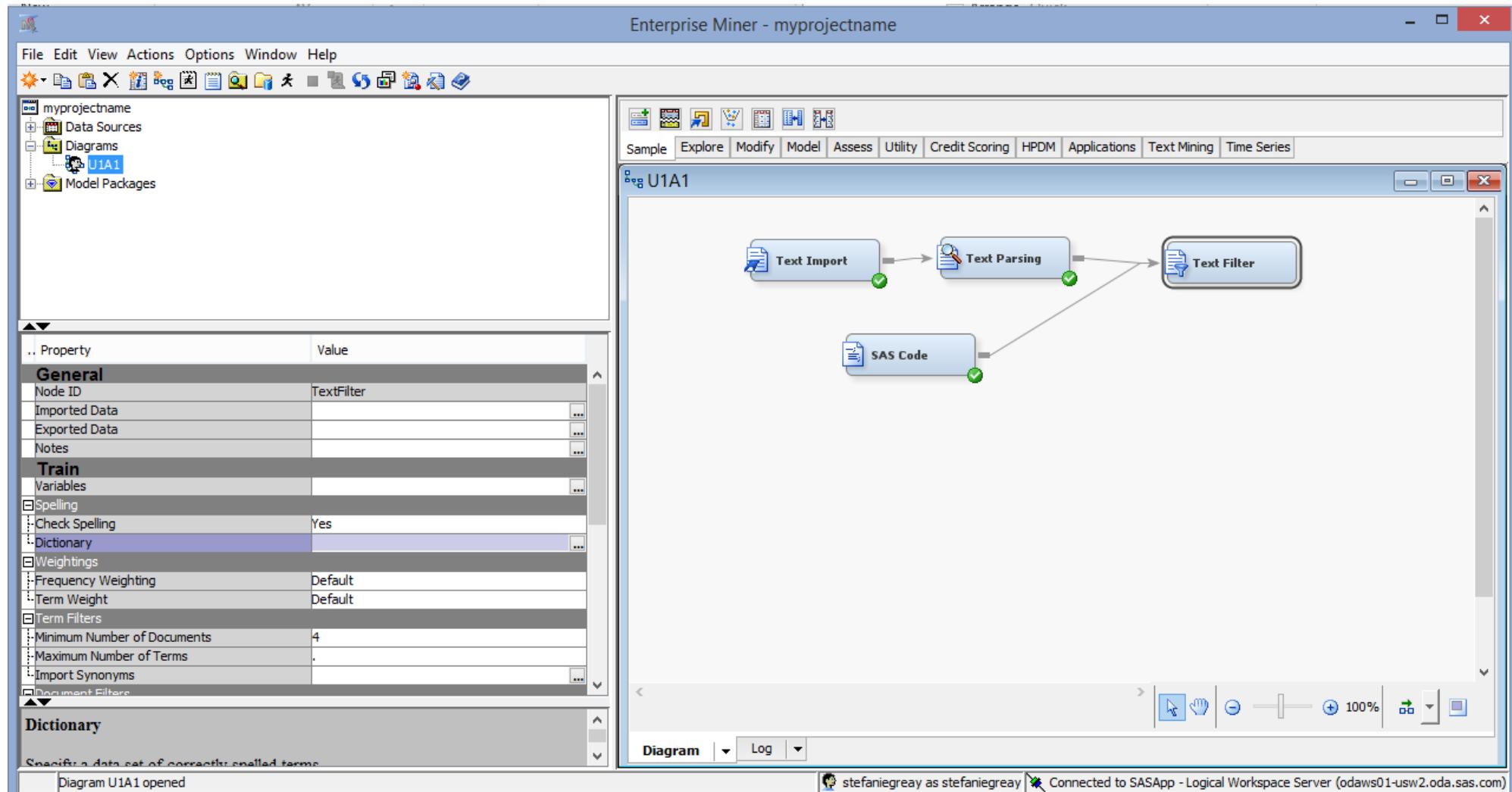
Drag the Text Filter node to the right of your other path and nodes, and connect it to the Text Parsing node using the box and arrow.



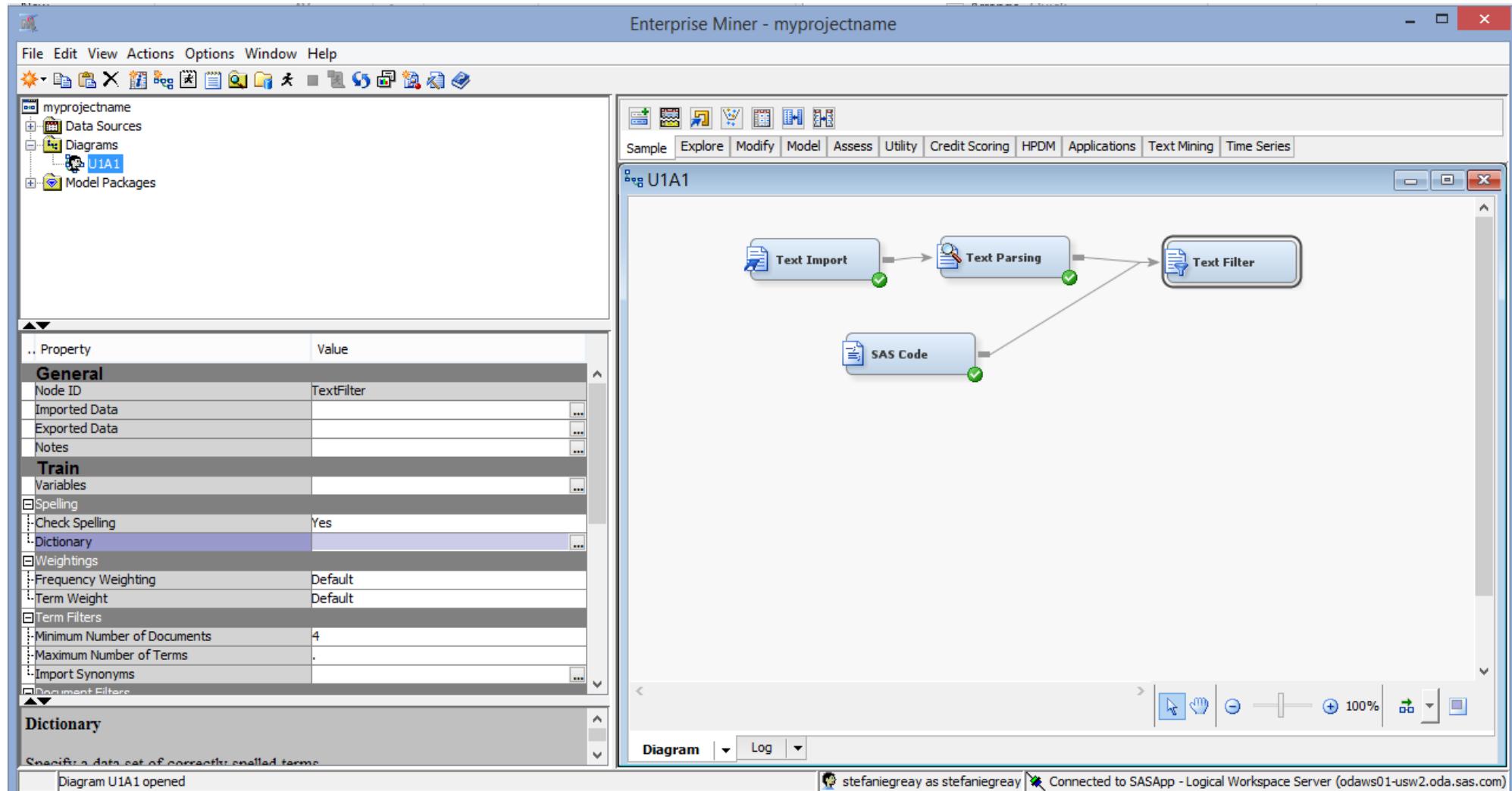
Be sure to connect your SAS code node to your Text Filter node the same way.



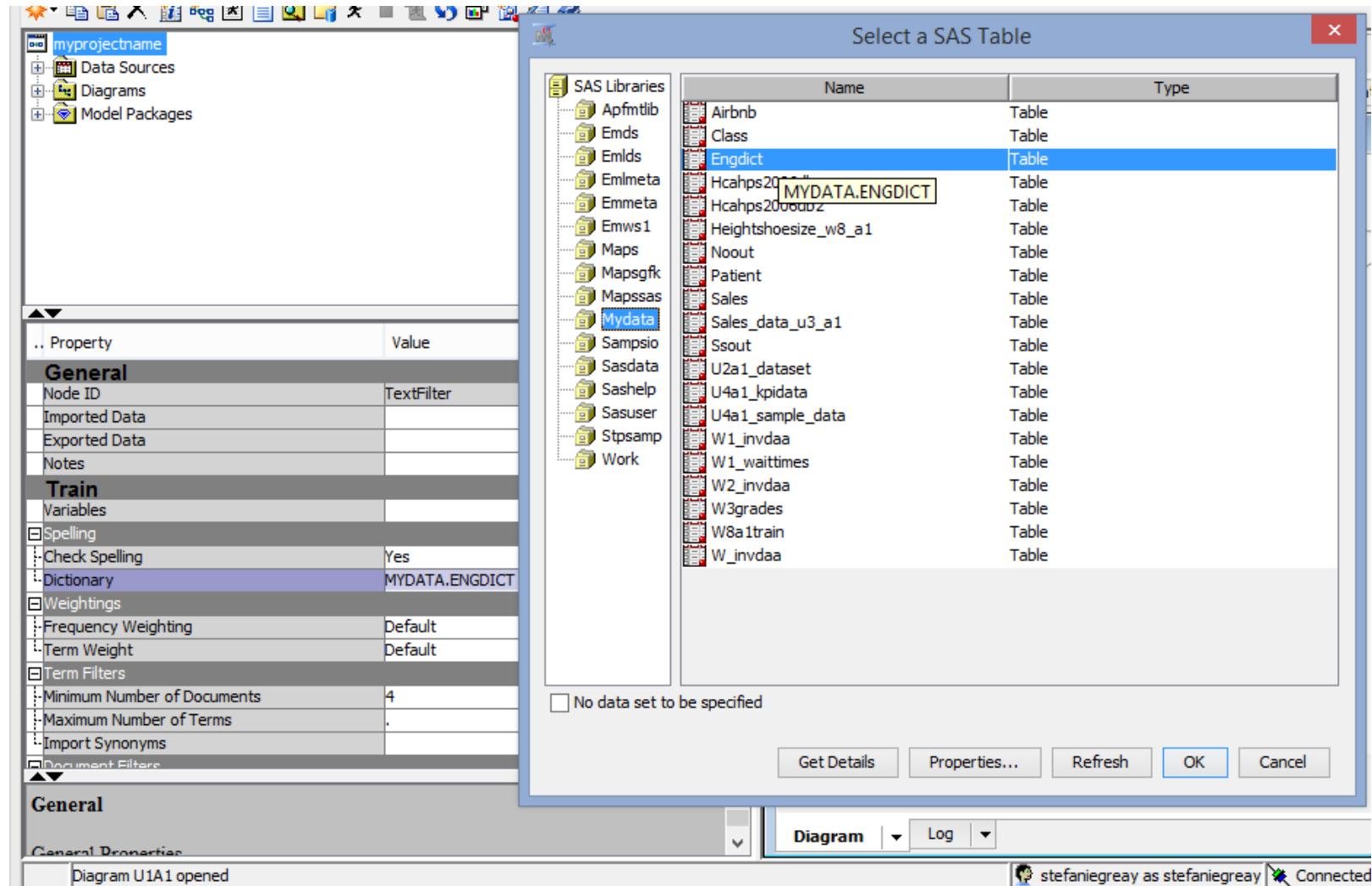
Click on the Text Filter node, and refer to the left menu.



Choose “Yes” for the Check Spelling option and click on the ellipses to select the Dictionary you just created.



Select the libname and the dataset that you just created for the dictionary, and click “OK.”



# Complete and submit your Unit 2 Assignment 1 assignment

Selecting to check spelling and adding the English dictionary you created are all you have to do to complete this assignment. No additional work in the Text Filter Node is required until next unit.



Follow the steps outlined in the text below to evaluate the success of the import of the pdf files.

- Chakraborty, G., Pagolu, M., & Garla, S. (2014). Text mining and analysis: practical methods, examples, and case studies using SAS. SAS Institute.

<https://capella.skillport.com/skillportfe/main.action?path=summary/BOOKS/59026>

- Direct link to Chapter 4 – Parsing and Extracting Features:

<https://capella.skillport.com/skillportfe/assetNonSSOLaunch.action?courseName= ss chapter:59026-786520828&courseType=7>

