

ANLT5070

Unit 5 Assignment 1 Tutorial

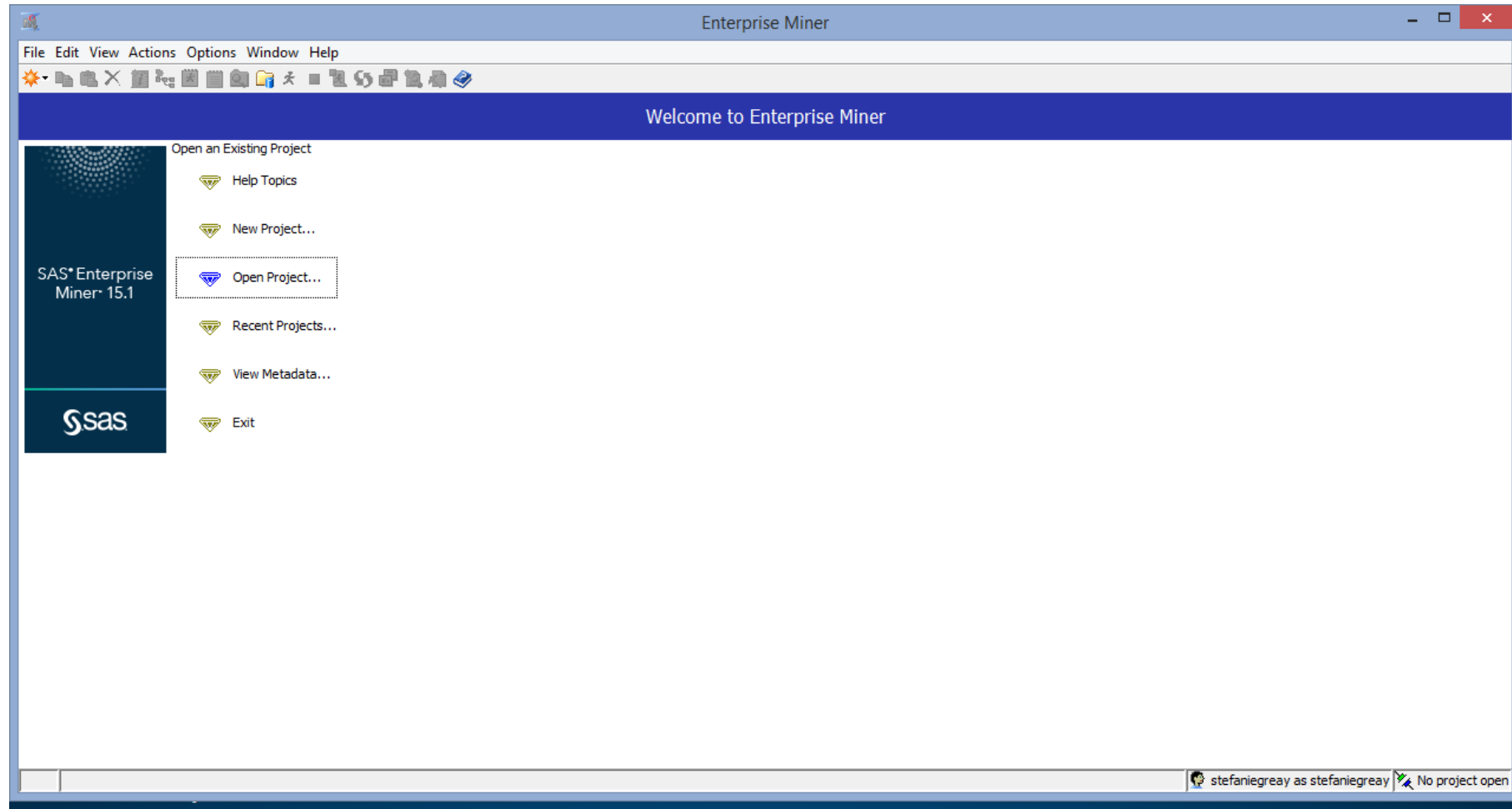


Open project created in Unit 1 and added to in Unit 2, 3, and 4.

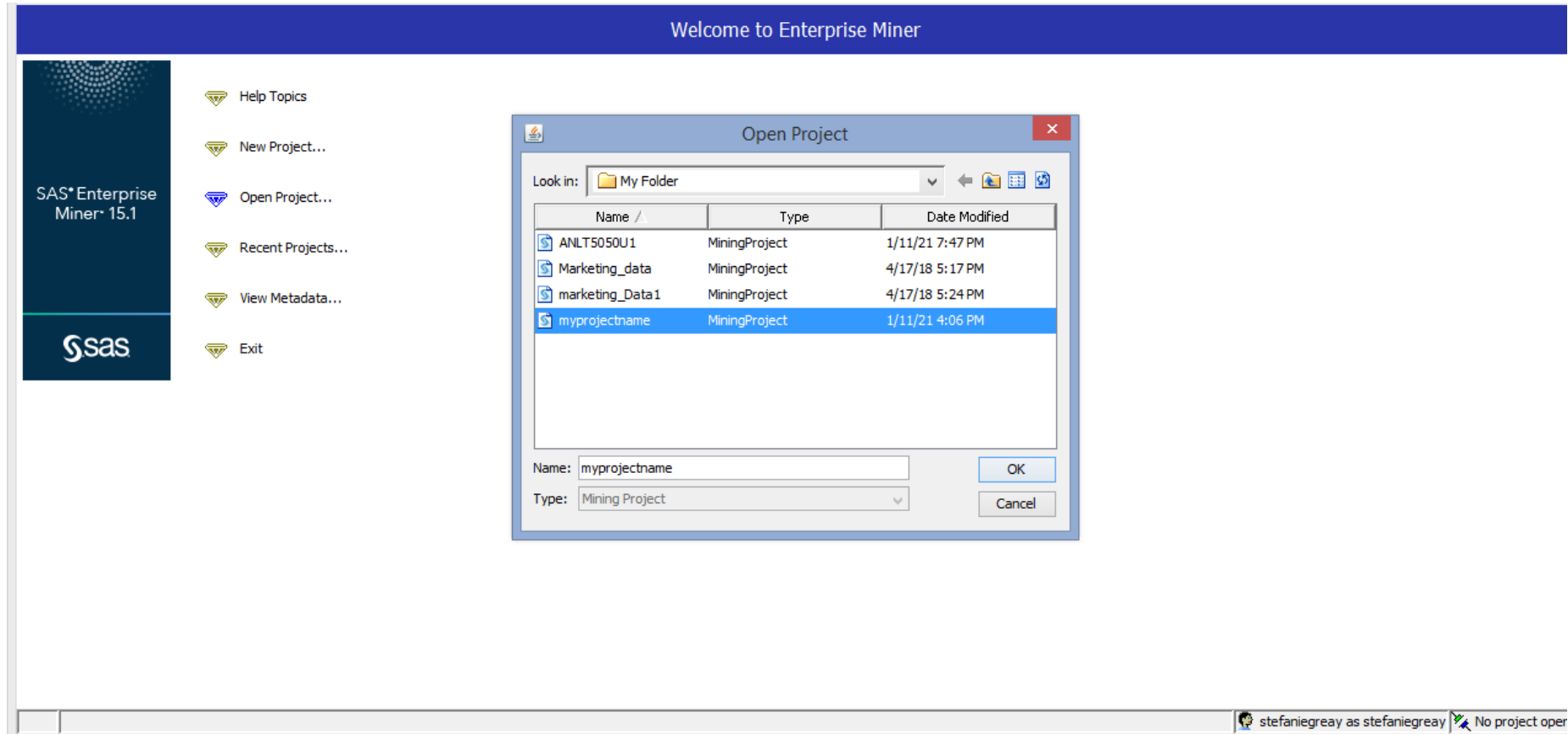
- This assignment is a continuation of the Unit 1 Assignment 1, Unit 2 Assignment 1, Unit 3 Assignment 1, and Unit 4 Assignment 1 assignments. If you have not successfully completed these, or you have lost your work for them, please re-watch and re-complete those steps first.
- This assignment focuses on the exploring the text topics portion of the text analytics project.



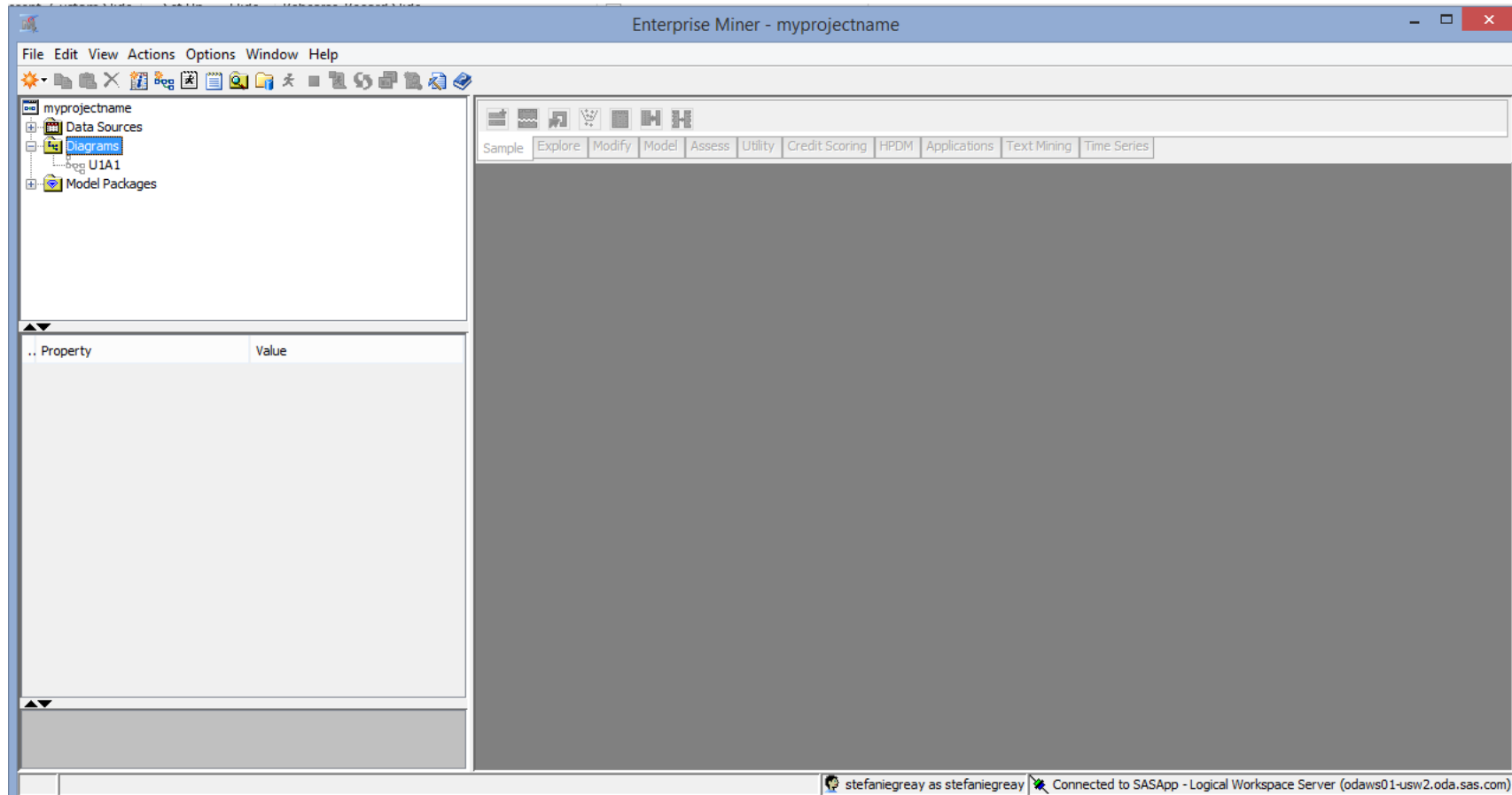
Click on open project



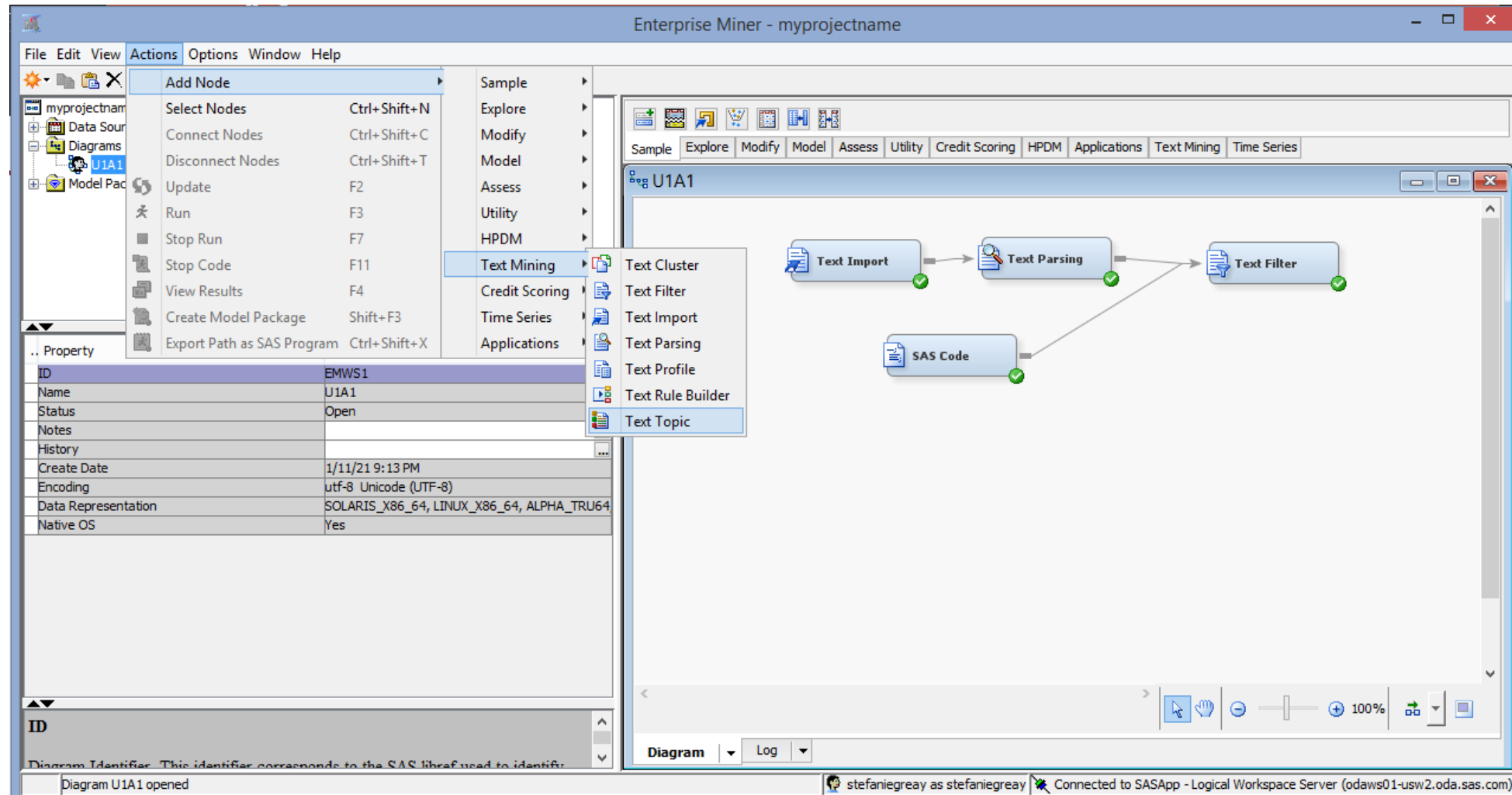
Select the project you created in U1A1, and click “OK” to open the project.



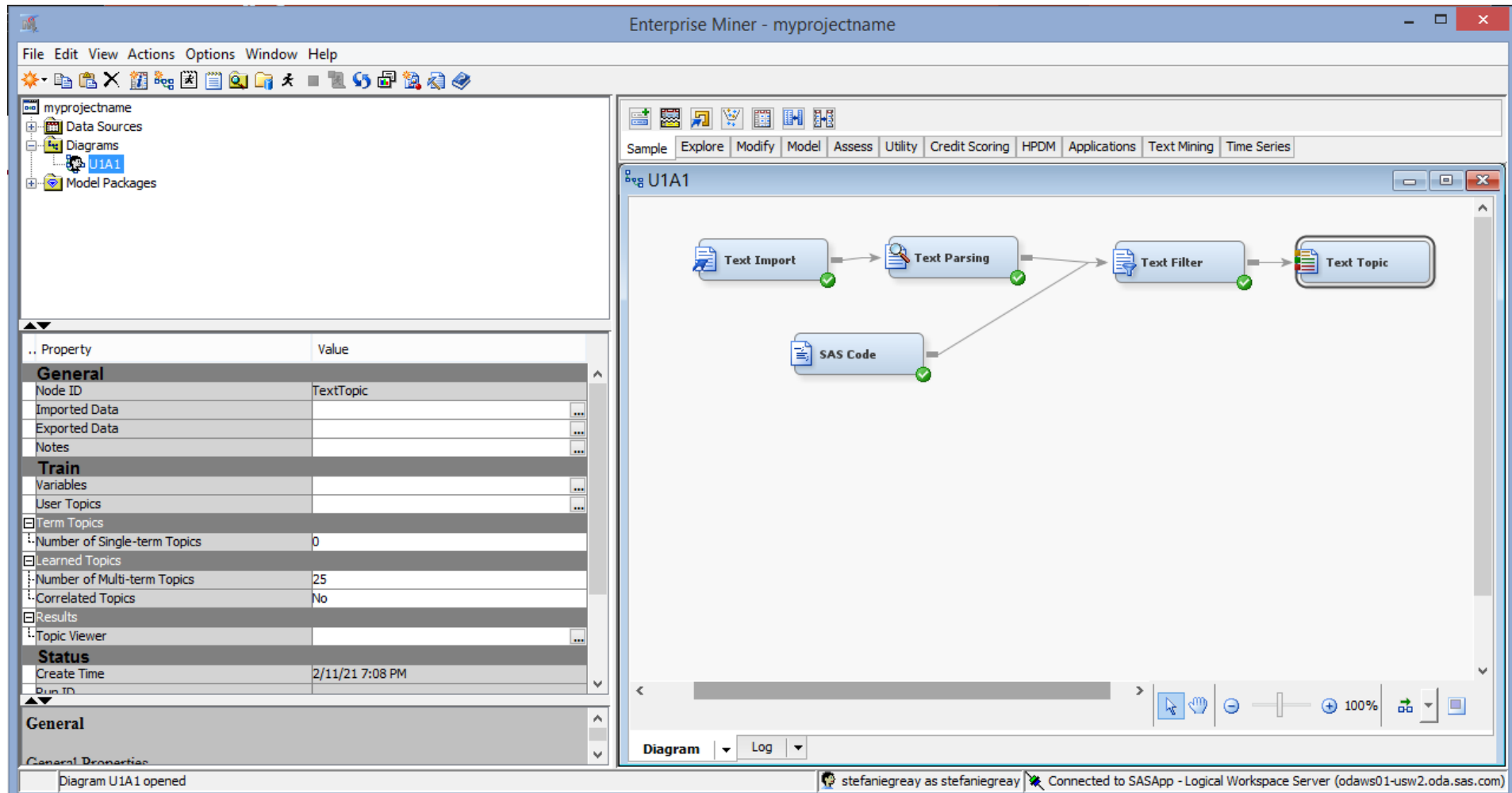
Click on “Diagrams,” navigate to the diagram you created, and double click on it to open it.



Click on “Actions” > “Add Node” > “Text Mining” > “Text Topic” to add the Text Topic node to your diagram



Connect the Text Topic node to the Text Filter node using the connection arrow.



Run the Text Topics node (and the path that precedes it) by right clicking on the Text Topics node and selecting “Run.”

The screenshot displays the SAS Enterprise Miner software interface. The main window, titled "Enterprise Miner - myprojectname", shows a workflow diagram in the center. The diagram consists of four nodes: "Text Import", "Text Parsing", "Text Filter", and "Text Topics". Arrows indicate a sequential flow from "Text Import" to "Text Parsing", then to "Text Filter", and finally to "Text Topics". A "SAS Code" node is also present, with an arrow pointing to the "Text Filter" node. Each node has a green checkmark next to it, indicating successful execution or completion. On the left side, there is a tree view showing the project structure, including "Data Sources", "Diagrams", "U1A1", and "Model Packages". Below the tree view, there is a table with properties and values for the selected node, "Text Topics". The table has two columns: "Property" and "Value". The properties listed are "Node ID" (TextTopic), "Imported Data", "Exported Data", "Notes", "Variables", "User Topics", "Term Topics", "Number of Single-term Topics" (0), "Learned Topics", "Number of Multi-term Topics" (25), "Correlated Topics" (No), "Results", "Topic Viewer", "Create Time" (2/11/21 7:08 PM), and "Status". A right-click context menu is open over the "Text Topics" node, showing options 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 in blue. At the bottom of the window, there is a status bar showing the user "stefaniegreay as stefaniegreay" and the connection "Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)".

Property	Value
General	
Node ID	TextTopic
Imported Data	
Exported Data	
Notes	
Train	
Variables	
User Topics	
Term Topics	
Number of Single-term Topics	0
Learned Topics	
Number of Multi-term Topics	25
Correlated Topics	No
Results	
Topic Viewer	
Status	
Create Time	2/11/21 7:08 PM
Status	



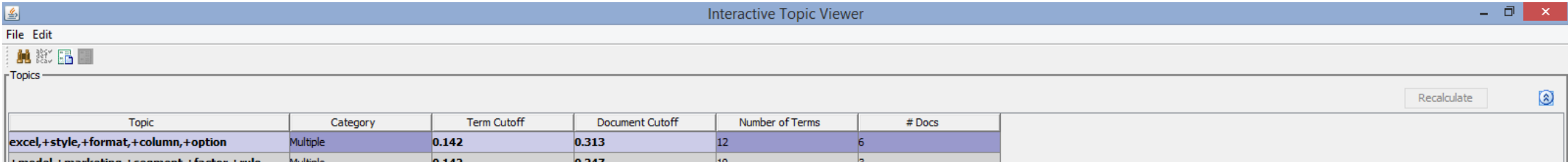
Click on the ellipses (3 dots) next to “Topic Viewer” to open the interactive topic viewer.

The screenshot displays the SAS Enterprise Miner interface. On the left, a tree view shows the project structure with 'myprojectname' expanded, containing 'Data Sources', 'Diagrams', 'U1A1', and 'Model Packages'. Below this, a table lists properties for 'U1A1', with 'Topic Viewer' highlighted and its ellipsis button circled in blue. The main area shows a workflow diagram for 'U1A1' with steps: Text Import, Text Parsing, Text Filter, and Text Topic, all marked with green checkmarks. A 'SAS Code' node is also present. The bottom status bar indicates 'Diagram U1A1 opened' and 'Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)'.

Property	Value
User Topics	
Term Topics	
Number of Single-term Topics	
Learned Topics	
Number of Multi-term Topics	25
Correlated Topics	No
Results	
Topic Viewer	...
Status	
Create Time	2/11/21 7:08 PM
Run ID	9a989091-47d5-1a41-a056-ece538037a4b
Last Status	Complete
Last Run Time	2/11/21 7:12 PM
Run Duration	0 Hr. 0 Min. 4.08 Sec.
Grid Host	
User-Added Node	No



The Topics table will display as part of the Topics Explorer window, along with the number of documents that contain that topic. You can explore the other panes (Terms and Documents) for each topic by clicking on the topic in the topic table.

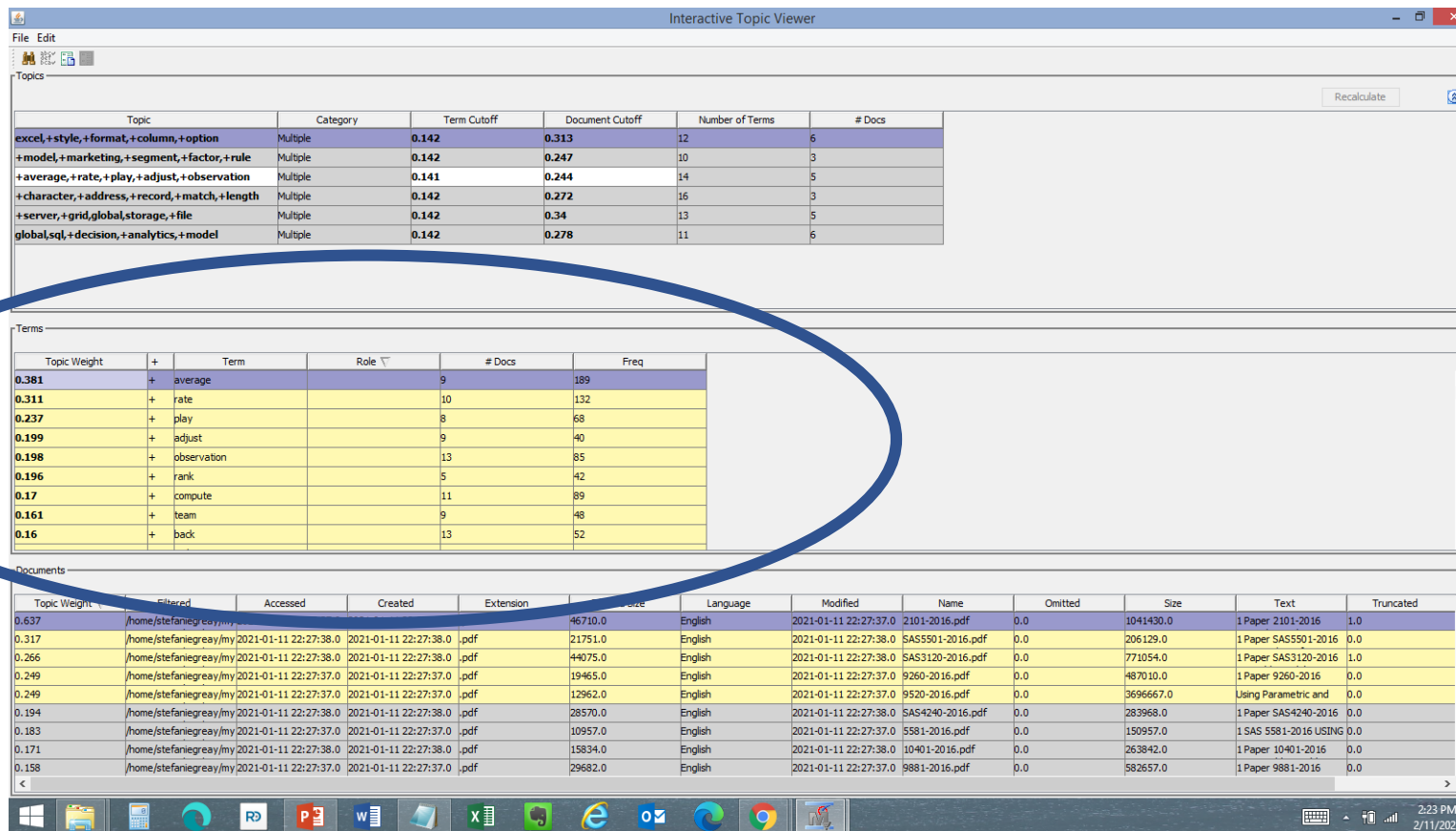


The screenshot shows a software window titled "Interactive Topic Viewer". It has a menu bar with "File" and "Edit". Below the menu bar is a toolbar with icons for file operations. The main area is labeled "Topics" and contains a table. The table has six columns: "Topic", "Category", "Term Cutoff", "Document Cutoff", "Number of Terms", and "# Docs". The first row of data is highlighted in blue and contains the following values: "excel,+style,+format,+column,+option", "Multiple", "0.142", "0.313", "12", and "6". A "Recalculate" button is located in the top right corner of the table area.

Topic	Category	Term Cutoff	Document Cutoff	Number of Terms	# Docs
excel,+style,+format,+column,+option	Multiple	0.142	0.313	12	6
model,+marketing,+segment,+factor,+role	Multiple	0.142	0.313	10	2



The terms window summarizes the terms included in each of the topics, their weights, frequencies, and documents they show up in.



Interactive Topic Viewer

File Edit

Topics

Topic	Category	Term Cutoff	Document Cutoff	Number of Terms	# Docs
excel,+style,+format,+column,+option	Multiple	0.142	0.313	12	6
+model,+marketing,+segment,+factor,+rule	Multiple	0.142	0.247	10	3
+average,+rate,+play,+adjust,+observation	Multiple	0.141	0.244	14	5
+character,+address,+record,+match,+length	Multiple	0.142	0.272	16	3
+server,+grid,global,storage,+file	Multiple	0.142	0.34	13	5
global,sql,+decision,+analytics,+model	Multiple	0.142	0.278	11	6

Recalculate

Terms

Topic Weight	+	Term	Role	# Docs	Freq
0.381	+	average		9	189
0.311	+	rate		10	132
0.237	+	play		8	68
0.199	+	adjust		9	40
0.198	+	observation		13	85
0.196	+	rank		5	42
0.17	+	compute		11	89
0.161	+	team		9	48
0.16	+	back		13	52

Documents

Topic Weight	Filtered	Accessed	Created	Extension	Size	Language	Modified	Name	Omitted	Size	Text	Truncated
0.637	/home/stefaniegray/my	2021-01-11 22:27:38.0	2021-01-11 22:27:38.0	.pdf	46710.0	English	2021-01-11 22:27:37.0	2101-2016.pdf	0.0	1041430.0	1 Paper 2101-2016	1.0
0.317	/home/stefaniegray/my	2021-01-11 22:27:38.0	2021-01-11 22:27:38.0	.pdf	21751.0	English	2021-01-11 22:27:38.0	SAS5501-2016.pdf	0.0	206129.0	1 Paper SAS5501-2016	0.0
0.266	/home/stefaniegray/my	2021-01-11 22:27:38.0	2021-01-11 22:27:38.0	.pdf	44075.0	English	2021-01-11 22:27:38.0	SAS3120-2016.pdf	0.0	771054.0	1 Paper SAS3120-2016	1.0
0.249	/home/stefaniegray/my	2021-01-11 22:27:37.0	2021-01-11 22:27:37.0	.pdf	19465.0	English	2021-01-11 22:27:37.0	9260-2016.pdf	0.0	487010.0	1 Paper 9260-2016	0.0
0.249	/home/stefaniegray/my	2021-01-11 22:27:37.0	2021-01-11 22:27:37.0	.pdf	12962.0	English	2021-01-11 22:27:37.0	9520-2016.pdf	0.0	3696667.0	Using Parametric and	0.0
0.194	/home/stefaniegray/my	2021-01-11 22:27:38.0	2021-01-11 22:27:38.0	.pdf	28570.0	English	2021-01-11 22:27:38.0	SAS4240-2016.pdf	0.0	283968.0	1 Paper SAS4240-2016	0.0
0.183	/home/stefaniegray/my	2021-01-11 22:27:37.0	2021-01-11 22:27:37.0	.pdf	10957.0	English	2021-01-11 22:27:37.0	5581-2016.pdf	0.0	150957.0	1 SAS 5581-2016 USING	0.0
0.171	/home/stefaniegray/my	2021-01-11 22:27:38.0	2021-01-11 22:27:38.0	.pdf	15834.0	English	2021-01-11 22:27:38.0	10401-2016.pdf	0.0	263842.0	1 Paper 10401-2016	0.0
0.158	/home/stefaniegray/my	2021-01-11 22:27:37.0	2021-01-11 22:27:37.0	.pdf	29682.0	English	2021-01-11 22:27:37.0	9881-2016.pdf	0.0	582657.0	1 Paper 9881-2016	0.0



You may review the initial results by right clicking on the “Text Topic” node and selecting “Results.”

The screenshot displays the Enterprise Miner application window titled "Enterprise Miner - myprojectname". The interface includes a menu bar (File, Edit, View, Actions, Options, Window, Help), a toolbar, and a left-hand pane with a project tree showing "myprojectname", "Data Sources", "Diagrams", "U1A1", and "Model Packages".

The main workspace shows a workflow diagram for "U1A1" with the following nodes: "Text Import", "Text Parsing", "Text Filter", "Text Topic", and "SAS Code". Arrows indicate the flow from "Text Import" to "Text Parsing", then to "Text Filter", then to "Text Topic", and finally to "SAS Code". Each node has a green checkmark icon below it.

A right-click context menu is open over the "Text Topic" node. The menu options are: "Edit Variables...", "Update", "Run", "Create Model Package...", "Results..." (highlighted), "Export Path as SAS Program", "Cut", "Copy", "Delete", "Rename", "Select All", "Select Nodes", "Connect Nodes", and "Disconnect Nodes".

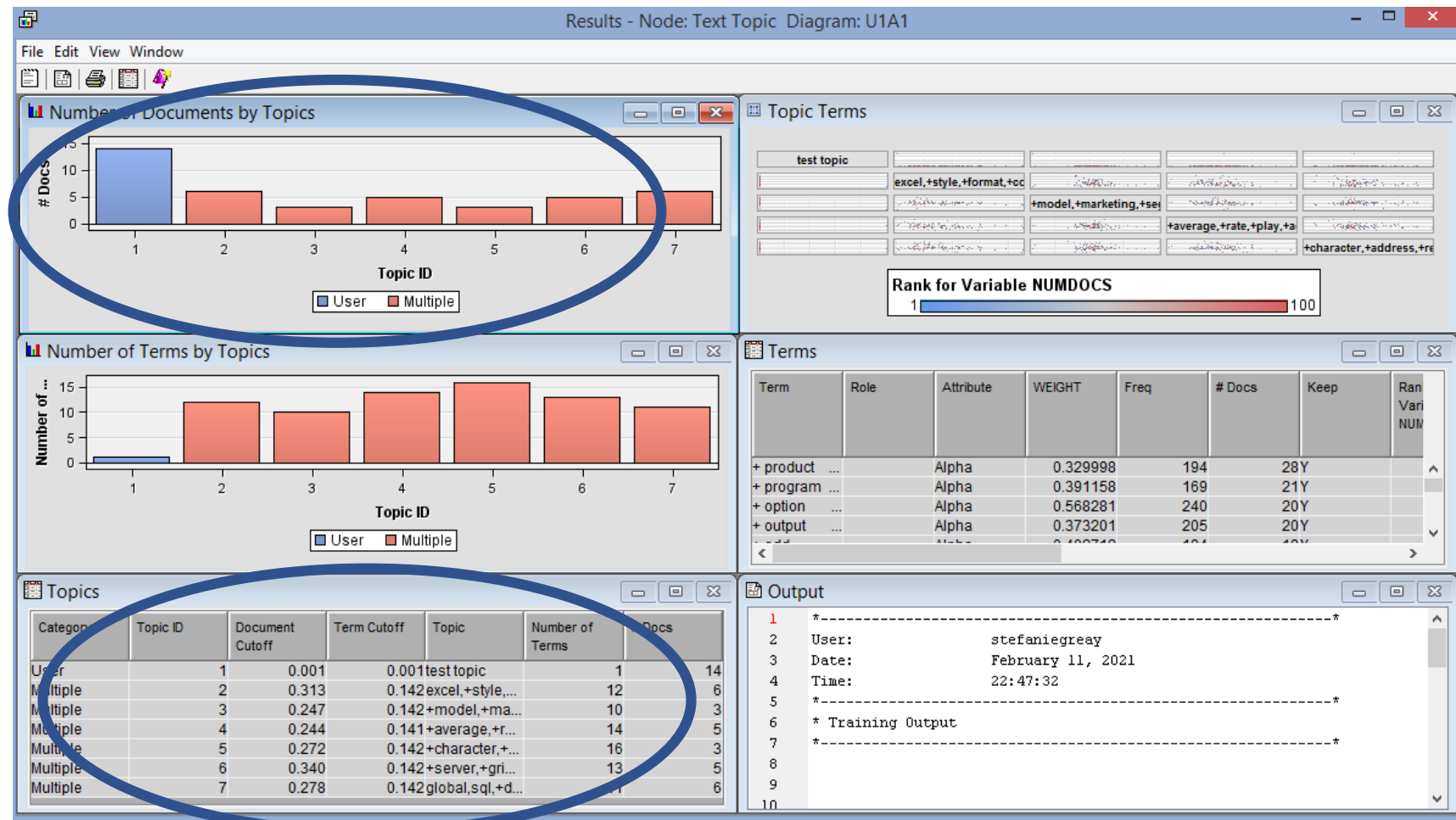
Below the workflow diagram, there is a "Property" table with the following data:

Property	Value
General	
Node ID	TextTopic
Imported Data	
Exported Data	
Notes	
Train	
Variables	
User Topics	
Term Topics	
Number of Single-term Topics	0
Learned Topics	
Number of Multi-term Topics	25
Correlated Topics	No
Results	
Topic Viewer	
Status	
Create Time	2/11/21 7:08 PM
Run ID	888957f3-0c9a-4b4a-905c-8016b6b8dd3c
Topic Viewer	

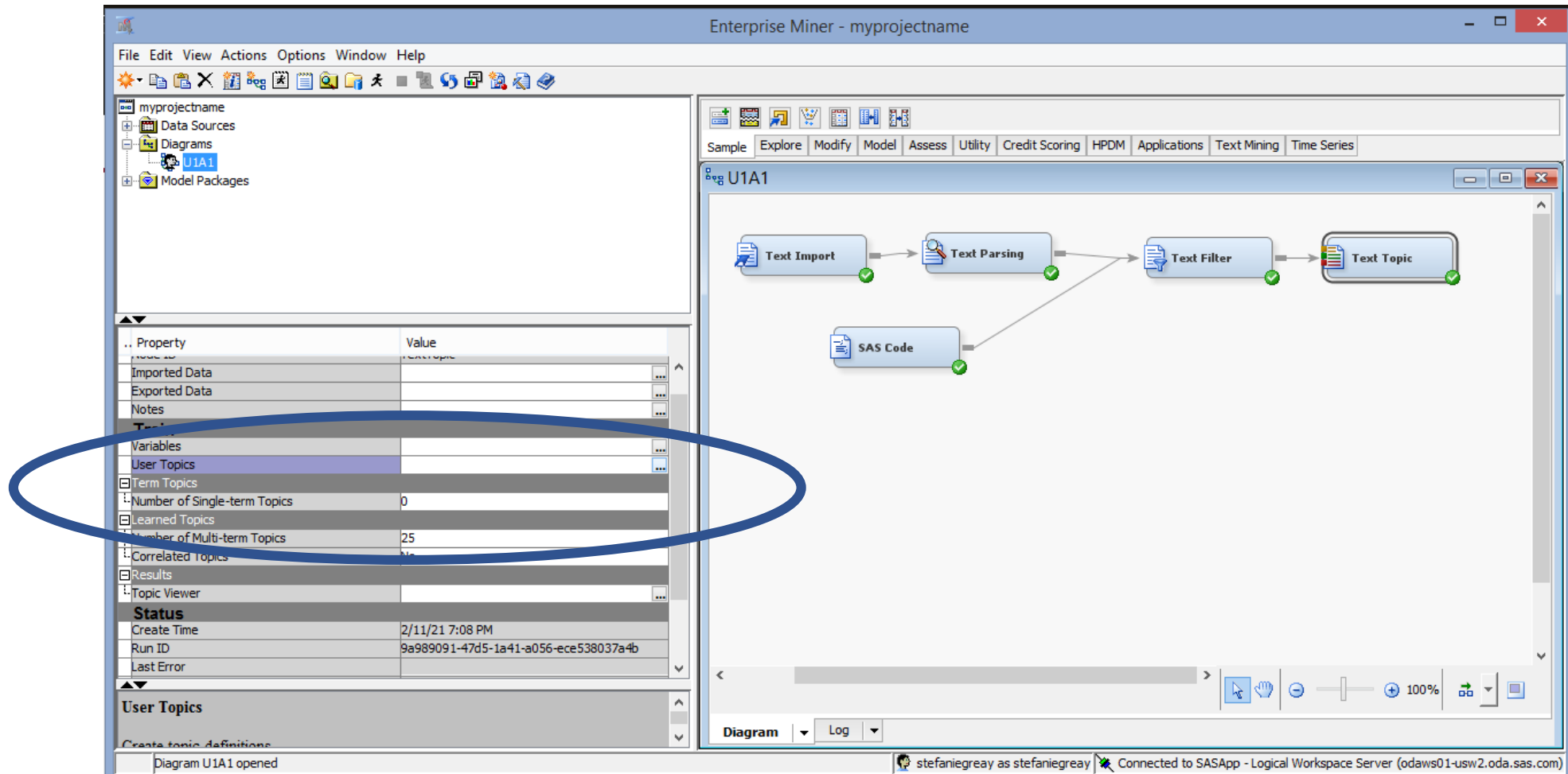
At the bottom of the window, a status bar shows "Diagram | Log" and a connection status: "stefaniegreay as stefaniegreay Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)".



In the results is where you will find the Topics table and the Number of Documents by Topic chart. Enlarge each of these individually and take a screenshot, as requested in the assignment. (Yours should not have a “user” category topic yet.)



To add user-defined topics, return to the main menu pane of the Text Topics node by clicking on the Text Topics node, and then click the ellipses next to “User Topics.”



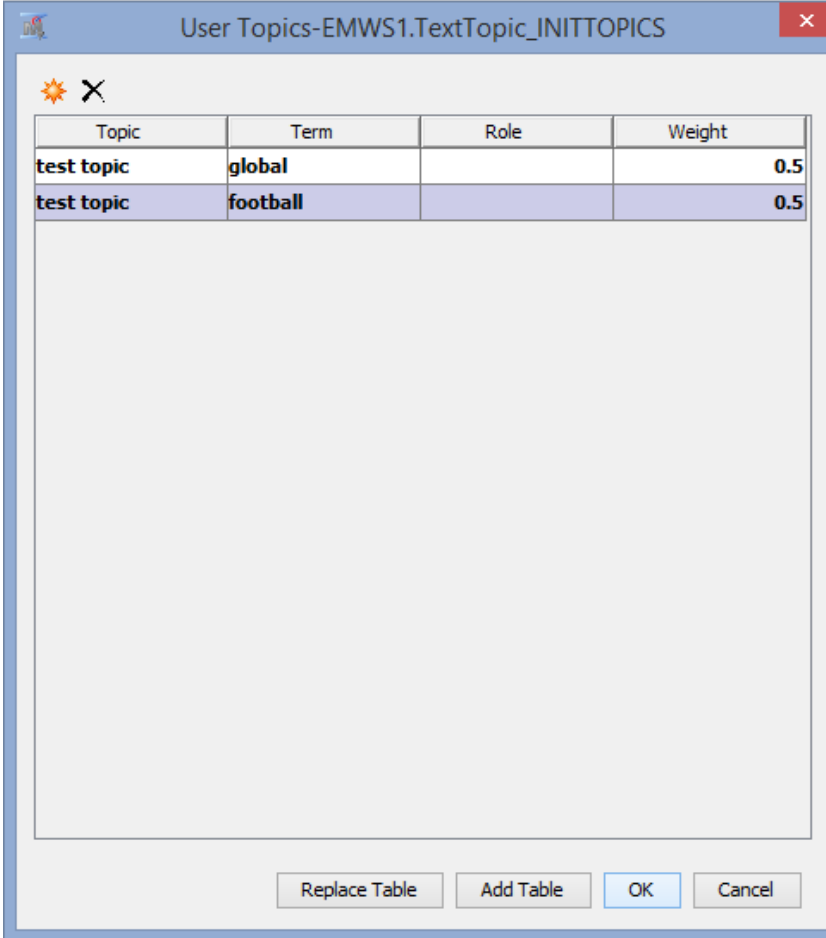
You can add your custom multi-term topic in this user topics menu.

The screenshot shows a software window titled "User Topics-EMWS1.TextTopic_INITTOPICS". Inside the window, there is a table with four columns: "Topic", "Term", "Role", and "Weight". The table is currently empty. At the bottom of the window, there are four buttons: "Replace Table", "Add Table", "OK", and "Cancel".

Topic	Term	Role	Weight
-------	------	------	--------



Add one row per term per topic, and be sure to spell and capitalize your topic the same in each row. Click “OK” to add the user topic. (DO NOT USE THIS TEST TOPIC...please create your own.)



Topic	Term	Role	Weight
test topic	global		0.5
test topic	football		0.5



Run the Text Topics node (and the path that precedes it) by right clicking on the Text Topics node and selecting “Run.”

The screenshot displays the SAS Enterprise Miner software interface. The main window, titled "Enterprise Miner - myprojectname", shows a workflow diagram in the center. The diagram consists of four nodes: "Text Import", "Text Parsing", "Text Filter", and "Text Topics". Arrows indicate a sequential flow from "Text Import" to "Text Parsing", then to "Text Filter", and finally to "Text Topics". A "SAS Code" node is also present, with an arrow pointing to the "Text Filter" node. Each node has a green checkmark next to it, indicating successful execution or completion. On the left side, there is a tree view showing the project structure, including "Data Sources", "Diagrams", and "Model Packages". Below the tree view, there is a property window for the selected node, "Text Topics". This window has tabs for "General", "Train", "Results", and "Status". The "General" tab is currently selected, showing properties such as "Node ID" (TextTopic), "Imported Data", "Exported Data", "Notes", "Variables", "User Topics", "Term Topics", "Number of Single-term Topics" (0), "Learned Topics", "Number of Multi-term Topics" (25), "Correlated Topics" (No), "Topic Viewer", "Create Time" (2/11/21 7:08 PM), and "Run ID". A context menu is open over the "Text Topics" node, showing options 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 in blue. At the bottom of the interface, there is a status bar showing the user "stefaniegreay as stefaniegreay" and the connection "Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)".



Click on the ellipses next to “Topic Viewer” to review the topics.

The screenshot displays the SAS Enterprise Miner software interface. The main window, titled 'Enterprise Miner - myprojectname', shows a project tree on the left with 'myprojectname' expanded, containing 'Data Sources', 'Diagrams', 'U1A1', and 'Model Packages'. The 'U1A1' diagram is selected, showing a workflow with five steps: 'Text Import', 'Text Parsing', 'Text Filter', 'Text Topic', and 'SAS Code'. Each step has a green checkmark indicating success. The 'Text Filter' step is currently selected. Below the workflow, the 'Topic Viewer' properties panel is visible, showing a table of properties and values.

Property	Value
General	
Node ID	TextTopic
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
User Topics	...
Term Topics	...
Number of Single-term Topics	0
Learned Topics	...
Number of Multi-term Topics	25
Correlated Topics	No
Results	...
Topic Viewer	...
Status	
Create Time	2/11/21 7:08 PM
Run ID	6d0390c0-ba27-6b47-80a2-87b905a1dc37

The 'Topic Viewer' section at the bottom of the properties panel is currently empty. The status bar at the bottom of the window shows the user 'stefaniegreay' is connected to the SASApp - Logical Workspace Server.



You may review the results by right clicking on the “Text Topic” node and selecting “Results.”

The screenshot displays the Enterprise Miner application window titled "Enterprise Miner - myprojectname". The interface includes a menu bar (File, Edit, View, Actions, Options, Window, Help), a toolbar, and a left-hand pane with a project tree showing "myprojectname", "Data Sources", "Diagrams", "U1A1", and "Model Packages".

The main workspace shows a workflow diagram for "U1A1" with the following nodes: "Text Import", "Text Parsing", "Text Filter", "Text Topic", and "SAS Code". Arrows indicate the flow from "Text Import" to "Text Parsing", then to "Text Filter", and finally to "Text Topic". A "SAS Code" node is also connected to the "Text Parsing" node. All nodes have a green checkmark, indicating they are successful or ready.

A context menu is open over the "Text Topic" node, listing the following options: "Edit Variables...", "Update", "Run", "Create Model Package...", "Results..." (highlighted), "Export Path as SAS Program", "Cut", "Copy", "Delete", "Rename", "Select All", "Select Nodes", "Connect Nodes", and "Disconnect Nodes".

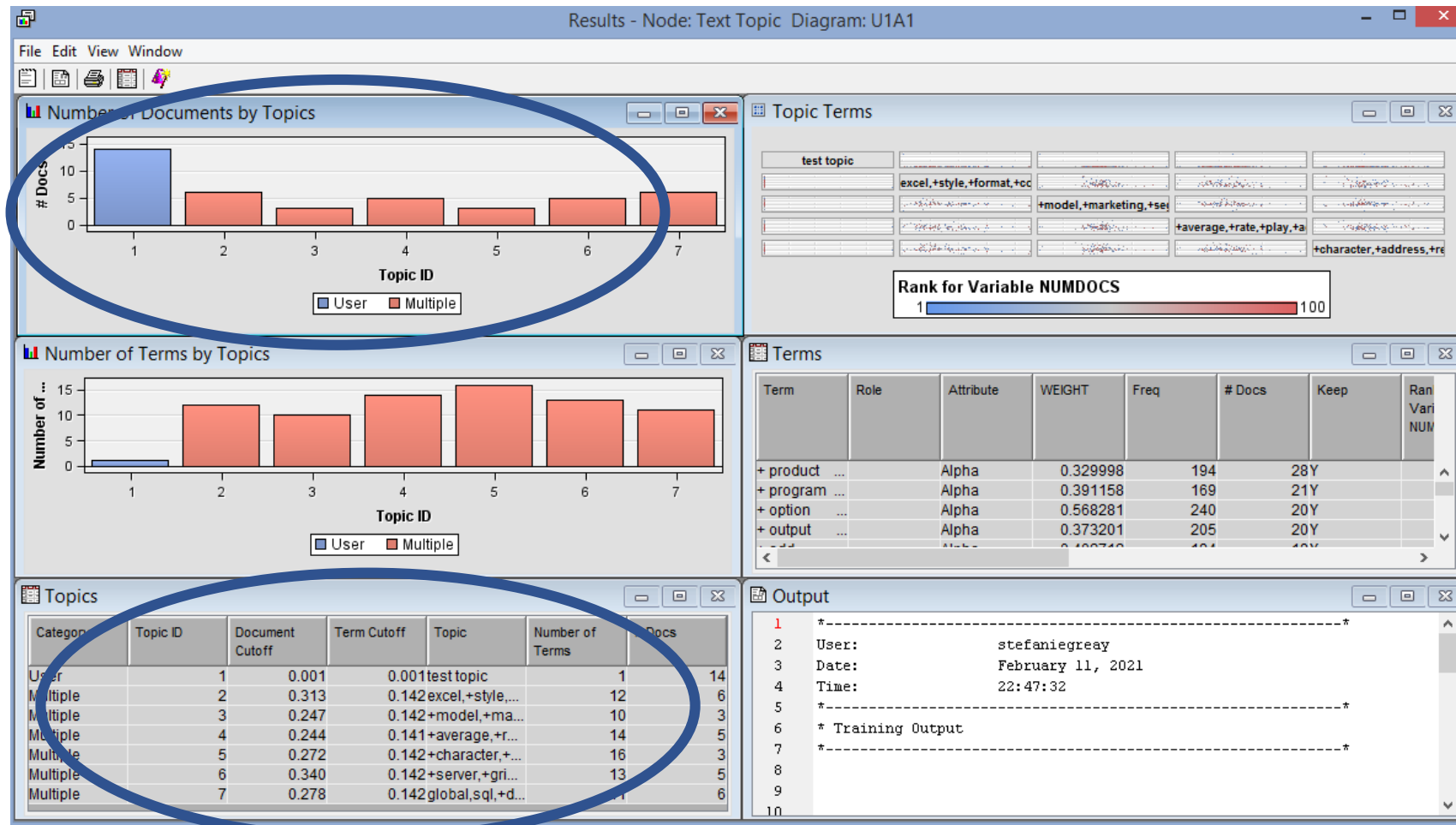
Below the workflow diagram, there is a "Property" table with columns "Property" and "Value".

Property	Value
General	
Node ID	TextTopic
Imported Data	
Exported Data	
Notes	
Train	
Variables	
User Topics	
Term Topics	
Number of Single-term Topics	0
Learned Topics	
Number of Multi-term Topics	25
Correlated Topics	No
Results	
Topic Viewer	
Status	
Create Time	2/11/21 7:08 PM
Run ID	888957f3-0c9a-4b4a-905c-8016b6b8dd3c
Topic Viewer	
Open the Interactive Topic Viewer	

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



In the results is where you will find the Topics table and the Number of Documents by Topic chart. Enlarge each of these individually and take a screenshot, as requested in the assignment.



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 6 – Clustering and Topic Extraction:
<https://capella.skillport.com/skillportfe/assetNonSSOLaunch.action?courseName= ss chapter:59026-342485391&courseType=7>
(Text Topic Node is at the bottom of the chapter)
- SAS Documentation on the Text Topic Node:
<https://documentation.sas.com/?docsetId=tmref&docsetTarget=n1sw3r97922sgmn1tqbwayrgn9eo.htm&docsetVersion=14.3&locale=en>

