# Step-by-Step SAS Text Miner Instructions for Text Clustering

We use SAS Text Miner to conduct text clustering. Text Miner is a component of SAS Enterprise Miner. This material is based on Enterprise Miner 14.1.

• <u>Practice Data</u>: Movie Text Reviews (45MB; Refer to the data description to understand the data.)

Download this dataset into your data directory folder.

To open <u>SAS Text Miner</u>, choose the SAS program named *Enterprise Miner*.



### Choose New Project.

Project Name: type in "movie"

SAS Server Directory – Type in the address of your target directory folder, which will store all the results files from SAS once the whole procedure is completed.

🕵 Enterprise Miner	
File Edit View Actions Options Window Help	
※ № 晚 X 圖 № 圖 圖 @ G ★ ■ 2 55 副 2 4 0	
SAS <sup>®</sup> Whelp Topics	
Niner* 14.1 Rew Project	
🤝 Open Project	
Recent Projects	
built by	
powered by 🔣 Create New Project Step 1 of 2 Specify Project Name and Server Dir 🗙	
SAS         SAS*         Enterprise         Miner~ 14.1         Project Name         movie         SAS Server Directory         I\Innovation Analytics\EMiner Project\Ex1 Text Clustering Browse	
< Back Next > Cancel	

## Choose *Next > Finish*. Now, you will see a workspace named "Enterprise Miner – movie."

K Enterprise Miner - movie	- 0
File Edit View Actions Options Window Help	
🚱 🐜 赵 X 🏭 😽 🗐 🚉 🕞 초 = 📜 55 🗐 🗒 🖏 🤣	
E Book Sources In Ed Data Sources Ed Data Sources Ed Data Sources Ed Model Packages	Sample Diptirre Modify Model Assess (Utility HPDM Applications   Text Mixing   Time Series)
AV	
Property Value	
Name movie	
Project Start Code	
Created 1/13/18 5:50 PM	
Server	
Grid Available No Path E:\Teaching\Innecation Apahtics\E	
Metadata Folder Path	
Max. Concurrent Tasks Default	
~	
Name	
Project Name	

#### To set up our data access, *File < New < Library < Create New Library*

🔣 Enterprise Miner - movie File Edit View Actions Options Window Help movie Data Sources Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series 🛋 Library Wizard -- Step 1 of 3 Select Action  $\times$ . Property Select an action – Value Create New Library .. Property Name Project Start Code Project Macro Variables Created Server Grid Available Path Metadata Folder Path Max. Concurrent Tasks novie Modify Library Delete Library 1/13/18 5:50 PM No F:\Teaching\Innovation Analytics\{ Default < Back Next > Cancel

## Library Wizard

*Name*: movie\_da *Path*: Type in (your data directory folder address).

■ Enterprise Miner - movie File Edit View Actions Opt	ions Window Help			
🔅 🖿 🖿 🗶 🎢 🏟 🖉	🔓 ž = 🗓 S5 🖥 🗎 🖓 🤣			
movie				
Diagrams		Sampl	Explore Modify Model Assess Utility HPDM Applications Text Mining Time	ne Series
The mover rackages			🔣 Library Wizard Step 2 of 3 Create or Modify	×
A <b>V</b>			Name	Engine
Property	Value		movie da	BASE ~
Name	movie		_	
Project Start Code				
Project Macro Variables				
Created	1/13/18 5:50 PM		- Library information	
Server			Path	
Grid Available	No		Faul	
Path Matadata Calder Dath	F:\Teaching\Innovation Analytics\t		F:\Teaching\Innovation Analytics	Browse
Max Concurrent Tasks	Default			
Max. concurrent rasks	Derauit			
			Options	
			< Pack N	(avt > Cancol
			< BdOk N	Cancer

Hit *Next* > *Finish*.

Place the mouse pointer on Data Sources.

🚿 Enterprise Miner - movie		
File Edit View Actions Optio	ns Window Help	
🔅 🗎 🐂 💥 🎢 🍓 🗮 🗐	🖥 🖈 🔳 🛍 Sy 🗬 🛍 🔕 🤗 👘	
movie movie mathematical definition mathematical defi		Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
Model Packages	Value	

Find Create Data Source (with right click on Data Sources).

膩 Enterprise Miner - movie				
File Edit View Actions Options W	indow Help			
🔆 🖿 💼 📉 🏭 🍓 🗷 📋 🎑 🕞 🖈	= 11 55 🗗 11 🗟 🤣 👘	·		
movie Data Sources Diagrams Model Packages	urce	Sample Explore	Modify Model	Assess Ut
Property Value	ρ	r		
Image: Second structure       Image: Second structure         File Edit View Actions Options Window Help         Image: Second structure	Sample Explore Modify Model Assess	Utility   HPDM   Applications   Text	Mining Time Series	
- Property Value	-			
The second secon	e Wizard Step 1 of 8 Metadata Source			×
	Select a metadata source Source : SAS Table		~	

< Back Next > Cancel

### Hit Next.

🛋 Enterprise Miner - movie	
File Edit View Actions Options Window Help	
🌞 - 🐂 🛍 🗡 🏭 🝓 诺 🗐 🍳 🕞 🧶 🔳 📜 55 🖃 📜 🏹 🤗 👘	
movie     movie     Data Sources     Diagrams     Common Decimates	Image: Sample         Explore         Modify         Model         Assess         Utility         HPDM         Applications         Text Mining         Time Series
H Model Packages	
Property Value	
Data Source Wiz	zard Step 2 of 8 Select a SAS Table X Select a SAS table Table : Browse.
	< Back Next > Cancel

Choose the target dataset in the created library.



### Hit OK > Next.

For the rest of the steps in the wizard, choose the default options. Then, you can see the target dataset in the *Data Sources* section.

風 Enterprise Miner - movie											
File Edit View Actions Option	ns Window Help										
🔅 🗎 🐂 🗶 🎢 🍓 🖈 🗐 🚉	j 🖈 🔳 🕄 🗗 🖺 🖓 🤣										
movie     movie     movie     MoviETEXTREVIEWS     MoviETEXTREVIEWS     Model Packages		Sample	Explore	Modify I	Model )	Assess	Utility	HPDM	Applications	Text Mining	Time Series
▲▼											
Property	Value										
TD	movietextreviews										
Name	MOVIETEXTREVIEWS										
Variables											
Decisions											
Role	Raw										
Notes											
Library	MOVIE DA										
Table	MOVIETEXTREVIEWS										
Sample Data Set											
Size Type											
Sample Size											
Туре	DATA										
No. Obs	2799										
No. Cols	10										
No. Bytes	45859840										
Segment											
Created By	Sangkil										
Create Date	1/13/18 6:00 PM										
Modified By	Sangkil										
Modify Date	1/13/18 6:00 PM										
Scope	Local										

# Now, to set up <u>Text Miner</u>, *Diagrams < Create Diagram* (with right clock on *Diagrams*)

Interprise Miner - movie	
File Edit View Actions Options Window Help	
🌣 - 🐚 🛍 🗙 🏭 😹 🗐 💁 🎧 🖈 🔳 🐮 55 🖃 🐩 🎝 🤣	
movie     Data Sources     MOVIETEXTREVIEWS	Sample       Explore       Modify       Model       Assess       Utility       HPDM       Applications       Text Mining       Time Series
Argenting Create Diagram     Import Diagram from XML	
. Property Value	

# Diagram Name: Enter 'movie\_go.'

風 Enterprise Miner - movie										
File Edit View Actions Option	s Window Help									
🔆 🖿 🛍 🗙 🏭 🍓 🗷 📋 🚉 😭	* = 🗓 S) 🗗 🗎 🎝 🤗 👘									
movie     mil Data Sources     mil Data MovIETEXTREVIEWS     mil MovIETEXTREVIEWS     mil MovIETEXTREVIEWS			Explore	Modify	<b>M</b> odel	Assess	Utility	HPDM	Applications	Text Mining
Model Packages	Value	-								
Hoperty	Value									
				Create	New D	Diagram	1		×	
				Dia	gram N	ame:				
				mo	ovie_go				_	
							ОК	(	Cancel	

Hit OK. Then, you will see a white workspace named "movie\_go."

🐧 Enterprise Miner - movie											
File Edit View Actions Optic	ons Window Help										
🔆 - 🖿 🕼 🗙 🏭 🍓 🗷 🗐 🚉 (	🙀 ž 🔳 🛍 😏 📲 🛄 🏹 🤗 👘										
movie  Data Sources  Sources Sources  Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources Sources		Sample	Explore	Modify	Model	Assess	Utility	HPDM	Applications	Text Mining	Time Series
Property	Value										
ID	EMWS1										
Name	movie_go										
Status	Open										
Notes											
History											
Create Date	1/13/18 6:02 PM										
Encoding	wlatin1 Western (Windows)										
Data Representation	WINDOWS_64										
Induve 05	res										

Drag the target data onto the workspace.

Enterprise Miner - movie     Eile Edit View Astions On	tions Window Holp										
rile Edit View Actions Op											
movie		Sample	Explore	Modify	<b>∐-</b> Model	Assess	Utility	HPDM	Applications	Text Mining	Time Series
Model Packages		<sup>8</sup> ल्ड mo	vie_go								
Property	Value										
General	^										
Node ID	Ids										
Imported Data											
Exported Data				HTT N	10VIETE>	TREV					
Notes				1	EWS						
Train											
Output Type	View										
Role	Raw										
Rerun	No										
Summarize	No										
Drop Map Variables	Yes										
Columns											
Variables											
Decisions											
- Refresh Metadata											
Advisor	Basic										
<sup>L</sup> Advanced Options											
Data											
Data Selection	Data Source										
Sample	Default										
<sup>L</sup> Sample Options											
Data Source											
Data Source	MOVIETEXTREVIEWS										
<sup>L</sup> Data Source Properties											
New Table											
Table Name											
Variable Validation	Strict										
<sup>L</sup> New Variable Role	Reject										
Metadata											
Table	MOVIETEXTREVIEWS										
	MOV/IE DA										
General											
General Properties											

Choose the *Text Mining* tab in the top menu.

🛒 Enterprise Miner - movie					
File Edit View Actions Option	ns Window Help				
🔅 🗈 🛍 🗙 🏭 🍓 🖾 🗎	ij 🖈 🔳 📜 📢 🗗 🗽 🦓 🥔				
movie					
MOVIETEXTREVIEWS		Sample Expl	Fext Parsing	Applications Text Mining	Time Series
movie_go  Model Packages		<sup>ଌ</sup> ଞ୍ଚ movie_go	Parse text in a document collection.		
<b>AV</b>					
Property	Value				
General	^				
Node ID Id	ds				
Imported Data					
Notes			MOVIETEXTREV		
Train					
Output Type V	ïew				
Role R	aw				

Choose the *Text Parsing* node and drag it onto the workspace.

🛒 Enterprise Miner - movie											
File Edit View Actions Ont	tions Window Help										
The Edit view Actions opt											
🐥 - 🗈 🖫 🗙 📆 🐜 🛒 🗒 🧟	] 请 X 🔳 📜 😏 🖬 😫 🦪 🤣										
📼 movie											
🖨 🛅 Data Sources			2 🛤								
MOVIETEXTREVIEWS		Sample	e Explore	Modify	Model	Assess	Utility	HPDM	Applications	Text Mining	Time Series
🖶 🖳 Diagrams				,			,				
movie_go		Pe8 mo	vie_go								
🗄 💽 Model Packages											
		1									
Property	Value										
General	^										
Node ID	TextParsing										
Imported Data				_							
Exported Data					IOVIETE)	TREV		6	Canal Dancia	]]	
Notes				<b>1</b>	EWS	J			P Text Parsi	"g	
Train											
Variables											
Parse											
Parse Variable	Ex all also										
Language	English										
Different Parts of Cheese	Vac										
Noun Crounc	Voc										
- Multi-word Terms	SASHELP ENC. MULTI										
Find Entities	None										
Custom Entities	- Hone										
Ignore Parts of Speech	'Aux' 'Conj' 'Det' 'Interj' 'Part'										
Ignore Types of Entities											
L 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	1/12/18 6:04 PM										
General											

Choose the *Text Filter* node and drag it onto the workspace.

Spelling

of Docun of Term

膩 Enterprise Miner - movie						
File Edit View Actions Optic	ons Window Help					
🔅 🖻 🛍 🗙 🏭 🍓 🖹 📋 💁	请 * 🔳 📜 55 🗗 🗽 🏹 🤗					
📼 movie 🖶 💼 Data Sources						
MOVIETEXTREVIEWS		Sar	Text Filter	Applications	Text Mining	Time Series
		800	Apply filters to reduce the number of terms			
🗄 🖻 Model Packages			or documents included in a text analysis			
			The Text Filter node must be preceded by			
			a Text Parsing node and may be preceded by			
			by Text Filter and Text Tenis nodes			
A <b>V</b>			by Text Filter and Text Topic flodes.			
Property	Value					
General		~				
Node ID	TextParsing					
Exported Data				6		
Notes			IEWS	🎽 Text Parsin	g	
Train						
Variables						
Parse Variable						
Language	English					
Detect						
Pointerent Parts of Speech	Yes					
Multi-word Terms	SASHELP.ENG_MULTI					
Find Entities	None					
•						
風 Enterprise Miner - movie						
File Edit View Actions Opt	tions Window Help					
🔅 - 🗈 🛍 🗙 🏭 🗮 🗐 🍳	] 🕞 夫 🔳 📜 😏 🗗 🗽 🏹 🥔					
= movie						
MOVIETEXTREVIEWS		Sample Ex	plore Modify Model Assess Utility HPDM Applications Text Mi	ning Time Series	5	
🖨 🚾 Diagrams		her mouio			_	
Model Packages		-va movie_g	go			
AT						
Property	Value					
General Node ID	TextFilter ^					
Imported Data	in continued			C		2
Exported Data Notes			Text Parsing		Text Filter	
Testa						9

Choose the *Text Cluster* node and drag it onto the workspace.

🛋 Enterprise Miner - movie						
File Edit View Actions Opti	ons Window Help					
🔅 🖹 🖹 🙈 🗙 🎬 🖓 🏟	🔓 A 🔳 📜 🐓 🗗 🗽 🏈					
I movie I Division Sources I Di Moviet Extreviews I Diagrams I I Movie go II Model Packages		Sa Beg	Text Cluster The Text Cluster node is used to perform a clustering analysis on a document collection. The Text Cluster node must be preceded by Text Parsing and Text Filter nodes and may also be preceded by a Text	Applications	Text Mining	Time Series
Property	Value		Topic node.			
General Node ID Imported Data Exported Data Voreables Spelling Occessform Occessform Verlahtnas Frequency Weighting Term Veight BTerm Stillers Maximum Number of Terms	No realized for the second sec			Text Parsi	ng	Text Filter

Enterprise Miner - movie		
File Edit View Actions Opt	tions Window Help	
<u>*-</u> ⊳ @ <b>∨</b> ₩ ≥ [2] ~ 6		
	」 <u>□</u> ] × = ≝ ♥) ⊡* ⊇, ∛⊴ ≪	
movie		
		Company Service Media Anna Weller UDDM Analysis a surger Time Contra
		Sample Explore Modily Model Assess Julity HPDM Applications Text Mining Time Series
and movie go		<sup>8</sup> eg movie go
🗄 💽 Model Packages		
A.		
Property	Value	
General		
Node ID	TextCluster	
Imported Data		
Exported Data		MOVIETEXTREV Text Cluster
Notes		I IEWS I IEXT Parsing
Train		
Variables		
Transform		
SVD Resolution	Low	
<sup>L</sup> Max SVD Dimensions	100	
Cluster		
Exact or Maximum Number	Maximum	
Number of Clusters	40	
Cluster Algorithm	Expectation-Maximization	
Descriptive Terms	15	
Status		
Create Time	1/13/18 6:06 PM	
Run ID		
Last Error		
Last Status		
Last Run Time		
Run Duration		
Grid Host		
User-Added Node	No	

18

Link the data to the *Text Parsing* node. Link the *Text Parsing* node to the *Text Filter* node. Link the *Text Filter* node to the *Text Cluster* node.

🕂 Enterprise Miner - movie		
File Edit View Actions Optio	ns Window Help	
🔅 🖻 🖹 🗙 🎢 🍓 🗷 📋 🍳	🧃 🖈 🔳 📜 👀 🔐 😫 🏈 🥔	
movie     Data Sources     MOVIETEXTREVIEWS     Diagrams		Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
- We movie go		موة wovie_go
Property	Value	
General	TestChates	
Imported Data	rexicluster	
Exported Data		
Notes		I lext Parsing
Train		
SVD Resolution	Low	
Max SVD Dimensions	100	
E Cluster		
Exact or Maximum Number	Maximum	

To run the required text clustering analysis,

Click the *Text Cluster* node on the workspace. Then, you will see the detailed operation information of the node on the lower left side.

Train < Cluster

*Exact or Maximum Number*: Choose '*Exact*.' Determine your *Number of Clusters*. (2 in our case) Determine your number of *Descriptive Terms*. (15 is the default option.)

🔣 Enterprise Miner - movie			
File Edit View Actions Optic	ons Window Help		
🔅 🗈 🛍 🗙 🏭 😹 🗐 🍳	🚔 🖈 🔳 📜 😏 🔂 🎇 🌏 🤣		
movie     Data Sources     MOVIETEXTREVIEWS     Go Jiagrams     Go Movie_go     Model Packages		A      A	Text Mining Time Series
AY			
Property	Value		
General			
Node ID	TextCluster		
Imported Data			
Exported Data		MOVIETEXTREV	Text Filter
Notes		IEWS IEWS	
Train			
Variables			
Transform			
SVD Resolution	Low		
Max SVD Dimensions	100		
Cluster			
Exact or Maximum Number	Exact		
Number of Clusters	2		
Cluster Algorithm	Expectation-Maximization		
Descriptive Terms	15		
Status			
Create Time	1/13/18 6:06 PM		
Run ID			
Last Error			
Last Status			
Last Run Time			
Run Duration			
Grid Host			
User-Added Node	No		

After placing the mouse pointer on the Text Cluster nod, choose *Run* with right click.

🔣 Enterprise Miner - movie			
File Edit View Actions Opt	ions Window Help		
🌞 🗈 🖹 🗙 🕅 🖦 🗷 🗐 🏹	i 🖬 🖈 🗉 📜 😘 🖗 😒 🔕 🥔		
movie     Data Sources     MovIETEXTREVIEWS     Diagrams     Model Packages		Sample Explore Model Assess Utility HPDM Applications Text Mining Time Series	
A.			
Property	Value		
General			
Node ID	TextCluster		
Imported Data			
Notes		MOVIETEXTREV = Stext Parsing = Text Filter = Steven Text	Cluster
Train			Edit Variables
Variables			
Transform			5 Update
SVD Resolution	Low		★ Run
<sup>L</sup> Max SVD Dimensions	100		2 Create Medel Backage
Cluster			Cleate Model Fackage
Exact or Maximum Number	Exact		Results
Number of Clusters	2		Export Path as SAS Program
Cluster Algorithm	Expectation-Maximization		
*Descriptive Terms	15		Cut
Status Create Time	1/12/18 6:06 PM		Copy
Run ID	1/13/18 0.00 PM		copy
Last Error			Delete
Last Status			Delete
Last Run Time			Rename
Run Duration			
Grid Host			Select All
User-Added Node	No		Select Nodes
			Course the dee
			Connect Nodes
			Disconnect Nodes
		1	
- Annual of Chiefore			

It may take some time to generate the text clustering results (usually, a few minutes).

🔣 Enterprise Miner - movie										
File Edit View Actions Opt	ions Window Help									
🔅 - 🗈 🛍 🗙 🏭 🍓 🗷 📋 🍳	🕞 * = 🖹 😘 🗗 🖺 🏹 🤗									
movie     Data Sources     MOVIETEXTREVIEWS		Sample Explore Modify	Model Assess	Utility HPDM	Applications	Text Mining	Time Series			
movie_go		Reg movie .go								
🗄 🖻 Model Packages										
Property	Value									
General										
Node ID	TextCluster									
Imported Data		_								
Exported Data		DOD N	IOVIETEXTREV	>	Text Parsin		Text Fi	ilter		Cluster
Notes			EWS	- 1		.,	-3			
Train										
Variables										
ETransform										
SVD Resolution	Low									
Max SVD Dimensions	100									
ECluster										
Exact or Maximum Number	Exact									
Number of Clusters	2									
Cluster Algorithm	Expectation-Maximization							Confirmation		X
* Descriptive Terms	15							commutor		~~
Status									nt to run this nath	2
Create Time	1/13/18 6:06 PM							2 00 you wa	ne to run ans pau	
Run ID								Diagram:n	novie_go	
Last Error								Path T	ovt Cluster	
Last Status								1001. 1	ext cluster	
Last Run Time									٦	
Run Duration								Yes	No	
Grid Host										
User-Added Node	No									
1										

Hit *Yes* to see some results. The results you see here is only a small fraction of the results the program have generated, which can be found in the following directory address:

(Your project target directory folder address) < movie < Workspaces < EMWS1

EMWS1					
File Home Share View					
$\leftarrow$ $\rightarrow$ $\checkmark$ $\uparrow$ $\blacksquare$ > This PC > SM2015 (F:) > Teac	hing > Innovation Analytics > EMiner Project	ct > Ex1 Text Clustering > m	novie > Workspace	s > EMWS1	
> 📕 UNCC	Name	Date modified	Туре	Size	
📙 System Repair	em doraph	1/13/2018 6:11 PM	SAS Data Set	128 KB	
✓ 🥪 SM2015 (F:)	em nodeid	1/13/2018 6:11 PM	SAS Data Set	128 KB	
> 📙 Personal	ids	1/13/2018 6:11 PM	SAS Catalog	41 KB	
> 📙 Research	🛄 em_history	1/13/2018 6:11 PM	SAS Data Set	128 KB	
V Teaching	🔜 textcluster	1/13/2018 6:11 PM	SAS Catalog	57 KB	
Innovation Analytics	textcluster_exportset	1/13/2018 6:11 PM	SAS Data Set	128 KB	
EMiner Project	textcluster_emreport	1/13/2018 6:11 PM	SAS Data Set	128 KB	
Evi Text Clustering	textcluster_emtrainvariable	1/13/2018 6:11 PM	SAS Data Set	128 KB	
* Extrex clustering	textcluster_cmeta_train	1/13/2018 6:11 PM	SAS Data Set	128 KB	
<ul> <li>Inovie</li> </ul>	textcluster_eminfo	1/13/2018 6:11 PM	SAS Data Set	128 KB	
DataSources	textcluster_graph_table	1/13/2018 6:11 PM	SAS Data Set	208 KB	
HPDM	textcluster_clusters	1/13/2018 6:11 PM	SAS Data Set	136 KB	
Meta	textcluster_train	1/13/2018 6:11 PM	SAS Data Set	36,096 KB	
Reports	textcluster_svd_u	1/13/2018 6:11 PM	SAS Data Set	3,072 KB	
📜 System	textcluster_tmconfig	1/13/2018 6:11 PM	SAS Data Set	128 KB	
✓	textcluster_docs	1/13/2018 6:11 PM	SAS Data Set	1,152 KB	
V 📜 EMWS1	textcluster_emoutstat	1/13/2018 6:11 PM	SAS Data Set	128 KB	
HPDM	textcluster_terms	1/13/2018 6:11 PM	SAS Data Set	10,304 KB	
	textcluster_svd_s	1/13/2018 6:11 PM	SAS Data Set	128 KB	
Sutom	textcluster_last_tm_nodes	1/13/2018 6:11 PM	SAS Data Set	128 KB	
- System	textcluster_importset	1/13/2018 6:11 PM	SAS Data Set	128 KB	
> Entruster	textcluster_variableset	1/13/2018 6:11 PM	SAS Data Set	128 KB	
> 📕 lextFilter	textfilter	1/13/2018 6:11 PM	SAS Catalog	57 KB	
> 📕 TextParsing	textfilter_emreport	1/13/2018 6:11 PM	SAS Data Set	192 KB	
> 📙 MBA Hybrid Course Development	textfilter_emtrainvariable	1/13/2018 6:11 PM	SAS Data Set	128 KB	

Because there are so many files generated, finding the information we need requires some exploration. To bring up the required data to fill in the three tables in the exercise, let's use *SAS Enterprise Guide*. (Because I cannot explain all the basic operation functions of Enterprise Guide here, I will cover only the most direct functions.)



Find the data file named 'taxtcluster\_textclusters' in the output file and drag it onto the *Process Flow* section of Enterprise Guide. This dataset has the information for Table 1.



Next, find the data file named 'taxtcluster\_train in the output file and drag it onto the *Process Flow* section of Enterprise Guide. This dataset has the information for Table 2 & Table 3.

and lose	· · ·	r toin +												A X and		300 × × + +		2 A 1993 - 1993		1 A Red	5 -
Process Film	E Titer e	no fort Balle	er Where   Del	a · Denotier · 1	Stands = desalvers =	Export + Send So	- 181							Inter inter	ANN N1. [17	J & V . W. & 3	A L'E. C. EE A. I	- AndbCcl AaBbC	Cel AaBbCel AaBbCel	AaBbCel	phane
brablastiar_clasters		Test La	the d an annu a		A much Chart	O Der stert O	instant of		A state	Pavlan Hod		TestOurne 5	Test: n	7000 J B I	U X,	x' 泉・芝・ <u>A</u> ・覆:	(日本) - 대 藤田三三〇(1)	· Emphanis Tribudin	ng1 Tikindog2 Thioreal	Streep + 1: Sele	e(3.%
24 Martine Jan		- sch	C a main		Contraction of the local division of the loc	·	Canada a	Long a	w uten	4 6 6 April 1	a month	VDr		ristont a		Tot:	a tracet		1941	1 100	-
	1 100	e case.		1.2	-	1				202 0		0.0041030113	4.00	B. (1 D. D.	A D D						
	1	the face			-			14	0.7302941126	111.0		0.0030110140	- 22								
	3 100	nd cillicity.	1	1.7		7		21	0.0011420001	317 0	1	0.1/10/10/04	1.20								
	a The	e is soften.		1.2		2		1		279.0		0.421307061	4.5								
	6 84	reporable L.		1.2	4		-46	60	0.7063423529	135-0		8 708260177	-01								
	2 The	wis withing.		1.2	2	6			0	182 0	3	0.614000400	4.6								
	2 1wl	the the first.	1	1.2	3	2	4	7	0.0714200714	252 1	1	0.7375036275	412								
	D The	in to some .	1	1.2	7	10	2	2		169 0	1	0.6724633241	4.20			Next, fo	red the data file marned 'textchaster	train in the output file and	Litrar it onto the		
	10 A.F	orsught ch.	1	1.2	1	5				272 0		0.6399616222	0.22			Process	Flow section of Entermise Guide.	This dataset has the inform	mation for Table 7.8		
	11 214	pane. They,	T	3.4	0	4				383 1	11	0.6965854675	-12			Table 2	The second of Environment Only				
	12 Spo	Ini Alet	* *	3.4	1		0.			799 1		0.3055546622	-00								
	(3) 1 he	led this ma.	2	3.4	1		14	-20	0.7	60.0	0	0.551258627	-0.1			E					
	14 The	scanvi.	2	3.4	1	3	2	2		79.0		0.6413118367	-4.95								
	11 Ales	méra Conz.	1	3.4	3		N	2	05	952 0	15	0.2535589411	-204								
	10 Ales	andra Ocea.		3.4	3	10	4		0.8	314.0		0.4256089643	4.05								
	17 860	, what I W.		3.4		3	N		0.5	369.0	.17	0.1268567929	- 51								
	10 144	6 espectes.		3.4		10	2		2.000040808.7			0.29676291/2	- 90								
	19	PULLION.								400 5		0.523,9430,962	- 22								
	20 10.0	tion, cannot t		3.4	1					122.0	10	0.0007000000	1.00								
	22 14	entering 2		14	1		- 2	-		117.0		0.023/20022									
	27 100	di sin di		14						400.0	11	0.1004010071	2.00								
	55 The	many had.		2.4	7	1	16	24	1.000000007	236.0	24	0.4515299982	1.00								
	SC 44.	the limit a		5.1	6	2			6.10CM30C7	962.0	16	IL SOLICEPTIN	4.80								
	25 6.00	n Sandiar		5.2	2			1	0.5	439.0		0.3590755405	4.85								
	27 90 7	Int Dates	1	5.2	0	1	3		0.4285114286	92.0	27	0.4711394022	4.00								
	25 652	echel \$2.54	T	5.2	0	3	1	2	85	244.0	20	0.3509640547	4.10								
	20 144	ne stat ly.	T	5.2	1	8				441 1	22	0.4331713546	4.00								
	30 I M	rit give any .	1	6 Z	1	6	2	3	1.0000000027	81.0	30	0.4953820483	-0.14								
	25 Ibs	inte almi.	1	5 Z	1	2	0.5	2		61.0	31	0.9606838362	4.12	6455 27 OF 27 - 585	ADDDS TH	CHEVEN LINES CONCO.				1 1 5	
	32 A.G	preedy in w.:	. 1	5.2	1				0.25	117:0	32	0.2914422147	4.10		-		Contraction of the local division of the loc				and the second state of the second state
	33 WM	le frie movi.	1	5.2	2	3	2		0.25	143 0	30	0.5455775403	4.15	- CI	0	🔞 🖉 🗵	K3 💦 M3 🖂	1.00		A 10 40 A 🖬	1045.00
	24 190	e tho movi.		2.3	0	8	2.		0.675	184 0	34	0.7280770738	0.017	-	_						17 PERMIE
	25 1 10	ed during C.	1	7.3	0	9	3	10	- 0.9	195 0	35	0.221452169	0.080								
	20 114	dy say this.	1	7.3	÷	6	1	2	0.5	190-0	20	0.2562808325	0.010								
	57 1.60	et nesd kos.	1	7.5			2	4	- 05	182.0	37	0.4336542366	410.								
	¢ [												3								

#### For Table 2 information, choose Tasks > Describe > Summary Statistics

🞯 SAS Enterprise Guid	de										
File Edit View	Tasks	Favorites	Progra	m	Tools	Help	1	- 🛛 🕼	84		ß
Project Tree	Br	owse		- x	textclu	uster_tra	in •				
Process Flow	D	ata	•		🐺 Filte	er and S	ort 🖷 Q	uery Buil	der 🍸	Wh	ere
textcluster_clu	D	escribe	•		List Da	ita			er_cl	12	Mo
	G	aph	•	Σ	Summ	ary Stati	stics Wiza	ard	1		
	A	NOVA	•	Σ	Summ	ary Stati	stics		1		
	Re	egression	•		Summ	ary Table	es Wizard	ł	1		
	M	ultivariate Invival Analysi	· ·		Summ	ary Table	es		1		
		invival Analysi	<u> </u>		List Re	port Wiz	zard				
		apability			Charac	terize D	ata		1		
	E Pa	reto Chart	•	đh	Distrib	ution Ar	nalysis		1		
		neto chart			One-W	/ay Freq	uencies		1		
		me Series	•		lable /	Analysis.			_ <u>'</u>		
		ata mining	•		12	Spoiler A	s. They		1		
	0	LAP	•		13	I hated th	nis mo		2		
	Ta	sk Templates	•		14	The scre	enwrit		2		
					15	Alejandro	Gonz		1		
					16	Alejandro	Gonz		1		

Choose the three variables indicated in Table 2 from the *Name* section and drag them onto the *Analysis variables* section.

Next, place the TextCluster\_cluster\_ variable from the *Name* section and drag it onto the *Classification variables* section.

Project Tree • • Summary Sta	atistics •	
E Process Flow	🛄 Code 🗐 Log 🙎 Results	_
Important       ∑       Important         Important       ∑       Summary Statistics         Important       ∑       Sumary Statistics         Important       ∑	Code Code Code Code Code Code Code Code	× 
	(TeXUsing SVG      (TeXUsi	
	TextCluster_SVD9     TextCluster_SVD10     V	
	Select a role to view the context help for that role.	~

Hit *Run*. Then, you will get the information you need for Table 2.

For Table 3 information, choose Tasks > Describe > One-Way Frequencies



Choose Genre from the *Name* section and drag it onto the *Analysis variables* section. Next, place the TextCluster\_cluster\_ variable from the *Name* section and drag it onto the *Group analysis by* section.



Hit *Run*. Then, you will get the information you need for Table 3.

Save the whole content of your work on Enterprise Guide in the name of "Ex 1 Text Clustering" as a project by choosing

#### *File > Save Project As*

	New Open Close Pr								
→ () →	Open Close Pr						▶ y Fr	equ	encies •
( 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Close Pr						Dat	talí	🗓 Code
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	crose i r	oject					esh		-• Modifv 1
2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Save Pro	ject			(	Ctrl+S			
∰ 9 ¶   ▶ F 9 1	Save Pro	ject As							
₩    >       	Save All				Ctrl+S	nift+S			
► F S E	Import [	Data							
S	Run Proj	ect							
E	Schedul	e Projec	t						
5	Export						•		
	Send To						•		
H I	Publish								
da P	Page Se	tup for S	AS Repo	ort - One-W	ay Frequenc	ies			
F	Print Pre	view for	SAS Re	port - One-\	Nay Frequer	icies			
<u> </u>	Print SA	S Report	t - One-\	Nay Frequer	ncies (	Ctrl+P			
	Project I	Propertie	es						
F	Recent F	Projects					•		
F	Recent F	rogram	s				•		
E	Exit								

ct Tree 🔹	One-Way Frequencies -	
Process Flow	ImputData I Code I Log S Results S Refresh I Modify Task   Export - Send To - Create - Publish   III Properties	
Σ Summary Statistics One-Way Frequencies	Save	- <b>D</b> X
	Save in: La Teaching V 4 V 10 11	· • \$5
	Image: Desktop     Name     Date mod       Image: My Documents     1/14/2018       Image: My Computer     Image: My Network Places       Image: Servers     Image: Servers	ified Type Size 11:05 AM File folder 6:38 AM File folder 7 9:53 AM File folder
	File name: Ex 1 Text Clustering	~
	Files of type: Enterprise Guide Project Files (*.egp)	×

Now, you have obtained all the analysis results for **Task A: Text Clustering – 2 Clusters** in the exercise. You should use necessary information to actually complete Task A. Next, to do **Task B: Text Clustering – 3 Clusters**, go back to the workspace named movie\_go on *Enterprise Miner*.

Point to the *Text Cluster* node and choose *Copy* with right click.



Move the mouse pointer out of the node and place it at any empty space spot and choose *Paste* with right click.

🔣 Enterprise Miner - me	ovie						
File Edit View Actions	Options Window Help						
🔅 - 🐚 🛝 🗙 🏭 🍓 🗷 (	1 9 G * = 1 9 F 1 3 4						
movie							
MOVIETEXTREVIE	WS	a La prince Matter Made Assess Utility UDDM Applications Text Mising Time Social					
🖶 🖻 Diagrams		Sample Copore modely model Assess daity in bin Applications reachining Time Series					
- movie_go		ere movie_go					
Model Packages							
A <b>V</b>							
Property	Value						
ID	EMWS1						
Name	movie_go						
Status	Open						
Notes							
History		I I I I I I I I I I I I I I I I I I I	er				
Create Date	1/13/18 6:02 PM		<b>v</b>				
Encoding	wlatin1 Western (Windows)						
Data Representation	WINDOWS_64						
Native US	Yes						
		Add Node >					
		Paste					
		Select All					
		Select Nodes					
		Layout >					
		Zoom >					
		Copy Diagram to Clipboard					
		Reset Diagram					

Then, you will get a new *Text Cluster* node.

🛋 Enterprise Miner - movie									
File Edit View Actions Optic	ns Window Help								
🔅 - 🗈 🛍 🗙 🏭 🍓 🗷 🗐 🍳	🧃 🖈 🔳 📜 😏 🔂 🍇 🌏 🥔								
movie     movie     Jota Sources     Log MoVIETEXTREVIEWS     Molagrams     Log movie.go     Model Packages		Sample Exp	V Nodel Asse	ess Utility HPDI	N Applications Text Mir	ning Time Seri	es		
A <b>V</b>									
Property	Value								
General Node ID Imported Data Exported Data Notes Train Variables ETransform	TextCluster2			EV	Text Parsing		Text Filter	5	Text Cluster
SVD Resolution Max SVD Dimensions Cluster Exact or Maximum Number Number of Clusters Cluster Algorithm Descriptive Terms Status Status	Low 100 Exact 2 Expectation-Maximization 15								Text Cluster
Create Time	1/14/18 11:14 AM								

Place the pointer on the new node and choose *Rename* with right click.

Enterprise Miner - movie	2								
File Edit View Actions On	tions Window Help								
🌣 - 🐚 🖏 🗙 🌃 🍓 🗷 🗐 🙆	u 🕞 🖈 🗉 📜 👀 🛱 🗽 🎝 🤣								
movie     Data Sources     MOVIETEXTREMENS									
Diagrams		Sample Explore Modify Model Assess Utility HPUM Applications Text Mining Time Senes							
movie_go		Beg movie_go							
🕀 🖭 Model Packages									
Property	Value								
General									
Node ID	TextCluster2								
Imported Data									
Exported Data									
Notes		Text Parsing Text Filter	r _						
Train									
Variables									
Transform									
SVD Resolution	Low								
Max SVD Dimensions	100	To rule							
Exact or Maximum Number	Exact		Edit Variables						
Number of Clusters	2								
Cluster Algorithm	Expectation-Maximization		Undate						
Descriptive Terms	15		A Due						
Status			Kun						
Create Time	1/14/18 11:14 AM		Create Model Package						
Kun ID			Results						
Last Clature			Export Dath as EAS Drogram						
Last Status			Export Path as SAS Program						
Run Duration			Cut						
Crid Host			Cut						
User-Added Node	No		Сору						
			Delete						
			Rename						
			Select All						
			SCIELL AII						
A <b>T</b>			Select Nodes						
General			Connect Nodes						
			and a state of the						
General Properties			Disconnect Nodes						

Type in "Text Cluster (3 clusters)" and hit OK.

🕂 Enterprise Miner - movie					
File Edit View Actions Options Window Help					
🔅 🖻 🖏 🗙 🏭 🐜 🗷 🗐 🏩 🎧 🖈 💷 📜 😏 🗗 🎇 🏹 🤇	9				
movie     movie     MoVIETEXTREVIEWS     MoVIETEXTREVIEWS     MoVIETEXTREVIEWS     MoVIETEXTREVIEWS     MoVIE of Partnerse	Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series				
Property Value					
Property Value      General Node ID Imported Data Exported Data Exported Data Correct Data Exported Data Train Variables  Train Variables  Train Syst Resolution Low Max SVD Dimensions 100 Cluster Exact Cluster Low Cluster Algorithm Expectation-Maximization Descriptive Terms 15 Status Create Time 1/14/18 11:14 AM Run ID Last Status Last Status Last Status Last Status Last Status Last Algorithm Expectation Carl Maxime Last Status Last Status Last Algorithm Expectation Carl Maxime Last Status Last Algorithm Expectation Carl Maxime Last Algorithm Expectation Carl Maxime Last Algorithm Expectation E	Text Parsing Text Filter Text Cluster ()				

Change Number of Clusters from 2 to 3 (at left).



Connect *Text Filter* to *Text Cluster (3 clusters)* with an arrow.

🔣 Enterprise Miner - movie					
File Edit View Actions Options Window Help					
🔅 🗎 🖄 🗙 🎢 🍓 🖉 📋	] 🕞 * 🔳 📜 😏 🗗 🚉 🌏 🧇				
movie     MovieTeXTREVIEWS     MovieTeXTREVIEWS     MovieTeXTREVIEWS     MovieTeXTREVIEWS     MovieTeXTREVIEWS		Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series			
mouer Fackages					
Property	Value				
General Node ID Imported Data Exported Data Notes Train Uvariables Birtensform Max SVD Dimensions Cluster Exact or Maximum Number Number of Clusters Cluster Number of Clusters Cluster Algorithm Descriptive Terms Status Create Time Run ID Last Status Last Sta	TextCluster2	Text Parsing Text Filter			

Place the pointer on the new node and choose *Run* with right click.



The rest of the procedure should remain the same with the previous 2-cluster case. Lastly, in the same output folder,

(Your project target directory folder address) < movie < Workspaces < EMWS1

you will see the newly added data files with number 2 (with the second node operation this time) such as "textcluster2\_clusters" and "textcluster2\_train" this time.

