Simon Lizarazo
Simon13.

Homework Assignment 6 [30 pts]

STAT437 Unsupervised Learning – Fall 2023

Due: Friday, October 6 on Canvas at 11:59pm CST.

Problems	Points
1.1	0.25
1.2	0.25
1.3	1
1.4	1
2.1	1.5
2.2	2
2.3	2
2.4	1.5
3.1	1.5
3.2	1.5
3.3	1.5
3.4	0.5
3.5	0.75
3.6.1.	1
3.6.2	0.5
3.6.3	1
3.6.4	0.5
4.1.1	0.75
4.1.2	0.75
4.2.1	0.75
4.2.2	0.75
4.3.1	0.75
4.3.2	1
4.4	1
5	3
6	3

Questions 1-4: See Jupyter notebook

Question 5: The dataset below is comprised of 10 professors in the UIUC Statistics department. Suppose we'd like to cluster this dataset using k-modes with k=2 clusters. The *current* cluster modes are shown below. What would be the *new cluster modes* found in the next iteration of the k-modes clustering algorithm. Show your work.

- *Hint: If there's a tie in the cluster assignment step, assign the person to the mode with the lowest index.*
- *Hint: If there's a tie in the centroid/mode update step, select the attribute value with the highest alphabetical order (ie. A>B).*

	PhD	Sex	Generation
Tori Ellison	Operations Research	Female	millennial
Karle Flanagan	Statistics Education	Female	millennial
Kelly Findley	Statistics Education	Male	millennial
Julie Deeke	Statistics	Female	millennial
Chris Kinson	Statistics	Male	millennial
Jeff Douglas	Statistics	Male	Gen X
Bo Li	Statistics	Female	Gen X
Steve Culpepper	Educational Psychology	Male	Gen X
Dave Zhao	Statistics	Male	millennial
Vimal Rao	Educational Psychology	Male	millennial

Hamming Distance

Mode 1	Mode 2	Mode
2	1	Mode 2
2	0	Mode 2
1	1 (TIE)	Mode 1
1	1 (TIE)	Mode 1
0	2	Mode 1
1	3	Mode 1
2	2 (TIE)	Mode 1
2	3	Mode 1
0	2	Mode 1
1	2	Mode 1

	PhD	Sex	Generation
Mode 1	Statistics	Male	millennial
Mode 2	Statistics Education	Female	millennial

New Mode

Mode 1 — Statistics — Male — Millenial

Mode 2 — Operational Research — Female — Millenial

Note: For mode 2, TIE in PhD.

Question 6: In the page below 4 categorical datasets are listed. Each of these datasets was put into a Hamming distance matrix. Each hamming distance matrix was then used as input into the t-SNE algorithm producing the following 4 sets of t-SNE plots shown in the pages below.

Match each of the 4 datasets 1-4 to the corresponding t-SNE plot sets A-D.

Hint: Some points may be completely overlapping in some of the t-SNE plots below.

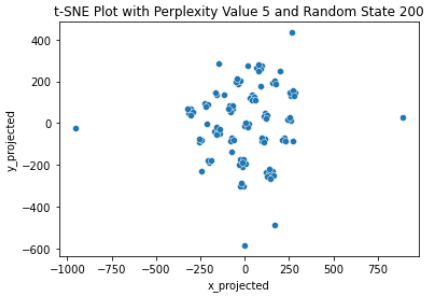
Explanations not required, but may help for partial credit if you get it wrong.

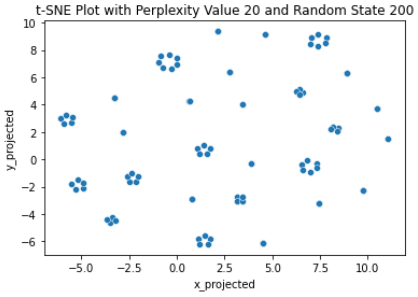
Dataset 1			Dataset 2			Dataset 3			Dataset 4		
pet	fav jonas bro	is a hotdog a sandwich?	pet	fav jonas bro	is a hotdog a sandwich?	pet	fav jonas bro	is a hotdog a sandwich?	pet	fav jonas bro	is a hotdog a sandwich?
cat	nick	yes	cat	nick	no	cat	joe	no	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	no	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	no	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	no	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	no	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	not_sure	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	not_sure	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	not_sure	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	yes	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	yes	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	joe	yes	cat	nick	yes
cat	nick	yes	cat	nick	no	cat	kevin	no	cat	nick	no
cat	nick	yes	cat	nick	no	cat	kevin	no	cat	nick	no
cat	nick	yes	cat	nick	no	cat	kevin	not_sure	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	not_sure	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	not_sure	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	yes	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	yes	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	yes	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	yes	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	yes	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	kevin	yes	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	nick	no	cat	nick	no
cat	nick	yes	cat	nick	not_sure	cat	nick	no	cat	nick	not_sure
cat	nick	yes	cat	nick	not_sure	cat	nick	no	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	no	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	not_sure	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	not_sure	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	not_sure	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	not_sure	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	not_sure	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	yes	cat	nick	not_sure
cat	nick	yes	cat	nick	yes	cat	nick	yes	cat	nick	not_sure
dog	kevin	no	dog	kevin	no	dog	joe	no	dog	kevin	no
dog	kevin	no	dog	kevin	no	dog	joe	no	dog	kevin	no
dog	kevin	no	dog	kevin	no	dog	joe	no	dog	kevin	no
dog	kevin	no	dog	kevin	no	dog	joe	no	dog	kevin	no
dog	kevin	no	dog	kevin	no	dog	joe	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	no	dog	joe	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	no	dog	joe	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	joe	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	joe	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	joe	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	joe	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	joe	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	no	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	no	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	yes	dog	kevin	no
dog	kevin	no	dog	kevin	not_sure	dog	kevin	yes	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	kevin	yes	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	no	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	no	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	no	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	no	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	no	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	not_sure	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	yes	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	yes	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	yes	dog	kevin	no
dog	kevin	no	dog	kevin	yes	dog	nick	yes	dog	kevin	no
fish	joe	not_sure	fish	joe	no	fish	joe	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	joe	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	no	fish	kevin	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	not_sure	fish	kevin	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	kevin	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	kevin	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	no	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	not_sure	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	yes	fish	joe	not_sure
fish	joe	not_sure	fish	joe	yes	fish	nick	yes	fish	joe	not_sure

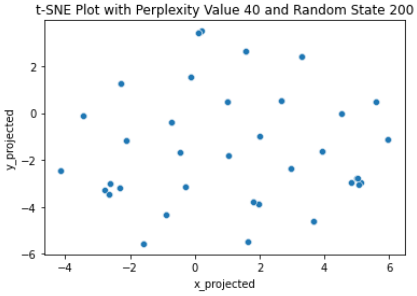
SUMMARY

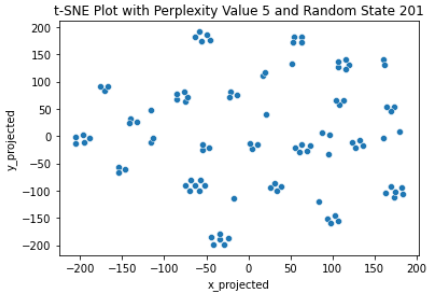
DATASET	PLOTSET
1	B
2	C
3	A
4	D.

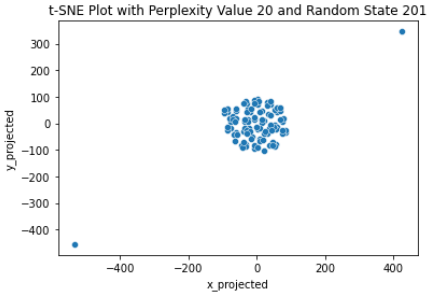
Plot Set A DATASET 3

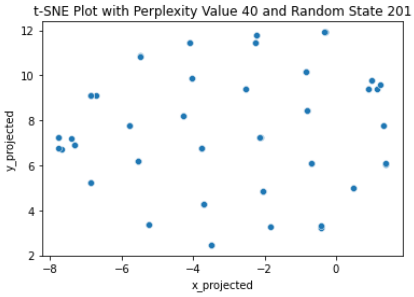


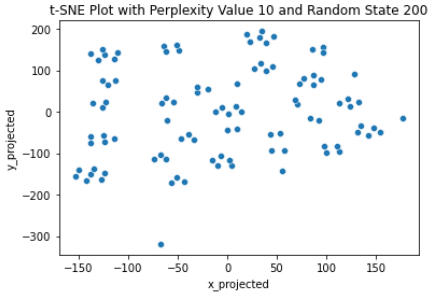


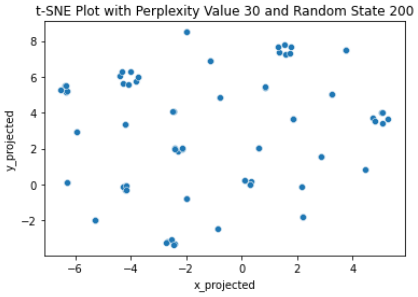


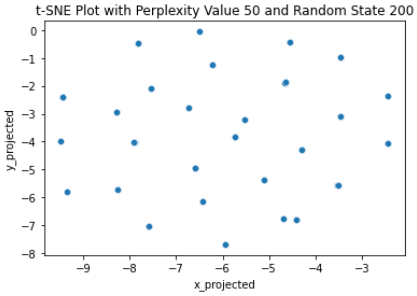


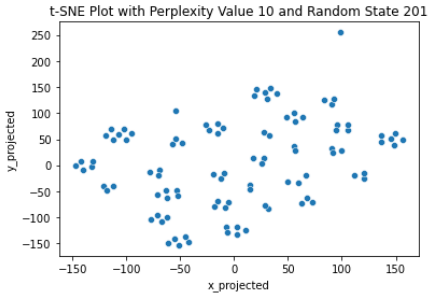


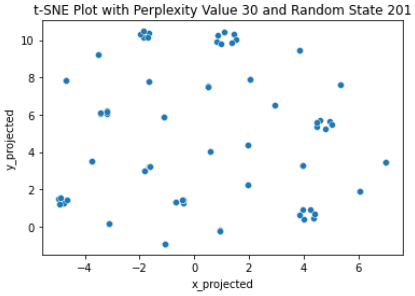


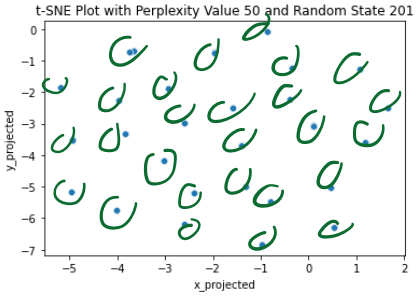




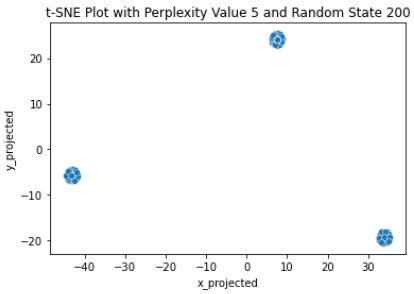


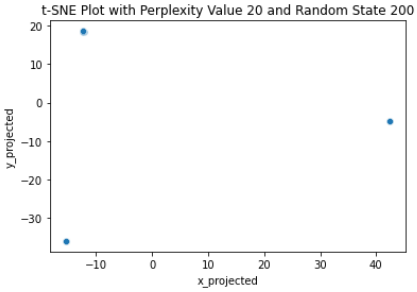


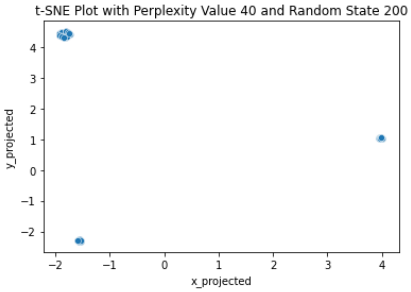


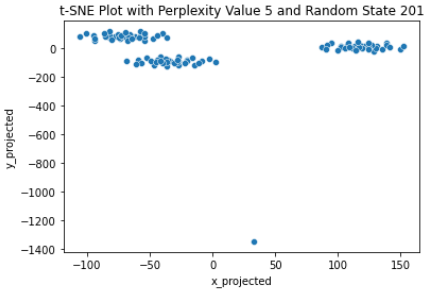


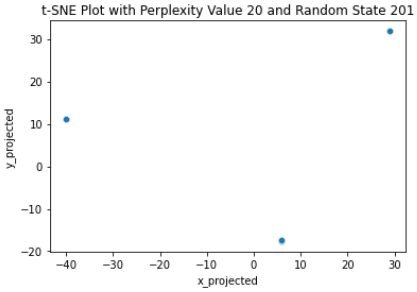
Plot Set B DATASET 1.

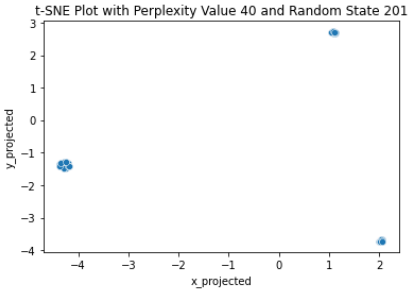


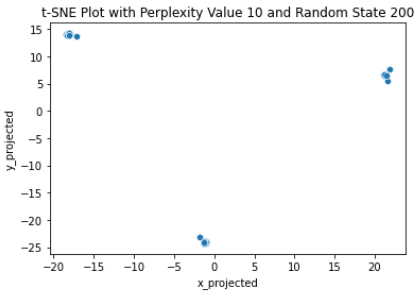


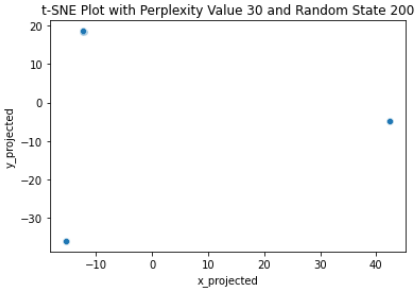


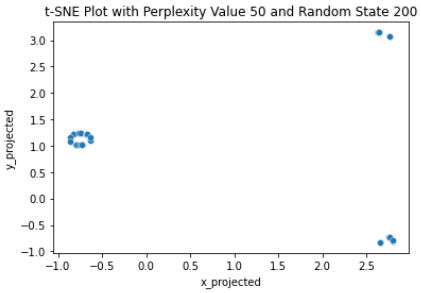


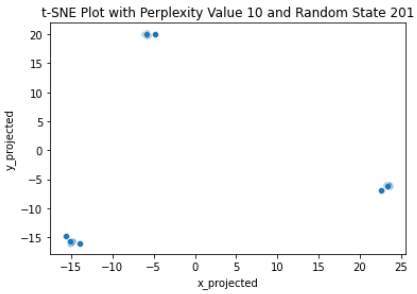


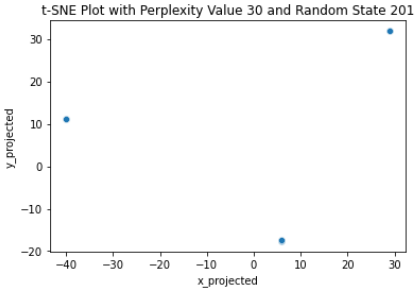


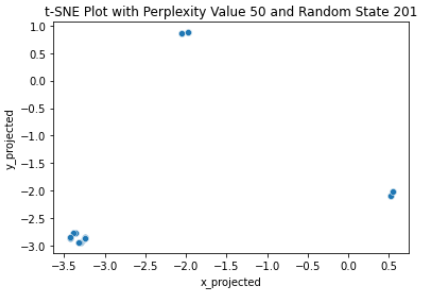




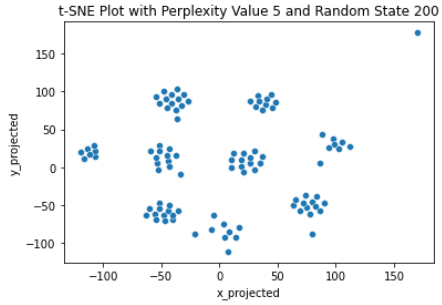


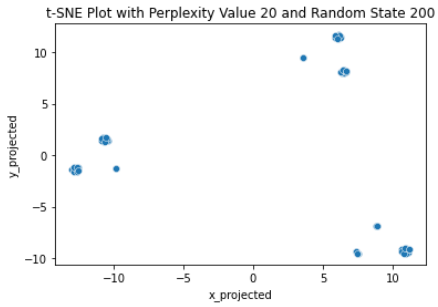


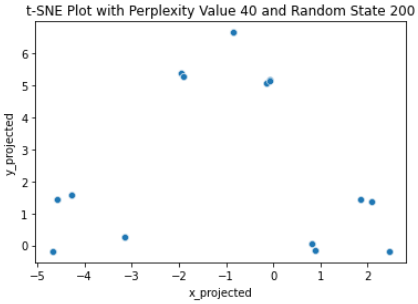


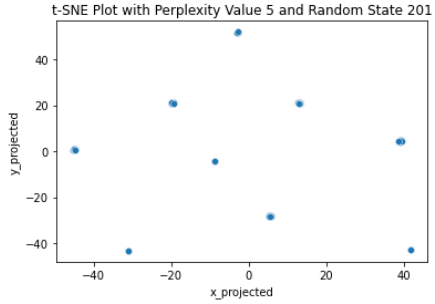


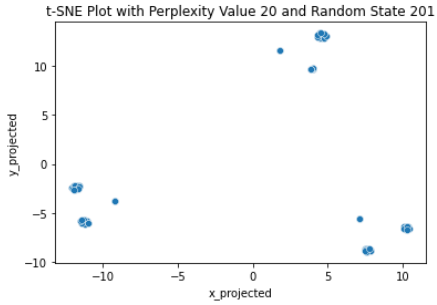
Plot Set C

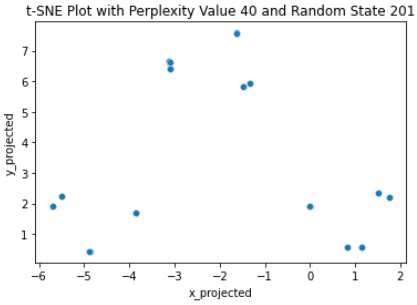

t-SNE Plot with Perplexity Value 5 and Random State 200
y_projected
x_projected

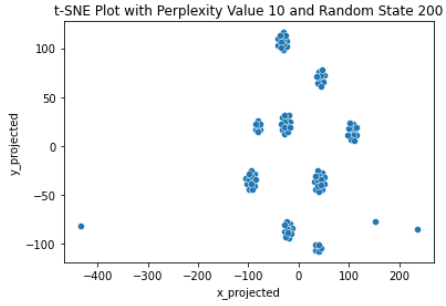

t-SNE Plot with Perplexity Value 20 and Random State 200
y_projected
x_projected

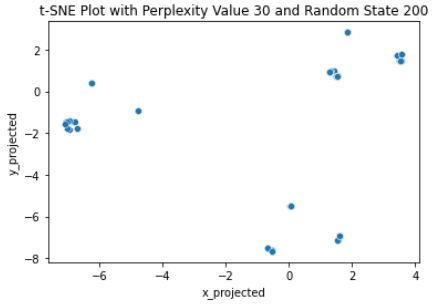

t-SNE Plot with Perplexity Value 40 and Random State 200
y_projected
x_projected

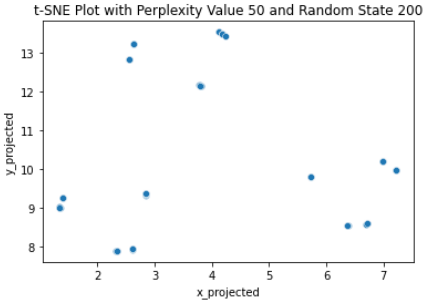

t-SNE Plot with Perplexity Value 5 and Random State 201
y_projected
x_projected

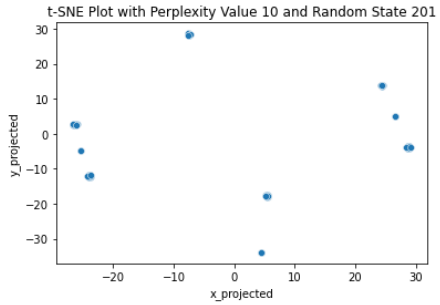

t-SNE Plot with Perplexity Value 20 and Random State 201
y_projected
x_projected

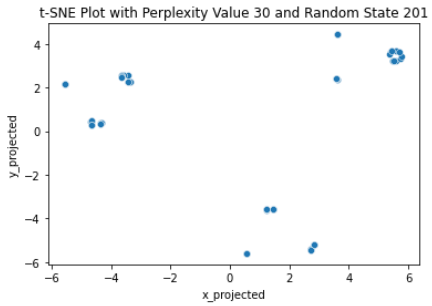

t-SNE Plot with Perplexity Value 40 and Random State 201
y_projected
x_projected

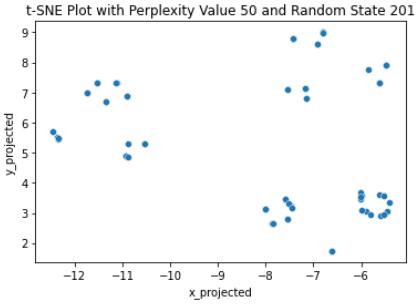

t-SNE Plot with Perplexity Value 10 and Random State 200
y_projected
x_projected


t-SNE Plot with Perplexity Value 30 and Random State 200
y_projected
x_projected


t-SNE Plot with Perplexity Value 50 and Random State 200
y_projected
x_projected


t-SNE Plot with Perplexity Value 10 and Random State 201
y_projected
x_projected


t-SNE Plot with Perplexity Value 30 and Random State 201
y_projected
x_projected


t-SNE Plot with Perplexity Value 50 and Random State 201
y_projected
x_projected

Plot Set D DATASET 4