Muhammad Hashaam Shahid 2491488

Part 1

Q Table for KNN statistics

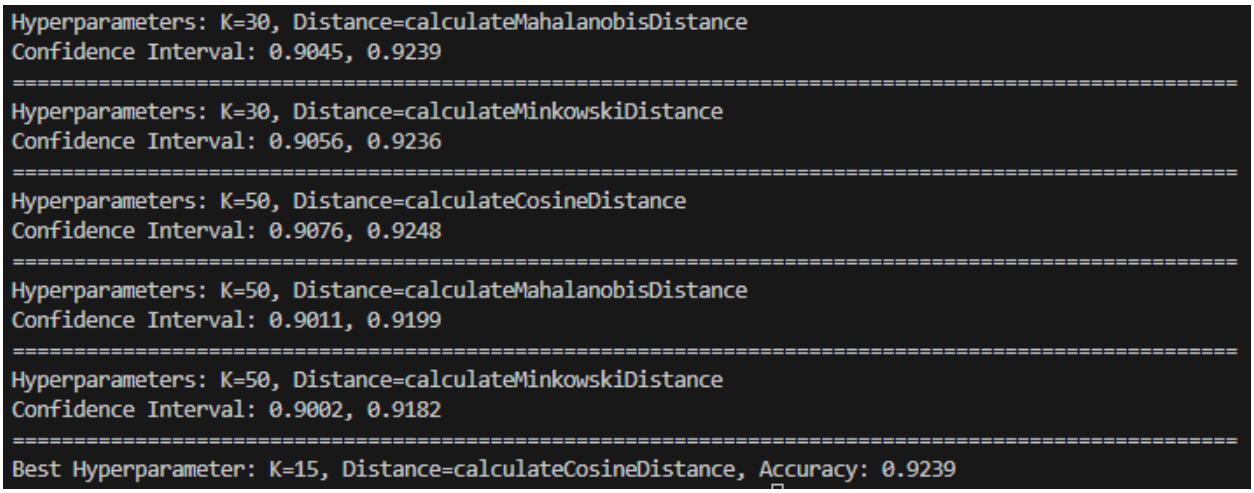
K	Distance	Confidence Interval
1	Cosine	0.9133, 0.9133
1	Mahalanobis	0.8917, 0.9083
1	Minkowski	0.9014, 0.9208
5	Cosine	0.9092, 0.9274
5	Mahalanobis	0.9068, 0.9225
5	Minkowski	0.9108, 0.9248
10	Cosine	0.9156, 0.9301
10	Mahalanobis	0.9085, 0.9248
10	Minkowski	0.9128, 0.9287
15	Cosine	0.9163, 0.9315
15	Mahalanobis	0.9068, 0.9257
15	Minkowski	0.9103, 0.9284
30	Cosine	0.9125, 0.9294
30	Mahalanobis	0.9045, 0.9239

Muhammad Hashaam Shahid 2491488

30	Minkowski	0.9056, 0.9236
50	Cosine	0.9076, 0.9248
50	Mahalanobis	0.9011, 0.9199
50	Minkowski	0.9002, 0.9182

```
Hyperparameters: K=1, Distance=calculateCosineDistance
Confidence Interval: 0.9133, 0.9133
================================================================================
Hyperparameters: K=1, Distance=calculateMahalanobisDistance
Confidence Interval: 0.8917, 0.9083
================================================================================
Hyperparameters: K=1, Distance=calculateMinkowskiDistance
Confidence Interval: 0.9014, 0.9208
================================================================================
Hyperparameters: K=5, Distance=calculateCosineDistance
Confidence Interval: 0.9092, 0.9274
================================================================================
Hyperparameters: K=5, Distance=calculateMahalanobisDistance
Confidence Interval: 0.9068, 0.9225
================================================================================
Hyperparameters: K=5, Distance=calculateMinkowskiDistance
Confidence Interval: 0.9108, 0.9248
================================================================================
Hyperparameters: K=10, Distance=calculateCosineDistance
Confidence Interval: 0.9156, 0.9301
================================================================================
Hyperparameters: K=10, Distance=calculateMahalanobisDistance
Confidence Interval: 0.9085, 0.9248
================================================================================
Hyperparameters: K=10, Distance=calculateMinkowskiDistance
Confidence Interval: 0.9128, 0.9287
================================================================================
Hyperparameters: K=15, Distance=calculateCosineDistance
Confidence Interval: 0.9163, 0.9315
================================================================================
Hyperparameters: K=15, Distance=calculateMahalanobisDistance
Confidence Interval: 0.9068, 0.9257
================================================================================
Hyperparameters: K=15, Distance=calculateMinkowskiDistance
Confidence Interval: 0.9103, 0.9284
================================================================================
Hyperparameters: K=30, Distance=calculateCosineDistance
Confidence Interval: 0.9125, 0.9294
================================================================================
```

Muhammad Hashaam Shahid 2491488



Best Hyperparameter=15, Distance=cosine distance

//my vals might be wrong due to me taking too many avgs and then //taking confidence interval kindly check the confidence interval out

Q. In addition, please add some comments on how you have picked the best-performing hyperparameter values.

I used grid search to check for the best hyperparameter. Choose these values of K as it was written in the doc to use these K values (5,10,30) also used 1 15 50 as they are odd numbers except 50 which was taken randomly.

Muhammad Hashaam Shahid 2491488

Part 2

Q. Table for K means Statistics.

Dataset	K	Avg Time for each iteration (seconds)	Confidence Interval
1	2	0.1450745391845703	162.10629413850302, 162.10629413850302
1	3	0.19508790969848633	94.03403416666055, 105.43633217856143
1	4	0.2182069516181946	51.15293305705194, 51.15293305705201
1	5	0.2878965592384338	23.37061388522963, 23.370613885229645
1	6	0.4707791042327881	22.437489783137792, 22.49221516222954
1	7	0.6453157329559327	21.609008567166153, 21.707750877652163
1	8	0.7662527489662171	20.853460953087186, 20.944733368513134
1	9	0.9712622976303102	20.049844777264976, 20.210751461867613
1	10	1.0783032941818238	19.403538520048688, 19.540546657048463
2	2	0.07326421976089478	153.04269642371062, 153.04269642371062

Muhammad Hashaam Shahid 2491488

2	3	0.11963428020477296	22.498196679988382, 22.498196679988396
2	4	0.2536673140525818	13.58659308541042, 13.586593085410428
2	5	0.4083128833770752	11.41605752088716, 11.469522383929116
2	6	0.5927012801170349	9.340166427616774, 9.4658237700512
2	7	0.7909860801696778	7.611395770242431, 8.392713947241056
2	8	0.9754738140106202	6.686070107788327, 7.427350396693528
2	9	0.9729329013824464	6.138032178269513, 6.187340312234693
2	10	1.1465628266334533	5.5674568919171525, 5.670579823413667

Muhammad Hashaam Shahid 2491488

```
Dataset: 1 - K: 2
Avg of Loss: 162.10629413850302
Confidence Interval: (162.10629413850302, 162.10629413850302)
Avg Running Time: 0.1450745391845703 seconds
------------------------------------------------------------
Dataset: 1 - K: 3
Avg of Loss: 99.73518317261099
Confidence Interval: (94.03403416666055, 105.43633217856143)
Avg Running Time: 0.19508790969848633 seconds
------------------------------------------------------------
Dataset: 1 - K: 4
Avg of Loss: 51.152933057051975
Confidence Interval: (51.15293305705194, 51.15293305705201)
Avg Running Time: 0.2182069516181946 seconds
------------------------------------------------------------
Dataset: 1 - K: 5
Avg of Loss: 23.370613885229638
Confidence Interval: (23.37061388522963, 23.370613885229645)
Avg Running Time: 0.2878965592384338 seconds
------------------------------------------------------------
Dataset: 1 - K: 6
Avg of Loss: 22.464852472683667
Confidence Interval: (22.437489783137792, 22.49221516222954)
Avg Running Time: 0.4707791042327881 seconds
------------------------------------------------------------
Dataset: 1 - K: 7
Avg of Loss: 21.65837972240916
Confidence Interval: (21.609008567166153, 21.707750877652163)
Avg Running Time: 0.6453157329559327 seconds
------------------------------------------------------------
Dataset: 1 - K: 8
Avg of Loss: 20.89909716080016
Confidence Interval: (20.853460953087186, 20.944733368513134)
Avg Running Time: 0.7662527489662171 seconds
------------------------------------------------------------
Dataset: 1 - K: 9
Avg of Loss: 20.130298119566294
Confidence Interval: (20.049844777264976, 20.210751461867613)
Avg Running Time: 0.9712622976303102 seconds
------------------------------------------------------------
Dataset: 1 - K: 10
Avg of Loss: 19.472042588548575
Confidence Interval: (19.403538520048688, 19.540546657048463)
Avg Running Time: 1.0783032941818238 seconds
------------------------------------------------------------
```

Muhammad Hashaam Shahid 2491488

```
Dataset: 2 - K: 2
Avg of Loss: 153.04269642371062
Confidence Interval: (153.04269642371062, 153.04269642371062)
Avg Running Time: 0.07326421976089478 seconds
------------------------------------------------------------
Dataset: 2 - K: 3
Avg of Loss: 22.49819667998839
Confidence Interval: (22.498196679988382, 22.498196679988396)
Avg Running Time: 0.11963428020477296 seconds
------------------------------------------------------------
Dataset: 2 - K: 4
Avg of Loss: 13.586593085410424
Confidence Interval: (13.58659308541042, 13.586593085410428)
Avg Running Time: 0.2536673140525818 seconds
------------------------------------------------------------
Dataset: 2 - K: 5
Avg of Loss: 11.442789952408138
Confidence Interval: (11.41605752088716, 11.469522383929116)
Avg Running Time: 0.4083128833770752 seconds
------------------------------------------------------------
Dataset: 2 - K: 6
Avg of Loss: 9.402995098833987
Confidence Interval: (9.340166427616774, 9.4658237700512)
Avg Running Time: 0.5927012801170349 seconds
------------------------------------------------------------
Dataset: 2 - K: 7
Avg of Loss: 8.002054858741744
Avg Running Time: 0.7909860801696778 seconds
------------------------------------------------------------
Dataset: 2 - K: 8
Avg of Loss: 7.0567102522409275
Confidence Interval: (6.686070107788327, 7.427350396693528)
Avg Running Time: 0.9754738140106202 seconds
------------------------------------------------------------
Dataset: 2 - K: 9
Avg of Loss: 6.162686245252103
Confidence Interval: (6.138032178269513, 6.187340312234693)
Avg Running Time: 0.9729329013824464 seconds
------------------------------------------------------------
Dataset: 2 - K: 10
Avg of Loss: 5.61901835766541
Confidence Interval: (5.5674568919171525, 5.670579823413667)
Avg Running Time: 1.1465628266334533 seconds
------------------------------------------------------------
```

Muhammad Hashaam Shahid 2491488

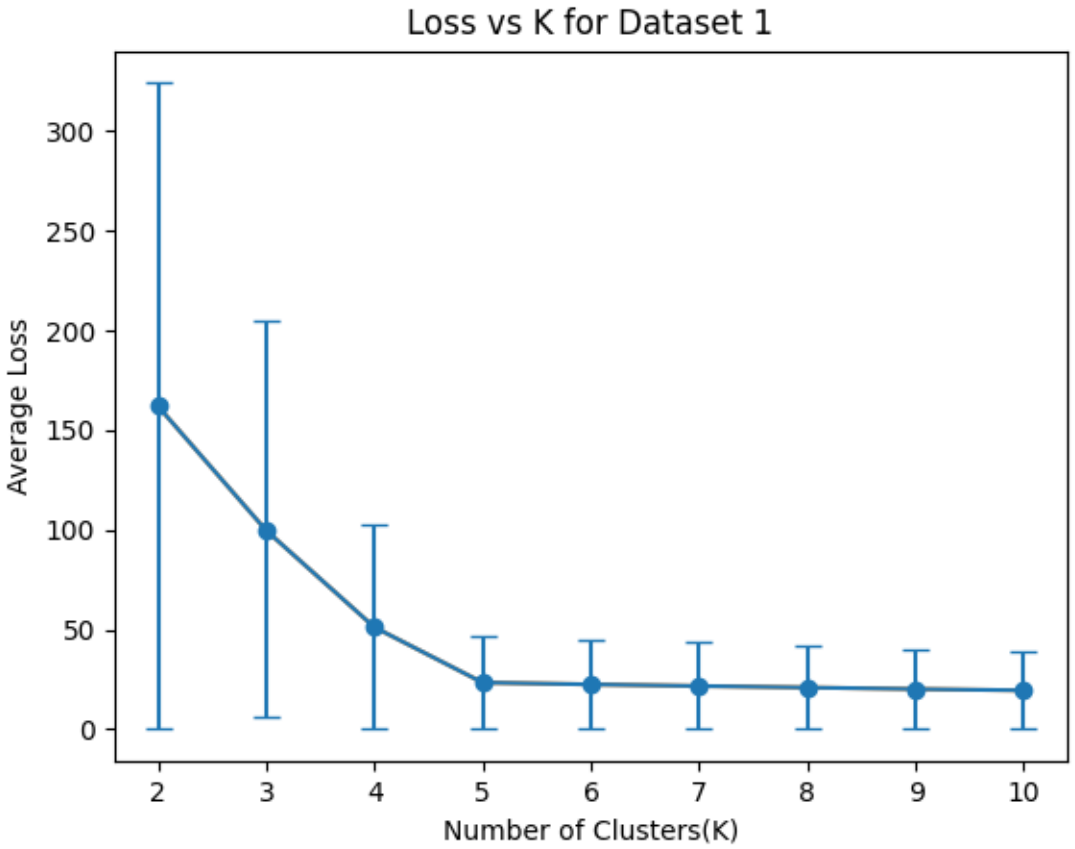


Figure 1 Kmean elbow for dataset 1

Muhammad Hashaam Shahid 2491488

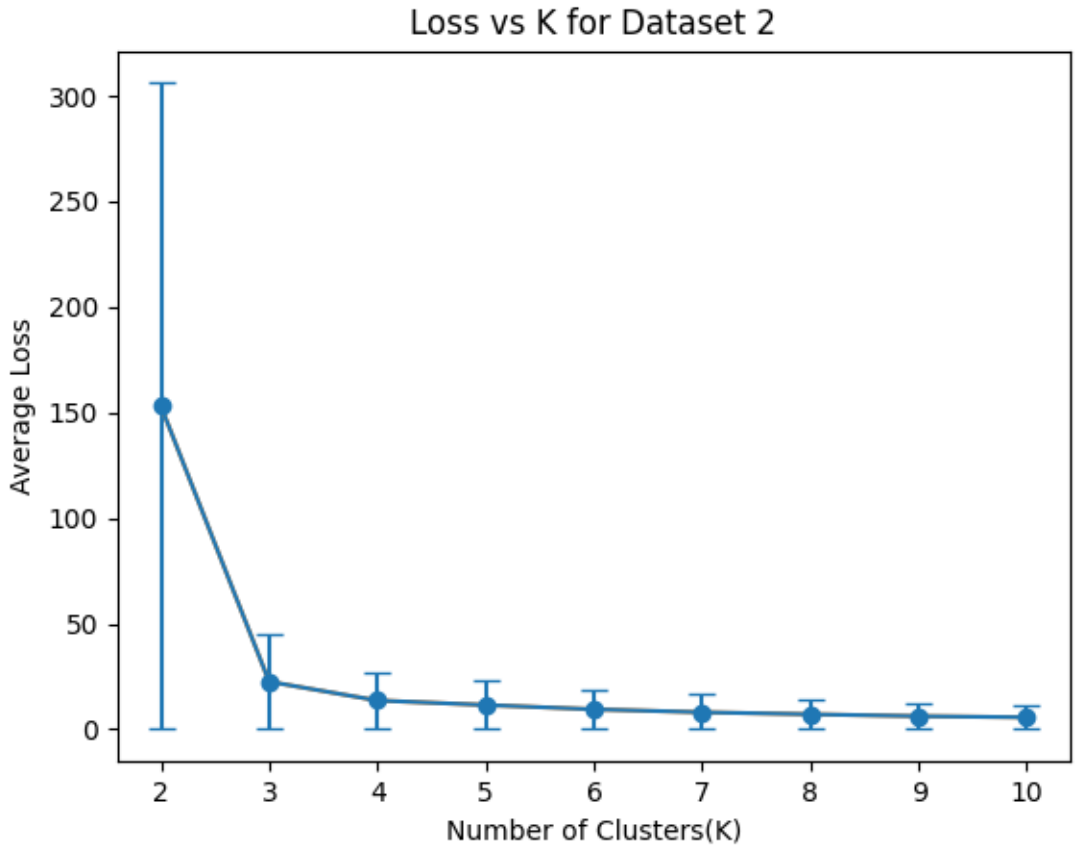


Figure 2 kmean elbow for dataset 2

Muhammad Hashaam Shahid 2491488

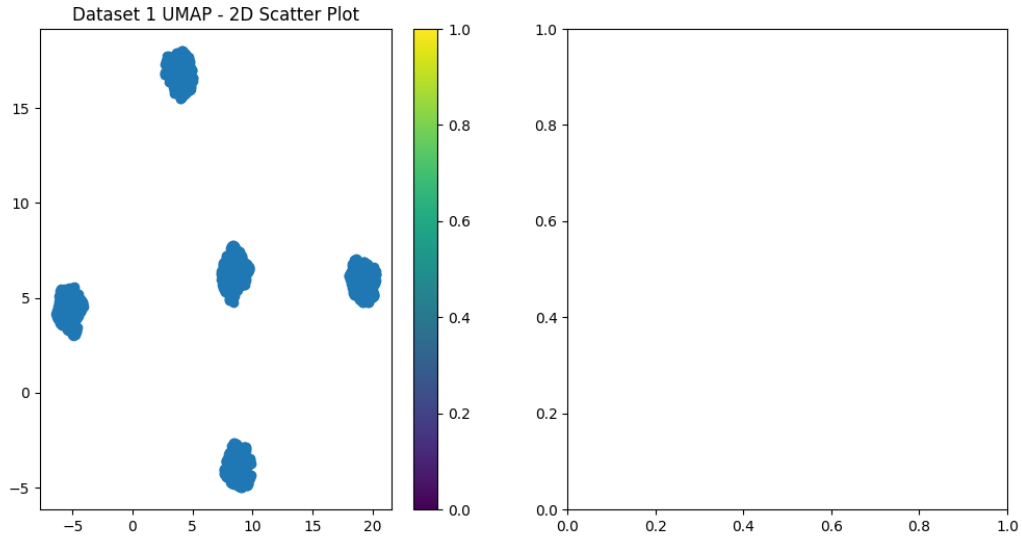


Figure 3 UMAP 2D scatter plot for dataset 1

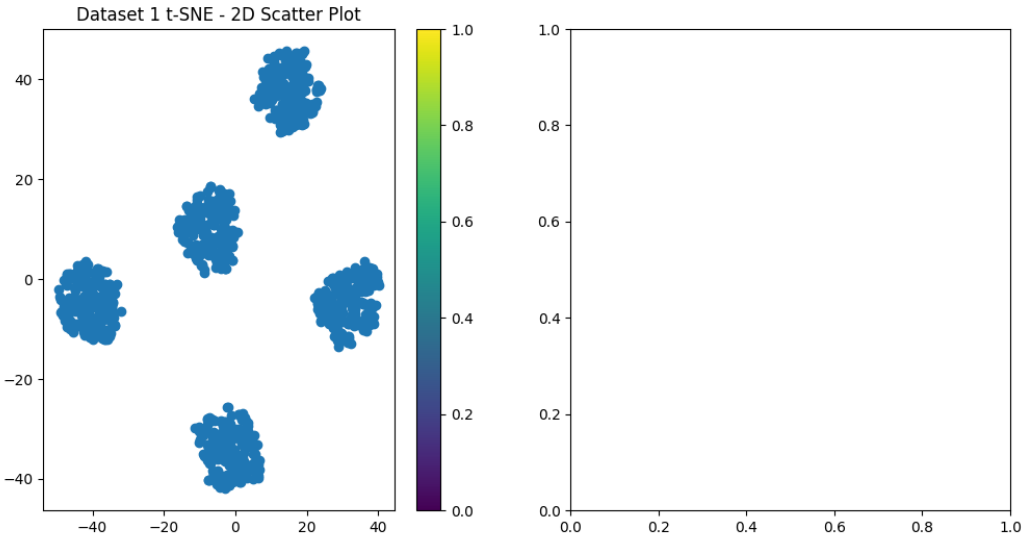


Figure 4 t-SNE 2D scatter plot for dataset 1

Muhammad Hashaam Shahid 2491488

Q. Table for K Medoid Statistics.

Dataset	K	Avg Time for each iteration (seconds)	Confidence Interval
1	2	5.88341675043106	314.31949469156496, 321.3132253775915
1	3	4.407943308353423	170.62010092107587, 205.82641046166177
1	4	4.002643322944642	89.10842189370602, 125.76086093219631
1	5	3.416951887607574	48.98948805206925, 78.07801461941094
1	6	2.620533857345581	41.60621342629213, 57.66935218229871
1	7	2.512315654754638	37.837347402027774, 40.35209262187131
1	8	2.3848349142074583	35.54675738084155, 37.286898274542516
1	9	2.1305950427055356	34.09753892710093, 34.86257548043844
1	10	1.9214842343330385	32.70714688243899, 34.1308850770947
2	2	4.484040660858154	3.5535468459129333, 3.5535468459129333
2	3	3.531221942901611	1.7179941763945688, 2.1984366784027944

Muhammad Hashaam Shahid 2491488

2	4	2.507490067481995	1.3472979264838678, 1.5208071512596624
2	5	1.8384440875053403	1.2037555068209558, 1.2431292802617164
2	6	2.0417592668533326	1.120721139237253, 1.182439480498465
2	7	1.7096552586555478	1.043219544180253, 1.0789433820167988
2	8	1.641682794094086	0.9785919148259836, 1.0743265073961537
2	9	1.3980965638160705	0.9287897939324123, 0.9788388257384555
2	10	1.1951286196708681	0.8964296910342137, 0.9651305463734706

Muhammad Hashaam Shahid 2491488

```
Dataset: 1 - K: 2
Avg of Loss: 317.8163600345782
Confidence Interval: (314.31949469156496, 321.3132253775915)
Avg Running Time: 5.88341675043106 seconds
------------------------------------------------------------
Dataset: 1 - K: 3
Avg of Loss: 188.22325569136882
Confidence Interval: (170.62010092107587, 205.82641046166177)
Avg Running Time: 4.407943308353423 seconds
------------------------------------------------------------
Dataset: 1 - K: 4
Avg of Loss: 107.43464141295117
Confidence Interval: (89.10842189370602, 125.76086093219631)
Avg Running Time: 4.002643322944642 seconds
------------------------------------------------------------
Dataset: 1 - K: 5
Avg of Loss: 63.53375133574009
Confidence Interval: (48.98948805206925, 78.07801461941094)
Avg Running Time: 3.416951887607574 seconds
------------------------------------------------------------
Dataset: 1 - K: 6
Avg of Loss: 49.63778280429542
Confidence Interval: (41.60621342629213, 57.66935218229871)
Avg Running Time: 2.620533857345581 seconds
------------------------------------------------------------
Dataset: 1 - K: 7
Avg of Loss: 39.09472001194954
Confidence Interval: (37.837347402027774, 40.35209262187131)
Avg Running Time: 2.512315654754638 seconds
------------------------------------------------------------
Dataset: 1 - K: 8
Avg of Loss: 36.41682782769203
Confidence Interval: (35.54675738084155, 37.286898274542516)
Avg Running Time: 2.3848349142074583 seconds
------------------------------------------------------------
Dataset: 1 - K: 9
Avg of Loss: 34.48005720376968
Confidence Interval: (34.09753892710093, 34.86257548043844)
Avg Running Time: 2.1305950427055356 seconds
------------------------------------------------------------
Dataset: 1 - K: 10
Avg of Loss: 33.41901597976685
Confidence Interval: (32.70714688243899, 34.1308850770947)
Avg Running Time: 1.9214842343330385 seconds
------------------------------------------------------------
```

Muhammad Hashaam Shahid 2491488

```
Dataset: 2 - K: 2
Avg of Loss: 3.5535468459129333
Confidence Interval: (3.5535468459129333, 3.5535468459129333)
Avg Running Time: 4.484040660858154 seconds
------------------------------------------------------------
Dataset: 2 - K: 3
Avg of Loss: 1.9582154273986816
Confidence Interval: (1.7179941763945688, 2.1984366784027944)
Avg Running Time: 3.531221942901611 seconds
------------------------------------------------------------
Dataset: 2 - K: 4
Avg of Loss: 1.434052538871765
Confidence Interval: (1.3472979264838678, 1.5208071512596624)
Avg Running Time: 2.507490067481995 seconds
------------------------------------------------------------
Dataset: 2 - K: 5
Avg of Loss: 1.223442393541336
Confidence Interval: (1.2037555068209558, 1.2431292802617164)
Avg Running Time: 1.8384440875053403 seconds
------------------------------------------------------------
Dataset: 2 - K: 6
Avg of Loss: 1.151580309867859
Confidence Interval: (1.120721139237253, 1.182439480498465)
Avg Running Time: 2.0417592668533326 seconds
------------------------------------------------------------
Dataset: 2 - K: 7
Avg of Loss: 1.061081463098526
Confidence Interval: (1.043219544180253, 1.0789433820167988)
Avg Running Time: 1.7096552586555478 seconds
------------------------------------------------------------
Dataset: 2 - K: 8
Avg of Loss: 1.0264592111110686
Confidence Interval: (0.9785919148259836, 1.0743265073961537)
Avg Running Time: 1.641682794094086 seconds
------------------------------------------------------------
Dataset: 2 - K: 9
Avg of Loss: 0.9538143098354339
Confidence Interval: (0.9287897939324123, 0.9788388257384555)
Avg Running Time: 1.3980965638160705 seconds
------------------------------------------------------------
Dataset: 2 - K: 10
Avg of Loss: 0.9307801187038421
Confidence Interval: (0.8964296910342137, 0.9651305463734706)
Avg Running Time: 1.1951286196708681 seconds
------------------------------------------------------------
```

Muhammad Hashaam Shahid 2491488

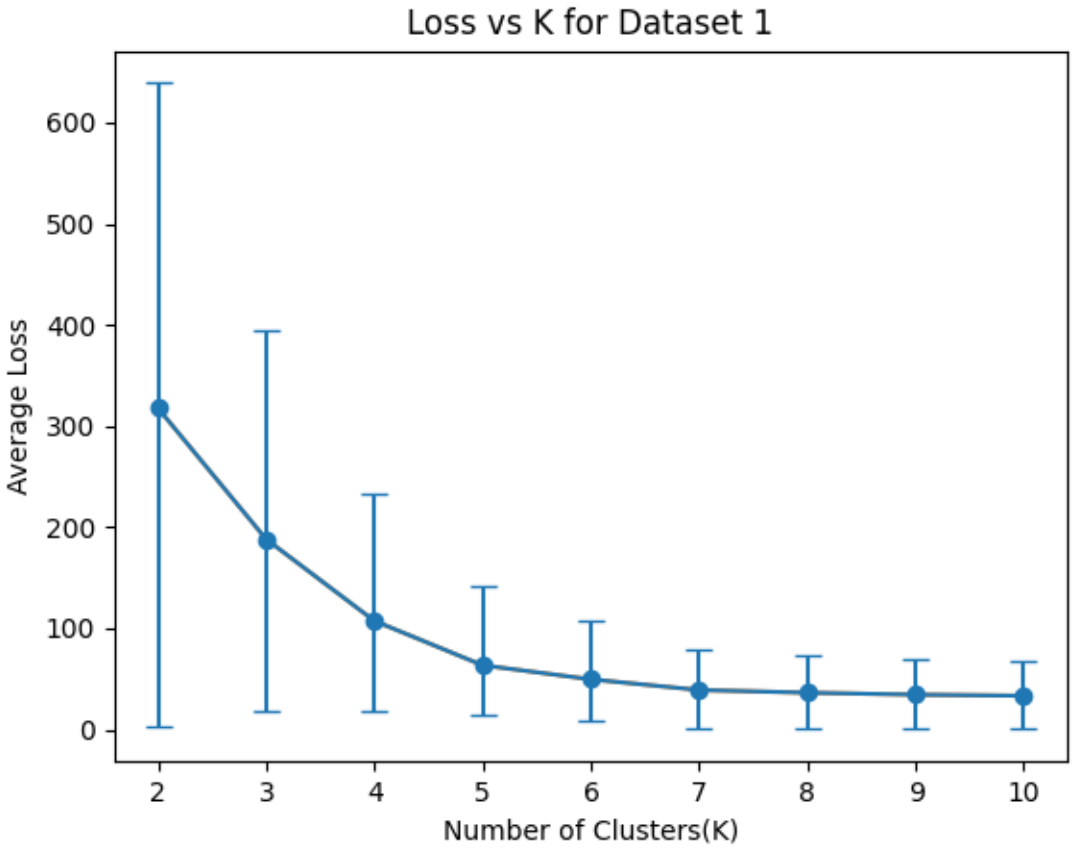


Figure 5 K medoid elbow for dataset 1

Muhammad Hashaam Shahid 2491488

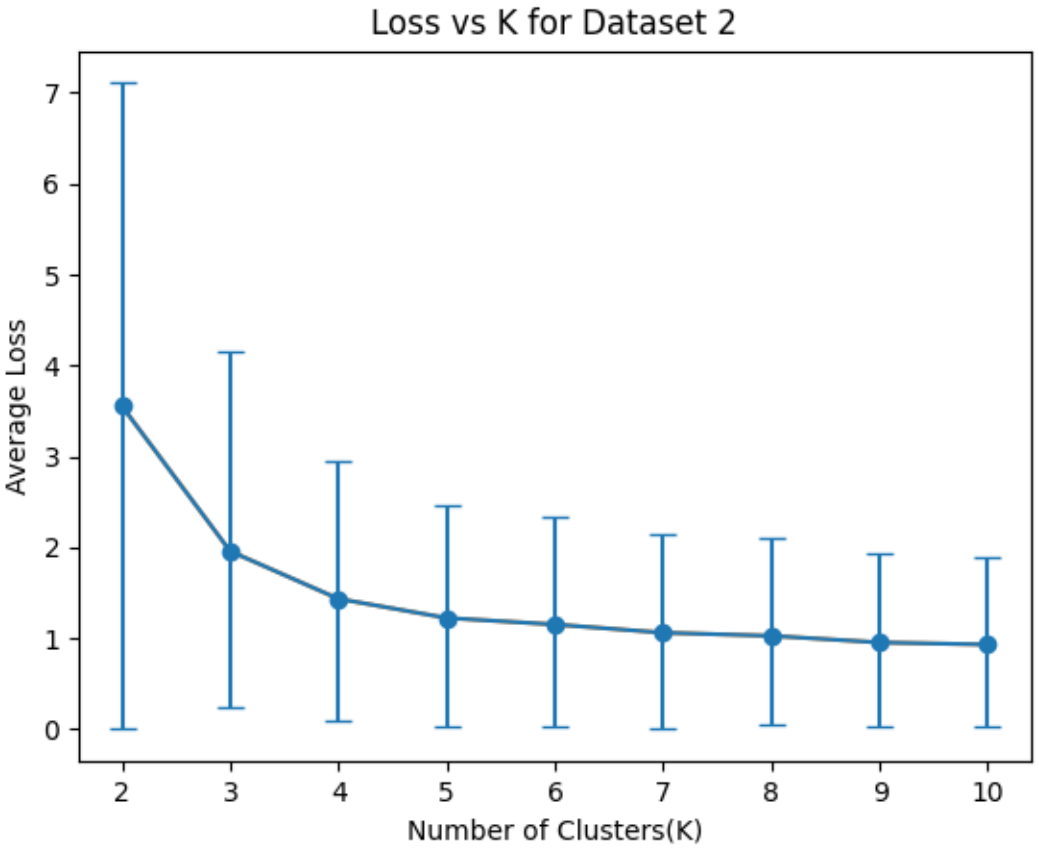


Figure 6 K Medoid elbow for dataset 2

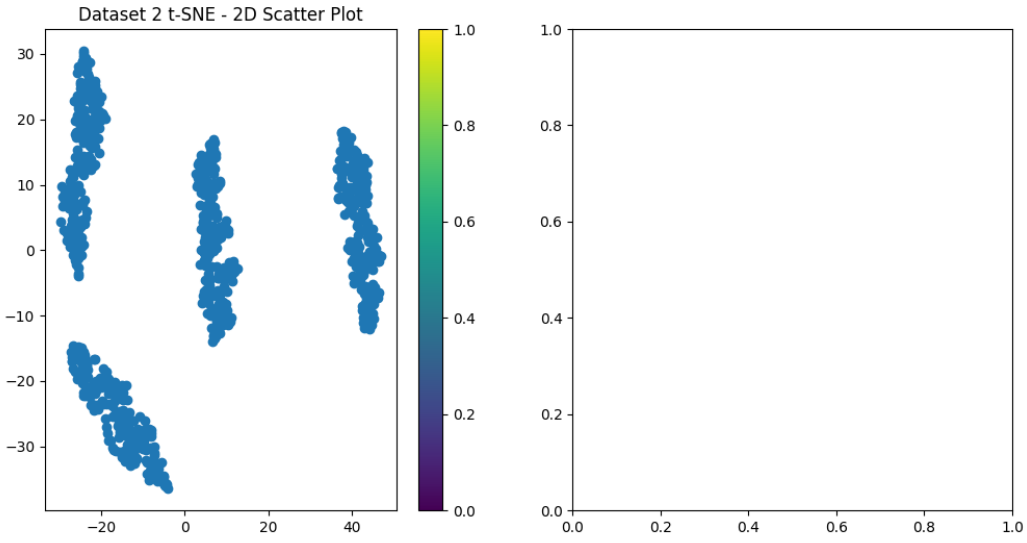


Figure 7 t-SNE 2D scatter plot for dataset 2

Muhammad Hashaam Shahid 2491488

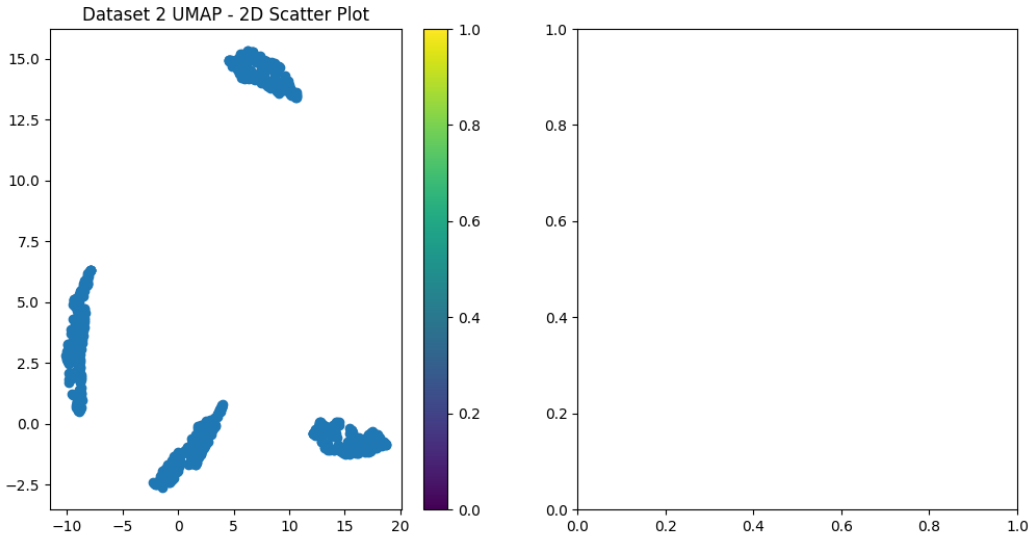


Figure 8 UMAP 2D scatter Plot for Dataset 2

Q. Also, please briefly discuss whether the number of clusters you have identified via the elbow method matches the number of clusters in visualizations for both datasets separately.

For dataset 1 clusters are easily recognizable using elbow method which is giving k=5 however for dataset 2 elbow method is making elbow at k=3 however meaningful clusters are 4.

Muhammad Hashaam Shahid 2491488

Part 3

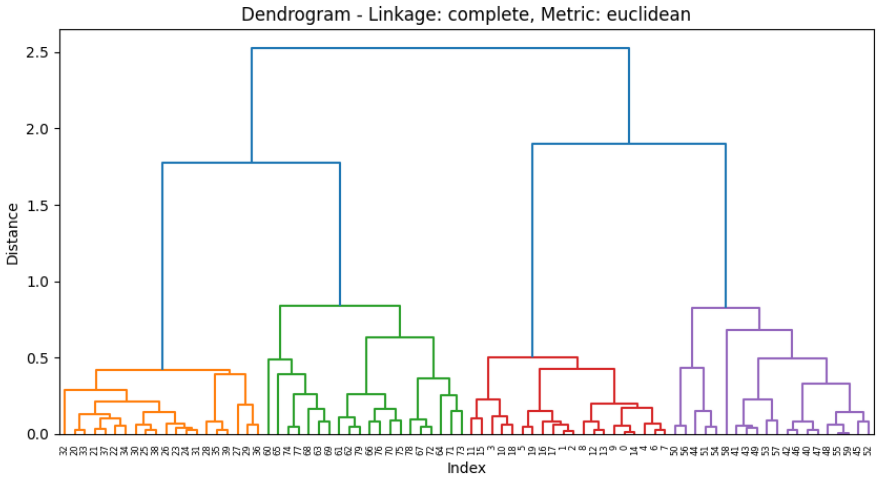


Figure 9 dendrogram Complete Euclidean

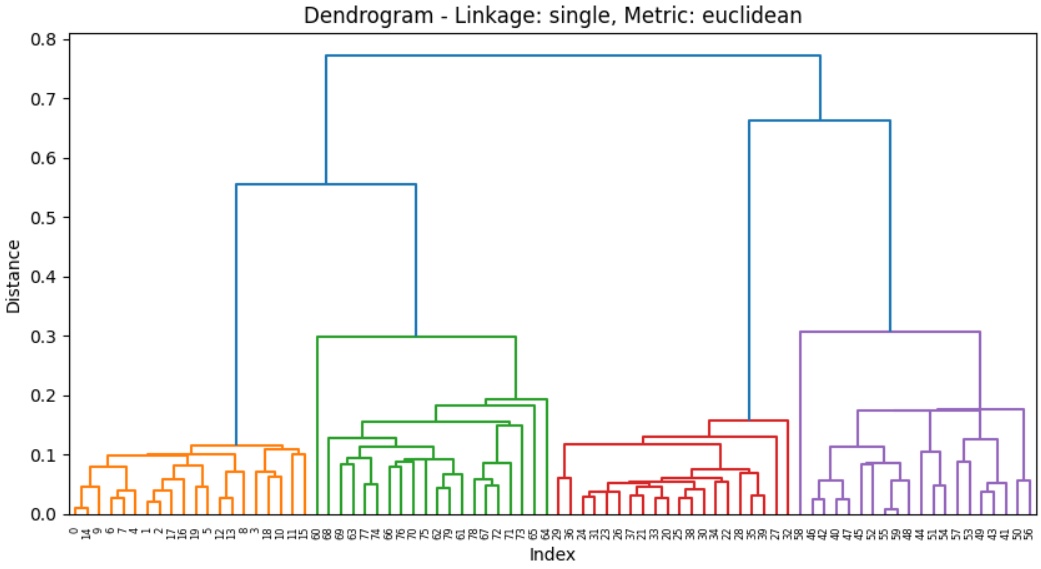


Figure 10 Dendrogram Single Euclidean

Muhammad Hashaam Shahid 2491488

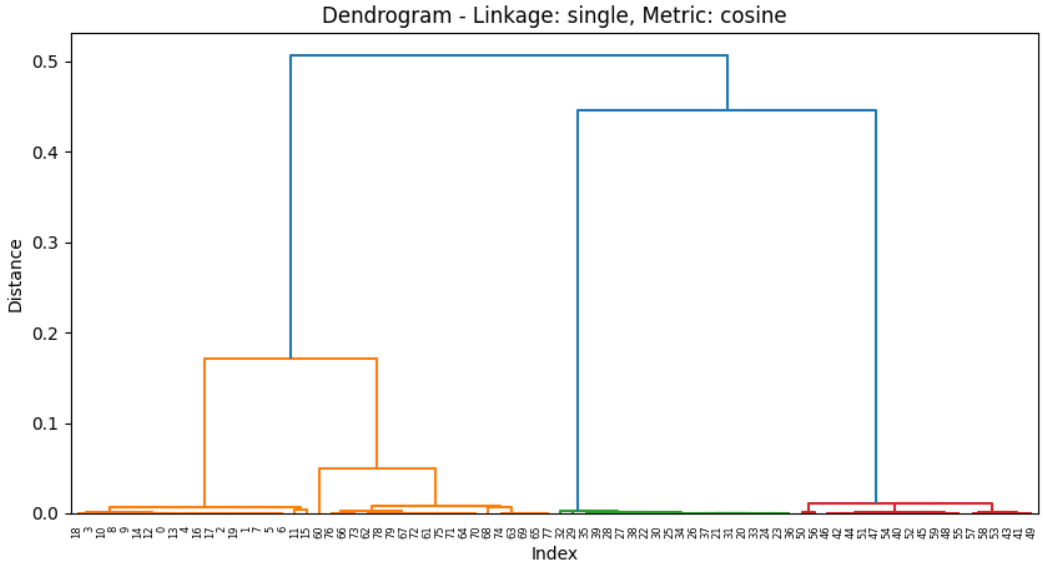


Figure 11 Dendrogram Single Cosine

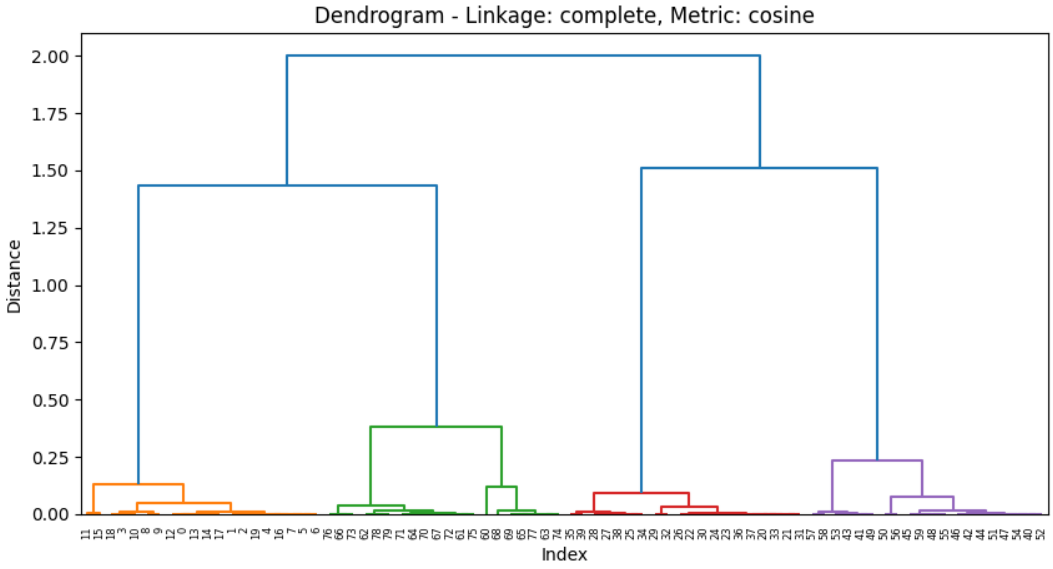


Figure 12 Dendrogram Complete Cosine

Muhammad Hashaam Shahid 2491488

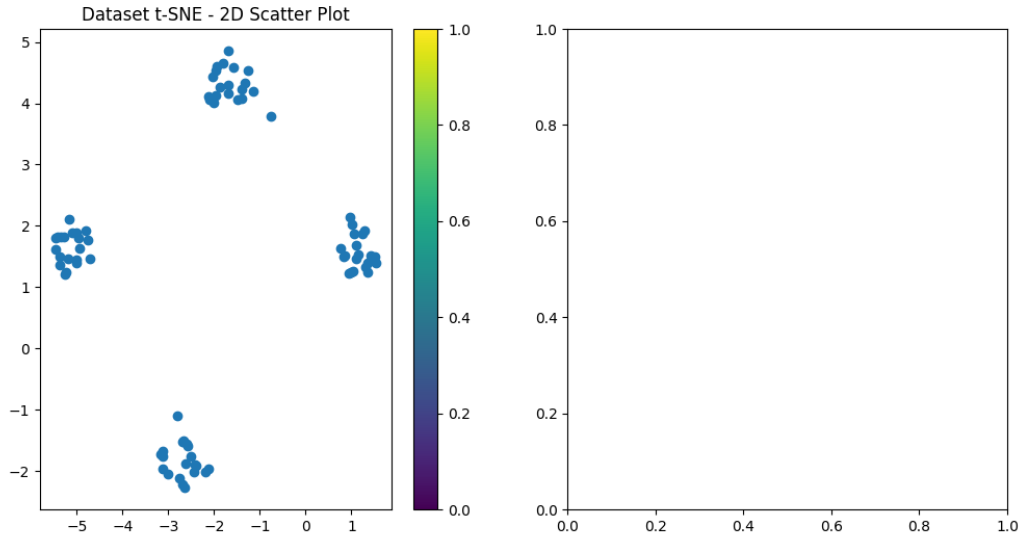


Figure 13 t-SNE 2D scatter plot for the dataset

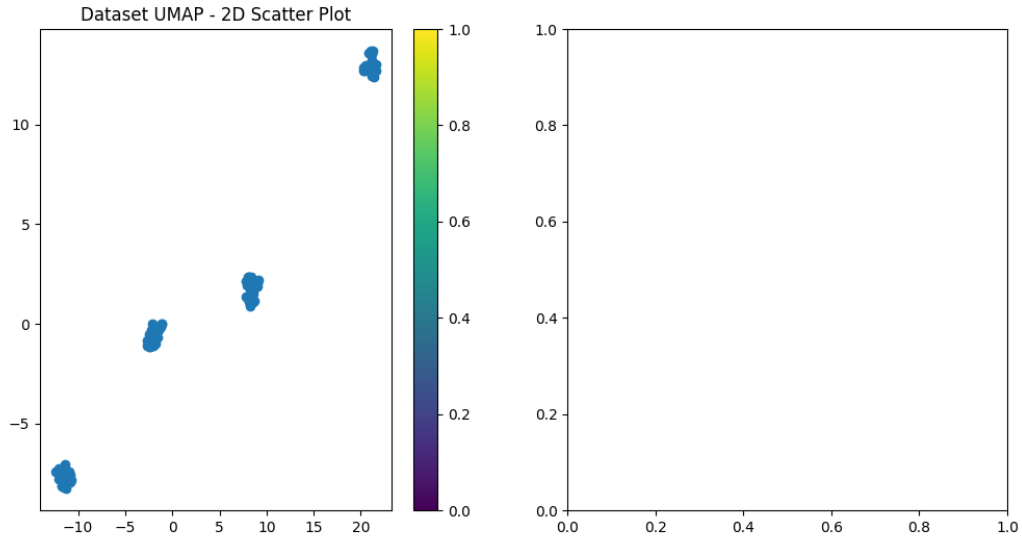


Figure 14 UMAP 2D scatter plot for the dataset

//getting 0 silhouette score