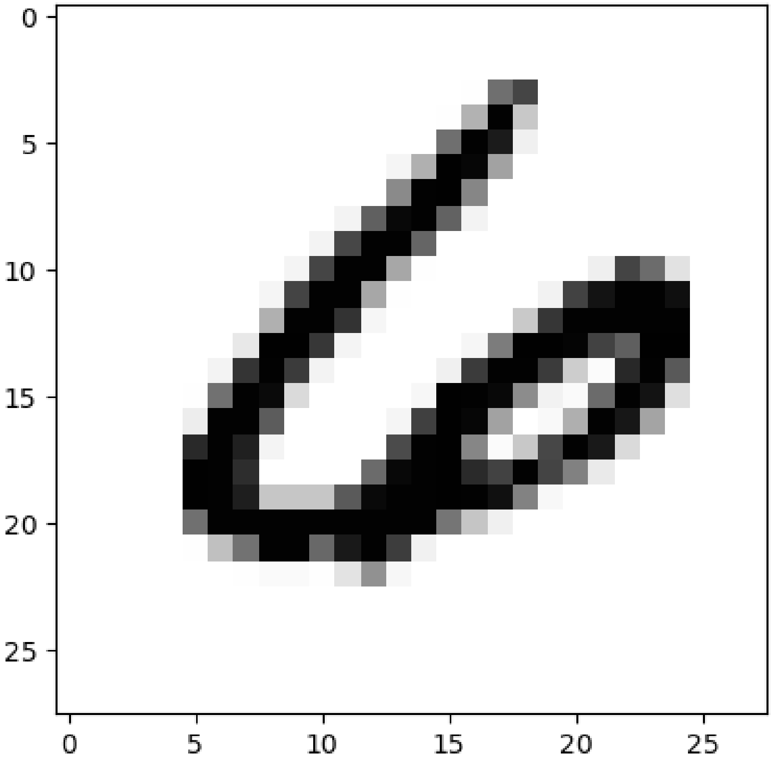
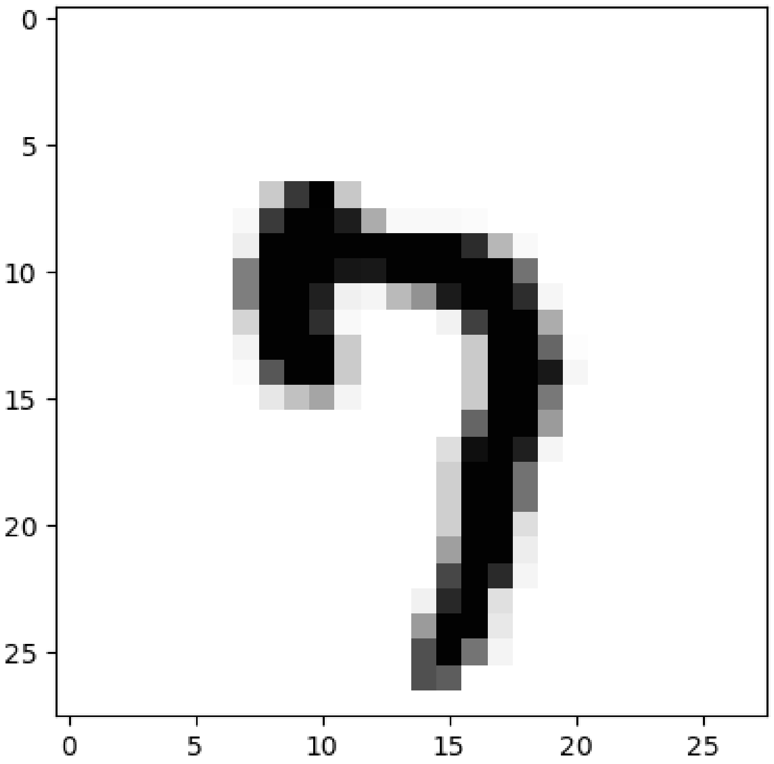
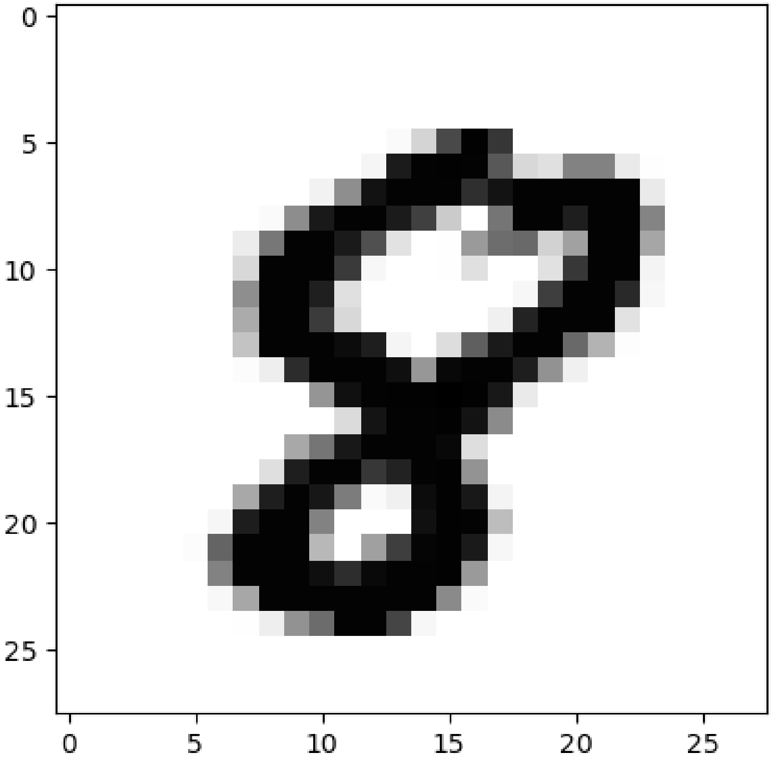
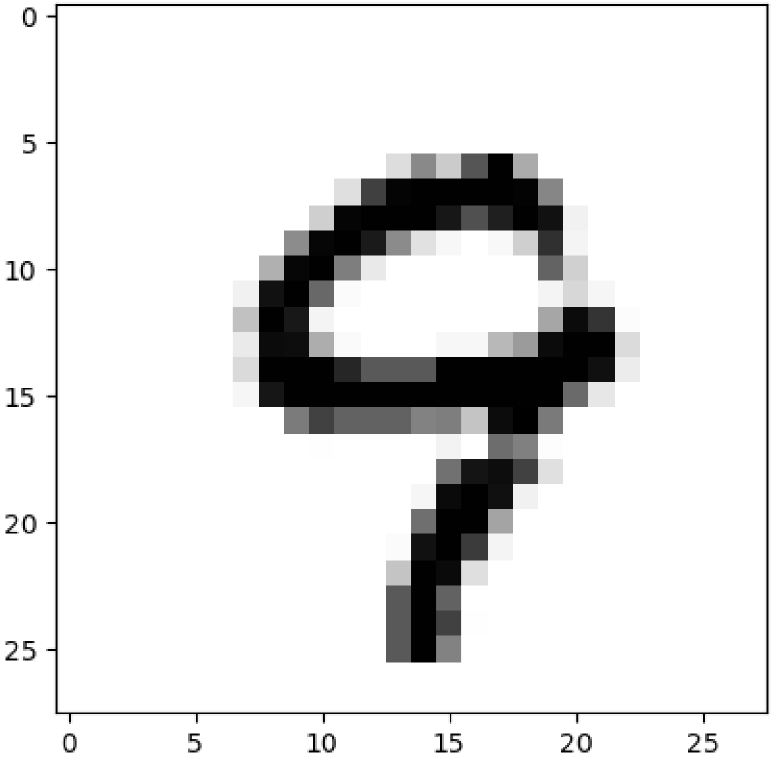
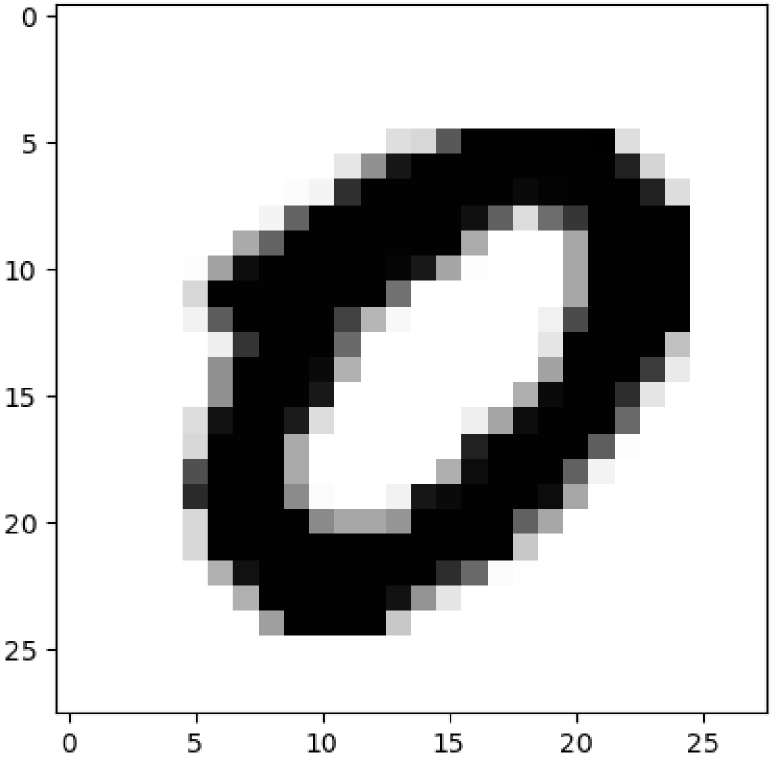
Q1.1

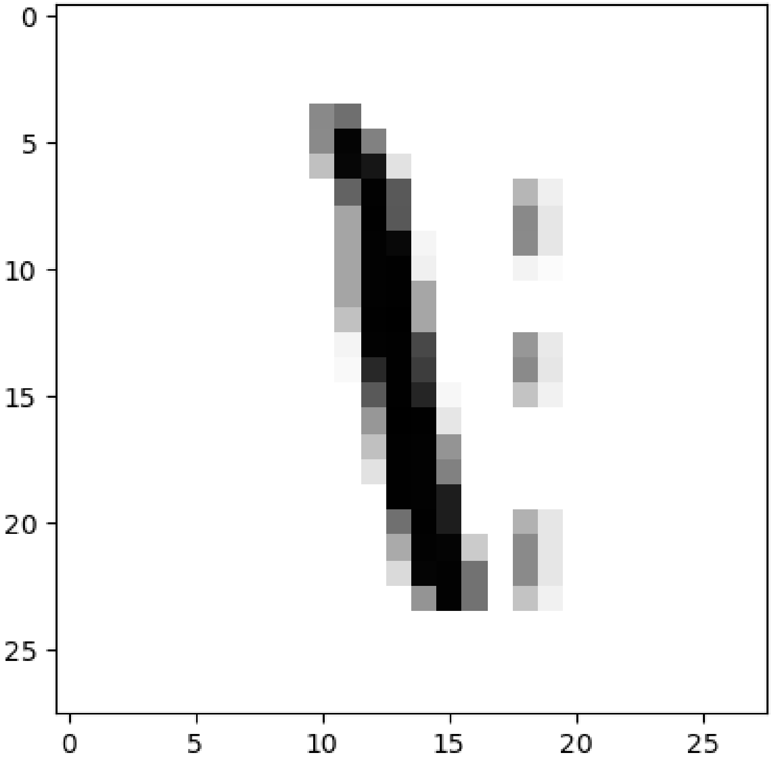


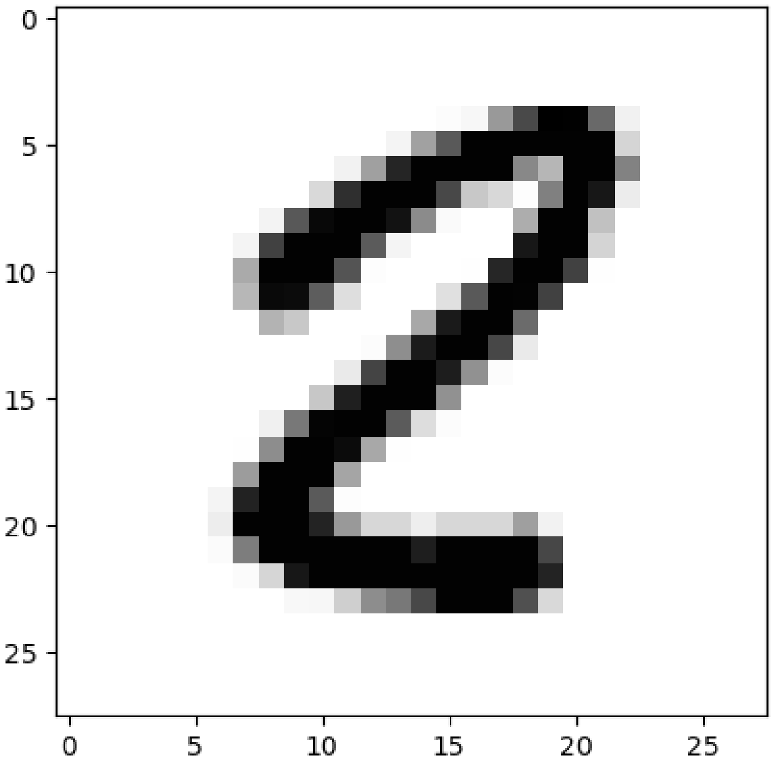

0
5
10
15
20
25
0
5
10
15
20
25

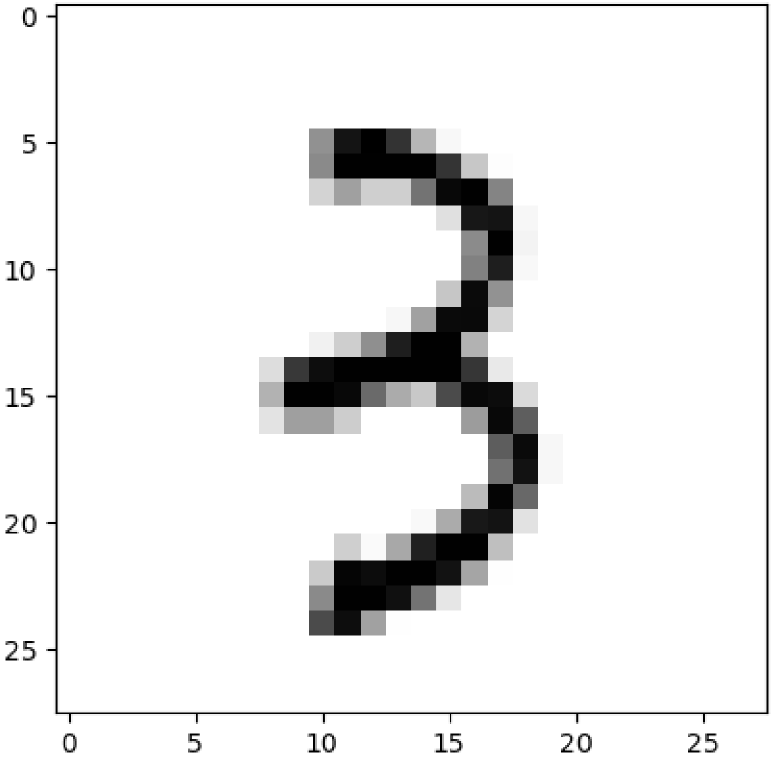

0
5
10
15
20
25
0
5
10
15
20
25

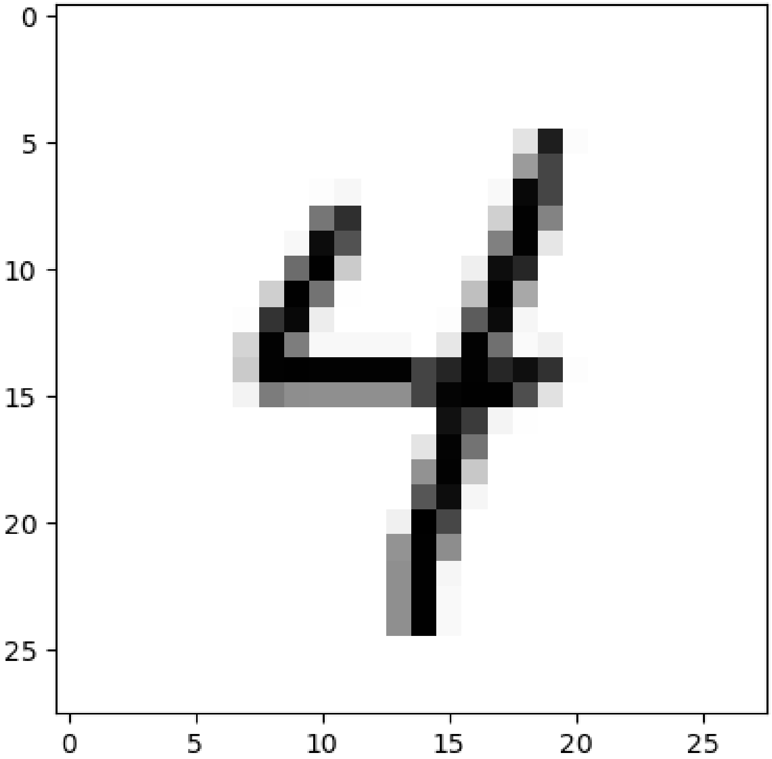

0
5
10
15
20
25
0
5
10
15
20
25

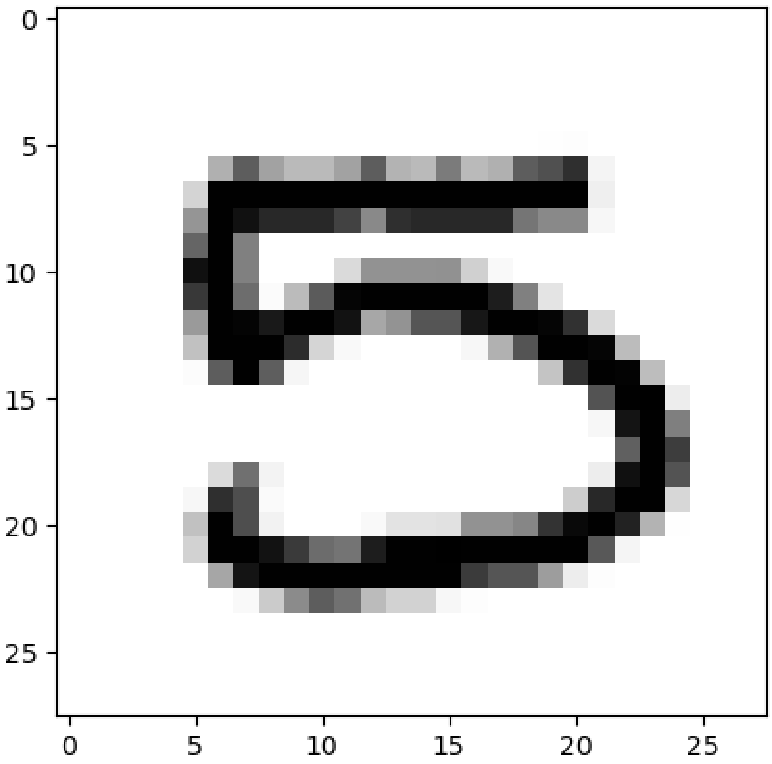

0
5
10
15
20
25
0
5
10
15
20
25


0
5
10
15
20
25
0
5
10
15
20
25

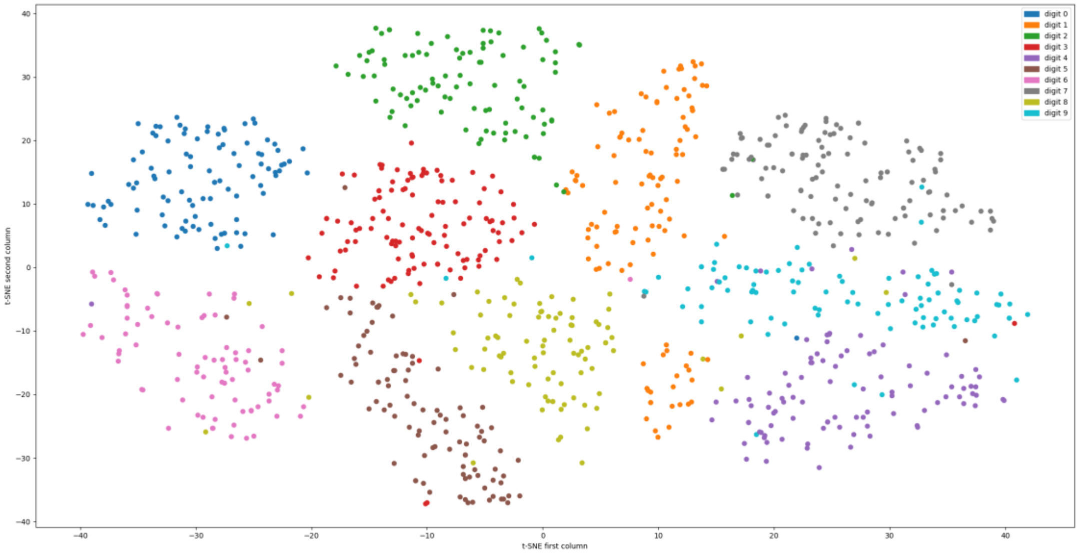

0
5
10
15
20
25
0
5
10
15
20
25


0
5
10
15
20
25
0
5
10
15
20
25

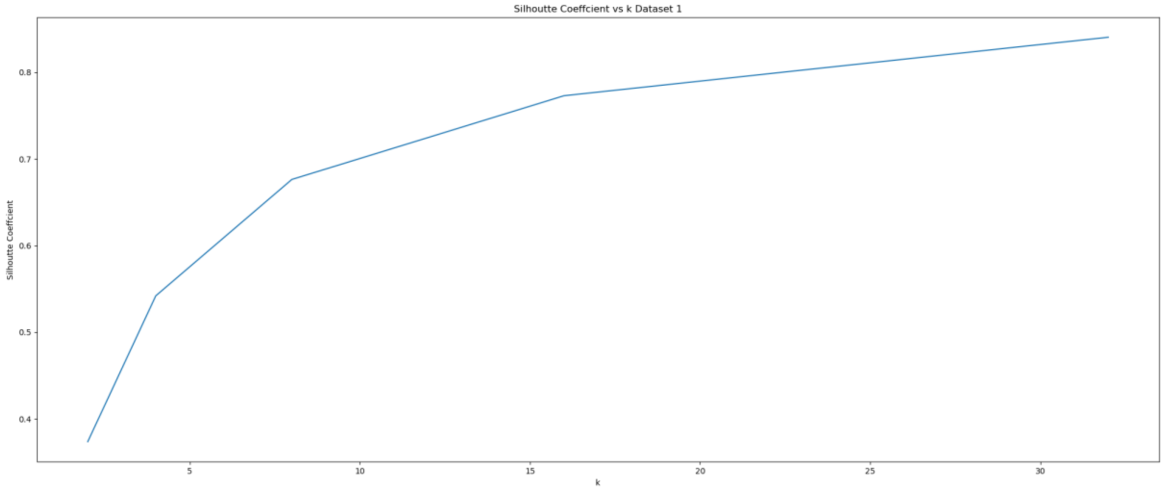

0
5
10
15
20
25
0
5
10
15
20
25

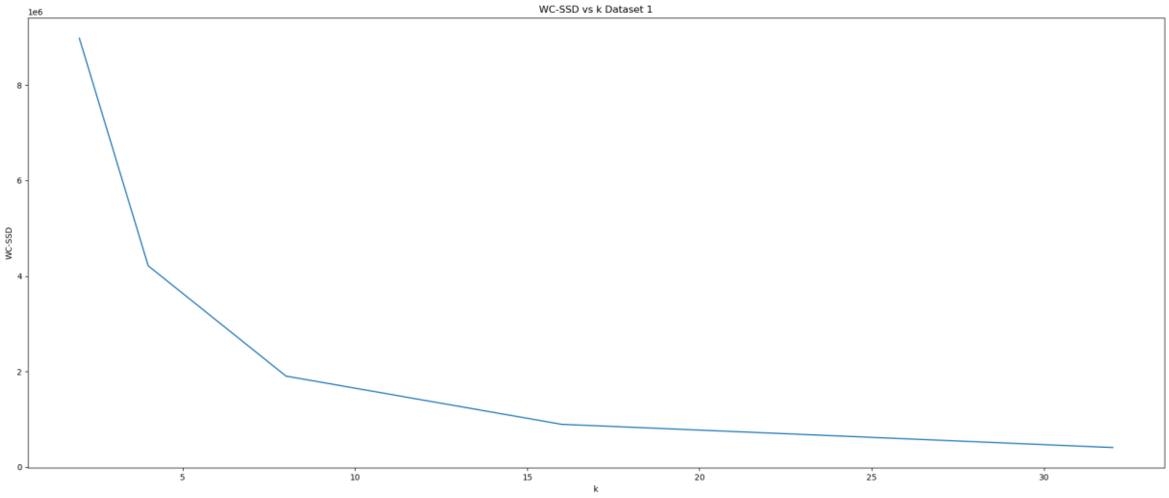

0
5
10
15
20
25
0
5
10
15
20
25

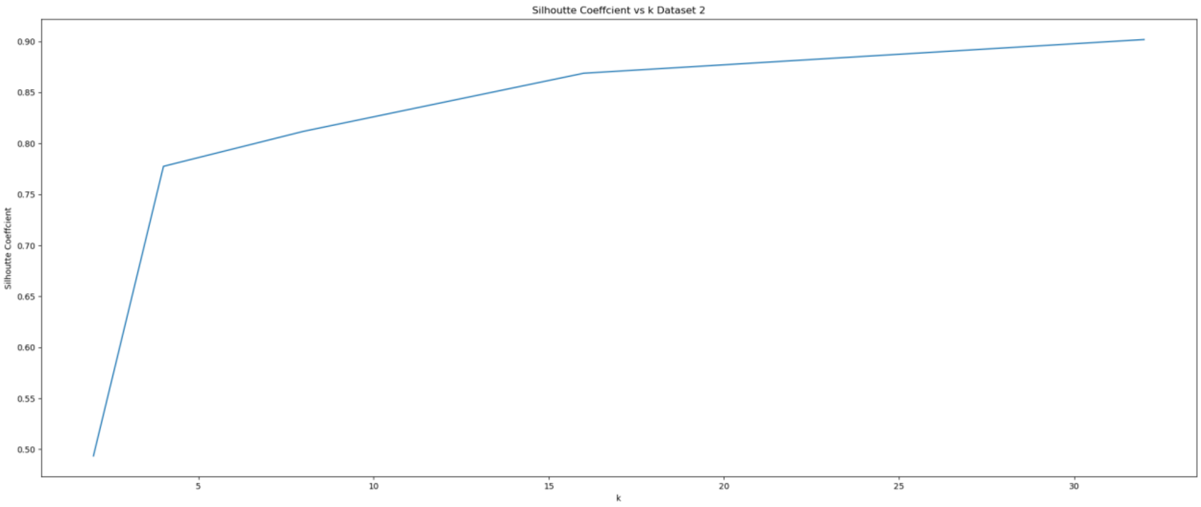
Q1.2

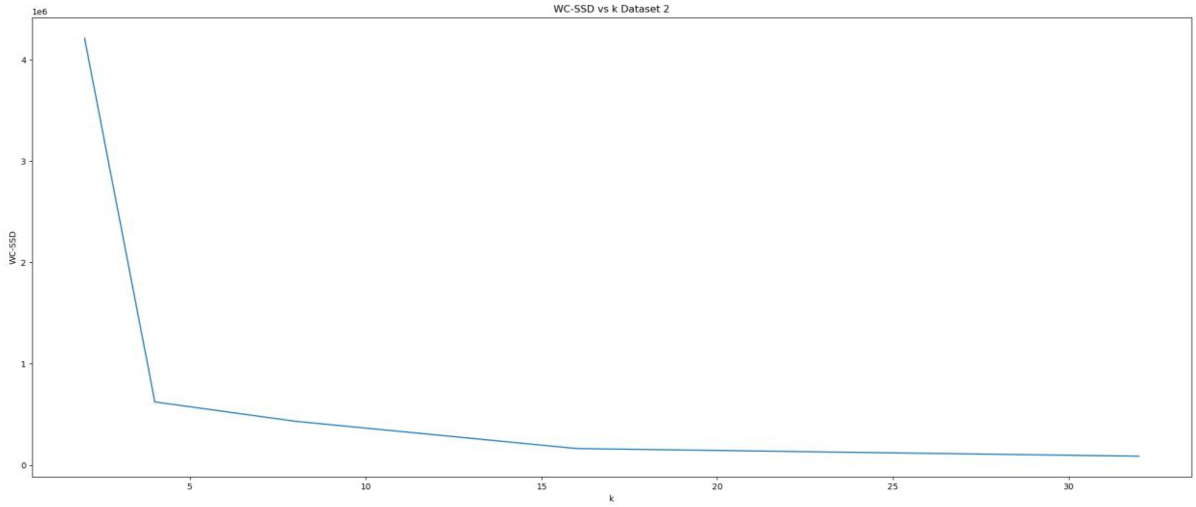

digit 0
digit 1
digit 2
digit 3
digit 4
digit 5
digit 6
digit 7
digit 8
digit 9
t-SNE first column
t-SNE second column
−40
−30
−20
−10
0
10
20
30
40
−40
−30
−20
−10
0
10
20
30
40

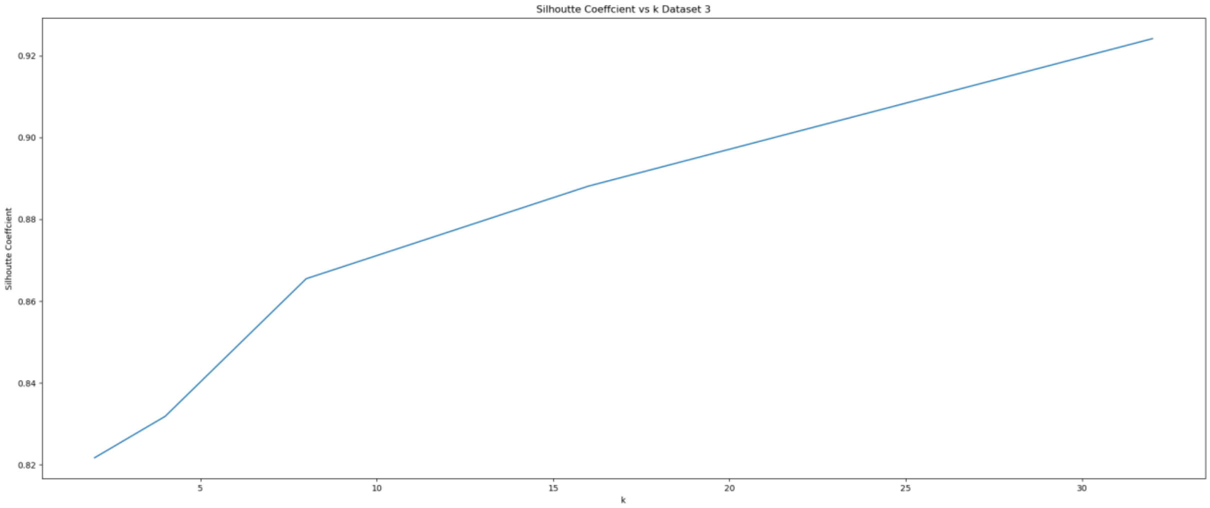
Q2.2.1

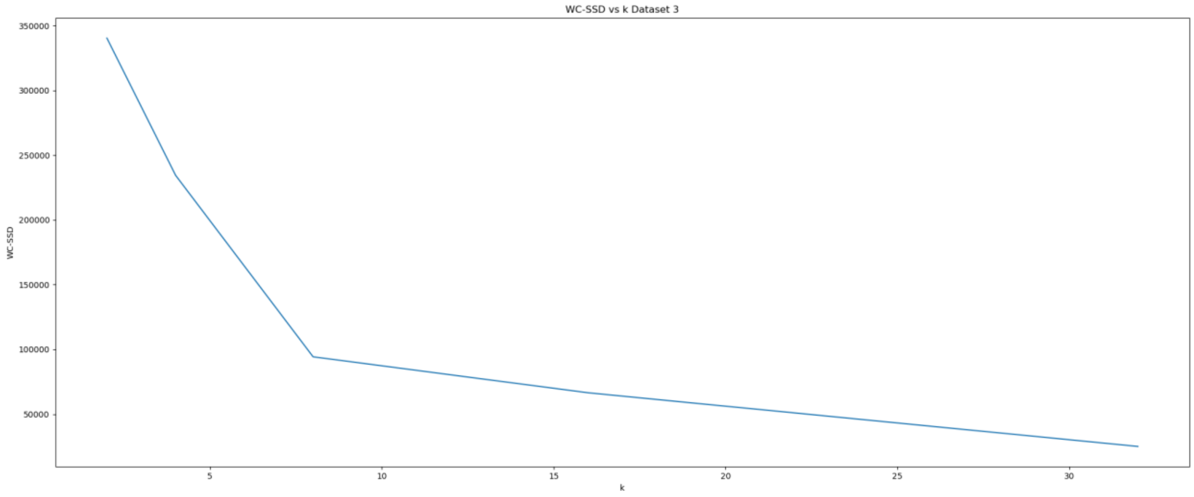

Silhoutte Coeffcient vs k Dataset 1
Silhoutte Coeffcient
0.4
0.5
0.6
0.7
0.8
5
10
15
20
25
30
k


WC-SSD vs k Dataset 1
1e6
WC-SSD
0
2
4
6
8
5
10
15
20
25
30
k


Silhoutte Coeffcient vs k Dataset 2
Silhoutte Coeffcient
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
5
10
15
20
25
30
k


WC-SSD vs k Dataset 2
1e6
4
3
2
1
0
WC-SSD
5
10
15
20
25
30
k


Silhoutte Coeffcient vs k Dataset 3
0.92
0.90
0.88
0.86
0.84
0.82
Silhoutte Coeffcient
5
10
15
20
25
30
k


WC-SSD vs k Dataset 3
350000
300000
250000
200000
150000
100000
50000
WC-SSD
5
10
15
20
25
30
k

Q2.2.2

We evaluate the elbow and knee point of WC-SSD and SC plots respectively, for each dataset. Beyond those points, the magnitude of gradient of WC-SSD or SC curve drastically decrease.

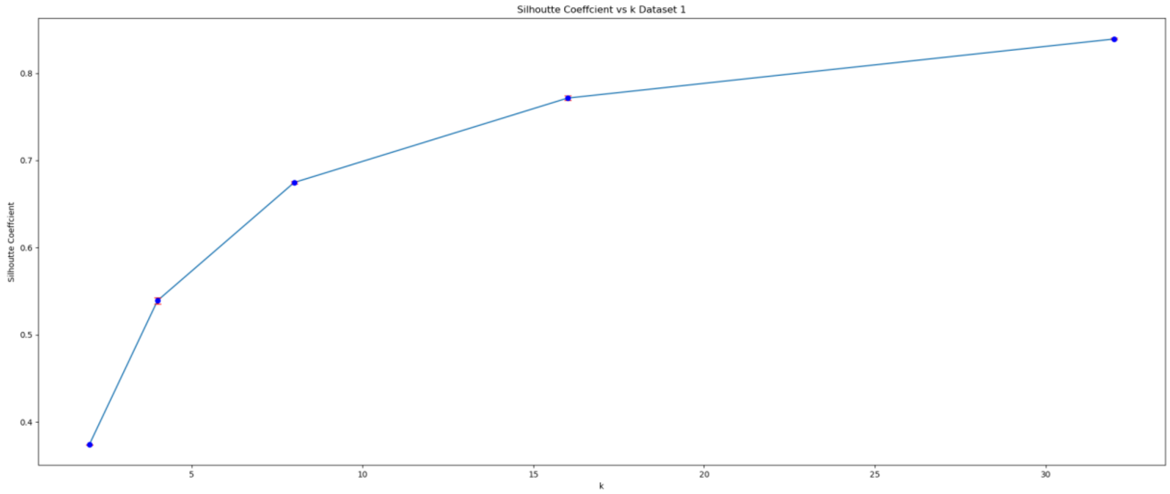
For dataset 1 we chose optimal K=16, for dataset 1 we chose optimal K=4, For dataset 3 we chose optimal K=8

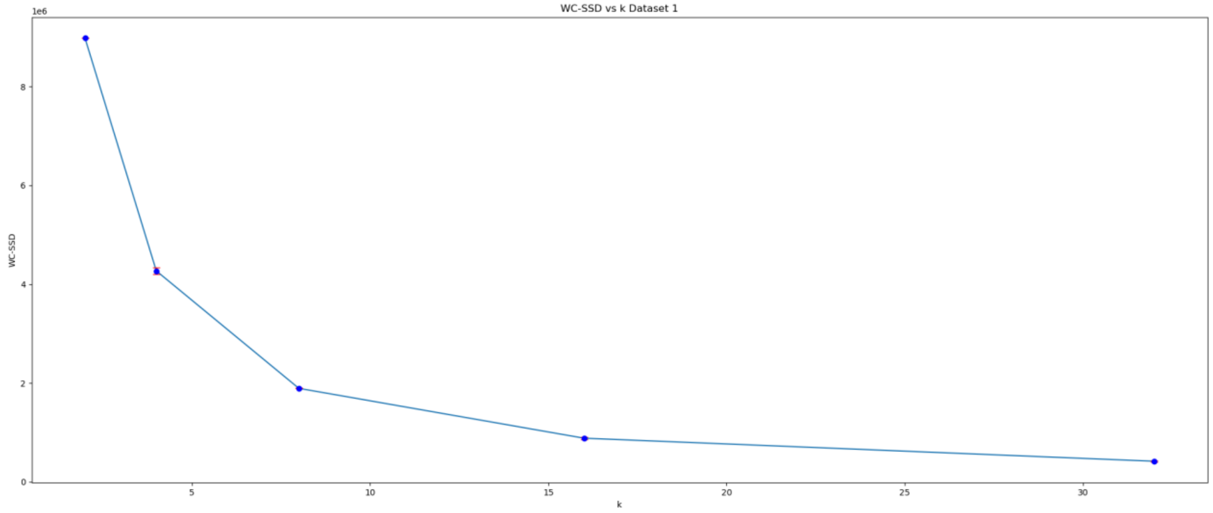
Discuss how the results compare across the two scores and the three versions of the data:

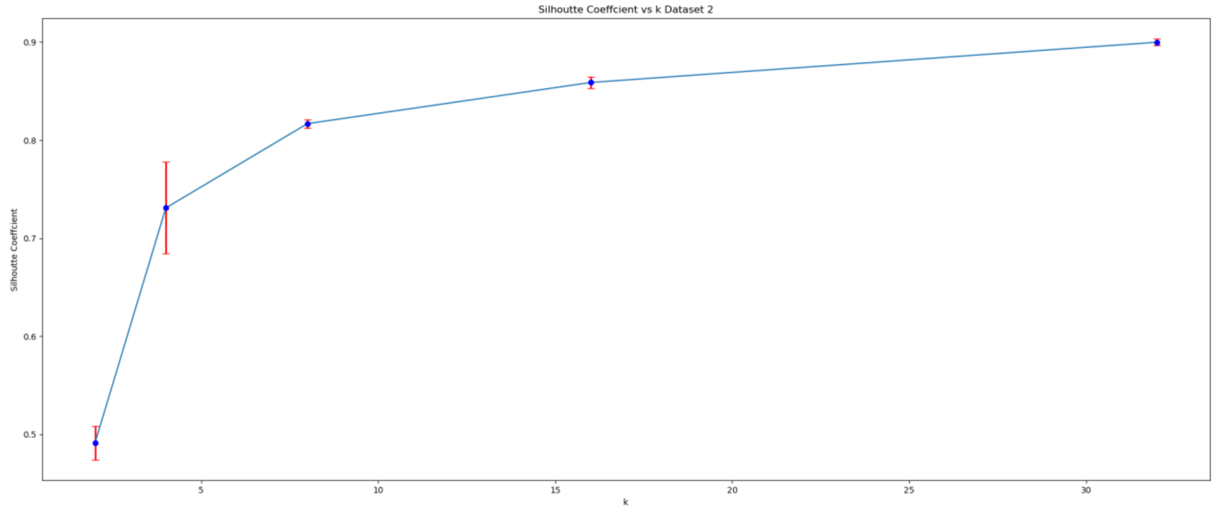
For the WC-SSD plots across the three datasets, its value decrease as the number of classes decreases. Th optimal K in dataset 1 and 2 are somewhat close to the real number of classes, while the optimal K in dataset 3 is not very close to the real number of classes, may be resulted from the uniform distribution of those sample spots in their cluster (loose cluster) which cause some ambiguity in k-means clustering evaluated upon WC-SSD.

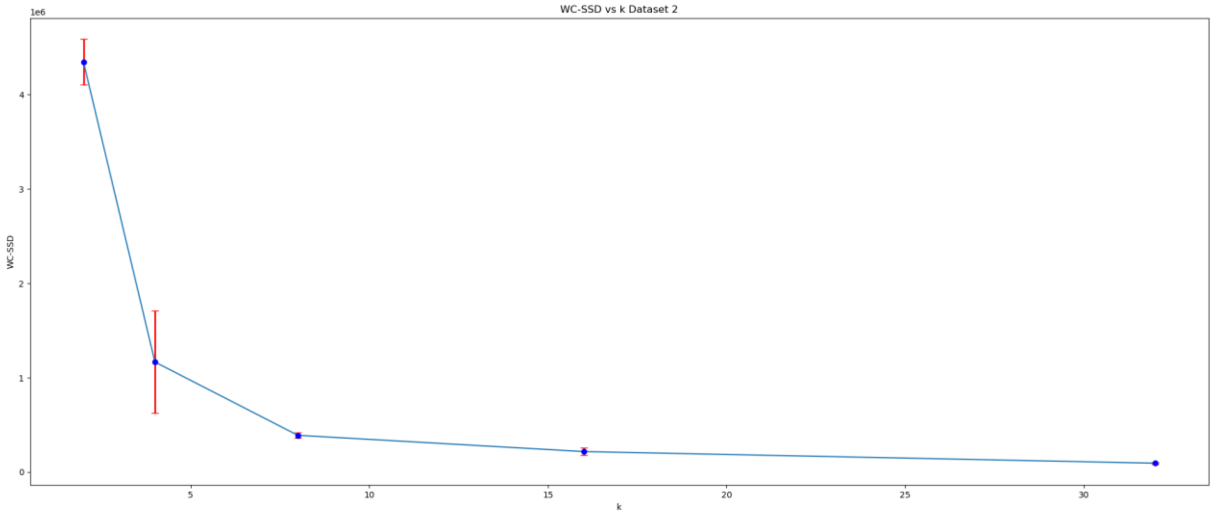
Q2.2.3

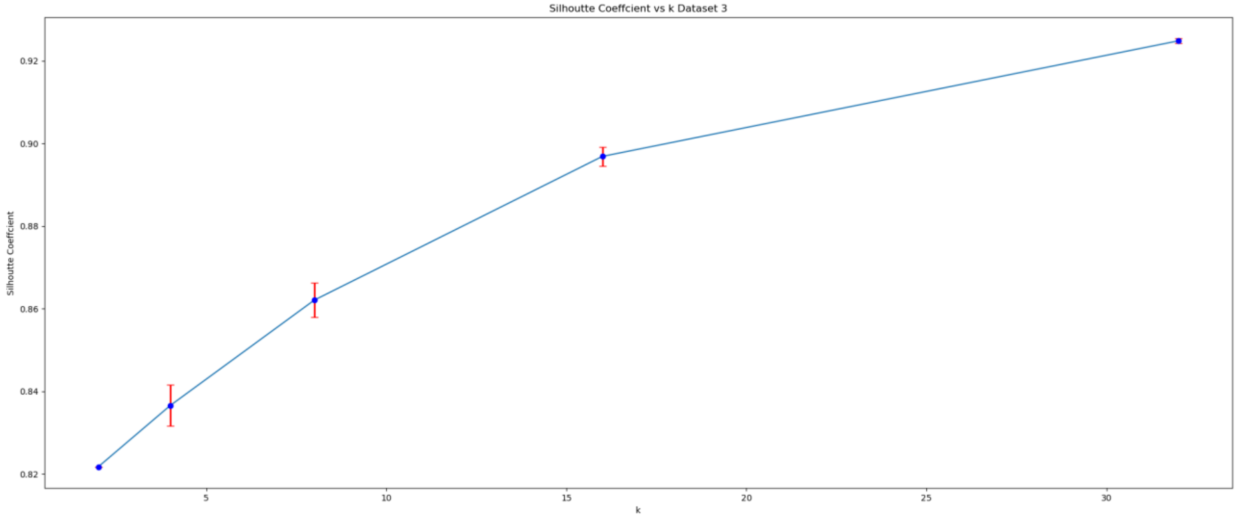
Please see the following plots

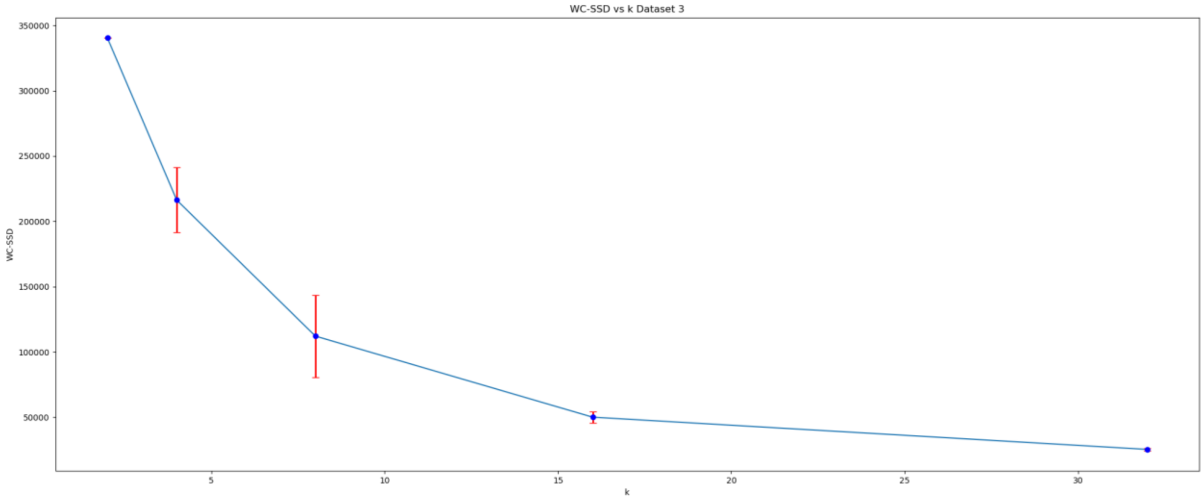





Silhoutte Coeffcient vs k Dataset 2
Silhoutte Coeffcient
0.9
0.8
0.7
0.6
0.5
5
10
15
20
25
30
k

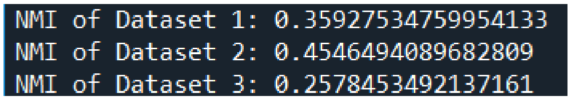

WC-SSD vs k Dataset 2
1e6
WC-SSD
4
3
2
1
0
5
10
15
20
25
30
k


Silhoutte Coeffcient vs k Dataset 3
Silhoutte Coeffcient
0.92
0.90
0.88
0.86
0.84
0.82
5
10
15
20
25
30
k



The plots show that kmeans is sensitive to the initial starting conditions. In each dataset the SC and WC-SSD plots show high standard deviation.

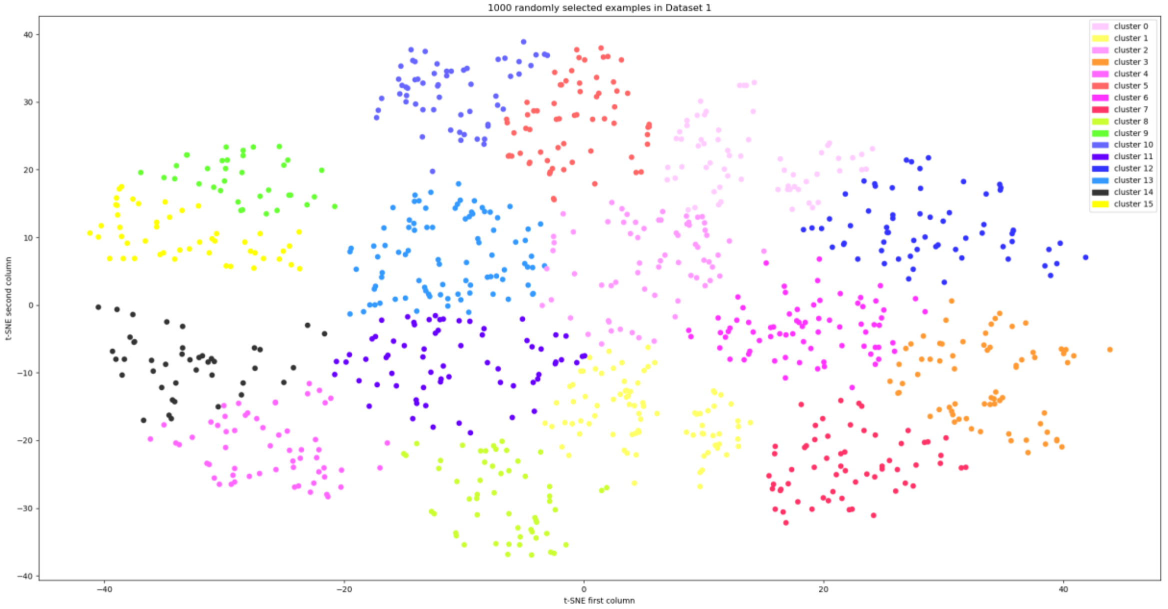
Q2.2.4

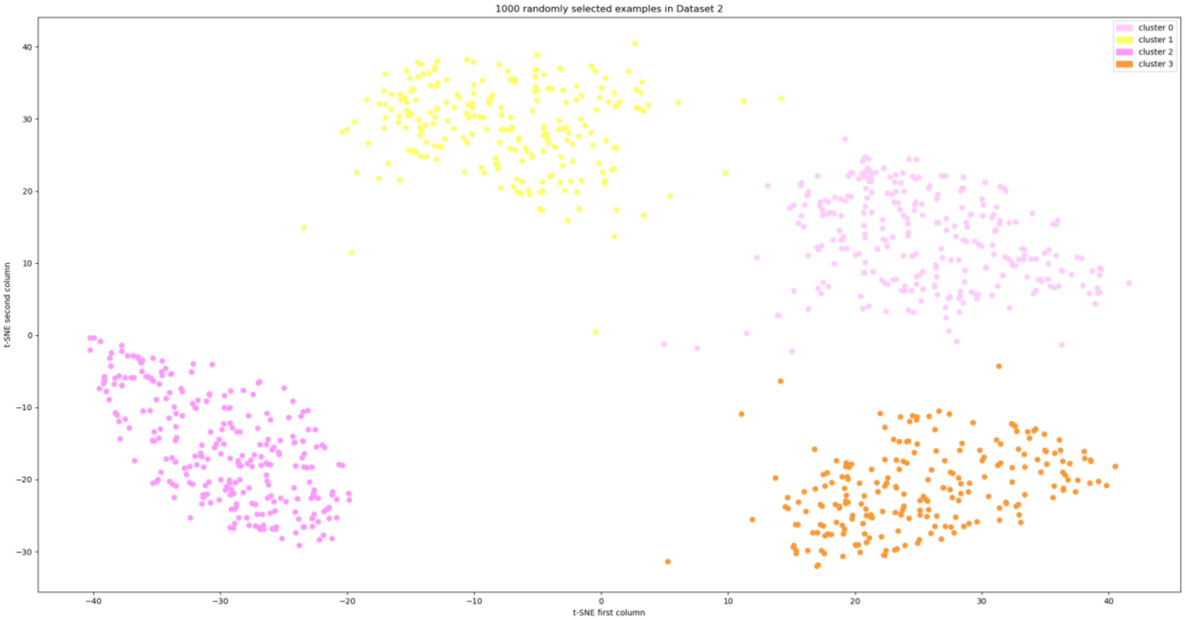


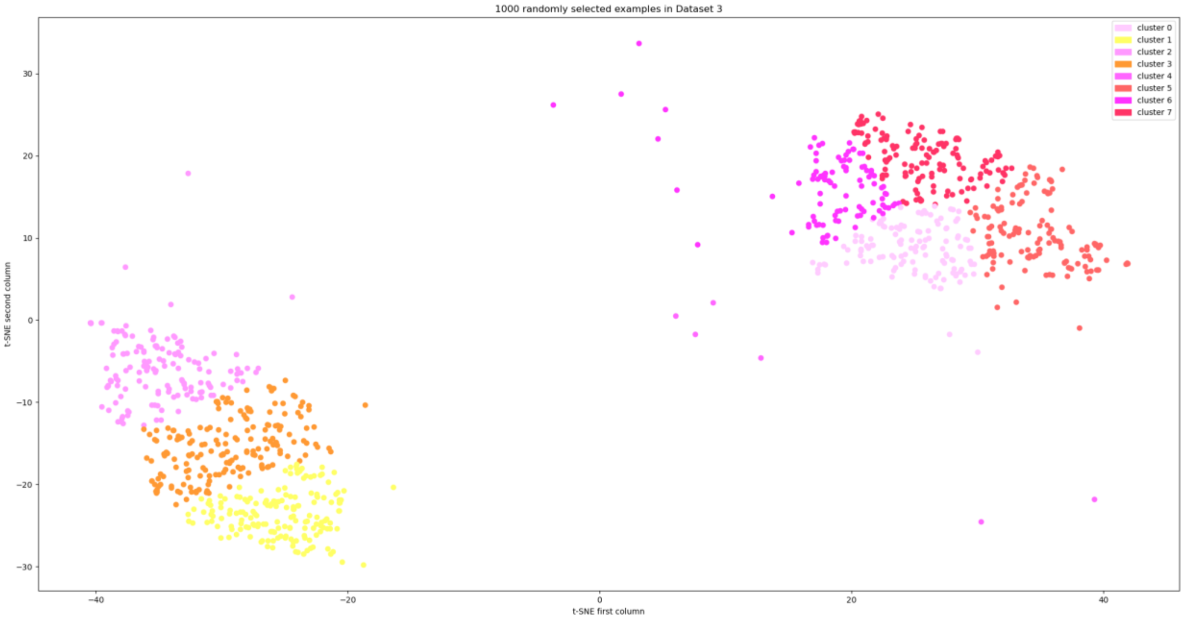
```
NMI of Dataset 1: 0.35927534759954133
NMI of Dataset 2: 0.4546494089682809
NMI of Dataset 3: 0.2578453492137161
```

K = 16,4,8 for dataset 1,2 and 3

See the following graph

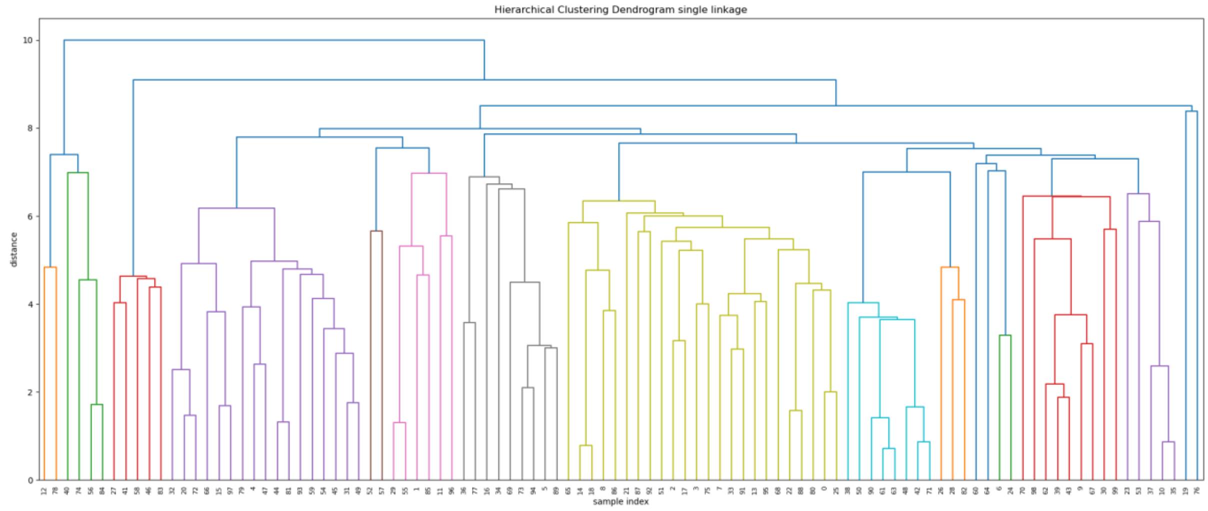


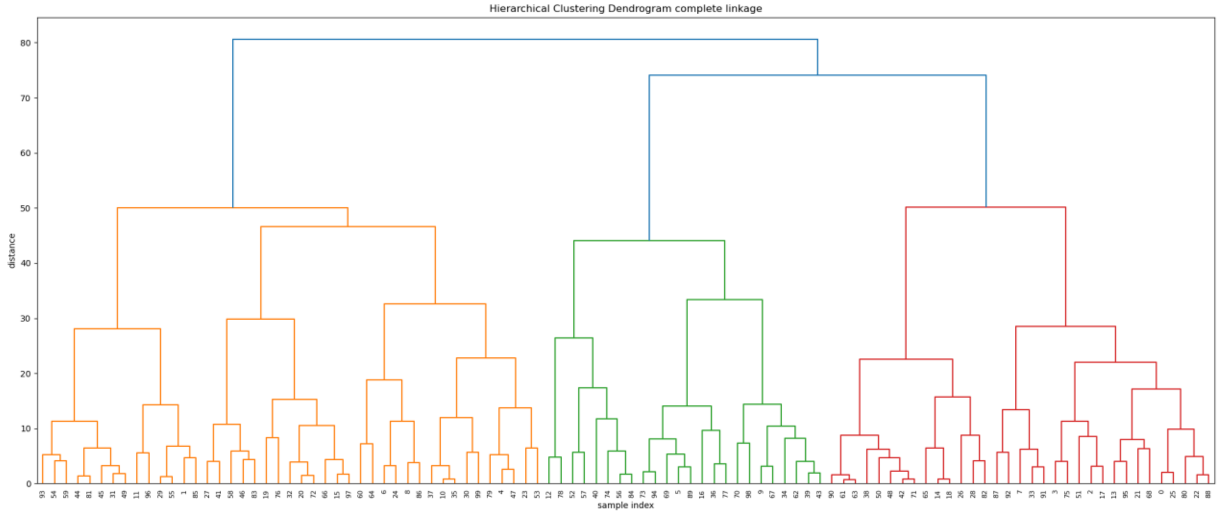


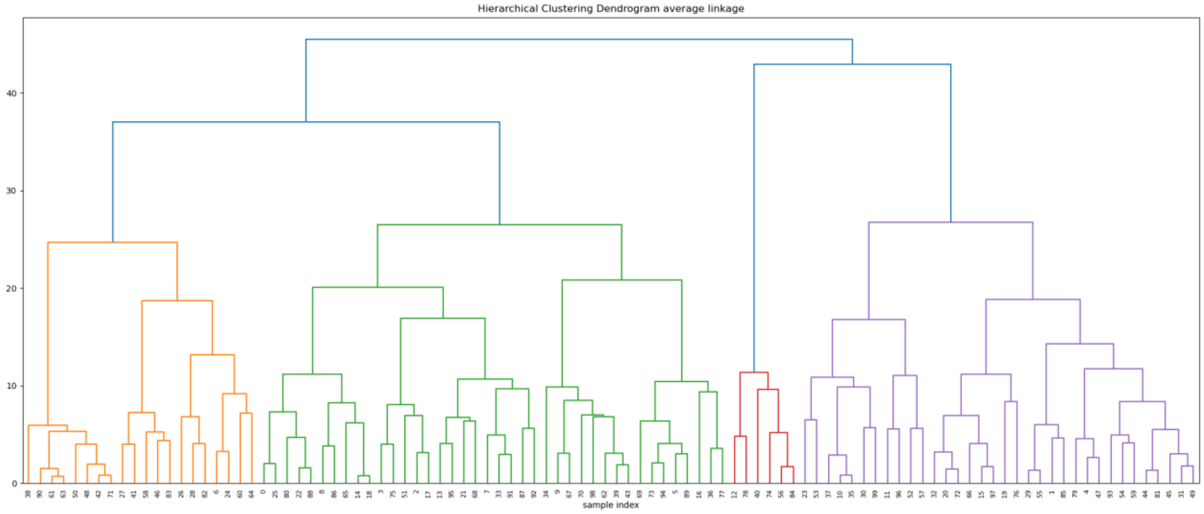


In general the NMI value should be larger if the number of clusters is smaller. In our case, dataset2 has the highest NMI because it has 16 clusters, but for dataset 3 since the clustering is far away from real label numbers, the NMI is a bit abnormal. In the scatter plot, the clusters are more distinguishable with less total cluster numbers. The most separable dataset is 2, then 3, then 1.

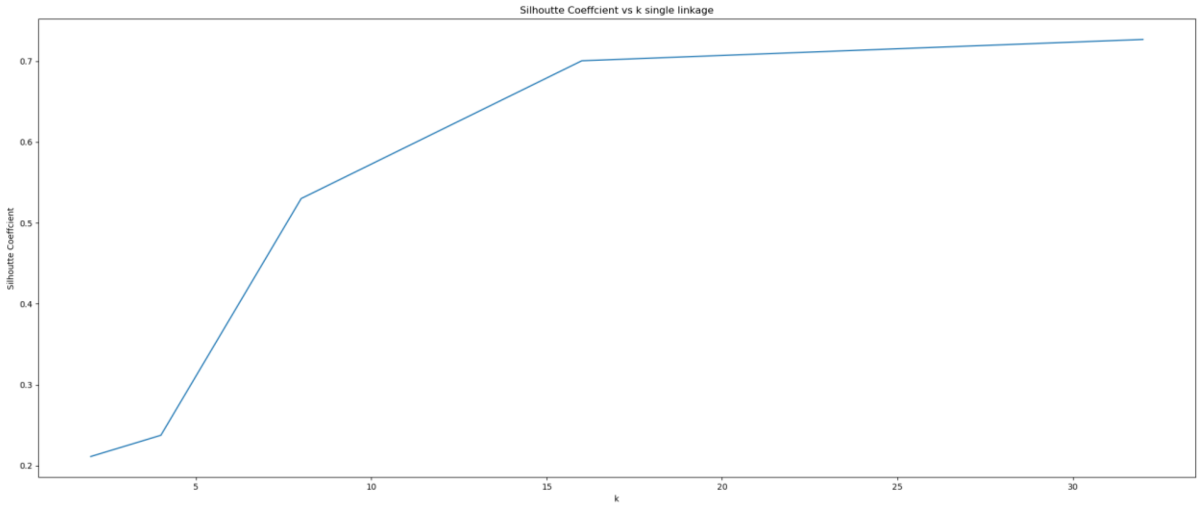
Q3.2, 3.3

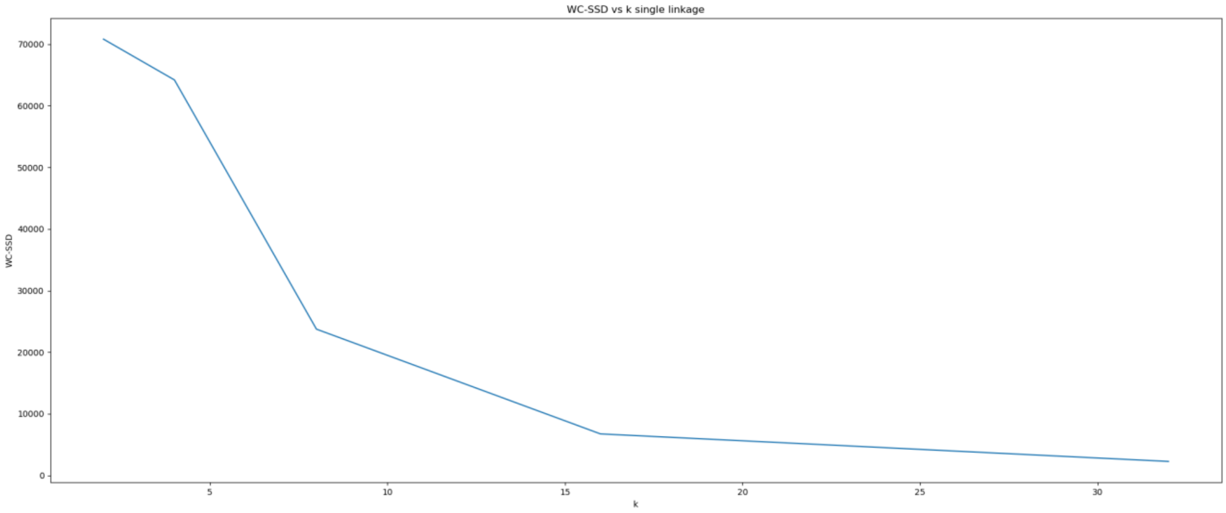

Hierarchical Clustering Dendrogram single linkage

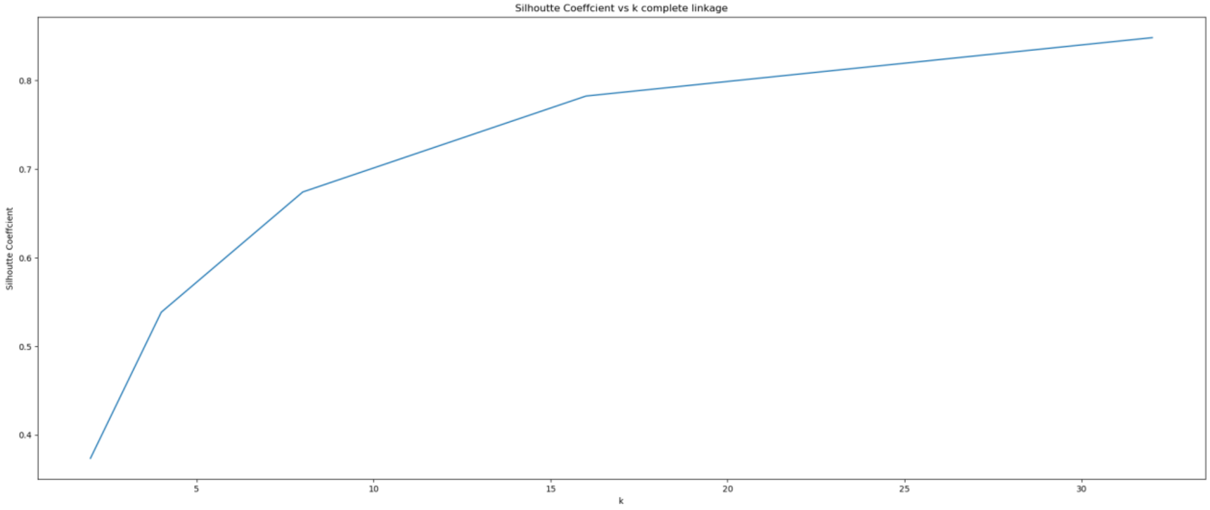

Hierarchical Clustering Dendrogram complete linkage

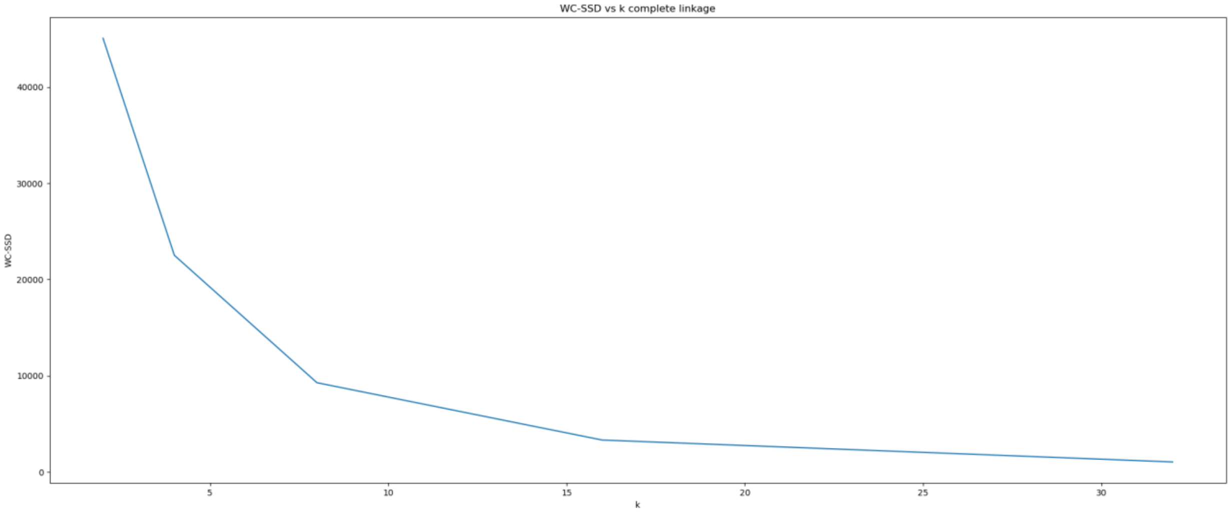

Hierarchical Clustering Dendrogram average linkage

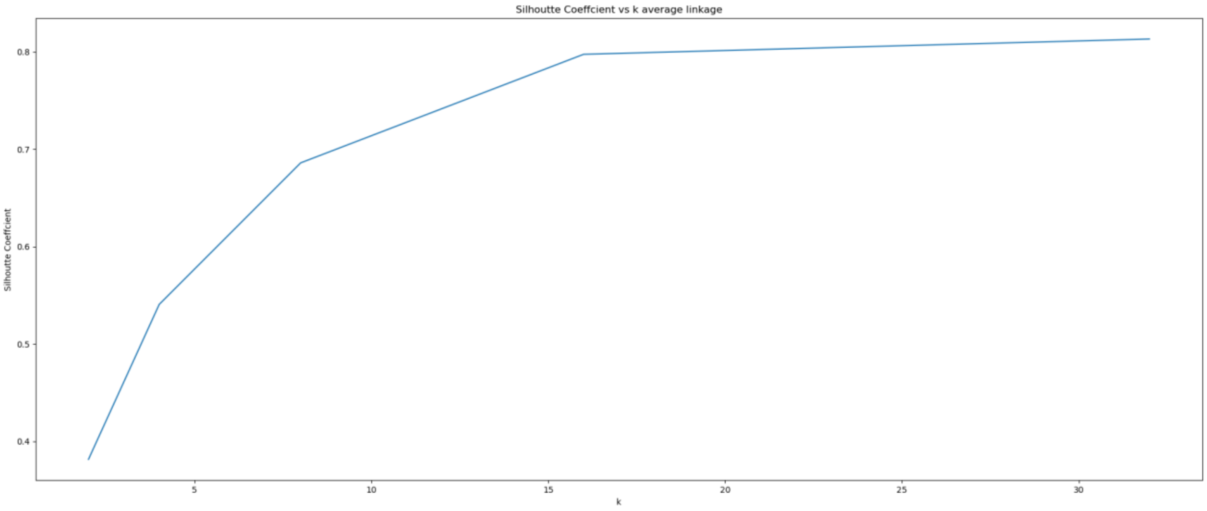
Q3.4

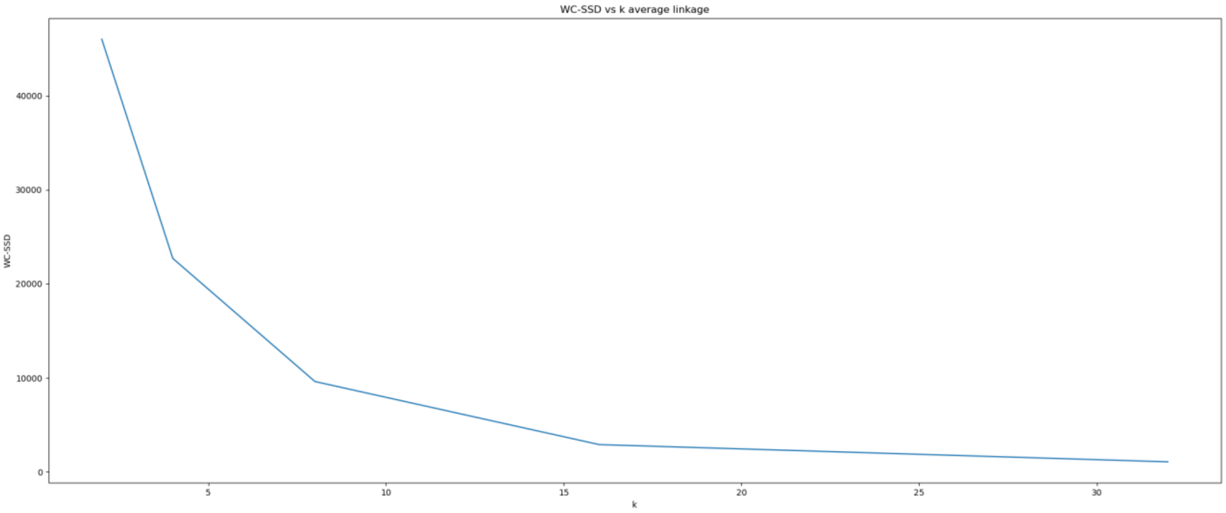

Silhoutte Coeffcient vs k single linkage
Silhoutte Coeffcient
0.2
0.3
0.4
0.5
0.6
0.7
5
10
15
20
25
30
k


WC-SSD vs k single linkage
WC-SSD
0
10000
20000
30000
40000
50000
60000
70000
5
10
15
20
25
30
k


Silhoutte Coeffcient vs k complete linkage
Silhoutte Coeffcient
0.4
0.5
0.6
0.7
0.8
5
10
15
20
25
30
k


WC-SSD vs k complete linkage
WC-SSD
40000
30000
20000
10000
0
5
10
15
20
25
30
k


Silhoutte Coeffcient vs k average linkage
Silhoutte Coeffcient
0.8
0.7
0.6
0.5
0.4
5
10
15
20
25
30
k


WC-SSD vs k average linkage
WC-SSD
40000
30000
20000
10000
0
5
10
15
20
25
30
k

Q3.5

Using the same method used in finding optimal K in kmeans in previous question, we decide the optimal K is 16 for all three likages. The value is the same as what we found using k-means for Dataset 1 in Section 2.

Q3.6

With K=16 for all linkages

Output:

```
single NMI:    0.3545881948493385
complete NMI:    0.369599539861315
average NMI:    0.390785830258648
```

The NMI score using average linkage is highest, the second highest is using complete linkage, then single linkage. It seems that the difference between the three scores are not significant and the results will vary a lot with different runs. Their average value is higher than what we got for NMI in using k-means for Dataset 1 in Section2