PET 328E Homework IV

May 29, 2023

Baran Gökgöz
050190441

1)

The certificates are given in the .zip file.

2a)

After the relevant clustering is conducted on the numerical columns only, the elbow is can be observed on Figure 2A where the we see the average distance form the centroid to decrease with a sudden curve. Based on the number of clusters that are given and the behavior of the curve, optimum number for k could could be 2 for SPE_shale data set.
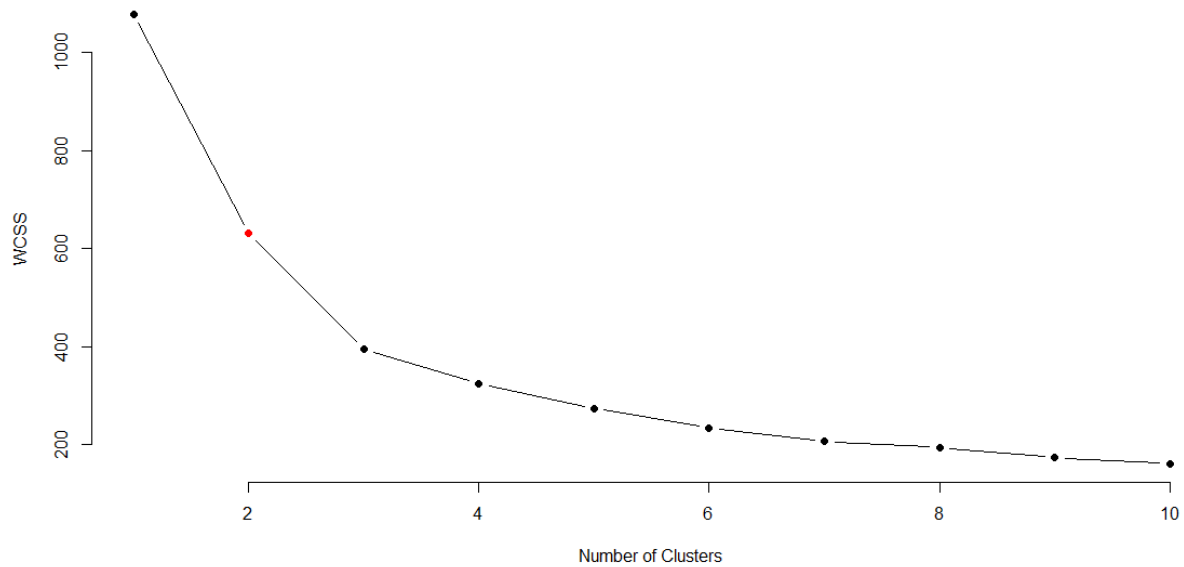
Figure 2A: Plot for K-Means clustering with elbow method

2b)

Cluster diagram is created by hierarchical clustering with minimum linkage and the resulting plot is given in Figure 2B. After trying several different clustering methods, a reasonable cut-off value of 3 decided as it gives out the most reliable information regarding the clusters due to the fact bigger clusters could become unnecessary and lesser clusters could become too specific to convey meaningful information.
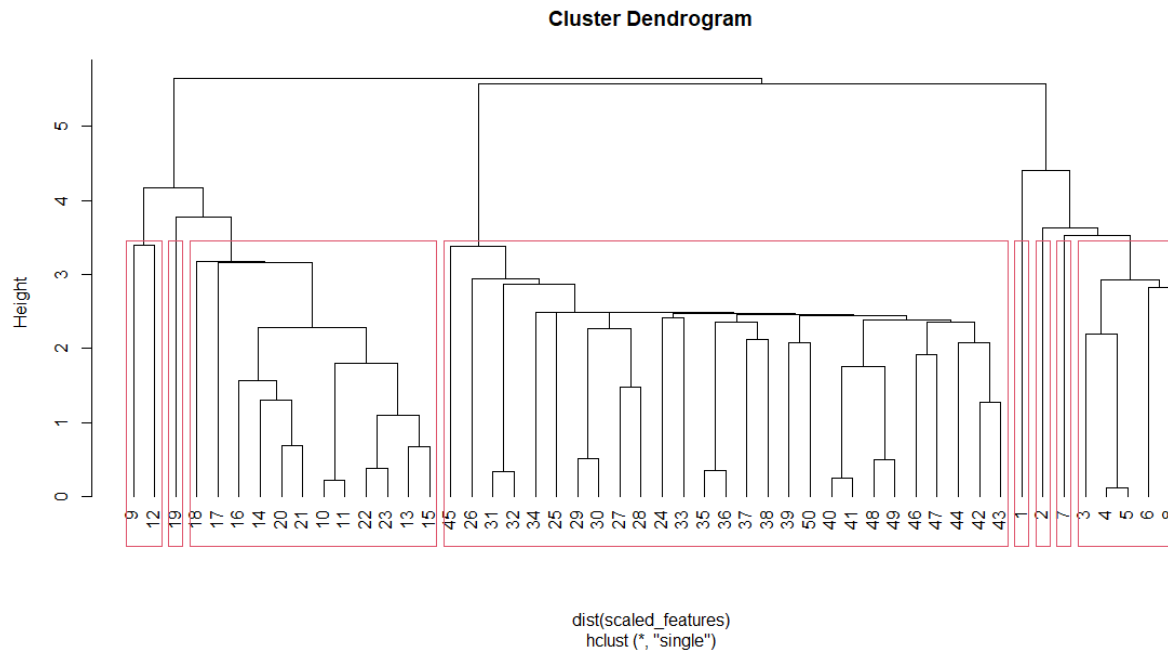
Figure 2B: Dendogram plot with clusters

3)

FNN package in R is used to detect the anomalies in the SPE_shale dataset. By using this package for the dataset, the anomalies with 5 neighbors were classes of 9 and 12; whereas, the anomalies with 10 neighbors were the class 12 only.

Bonus)

The link to the github repository: *https://github.com/BGokgoz/PET-328E---HW4*