Assignments for Lab-9

Data for Q1-Q4: s1 data

Data for Q5: iris data

1) Visualize s1 data and observe the distribution of data points. What kind of distribution exists? Is it clustered data? Yes, it is a clustered data. 14 clusters are the best.

2) Apply k-means and observe clusters. Use K Means from the modelling tab and then Graphboard

3) Insert some outliers in s1 data manually. Now apply k-means and compare the clusters with the clusters obtained in Q2. What difference do you get? Do you find variation in SSE's?

4) Apply k-means multiple times and observe the clusters and SSE? What difference do you get?

5) Apply k-means on IRIS data. Now first apply PCA on IRIS data and then apply k-means in reduced dimensions. Do you observe difference in clusters/SSE? With PCA, it performs worse, silhouette coefficient is lesser