

a) Compare and contrast K-means Clustering and Hierarchical Clustering.

	K_means Clustering	Hierarchical clustering
Selection of number of clusters	We need advance knowledge of K i.e. no. of clusters one want to divide your data before the model	One can stop at any number of clusters, one find appropriate by interpreting the dendrogram.
Cluster center	Mean as a cluster centre to represent each cluster.	Initially with N clusters and sequentially forms to a single cluster for agglomerative and vice versa for divisive approach
Computational power	Requires less comparatively	Requires more comparatively
Data size	Suitable for bigger datasets	Suitable for smaller datasets

b) Briefly explain the steps of the K-means clustering algorithm.

1. Start by choosing K random points the initial cluster centres.
2. Assign each data point to their nearest cluster centre. The most common way of measuring the distance between the points is the Euclidean distance.
3. For each cluster, compute the new cluster centre which will be the mean of all cluster members.
4. Now re-assign all the data points to the different clusters by taking into account the new cluster centres.
5. Keep iterating through the step 3 & 4 until there are no further changes possible.

c) How is the value of 'k' chosen in K-means clustering? Explain both the statistical as well as the business aspect of it.

The Highest value should be taken from Silhouette graph and the elbow point should be taken for the optimal K value.

Within-Cluster-Sum of Squared Errors sounds a bit complex. Let's break it down:

- The Squared Error for each point is the square of the distance of the point from its representation i.e. its predicted cluster center.
- The WSS score is the sum of these Squared Errors for all the points.
- Any distance metric like the Euclidean Distance or the Manhattan Distance can be used.

In the Silhouette Score reaches its **global maximum at the optimal k**. This should ideally appear as a peak in the Silhouette Value-versus-k plot.

d) Explain the necessity for scaling/standardisation before performing Clustering.

Scaling is required to be done because the columns in larger scales might get while developing the model compared to one which is at smaller scales and hence, the minmax scale or normalized scale needs to be done. This ensures the uniformity among all columns

e) Explain the different linkages used in Hierarchical Clustering.

Single Linkage : Here, the distance between 2 clusters is defined as the shortest distance between points in the two clusters

Complete Linkage : Here, the distance between 2 clusters is defined as the maximum distance between any 2 points in the clusters

Average Linkage : Here, the distance between 2 clusters is defined as the average distance between every point of one cluster to every other point of the other cluster.