1. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:

   Answer: 4

2. In which of the following cases will K-Means clustering fail to give good results?

   Answer: Option D. - 1,2 and 4

3. The most important part of -------- is selecting the variables on which clustering is based?

   Answer: Formulating the clustering problem

4. The most commonly used measure of similarity is the ------or its square?

   Answer: Euclidean distance

5. ----is a clustering procedure where all objects start out in one giant cluster. Clusters are formed by dividing this cluster into smaller and smaller clusters?

   Answer: Divisive clustering

6. Which of the following is required by K-means clustering?

   Answer: D. All answers are correct

7. The goal of clustering is to-

   Answer: Divide the data points into groups

8. Clustering is a-

   Answer: Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

   Answer: All of the above

10. Which version of the clustering algorithm is most sensitive to outliers?

    Answer: K-means clustering algorithm

11. Which of the following is a bad characteristic of a dataset for clustering analysis-

Answer: All of the above

12. For clustering, we do not require-

Answer: Labelled data

13. How is cluster analysis calculated?

Answer: Cluster analysis is simple and involves separation into groups. cluster analysis is a way to identify the groups.

The below steps are used for calculation:

1) Randomly assign K centres.
2) Calculate the distance of all the points from all the K centres and allocate the points to cluster based on the shortest distance. The model's *inertia* is the mean squared distance between each instance and its closest centroid. The goal is to have a model with the Lowes intertia.
3) Once all the points are assigned to clusters, recompute the centroids.
4) Repeat the steps 2 and 3 until the locations of the centroids stop changing and the cluster allocation of the points becomes constant.

14. How is cluster quality measured?

Answer: Cluster quality is measured by taking the average silhouette coefficient value of all objects in the data set.
Formula while building model: silhouette_score(X,y_kmeans)]

15. What is cluster analysis and its types?

Answer: Clustering is an unsupervised approach which finds a structure/pattern in a collection of unlabelled data. A cluster is a collection of objects which are "similar" amongst themselves and are "dissimilar" to the objects belonging to a different cluster.

A group of data points would comprise together to form a cluster in which all the objects would belong to the same group.

Types of Cluster Analysis:

There are a number of different methods to perform cluster analysis. Some of them are,

* **Hierarchical clustering:** It is an alternative approach which does not need us to give the value of K beforehand and also, it creates a beautiful tree-based structure for visualization.

Centroid-based Clustering: In this type of clustering, clusters are represented by a central entity, which may or may not be a part of the given data set.

Distribution-based Clustering: It is a type of clustering model closely related to statistics based on the modals of distribution. Objects that belong to the same distribution are put into a single cluster.

Density-based Clustering: In this type of clustering, clusters are defined by the areas of density that are higher than the remaining of the data set.