# CLUSTER 2

1. Hierarchical clustering is a method of cluster analysis that builds a hierarchy of clusters by either merging smaller clusters into larger ones (agglomerative) or splitting larger clusters into smaller ones (divisive). It differs from other clustering techniques like K-means because it doesn't require specifying the number of clusters beforehand and provides a visual representation of the hierarchy.

2. The two main types of hierarchical clustering algorithms are:

Agglomerative: It starts with individual data points as separate clusters and iteratively merges them into larger clusters.

Divisive: It begins with all data points in one cluster and recursively splits them into smaller clusters.

3. The distance between two clusters in hierarchical clustering is determined by various distance metrics, such as Euclidean distance, Manhattan distance, or correlation distance. These metrics quantify the dissimilarity between data points or clusters.

4. Determining the optimal number of clusters in hierarchical clustering can be challenging. Common methods include visual inspection of dendrograms, where you look for significant changes in the tree structure, and quantitative methods like the Elbow method, Silhouette analysis, or Gap statistics.

5. Dendrograms are tree-like diagrams that visually represent the hierarchy of clusters in hierarchical clustering. They are useful for understanding the relationships between data points or clusters and help identify the optimal number of clusters by observing the structure of the tree.

6. Yes, hierarchical clustering can be used for both numerical and categorical data. Distance metrics differ for each type of data. For numerical data, Euclidean or Manhattan distances are common, while for categorical data, methods like Gower's distance or Jaccard index are used.

7. Hierarchical clustering can be used to identify outliers or anomalies by examining the distance between data points and their proximity to clusters. Data points that are far from all clusters or have unusual patterns in the hierarchy may be considered outliers.