1. What is the difference between supervised and unsupervised learning? Give some examples to illustrate your point.

Ans: **Supervised learning has a labelled data but in unsupervised learning we don’t have a labelled data and create clusters of the data based on the similarities between them. Supervised learning example is a model which predicts salary based on experience and education and the example of unsupervised learning is to make groups of people based on their salary and expenses to be able to take specific actions for specific clusters.**

1. Mention a few unsupervised learning applications.

Ans:

**NetFlix movies recommendations based on region, language or interest**

**YT videos recommendation based on genre of watched videos previously.**

**Amazon product recommendation based on previously bought items.**

1. What are the three main types of clustering methods? Briefly describe the characteristics of each.

Ans: **K-Means Clustering**: we initialised centroids and keep changing the position of centroids depending on the average of datapoints within a cluster. The best K value is determined by the elbow method and K-means ++ parameter is used to prevent the formation of incorrect clusters by initializing them afar.

**Hierarchical Clustering**: each point is considered as a cluster and by finding the nearest cluster we merge the clusters and move towards making bigger clusters. We find the best value of K or clusters with the help of dendrogram where we find the longest vertical line where no horizontal line passes through it.

**DBScan Clustering:** it is mainly used to keep the noise or outliers in a separate clusters so they can be identified and removed if they are not required to solve a problem statement.

1. Explain how the k-means algorithm determines the consistency of clustering.

Ans: **the consistency is determined by calculating the Euclidean distance between the data points and its centroids. Centroids also shift based on the average value of the cluster points.**

1. With a simple illustration, explain the key difference between the k-means and k-medoids algorithms.

Ans: **K means uses Euclidean distance to minimise the total squared error and uses the average/mean value of the points for centroid position selection so If you need to use eucldian distance then use k mean otherwise K medioids.**

1. What is a dendrogram, and how does it work? Explain how to do it.

Ans: **Dendogram in Hierarchial Means algorithm is used to find the optimal value of K. in the dendogram, we try to find the longest horizontal line where no horizontal passes through it. After passing such verical line, we draw a horizontal line to find the intersect points to be able to find the value of K**

1. What exactly is SSE? What role does it play in the k-means algorithm?

Ans: **SSE stands for sum of squared error. In K means- it finds the distance between centroid and other points within in the cluster to find the error.**

1. With a step-by-step algorithm, explain the k-means procedure.

Ans: **Initialise K centroids randomly.**

**Then find the nearest points to form a cluster. Use the eucldian distance**

**Then find the average value of the points within the cluster to replace the centroids**

**Keep iterating this process until the centroid stops moving.**

1. In the sense of hierarchical clustering, define the terms single link and complete link.

Ans: **in single link: we merge two clusters whose two closest members have the smallest distance. In complete link: each points create its own cluster and the clusters are sequentially combined until all the points come inside one cluster.**

1. How does the apriori concept aid in the reduction of measurement overhead in a business basket analysis? Give an example to demonstrate your point.

Ans: **Apriori concept allows the company to find the association between different elements whether it be two or items or most visited pages on the website. If they need to find the items which are bought together then they can place those items in the same shelf in the store to save customers’ time.**