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

 Supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not. In supervised learning, the algorithm “learns” from the training dataset by iteratively making predictions on the data and adjusting for the correct answer.

unsupervised tasks inclused clustering, visualization, dimensionality reduction , and association rule learning while supervised learning can be text classification problems

**2. Mention a few unsupervised learning applications.**

*data exploration,*

*customer segmentation,*

*recommender systems,*

*target marketing campaigns, and.*

*data preparation and visualization*

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

The major types of cluster analysis are Centroid Based/ Partition Clustering, Hierarchical Based Clustering, Distribution Based Clustering,

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

Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

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

The main difference between K-Means and K-Medoids is that K-Means will form clusters based on the distance of observations to each centroid, while K-Medoid forms clusters based on the distance to medoids.

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

A dendrogram is a diagram that shows the attribute distances between each pair of sequentially merged classes. To avoid crossing lines, the diagram is graphically arranged so that members of each pair of classes to be merged are neighbors in the diagram. The Dendrogram tool uses a hierarchical clustering algorithm.

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

The SSE is defined as **the sum of the squared Euclidean distances of each point to its closest centroid**. Since this is a measure of error, the objective of k-means is to try to minimize this value.

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

Step 1: Choose the number of clusters k.

Step 2: Select k random points from the data as centroids. ...

Step 3: Assign all the points to the closest cluster centroid. ...

Step 4: Recompute the centroids of newly formed clusters. ...

Step 5: Repeat steps 3 and 4.

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

In Single Linkage, the distance between two clusters is the minimum distance between members of the two clusters. In Complete Linkage, the distance between two clusters is the maximum distance between members of the two clusters.

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

It is an analyzing technique based on the idea that if we buy an item then we are bound to buy or not-buy a group (or single) items. For example, if a customer is buying bread then the chances of him/her buying jam is more. This is represented by the following equation: Association Mining Rule.