1. Movie Recommendation systems are an example of:

**ii) Clustering**

Options: **a) 2 Only**

1. Sentiment Analysis is an example of**:**

**i) Regression ii) Classification** **iv) Reinforcement**

Options: **d) 1, 2 and 4**

1. Can decision trees be used for performing clustering?

**a) True**

1. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:

**i) Capping and flooring of variables**

Options: **a) 1 only**

1. What is the minimum no. of variables/ features required to perform clustering?

**b) 1**

1. For two runs of K-Mean clustering is it expected to get same clustering results?

**b) No**

1. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
2. **Yes**
3. Which of the following can act as possible termination conditions in K-Means? **i)** For a fixed number of iterations. **ii)** Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum. **iii)** Centroids do not change between successive iterations. **iv)** Terminate when RSS falls below a threshold.

Options: **d) All of the above**

1. Which of the following algorithms is most sensitive to outliers?

**a) K-means clustering algorithm**

1. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning): **i)** Creating different models for different cluster groups. **ii)** Creating an input feature for cluster ids as an ordinal variable. **iii)** Creating an input feature for cluster centroids as a continuous variable. **iv)** Creating an input feature for cluster size as a continuous variable.

Options: **d) All of the above**

1. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset? **a)** Proximity function used **b)** of data points used **c)** of variables used Options**: d) All of the above**
2. **Is K sensitive to outliers?**

The K-means clustering algorithm is sensitive to outliers because a mean is easily influenced by extreme values. K-medoids clustering is a variant of K-means that is more robust to noises and outliers. Instead of using the mean point as the center of a cluster, K-medoids use an actual point in the cluster to represent it.

Example: -

Because the mean, as a statistic, is generally sensitive to outliers.

The mean of 2,2,2,3,3,3,4,4,42,2,2,3,3,3,4,4,4 is 33

If we add a single 2323 to that, the mean becomes 55, which is larger than *any* of the other values.

Since in k-means, you will be taking the mean a lot, you wind up with a lot of outlier-sensitive calculations.

That's why we have the k-medians algorithm. It just uses the median rather than the mean and is less sensitive to outliers.

1. **Why is K means better?**

k-means is one of the simplest algorithms which uses an unsupervised learning method to solve known clustering issues. It works well with large datasets.

Other clustering algorithms with better features tend to be more expensive. In this case, k-means becomes a great solution for pre-clustering, reducing the space into disjoint smaller sub-spaces where other clustering algorithms can be applied.

1. **Is K means a deterministic algorithm?**

K-Means is its non-deterministic nature. K-Means starts with a random set of data points as initial centroids. This random selection influences the quality of the resulting clusters.

K-Means is non-deterministic because running an algorithm on the same data over and over could provide different results. However, FCS Express performs k-means clustering using a deterministic method to ensure consistent results.