**Machine Learning**

1. Movie Recommendation systems are an example of:

Ans: b) 1 and 2

2. Sentiment Analysis is an example of:

Ans: d) 1, 2 and 4

3. Can decision trees be used for performing clustering?

Ans: a) True

4. 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

ii) Removal of outliers

Ans: a) 1 only

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

Ans: b) 1

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

Ans: b) No

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?

Ans: a) Yes

8. Which of the following can act as possible termination conditions in K-Means?

Ans: d) All of the above

9. Which of the following can act as possible termination conditions in K-Means?

Ans: d) 1 and 3

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

Ans: a) K-means clustering algorithm

11. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):

Ans: d) A of the above

12. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

Ans: d) All of the above

13. Is K sensitive to outliers?

Ans: 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 uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points. Mean is greatly influenced by the outlier and thus cannot represent the correct cluster center, while medoid is robust to the outlier and correctly represents the cluster center.

14. Why is K means better?

Ans: 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. Relatively simple to implement. Scales to large data sets, guarantees convergence. IT can also warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

15. Is K means a deterministic algorithm?

Ans: The basic k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results.

**SQL**

1. Which of the following constraint requires that there should not be duplicate entries?

Ans: D) Unique

2. Which of the following constraint allows null values in a column?

Ans: D) None of them

3. Which of the following statements are true regarding Primary Key?

Ans: B) There can be duplicate values in a primary key column

4. Which of the following statements are true regarding Unique Key?

Ans: D) All of the above

5. Which of the following is/are example of referential constraint?

Ans: B) Foreign Key

6. How many foreign keys are there in the Supplier table?

Ans: c) 2

7. The type of relationship between Supplier table and Product table is:

Ans: A) one to many

8. The type of relationship between Order table and Headquarter table is:

Ans: c) one to one

9. Which of the following is a foreign key in Delivery table?

Ans: a) delivery id

10. The number of foreign keys in order details is:

Ans: d) 2

11. The type of relationship between Order Detail table and Product table is:

Ans: b) many to one

12. DDL statements perform operation on which of the following database objects?

Ans: c) Table

13. Which of the following statement is used to enter rows in a table?

Ans: a) Insert Into

14. Which of the following is/are entity constraints in SQL?

Ans: b) Unique, c) Primary Key

15. Which of the following statements is an example of semantic Constraint?

Ans: A) A blood group can contain one of the following values - A, B, AB and O and

b) A blood group can only contain characters

**Statistics**

1. What represent a population parameter?

Ans: c) both

2. What will be median of following set of scores (18,6,12,10,15)?

Ans: c) 12

3. What is standard deviation?

Ans: d) All of the above

4. The intervals should be \_\_\_\_\_\_ in a grouped frequency distribution

Ans: c) Both of these.

5. What is the goal of descriptive statistics?

Ans: c) Analyzing and Interpreting

6. A set of data organized in a participant by variables format is called

Ans: b) Dataset

7. In multiple regression,\_\_\_\_\_\_\_ dependent variables are used

Ans: a) 2 or more

8. Which of the following is used when you want to visually examine the relationship between 2 quantitative variables?

Ans: B) Scatte Plot

9. Two or more groups means are compared by using

Ans: d) Analysis of Variance

10. \_\_\_\_\_\_\_is a raw score which has been transformed into standard deviation units?

Ans: a) z-score

11. \_\_\_\_\_\_\_is the value calculated when you want the arithmetic average?

Ans: c) Mean

12. Find the mean of these set of number (4,6,7,9,2000000)?

Ans: d) 400005.2

13. \_\_\_\_\_\_\_ is a measure of central tendency that takes into account the magnitude of scores?

Ans: d) Mean

14. \_\_\_\_\_\_ focuses on describing or explaining data whereas \_\_\_\_\_\_involves going beyond immediate data and making inferences

Ans: a) Descriptive and Inferences

15. What is the formula for range?

Ans: d) H-L