SQL

1. Which of the following constraint requires that there should not be duplicate entries?
   D) Unique
2. Which of the following constraint allows null values in a column?
   D) None of them
3. Which of the following statements are true regarding Primary Key?
   A) Each entry in the primary key uniquely identifies each entry or row in the table
4. Which of the following statements are true regarding Unique Key?
   D) All of the above
5. Which of the following is/are example of referential constraint?
   B) Foreign Key
6. How many foreign keys are there in the Supplier table?
   D) 1
7. The type of relationship between Supplier table and Product table is:
   A) one to many
8. The type of relationship between Order table and Headquarter table is:
   B) many to one
9. Which of the following is a foreign key in Delivery table?
   B) supplier id
10. The number of foreign keys in order details is:
    D) 2
11. The type of relationship between Order Detail table and Product table is:
    A) one to many
12. DDL statements perform operation on which of the following database objects?
    C) Table
13. Which of the following statement is used to enter rows in a table?
    A) Insert in to
14. Which of the following is/are entity constraints in SQL?

    B) Unique
    C) Primary Key
15. Which of the following statements is an example of semantic Constraint?
    A) A blood group can contain one of the following values - A, B, AB and O.
    B) A blood group can only contain characters

STATISTICS

1. What represent a population parameter?
   A) none
2. What will be median of following set of scores (18,6,12,10,15)?
   C) 12
3. What is standard deviation?
   C) The square root of the variance
4. The intervals should be in a grouped frequency distribution
   C) Both of these
5. What is the goal of descriptive statistics?
   B) Summarizing and explaining a specific set of data
6. A set of data organized in a participant by variables format is called
   B) Data set
7. In multiple regression, independent variables are used
   A) 2 or more
8. Which of the following is used when you want to visually examine the relationship between 2 quantitative variables?
   B) Scatterplot
9. Two or more groups means are compared by using
   D) Analysis of variance
10. _____is a raw score which has been transformed into standard deviation units?
    A) Z-score
11. is the value calculated when you want the arithmetic average?
    C) mean
12. Find the mean of these set of number (4,6,7,9,2000000)?
    D) 400005.2
13. _____is a measure of central tendency that takes into account the magnitude of scores?
    D) Mean
14. _____focuses on describing or explaining data whereas___involves going beyond immediate data and making inferences
    A) Descriptive and inferences
15. What is the formula for range?
    D) H-L

MACHINE LEARNING

1. Movie Recommendation systems are an example of:
   i) Classification
   ii) Clustering
   iii) Regression
   Options:
   a) 2 Only
2. Sentiment Analysis is an example of:
   i) Regression
   ii) Classification
   iii) Clustering
   iv) Reinforcement
   Options:
   d) 1, 2 and 4
3. Can decision trees be used for performing clustering?
   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
   Options:
   a) 1 only
5. What is the minimum no. of variables/ features required to perform clustering?
   b) 1
6. For two runs of K-Mean clustering is it expected to get same clustering results?
   b) No
7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   a) Yes
8. 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 witha 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
9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm

10. 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

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
    d) All of the above

12. Is K sensitive to outliers?
    Yes, K means clustering is sensitive to outliers. This is because a mean is easily influenced by extreme values. It can also be used for outlier detection.
13. Why is K means better?
    K means is simple to implement. This also tends for high convergence. This can take large data also.
14. Is K means a deterministic algorithm?
    No, it is not a deterministic algorithm. It takes random set of data points as centroids at first and then works on other data by adding data points.