### Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:
   1. Classification
   2. Clustering
   3. Regression Options:
      1. 2 Only
      2. 1 and 2
      3. 1 and 3
      4. 2 and 3

**Ans: a**

1. Sentiment Analysis is an example of:
   1. Regression
   2. Classification
   3. Clustering
   4. Reinforcement Options:
2. 1 Only
3. 1 and 2
4. 1 and 3
5. 1, 2 and 4

**Ans: d**

1. Can decision trees be used for performing clustering?
2. True
3. False

**Ans: 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:
   1. Capping and flooring of variables
   2. Removal of outliers Options:
      1. 1 only
      2. 2 only
      3. 1 and 2
      4. None of the above

**Ans: a)** **1 only**

1. What is the minimum no. of variables/ features required to perform clustering?
2. 0
3. 1
4. 2
5. 3

**Ans: b) 1**

1. For two runs of K-Mean clustering is it expected to get same clustering results?
2. Yes
3. No

**Ans: b) No**

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

1. Yes
2. No
3. Can't say
4. None of these

**Ans: a)Yes**

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

* 1. For a fixed number of iterations.
  2. Assignment of observations to clusters does not change between iterations. Except for cases witha bad local minimum.
  3. Centroids do not change between successive iterations.
  4. Terminate when RSS falls below a threshold. Options:
     1. 1, 3 and 4
     2. 1, 2 and 3
     3. 1, 2 and 4
     4. All of the above

**Ans: All of the above**

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

1. K-means clustering algorithm
2. K-medians clustering algorithm
3. K-modes clustering algorithm
4. K-medoids clustering algorithm

**Ans: a) K-means clustering algorithm**

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

* 1. Creating different models for different cluster groups.
  2. Creating an input feature for cluster ids as an ordinal variable.
  3. Creating an input feature for cluster centroids as a continuous variable.
  4. Creating an input feature for cluster size as a continuous variable. Options:

1. 1 only
2. 2 only
3. 3 and 4
4. All of the above

**Ans: 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?

1. Proximity function used
2. of data points used
3. of variables used
4. All of the above

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

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

1. Is K sensitive to outliers?

Ans: The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. ... The group of points in the right form a cluster, while the rightmost point is an outlier. 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.

1. Why is K means better?

Ans:Guarantees convergence. Can warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters. Guarantees convergence. Can warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

1. 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.

The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. ... The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.

ASSIGNMENT

# WORKSHEET 2 SQL

## Q1 to Q13 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following constraint requires that there should not be duplicate entries?
   1. No Duplicity B) Different

C) Null D) Unique

**Ans: D) Unique**

1. Which of the following constraint allows null values in a column?
   1. Primary key B) Empty Value

C) Null D) None of them

**Ans: D) None of them**

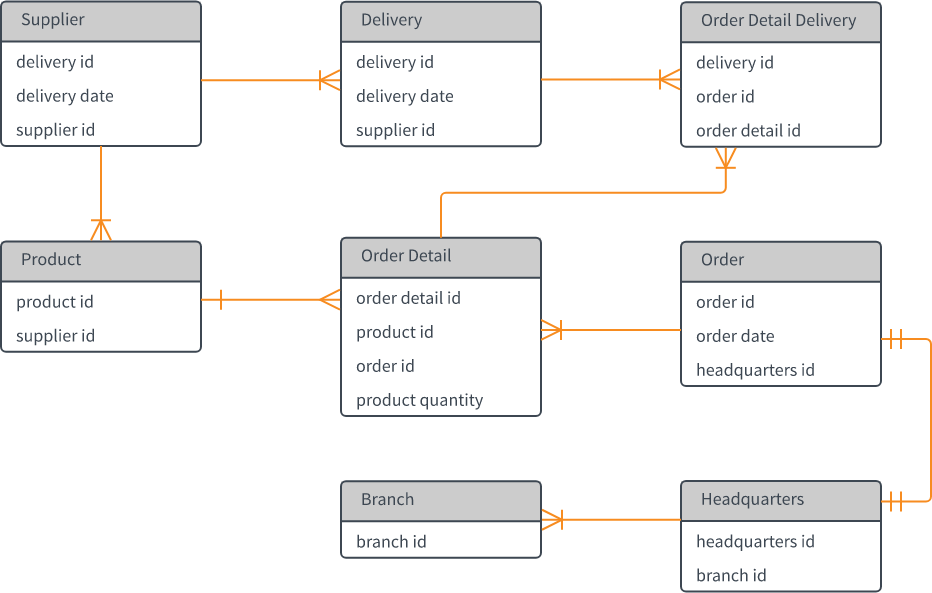
1. Which of the following statements are true regarding Primary Key?
   1. Each entry in the primary key uniquely identifies each entry or row in the table
   2. There can be duplicate values in a primary key column
   3. There can be null values in Primary key
   4. None of the above.

**Ans:A) Each entry in the primary key uniquely identifies each entry or row in the table**

1. Which of the following statements are true regarding Unique Key?
   1. There should not be any duplicate entries
   2. Null values are not allowed
   3. Multiple columns can make a single unique key together
   4. All of the above

**Ans: There should not be any duplicate entries**

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



* 1. Not Null B) Foreign Key

C) Referential key D) All of them

**Ans: B) Foreign Key**

## For Questions 6-13 refer to the below diagram and answer the questions:

1. How many foreign keys are there in the Supplier table?
   1. 0 B) 3

C) 2 D) 1

**Ans: C) 2**

1. The type of relationship between Supplier table and Product table is:
   1. one to many B) many to one

C) one to one D) many to many

**Ans: A)one to many**

1. The type of relationship between Order table and Headquarter table is:
   1. one to many B) many to one

C) one to one D) many to many

**Ans: D) many to many**

1. Which of the following is a foreign key in Delivery table?
   1. delivery id B) supplier id

C) delivery date D) None of them

**Ans: B) supplier id**

1. The number of foreign keys in order details is:
   1. 0 B) 1

C) 3 D) 2

**Ans: 1**

1. The type of relationship between Order Detail table and Product table is:
   1. one to many B) many to one

C) one to one D) many to many

**Ans: A) one to many**

1. DDL statements perform operation on which of the following database objects?
   1. Rows of table B) Columns of table

C) Table D) None of them

**Ans:C)Table**

1. Which of the following statement is used to enter rows in a table?
   1. Insert in to B) Update

C) Enter into D) Set Row

**Ans:A) Insert in to**

## Q14 and Q15 have one or more correct answer. Choose all the correct option to answer your question.

1. Which of the following is/are entity constraints in SQL?
   1. Duplicate B) Unique

C) Primary Key D) Null

**Ans: C) Primary Key**

1. Which of the following statements is an example of semantic Constraint?
   1. A blood group can contain one of the following values - A, B, AB and O.
   2. A blood group can only contain characters
   3. A blood group cannot have null values
   4. Two or more donors can have same blood group

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

# STATISTICS WORKSHEET-2

### Q1 to Q15 have only one correct answer. Choose the correct option to answer your question.

1. What represent a population parameter?
   1. SD
   2. mean
   3. both
   4. none

**Ans: B)mean**

1. What will be median of following set of scores (18,6,12,10,15)?
   1. 14
   2. 18
   3. 12
   4. 10

**Ans: C) 12**

1. What is standard deviation?
   1. An approximate indicator of how number vary from the mean
   2. A measure of variability
   3. The square root of the variance
   4. All of the above

**Ans: A) An approximate indicator of how number vary from the mean**

1. The intervals should be in a grouped frequency distribution
   1. Exhaustive
   2. Mutually exclusive
   3. Both of these
   4. None

**Ans: B) Mutually exclusive**

1. What is the goal of descriptive statistics?
   1. Monitoring and manipulating a specificdata
   2. Summarizing and explaining a specific set of data
   3. Analyzing and interpreting a set of data
   4. All of these

**Ans: B) Summarizing and explaining a specific set of data**

1. A set of data organized in a participant by variables format is called
   1. Data junk
   2. Data set
   3. Data view
   4. Data dodging

**Ans: C) Data Set**

1. In multiple regression, independent variables are used
   1. 2 or more
   2. 2
   3. 1
   4. 1 or more

**Ans: A) 2 or more**

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

* 1. Line graph
  2. Scatterplot
  3. Bar graph
  4. Pie graph

**Ans: F) Scatterplot**

1. Two or more groups means are compared by using
   1. analysis
   2. Data analysis
   3. Varied Variance analysis
   4. Analysis of variance

**Ans: A) analysis**

1. is a raw score which has been transformed into standard deviation units?
   1. Z-score
   2. t-score
   3. e-score
   4. SDU score

**Ans: Z-score**

1. is the value calculated when you want the arithmetic average?
   1. Median
   2. mode
   3. mean
   4. All

**Ans: C)mean**

1. Find the mean of these set of number (4,6,7,9,2000000)?
   1. 4
   2. 7

C) 7.5

D) 400005.2

**Ans: D) 400005.2**

1. is a measure of central tendency that takes into account the magnitude of scores?
   1. Range
   2. Mode
   3. Median
   4. Mean

**Ans: C) Median**

1. focuses on describing or explaining data whereas involves going beyond immediate data and making inferences
   1. Descriptive and inferences
   2. Mutually exclusive and mutually exhaustive properties
   3. Positive skew and negative skew
   4. Central tendency

**Ans:A) Descriptive and inferences**

1. What is the formula for range?
   1. H+L
   2. L-H
   3. LXH
   4. H-L

**Ans: L-H**