Assignment 2

Machine Learning

1. b) 1 and 2
2. d) 1, 2 and 4
3. a) True
4. a) 1 only
5. b) 1
6. b) No
7. a) Yes
8. d) All of the above
9. a) K-means clustering algorithm
10. d) All of the above
11. d) All of the above
12. The k-means algorithm is sensitive to the outliers. Such outliers can significantly influence the final cluster configuration and should be removed to obtain quality solutions.
13. Below points proves K means a better option:
    Relatively simple to implement
    Scales to large data sets
    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
14. 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. D) Unique
2. C) Null
3. A) Each entry in the primary key uniquely identifies each entry or row in the table
4. D) All of the above
5. B) Foreign Key
6. D) 1
7. D) many to many
8. B) many to one
9. B) supplier id
10. C) 3
11. D) many to many
12. C) Table
13. A) Insert in to
14. B) Unique C) Primary Key D) Null
15. B) A blood group can only contain characters and A) A blood group can contain one of the following values - A, B, AB and O.

Statistics

1. B) mean
2. C) 12
3. D) All of the above
4. C) Both of these
5. D) All of these
6. B) Data set
7. A) 2 or more
8. B) Scatterplot
9. D) Analysis of variance
10. A) Z-score
11. C) mean
12. D) 400005.2
13. D) Mean
14. A) Descriptive and inferences
15. D) H-L