Worksheet set 2 sql

1. **A) 2 Only** . Clustering only
2. **D) 1 , 2 and 4** . Regression , Classification and Reinforcement
3. **A) True.** Yes, decision trees can be used for performing clustering
4. **A) 1 only**. Capping and flooring is more desirable when the quantity of data points is less.
5. **B) 1.** Minimum no of features required to perform clustering is 1.
6. **B) No.** It usually runs on local minima and global minima so it is advised to run it multiple times.
7. **A) Yes**. After it has reached global minima it won't alter any iteration
8. **D) All the above.**
9. **A) K means clustering algorithm.** This is because it uses the mean of the data
10. **D) All of the above.**
11. **D) All of the above**

12. In K means clustering the data is clustered into various clustered centers depending upon the closeness of the data points. K means clustering thus becomes sensitive to outliers because outliers act as noise and therefore cause a loss in accuracy as the majority of the data is centered or clustered by taking a collective mean of the nearest data points. Therefore it is preferred to remove the outliers before performing this method.

13. K means it is better than many other alternatives because it is efficient. It centralizes the data and gives more importance to the majority data points instead of all the data points. This makes it highly efficient and it basically guarantees convergence. Also depending upon the data it has better adaptability and can create a different cluster without any issues. Thus it is highly preferred

14. Deterministic algorithm means a certain algorithm that with a particular given input will always give the same or similar kind of output. This non deterministic nature of k means algorithm , in which it adapts to the given input, centralizes various data points makes it a far better algorithm than other deterministic algorithm in ML.