Statistics worksheet-2	SQL worksheet-2	Machine Learning Assignment-2
Ans1) C	Ans1) D	Ans1) A
Ans2) C	Ans2) C	Ans2) D
Ans3) C	Ans3) A	Ans3) A
Ans4) C	Ans4) A	Ans4) A
Ans5) B	Ans5) B	Ans5) B
Ans6) B	Ans6) D	Ans6) B
Ans7) A	Ans7) C	Ans7) A
Ans8) B	Ans8) C	Ans8) D
Ans9) D	Ans9) B	Ans9) A
Ans10) A	Ans10) D	Ans10) D
Ans11) C	Ans11) C	Ans11) D
Ans12) D	Ans12) A	
Ans13) D	Ans13) A	
Ans14) A	Ans14) C and B	
Ans15) D	Ans15) A and B	

Machine Learning Assignment-2

Subjective Questions:

Ans12) Yes, k is sensitive to outliers as the k-means modify the centers of cluster by taking the average of all points related to data which are closer to each center of cluster. The all points of data are packed nicely it makes sense to average. But when you have outliers in the data, this may affect the average calculated of the whole cluster which in result push the center of cluster closer to outlier. As in k-means we take the mean sometimes the values are larger than any other values present in the data. Taking mean, winds up with a lot of outlier sensitive calculations.

Ans13) The k-means is better can be understand by its advantages that are:

It is simple to implement.

It scales to large data sets.

It guarantees convergence.

It can warm-start the positions of centroids.

It is easily adapts to new examples.

It generalizes to clusters of different shapes and sizes, such as elliptical clusters.

Ans14) The k-means is not a deterministic algorithm, it has only one drawback that is its non-deterministic nature. K-means comes up with the random selection of data points as initial centroids. The random selection of data points as initial centroids influences the quality of the resulting clusters. The several runs of the algorithm done on the dataset that is same produces output which is different with every single run done.