**STATISTICS WORKSHEET-2**

Answer No. 1: C

Answer No. 2: C

Answer No. 3: D

Answer No. 4: C

Answer No. 5: B

Answer No. 6: B

Answer No. 7: A

Answer No. 8: B

Answer No. 9: D (ANNOVA TEST)

Answer No. 10: A

Answer No. 11: C

Answer No. 12: D

Answer No. 13: C

Answer No. 14: A

Answer No. 15: D

**WORKSHEET 2 SQL**

Answer No. 1: D

Answer No. 2: A

Answer No. 3: B

Answer No. 4: C

Answer No. 5: B

Answer No. 6: D

Answer No. 7: A

Answer No. 8: C

Answer No. 9: B

Answer No. 10: D

Answer No. 11: B

Answer No. 12: C

Answer No. 13: A

Answer No. 14: B & C (UNIQUE, PRIMARY KEY)

Answer No. 15:

**MACHINE LEARNING**

Answer No. 1: B

Answer No. 2: D

Answer No. 3: A

Answer No. 4: A (Removal of outliers is not necessary if data points are few in number and in such scenario capping and flouring of variables is the most appropriate strategy).

Answer No. 5: B

Answer No. 6: B

Answer No. 7: A

Answer No. 8: D

Answer No. 9: D

Answer No. 10: A

Answer No. 11: D

Answer No. 12: D

Answer No. 13: The k-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values.

The k-means algorithm updates the cluster centers by taking the average of all the data points that are closer to each cluster center. When all the points are packed nicely together, the average makes sense. However, when you have outliers, this can affect the average calculation of the whole cluster. As a result, this will push your cluster center closer to the outlier. There are other clustering algorithms out there that are less sensitive to outliers.

Answer No. 14: k-means is one of the simplest algorithms which uses unsupervised learning method to solve known clustering issues. Other clustering algorithms with better features tend to be more expensive. In this case, k-means becomes a great solution for pre-clustering, reducing the space into disjoint smaller sub-spaces where other clustering algorithms can be applied. It works really well with large datasets but cannot be used in real world problems as it misses out information which it doesn’t understand.

Answer No. 15: 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. However, to ensure consistent results, *k*-means clustering using a deterministic method.