STATISTICS WORKSHEET-1

1. a
2. a
3. b
4. d
5. c
6. b
7. b
8. a
9. c
10. Normal Distribution is a bell like curve, which is always symmetrical at the mean. The shape of a normal distribution curve is determined by the mean and standard deviation. For normal distribution, mean is located in the horizontal axis and the standard deviation defines the spread of normal distribution.
11. Missing data can be seen through heat map. If the missing data rows are comparably very less to the total number of rows in the dataframe, the missing data rows can be dropped. If the missing data rows are more in number, the missing values can be replaced with mean or median or mode of that column. If the missing data is numerical, then it is replaced by mean or median of that column. If the missing value is string value, then it is replaced my mode of that column.
12. A/B Testing is type of a split testing which is a marketing experiment wherein you split your audience to test a number of variations of a campaign and determines which performs better. Here, we will show version A marketing content to half of your audience, and version B to the other half.
13. Generally, the mean imputation is not a good practice. If the independent variable and dependent variable are highly correlated to each other, by replacing the mean at the place of missing value will make the correlation nearly zero. This can significantly reduce the model's accuracy and bias the results.
14. Linear Regression is a linear model where the model assumes a linear relationship between the independent and dependent variable. After observing all the data points, a regression line is made across the data

points having least mean square error from the point to the regression line. With the help of this regression line, we can predict the output with given independent variable. The equation of this line is Y=mX+c where Y is the dependent variable, X is the independent variable, m is the slope of the line and c is the intercept.

15. The various branches of statistics are descriptive and inferential statistics.

- The descriptive statistics deals with the important characteristics of the data using the measures of central tendency like mean/median/mode and the measures of dispersion like range/standard deviation/variance.
- The inferential statistics is about using sample data from a large population and then making inferences from that sample. The goal of the inferential statistics is to draw conclusions from a sample and generalize them to the total population by using methodologies like hypothesis tests and by using probability theory.