ASSIGNMENT – 4 STATISTICS

1. d) All of the mentioned
2. a) Discrete
3. a) pdf
4. c) mean
5. c) empirical mean
6. b) standard deviation
7. c) 0 and 1
8. b) bootstrap
9. a) Frequency
10. A boxplot is a graphical representation of numerical data that displays a summary of the distribution of the data. It shows the median, quartiles, range, and outliers in the data. A histogram is a graphical representation of numerical data that displays the data as bars of various heights. It displays the frequency of the data, showing how many data points fall in each interval.
11. 1. Identify the goals of your project. What are the specific objectives you are trying to achieve? 2. Identify the key performance indicators (KPIs) associated with those objectives. These will be the metrics you will measure to gauge success. 3. Evaluate each KPI to determine how it can help you measure progress towards your objectives. 4. Choose the metrics that are most meaningful and relevant to your project. 5. Track your metrics regularly to get a better understanding of how your project is progressing. 6. Make sure to track the data over time to spot trends and make adjustments to your project as needed.
12. To assess the statistical significance of an insight, you should first calculate the p-value for the given insight. The p-value shows the probability that the observed effect could have occurred by chance. If the p-value is less than 0.05, then the insight is considered statistically significant. The lower the p-value, the more statistically significant the insight. Additionally, you should consider the sample size of the study and the confidence interval of the results. These can give you an indication of how reliable the results are.
13. 1. Exponential Distribution: This distribution is often used to model the time between events, such as the time between customer purchases. 2. Poisson Distribution: This distribution is used to model the number of events that occur within a certain time period, such as the number of customer inquiries

in a day. 3. Gamma Distribution: This distribution is used to model the total duration of events, such as the total duration of customer calls. 4. Weibull Distribution: This distribution is used to model the time to failure of a product, such as the time until a product breaks down.

14. An example where the median is a better measure than the mean is when there are a few extreme values that could skew the data. For example, if you were looking at the salaries of a group of workers, and one of the workers was a CEO making an extremely high salary compared to the rest of the workers, the mean salary would be much higher than the median salary. The median salary would be a better measure of the typical salary of the group, while the mean would be heavily affected by the extreme value.

15. The likelihood is a measure of how likely it is that an event will occur. It is usually expressed as a probability, which is a number between 0 and 1. A likelihood of 1 means that the event is certain to occur, while a likelihood of 0 means that the event is impossible.