



STATISTICS WORKSHEET 6

1. A,B
2. A
3. A
4. C
5. C
6. A
7. C
8. B
9. B
10. A boxplot and histogram are both graphical representations of data. A boxplot is a summary of a set of data containing the minimum, first quartile, median, third quartile, and maximum. A histogram is a graphical representation of the distribution of numerical data. It shows the number of observations within each given interval or bin
11. To select metrics, we need to understand what the problem is we are trying to solve, what are the questions we need to answer, and what are the goals we want to achieve. Then we need to identify the metrics that are directly related to solving the problem and answering the questions. After that we

will need to evaluate the metrics based on their relevance, accuracy, and ease of interpretation.

- 12. To assess the statistical significance of an insight, one can use statistical hypothesis testing. This involves formulating a null hypothesis and an alternative hypothesis and then using a test statistic and p-value to determine whether the insight is statistically significant or not.**
- 13. Examples of data that does not have a Gaussian distribution, nor log-normal are:**
 - Exponential Distribution**
 - Poisson Distribution**
 - Bernoulli Distribution**
 - Uniform Distribution**
- 14. An example where median is a better measure than mean is when the data has outliers. In such cases, the mean can be skewed by the outliers, while the median is not affected by them.**
- 15. Likelihood is a way to measure how well a set of parameters fits a given set of observations. It is a function of the parameters of the model and is used to estimate the parameters that are most likely to have generated the observed data. The higher the likelihood, the more likely the observed data is to have been generated by the given set of parameters.**