# 10. What is the difference between a boxplot and histogram?

# Histograms and box plots are graphical representations for the frequency of numeric data values.

# Histograms are used when we have to determine the probability distribution of data, while boxplots are used when comparing two data sets/ columns.

11. How to select metrics?

Metrics must be selected depending on the objective of the problem.Based on Regression or Classification, we can select the metric to be used example r squared for regression and accuracy for classification.

12. How do you assess the statistical significance of an insight?

Statistical significance can be accessed using hypothesis testing.Stating a null hypothesis which is usually the opposite of what we wish to test (classifiers A and B perform equivalently, Treatment A is equal of treatment B)

Then, we choose a suitable statistical test and statistics used to reject the null hypothesis

Also, we choose a critical region for the statistics to lie in that is extreme enough for the null hypothesis to be rejected (p-value)

We calculate the observed test statistics from the data and check whether it lies in the critical region  
Common tests:  
One sample Z test  
Two-sample Z test  
One sample t-test  
paired t-test  
Chi-squared test   
Anova

If the p-value is less than the alpha, We will reject the null in other words, the result is statistically significant.

13. Give examples of data that does not have a Gaussian distribution, nor log-normal.

Life data analysis(helps to measure time to failure rate).

14. Give an example where the median is a better measure than the mean.

Median tends to be better measure of central tendency when mean is being skewed by some large data points.

If the score of students in a class are 1,2,3,4,20  
So if we calculate the mean =1+2+3+4+20​/5=30/5​=6  
Median =3  
So, median is better or appropriate measure because 20 is much greater than other numbers and because of 20 the mean has come out to 6.

∴ Its better to take median than mean.

15. What is the Likelihood?

Likelihood can be defined as a measure of how well a statistical model fits to a sample data