**Question: 1** In one state, 52% of the voters are BJP, and 48% are Congress. In a second state, 47% of the voters are BJP, and 53% are Congress. Suppose simple random samples of 100 voters are surveyed from each state. What is the probability that the survey will show a greater percentage of BJP voters in the second state than in the first state?

**Question: 3** Name the Tool Techniques discussed in the Session for Data Handling w.r.t Data Science.

**Question: 4** Frame in your Words Why we use Sampling and describe their Types based on Session Discussion? (Not more than 100 Words).

**Question: 5** Can we apply a Machine Learning Algorithm directly with Raw Data? Yes / No. Justify (100 words)?

**Question: 6** Military Soldier records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of the 6 randomly selected Soldiers, 4 will recover?

**Question: 7:** Ten percent of computer parts produced by a certain supplier are defective. What is the probability that a sample of 10 parts contains more than 3 defective ones?

**Question: 8:** What is the difference between supervised and unsupervised machine learning?( 50 Words)

**Answer:** Supervised learning is a data mining task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object and the desired output value

**Question: 9:** What is Bayes’ Theorem? How is it useful in a machine learning context?

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

**Question: 10:** What is Normal Distribution? Write the Normal Distribution as a Mathematical Function?

* What’s the Z Value of getting a math SAT score of 575 or less, μ=500 and σ=50?
* Explain Process to check that the Data is Normally Distributed?

**Question 11:** Write differences between Descriptive and Inferential Statistics?

**Question 12:** 5000 orders received in the last month.3000 (60%) were from new customers.2000 (40%) were from old customers. So it looks like the new customers are doing better. Identify the Catch here?

**Question 13:** What is Data Engineering .What Role does Datatype Play in Data Analysis. Comment in not more than 100 words.

**Question 14:** Describe the impact of Outliers in a Data Set. What can be the potential impacts on the analysis? (Limit- 100 Words)

## Question 15: Does all data have a median, mode and mean? Comment in not more than 100 words.

## Answer: We use statistics such as the [mean](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#mean), [median](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#median) and [mode](http://www.mathsteacher.com.au/year8/ch17_stat/02_mean/mean.htm#mode) to obtain information about a [population](http://www.mathsteacher.com.au/year8/ch17_stat/01_data/surv.htm#pop) from our [sample](http://www.mathsteacher.com.au/year8/ch17_stat/01_data/surv.htm#sample) set of observed values.

## The mean (or average) of a set of data values is the [sum](http://www.mathsteacher.com.au/year8/ch01_arithmetic/01_basic/oper.htm#add) of all of the data values divided by the number of data values. The "median" is the "middle" value in the list of numbers.

## The median of a set of data values is the middle value of the data set when it has been arranged in ascending order.  That is, from the smallest value to the highest value.

## The mode of a set of data values is the value(s) that occurs most often. If no number in the list is repeated, then there is no mode for the list.