

# Understanding the Statistic of data science

- Introduction to Statistics
- Measures of Central Tendency I
- Measures of Central Tendency II
- Understanding the spread of data
- Data Distribution
- Introduction to Probability
- Probabilities of Discreet and Continuous Variables
- Central Limit Theorem and Normal Distribution I
- Central Limit Theorem and Normal Distribution II



VINAY

LEARN CONTRIBUTE CONQUER



VINAY

LEARN CONTRIBUTE CONQUER





VINAY

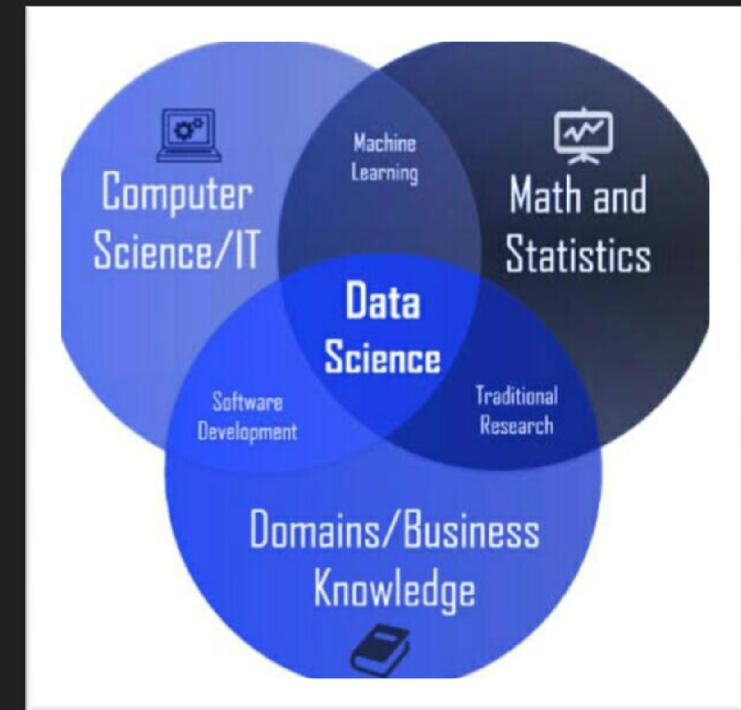
LEARN CONTRIBUTE CONQUER

# Introduction to statistic :-

Data science is using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions.

In other word's

Data science is A blend of Various tools algorithm's In machines learning principles with the goal To discover Hidden pattern in raw Data... ..





VINAY

LEARN CONTRIBUTE CONQUER



- Build intelligence & ability machine .
  - Pereform a sentiment analysis To predicate the outcome Of elections
  - Predicate tha characteristics of high LTV costumers .
  - Its help in coustumer sedimentation.



VINAY

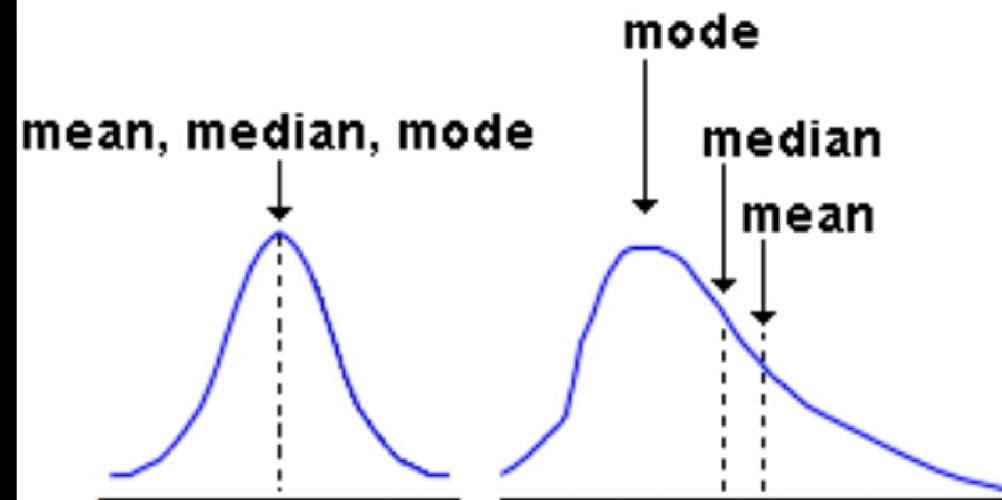
LEARN CONTRIBUTE CONQUER

# Measures of Central Tendency | :-

It represents the center point or typical value of a dataset.

The most values in a distribution fall and are also referred to as the central location of a distribution.

## Measures of Central Tendency

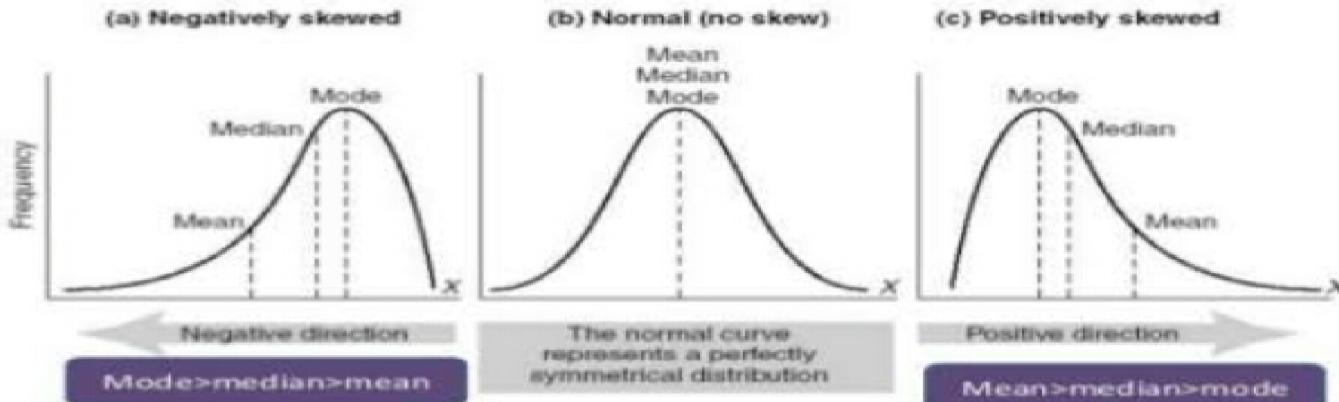




VINAY

LEARN CONTRIBUTE CONQUER

## Position of mean median mode



- The three most common measures of central tendency are the mean, median, and mode.
- These measures calculate the location of the central point using a different method.



# Mean, midean, mode :-

- **Mean**: Mean is the most common measure of central tendency. It is simply the sum of the numbers divided by the number of numbers in a set of data. This is also known as average.
- **Median**: Median is the number present in the middle when the numbers in a set of data are arranged in ascending or descending order. If the number of numbers in a data set is even, then the median is the mean of the two middle numbers.
- **Mode**: Mode is the value that occurs most frequently in a set of data.

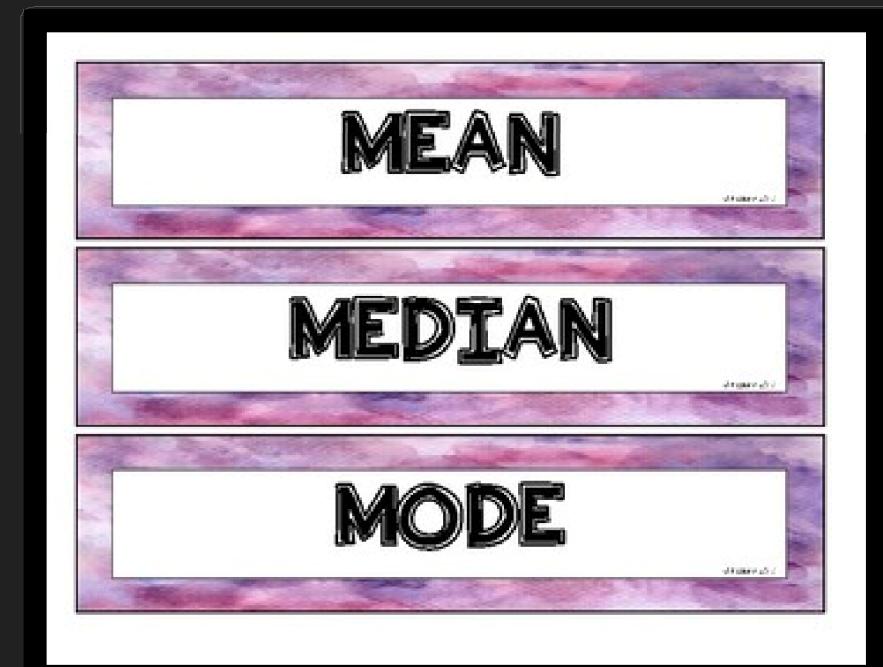


VINAY

LEARN CONTRIBUTE CONQUER

# Measures of Central Tendency II :-

- It is unique that a data set has one and only one mean.
- The value of the mean is influenced by all observations in the data set
- Measures of central tendency provide information about typical or average values of a data set.





# Type's of Mean :-

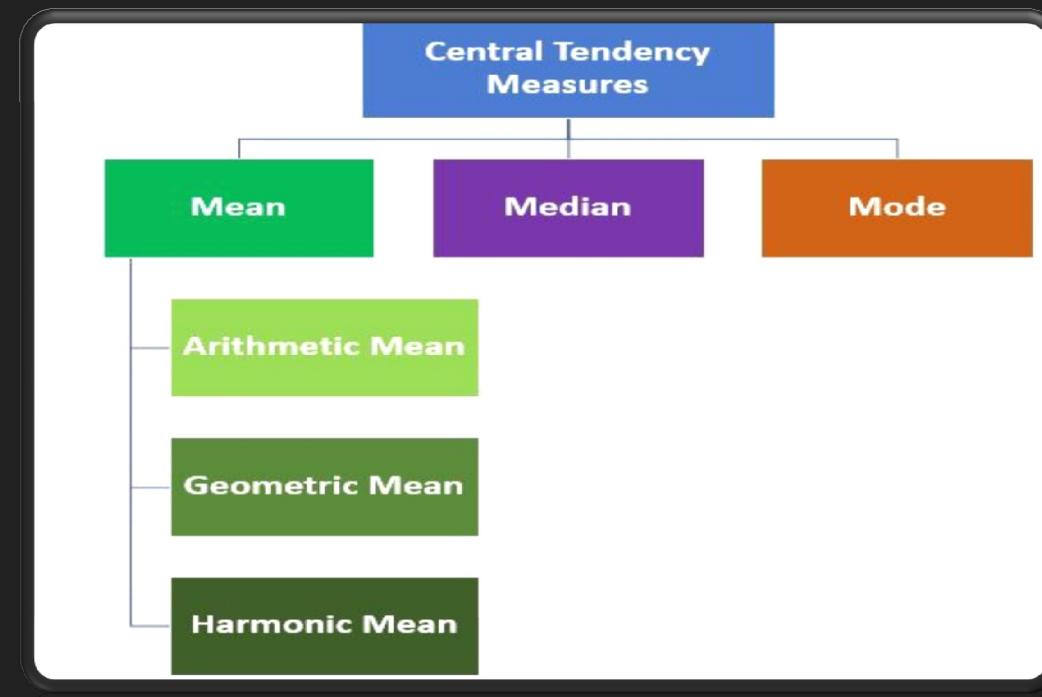
Mean is the most commonly used measure of central tendency . **Mean = Sum of the Given Data/Total number of Data )**

There are different types of mean

○ Arithmetic mean

○ Geometric Mean(GM)

○ Harmonic mean (HM)



# Understanding the spread of data:-

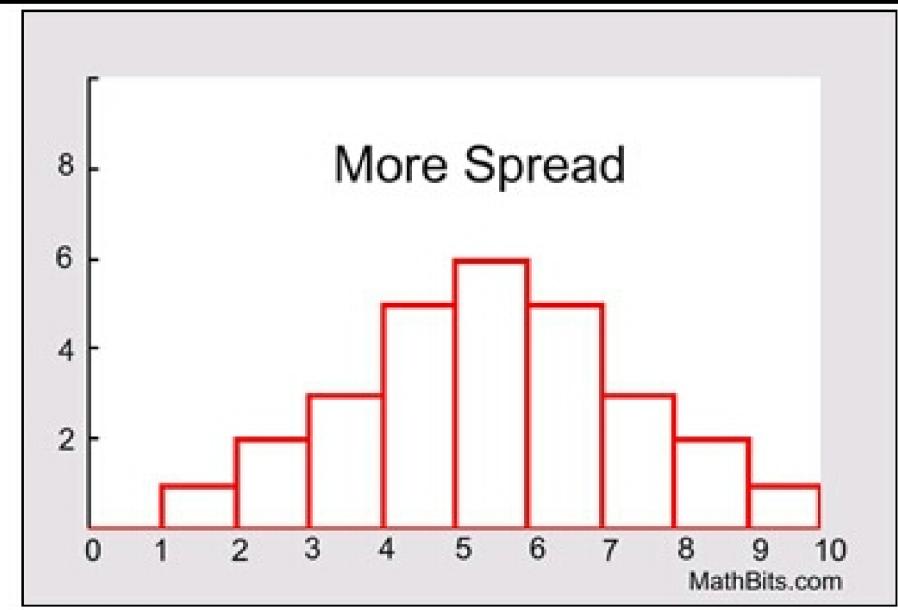
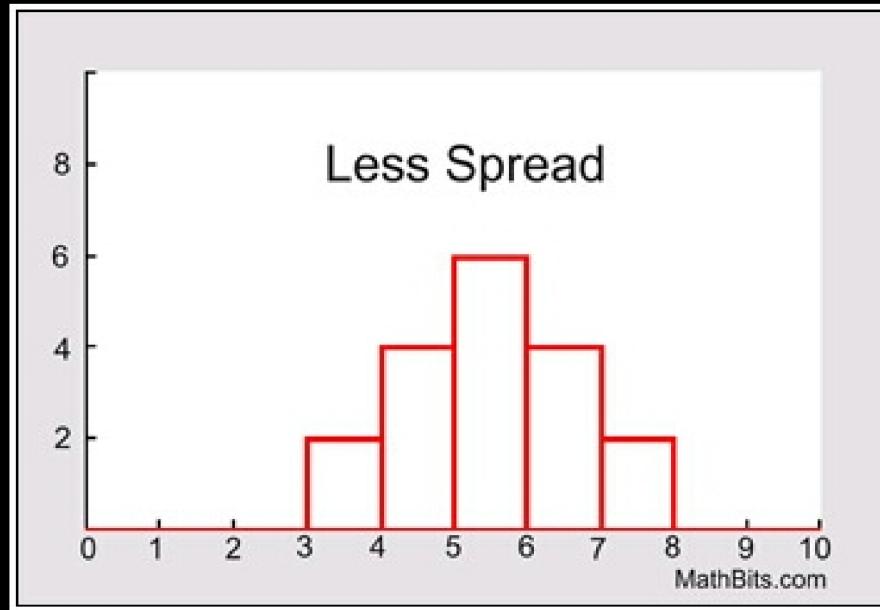
|    | A - Task Name                  | B - Start Dt   | C - End Dt | D        | E   | F - Status      | G         | H - Owner        |
|----|--------------------------------|----------------|------------|----------|-----|-----------------|-----------|------------------|
| 1  |                                |                |            |          |     |                 |           |                  |
| 2  | Project Name                   | [Project Name] |            |          |     | Start Date      | 1/1/2020  |                  |
| 3  | Project Manager                | Beth Bigidea   |            |          |     | Target End Date | 8/30/2020 |                  |
| 4  |                                |                |            |          |     |                 |           |                  |
| 5  | Task Name                      | Start Date     | End Date   | Duration | RYG | Status          | Priority  | Owner            |
| 6  | Product Plan                   | 1/1/2020       | 8/30/2020  | 242      | ●   | In Progress     | 🔴         | Beth Bigidea     |
| 7  | Product Use Case Definition    | 1/1/2020       | 6/1/2020   | 152      | ●   | Completed       | 🟡         | Beth Bigidea     |
| 8  | Business Requirements          | 1/1/2020       | 6/1/2020   | 152      | ●   | Not Started     | 🔴         | Roger Reports    |
| 9  | Launch Schedule                | 6/1/2020       | 8/30/2020  | 90       | ●   | On hold         | 🔴         | Crystal Codebase |
| 10 | Launch Event Plan              | 6/1/2020       | 8/30/2020  | 90       | ●   | Not Started     | 🟡         | Beth Bigidea     |
| 11 | Marketing Plan                 | 2/1/2020       | 8/1/2020   | 182      | ●   | In Progress     | 🟡         | Beth Bigidea     |
| 12 | Marketing Materials Plan       | 2/1/2020       | 8/1/2020   | 182      | ●   | In Progress     | 🟡         | Roger Reports    |
| 13 | Datasheets                     | 2/1/2020       | 4/1/2020   | 60       | ●   | Completed       | 🟡         | Peter Perfcycle  |
| 14 | ROI Whitepaper                 | 2/1/2020       | 4/1/2020   | 60       | ●   | In Progress     | 🟡         | Peter Perfcycle  |
| 15 | Website Design and Development | 5/1/2020       | 8/1/2020   | 92       | ●   | On hold         | 🟡         | Crystal Codebase |
| 16 | Web & Mobile Site Design       | 5/1/2020       | 7/1/2020   | 61       | ●   | Not Started     | 🟡         | Crystal Codebase |
| 17 | Web & Mobile Site Development  | 6/1/2020       | 8/1/2020   | 61       | ●   | Not Started     | 🟡         | Crystal Codebase |
| 18 | Web & Mobile Site Deployment   | 8/1/2020       | 8/30/2020  | 452      | ●   | In Progress     | 🟡         | Crystal Codebase |

- We can calculate spread in a variety of ways using different methods known as measures of spread.
- The spread in data can show us how much variation there is in the values of the data set



VINAY

LEARN CONTRIBUTE CONQUER



- The spread in data is the measure of how far the numbers in a data set are away from the mean or the median.
- If  $x$  is a number, then the difference " $x - \text{mean}$ " is called its deviation. Its deviations are used to calculate the standard deviation.



VINAY

LEARN CONTRIBUTE CONQUER



VINAY

LEARN CONTRIBUTE CONQUER

# Data Distribution :-

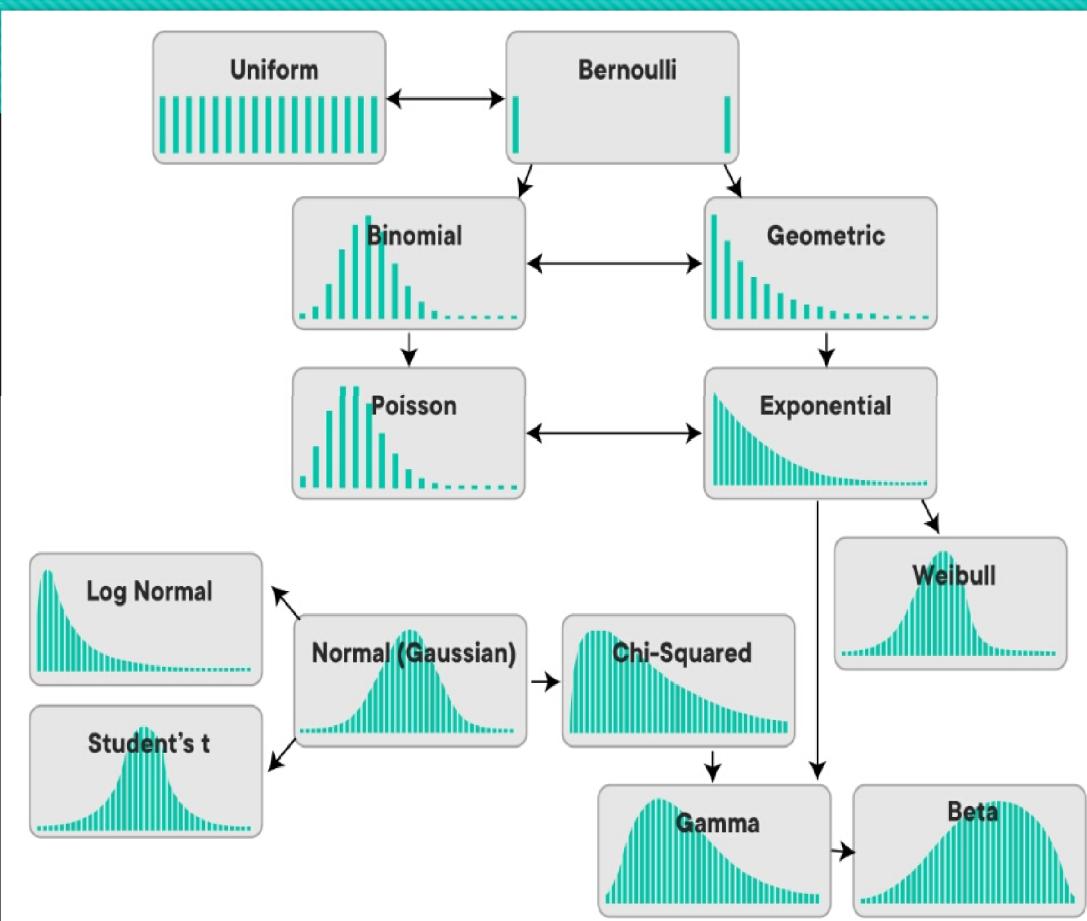
- A data distribution is a function or a listing which shows all the possible values (or intervals) of the data.
- the data in a distribution will be ordered from smallest to largest,





VINAY

LEARN CONTRIBUTE CONQUER



The function of a distribution that shows the density of the values of our data is called a probability density function .

Graphs and charts allow you to easily see both the values and the frequency.



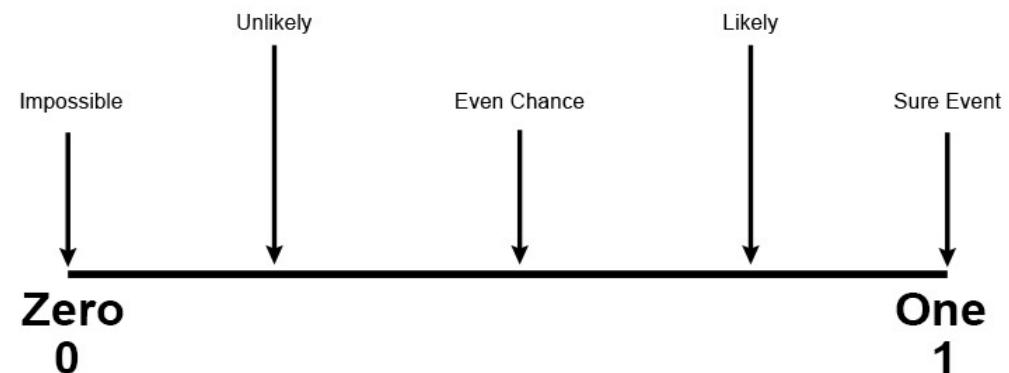
VINAY

LEARN CONTRIBUTE CONQUER

# Introduction to Probability :-

Probability is likel events are to happen. The probability of an event is a number between 0 and 1.

- The probability (P) that an event will happen is:
- $P = \frac{\text{Number of outcomes that will lead to that event}}{\text{Total number of possible outcomes}}$

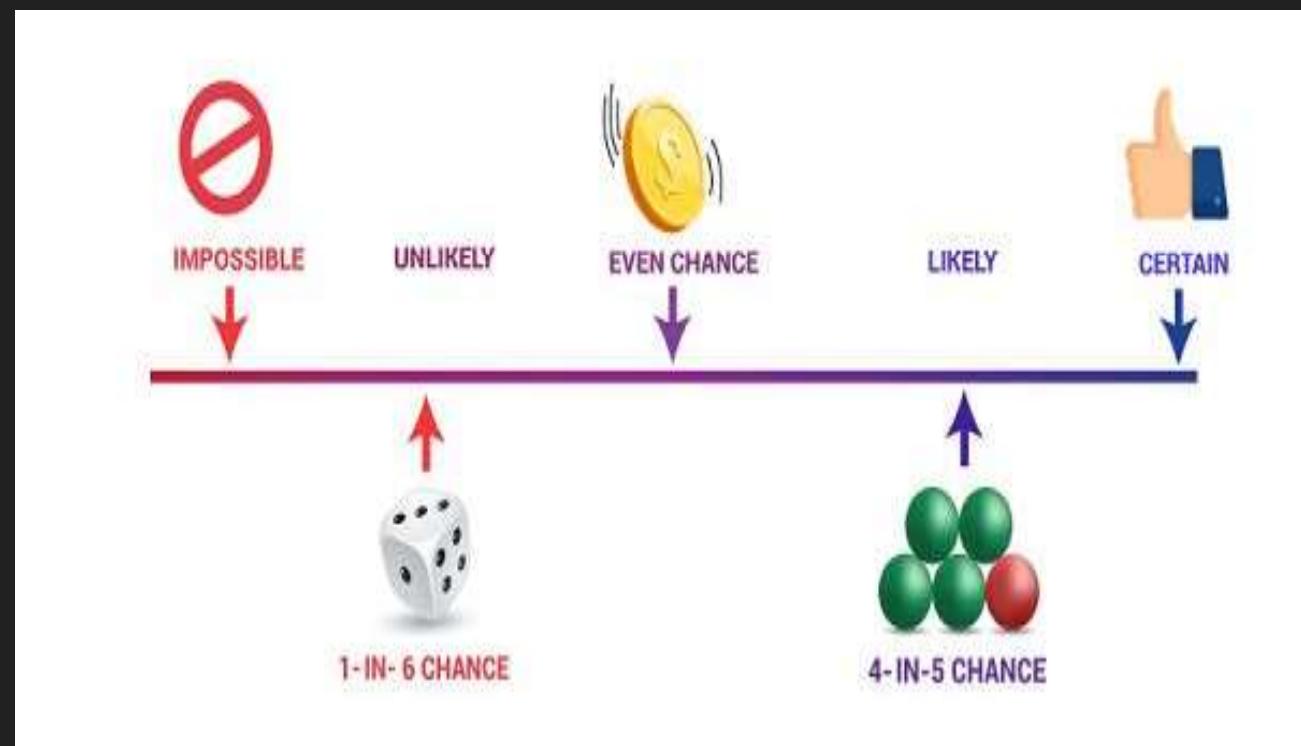




VINAY

LEARN CONTRIBUTE CONQUER

- There are three major types of probabilities:
- Theoretical Probability.
- Experimental Probability.
- Axiomatic Probability.



# Probabilities of Discrete and Continuous Variables:-



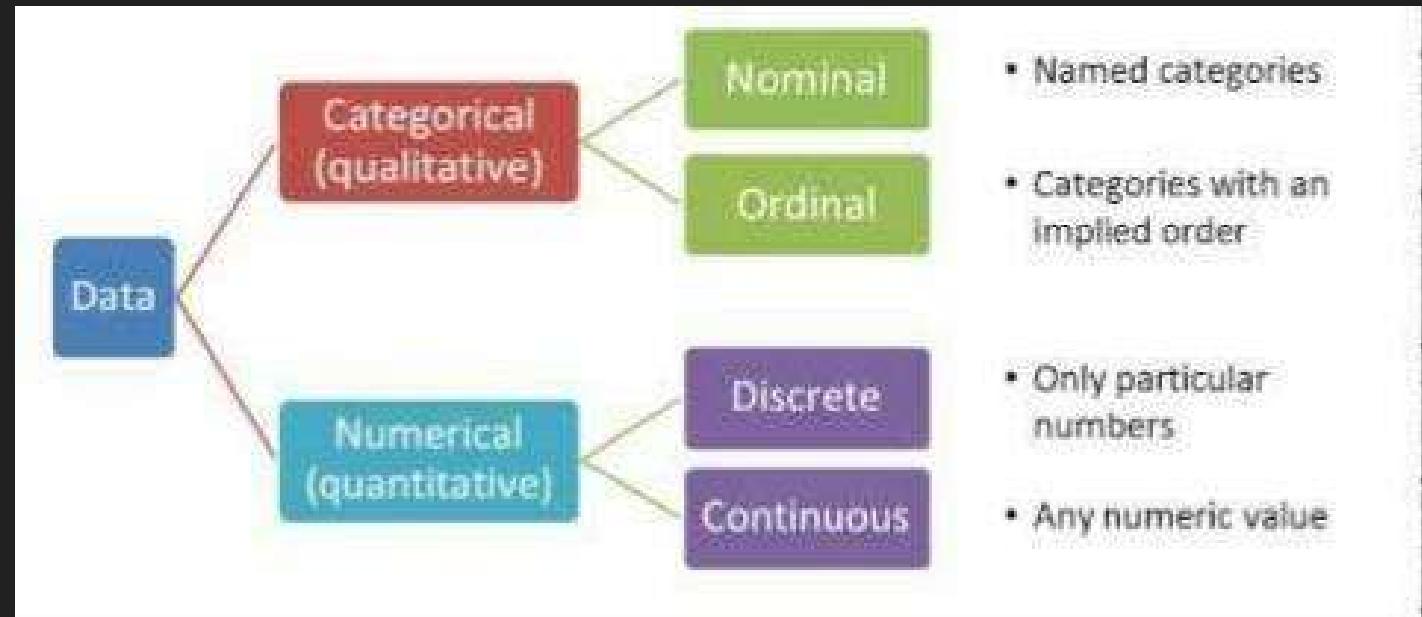
VINAY

LEARN CONTRIBUTE CONQUER

## Discrete:-

A set of data is said to be discrete if the values belonging to the set are distinct and separate.

countable  
nothing in between  
digital





VINAY

LEARN CONTRIBUTE CONQUER

## Continuous:-

A set of data is said to be continuous if the values belonging to the set can take on any value within a finite or infinite interval.

infinite  
always something between  
analog

Discrete  
Data

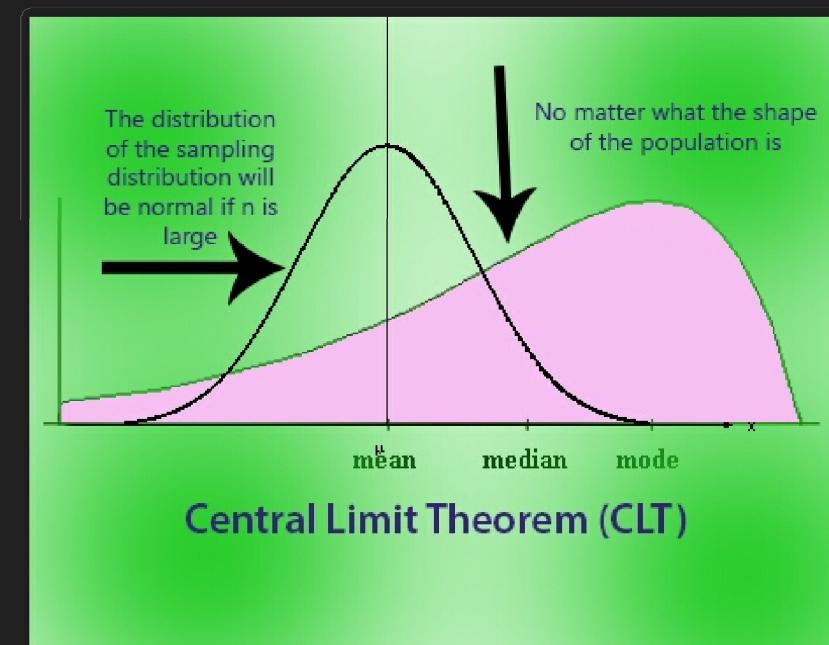
- 5 kids
- 96 workers
- 3 laptops

Continuous  
Data

- 3.25 kg
- 1.32 miles
- 7.25 inches

# Central Limit Theorem and Normal Distribution I :-

- The random variable  $X$  has a distribution (which may or may not be normal) with mean ( $\mu$ ) and standard deviation ( $\sigma$ ).
- Samples all of the same size  $n$  are randomly selected from the population of  $X$  values.



VINAY

LEARN CONTRIBUTE CONQUER



VINAY

LEARN CONTRIBUTE CONQUER



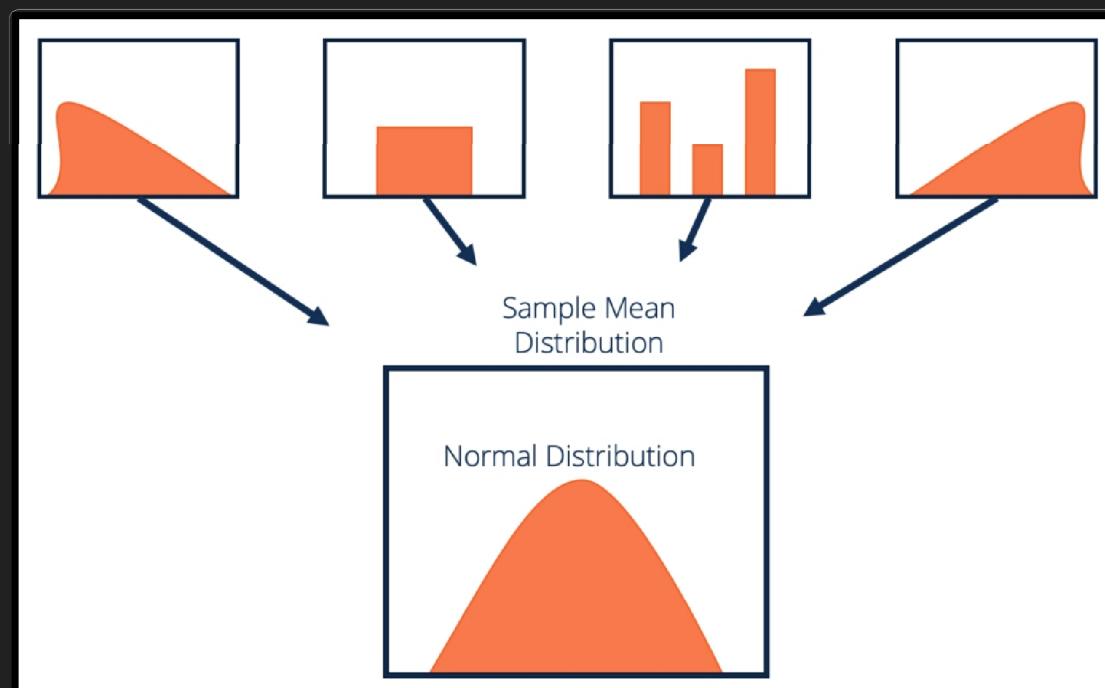
- In either case, the sampling distribution of sample means has a mean equal to the population mean
- The sampling distribution of sample means has a variance equal to  $1/n$  times the variance of the population and a standard deviation equal to the population standard deviation divided by the square root of  $n$ .

# Central Limit Theorem and Normal Distribution || :-



VINAY

LEARN CONTRIBUTE CONQUER



The mean of several draws from this distribution ( $T$ ) is

- random mean of( $\mu_e$ )
- standard deviation =  $\sigma / \sqrt{n}$
- normal for all samples if  $X$  is normal.



VINAY

LEARN CONTRIBUTE CONQUER



## -:Understanding the central limit theorem:-

$\mu$  is the population mean.  
 $\sigma$  is the population standard deviation.  
n is the sample size.

Central Limit Theorem exhibits a phenomenon where the average of the sample means and standard deviations equal the population mean and standard deviation.



VINAY

LEARN CONTRIBUTE CONQUER

# THANK - YOU

