

# ★ Statistics

- Data Analyst
  - Business Analyst
  - Data Scientist
  - Business Analyst
  - Data Scientist
  - Product Managers.
- Jobs

## Use Case :-

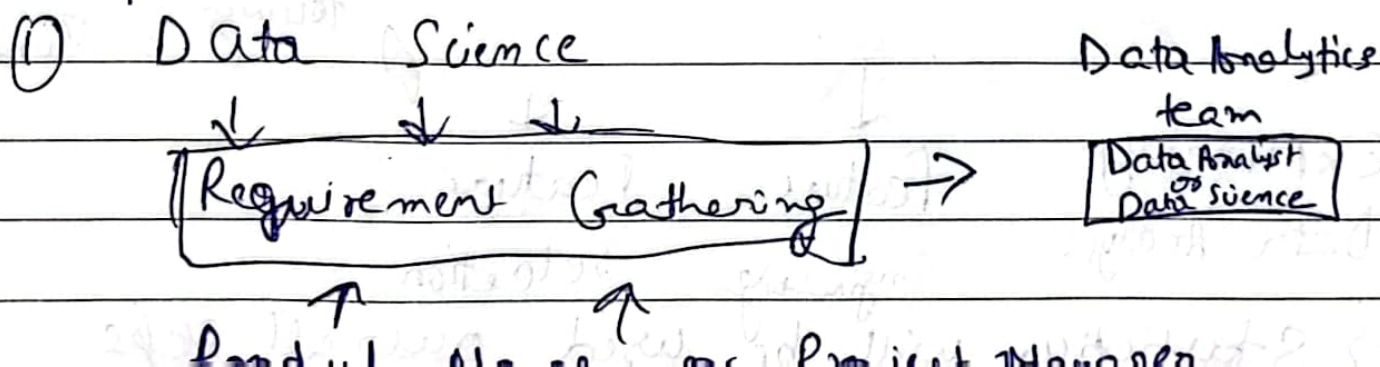
- ① ATM  $\square - 30km - \square - \square$
- ② Find the average size of the shark through  
- Out the word, blue whale
- ③ Amazon big billion day sale  
L w/c month should you select.

## Statistics

### Types of companies.

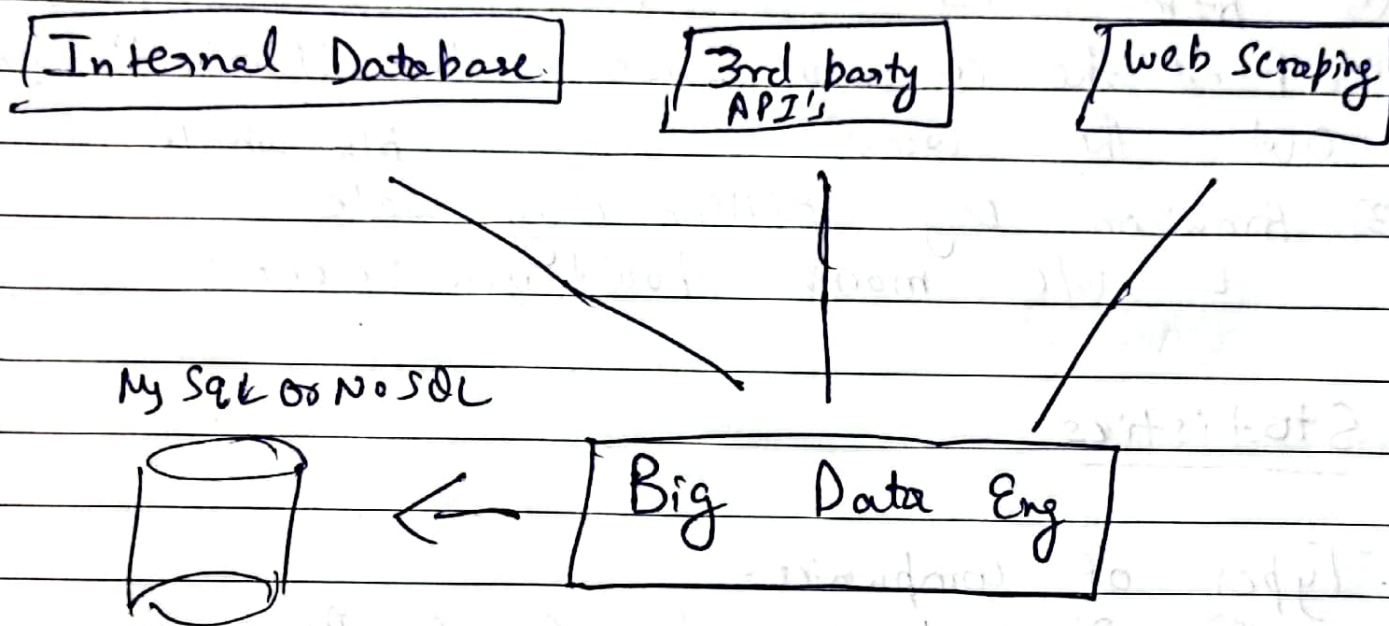
- ① Product based - Google, Apple, IBM
- ② Service based - HCL, Infosys

### Steps

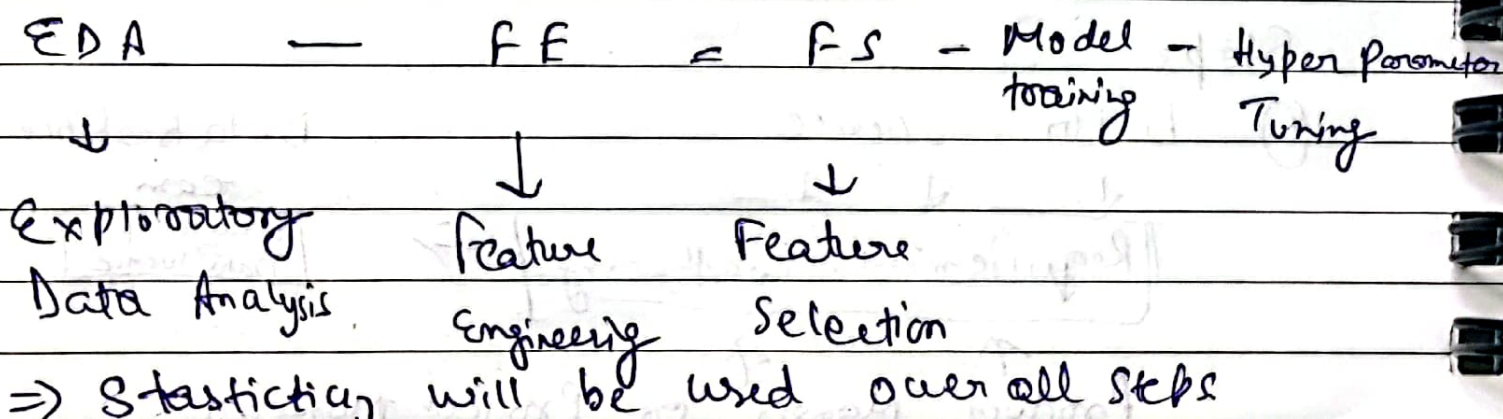


- Data Analytics team -
- ① Data Analyst
  - ② Data Scientist
  - ③ Big data Engineers
  - ④ Cloud Engineers

④ Cloud Engineers.



## Life cycle of DS Project





# Sampling Techniques

## ① Simple Random Sampling:

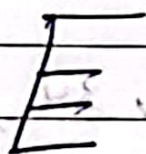
Every ~~one~~ member of the population ( $N$ ) has an ~~one~~ equal chance of being selected.  
for given sample ( $n$ )

## ② Stratified Sampling - when the population ( $N$ ) is split into non-overlapping groups (strata)

Gender  
    ↗ Male  
    ↘ Female

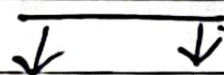
Education degree  
    ↗ High School  
    ↘ Masters  
    ↘ PhD

Blood Groups



Population

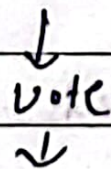
||



$< 18$

$\geq 18$

Not OK

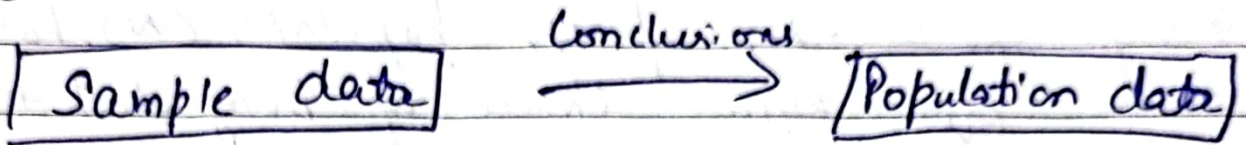


OK

Random Sampling

P-value

- ① z test
- ② T test
- ③ Chi square test
- ④ f test



By taking  $\uparrow$  conclusions of  $\uparrow$  - we make  $\uparrow$

★ let's say there are 20 classrooms in a university and you have collected the age of students in one classroom.

Age ( 21, 20, 18, 34, 17, 22, 24, 25, 24, 23, 22 )

Weight { - - - - - }

Descriptive Stats - what is the average age of students in the classroom?

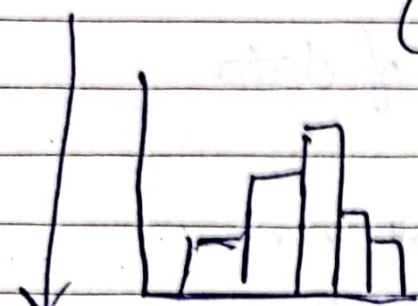
Relationship b/w Age & Gender?



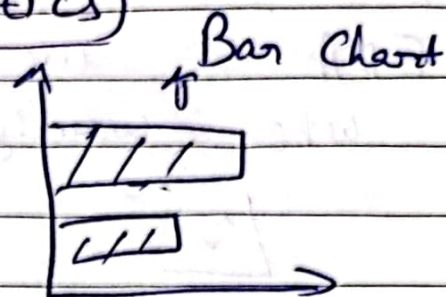
Questions (descriptive - stats) ->

1) what is avg. marks, avg no. of students passed, etc

(Descriptive Statistics)



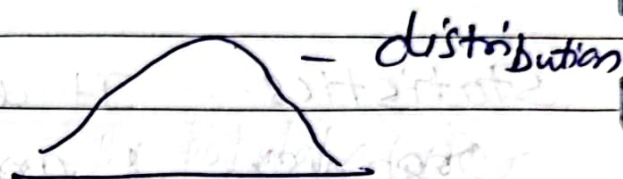
- Histogram



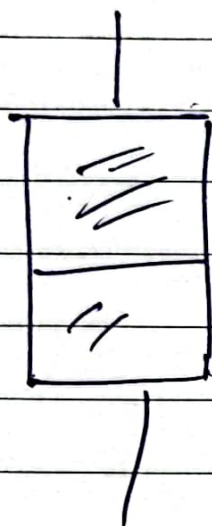
2) Percentage of students failed.



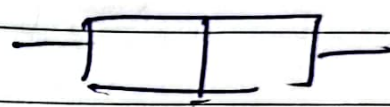
- Pie Chart



- distribution



- Candle stick



Box plot

Questions related to Inferential stats:-

University - 500 people

Class A -> 60 people

Sample data = Age = Average data

1) what is the relation b/w age of student in class with university students.

2) what % of students passed in whole class, not university

EDA - we do the Analysis of data when visualizing the data

 - Descriptive statistics

↓

Summarising the data (scatter diagram)

Statistics - It is the Science of collecting, organising & analysing the data.

12

## Statistics

↓  
Descriptive Stats

① Organising & summarising the data.

Inferential stats

② It consists of collecting sample data & making conclusion about population data.

↙  
{ Making conclusion using some experiments (by Hypothesis testing) }



(3) Systematic Sampling  $\rightarrow$  { Airport }

{ Credit Card }

Select every  $n^{\text{th}}$  individual out of population ( $N$ )

Quantitative Variable

Discrete variable

Continuous variable

Whole number

$\rightarrow$  Bank accounts

Height, weight, ages,

Rainfall

(1) What kind of variable is marital status?

Ans ~~Categorical~~ ~~Discrete~~ Variable  
Continuous

(2) What kind ... range ~~now~~ River length?

Ans Continuous

(3) What kind ... Movie length?

Ans Continuous.

(4) ... Pin code - Discrete

(5) ... ID? - Discrete