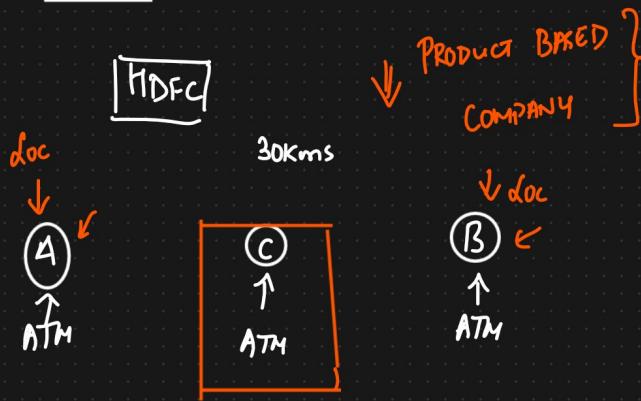


Statistics

Use case

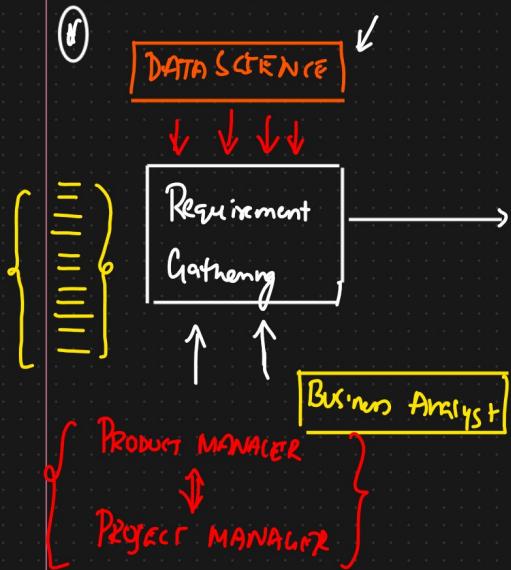


X Statistician → 5 years

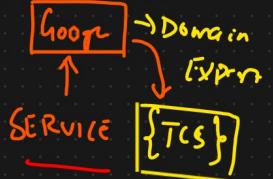


- ② Find the average size of the shark throughout the world?
- ③ Amazon Big Billion Day Sale {Intuit} → Which month should you select?

Statistics {Life cycle of DATA Scientific Project}



DATA ANALYST'S TEAM

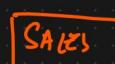


PRODUCT BASED

{ Google }

Domain knowledge

PRODUCT MANAGER
BA



YouTube, GPAY,
Google Ads, GMAIL

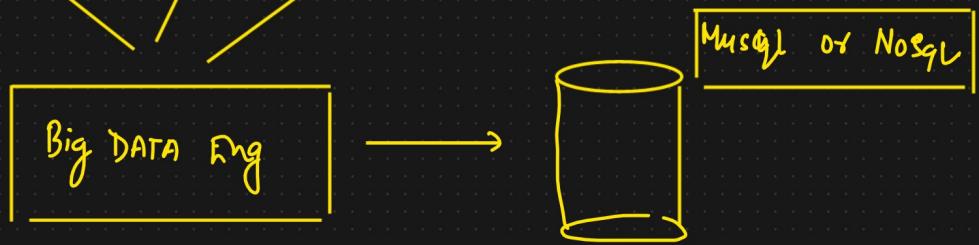
Domain Expertise

Product Manager

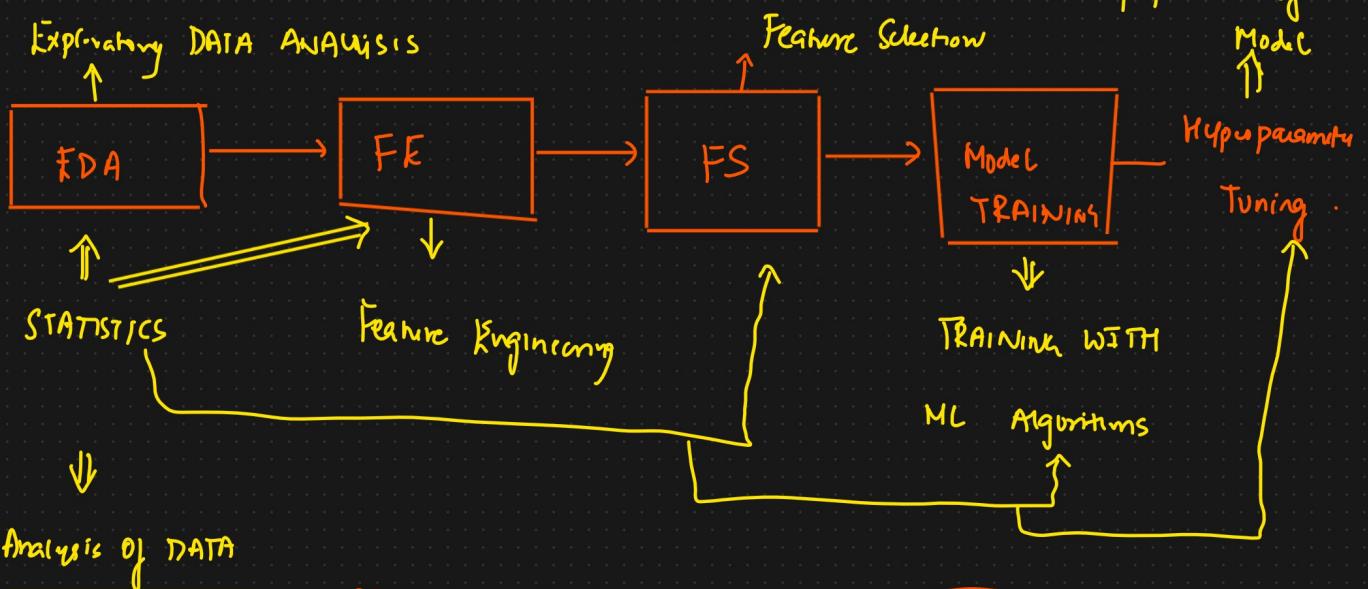
Internal
DATABASE

3rd party
API'S

Web Scraping



Life Cycle of DS Project



$$\text{Age} = \{12, 13, 14, 18, 20, 25\} \Rightarrow \text{Average Age} \Rightarrow \text{Measure of Central Tendency}$$

↓

DESCRIPTIVE STATS

Statistics = Defn : Statistics is the science of collecting, organising and analysing the data.

Data : "facts or pieces of information"

Eg: Ages of students in classroom

$$\{24, 25, 32, 29, 28\} \Rightarrow \text{Mean, Median, Mode}$$

Standard deviation

② Weights of students in classroom

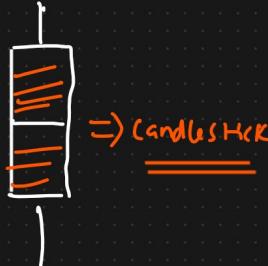
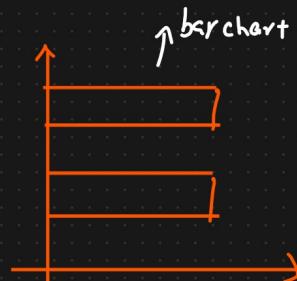
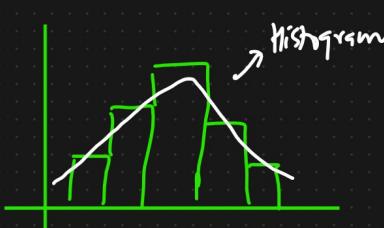
Descriptive Stats [EDA + FF]

Inferential Stats

① It consists of organising and summarizing the data.

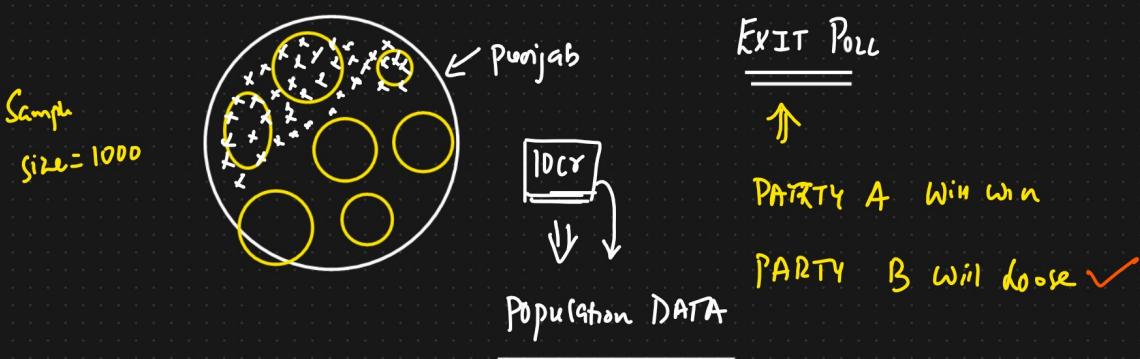
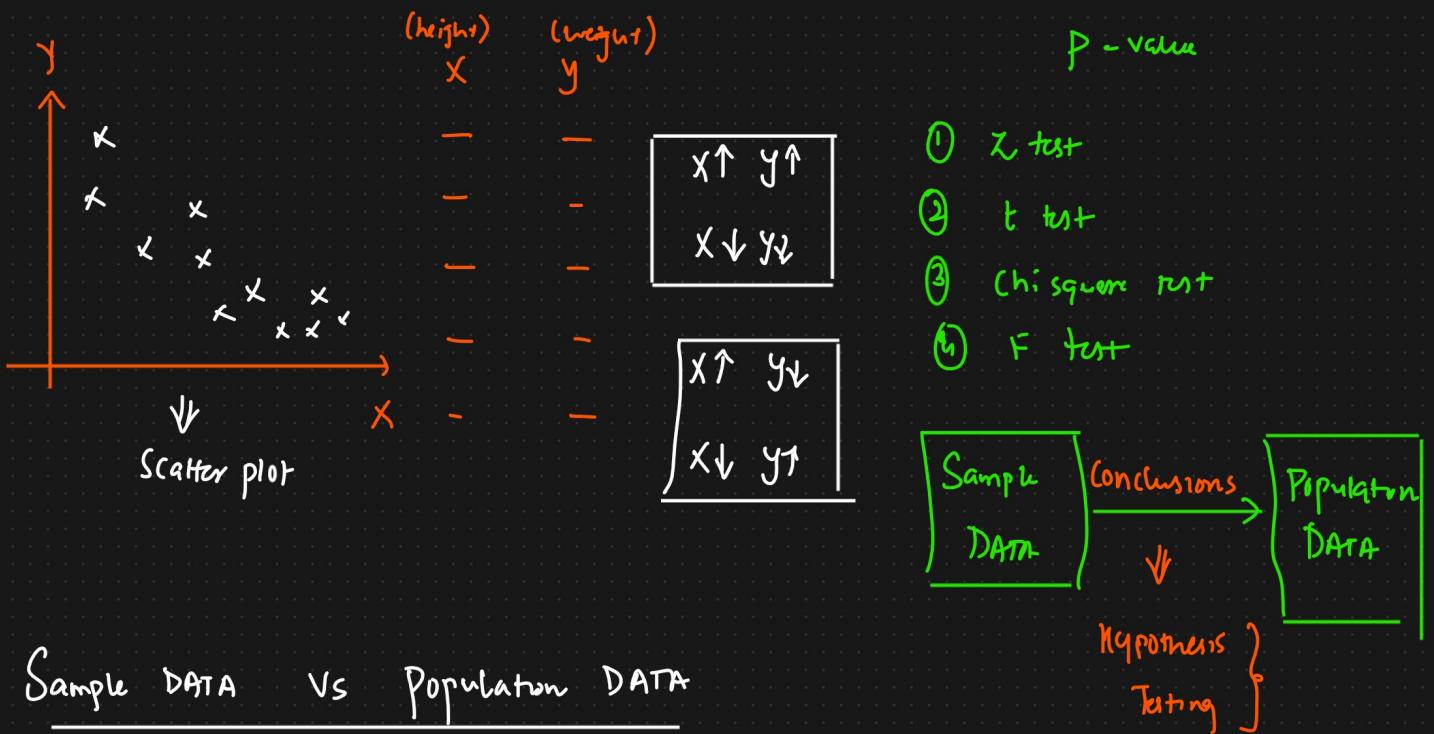
④ It consists of collecting sample data and making conclusion about population data using some experiments

Hypothesis Testing



Box plot

University → 500 people
CLASS A → 60 people
Sample data → Age → Average age of the entire university
Hypothesis Testing
C.I ⇒ Confidence Interval



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

Ages { 21, 20, 18, 34, 17, 22, 24, 25, 26, 23, 22 }

Weight { - - - - - }

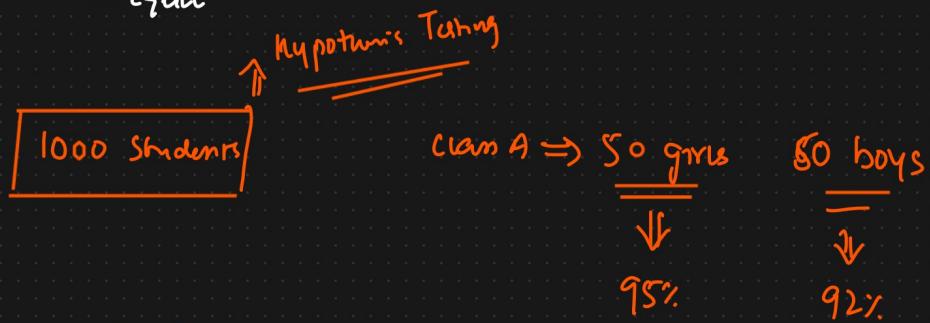
Descriptive Stats : What is the average age of students in the classroom?

Relationship between Age & Gender?

Inferential Stats : Are the average age of the students in the classroom

less than the average age of the students in the university?
↓

$\left\{ \begin{array}{l} \text{Greater} \\ \Downarrow \\ \text{Equal} \end{array} \right.$



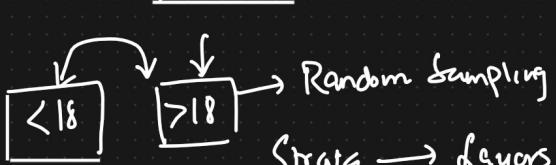
Choose a Sample
Sampling Techniques

Population (N) sample(n)

- ① Simple Random Sampling : Every member of the population (N) has an equal chance of being selected for your sample (n)



$n=1000$



Strata \rightarrow Layers \rightarrow Clusters \Rightarrow Groups

- ② Stratified Sampling

Gender

- Male
- Female

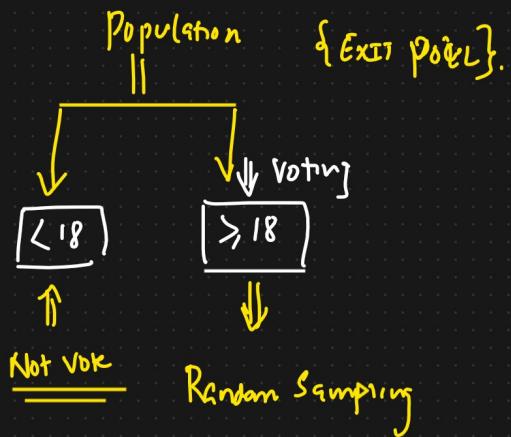
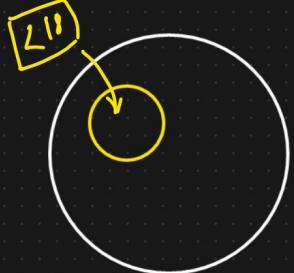
Education

- High School
- Master
- phd

Degree

Blood groups

-



③ Systamatic Sampling $\rightarrow \{ \text{AIRPORT} \}$ n^{th} person

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

Select every n^{th} individual out of Population (N)

④ Convenience Sampling \div Only those who are interested in the survey

Will only participate

inconvenient job for

$\left\{ \text{DATA SCIENCE SURVEY} \rightarrow \text{General AI Survey} \right\}$ a specific



\Downarrow
 $\left\{ \text{Fill the Form} \right\}$

① Survey Regarding New Technology \Rightarrow Convenience Sampling

② RBI Survey \Rightarrow Women \Rightarrow Stratified + Random Sampling \rightarrow Married Women

③ Credit Card \div Stratified + Random Sampling

① Variable : A variable is a property that can take any values

Eg: $age = 14$ Variables

$age = 25$ $AgeS = [24, 25, 26, 27, 28, 29] \Rightarrow$ Collection

$age = 100$

Two different types of Variable

① Quantitative Variable \rightarrow Measured Numerically {Mathematical Operation}.

Eg: Age, weight, height, rainfall(cm), temp, distance

② Qualitative Variables \rightarrow Categorical Variables {Based on some characteristics they are grouped together}.

Eg: Gender, Types of flowers, Types of Marbles

Quantitative Variable



Continuous Variable.

Eg: Whole number \rightarrow fixed

Eg: No. of Bank Accounts

$\{1, 2, 3, 4, 5\}$

$25/X$

Eg: Continuous \rightarrow Decimal values

Eg: Height, weight, ages, Rainfall

Speed

Eg: No. of children \div Whole numbers

Pincode = fixed

Categorical
variables.
Marbles
 \rightarrow Mixed
 \rightarrow Not Mixed }
} \uparrow

Assessment

Gender ? Categorical

① What kind of variable is Marital Status? Categorical variable

Length River length? Continuous

Movie duration? Continuous

Pincode ? Discrete

IQ ? Discrete

105.75, 90.5,

Pancard

Pincode

Fixed

Categorical

↓
[FE]

Ans no. of Categories

[AMLPN - - -]

360099
720058
560092

} \Rightarrow It is many?

Categorical

Variables

Continuous



Discrete ←

Continuous

Whole number

Bank Account = { 2, 3, 4, 5 }

Pincode = { }

Cities

5

Gender Pincode

M

F

[6]

Categorical

PAN ←

[]

[]

} Categorical

Profile Building

- ① LinkedIn
 - ② GitHub
 - ③ Instagram
 - ④ Resume
 - ⑤ Mock Interviews
- ⇒ Jobs, Opportunities
- X → Freelancing