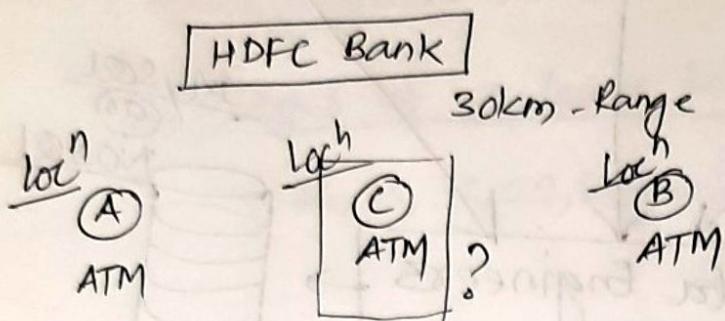


Use Case :

- ① Can we open new ATM @ location 'C'?



① Service Based Co.
{TCS}

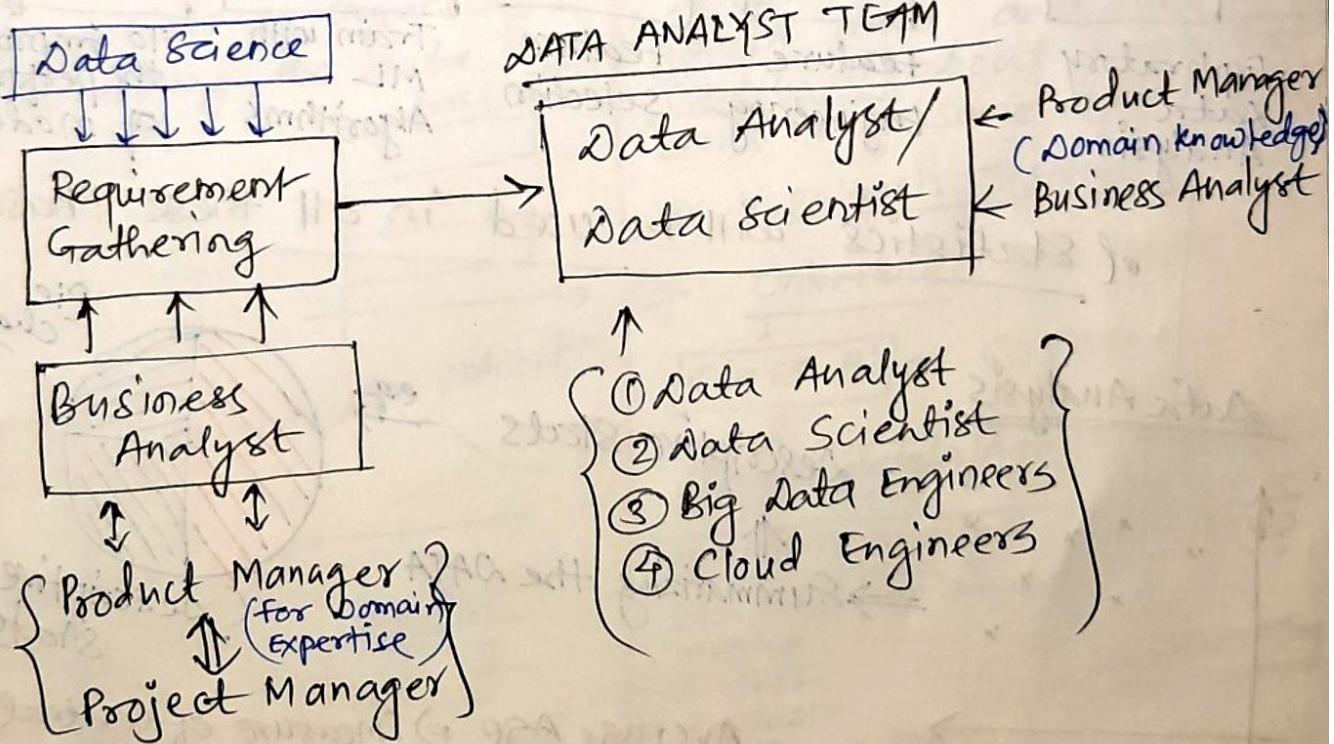
② Product Based Co.

- {Google}
- YouTube, G Pay, GMail, Google Ads, etc.

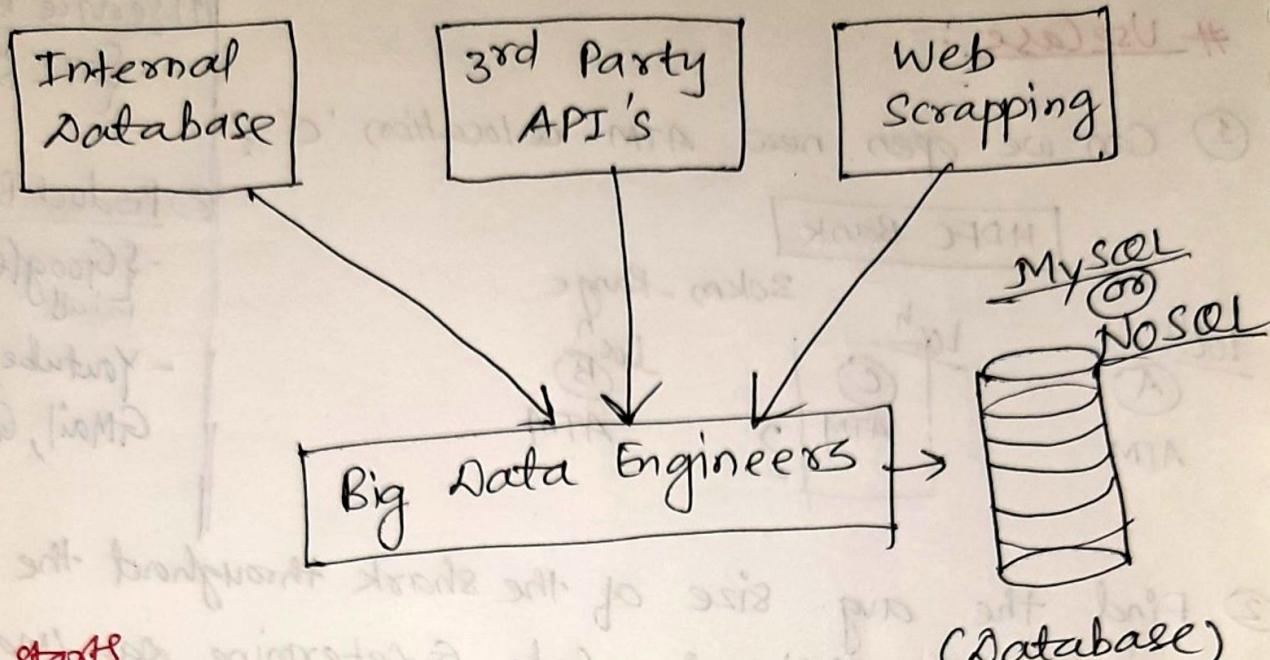
- ② Find the avg. size of the shark throughout the world?
 ③ Amazon Big Billion Day Sale (determine day / Month of sale)

STATISTICS

• Life Cycle of Data Science Project -

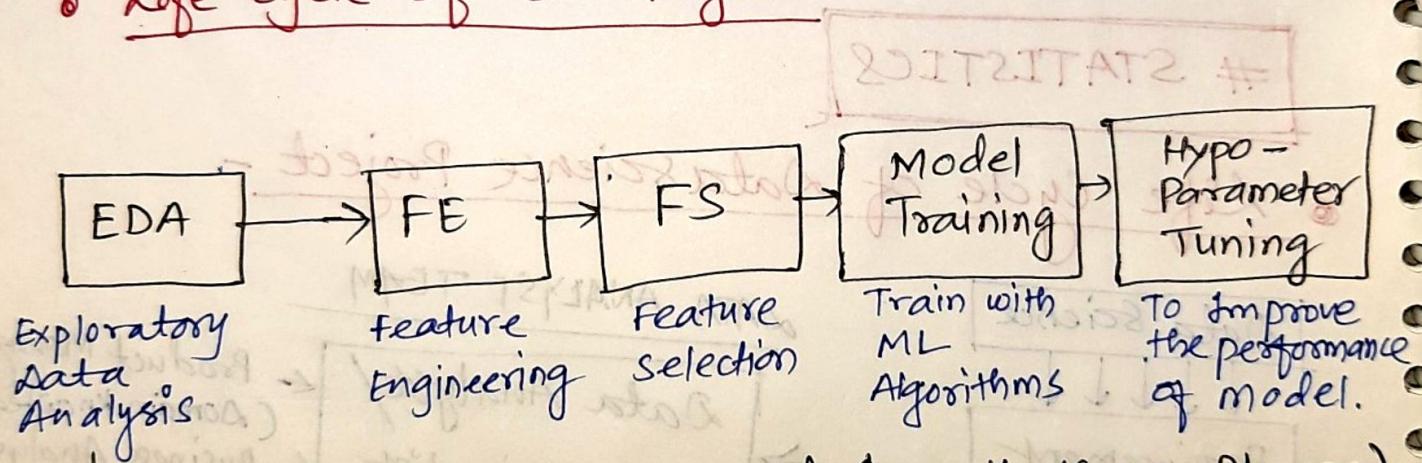


② database collection:



→ Now starts

• Life Cycle of DS. Project :-



(Statistics will be used in all these Phases)

Data Analysis

descriptive stats
Summarizing the DATA
eg →
Age = {12, 14, 13, 18, 20, 25} ⇒ Average Age ⇒ (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: ① Ages of students in classroom

{ 24, 25, 32, 29, 28 } \Rightarrow Mean, Median, Mode, Standard deviation...etc.

② weights of students in classroom:

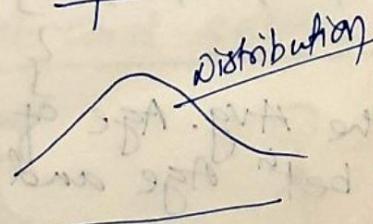
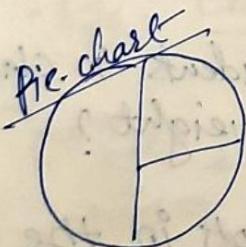
• Types:

Statistics

↓
(EDA + FE)

descriptive stats

- It consists of organising & summarizing the data



Box-Plot



Candlestick

inferential stats.

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

By Hypothesis Testing

e.g.: University = 500 People

[Class 'A'] = 60 People

↓
[Sample data] \Rightarrow [Age] \Rightarrow Avg. Age of entire university?

- Hypothesis testing:

④ C.I. = Confidence Interval

⑤ P-value

⑥ Z-test

⑦ T-test

⑧ Chi-Square Test

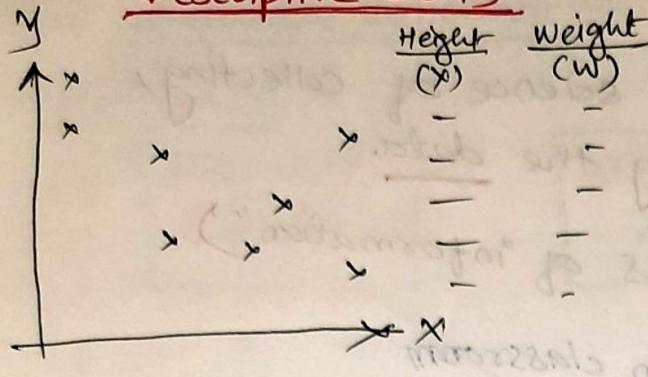
⑨ F-test

↓ yes.

By Hypothesis testing

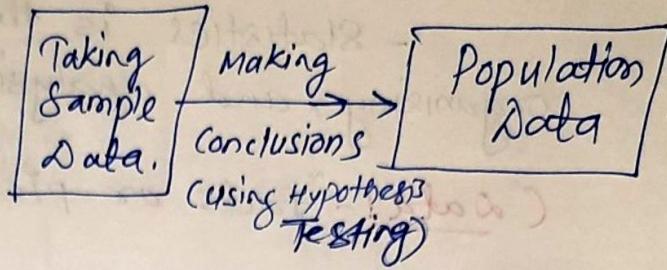
④

Descriptive stats :-

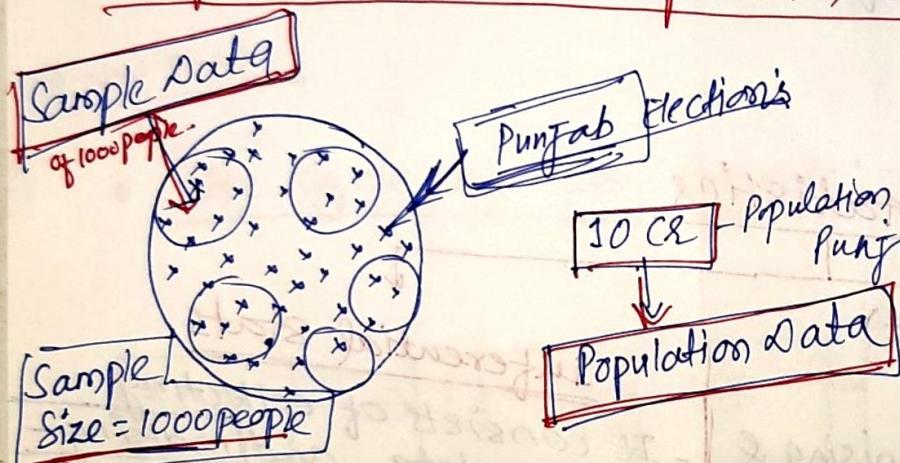


Re: $\begin{matrix} X \uparrow & Y \downarrow \\ X \downarrow & Y \uparrow \end{matrix}$ for Eg. only.

Inferential stats :-



Sample Data Vs Population Data :-



EXIT POLL

- Party 'A' will Win
- Party 'B' will Loose

Problem Statement

- Lets say there are 20 classrooms in a university and you have collected the Age's of 8 students in one classroom

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

Weight { --- --- --- --- --- }

- Descriptive stats :-
 - ① What is the Avg. Age of students in classroom,
 - ② Relationship betⁿ Age and weight?

- Inferential stats :-
 - ① Are the Avg. Age of the students in the classroom less than the Average Age of the students in the university?

② Data :- 1000 students in university :- Class 'A' \rightarrow 50 Girls - 50 Boys

③ So Girls have performed well throughout University 75% - 92%

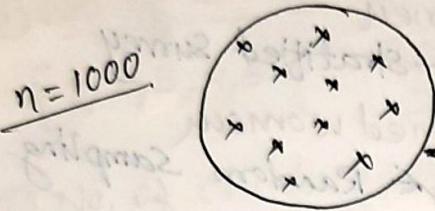
Sampling Techniques :-

(5)

$$\frac{\text{Population (N)}}{\text{Sample (n)}}$$

① Simple Random Sampling :-

- Every member of Population (N) has an equal chance of being selected for your sample (n)

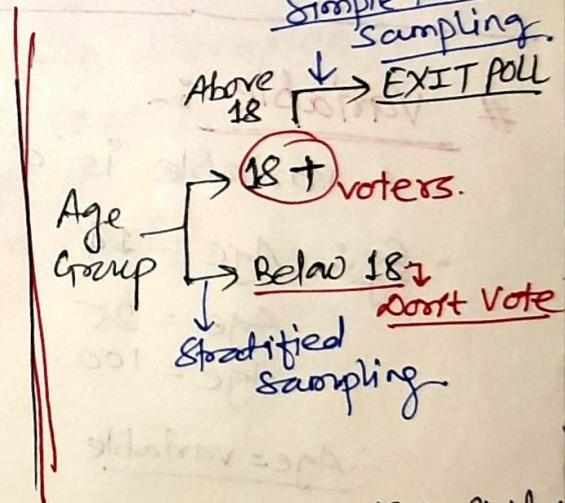


Simple Random Sampling

Eg:- EXIT POLL, SURVEY,
Movie Review, Lottery.... etc.

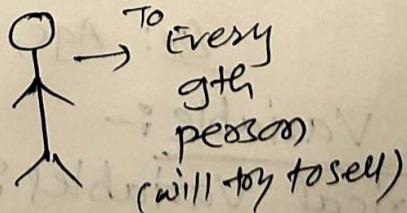
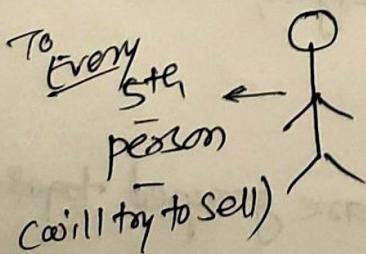
② Stratified Sampling :-

- Strata \rightarrow Layers \rightarrow Clusters \rightarrow GROUPS.
- Gender
 - \rightarrow Male.
 - \rightarrow Female.
- Education
 - \rightarrow High School
 - \rightarrow Masters
 - \rightarrow Ph.D.
- Blood groups
 - \rightarrow A
 - \rightarrow B
 - \rightarrow O
 - \rightarrow AB



③ Systematic Sampling :- - you select every n^{th} individual out of population (N).

- Eg.: credit card sellers @ AIRPORT: n^{th} person



⑥

④ Convenience Sampling :-

- Only those who are interested in the survey, will only participate.

Eg:- ① Survey Regarding New Technology.

↳ convenience sampling.

② RBI Conducted Survey → Women

↓ - stratified survey

Married women

↓ ↳ Random Sampling

③ Call for Credit Card sell.

↓ stratified (to only salaried person)

+ Random sampling

Variable :-

- A variable is a property that can take any values.

- Eg.: Age = 14

: Variables
↓

Age = 25

Ages - [24, 25, 26, 27, 28] ⇒ collection.

Age = 100.

∴ Age = variable

Types of Variable :-

① Quantitative Variable

→ Measured Numerically.
{ Mathematical Operations }

Ex: Age, weight, Height, Distance...etc.

② Qualitative Variable :-

2. ④ Categorical Variables :-

{ Based on some characteristics they are grouped together }

Ex:- Gender, Types of flowers, Types of Movies.

Quantitative Variable

Discrete Variable

Eg.: Whole Number

E.g. ① No. of Bank Accounts.

$\{1, 2, 3, 4, 5\}$ $(2.5) \times$
not possible.

② No. of children's :-

Continuous Variable

Eg.: Continuous Number.

Eg.: Height, weight, Ages,
Rainfall (CM), speed,

etc. (can be decimal/
fraction)

Assignment #1

Q. ① what kind of variable is Marital status?

② " Ganga River Length?

③ " Movie Duration?

④ " Pin code?

⑤ " IQ ?

⑥ " Pan card?

Ans. ① - Categorical variable.

② - Continuous variable

③ - Continuous variable

④ - Discrete variable.

⑤ - ~~Continuous variable~~

Discrete Variable - IQ

⑥ - Categorical variable.