

Hare Krishna:-

Statistics:-

Statistician

Use case:-

↳ 5 years.

[HDFC]

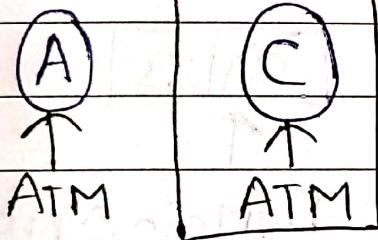
↳ Data Analyst

30 KMS

↳ Business Analyst

↳ Data Scientist

↳ Product Managers

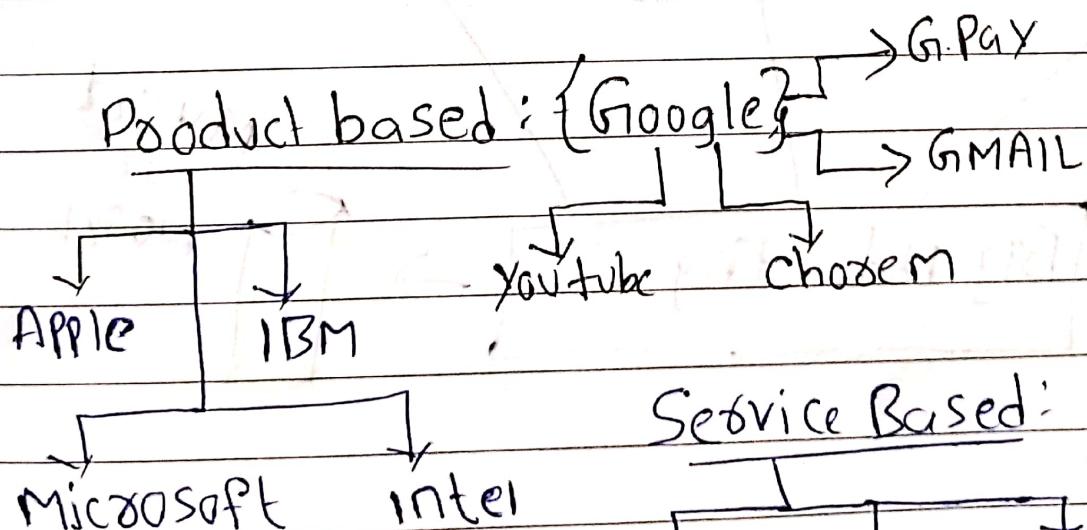


① DATA ANALYST

② DATA SCIENTIST

Product Based
company

Service Based
company



Service Based:

TCS
HCL
infosys
Matrikas

Statistics:-

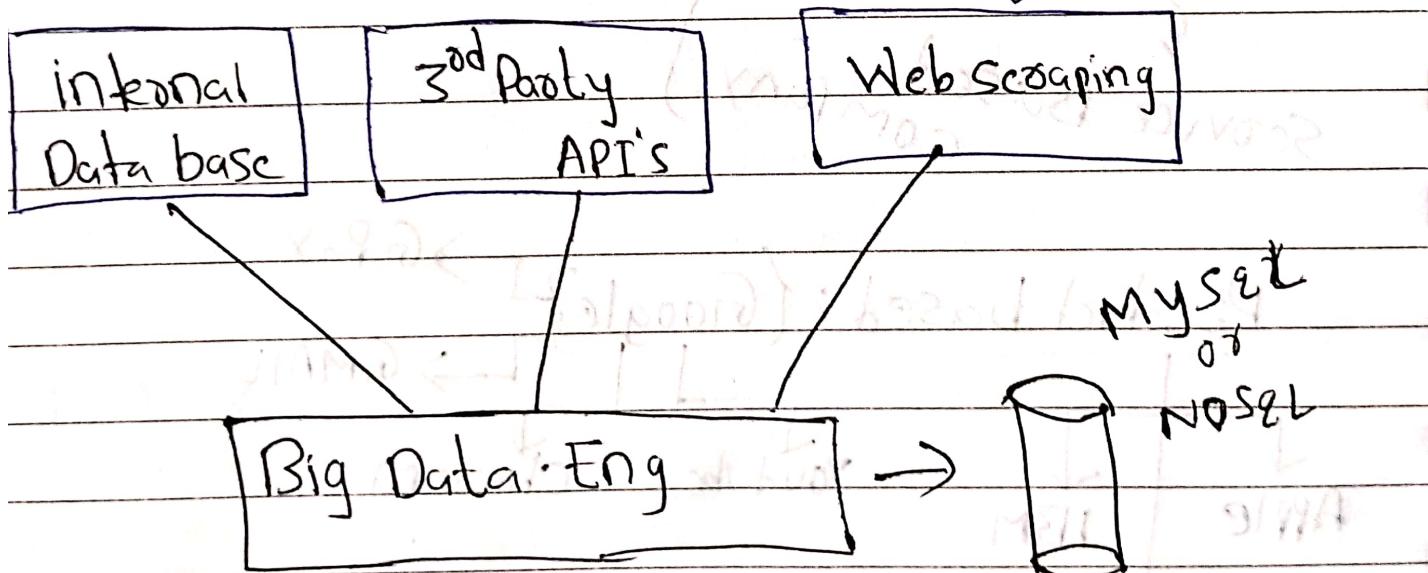
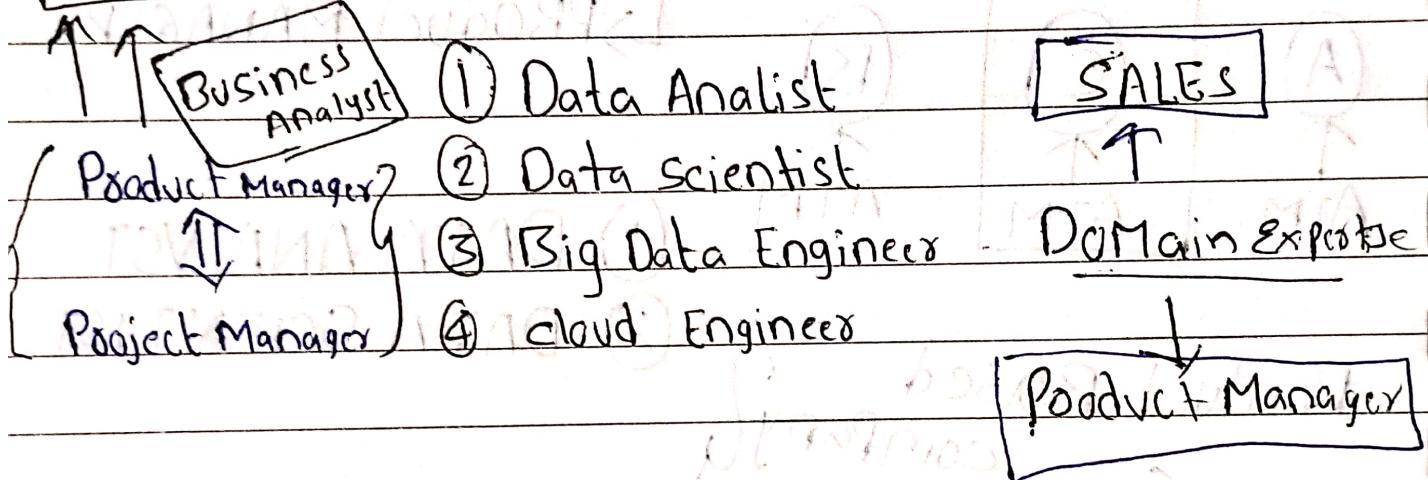
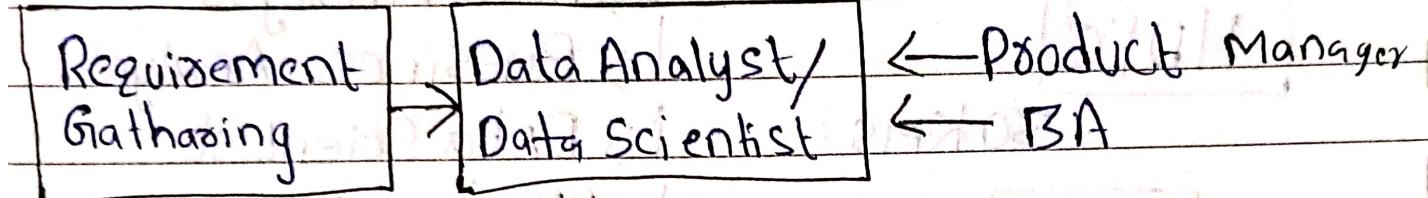
life cycle of data Science

Project

DATA SCIENCE

Domain

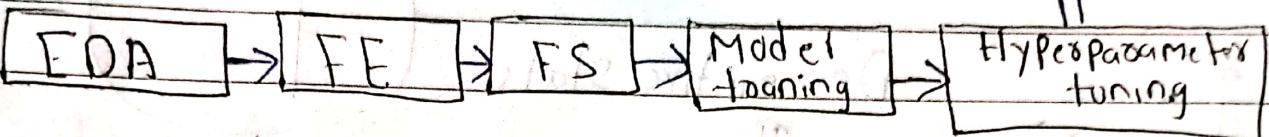
Knowledge



Date: / /

Life cycle of DS Projects.

improve the
Project model



EDA :- Exploratory Data Analysis

FE :- Feature Engineering

FS :- Feature Selection

Statistics are used in

→ EDA

→ FA

→ FS

→ Model training

→ Hyperparameter tuning.

Product base Companies:- focuses on making
the highest quality product for its customers
Ex:- Microsoft, Google, Cisco, Adobe, Apple etc

Service based companies:- Provide required
service to the another company products

Ex:-

TCS, Infosys, IBM, Wipro, Accenture
etc.

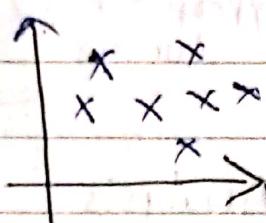
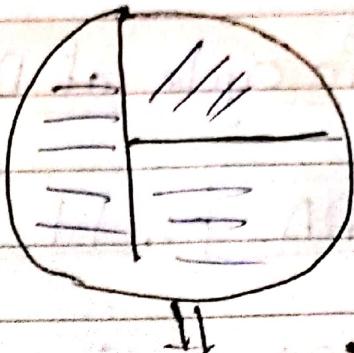
MATRIX



Scanned with OKEN Scanner

Date: 1/1

Analysis of Data:



Descriptive stats
↓

Summarizing the data

Age = {12, 13, 14, 18, 20, 25} \Rightarrow Average Age

Descriptive stats

descriptive
stats

Measure the central

Tendency

Definition of Statistics:-

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

Data:- "Parts or Pieces of information"

Eg.: Ages of students in classroom

{24, 25, 22, 29, 28} \Rightarrow Mean, median,

Standard

deviation

Mode

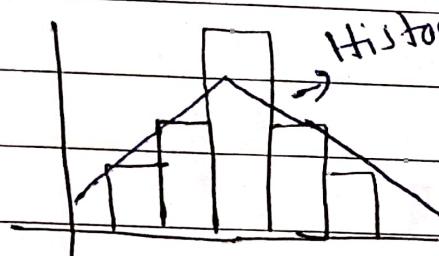
MATRIXAS

Statistics

Descriptive stats

[EDA + FE]

- ① it consists of organising and summarizing the data



Hypothesis testing

inferential stats

- ① it consists of collecting sample data and making conclusion about population data using sum experiment

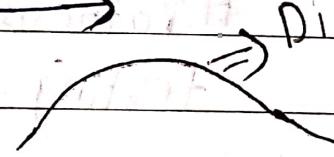
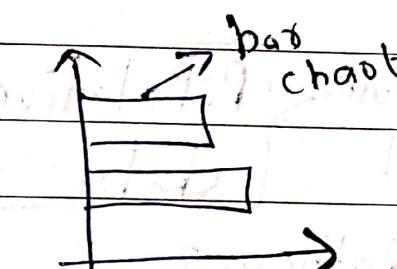
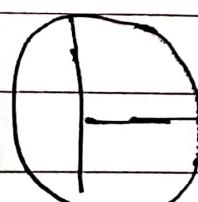
University → 500 people

Class A → 60 people

Sample data ⇒ Age

Average age of entire university

↓
Hypothesis testing



⇒ candlestick

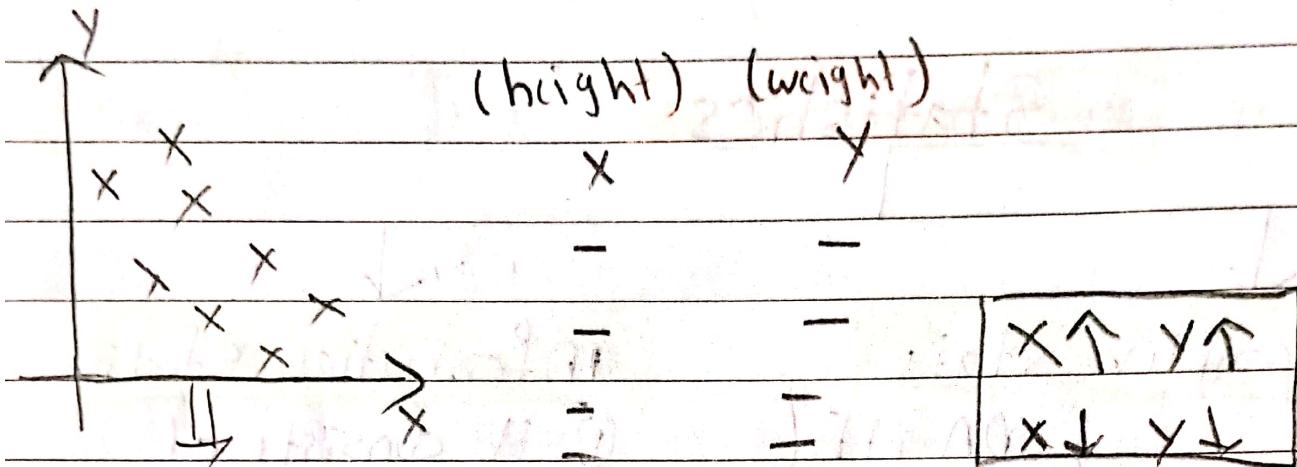
C.I ⇒

confidence interval



Box plot.

Date: / /

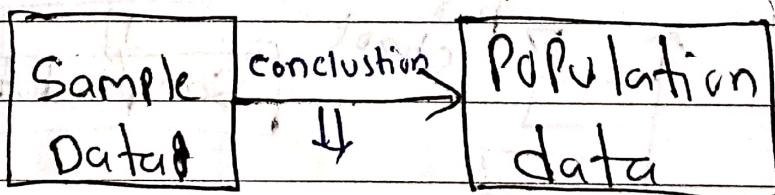


Scatter plot

$x \uparrow y \uparrow$
$x \downarrow y \uparrow$

P-Value

- ① Z-test
- ② t-test
- ③ Chi-square test
- ④ F-test



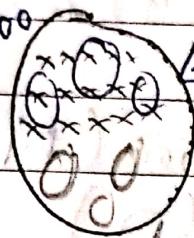
Hypothesis
testing

Date: / /

Sample data & Population data

Sample

size = 100



Punjab

100%

Exit Poll:

Party A will win

Party B will loose

Population data

Eg: let say that 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 b/w Age & gender?

Inferential stats:-

Are the average age of the students in the classroom less than the average age of the student in the

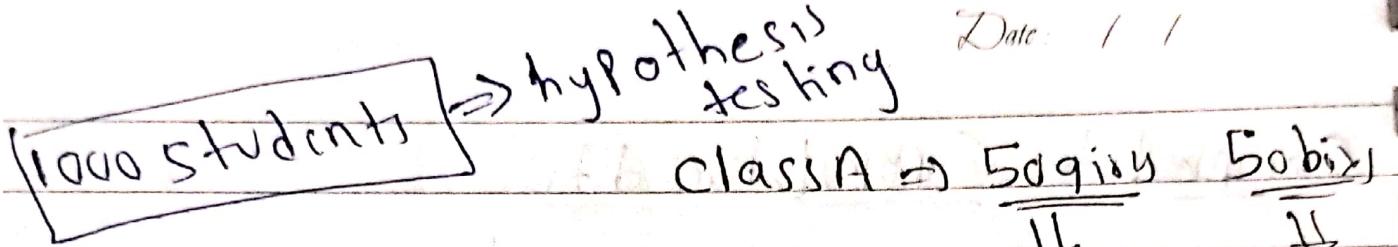
{ 1st
got
tre... }

University
Matrikas



Scanned with OKEN Scanner

Date: / /

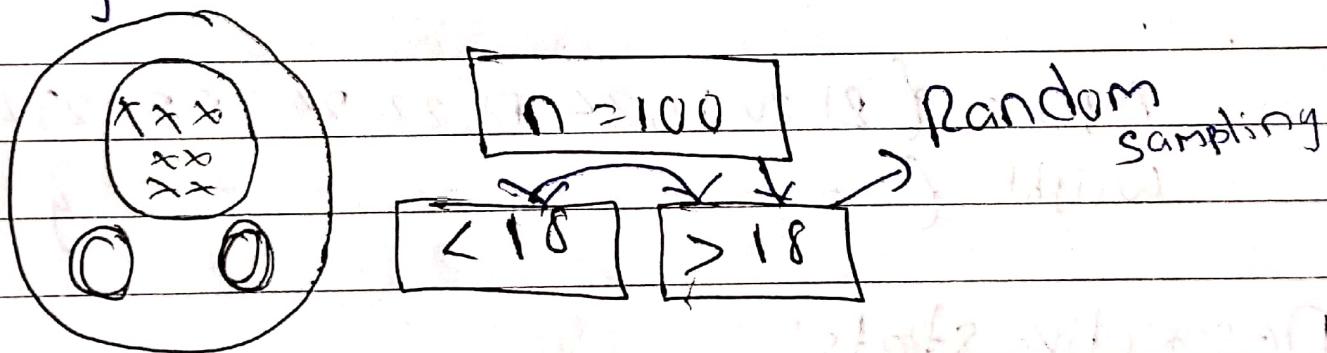


choose a sample: $\frac{11}{11}$ $\frac{11}{11}$
95% 92%

Sampling Techniques: Population (N)
Sample (n)

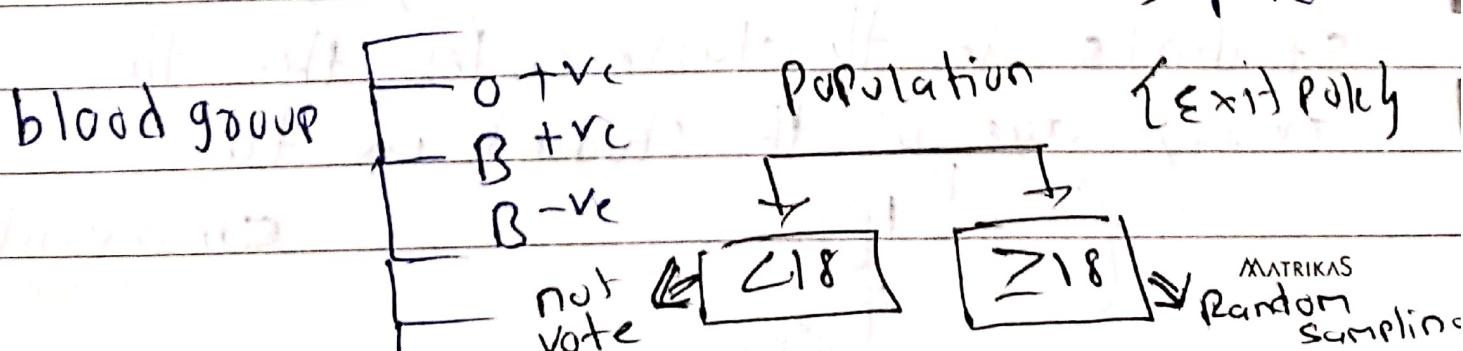
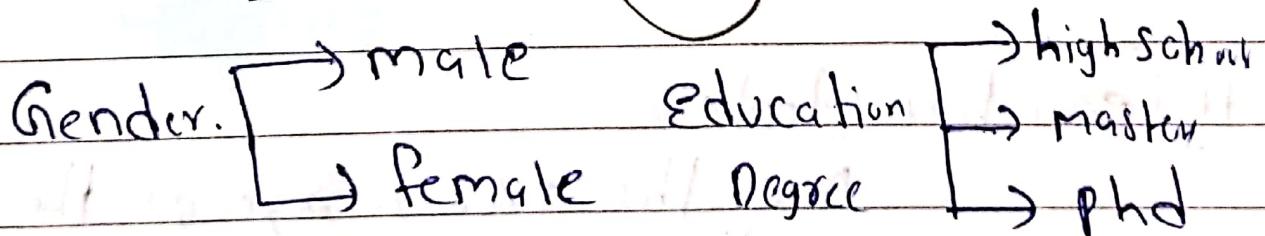
① Simple Random Sampling:

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



Strata \rightarrow layers \rightarrow clusters

② Stratified Sampling:



Date: / /

③ Systematic Sampling:

Select any n^{th} person

out of population (N)

$\rightarrow \{ \text{Air Port} \}$

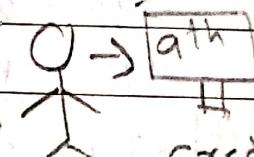
\searrow selling

5^{th} person

$\{ \text{credit card} \}$

Select every n^{th} individual

out of population (N)



* Convenience Sampling:

only those who are interested in the survey will only

participate

① Survey Regarding New technology:

\Rightarrow Convenience

② RBI Survey \Rightarrow Women

\Rightarrow married

\Rightarrow Sampling

stratified

+

Random Sampling

③ credit card!

stratified +

Random Sampling

MATRIXAS

Date: / /

① Variable: A Variable is a property that can take any variables

Eg: age = 14 Variable

age = 25 Age: [24, 25, 26, 27, 28]

age = 100 collection

Two different type of Variable

① Quantitative Variable

Measured numerically (mathematical operation)

Ex: age, weight, height, dist, rainfall, temp,

② Qualitative Variable:

Categorical Variables

(based on some characteristic they are grouped together)

Eg: Gender, types of flowers,

types of movies

Quantitative Variable

Discrete Variable

Continuous Variable

Discrete Variable:

Eg:- Whole number → fixed

Eg:- N.o of Bank accounts

{1, 2, 3, 4} 2.5 X

Eg:- N.o of children :- whole n.o

Continuous Variable:

Eg:- continuous → decimal values

Eg:- height, weight, age, speed

marital → named
not marital → not named

Categorical variable

Quantitative Variable

Discrete Variable

Continuous Variable

whole numbers

Bank account = {1, 2, 3, 4}

pin code = {---}