

→ Statistics day 1

Page No.:

Date: / /

why need Statistics :-

example :

(A)

(C)

(B)

A and B are location at certain distance we need find that is at point C do need to open a new Atm or not :- done by Statistics

definition :-

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

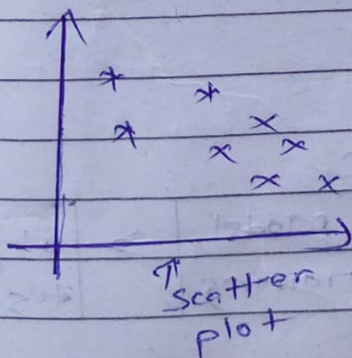
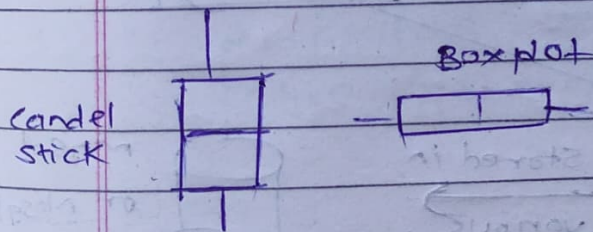
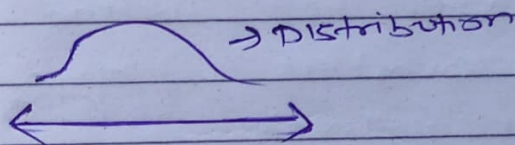
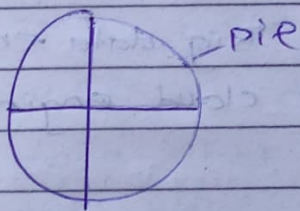
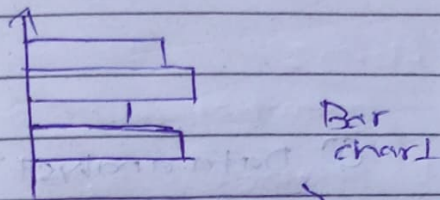
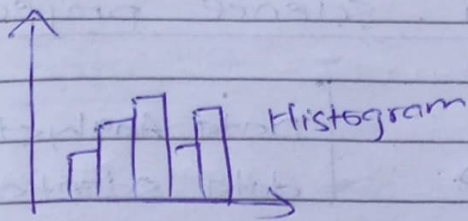
Types of Statistics :-

Descriptive

- It consist of organising and summarizing the data

Inferential

- It consist of collecting sample data and making conclusions about population data using some experiments

DiscriptiveInferential

we make conclusion
using hypothesis
testing

ex:-

university \rightarrow 500 Student

Class A \rightarrow 60 people

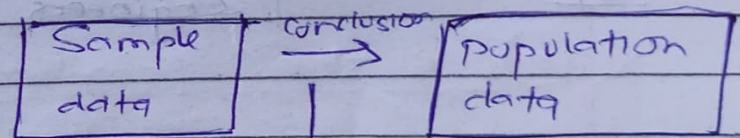
Sample data = Age = Avg age
of entire
university

(1) z-test

(2) t-test

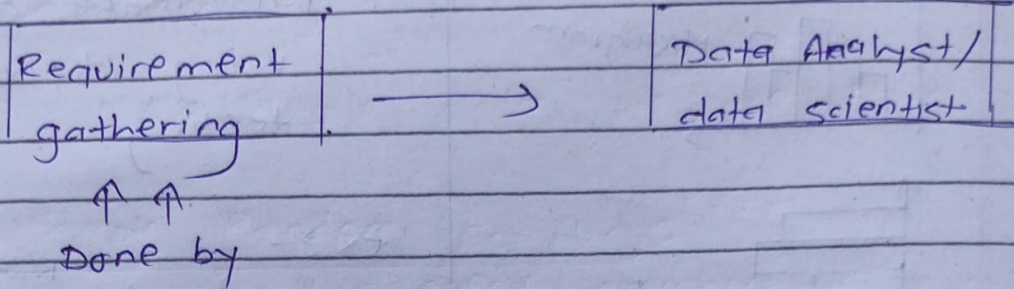
(3) chi-square

(4) f-test



done by
hypothesis
testing

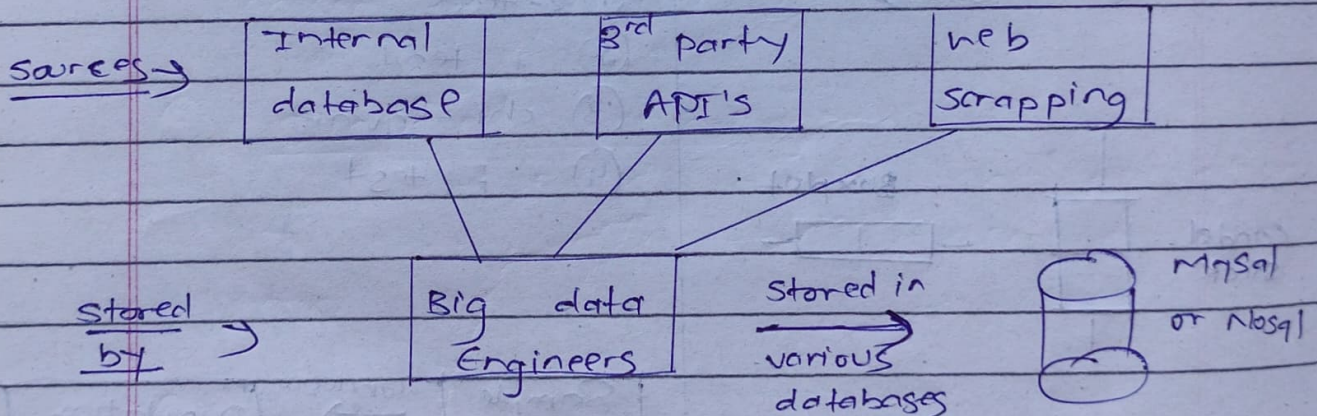
Life cycle of data science project:-



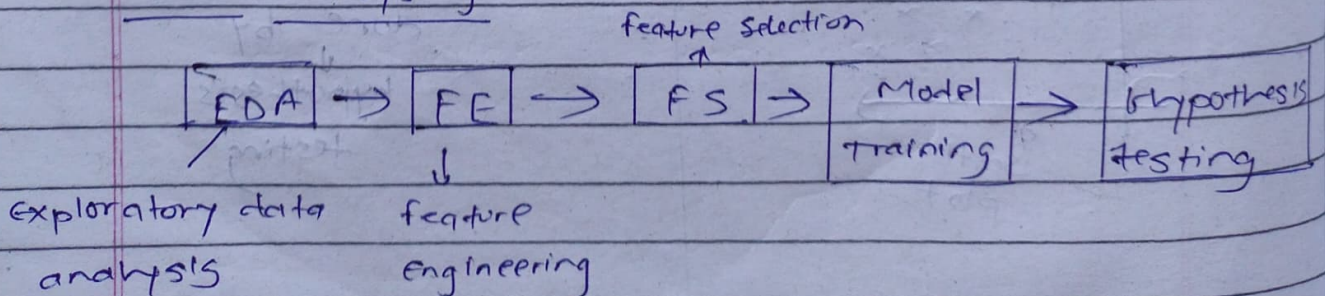
① Product manager
② Business analyst
both prepare list of required things

- ① Data analyst
- ② Data scientist
- ③ Big data engineers
- ④ cloud engineers

* for collecting data:-



* for analysing:-



** in all analysis process statistic is required

data :- It is fact, figure or values that can be processed to make information

Types of data

population (N)	Sample (n)
→ It is called as universal data or data that entirely we have	→ It is part of population upon which we apply our analysis techniques to get information

example :-

In exit pole we predict chance of winning a party by taking certain & specified data and analyse it with entire (population) data

* K. Sampling Techniques :-

To take sample data we use 4 methods.

① Simple Random Sampling :-

- In this every member of population has an equal chance of getting selected for sample data (n).

② Systematic Sampling :-

- It is similar to Simple random sampling. Every member of the population is listed with a number but instead of randomly generating numbers, individuals are chosen at regular intervals.

③ Stratified sampling :-

- involves dividing the population into subpopulation that differ in important ways.

④ cluster Sampling :-

- also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample.

Variable :-

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

e.g. age = 14 ages = [19, 21, 24, 30, 35]

↓ ↓

it is variable it is variables

Types of variables

qualitative	quantitative
- Categorical	- Numerical
- e.g. Gender, department	- e.g. Age, height

quantitative variable

Discrete	Continuous
<ul style="list-style-type: none"> - it is whole number - It cannot divide into finer parts - e.g. no. of pens, no. of students 	<ul style="list-style-type: none"> - it is floating point number - It can divide into finer parts. e.g. temperature of a day, height
	<p>④ we can express height in m then to cm \rightarrow mm \rightarrow as much we want</p> <p>\therefore it is continuous</p>