18/04/2023

Statistics

X statisticians ⟶ 5 years

→ Data analylist ✓

— Data scieatist ✓

— Business analyst ✓

$\boxed{\text{product managers}}$ - Domain expertise

TCS - Servic

gogle - prod

↓ creating own product

Usecase :- working

Bank → HDFC

30 km.

✓ new one ✓

(A) (C) (B)

Location Location

ATM ATM ATM

Data engineering

→ Opening new ATM in C How we can be come to conclusion we built ATM?

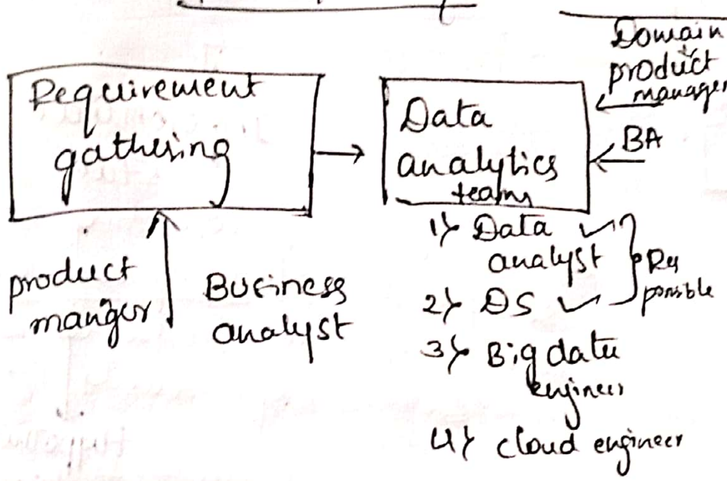
1) find the average size of the shark throught the world? people

2) Amazon big billion day save {intuit} which month should you sell for the

Statistics

1) Defn
2) type of statistics
3) Life cycle of Datascientist
4) sample data(n) v/s population data(N)
5) sampling technique
6) variable types

Life cycle of Datascience



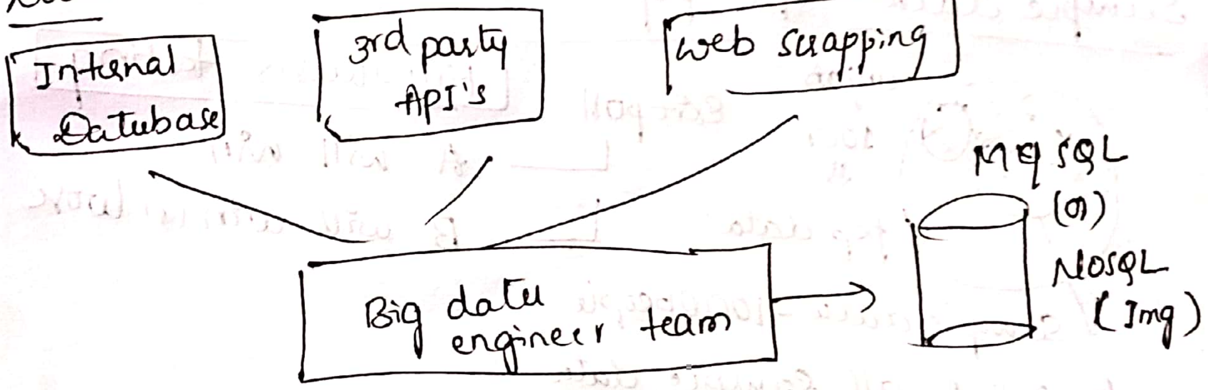
Statistics

It is The science of collecting, organising & analyzing the data.

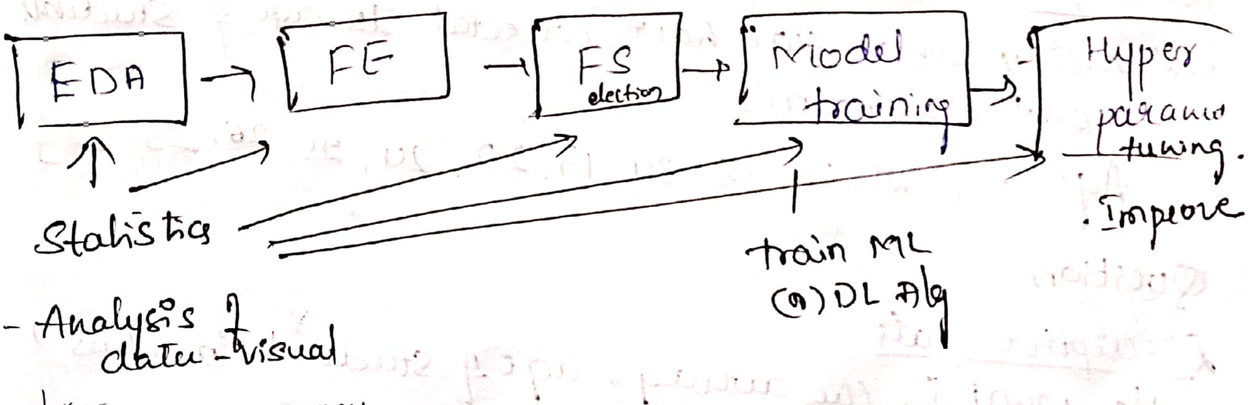
Data :- facts (or) pieces of information.

Eg:- Age of students - { 24, 25, 36 }

Data



Datascientist :- project lifecycle



- Analysis of data - visual

L×× → summary data | Discrite Statistics

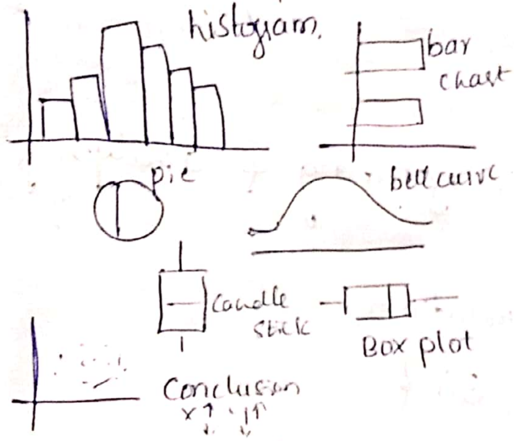
Age = { 1, 2, 5, 8 } — Discriptive, measure of central tendency

Statistics

Discriptive Stats [EDA, FE] Extensively used

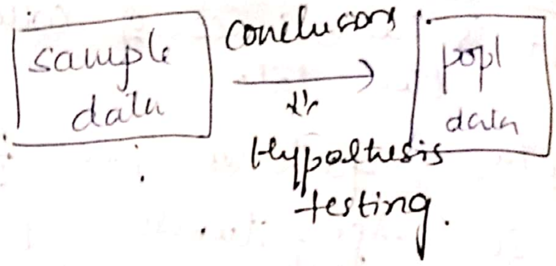
Inferential Stats

1) It consist of organizing & summarizing using all plots



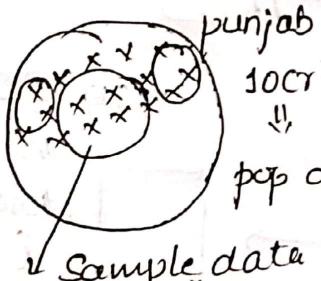
→ It consist of collecting sample data & making conclusion about population data using some experiments → Hypothesis test

→ Average calculation



(n)

Sample data v/s population data (N)



Exit poll

Hypothesis testing

└── A will win

└── B will win (a) lose

Sample data - 1000 people

→ Average of all sample data

Example:- Let say there are 20 classroom in the university and you have collected the age of students in one classroom?

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

Question

Discriptive stats

1) what is the average age of student in class?

2) Relationship b/w age & gender?

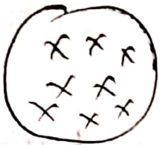
Inferential stats:-

1) Are the avesage of the student in the classrom less than the average age of student in the university?
[comparing - hypothesis test]

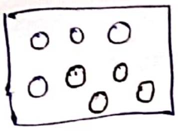
2) Average marks
girls 95%
50 boys 92%
1000 students

Sampling Techniques:-

1) simple random sampling:- Every member of the population (N) has an equal chance of being selected for your sample (n).

n=1000 

Exit poll
general servey } Random sampling

Ex:- 

select the marble, each marble have equal chance of selecting.

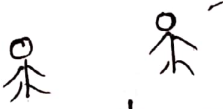
2) stratified sampling:- [strata → layers → cluster → group]

① Gender [F M
② education degree [High school master

electing
③ Exit poll
vote >18 → Random sampling
not <18 vote →
→ R.
stratified sampling

④ blood group.

3) systematic sampling:-
{credit card} → select every nth indivias out of population (N).

5th person  9th person

→ There is slm way of selecting. (nth)

4) convience sampling: only those who are interested in survey will only participate.

Ex: DS survey → general AI. who interested :

→ ineuron job for specific role - those who are really interested those. only fill.

1) survey regarding new technology.
↳ convience sampling.

RBI → survey → women. ⇒ stratified + Random
↳ Married

2) credit card call
↳ stratified + Random
↑
salary.

Variable :- It is a property that can take any value.

Ex: Age - 14
A = 24

variables
Ages = [1, 2, 3]

2- types of variable

1) Quantitative variables - measure Numerically.
{ maths operation }
Ex: Age, weight, temp, distance

2) Qualitative variables -[categorical Variable]
based on some characteristics they grp together
Ex: gender, flower, car, movies, dept.

Quatitative

Discrete variable [many no of vaiable] categorical)

Ex. whole no

- No of bank acc
{1.2.3.4}

-> limit to no again repeat wrt data points

continous vaiab

Ex continous.

Ex:- Height, age, speed, Rainfall

-> decimal value.

1) Marital status ? - categorical [no of maried people discrete]

2) ganga river length ? continous

3) movie duration ? continous

4) pincode ? [descrete - continous] not categorical bcz

5) IQ ? continous [discrete) more no so not able analyze instead

6) gender ? categorical descrete - Feature engineering

7) country ? - discrete