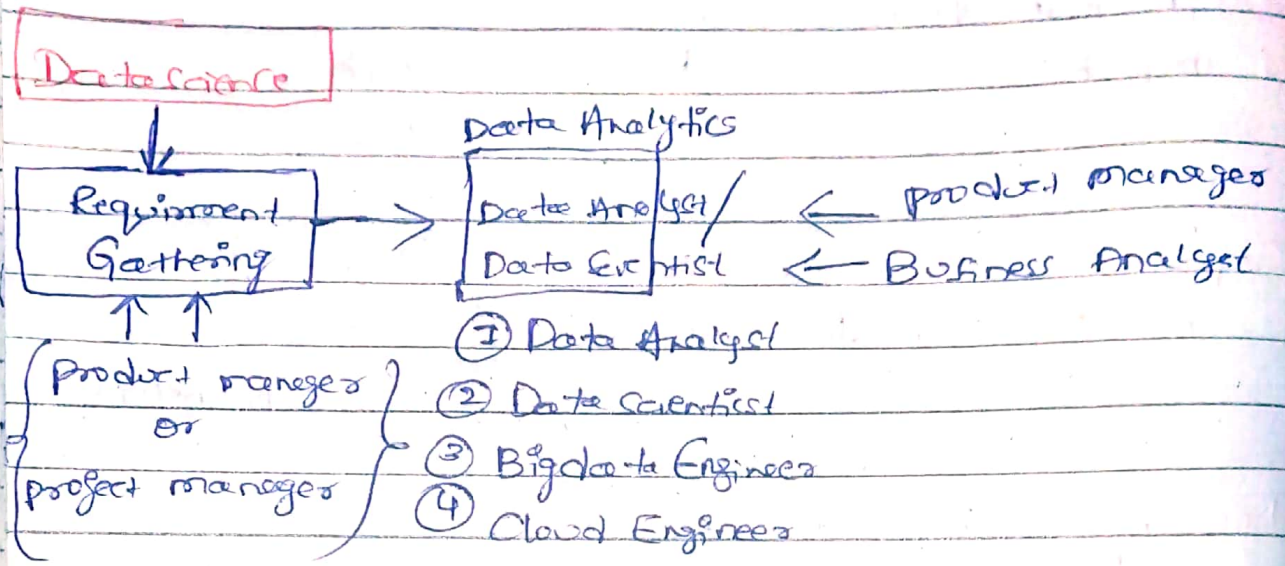
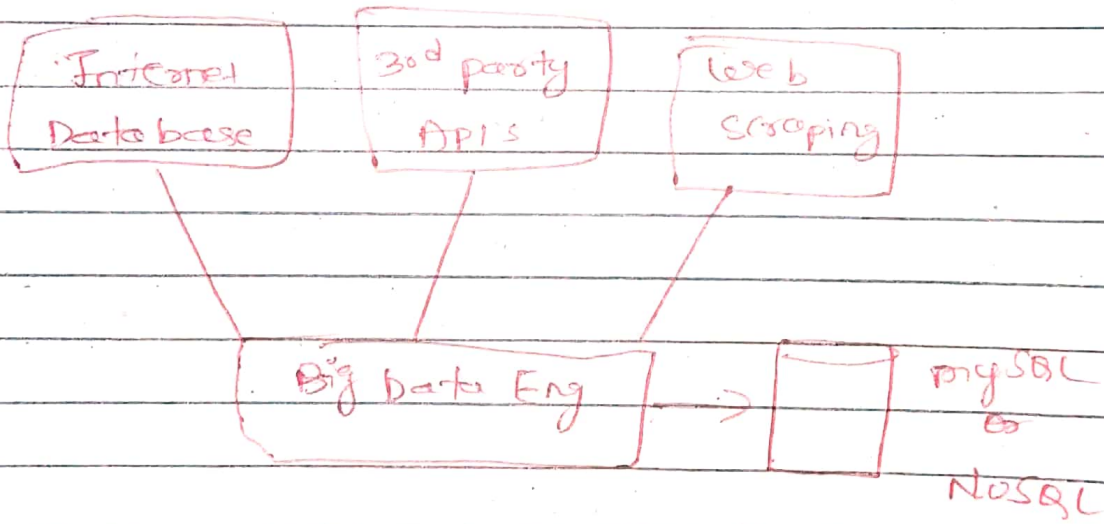
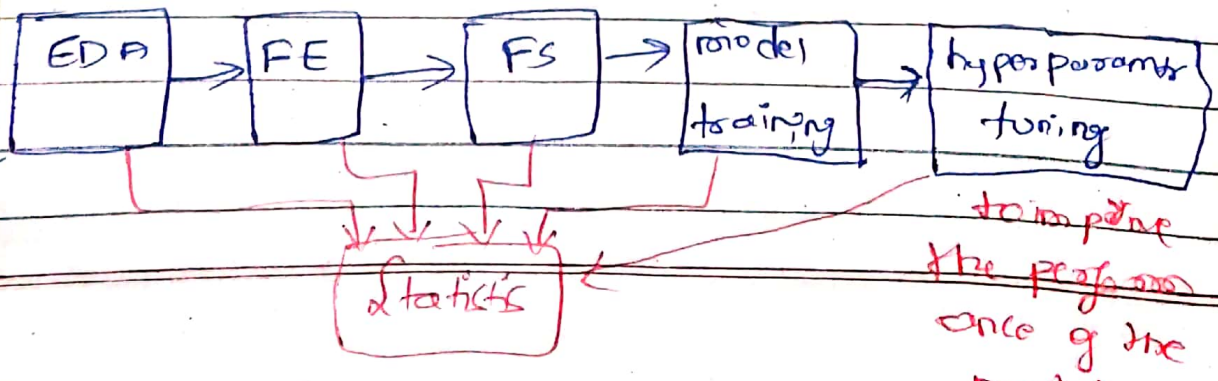
Statistics:- Day 1:-

Life-cycle of Data Science project





Life cycle of DS Project



Statistics :-

Statistics is the science of Collecting, Organizing and analysis the data.

Data :- "facts or pieces of information"

eg :- Age of student in Classroom

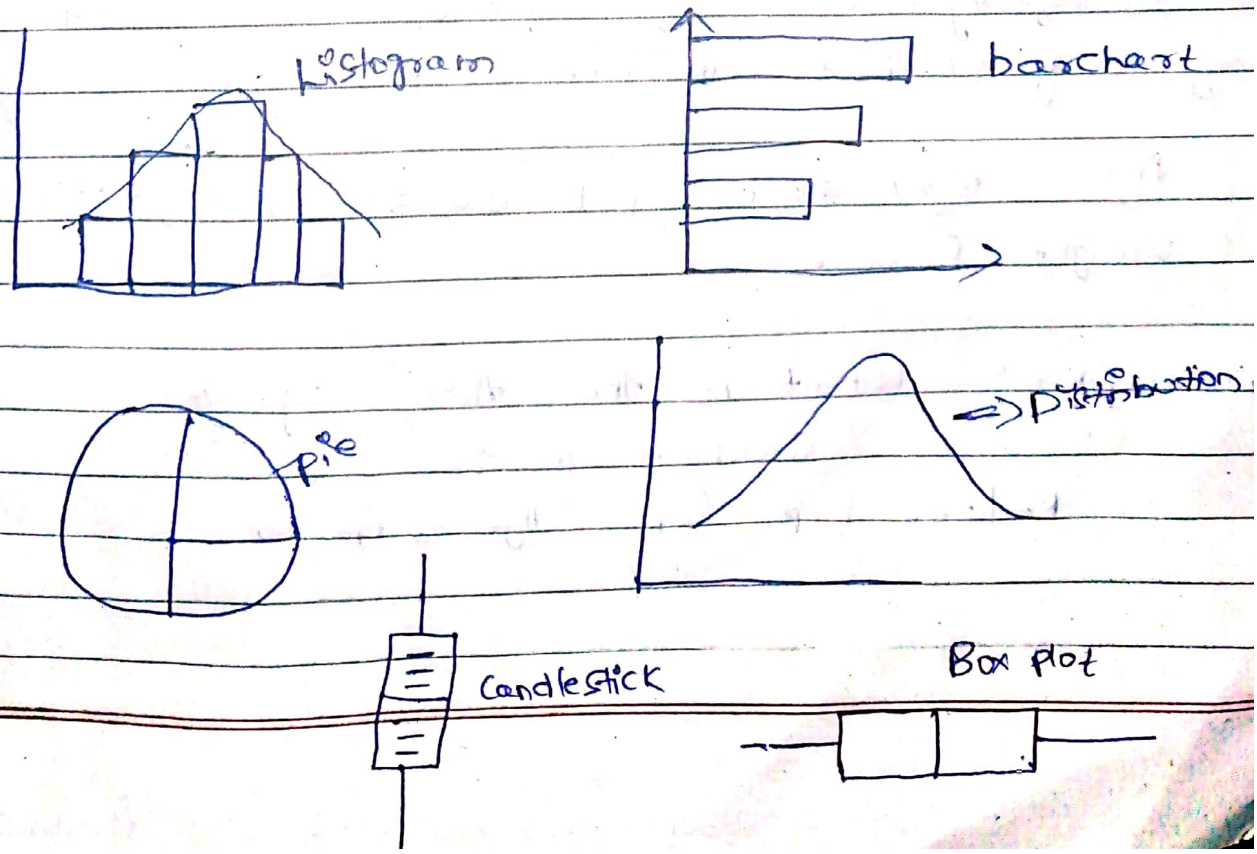
{24, 25, 32, 29, 28} => mean, mode, median, standard deviation

Types of statistics :-

- Descriptive
- Inferential

Descriptive :-

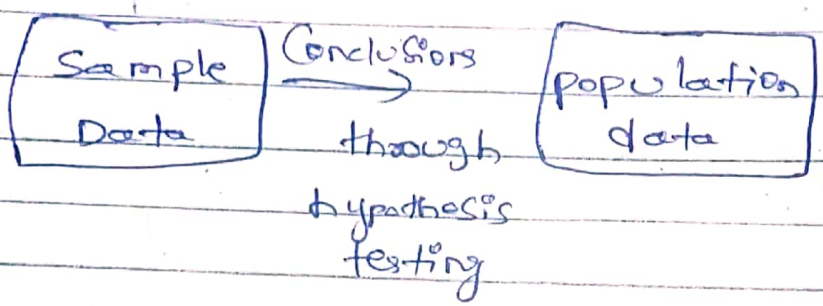
It Consists of Organizing and Summarizing the data using below graphs



Inferential Stats :-

It Consist of Collecting Sample data and making Conclusion about Population data using Some Experiments

making Conclusion → Hypothesis testing



Sample Data Vs population Data

Ex: lets Say there are 20 classrooms in a university and you have collected the age of Students in one classroom

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

Weight = { - - - - - }

Descriptive: What is the Avg age of Student in the Classroom,

Relationship b/n Age & Gender

Population (N) Sample (n)

Inferential Stats :-

Are the avrg age of the Student in the Classroom less than the avrg age of the Student in the University

Different types of Sampling techniques :-

(1.) Simple Random Sampling :

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

Ex: Random Sample (Lottery)

(2) Stratified Sampling :-

Strata → layers → clusters → Groups

Gender → male, female

Blood Groups → AB+, B+, O+

(3) Systematic Sampling :-

Selecting every nth individual out of population (N)

(4) Convenience Sampling :

Only those who are interested in the survey will only participate

① Survey Regarding New Technology → Convenience Sampling

② Credit Card → Stratified + Random Sampling

Variable :-

a property that can take any values

Eg :- age = 14
age = 35
age = 100
} Variable

Variables :

Ages = [24, 25, 36 42] =

2 Different types of Variable :

① Quantitative Variable

⇒ measured Numerically (mathematical operation)

Distance, age, weight, temperature,

② Qualitative Variables (Categorical Variables)

Based on some characteristics they are grouped together

Eg: Gender, Types of flowers, Types of movie

Quantitative Variable:

Discrete Variable (fixed values)	Continous Variable (Decimal values)
eg:- Whole number	eg:- Continous
eg: ① No of bank accounts {1, 2, 3, 4, 5} (2.5 X)	Ex:- Height, weight, ages, Rainfall Speed
② nbr of Children	

Assignment:-

1. What kind of Variable is marital status ?
   Categorical

" " " Gauge River length ⇒ Continous

" " " Movie duration ⇒ Continous

" " " Pincode ⇒ Discrete

IQ ⇒ Discretes