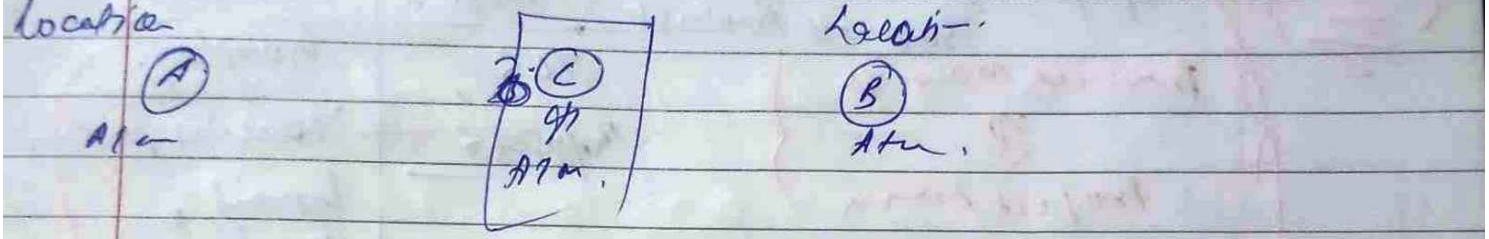
Statistics
- Data Analyst
- Business Analyst
- Data Scientist
- Product Manager

use-case-1

Bank → HDFC



↳ whether they should add this atm at C or not.

This task is for Data Analyst & Data Scientist

use-case-2 Amazon

Find the avg age of the shares throughout the world?

use-case-3 Amazon Big Billion Day Sale. {Instinct}

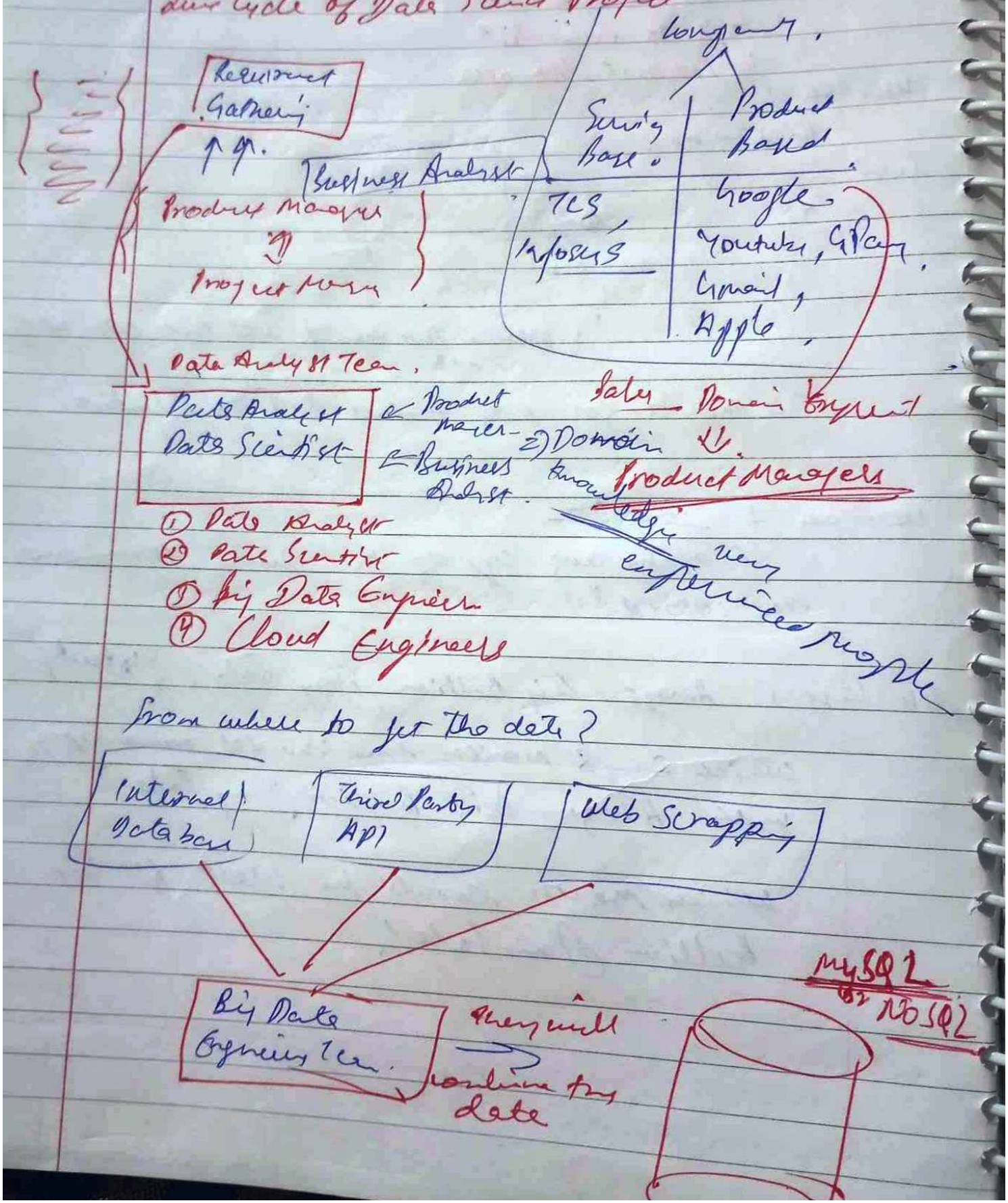
Which Day & month they should come up with Amazon Big Billion Day Sale?

Which month should be select for Big Billion Day Sale!

Statistics.

- We will learn of Data Analyst, Data Scientist,

Life Cycle of Data Science Project :-



Life Cycle of Data Science Projects

Train with ML Algorithms

EDA → Feature Engineering → Feature Selection → Model Training

Exploratory Data Analysis

↓

Hyper Parameter Tuning → Deploying.

(Improve the performance of Model)

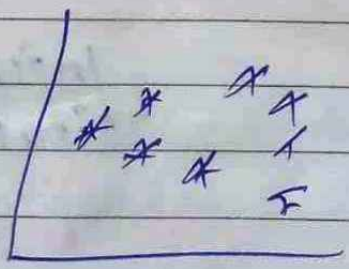
Where is Statistics?

In EDA, In Feature Engineering, In FS, model Training

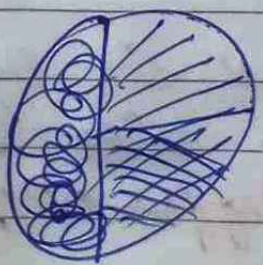
In EDA we do Analysis of Data,

Analysis is done by visualizing the data.

Descriptive Stats.



⇒ Summarizing the data.



⇒ Descriptive Stats.

Age = {12, 13, 14, 14, 20, 25} ⇒ Avg age.

This is part of Descriptive Stats.
We called it measure of central Tendency

Statistics →

Defination → It is a science of collecting, Organising & analysing the data.

Data → facts or piece of information

Eg. Ages of Student in Classroom.

{24, 25, 32, 29, 28}

We can find mean, mode, Median, Standard Deviation.

Eg. Weight of Student in Classroom.

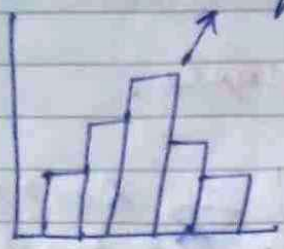
STATISTICS

Descriptive Statistics

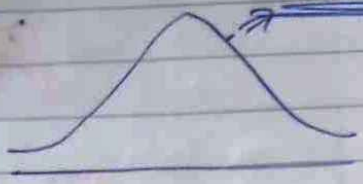
Inferential Statistics

Descriptive Stats :- It consit of organising and summarizing the data. using all the below different type of graphs for eg:

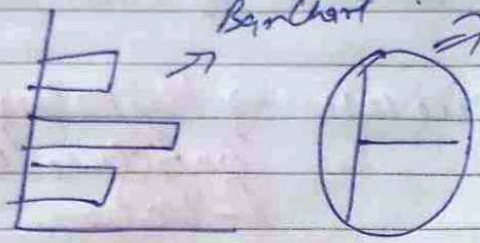
Histogram.



Date Distribution



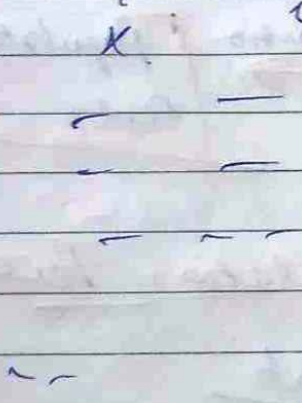
Bar chart. ⇒ Pie

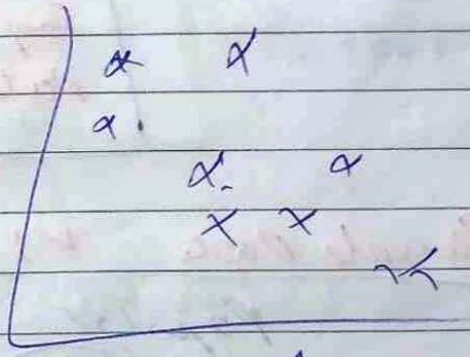


Candle Stick.

Box Plot.

Height Weight





Scatter plot.

INFERENTIAL STATISTICS

It consist of collection sample data and making conclusion about population data. using some experiments.

↳ Hypothesis Testing

University → 500 People

Class A → 60 people.

Sample data → Age ⇒ Average age of entire University

Technique use → Hypothesis test
Confidence Interval
P-Value.

Experiment
1. Z-test
2. T-test
3. Chi Square test.
4. F-test

Sample data → Conclusion → Population data.

Sample Data vs Population Data.

Punjab

Popular

10 cr

10 cr is Population data.

Sample

10k 10k 10k 10k 10k

Exit Poll

Party A

Party B

takes Sample data and come for conclusion for whole population data.

Population (N) Sample (n)

Date

Q Lets say there are 20 classroom in a university & you have collected the ages of students in one classroom.

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

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

Descriptive stats → What is the avg of student in the classroom?

→ Relationship b/w age & weight.

Ans Inferential Statistics :-

Are the avg age of the students in the classroom less than avg age of students in university?

1000 students in University

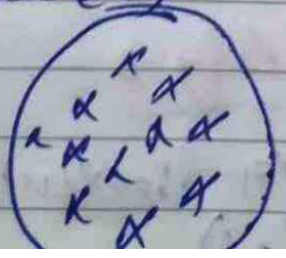
Class A → 50 girls 50 boys

Avg Marks 95% 92%

Can we say girls performed well with boys

Ans What are different Sampling techniques?

① Simple Random Sampling : Every member of the population (N) has equal chance of being selected for your sample (n)



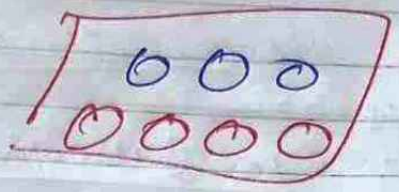
Exit poll → Yes we can use.

Random Sample

General Surveys,
Mobile Reviews,
Lottery
Jackpot.

N = P

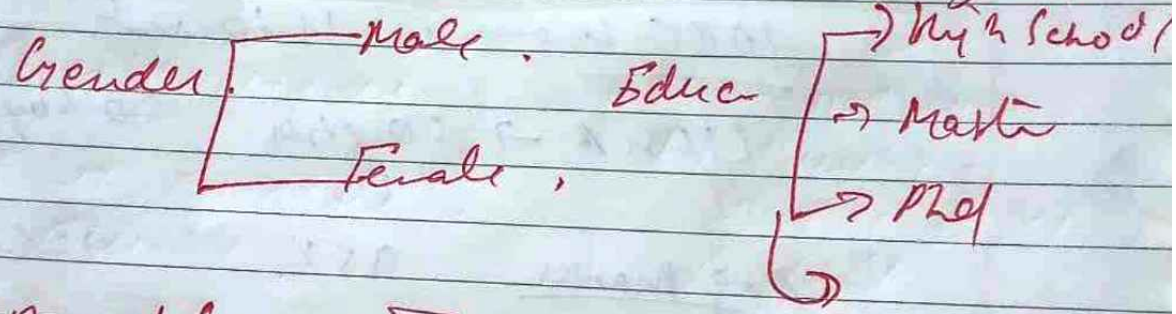
n = 3

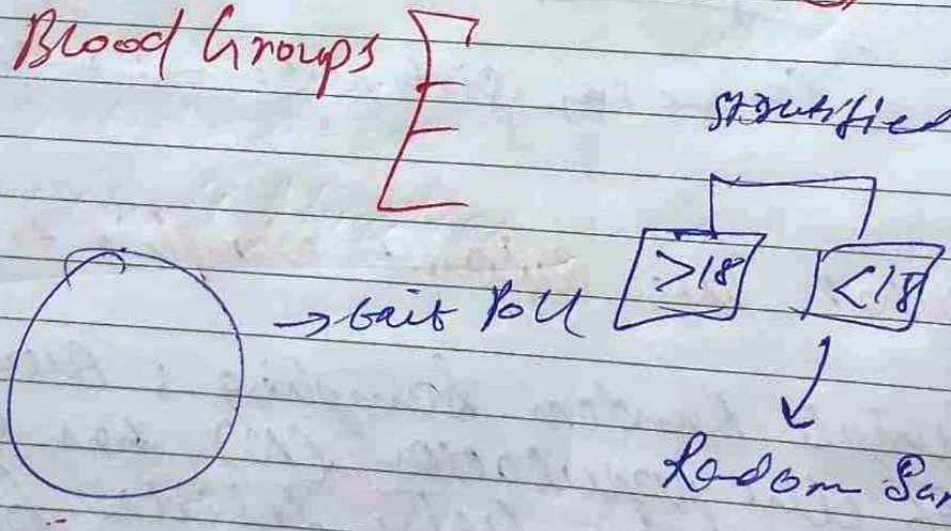


Every possible has equal probability of selection.

② Stratified Sampling (Grouping)

Strata → Men → Layers → Cluster → Groups





③ Systematic Sampling :→

Select every nth individual Or % Population (N).

{AIRLOR}

nth person

{CREDIT CARD}

→ every nth person whether person need credit or not.

every 5th person, also go to approach,

Convenience Sampling → Only those who are interested in survey will only participate in

Person who is relevant to that part we will sample them only.

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

Eg. age = 14
age = 25
age = 100

Variables →
Ages = [24, 25, 26, 27, 28, 29] ⇒ Collection

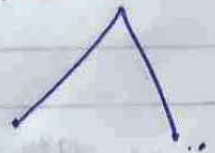
Two types of Variables:

① Quantitative variable → It is measure numerical. {add, sub, mathematical oper

Eg. Age, Weight, No. of students

Date

(2) Qualitative Variable → Categorical Variable



Categorical Variable.

Eg. Gender, Type of Flowers, Type of cars, Type of movie.

Based on some characteristics they are grouped together.

Quantitative Variable.

Discrete Variable

Eg. it cannot be decimal.

Eg. No. of Bank Ac, No. of children in a family

Here 'Whole Numbers'

Continuous

Eg. Height, weight, age, Rainfall, Speed,

Assignment (1) → What kind of variable is marital status.

Categorical Variable.

(2) Gaya zinc Eqn → Continuous.