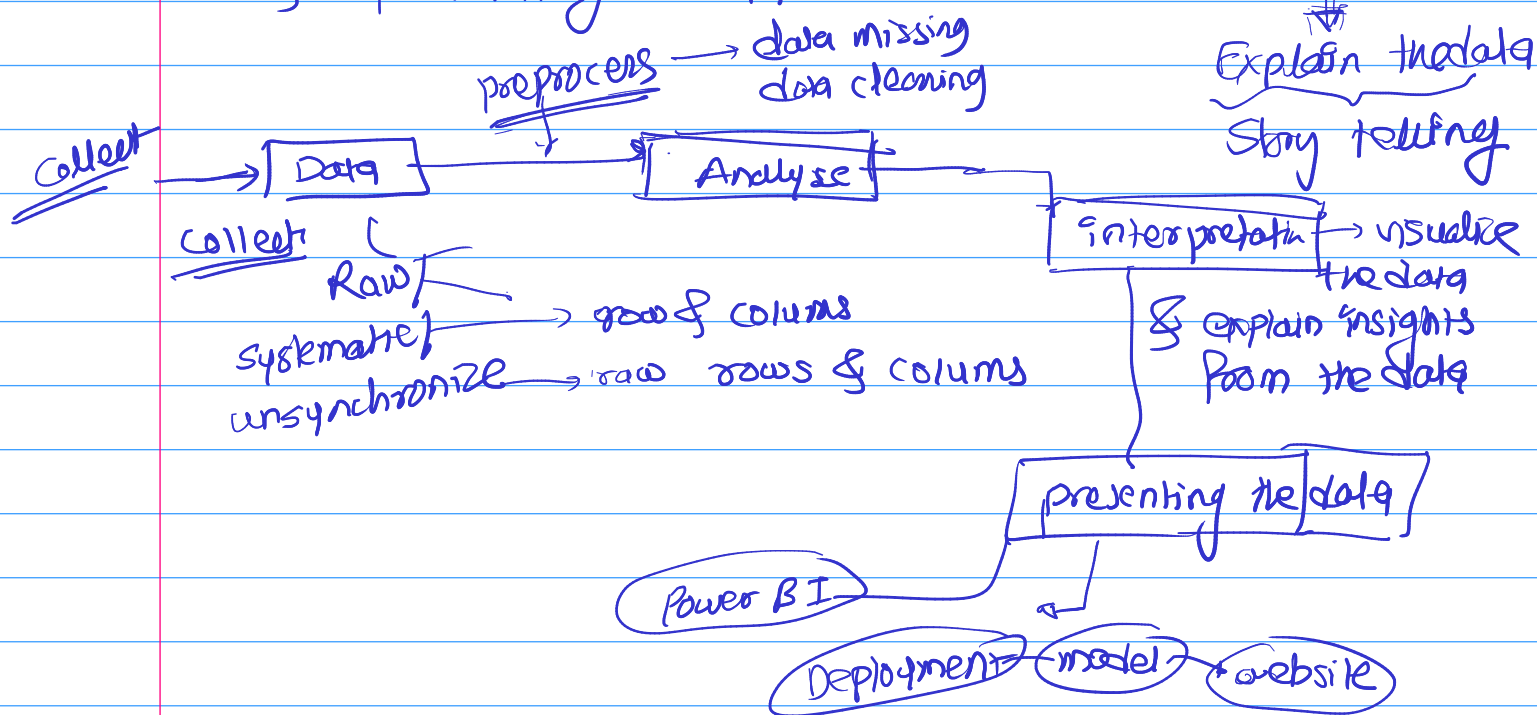
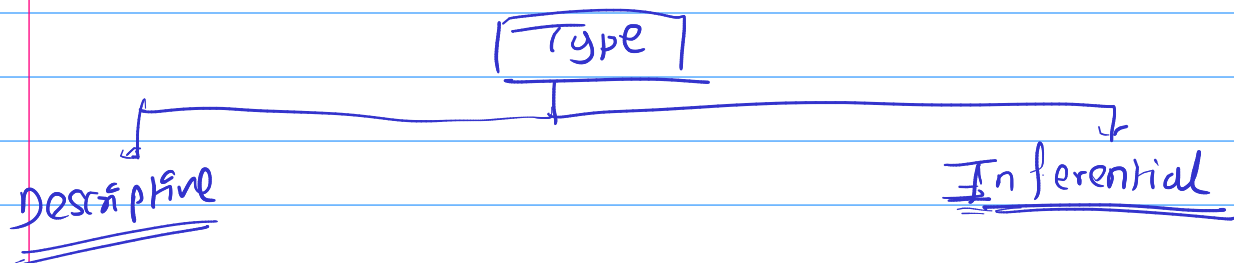


Statistics ? \Rightarrow Statistics is a branch of mathematics that involves collecting, analysing, interpreting & presenting data.



It provide tools and methods to understand & make sense of large amount of data and to draw Conclusion and make decision based on data.

— business, Economics, Social science, medicine & engineering, finance, etc.



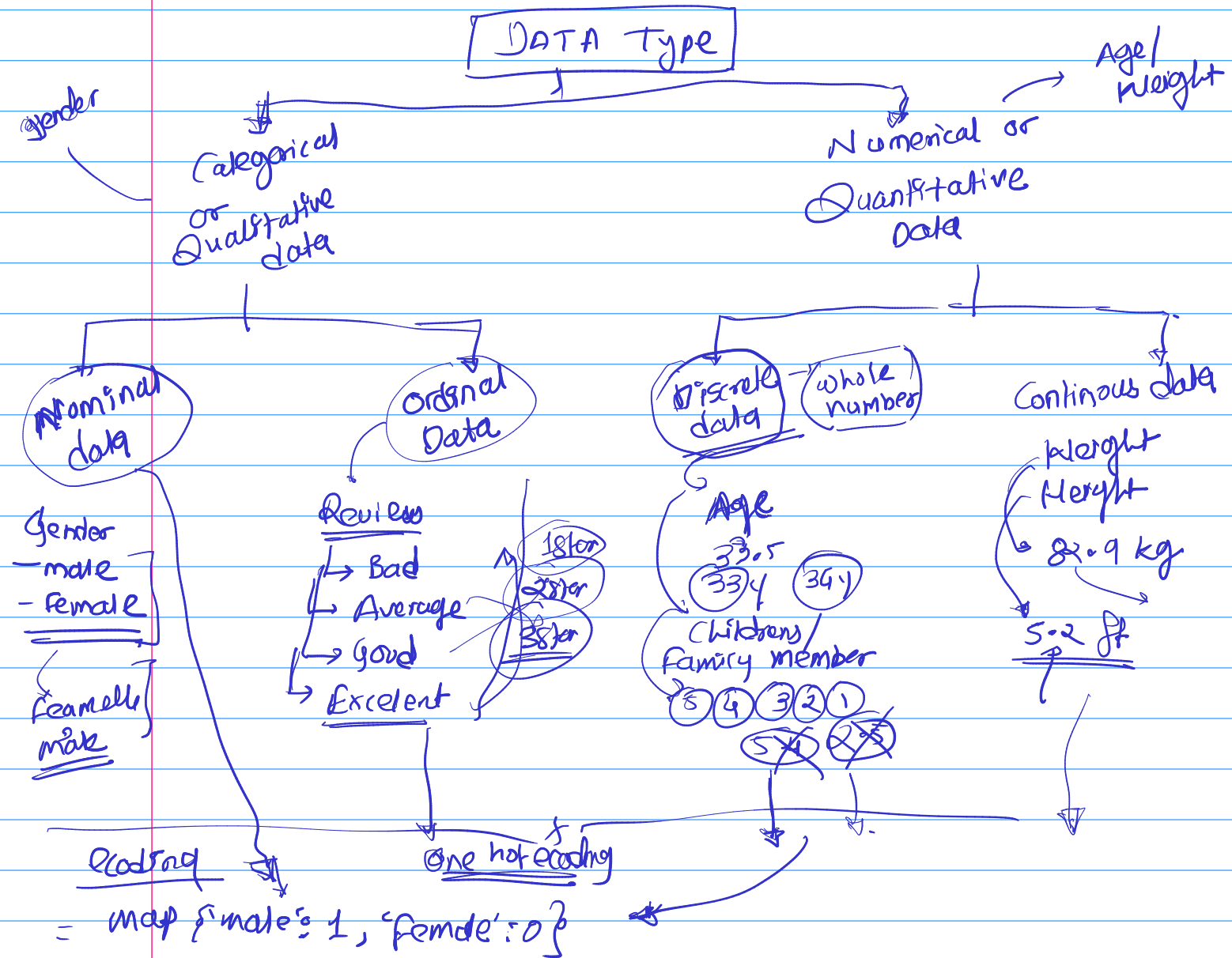
deals with,
collecting, organize, analyse,
interpretation & presentation -
focus \rightarrow on summarizing &
describing the feature of
data.

It deals with making
Conclusion & prediction about
a population based on Sample.

Population & Sample
(N) (n)

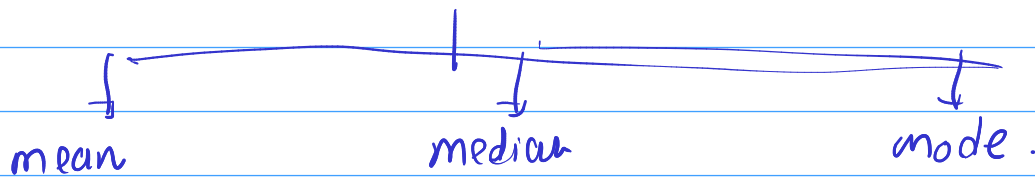
Population refer the entire group of individuals or object that we interested in studying.

Sample — on th. is a subset of a population. It is a smaller group of individual or object.

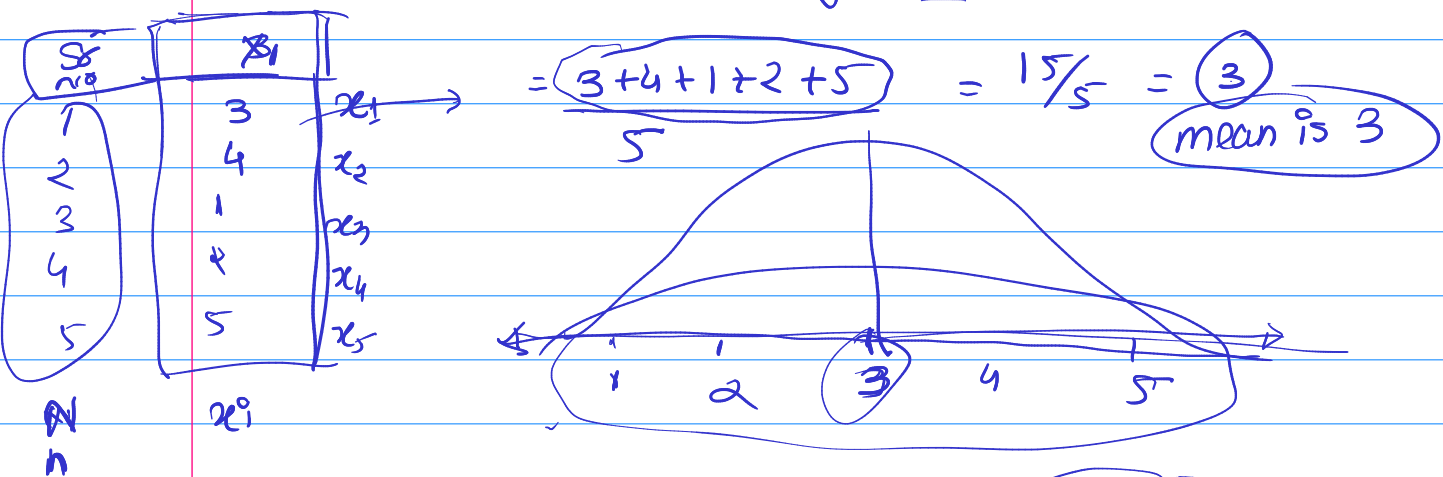


① measure of Central Tendency → center of the data

is a statistical measure that represent a typical or center value for a dataset



① Mean :- The mean is the sum of all value in dataset divided by the number of value.



Sam Symbol

Population μ

$$\mu = \sum_{i=1}^N \frac{x_i}{N}$$

Sample \bar{x}

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

② Median — is the middle value in the dataset when the data is arranged in order

(Average)

1st step - Sort the data

ex = 3, 4, 1, 2, 5

sort \Rightarrow 1, 2, 3, 4, 5

1, 2, 3, 4, 5, 6

$$= \frac{\text{add no.} + 1}{2}$$

$$= \frac{5+1}{2} = 3$$

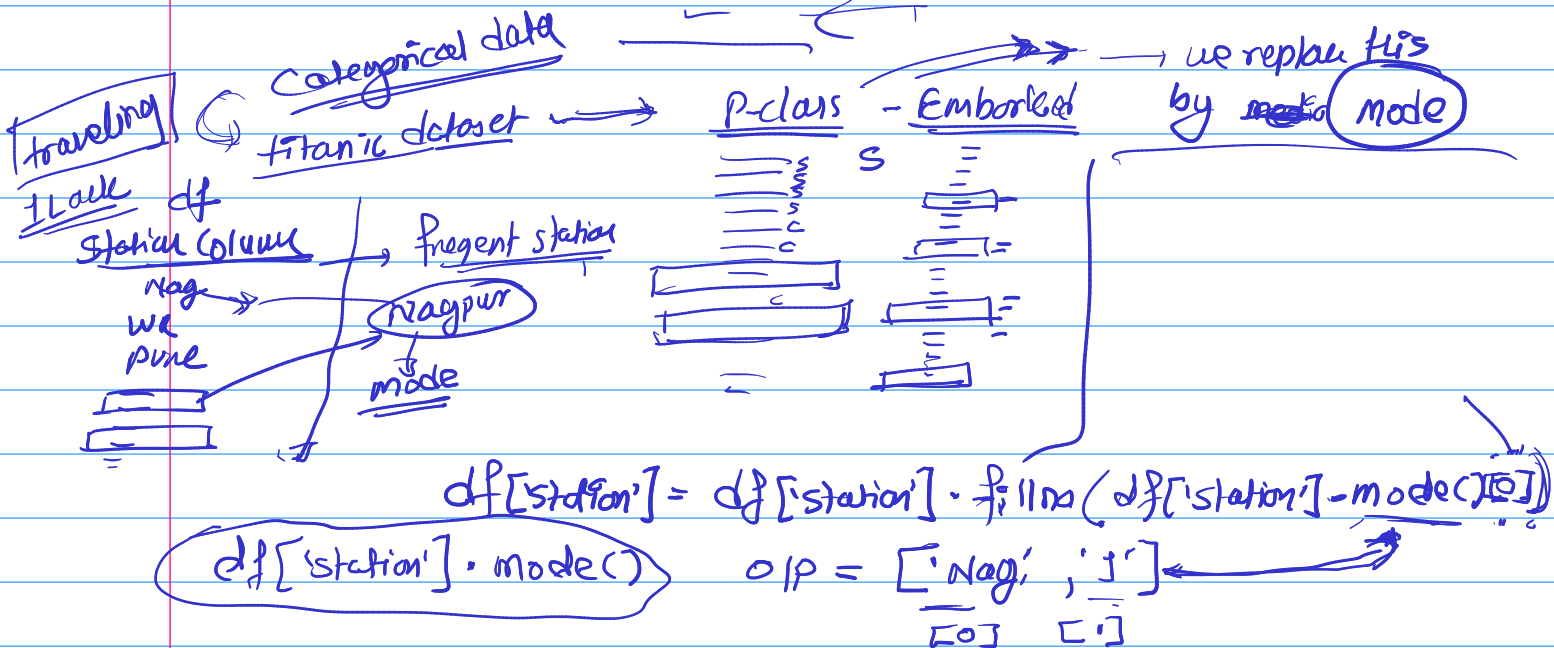
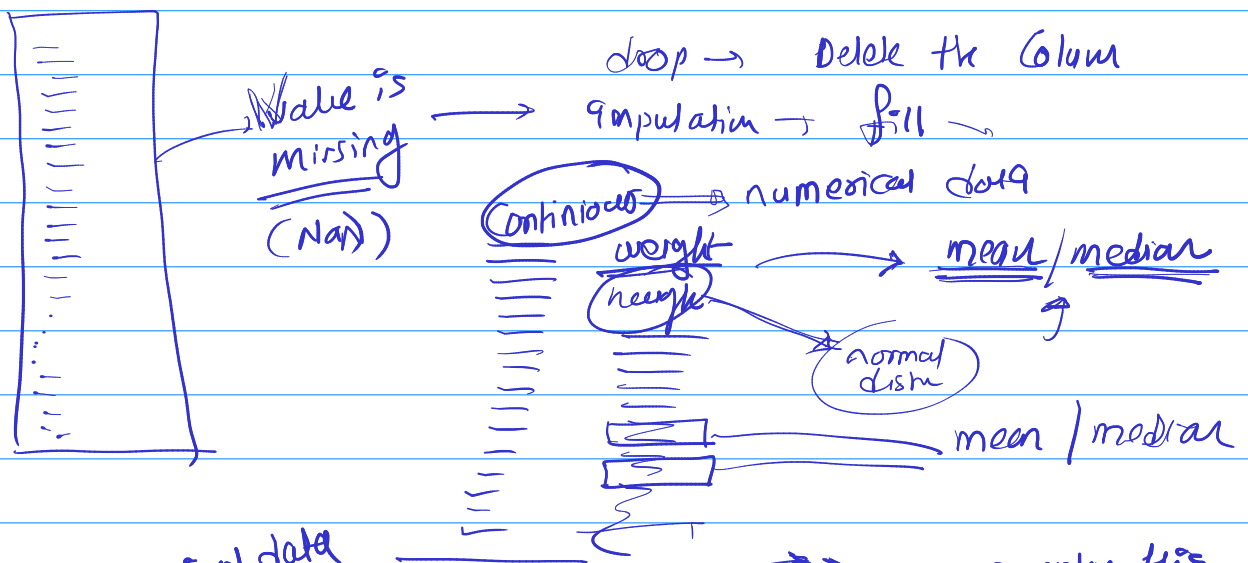
$$= \frac{6}{2} = 3$$

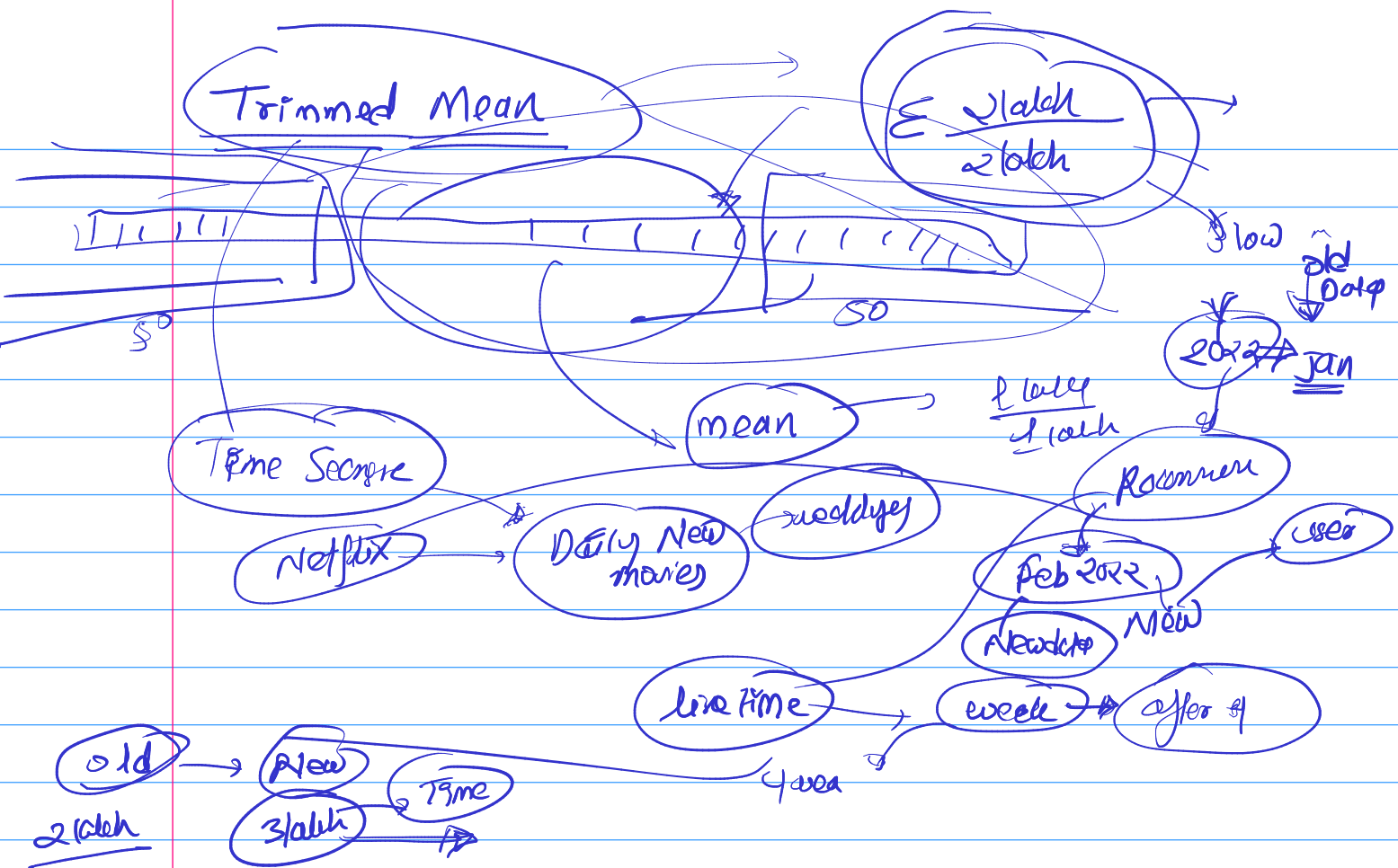
③ Mode - The value that appears the most frequently in dataset

ex = 1, 1, 2, 1, 3, 4, 1, 5, 6, 2, 1

1 → 5 time → most frequent value is 1
2 → 2 time
3 → 1 time

mode = 1

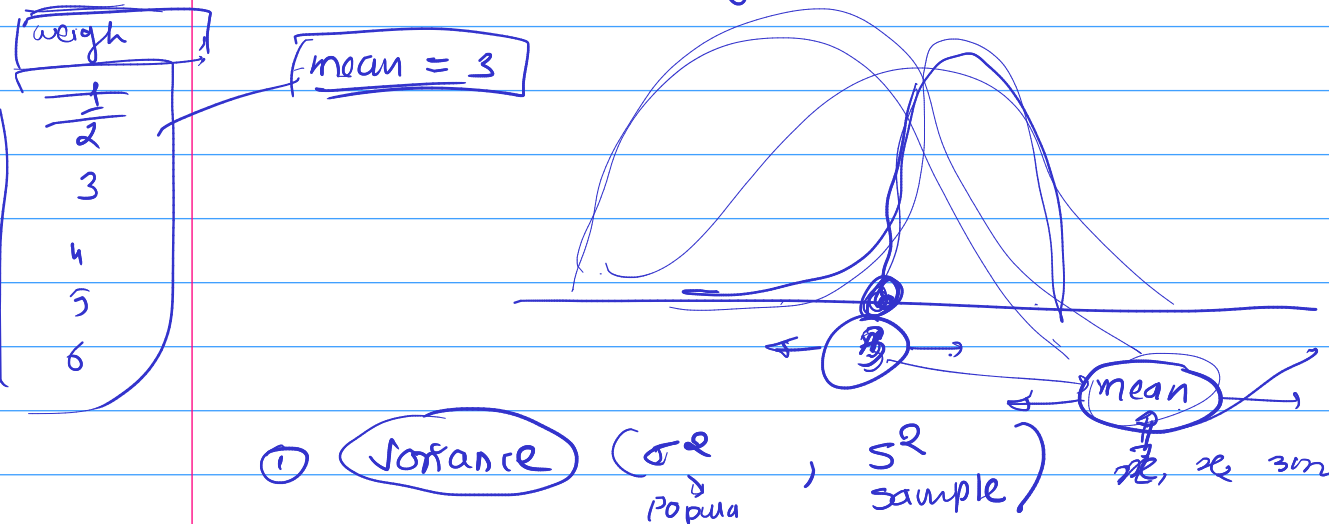




② Measure of Dispersion → Spread

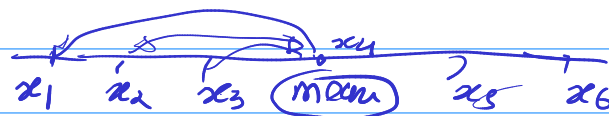
↳ is a statistical measure that describe the spread or variability of data -

why? \Rightarrow we need to find how the data is distributed around central tendency (mean, median, & mode)



① Variance (σ^2 Popula, s^2 sample)

— The variance is the average of squared difference between each datapoint & the mean.



$$\sigma^2 = \frac{\sum_{i=1}^N ((x_i - \mu)^2)}{N}$$

Population Variance

$$s^2 = \frac{\sum_{i=1}^n ((x_i - \bar{x})^2)}{n-1}$$

Sample Variance

Bessel correction
degree of freedom



$$= \frac{(x_i - \mu)^2}{N} = \text{spread}$$

correct
 $n-2$ $n-5$
 $n-3$ $n-100$
 $n-9$ $n-200$
 $\rightarrow n-1$

