Tomorrow we have class

Generative AI ========= > ML/DL/Python

Generative AI certification

Python certification

1. Data related concepts
2. Data is divided into two types

* Numerical
* Categorical

1. Numerical data

* Continues : height weight temp
* Discrete : counting numbers

1. Levels of data

* Nominal
* Ordinal
* Interval
* Ratio

1. Population
2. Sample
3. Population is never exist
4. Resource /Money/Time ====== sample
5. How to represent the data
6. Frequency table
7. Bar graph
8. Pie chart
9. Frequency distribution table
10. Histogram
11. Distribution plot

Data distributon:

First will have raw data === class intervals === how many observations fall in that interval

Histogram

On free line on top of histogram ======== distribution plot

I want understand more about my data

1. Centre of the data
2. How data varies
3. Central tendency

* Mean
* Median
* Mode

1. Data distribution/varies/flows/deviation

* Range
* Mean deviation
* Absolute mean deviation
* Variance
* Standard deviation

1. Central tendency
2. Mean:

Average

Virat Kohli average in ODI is 60 =========

Approximately on every match he scored 60 runs

If he play next match we can expect 60 runs

600

89.5% in SSC?

out of 100% get 89.5%

approximately each subject get above 89,5 marks

out of 100 he score 89.5 marks

there are 6 subjects

91 92 89 90 94 81

Average = sum of observations/total number of observations

= 91+92+89+90+94+81/6= 537/6= 89.5

On of average in each subject I got 89 marks approximately

1. Median

Mid value

Middle point of the data

1,6,9,20,21,26,13

What is the middle point

1. Make the numbers are in ascending or descending order

1 ,6, 9, 13, 20, 21, 26

Median= 50 percentile of data

Middle point of data

If raw data observation are even number ?

1 ,6, 9, 13, 20, 21, 26,30,

Average of two numbers = (13+20) /2= 33/2= 16.5

1 ,6, 9, 13, 20, 21, 26 , 30,30,30,30,30

Assume consider total data is 100 percentile

50 percentile is mid point

**Mean vs Median**

**Both will explain about centre point**

**Both will not explain about min and max points**

**Is both concepts effect by min and max value?**

**Assume that Indian income**

**50k , 60k, 70k,80k, 100k**

**What is the average: 50+60+70+80+100/5= 360/5= 70k**

**What is the median = 70k**

**Assume that Indian income**

**50k , 60k, 70k,80k, 100k, 100crs**

**What is the average: 50+60+70+80+100+10000k/5= 360/5= 7crs**

**Median = 70+80/2= 75k**

**With huge observation either min or max : Mean will affect**

**Median will not affect**

**This huge observations = outliers**

**Ambani is a outlier**

**India ========== middile class is more**

**Zim ========== poverty is more**

**Poverty number is pulling down middile class number when you do average**

**Rich number is pulling up middild class number when you do average**

**Outlier can be min value or can be max value**

**House broker**

**3 house**

**2cr 2.5cr 1.5cr**

**1.5 2 2.5 200crs**

**To develop any model =========== good data**

**How to deal outliers ?**

**Separate session is required**

**If the data provided is more of Outliers , how to deal with it ? Does it has a processing method ?**

1. **Mode:**

**Most occurrence of number**

**1,1,2,2,2,3,3,3,3,4,5,6,7**

**Mode : 3**

**Bar graph and histogram**

**As age progress we have more health problems**

**completing 10th to graduation how many year it takes histogram**

**whenever age increases , strength will decrease**

**u are histogtam and ur father is bar graph u will find more fnds and get attached to ur friend and ur father will work and they don't have time to speak so they get saperated from the family**

**My parents have two child's my uncle have two my material uncle have 3 so three parents are class and their child are frequency we can plot it bar graph**

**Age is directly proportional to Income but indirectly proportional to Energy**

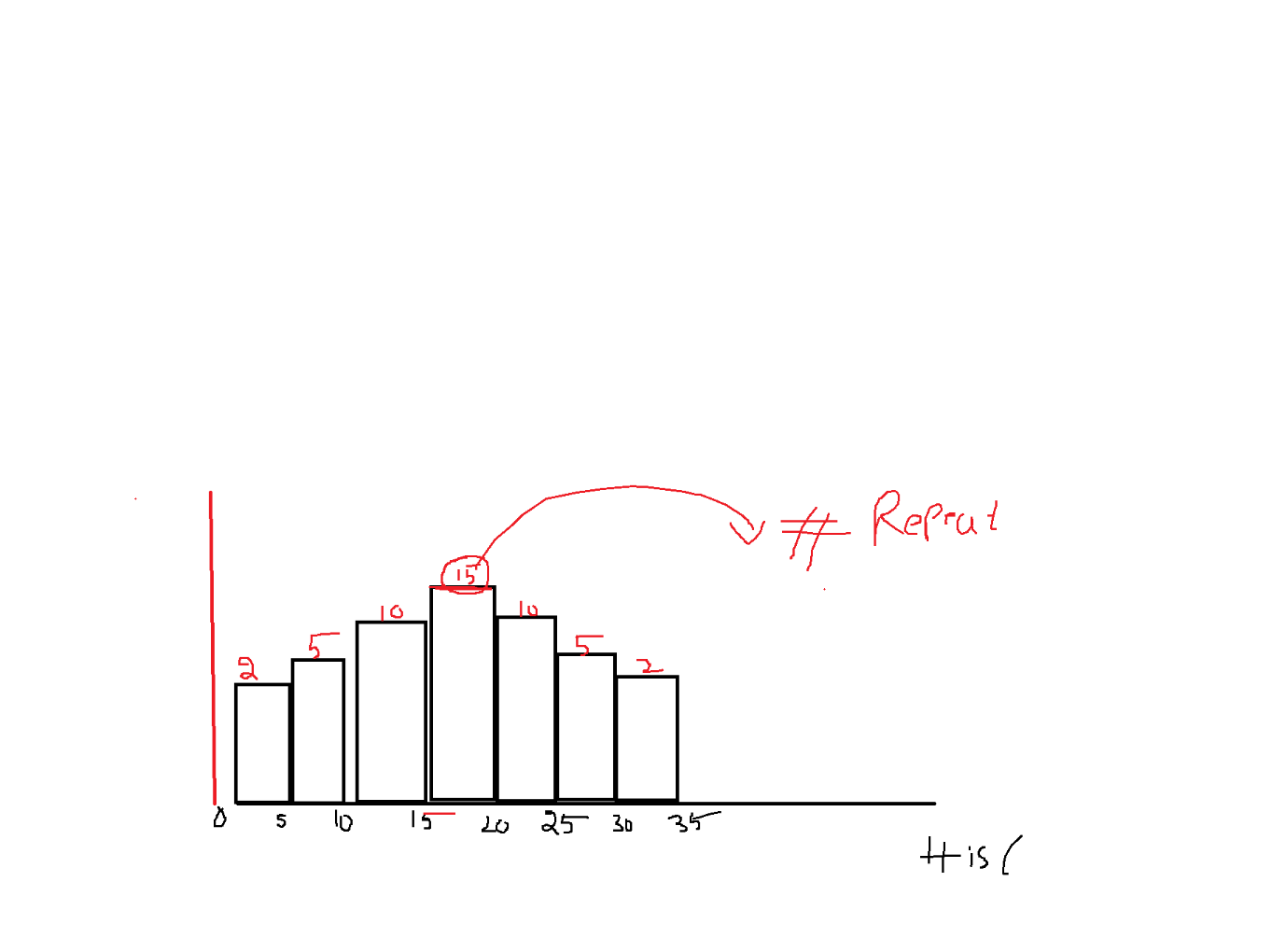
**in 10th 1 year(15 year old) in 12th it take 2 year(16-17 year old) and in gradution it takes 3years(17-20year old) year**

**Think about this?**

**Inner meaning of bar histography**

**Explain ========= tell the story**

**When age is less, we are dependent. When age is medium we are independent again when age is more then we are again dependent**

****

**There is a number between 15 and 20 repeated 15 times**

**Mode is : most reptead number**

**13 ================== 15 times**

**13 is mode ==========? From the graph**

**It provides quick single point idea**

**When the data has numerical data**

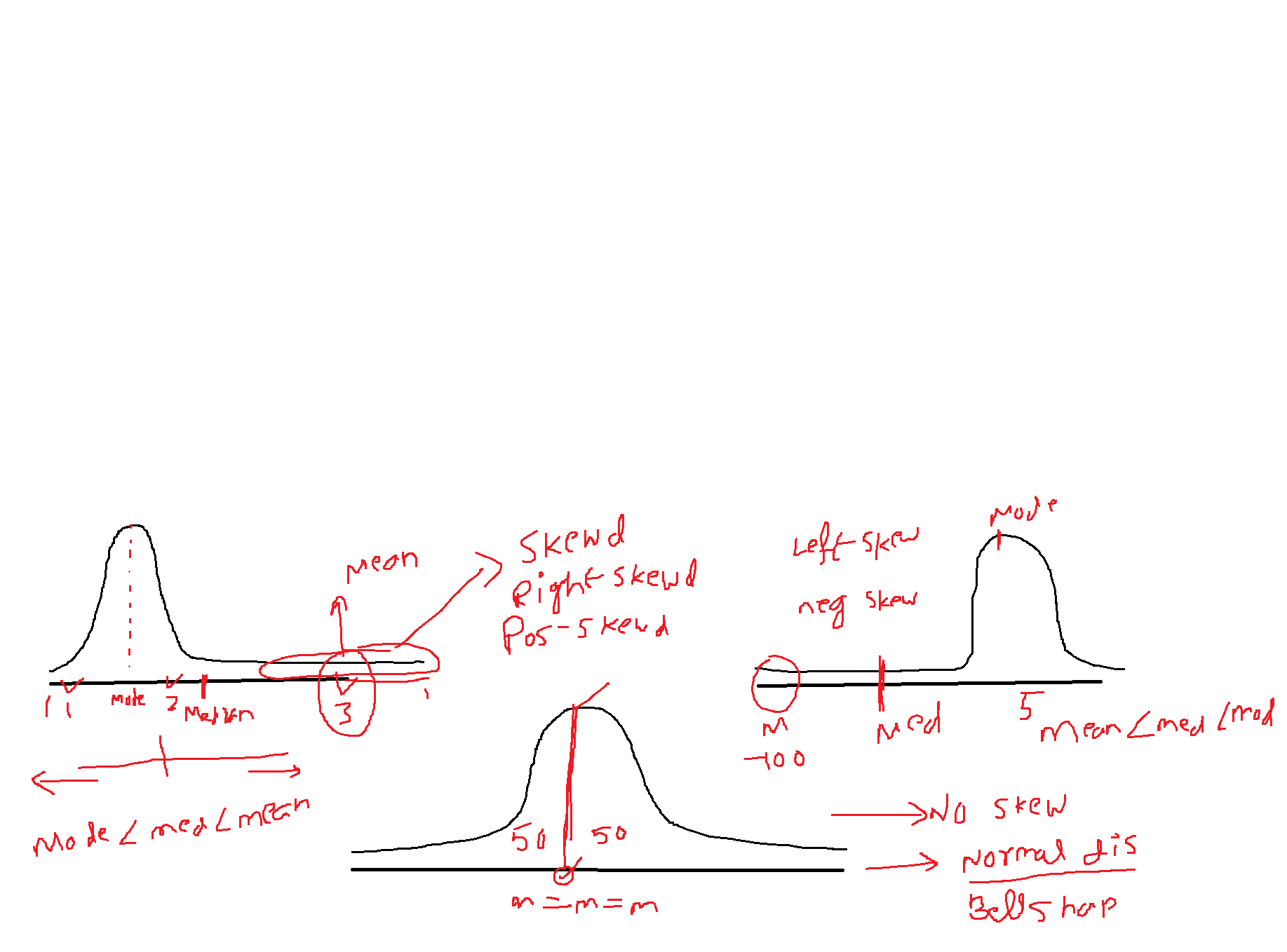
**Whn the data has categorical data**

**10 times you coin a toss**

**H,H,H,H,H,H,T,T,H,H**

**11the time ============ H**

1. **Mean will affect with huge observations either negative or positive value**
2. **Median 50 percentile data**
3. **Mode highest peak of distribution**

****

**We have total 3 distributions**

1. **Left skewed**

* **Imagine real axis left side : negative value**
* **Negative skewd**
* **Skewd means pulling the data**
* **Data pull or skew happen because of outliers**
* **Mean will affect by outliers**
* **Median always 50perenctile data**
* **Mode is highest peak**
* **Mean < Median<Mode**

1. **Right skewd:**

* **Postive skewd**
* **Mean>Median>Mode**

1. **No skew**

* **Normal distribution**
* **Bell shaped curve**
* **50 % of data both sides**
* **Mean= Median=Mode**
* **All the maths developed by assumption on data follows Normal distribution**
* **Any type of distribution will convert into Normal distribution only**
* **Then will apply maths**
* **Any exam : GMAT/GATE/CAT**

1. Class notes
2. PPT
3. Text books