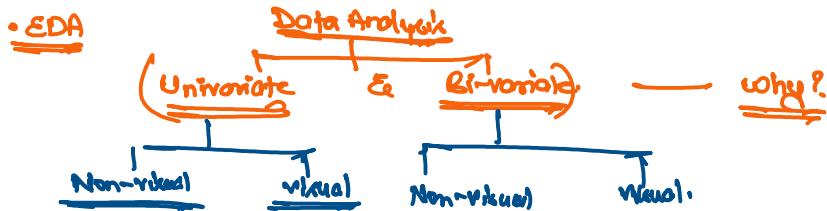


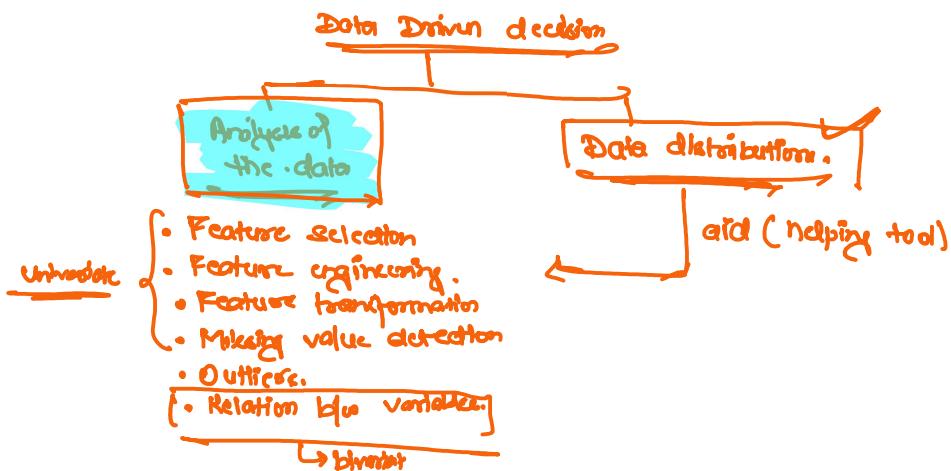
Monday, 8 December 2025 6:01 PM

Recap.

- Python
 - problem solving skills.
 - OOPs (creation).
 - Insight OOPs in real time.

Why?

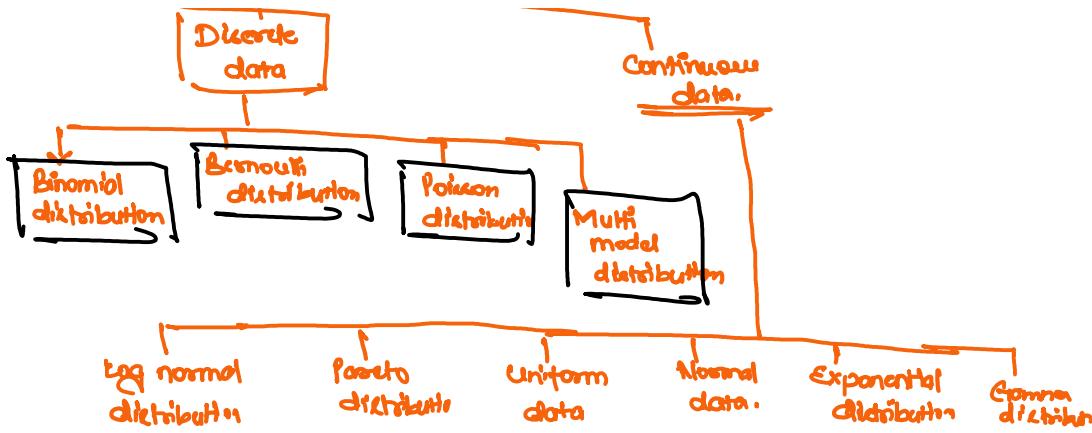
- To understand the data
- To gain insights from the data
- find relationship b/w two - variables { bi-variate }

* Data Distribution

skewness → left skewed or right skewed.
↳ property of the distribution.

- Gaussian distribution } most common form.
 - Normal distribution
- ↳ continuous data

Data.



Q-3 distributions

o Bernoulli Distribution

My data follow Bernoulli distribution.

↳ It exactly contains 2 possible values.

↳ One value represent a success
↳ Other value represent a failure

Ex.

Yes/No
0/1
True/False

Cardinality. → no. of unique value present in column
↳ In bernoulli distribution ↪ 2.

↳ bernoulli distribution

In logistic regression,

↳ by default → we cannot perform multi-class classification.
↳ it can only perform binary classification

why?

Assumption of vanilla (Basic) logistic regression

- Target variable has a cardinality = 2.

Why to learn data distribution.

- Better understanding of ML.
- Stronger at data analysis.

The more you understand distribution, the deeper your understanding of data science becomes.

Applications

- Better data driven decision. ✓
- Better outlier handling. ✓
- Better feature engineering. ✓
- Stronger ML modelling. ✓

Q&A - Q1

820 - 30 → 190

50 - 45 → 5

- Data distribution is not optional. → they are foundational.

In sklearn

Logistic regression (basic) → binary classification

→ multi-class classification.

Addition strategy

↳ OVR (One v/s Rest). }
↳ OVO (One v/s One). },

Tonger

(2) unique value

Let half. → one v/s Rest

1 v/s (2+2)

→ 2/2

Basic logistic regression inherently relies on mapping the tonger variable to Bernoulli distribution, which requires only two classes.
It fails to do multi-class classification in one directly.

We have to use the OVR and OVO mechanism along with logistic regressions to do multi-class classification.

• Real-life Example

Pizza Store dataset:

Order-id	Pizza	Soft drink
1	Margarita	Yes → Success
2	Veggie Delight	No → Failure
3	Panzer Special	Yes
4	Margarita, Four cheese	No
5		No

→ cardinality = 2,
(Bernoulli's distn)

(5)

(2)

- What is the probability of people buying a soft drink with their pizza.

$$P(\text{soft drink} = \text{Yes})$$

$$P(\text{event}) = \frac{\text{No. of favourable outcome}}{\text{Total no. of outcome}}$$

Favourable outcome = 2.

Total no. of outcome = 5.

$$P(\text{soft drink} = \text{Yes}) = 2$$

5

- Probability of failure ($\text{soft drink} = \text{No}$)

$$\begin{aligned} P(\text{Failure}) &= 1 - P(\text{Success}) \\ &= 1 - \frac{3}{5} \\ &= \frac{2}{5}. \end{aligned}$$

In the next 2 orders

Q what is the probability that exactly 2 customers will order a soft drink with their pizza.

→ what is the probability that 2 out of 3 customers will order a soft drink with their pizza. → combination → Bernoulli distribution
 \rightarrow Binomial distribution

Binomial distribution

- A collection of Bernoulli trials.

Bernoulli trial → One event with two outcomes (Success/Failure).

Formula

$$P(\text{exactly } m \text{ successes in } n \text{ trials}) = {}^n C_m p^m (1-p)^{n-m}.$$

where $n = \text{no. of trials}$

$m = \text{no. of successful trials}$.

$p = \text{probability of success}$

$(1-p) = \text{probability of failure}$.

$${}^n C_m = \frac{n!}{m! (n-m)!}$$

$$n = 3$$

$$m = 2$$

$$p = \frac{3}{5}$$

$$(1-p) = \frac{2}{5}$$

$${}^3 C_2 \cdot \left(\frac{3}{5}\right)^2 \left(\frac{2}{5}\right)^{3-2}.$$

$$\frac{3!}{2! 1!} \cdot \left(\frac{3}{5}\right)^2 \left(\frac{2}{5}\right)^1$$

$$\frac{3 \times 2 \times 1}{2 \times 1} \times \frac{4}{25} \times \frac{3}{5}$$

$$\Rightarrow \frac{3 \times 2 \times 1}{25 \times 5} = \frac{3}{125} \approx 0.024$$

$\rightarrow 90 \text{ m}$ $\rightarrow 1\text{m}$

o o o o w w w w