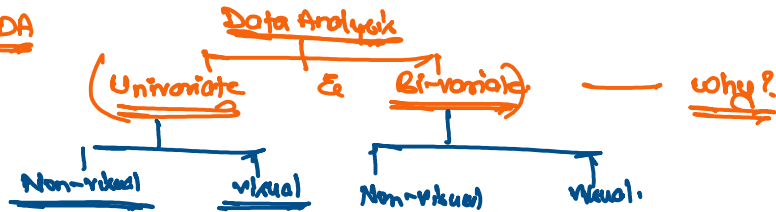


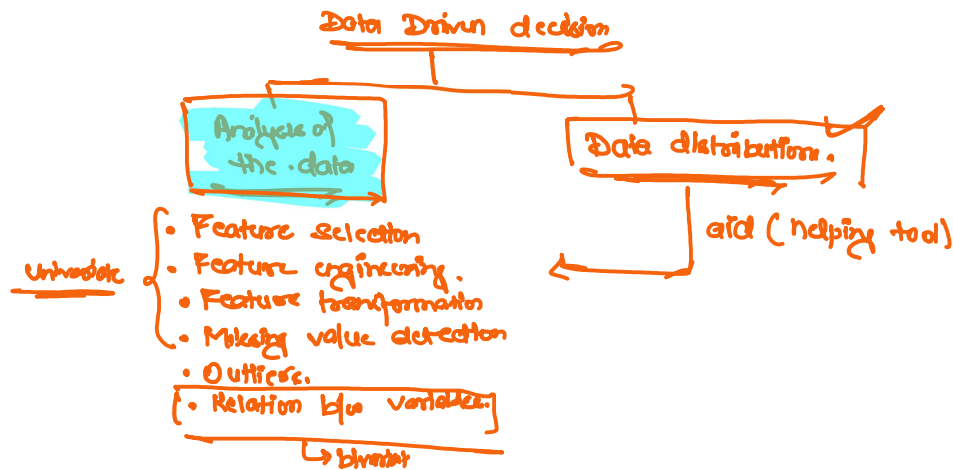
Monday, 8 December 2025 6:01 PM

Recap.

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

EDAWhy?

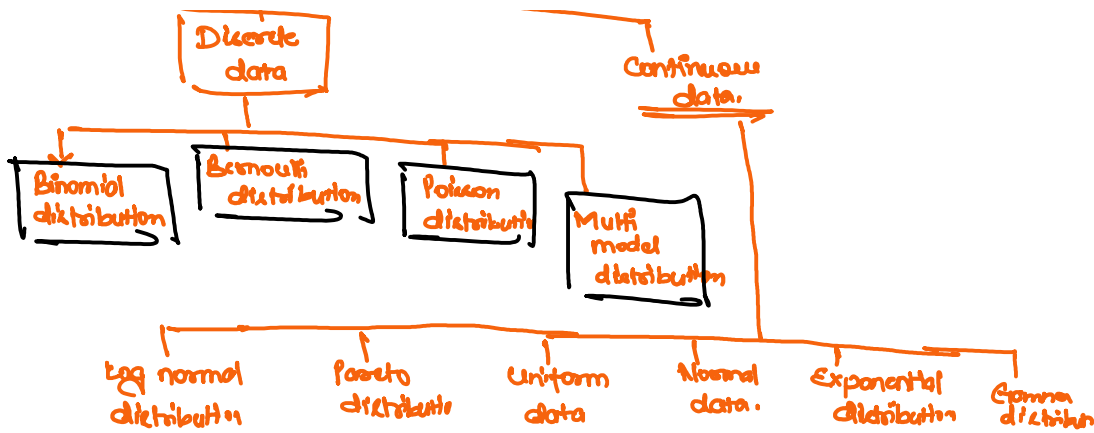
- To understand the data
  - To gain insights from the data
  - find relationship b/w two variables
- } Univariate.  
 } Bivariate.

\* Data Distribution

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

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

Data



## 2-3 distributions

### o Bernoulli Distribution

My data follows Bernoulli distribution

↳ It exactly contains 2 possible values.

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

eg.

Yes/No  
0/1  
True/False

Cardinality. → no. of unique value present in a 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.

### o Application

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

age - age

20-30 → 190

50-60 → 200

- Data distribution is not optional.  $\rightarrow$  they are foundational.

In sklearn

Logistic regression (basis)  $\rightarrow$  Binary classification

addition strategies  $\rightarrow$  multi-class classification.

$\left. \begin{array}{l} \rightarrow \text{OVR (One v/e Rest).} \\ \rightarrow \text{OVO (One v/e One).} \end{array} \right\}$

Target

(2) unique value

Let's talk.

$\rightarrow$  one v/e Rest

1 v/e (2 v 2)

$\rightarrow$  2/2

Basic logistic regression inherently relies on mapping the target variable to Bernoulli distribution, which requires only two classes. If talk to do multi-class classification is used directly.

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

### • Real time Example

Pizza store dataset.

<u>Order-id</u>	<u>Pizza.</u>	<u>Soft drink</u>
1	Margherita	Yes $\rightarrow$ Success
2	Veggie Delight	No $\rightarrow$ Failure
3	Panzer Special	Yes
4	Margherita	No
5	Form house	No

Cardinality = 2.  
(Bernoulli's distri)

(5) { } { } (2)

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

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

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

Favourable outcomes = 2.

Total no. of outcomes = 5.

$$P(\text{soft drink} = \text{Yes}) = \frac{2}{5}$$

5

- Probability of failure (soft drink = No)

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

In the next 3 orders  
 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 → Binomial distribution

- Binomial distribution

- A collection of Bernoulli trials.

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

### Formula

$$P(\text{exactly } x \text{ success in } n \text{ trials}) = {}^nC_x p^x (1-p)^{n-x}.$$

where  $n$  = no. of trials

$x$  = no. of successful trials.

$p$  = probability of success

$(1-p)$  = probability of failure.

$${}^nC_x = \frac{n!}{x! (n-x)!}$$

$$n = 3$$

$$x = 2.$$

$$p = \frac{2}{5}.$$

$$(1-p) = \frac{3}{5}.$$

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

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

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

$$\Rightarrow \frac{3 \times 4 \times 3}{25 \times 5} = \frac{36}{125} \approx 0.288$$

→ 90 ml → 100 ml.

→ 90.0% 20 100%