

Dataframe

eid	esal	ename
1	1000	Rashmi

Dataframe

Series

length-room	breath-room	price
:	:	:
:	:	:
:	:	:

area	price

ML algo

⇒ UR →

dataset1 ⇒

	eid	ecity	esal
→ 1		Mumbai	1000
→ 2		Mumbai	2000
→ 3		Bangalore	2000
→ 4		Mumbai	1000

Group by ecity

SQL ☹️ wrong
Correct ☺️
pandas

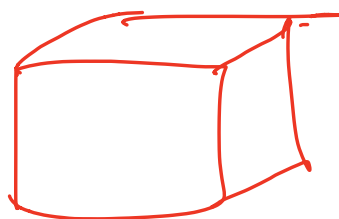
select * from dataset1 group by ecity;

↓ correct

group	eid	esal
Bangalore	3	2000
Mumbai	1	1000
	2	2000
	4	1000

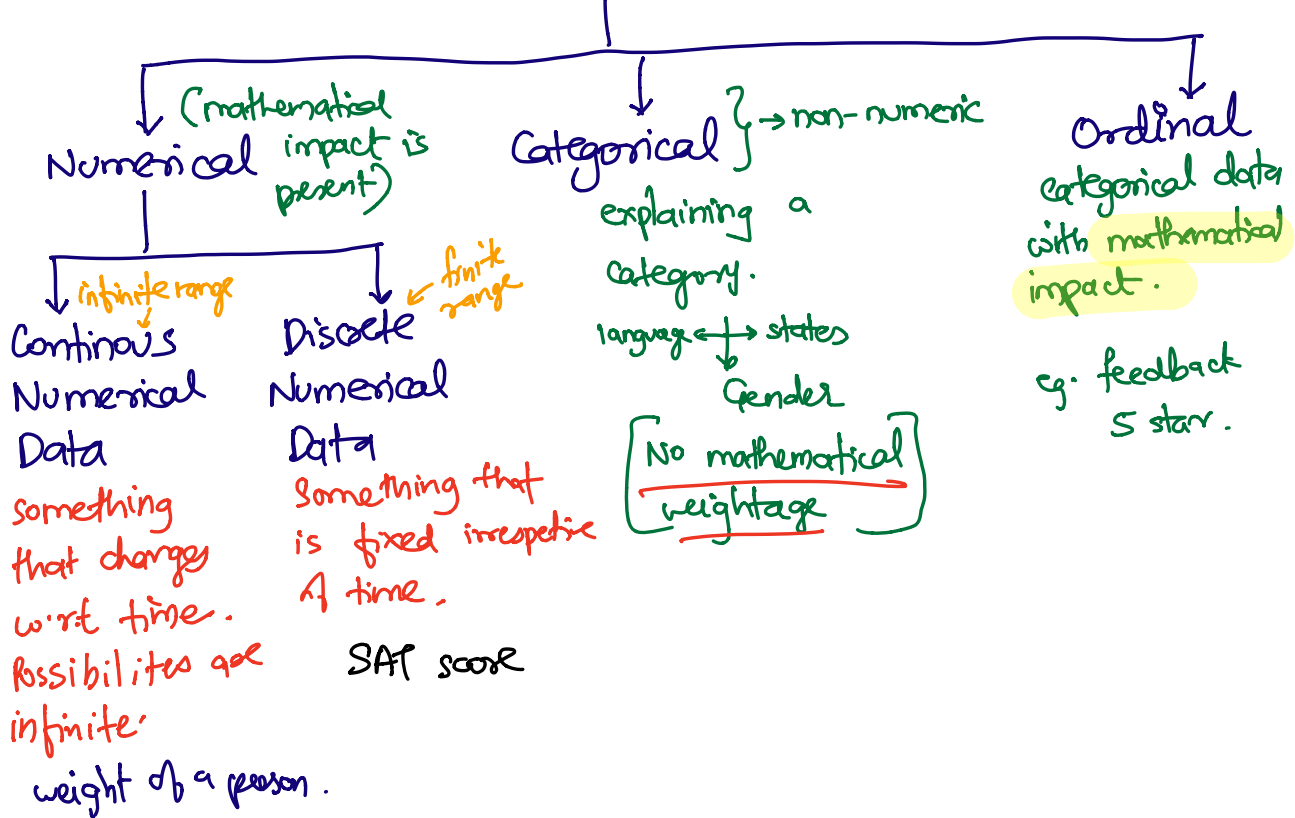
3d table → 3d table

CUBE



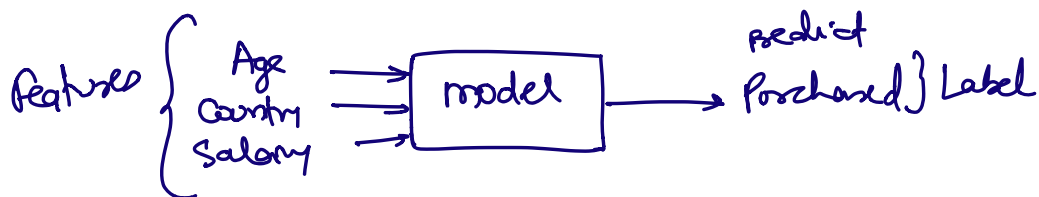
2D

Types of data



Machine Learning.

Features → independent variables → inputs
Label → dependent variables → output



Dummy Variables

dataset \Rightarrow

Country	sal
India	10000
US	20000
India	10000
Africa	10000
US	10000

mathematical
weightage \rightarrow

country	sal
1	10000
2	20000
1	10000
0	10000
2	10000

(Africa)	(India)	(US)	Sal
0	1	0	10000
0	0	1	20000
0	1	0	10000
1	0	0	10000
0	0	1	10000

To create dummy variables following is the algorithm:

- Get the unique values of categorical column and arrange the same in ascending order

0	1	2
Africa	India	US

- Use index of the list as reference and replace the values (Label Encoding)

replacing the values with the index values of the unique list.

- Perform OHE One Hot Encoding to create dummy-variables. OHE remove mathematical weightage introduced in steps.