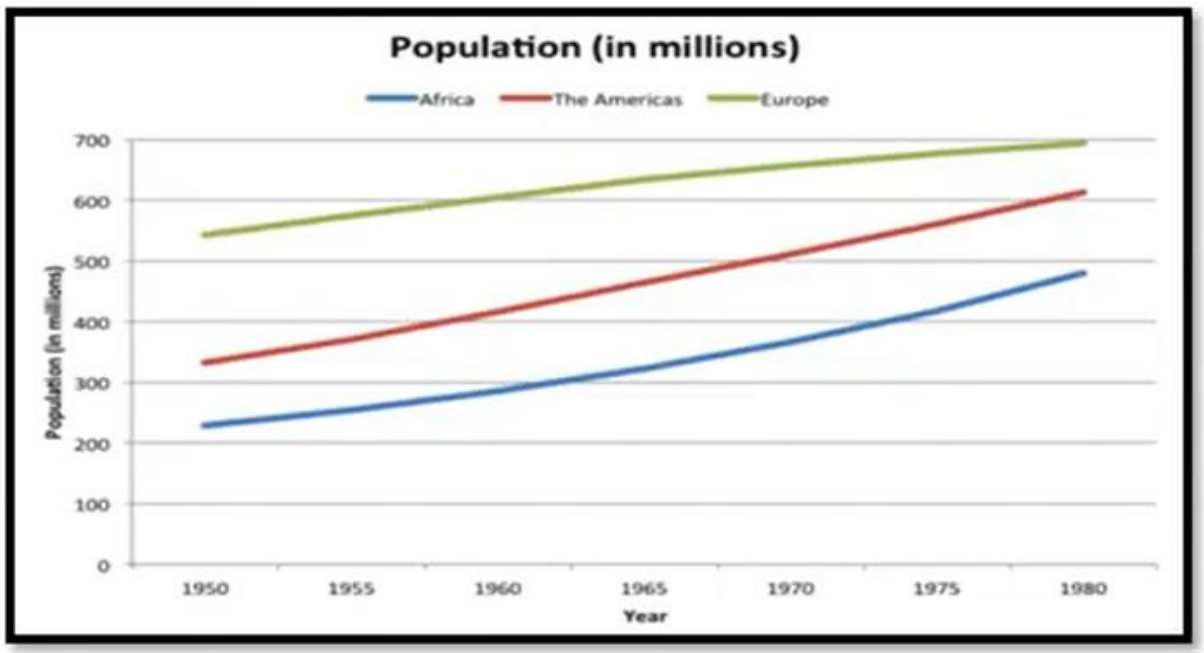
Multiple Linear Regression

Multiple Linear Regression

Dependent variable (DV) Independent variables (IVs)

$$y = b_0 + b_1*x_1 + b_2*x_2 + \ldots + b_n*x_n$$

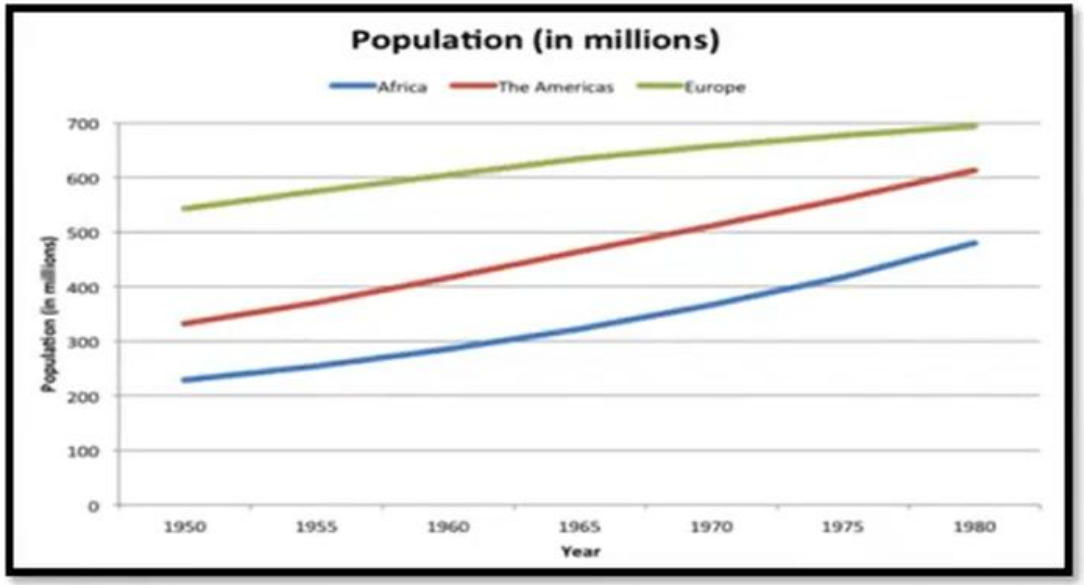


Comparison

Multiple Linear Regression

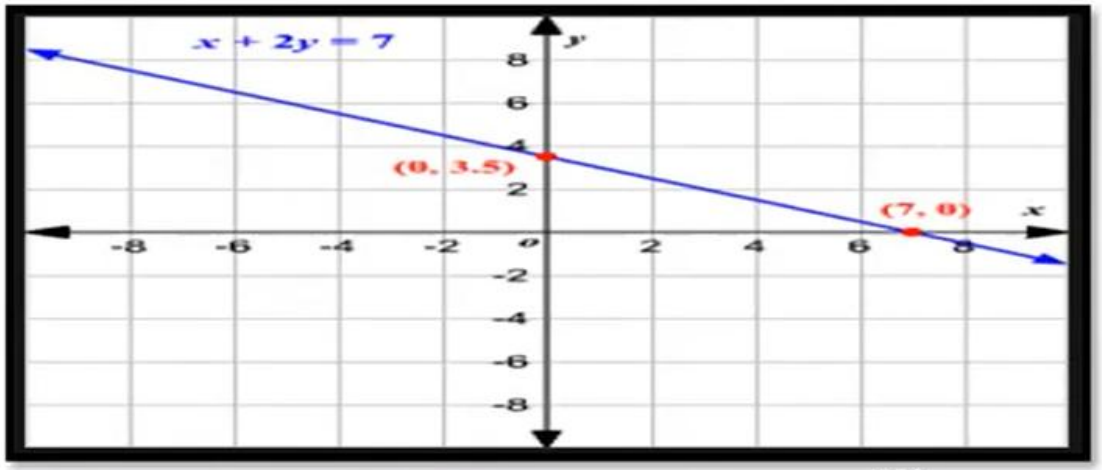
Dependent variable (DV) Independent variables (IVs)

$$y = b_0 + b_1*x_1 + b_2*x_2 + \ldots + b_n*x_n$$



Simple Linear Regression

$$y = b_0 + b_1*x_1$$



Problem Statement
AI in Business Intelligence/Analyst

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12

stage 1 → ML

stage 2 → supervised

stage 3 → Regression

Types of Column in the dataset

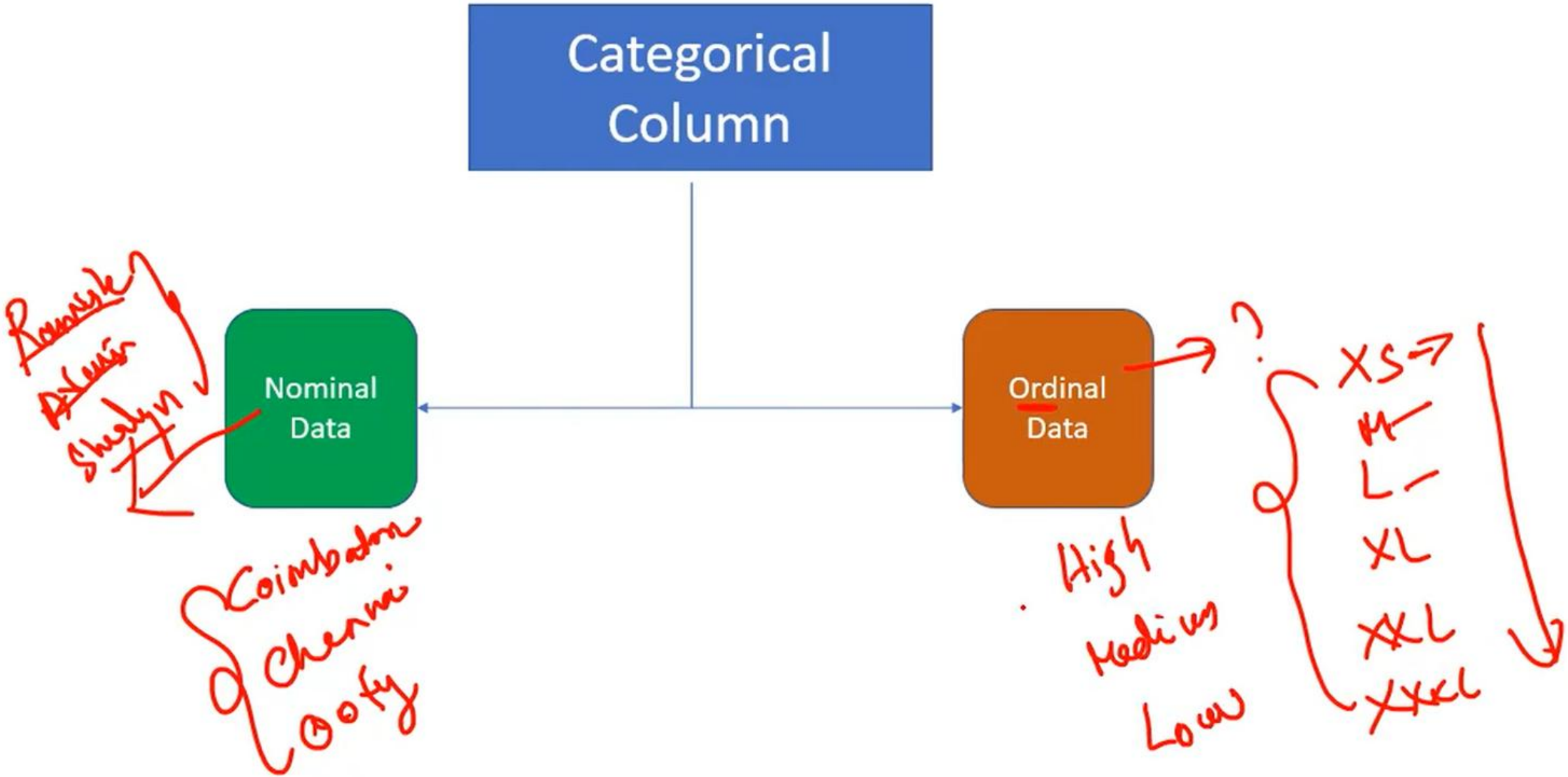
	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12

C- categorical value

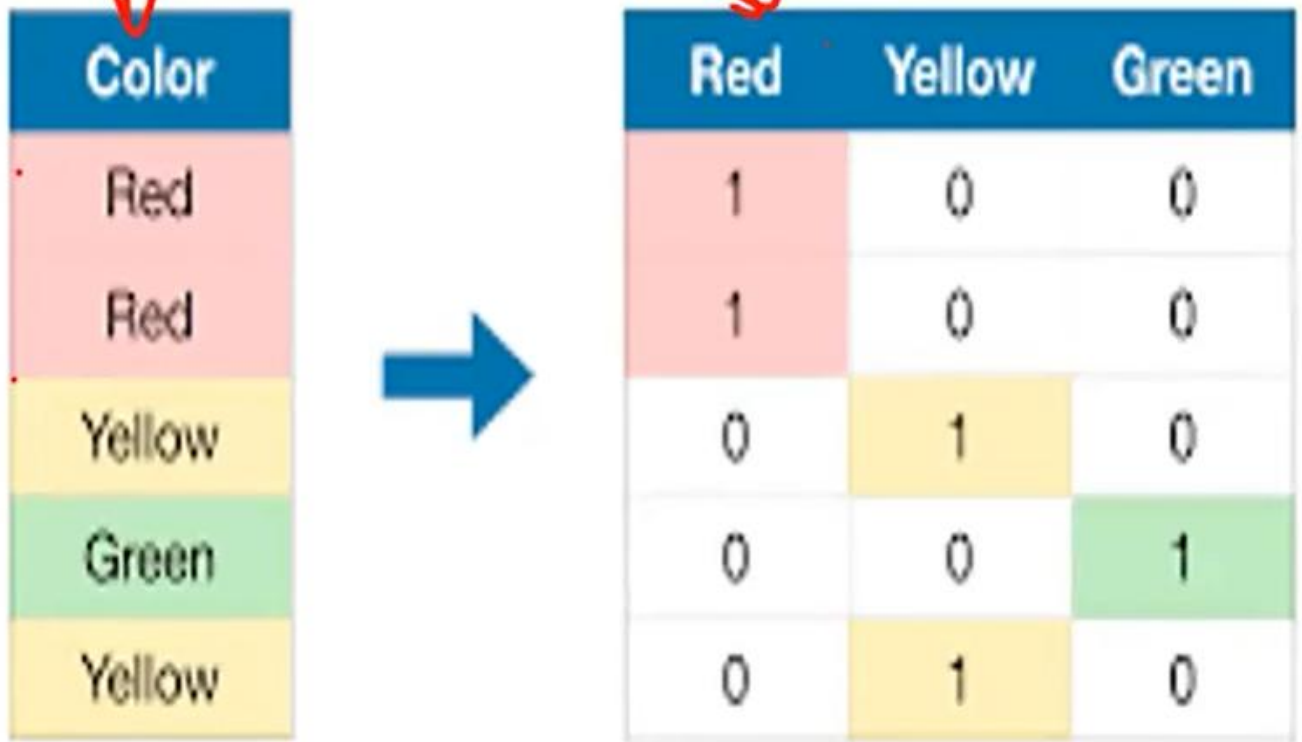
Converted in numeric value

Python → AI → categorical data

How to handle Categorical Column



Categorical Column- Nominal –One Hot Encoding



Color
Red
Red
Yellow
Green
Yellow

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0

Categorical Column- Ordinal –Mapping- Label Encoder

No Expansion

SAFETY-LEVEL (TEXT)	SAFETY-LEVEL (NUMERICAL)
None	0
Low	1
Medium	2
High	3
Very-High	4