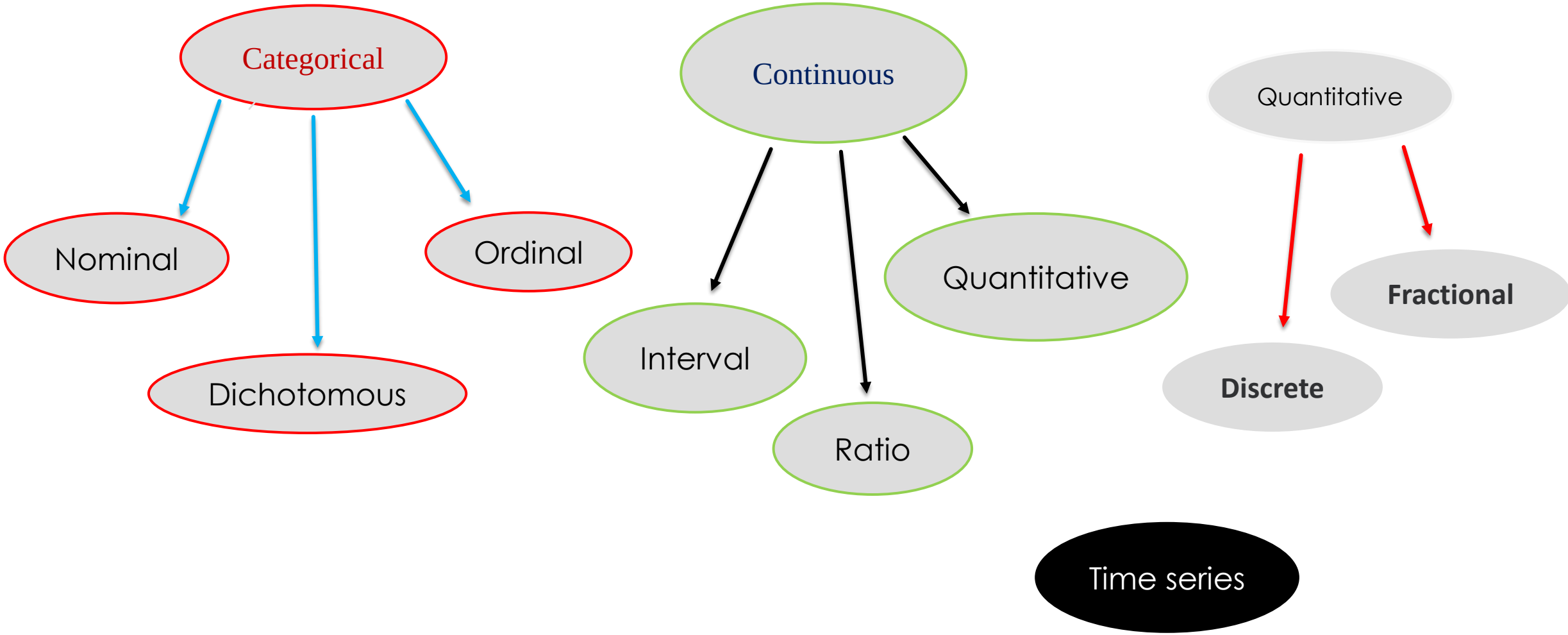
Feature Engineering: Let's talk about-

- Types of Variables
- Measure of Central Tendency
- Encoding Techniques
- Handle NaN Value
- Implementing using Python

Types of Variables in Data Science



Categorical & Continuous Variable

AI Quest

Nominal

1. Colors Names
2. Location Names
3. Car Names

Dichotomous

1. Yes, No
2. Male, Female
3. Left, Right

Ordinal

1. Good >Average> Below
2. High> Medium> Low
3. Satisfied> Neutral> Dissatisfied

Measures of Central Tendency

- **Mean:** The mean is the most popular and well known measure of central tendency.

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

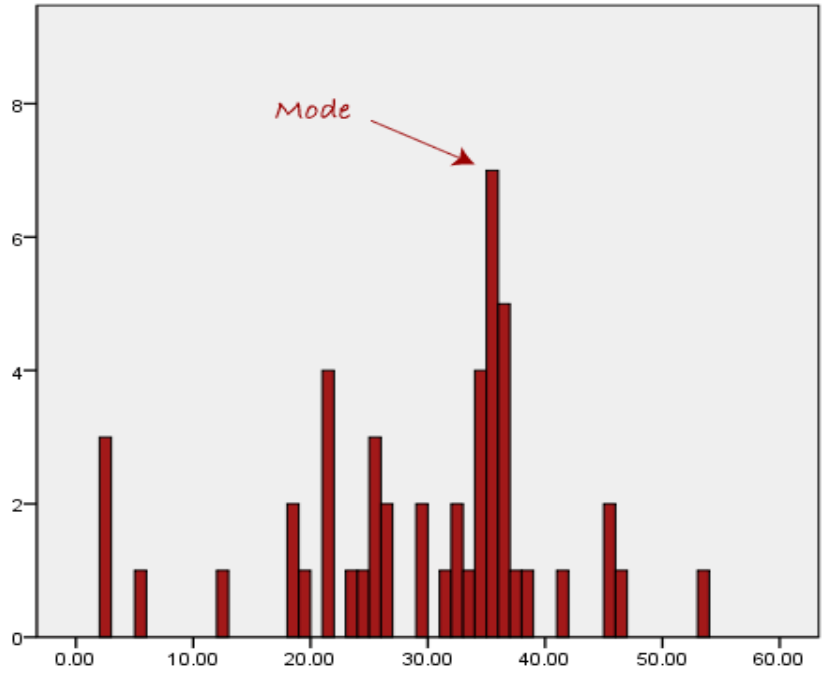
- **Median:** The median is the middle score for a set of data that has been arranged in order of magnitude.

65	55	89	56	35	14	56	55	87	45	92

We first need to rearrange that data into order of magnitude (smallest first):

14	35	45	55	55	**56**	56	65	87	89	92

- **Mode:** The mode is the most frequent score in our data set.



Encoding Techniques in Machine Learning Model

- What is encoding?
- Why it is very important?

Encoding Techniques in Machine Learning Model

Encoding is a technique of converting categorical variables into numerical values so that it could be easily fitted to a machine learning model.

Original Data

Team	Points
A	25
A	12
B	15
B	14
B	19
B	23
C	25
C	29

One-Hot Encoded Data

Team_A	Team_B	Team_C	Points
1	0	0	25
1	0	0	12
0	1	0	15
0	1	0	14
0	1	0	19
0	1	0	23
0	0	1	25
0	0	1	29

Encoding Techniques in Machine Learning Model

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.10	Dhaka	192261.83
1	162597.70	151377.59	443898.53	Ctg	191792.06
2	153441.51	101145.55	407934.54	Rangpur	191050.39
3	144372.41	118671.85	383199.62	Dhaka	182901.99
4	142107.34	91391.77	366168.42	Rangpur	166187.94
5	131876.90	99814.71	362861.36	Dhaka	156991.12
6	134615.46	147198.87	127716.82	Ctg	156122.51

Fig: Before Encoding

Encoding Techniques in Machine Learning Model

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.10	1	192261.83
1	162597.70	151377.59	443898.53	0	191792.06
2	153441.51	101145.55	407934.54	2	191050.39
3	144372.41	118671.85	383199.62	1	182901.99
4	142107.34	91391.77	366168.42	2	166187.94

Fig: After Encoding

Encoding Techniques in Machine Learning Model

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.10	Dhaka	192261.83
1	162597.70	151377.59	443898.53	Ctg	191792.06
2	153441.51	101145.55	407934.54	Rangpur	191050.39
3	144372.41	118671.85	383199.62	Dhaka	182901.99
4	142107.34	91391.77	366168.42	Rangpur	166187.94

Fig: Before Encoding

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.10	1	192261.83
1	162597.70	151377.59	443898.53	0	191792.06
2	153441.51	101145.55	407934.54	2	191050.39
3	144372.41	118671.85	383199.62	1	182901.99
4	142107.34	91391.77	366168.42	2	166187.94

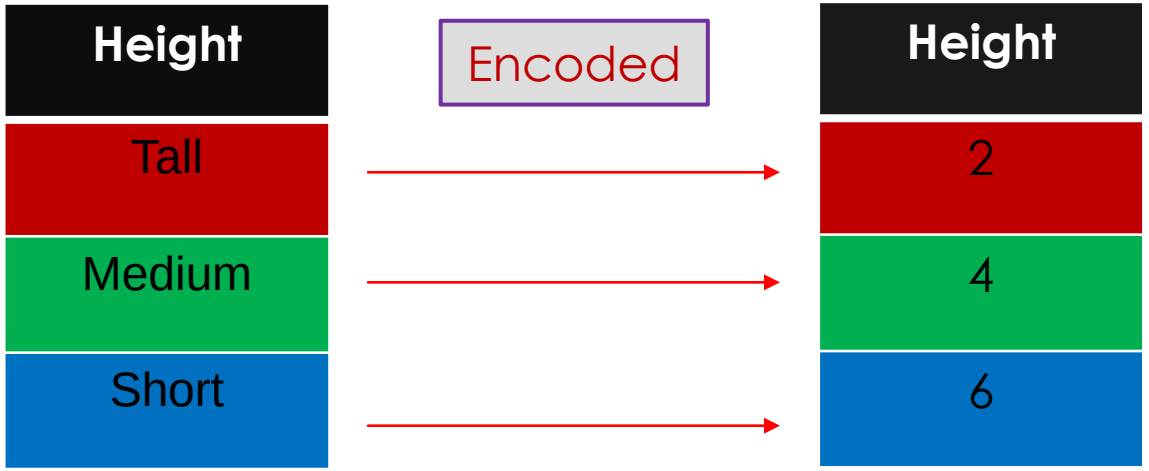
Fig: After Encoding

Encoding Techniques in Machine Learning Model

- Without Use Any Encoding Techniques
- Label Encoding
- One-Hot Encoding
- Ordinal Encoding

Encoding Techniques in Machine Learning Model

Without Use Any Encoding Techniques



Label Encoding

	Marketing Spend	Administration	Transport	Area
0	114523.61	136897.80	471784.100000	Dhaka
1	162597.70	151377.59	443898.530000	Ctg
2	153441.51	101145.55	407934.540000	Rangpur
3	144372.41	118671.85	383199.620000	Dhaka
4	142107.34	91391.77	366168.420000	Rangpur
5	131876.90	99814.71	362861.360000	Dhaka
6	134615.46	147198.87	127716.820000	Ctg
7	130298.13	145530.06	323876.680000	Rangpur
8	120542.52	148718.95	311613.290000	Dhaka

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.100000	1	192261.83
1	162597.70	151377.59	443898.530000	0	191792.06
2	153441.51	101145.55	407934.540000	2	191050.39
3	144372.41	118671.85	383199.620000	1	182901.99
4	142107.34	91391.77	366168.420000	2	166187.94
5	131876.90	99814.71	362861.360000	1	156991.12
6	134615.46	147198.87	127716.820000	0	156122.51
7	130298.13	145530.06	323876.680000	2	155752.60
8	120542.52	148718.95	311613.290000	1	152211.77

One-Hot Encoding

	Marketing Spend	Administration	Transport	Area
0	114523.61	136897.80	471784.100000	Dhaka
1	162597.70	151377.59	443898.530000	Ctg
2	153441.51	101145.55	407934.540000	Rangpur
3	144372.41	118671.85	383199.620000	Dhaka
4	142107.34	91391.77	366168.420000	Rangpur
5	131876.90	99814.71	362861.360000	Dhaka
6	134615.46	147198.87	127716.820000	Ctg
7	130298.13	145530.06	323876.680000	Rangpur
8	120542.52	148718.95	311613.290000	Dhaka

	Ctg	Dhaka	Rangpur
0	0	1	0
1	1	0	0
2	0	0	1
3	0	1	0
4	0	0	1
5	0	1	0

One-Hot Encoding

	Marketing Spend	Administration	Transport	Area
0	114523.61	136897.80	471784.100000	Dhaka
1	162597.70	151377.59	443898.530000	Ctg
2	153441.51	101145.55	407934.540000	Rangpur
3	144372.41	118671.85	383199.620000	Dhaka
4	142107.34	91391.77	366168.420000	Rangpur
5	131876.90	99814.71	362861.360000	Dhaka
6	134615.46	147198.87	127716.820000	Ctg
7	130298.13	145530.06	323876.680000	Rangpur
8	120542.52	148718.95	311613.290000	Dhaka

	Marketing Spend	Administration	Transport	Dhaka	Rangpur
0	114523.61	136897.80	471784.100000	1	0
1	162597.70	151377.59	443898.530000	0	0
2	153441.51	101145.55	407934.540000	0	1
3	144372.41	118671.85	383199.620000	1	0
4	142107.34	91391.77	366168.420000	0	1
5	131876.90	99814.71	362861.360000	1	0
6	134615.46	147198.87	127716.820000	0	0
7	130298.13	145530.06	323876.680000	0	1
8	120542.52	148718.95	311613.290000	1	0

Ordinal Encoding

	Marketing Spend	Administration	Transport	Area
0	114523.61	136897.80	471784.100000	Dhaka
1	162597.70	151377.59	443898.530000	Ctg
2	153441.51	101145.55	407934.540000	Rangpur
3	144372.41	118671.85	383199.620000	Dhaka
4	142107.34	91391.77	366168.420000	Rangpur
5	131876.90	99814.71	362861.360000	Dhaka
6	134615.46	147198.87	127716.820000	Ctg
7	130298.13	145530.06	323876.680000	Rangpur
8	120542.52	148718.95	311613.290000	Dhaka

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.10	0.0	192261.83
1	162597.70	151377.59	443898.53	1.0	191792.06
2	153441.51	101145.55	407934.54	2.0	191050.39
3	144372.41	118671.85	383199.62	0.0	182901.99
4	142107.34	91391.77	366168.42	2.0	166187.94

Now, Let's do it with PYTHON!