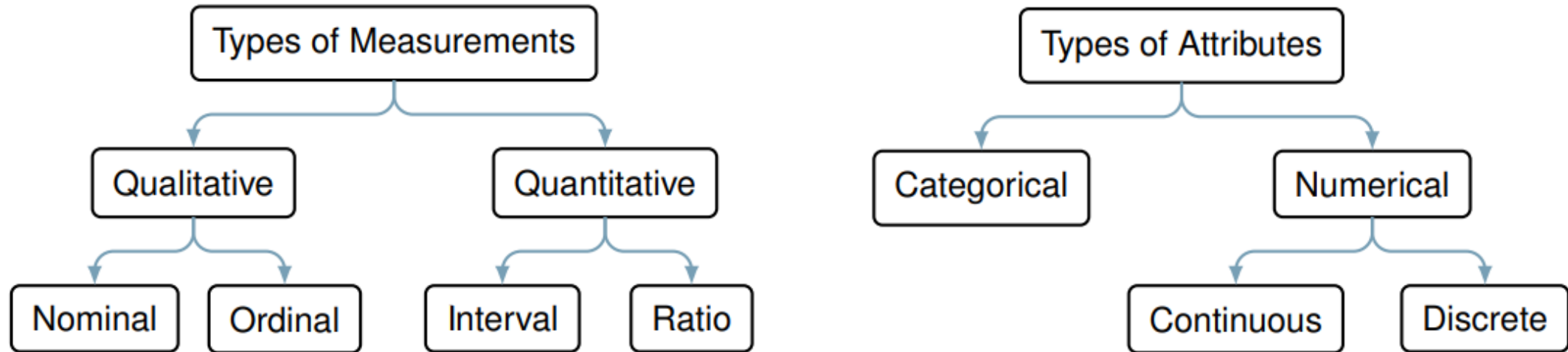


- Types of Variables
- The measure of Central Tendency
- Encoding Techniques
- Handle NaN Value
- Implementing using Python

Two different views:



- *Qualitative* measurements describe an attribute without providing a size or quantity.
- *Quantitative* measurements, often also called *numerical* attributes, are quantitatively measured and often represented in integers or real values.

Measures of Central Tendency

Mean, Median, Mode

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

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

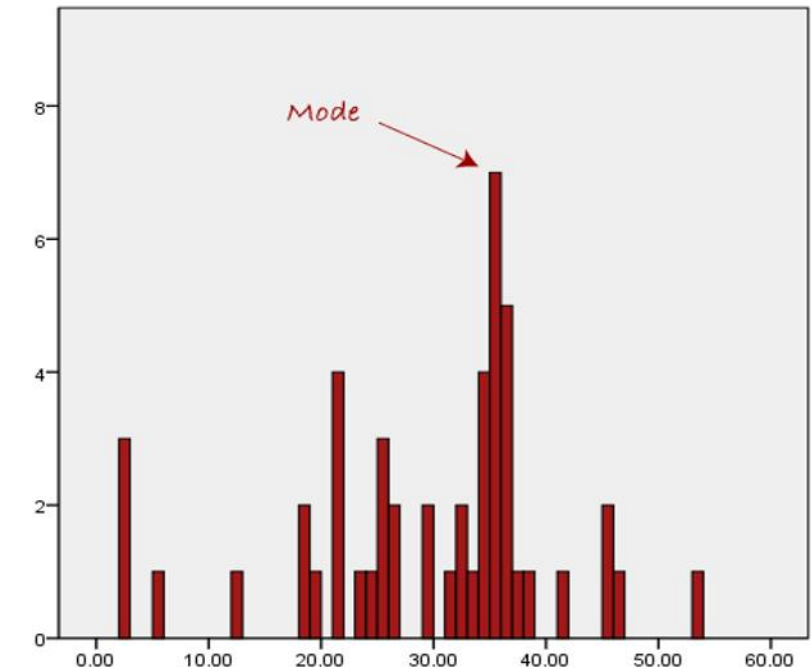
- **Median:** The median is the middle score for a set of data that has been arranged in order of magnitude.

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

Original Data		One-Hot Encoded Data			
Team	Points	Team_A	Team_B	Team_C	Points
A	25	1	0	0	25
A	12	1	0	0	12
B	15	0	1	0	15
B	14	0	1	0	14
B	19	0	1	0	19
B	23	0	1	0	23
C	25	0	0	1	25
C	29	0	0	1	29

	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

BEFORE

	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

AFTER

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

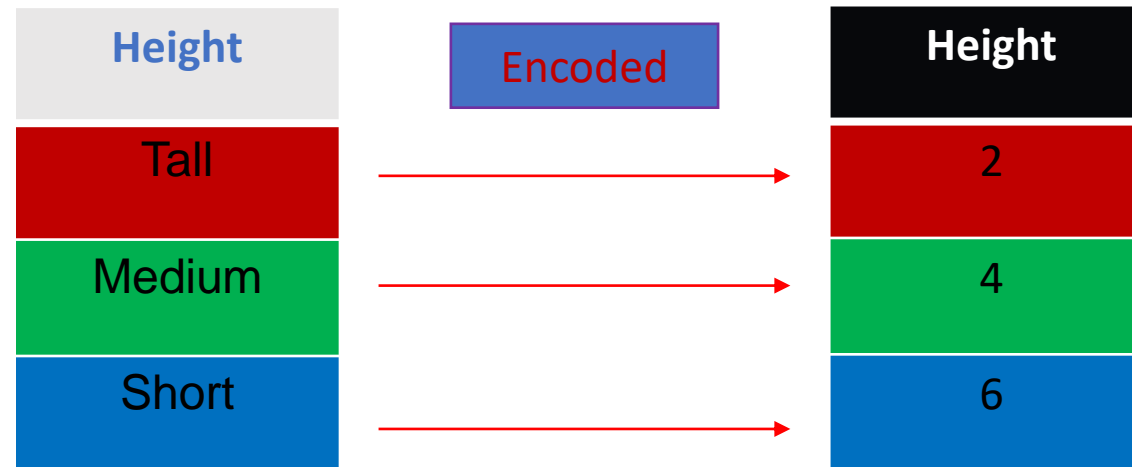
BEFORE

AFTER

- Without the use of Any Encoding Techniques
- Label Encoding
- One-Hot Encoding
- Ordinal Encoding

- ❖ **Label encoding** is the simplest of the three techniques. It simply assigns a numerical value to each category. For example, the categories "Dhaka," "Ctg." and "Rangpur" might be assigned the values 1, 2, and 3, respectively.
- ❖ **One-hot encoding** is a more sophisticated technique that creates a new binary variable for each category. Each category is then represented by a vector of 0s and 1s, where the 1 is in the position corresponding to that category. For example, the category "Dhaka" would be represented by the vector [1, 0, 0], the category "Ctg." would be represented by the vector [0, 1, 0], and the category "Rangpur" would be represented by the vector [0, 0, 1].
- ❖ **Ordinal encoding** is a technique that preserves the order of the categories. It assigns numerical labels to each category in a meaningful order. For example, you might assign the labels 1, 2, and 3 to the categories "Dhaka," "Ctg." and "Rangpur," respectively.

REPLACE FUNCTION



LABEL ENCODER

	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 ENCODER

	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 ENCODER (DUMMY VAR TRAP)

	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 ENCODER

	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