In [1]:

```
import pandas as pd
dataset=pd.read_csv(r"C:\Users\RacksonsIT\Desktop\Data Science Session\Batch 8 Symboisis 6
```

In [2]:

```
dataset
```

Out[2]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	RE
0	Owner	0	0	2	BHK	1300.236407	
1	Dealer	0	0	2	BHK	1275.000000	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
4	Dealer	1	0	2	BHK	999.009247	
...	...	...	...	...	...	...	
29446	Owner	0	0	3	BHK	2500.000000	
29447	Owner	0	0	2	BHK	769.230769	
29448	Dealer	0	0	2	BHK	1022.641509	
29449	Owner	0	0	2	BHK	927.079009	
29450	Dealer	0	1	2	BHK	896.774194	

29451 rows × 12 columns

In [3]:

```
dataset["POSTED_BY"].unique()
```

Out[3]:

```
array(['Owner', 'Dealer', 'Builder'], dtype=object)
```

In [4]:

```
ab=[]
for x in dataset["POSTED_BY"]:
    if x=='Owner':
        ab.append(0)
    elif x=='Dealer':
        ab.append(1)
    else:
        ab.append(2)
print(ab)
```

```
[0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1,
1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 2,
1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0,
0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1,
0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1,
1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0,
1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1,
1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0,
0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1,
1, 1, 0, 1, 1, 1, 2, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 2, 1, 1, 1,
1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 2, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1,
0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 2, 1, 1, 0, 0,
1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0,
1, 0, 1, 1, 0, 2, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 2, 1, 0, 0, 0, 1, 1, 1, 1,
1, 0, 1, 0, 1, 1, 2, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 2, 0, 1, 1, 1, 0,
2, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1,
1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1,
1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1,
1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 2,
```

In [5]:

```
dataset["Converted"]=ab
```

In [6]:

```
dataset
```

Out[6]:

Y_TO_MOVE	RESALE	City	LONGITUDE	LATITUDE	TARGET(PRICE_IN_LACS)	Converted
1	1	Bangalore	12.969910	77.597960	55.0	0
1	1	Mysore	12.274538	76.644605	51.0	1
1	1	Bangalore	12.778033	77.632191	43.0	0
1	1	Ghaziabad	28.642300	77.344500	62.5	0
0	1	Kolkata	22.592200	88.484911	60.5	1
...	...	...	...	...	...	...
1	1	Agra	27.140626	78.043277	45.0	0
1	1	Lake View Recidency	39.945409	-86.150721	16.0	0
1	1	Jaipur	26.928785	75.828002	27.1	1
1	1	Chennai	12.900150	80.227910	67.0	0
1	1	Jaipur	26.832353	75.841749	27.8	1

In [8]:

```
ab1=[]
for x in dataset["TARGET(PRICE_IN_LACS)"]:
    x1=x+10
    ab1.append(x1)
dataset["Updated_Price"]=ab1
```

In [9]:

```
dataset
```

Out[9]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	RE
0	Owner	0	0	2	BHK	1300.236407	
1	Dealer	0	0	2	BHK	1275.000000	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
4	Dealer	1	0	2	BHK	999.009247	
...	...	...	...	...	...	...	
29446	Owner	0	0	3	BHK	2500.000000	
29447	Owner	0	0	2	BHK	769.230769	
29448	Dealer	0	0	2	BHK	1022.641509	
29449	Owner	0	0	2	BHK	927.079009	
29450	Dealer	0	1	2	BHK	896.774194	

29451 rows × 14 columns

In [10]:

```
del dataset['TARGET(PRICE_IN_LACS)']
```

In [11]:

```
dataset
```

Out[11]:

	POSTED_BY	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQUARE_FT	RE
0	Owner	0	0	2	BHK	1300.236407	
1	Dealer	0	0	2	BHK	1275.000000	
2	Owner	0	0	2	BHK	933.159722	
3	Owner	0	1	2	BHK	929.921143	
4	Dealer	1	0	2	BHK	999.009247	
...	...	...	...	...	...	...	
29446	Owner	0	0	3	BHK	2500.000000	
29447	Owner	0	0	2	BHK	769.230769	
29448	Dealer	0	0	2	BHK	1022.641509	
29449	Owner	0	0	2	BHK	927.079009	
29450	Dealer	0	1	2	BHK	896.774194	

29451 rows × 13 columns

In [13]:

```
dataset.insert(1,"New_Update",ab1)
```

In [14]:

```
dataset
```

Out[14]:

	POSTED_BY	New_Update	UNDER_CONSTRUCTION	RERA	BHK_NO.	BHK_OR_RK	SQ
0	Owner	65.0	0	0	2	BHK	13
1	Dealer	61.0	0	0	2	BHK	12
2	Owner	53.0	0	0	2	BHK	9
3	Owner	72.5	0	1	2	BHK	9
4	Dealer	70.5	1	0	2	BHK	9
...	...	...	...	...	...	...	
29446	Owner	55.0	0	0	3	BHK	25
29447	Owner	26.0	0	0	2	BHK	7
29448	Dealer	37.1	0	0	2	BHK	10
29449	Owner	77.0	0	0	2	BHK	9
29450	Dealer	37.8	0	1	2	BHK	8

29451 rows × 14 columns

In [15]:

```
dataset.isnull().any()
```

Out[15]:

```
POSTED_BY             False
New_Update            False
UNDER_CONSTRUCTION    False
RERA                  False
BHK_NO.               False
BHK_OR_RK             False
SQUARE_FT             False
READY_TO_MOVE         False
RESALE                False
City                   True
LONGITUDE             False
LATITUDE              False
Converted             False
Updated_Price         False
dtype: bool
```

In [16]:

```
dataset.isnull().sum()
```

Out[16]:

```
POSTED_BY             0
New_Update            0
UNDER_CONSTRUCTION    0
RERA                  0
BHK_NO.               0
BHK_OR_RK             0
SQUARE_FT             0
READY_TO_MOVE         0
RESALE                0
City                  9
LONGITUDE             0
LATITUDE              0
Converted             0
Updated_Price         0
dtype: int64
```

In [17]:

```
data=pd.read_csv(r"D:\My Work\Top_Dataset\weather.csv")
```

In [18]:

```
data
```

Out[18]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	Windl
0	8.0	24.3	0.0	3.4	6.3	NW	30.0	
1	14.0	26.9	3.6	4.4	9.7	ENE	39.0	
2	13.7	23.4	3.6	5.8	3.3	NW	85.0	
3	13.3	15.5	39.8	7.2	9.1	NW	54.0	
4	7.6	16.1	2.8	5.6	10.6	SSE	50.0	
...	...	...	...	...	...	...	...	
361	9.0	30.7	0.0	7.6	12.1	NNW	76.0	
362	7.1	28.4	0.0	11.6	12.7	N	48.0	
363	12.5	19.9	0.0	8.4	5.3	ESE	43.0	
364	12.5	26.9	0.0	5.0	7.1	NW	46.0	
365	12.3	30.2	0.0	6.0	12.6	NW	78.0	

366 rows × 22 columns

In [20]:

```
data.isnull().sum()
```

Out[20]:

```
MinTemp           0
MaxTemp           0
Rainfall          0
Evaporation       0
Sunshine          3
WindGustDir       3
WindGustSpeed     2
WindDir9am       31
WindDir3pm        1
WindSpeed9am      7
WindSpeed3pm      0
Humidity9am       0
Humidity3pm       0
Pressure9am       0
Pressure3pm       0
Cloud9am          0
Cloud3pm          0
Temp9am           0
Temp3pm           0
RainToday         0
RISK_MM           0
RainTomorrow      0
dtype: int64
```

In [21]:

```
data['WindDir9am'].fillna(111,inplace=True)
```

In [22]:

```
data
```

Out[22]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	Windl
0	8.0	24.3	0.0	3.4	6.3	NW	30.0	
1	14.0	26.9	3.6	4.4	9.7	ENE	39.0	
2	13.7	23.4	3.6	5.8	3.3	NW	85.0	
3	13.3	15.5	39.8	7.2	9.1	NW	54.0	
4	7.6	16.1	2.8	5.6	10.6	SSE	50.0	
...	...	...	...	...	...	...	...	...
361	9.0	30.7	0.0	7.6	12.1	NNW	76.0	
362	7.1	28.4	0.0	11.6	12.7	N	48.0	
363	12.5	19.9	0.0	8.4	5.3	ESE	43.0	
364	12.5	26.9	0.0	5.0	7.1	NW	46.0	
365	12.3	30.2	0.0	6.0	12.6	NW	78.0	

366 rows × 22 columns

In [23]:

```
data.isnull().sum()
```

Out[23]:

```
MinTemp          0
MaxTemp          0
Rainfall         0
Evaporation      0
Sunshine         3
WindGustDir      3
WindGustSpeed    2
WindDir9am       0
WindDir3pm       1
WindSpeed9am     7
WindSpeed3pm     0
Humidity9am      0
Humidity3pm      0
Pressure9am      0
Pressure3pm      0
Cloud9am         0
Cloud3pm         0
Temp9am          0
Temp3pm          0
RainToday        0
RISK_MM          0
RainTomorrow     0
dtype: int64
```

In [24]:

```
data['WindDir9am'].unique()
```

Out[24]:

```
array(['SW', 'E', 'N', 'WNW', 'SSE', 'SE', 'S', 'WSW', 'NNE', 'NNW',
       'ENE', 'SSW', 'NW', 'ESE', 'NE', 'W', 111], dtype=object)
```

In [ ]:

```

```