

Business Problem - Predict the Price of Bangalore House

Using Decision Tree Regression - Supervised Machine Learning Algorithm

Load Libraries

In [1]:

```
import pandas as pd
```

Load Data

In [2]:

```
df = pd.read_csv(r"D:\machine_learning\ML_by_Avanish\bg_house_price_data.csv")
```

In [3]:

```
df
```

Out[3]:

	bath	balcony	price	total_sqft_int	bhk	price_per_sqft	area_typeSuper built-up Area	area_typeBuilt- up Area
0	3.0	2.0	150.00	1672.0	3	8971.291866	1	0
1	3.0	3.0	149.00	1750.0	3	8514.285714	0	1
2	3.0	2.0	150.00	1750.0	3	8571.428571	1	0
3	2.0	2.0	40.00	1250.0	2	3200.000000	1	0
4	2.0	2.0	83.00	1200.0	2	6916.666667	0	0
...
7115	3.0	2.0	325.00	2900.0	3	11206.896552	1	0
7116	3.0	1.0	84.83	1780.0	3	4765.730337	1	0
7117	2.0	1.0	48.00	880.0	2	5454.545455	0	0
7118	2.0	1.0	55.00	1000.0	2	5500.000000	0	0
7119	2.0	1.0	78.00	1400.0	3	5571.428571	0	0

7120 rows × 108 columns

In [4]:

```
df.head()
```

Out[4]:

	bath	balcony	price	total_sqft_int	bhk	price_per_sqft	area_typeSuper built-up Area	area_typeBuilt- up Area	are
0	3.0	2.0	150.0	1672.0	3	8971.291866	1	0	
1	3.0	3.0	149.0	1750.0	3	8514.285714	0	1	
2	3.0	2.0	150.0	1750.0	3	8571.428571	1	0	
3	2.0	2.0	40.0	1250.0	2	3200.000000	1	0	
4	2.0	2.0	83.0	1200.0	2	6916.666667	0	0	

5 rows × 108 columns

Split Data

In [5]:

```
X = df.drop('price', axis=1)
y = df['price']

print('Shape of X = ', X.shape)
print('Shape of y = ', y.shape)
```

Shape of X = (7120, 107)

Shape of y = (7120,)

In [6]:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=51)

print('Shape of X_train = ', X_train.shape)
print('Shape of y_train = ', y_train.shape)
print('Shape of X_test = ', X_test.shape)
print('Shape of y_test = ', y_test.shape)
```

Shape of X_train = (5696, 107)

Shape of y_train = (5696,)

Shape of X_test = (1424, 107)

Shape of y_test = (1424,)

#Decision Tree Regression - ML Model Training

In [8]:

```
from sklearn.tree import DecisionTreeRegressor
```

In [9]:

```
regressor = DecisionTreeRegressor(criterion='mse')  
regressor.fit(X_train, y_train)
```

Out[9]:

DecisionTreeRegressor()

In [10]:

```
regressor.score(X_test, y_test)
```

Out[10]:

0.890945321471055

Predict the value of Home

In [11]:

```
X_test.iloc[-1, :]
```

Out[11]:

bath	2.000000
balcony	0.000000
total_sqft_int	1566.000000
bhk	2.000000
price_per_sqft	11494.252874
...	
location_Hosur Road	0.000000
location_Horamavu Banaswadi	0.000000
location_Domlur	0.000000
location_Mahadevpura	0.000000
location_Tumkur Road	0.000000

Name: 43, Length: 107, dtype: float64

In [12]:

```
regressor.predict([X_test.iloc[-1, :]])
```

Out[12]:

array([171.])

In [13]:

```
y_test.iloc[-1]
```

Out[13]:

180.0

In [14]:

```
pred = regressor.predict(X_test)
pred
```

Out[14]:

```
array([ 83.   ,  39.95, 120.   , ...,  32.64,  62.5 , 171.   ])
```

In [15]:

```
y_test
```

Out[15]:

```
2435      80.00
3113      40.00
426      120.00
1124      79.00
1161      45.00
...
2078      28.34
6855      84.00
4381      32.00
3862      63.00
43      180.00
Name: price, Length: 1424, dtype: float64
```

In []: