1. Using Linear Regression predicting the price for a 6000sqft building

In [1]:

```
import pandas as pd
```

In [2]:

```
dataset = pd.read_excel("dataset/Linear Regression.xlsx",sheet_name="Linear Regression")
```

In [3]:

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21613 entries, 0 to 21612
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   price        21613 non-null  int64  
 1   sqft_living  21613 non-null  int64  
 2   bedrooms     21613 non-null  int64  
 3   bathrooms    21613 non-null  float64
 4   floors       21613 non-null  float64
dtypes: float64(2), int64(3)
memory usage: 844.4 KB
```

So the dataset is good and no null cells are present.

In [4]:

```
dataset.head()
```

Out[4]:

	price	sqft_living	bedrooms	bathrooms	floors
0	221900	1180	3	1.00	1.0
1	538000	2570	3	2.25	2.0
2	180000	770	2	1.00	1.0
3	604000	1960	4	3.00	1.0
4	510000	1680	3	2.00	1.0

In [5]:

```
import statsmodels.api as sm
```

==

1) First finding out the effect of sqft on price

```
    here price is the dependent variable (y) and the sqft is the independent variable
  (x)
```

In [6]:

```
y = dataset.price
x = dataset.sqft_living
```

In [7]:

```
x1 = sm.add_constant(x)

simple_regress = sm.OLS(y,x1)

result = simple_regress.fit()

result.summary()
```

Out[7]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.493
Model:	OLS	**Adj. R-squared:**	0.493
Method:	Least Squares	**F-statistic:**	2.100e+04
Date:	Mon, 10 Aug 2020	**Prob (F-statistic):**	0.00
Time:	01:19:58	**Log-Likelihood:**	-3.0027e+05
No. Observations:	21613	**AIC:**	6.005e+05
Df Residuals:	21611	**BIC:**	6.006e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-4.358e+04	4402.690	-9.899	0.000	-5.22e+04	-3.5e+04
sqft_living	280.6236	1.936	144.920	0.000	276.828	284.419

Omnibus:	14832.490	**Durbin-Watson:**	1.983
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	546444.709
Skew:	2.824	**Prob(JB):**	0.00
Kurtosis:	26.977	**Cond. No.**	5.63e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.63e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

B0 = -4.358e+04

B1 = 280.6236

x = 6000

Simple Linear Regression :

```
    y = B0 + B1x
```

In [8]:

```
B0 = -4.358e+04
B1 = 280.6236
x  = 6000
y = B0 + (B1*x)
print(y)
```

```
1640161.6
```

So the predicted price for 6000sqft building would be 16,40,161.6

==

2) Now finding out the effect of number of bedrooms over price

```
    here price is the dependent variable (y) and the no. of bedrooms is the independent
  variable (x)

    lets find out the price for 5 bedroom apartment
```

In [9]:

```
dataset.head(1)
```

Out[9]:

	price	sqft_living	bedrooms	bathrooms	floors
0	221900	1180	3	1.0	1.0

In [10]:

```
y = dataset.price
x = dataset.bedrooms

x1 = sm.add_constant(x)
simple_regress = sm.OLS(y,x1)
result = simple_regress.fit()
result.summary()
```

Out[10]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.095
Model:	OLS	**Adj. R-squared:**	0.095
Method:	Least Squares	**F-statistic:**	2271.
Date:	Mon, 10 Aug 2020	**Prob (F-statistic):**	0.00
Time:	01:19:58	**Log-Likelihood:**	-3.0652e+05
No. Observations:	21613	**AIC:**	6.131e+05
Df Residuals:	21611	**BIC:**	6.131e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.298e+05	8931.866	14.533	0.000	1.12e+05	1.47e+05
bedrooms	1.217e+05	2554.304	47.651	0.000	1.17e+05	1.27e+05

Omnibus:	18859.406	**Durbin-Watson:**	1.961
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1199044.953
Skew:	3.904	**Prob(JB):**	0.00
Kurtosis:	38.644	**Cond. No.**	14.2

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

B0 = 1.298e+05

B1 = 1.217e+05

x = 5

Simple Linear Regression :

y = B0 + B1x

In [11]:

```
B0 = 1.298e+05
B1 = 1.217e+05
x = 5
y = B0 + (B1*x)
print(y)
```

```
738300.0
```

So the predicted price for 5 bedroom apartment would be 7,38,300.0

==

3) Now finding out the effect of number of bathrooms over price

```
    here price is the dependent variable (y) and the no. of bathrooms is the independen
  t variable (x)

    lets find out the price for 3 bathroom apartment
```

In [12]:

```
dataset.head(1)
```

Out[12]:

	price	sqft_living	bedrooms	bathrooms	floors
0	221900	1180	3	1.0	1.0

In [13]:

```
y = dataset.price
x = dataset.bathrooms

x1 = sm.add_constant(x)
simple_regress = sm.OLS(y,x1)
result = simple_regress.fit()
result.summary()
```

Out[13]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.276
Model:	OLS	**Adj. R-squared:**	0.276
Method:	Least Squares	**F-statistic:**	8229.
Date:	Mon, 10 Aug 2020	**Prob (F-statistic):**	0.00
Time:	01:19:59	**Log-Likelihood:**	-3.0412e+05
No. Observations:	21613	**AIC:**	6.082e+05
Df Residuals:	21611	**BIC:**	6.083e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.071e+04	6210.669	1.724	0.085	-1465.059	2.29e+04
bathrooms	2.503e+05	2759.528	90.714	0.000	2.45e+05	2.56e+05

Omnibus:	17286.361	**Durbin-Watson:**	1.960
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	887102.348
Skew:	3.458	**Prob(JB):**	0.00
Kurtosis:	33.614	**Cond. No.**	7.75

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

B0 = 1.071e+04

B1 = 2.503e+05

x = 3

Simple Linear Regression :

y = B0 + B1x

In [14]:

```
B0 = 1.071e+04
B1 = 2.503e+05
x = 3
y = B0 + (B1*x)
print(y)
```

```
761610.0
```

So the predicted price for 3 bathroom apartment would be 7,61,610.0

==

4) Now finding out the effect of number of floors over price

```
    here price is the dependent variable (y) and the no. of floors is the independent v
  ariable (x)

    lets find out the price for 5 floor building
```

In [15]:

```
dataset.head(1)
```

Out[15]:

	price	sqft_living	bedrooms	bathrooms	floors
0	221900	1180	3	1.0	1.0

In [16]:

```
y = dataset.price
x = dataset.floors

x1 = sm.add_constant(x)
simple_regress = sm.OLS(y,x1)
result = simple_regress.fit()
result.summary()
```

Out[16]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.066
Model:	OLS	**Adj. R-squared:**	0.066
Method:	Least Squares	**F-statistic:**	1526.
Date:	Mon, 10 Aug 2020	**Prob (F-statistic):**	1.58e-322
Time:	01:19:59	**Log-Likelihood:**	-3.0687e+05
No. Observations:	21613	**AIC:**	6.137e+05
Df Residuals:	21611	**BIC:**	6.138e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.792e+05	7101.851	39.313	0.000	2.65e+05	2.93e+05
floors	1.746e+05	4469.727	39.060	0.000	1.66e+05	1.83e+05

Omnibus:	19377.136	**Durbin-Watson:**	1.973
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1260093.162
Skew:	4.077	**Prob(JB):**	0.00
Kurtosis:	39.507	**Cond. No.**	6.37

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

B0 = 2.792e+05

B1 = 1.746e+05

x = 5

Simple Linear Regression :

y = B0 + B1x

In [17]:

```
B0 = 2.792e+05
B1 = 1.746e+05
x = 5
y = B0 + (B1*x)
print(y)
```

```
1152200.0
```

So the predicted price for 5 floor building would be 11,52,200.0

===

5) Now using the multiple regression method finding out the price of a 6000sqft building with 10 bedrooms, 10 bathrooms, 5 floors

In [18]:

```
dataset.head(1)
```

Out[18]:

	price	sqft_living	bedrooms	bathrooms	floors
0	221900	1180	3	1.0	1.0

In [19]:

```
y = dataset.price
x = dataset[['sqft_living','bedrooms','bathrooms','floors']]

x1 = sm.add_constant(x)
simple_regress = sm.OLS(y,x1)
result = simple_regress.fit()
result.summary()
```

Out[19]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.507
Model:	OLS	**Adj. R-squared:**	0.507
Method:	Least Squares	**F-statistic:**	5554.
Date:	Mon, 10 Aug 2020	**Prob (F-statistic):**	0.00
Time:	01:19:59	**Log-Likelihood:**	-2.9996e+05
No. Observations:	21613	**AIC:**	5.999e+05
Df Residuals:	21608	**BIC:**	6.000e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.467e+04	7679.122	9.724	0.000	5.96e+04	8.97e+04
sqft_living	309.3932	3.087	100.228	0.000	303.343	315.444
bedrooms	-5.785e+04	2347.323	-24.644	0.000	-6.24e+04	-5.32e+04
bathrooms	7853.5235	3814.223	2.059	0.040	377.365	1.53e+04
floors	200.4943	3775.505	0.053	0.958	-7199.774	7600.763

Omnibus:	14450.413	**Durbin-Watson:**	1.985
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	494760.938
Skew:	2.739	**Prob(JB):**	0.00
Kurtosis:	25.790	**Cond. No.**	1.04e+04

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.04e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

B0 = 7.467e+04

B1 = 309.3932 -------- x1 = 6000

B2 = -5.785e+04 ----- x2 = 10

B3 = 7853.5235 ------ x3 = 10

B4 = 200.4943 -------- x4 = 5

Simple Linear Regression :

y = B0 + B1x1 + B2x2 + B3x3 + B4x4

In [20]:

```
B0 = 7.467e+04
B1 = 309.3932
x1 = 6000

B2 = -5.785e+04
x2 = 10

B3 = 7853.5235
x3 = 10

B4 = 200.4943
x4 = 5

y = B0 + (B1*x1) + (B2*x2) + (B3*x3) + (B4*x4)
print(y)
```

```
1432066.9065
```

So the predicted price of a 6000sqft building with 10 bedrooms, 10 bathrooms, 5 floors is 14,32,066.9065