Predicting Home Prices Using Simple Linear Regression

```
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         from sklearn import linear_model
```

```
In [40]: Main_file = pd.read_csv(r"D:\NEEL_FOLDER\Data Science\Linear_Regression_CodeBa
         sic\homeprices.csv")
         Main_file
```
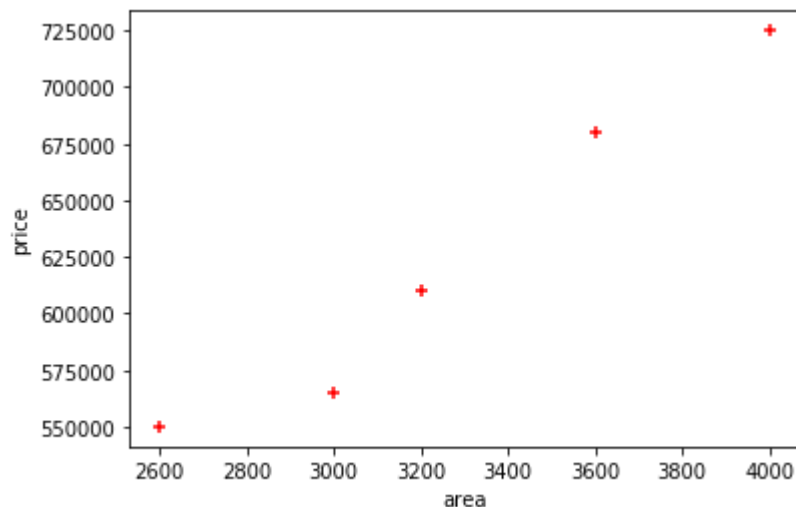
Out[40]:

	area	price
0	2600	550000
1	3000	565000
2	3200	610000
3	3600	680000
4	4000	725000

Lets plot scatterplot to see the distribution of the dataset.

```
In [9]:  %matplotlib inline
         plt.xlabel('area')
         plt.ylabel('price')
         plt.scatter(Main_file.area,Main_file.price,color='red',marker='+')
```

Out[9]: <matplotlib.collections.PathCollection at 0xf05cbf0>

Divide the dataset into two different dataframe i.e area & price

```
In [13]: area = Main_file.drop('price',axis='columns')
         area
```

Out[13]:

	area
0	2600
1	3000
2	3200
3	3600
4	4000

```
In [15]: price = Main_file.price
         price
```

```
Out[15]: 0    550000
         1    565000
         2    610000
         3    680000
         4    725000
         Name: price, dtype: int64
```

Create a Linear regression object(#1) and then fit the data(#2) (fitting the data means you are training the linear regression model using the available data points)

```
In [17]: #1
         reg = linear_model.LinearRegression()
         #2
         reg.fit(area,price)
```

```
Out[17]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=Fals
         e)
```

```
In [39]: reg.predict([[3300]])
```

```
Out[39]: array([628715.75342466])
```

```
In [25]: print('Value of m', reg.coef_)
         print('Value of b', reg.intercept_)
         print('y= mx+b is',135.78767123*3300+180616.43835616432)
```

```
         Value of m [135.78767123]
         Value of b 180616.43835616432
         y= mx+b is 628715.7534151643
```

Now we have a file('area') with a list of area and we want to predict the prices for those respective areas.

```
In [27]: New_file = pd.read_csv(r"D:\NEEL_FOLDER\Data Science\Linear_Regression_CodeBas
         ic\area.csv")
         New_file.head(5)
```

Out[27]:

	area
0	1000
1	1500
2	2300
3	3540
4	4120

```
In [28]: reg.predict(New_file)
```

```
Out[28]: array([ 316404.10958904,  384297.94520548,  492928.08219178,
                 661304.79452055,  740061.64383562,  799808.21917808,
                 926090.75342466,  650441.78082192,  825607.87671233,
                 492928.08219178, 1402705.47945205, 1348390.4109589 ,
                1144708.90410959])
```

Create a new column in data frame.

```
In [ ]: p = reg.predict(New_file)
```

```
In [ ]: New_file['prices'] = p
```

In [35]: New_file

Out[35]:

	area	prices
0	1000	3.164041e+05
1	1500	3.842979e+05
2	2300	4.929281e+05
3	3540	6.613048e+05
4	4120	7.400616e+05
5	4560	7.998082e+05
6	5490	9.260908e+05
7	3460	6.504418e+05
8	4750	8.256079e+05
9	2300	4.929281e+05
10	9000	1.402705e+06
11	8600	1.348390e+06
12	7100	1.144709e+06

Create a new file ' prediction.csv '.

In [37]: New_file.to_csv('prediction.csv',index=False)

(Source:Codebasic youtube channel)