Predicting Home Prices With Multiple Variable

In [32]:
```
import pandas as pd
import numpy as np
from sklearn import linear_model
```

In [33]:
```
#hp = house price
hp = pd.read_csv(r'D:\NEEL_FOLDER\Data Science\Linear_Regression_CodeBasic\homeprices_Multi.csv')
hp
```

Out[33]:

	area	bedrooms	age	price
0	2600	3.0	20	550000
1	3000	4.0	15	565000
2	3200	NaN	18	610000
3	3600	3.0	30	595000
4	4000	5.0	8	760000
5	4100	6.0	8	810000

Their is a null value in bedrooms column. Fill this null value with median value.

In [34]:
```
# just for curiosity
median = hp['bedrooms'].median()
mean = hp['bedrooms'].mean()
print('median =', a)
print('mean = ', b)
```

```
median = 4.0
mean =  4.2
```

In [37]:
```
hp.bedrooms = hp.bedrooms.fillna(hp.bedrooms.median())
hp
```

Out[37]:

	area	bedrooms	age	price
0	2600	3.0	20	550000
1	3000	4.0	15	565000
2	3200	4.0	18	610000
3	3600	3.0	30	595000
4	4000	5.0	8	760000
5	4100	6.0	8	810000

As the data is cleaned, lets create the model.

In [39]:
```
reg = linear_model.LinearRegression()
reg.fit(hp[['area','bedrooms','age']],hp.price)
# area,bedrooms and age are independent variable and price is dependent variable.
```

Out[39]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [40]:
```
reg.coef_
```

Out[40]:
```
array([  112.06244194, 23388.88007794, -3231.71790863])
```

In [41]:
```
reg.intercept_
```

Out[41]:
```
221323.0018654043
```

Predict:

1.Price of home with 3000 sqr ft area, 3 bedrooms, 40 year old

2.Price of home with 2500 sqr ft area, 4 bedrooms, 5 year old

In [43]:
```
reg.predict([[3000,3,40]])
```

Out[43]:
```
array([498408.25158031])
```

In [44]:
```
reg.predict([[2500,4,5]])
```

Out[44]:
```
array([578876.03748933])
```

Source:Codebasic youtube channel