In [23]:
```
# Importing all the necessary libraries
import pandas as pd
import numpy as np
from sklearn import linear_model
```

In [22]:
```
# Reading the file with Pandas
df = pd.read_csv(r'C:\Users\Desktop\areas_multiplevariables.csv')
```

In [24]:
```
# Look at the file which is read
df
```

Out[24]:

	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

In [26]:
```
#Data Preprocessing: Fill NA values with median value of a column

df.bedrooms.median()

df.bedrooms = df.bedrooms.fillna(df.bedrooms.median())
df
```

Out[26]:

	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

In [27]:
```
# Create linear regression object.
reg = linear_model.LinearRegression()
# Fit the data.in other words , training the model based on the available model.
#  1st argument is has to be a 2D array.
reg.fit(df.drop('price',axis='columns'),df.price)
```

Out[27]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

In [30]:
```
# Find price of home with 3000 sqr ft area, 3 bedrooms, 40 year old
reg.predict([[3000, 3, 40]])
```

Out[30]:
```
array([498408.25158031])
```

In [31]:
```
# Find price of home with 2500 sqr ft area, 4 bedrooms, 5 year old
reg.predict([[2500, 4, 5]])
```

Out[31]:
```
array([578876.03748933])
```

In [28]:
```
# coefficient of the linear regression line
reg.coef_
```

Out[28]:
```
array([  112.06244194, 23388.88007794, -3231.71790863])
```

In [32]:
```
# intercept of the linear regression line
reg.intercept_
```

Out[32]:
```
221323.00186540408
```