

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

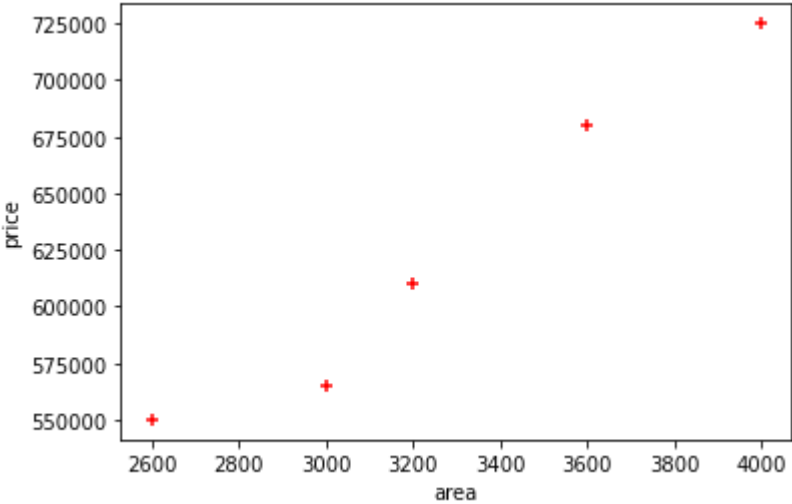
```
In [2]: df = pd.read_csv('homeprices.csv')
df
```

Out[2]:

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

```
In [3]: %matplotlib inline
plt.xlabel('area')
plt.ylabel('price')
plt.scatter(df.area,df.price,color='red',marker='+')
```

Out[3]: <matplotlib.collections.PathCollection at 0x23a36b054f0>



```
In [4]: new_df = df.drop('price',axis='columns')
new_df
```

Out[4]:

	area
0	2600
1	3000
2	3200
3	3600
4	4000

```
In [5]: price = df.price
price
```

Out[5]:

0	550000
1	565000
2	610000
3	680000
4	725000

Name: price, dtype: int64

```
In [6]: reg = linear_model.LinearRegression()
reg.fit(new_df,price)
```

Out[6]: LinearRegression()

```
In [7]: reg.predict([[3300]])
```

Out[7]: array([628715.75342466])

```
In [10]: reg.predict([[5000]])
```

Out[10]: array([859554.79452055])

```
In [11]: area_df = pd.read_csv("areas.csv")
area_df.head(3)
```

Out[11]:

	area
0	1000
1	1500
2	2300

```
In [14]: p = reg.predict(area_df)
p
```

Out[14]:

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])		

```
In [15]: area_df['prices']=p
area_df
```

Out[15]:

	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