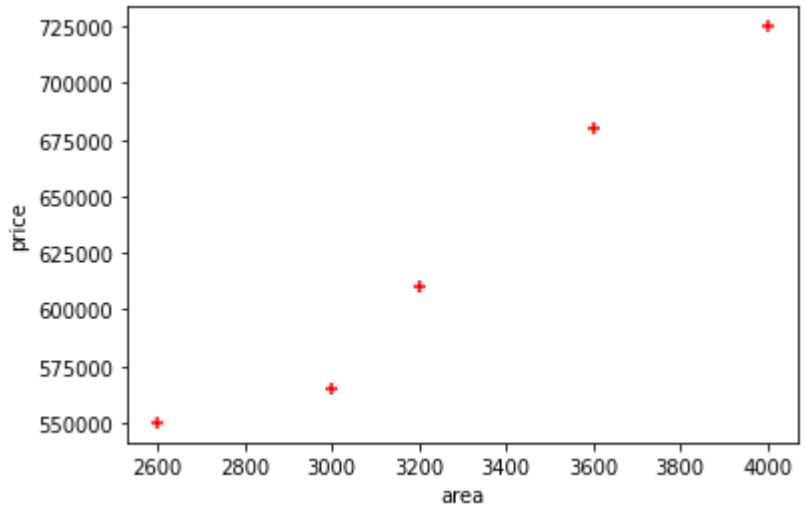
```
import pandas as pd
import numpy as np
from sklearn import linear_model
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/homeprices.csv')
df
```

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

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

```
<matplotlib.collections.PathCollection at 0x7f4ad5028250>
```




```
new_df = df.drop('price',axis='columns')
new_df
```

	area
0	2600
1	3000
2	3200

```
price = df.price
price
```

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

```
# Create linear regression object
reg = linear_model.LinearRegression()
reg.fit(new_df,price)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
reg.predict([[3300]])
```

```
array([628715.75342466])
```

```
reg.coef_
```

```
array([135.78767123])
```

```
reg.intercept_
```

```
180616.43835616432
```

```
3300*135.78767123 + 180616.43835616432
```

```
628715.7534151643
```

```
reg.predict([[5000]])
```

```
array([859554.79452055])
```

```
area_df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/areas.csv")
area_df.head(3)
```

	area
0	1000
1	1500
2	2300

```
p = reg.predict(area_df)
p
```

```
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])
```

```
area_df['prices']=p
area_df
```

	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

```
area_df.to_csv("prediction.csv")
```

✓ 0s completed at 13:32