

homeprices.csv

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

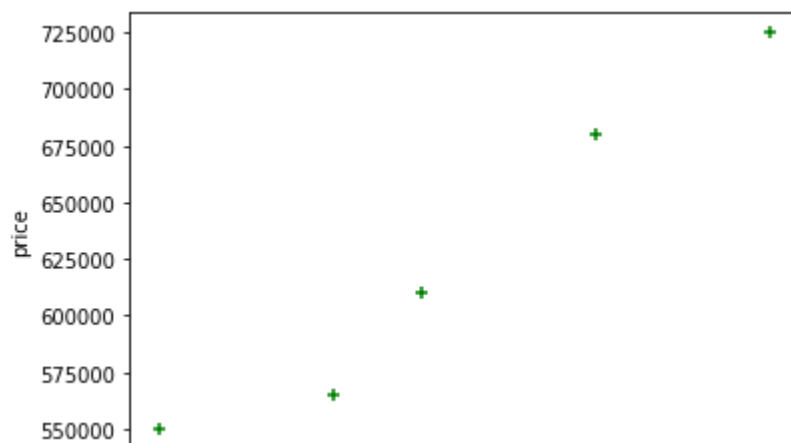
```
import numpy as np
import pandas as pd
from sklearn import linear_model
import matplotlib.pyplot as plt
print('homeprices.csv')
df = pd.read_csv('homeprices.csv')
df
```

homeprices.csv

	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='green',marker='+')
```

<matplotlib.collections.PathCollection at 0x7f3828d49690>





```
0 2600
1 3000
2 3200
3 3600
4 4000
```

```
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=Fa
```

```
# predicting price against 3300 as input area
reg.predict([[3300]])
```

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

```
# calculating regression coefficient
reg.coef_
```

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

```
# calculating intercept
reg.intercept_
```

```
180616.43835616432
```

```
# predicting prices against all areas
p = reg.predict(new_df)
p
```

4	4120
5	4560
6	5490
7	3460
8	4750
9	2300
10	9000
11	8600
12	7100

areas.csv

Double-click (or enter) to edit

```
# reading unknown areas for prediction
area_df = pd.read_csv('areas.csv')
area_df
```

	area
0	1000
1	1500
2	2300
3	3540
4	4120
5	4560
6	5490

```
area_df['price']=m
```

```
# saving prediction into new file  
area_df.to_csv("prediction.csv")
```