In [1]:

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

In [2]:

```
df = pd.read_csv("home.csv (1).csv")
```

In [3]:

```
df
```

Out[3]:

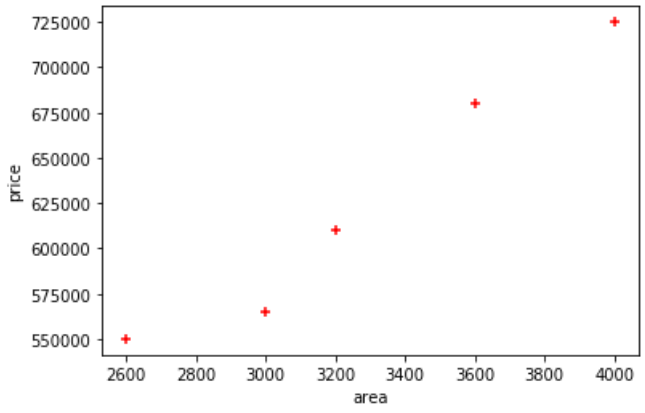
| | area | price |
|---|---|---|
| 0 | 2600 | 550000 |
| 1 | 3000 | 565000 |
| 2 | 3200 | 610000 |
| 3 | 3600 | 680000 |
| 4 | 4000 | 725000 |

In [4]:

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

Out[4]:

```
<matplotlib.collections.PathCollection at 0x48ef270>
```




In [5]:

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

Out[5]:

| | area |
|---|---|
| 0 | 2600 |
| 1 | 3000 |
| 2 | 3200 |

| | |
|---|---|
| 3 | 3600 |
| 4 | 4000 |

In [6]:

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

Out[6]:

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

In [7]:

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

Out[7]:

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

In [8]:

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

Out[8]:

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

In [9]:

```
reg.coef_
```

Out[9]:

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

In [10]:

```
reg.intercept_
```

Out[10]:

```
180616.43835616432
```

In [12]:

```
3300*135.78767123 + 180616.43835616432
```

Out[12]:

```
628715.7534151643
```

In [13]:

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

Out[13]:

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

In [16]:

```
area_df = pd.read_csv("areas.csv")
area_df.head(5)
```

Out[16]:

| | area |
|---|---|
| **0** | 1000 |
| **1** | 1500 |
| **2** | 2300 |
| **3** | 3540 |
| **4** | 4120 |

In [17]:

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

Out[17]:

```
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 [18]:

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

Out[18]:

| | 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 |

In [ ]:

```

```