In [1]:
```
# Experiment-6: Implement linear regression using python
```

In [2]:
```
import pandas as pd
```

In [3]:
```
df=pd.read_csv("F:\Lecture-1\ML Notes BTech III Yr\CSE-B ML-LAB\datasets\homeprices.csv")
df
```

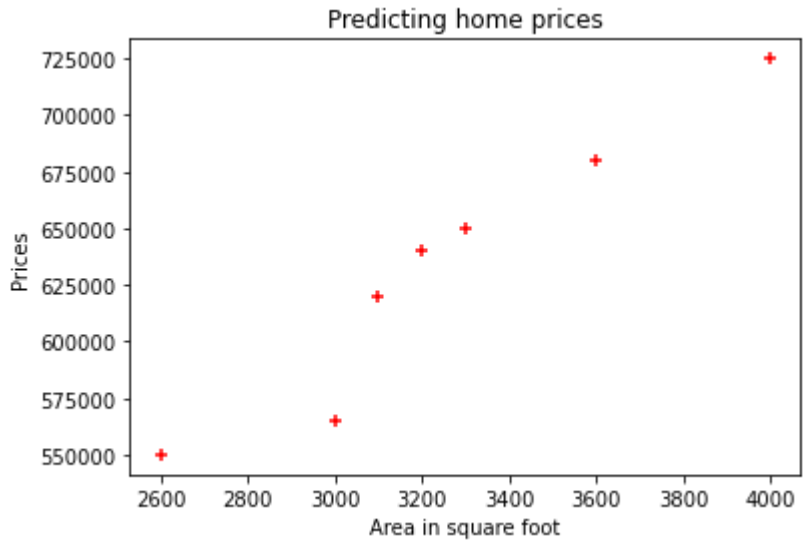
Out[3]:

	area	price
0	2600	550000
1	3000	565000
2	3100	620000
3	3200	640000
4	3300	650000
5	3600	680000
6	4000	725000

In [4]:
```
import matplotlib.pyplot as plt
```

In [5]:
```
plt.title("Predicting home prices")
plt.xlabel("Area in square foot")
plt.ylabel("Prices")
plt.scatter(df.area,df.price,color='red',marker='+')
```

Out[5]:
```
<matplotlib.collections.PathCollection at 0x27d0414da60>
```



In [6]:
```
x=df[['area']]
x
```

Out[6]:

	area
0	2600
1	3000
2	3100
3	3200
4	3300
5	3600
6	4000

In [7]:
```
y=df[['price']]
y
```

Out[7]:

	price
0	550000
1	565000
2	620000
3	640000
4	650000
5	680000
6	725000

In [8]:
```
from sklearn.model_selection import train_test_split
```

In [9]:
```
X_train, X_test, y_train, y_test=train_test_split(x,y,test_size=0.2)
```

In [10]:
```
len(X_train)
```

Out[10]: 5

In [11]:
```
len(X_test)
```

Out[11]: 2

In [12]:
```
X_train
```

Out[12]:

	area
5	3600
0	2600
2	3100
1	3000
3	3200

In [13]:
```
X_test
```

Out[13]:

	area
6	4000
4	3300

In [14]:
```
from sklearn.linear_model import LinearRegression
```

In [15]:
```
model=LinearRegression()
```

In [16]:
```
model.fit(X_train,y_train)
```

Out[16]: LinearRegression()

In [17]:
```
model.predict(X_test)
```

Out[17]:
```
array([[736480.76923077],
       [638884.61538462]])
```

In [18]:
```
model.score(X_test,y_test)
```

Out[18]:
```
0.9092053911900063
```

In [19]:
```
plt.title("Predicting home prices")
plt.xlabel("Area in square foot")
plt.ylabel("Prices")
plt.scatter(df.area,df.price,color='red',marker='+')
plt.plot(df.area,model.predict(df[['area']]),color='blue')
```

Out[19]:
```
[<matplotlib.lines.Line2D at 0x27d06c582e0>]
```