In [1]:
```
import pandas as pd
```

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

In [3]:
```
df
```

Out[3]:

	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1
5	56	1
6	55	0
7	60	1
8	62	1
9	61	1
10	18	0
11	28	0
12	27	0
13	29	0
14	49	1
15	55	1
16	25	1
17	58	1
18	19	0
19	18	0
20	21	0
21	26	0
22	40	1
23	45	1
24	50	1

	age	bought_insurance
25	54	1
26	23	0

In [4]:
```
df.head()
```

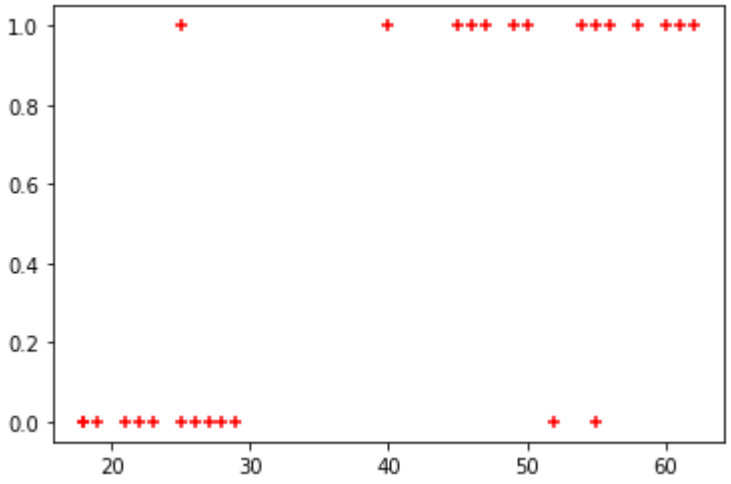
Out[4]:

	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

In [5]:
```
import matplotlib.pyplot as plt
%matplotlib inline
```

In [6]:
```
plt.scatter(df.age,df.bought_insurance,marker='+',c='r')
```

Out[6]:
```
<matplotlib.collections.PathCollection at 0x28c69a29580>
```



In [7]:
```
from sklearn.model_selection import train_test_split
```

In [ ]:
```

```

In [8]:
```
x_train, x_test, y_train, y_test=train_test_split(df[['age']],df.bought_insurance,test_size=0.2)
```

In [9]:
```
x_train.head()
```

Out[9]:

	age
24	50
8	62
9	61
1	25
12	27

In [10]:
```
from sklearn.linear_model import LogisticRegression
```

In [11]:
```
model=LogisticRegression()
```

In [12]:
```
model.fit(x_train,y_train)
```

Out[12]:
```
LogisticRegression()
```

In [13]:
```
model.predict(x_test)
```

Out[13]:
```
array([0, 0, 0, 0, 1, 1], dtype=int64)
```

In [14]:
```
y_test.values
```

Out[14]:
```
array([0, 0, 0, 0, 1, 1], dtype=int64)
```

In [15]:
```
model.score(x_test,y_test)
```

Out[15]:
```
1.0
```

In [16]:
```
model.predict_proba(x_test)
```

Out[16]:
```
array([[0.87371338, 0.12628662],
       [0.75772149, 0.24227851],
       [0.90674058, 0.09325942],
       [0.73629574, 0.26370426],
       [0.12761321, 0.87238679],
       [0.31260134, 0.68739866]])
```

In [19]:
```
model.predict([[40]])
```

Out[19]:
```
array([1], dtype=int64)
```

In [ ]:
```

```