In [1]:
```python
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

In [2]:
```python
df = pd.read_csv("insurance_data.csv")
df.head()
```

Out[2]:
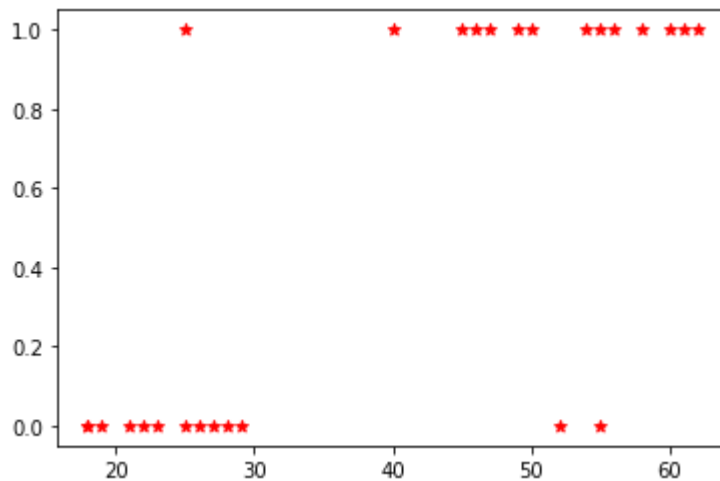
	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

In [3]:
```python
df.shape
```

Out[3]: (27, 2)

In [4]:
```python
plt.scatter(df.age, df.bought_insurance, marker = '*', color = 'red')
```

Out[4]: <matplotlib.collections.PathCollection at 0x1998dab0760>

In [5]:
```python
from sklearn.model_selection import train_test_split
```

In [6]:
```python
X_train, X_test, y_train, y_test = train_test_split(df[['age']], df.bought_insurance, t
```

In [7]:
```python
X_test
```

Out[7]:

	age
17	58
6	55
14	49
11	28

	age
10	18
22	40
8	62
15	55
12	27

In [8]:
```python
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
```

In [9]:
```python
model.fit(X_train, y_train)
```

Out[9]: LogisticRegression()

In [13]:
```python
y_predict = model.predict(X_test)
y_predict
```

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

In [14]:
```python
y_test
```

Out[14]:
```
17    1
6     0
14    1
11    0
10    0
22    1
8     1
15    1
12    0
Name: bought_insurance, dtype: int64
```

In [11]:
```python
model.predict_proba(X_test)
```

Out[11]:
```
array([[0.0467059 , 0.9532941 ],
       [0.0704143 , 0.9295857 ],
       [0.15330356, 0.84669644],
       [0.79266245, 0.20733755],
       [0.94231904, 0.05768096],
       [0.40088198, 0.59911802],
       [0.02667478, 0.97332522],
       [0.0704143 , 0.9295857 ],
       [0.81552018, 0.18447982]])
```

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

Out[15]: 0.8888888888888888

In [ ]: