

```
main.py x insurance_data.csv +
main.py > ... Form

1 import pandas as pd
2 from matplotlib import pyplot as plt
3 from sklearn.model_selection import train_test_split
4 from sklearn.linear_model import LogisticRegression
5
6
7 df = pd.read_csv("insurance_data.csv")
8 print(df.head())
9 plt.scatter(df.age, df.bought_insurance, marker='+', color='red')
10 plt.show()
11 X_train, X_test, y_train, y_test = train_test_split(df[['age']], df.bought_insurance, train_size=0.8)
12
13 print(X_test)
14
15
16 model = LogisticRegression()
17 model.fit(X_train, y_train)
18
19 y_predicted = model.predict(X_test)
20 print(y_predicted)
21
22 print(model.predict_proba(X_test))
23 print(model.score(X_test, y_test))
```

```

    age  bought_insurance
0    22             0
1    25             0
2    47             1
3    52             0
4    46             1
<matplotlib.collections.PathCollection object at 0x7ff54a15d000>
    age
8    62
24   50
1    25
2    47
9    61
7    60
[1 1 0 1 1 1]
[[0.07573248 0.92426752]
 [0.24669828 0.75330172]
 [0.854476   0.145524  ]
 [0.31649509 0.68350491]
 [0.08422056 0.91577944]
 [0.09356367 0.90643633]]
1.0

```

