In [1]:
```
from sklearn.datasets import make_classification
from matplotlib import pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
import pandas as pd
```

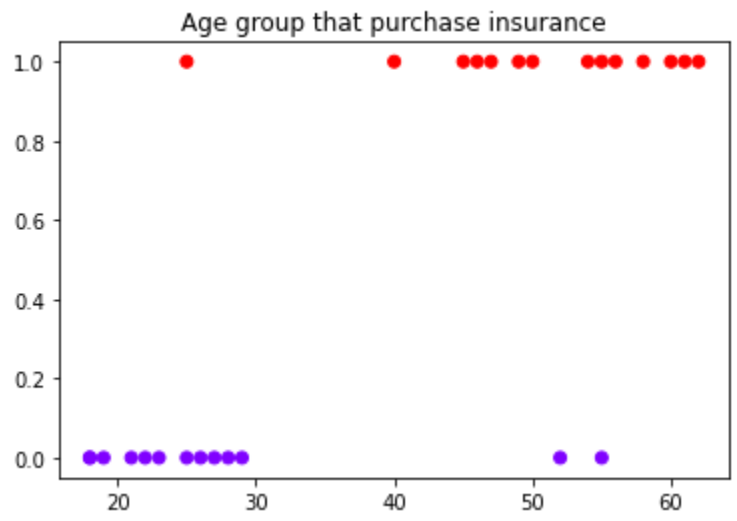
In [2]:
```
df = pd.read_excel("Insurance.xlsx")
df.head()
```

Out[2]:

	Age	Hasinsurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

In [4]:
```
x = list(df['Age'])
y = list(df['Hasinsurance'])
```

In [5]:
```
plt.scatter(x, y,c=y, cmap='rainbow',marker='o')
plt.title('Age group that purchase insurance')
plt.show()
```



In [6]:
```
from sklearn.model_selection import train_test_split
```

In [7]:
```
df.shape
```

Out[7]:
```
(27, 2)
```

In [15]:
```
x_train, x_test, y_train, y_test = train_test_split(df[['Age']],df.Hasinsurance,test_size=0.1)
```

In [16]:
```
x_test
```

Out[16]:

	Age
13	29
17	58
14	49

In [17]:
```
model = LogisticRegression()
```

In [18]:
```
model.fit(x_train, y_train)
```

Out[18]:
```
LogisticRegression()
```

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

Out[19]:
```
array([0, 1, 1], dtype=int64)
```

In [20]:
```
model.score(x_test, y_test)
```

Out[20]:
```
1.0
```

In [21]:
```
model.predict_proba(x_test)
```

Out[21]:
```
array([[0.77695064, 0.22304936],
       [0.08529966, 0.91470034],
       [0.22289917, 0.77710083]])
```

Interpretation of probability: At age 29, the person is unlikely to purchase insurance whereas, higher age group i.e. 49 and 58, are most likely to purchase insurance.