6/7/2020 Untitled50

```
In [1]: # importing the necessary libraries
    import pandas as pd
    from matplotlib import pyplot as plt
    %matplotlib inline
```

In [2]: # read the data file

df = pd.read\_csv(r"C:\Users\309962\Desktop\insurance\_data.csv")
 df.head()

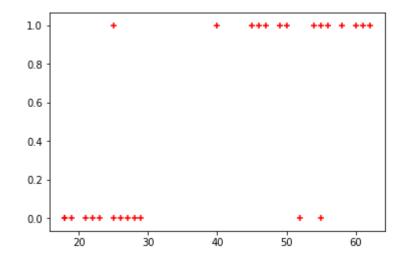
## Out[2]:

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

```
In [3]: # plot data on a scatter plot

plt.scatter(df.age,df.bought_insurance,marker='+',color='red')
```

## Out[3]: <matplotlib.collections.PathCollection at 0x168b70b9390>



In [4]: # create a train and test dataset from sklearn
from sklearn.model\_selection import train\_test\_split

6/7/2020 Untitled50

```
In [5]: X_train, X_test, y_train, y_test = train_test_split(df[['age']],df.bought_insuran
         C:\ProgramData\Anaconda3\lib\site-packages\sklearn\model selection\ split.py:20
         26: FutureWarning: From version 0.21, test size will always complement train si
         ze unless both are specified.
           FutureWarning)
In [6]: X_test
Out[6]:
             age
              55
          15
          25
              54
          24
              50
In [7]: # create a logistic regression model object
         from sklearn.linear model import LogisticRegression
         model = LogisticRegression()
In [8]: # train the model object with train dataset
         model.fit(X train, y train)
Out[8]: LogisticRegression(C=1.0, class weight=None, dual=False, fit intercept=True,
                   intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                   penalty='12', random_state=None, solver='liblinear', tol=0.0001,
                   verbose=0, warm start=False)
In [9]: X_test
Out[9]:
             age
          15
              55
          25
              54
          24
              50
         # Check the predicted value from machine learning model
In [10]:
         y predicted = model.predict(X test)
In [11]: # Predict the propabablity of purchasing an insurance policy
         model.predict proba(X test)
Out[11]: array([[0.27501496, 0.72498504],
                [0.28482804, 0.71517196],
                [0.32609841, 0.67390159]])
```

6/7/2020 Untitled50