**Practical: 4**

**Write a program to predict total payment for given number of claims on Swedish auto insurance dataset using linear regression**

**Code:**

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv('insurance.csv',

skiprows=list(range(0, 5)),

header=None,

names=['Claims', 'Payments'])

print(df.head())

print(df.shape)

plt.plot(df['Claims'],

df['Payments'],

'bo',

markersize=10,

label='Original Data')

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

x = df['Claims'].values

y = df['Payments'].values

# Convert to 2D array

x = x.reshape(-1, 1)

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x,

y,

test\_size=0.33,

random\_state=42)

model = LinearRegression()

model.fit(x\_train, y\_train)

y\_pred = model.predict(x\_test)

plt.plot(x\_train, y\_train, 'ro', label='Training Data')

plt.plot(x\_test, y\_pred, 'g', label='Predicted Data')

print("Model Score", model.score(x\_test, y\_test))

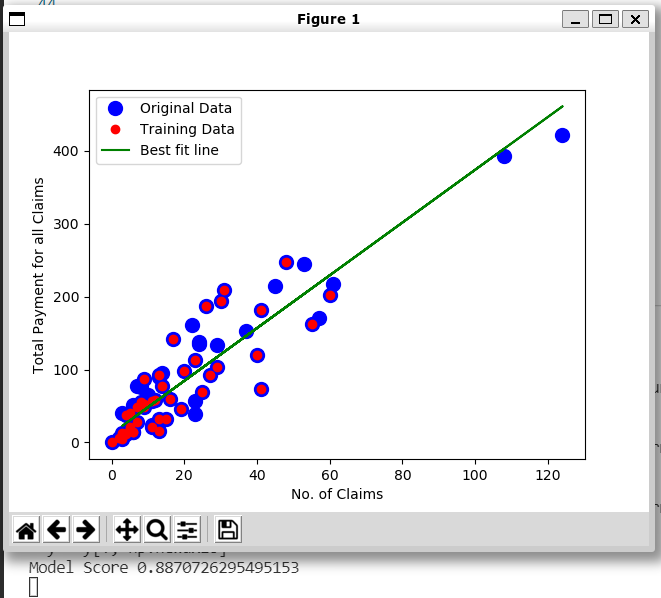
plt.xlabel('No. of Claims')

plt.ylabel('Total Payment for all Claims')

plt.legend()

plt.show()

**Output:**

****