PRACTICAL NO : 04

DATA ANALYTICS 1

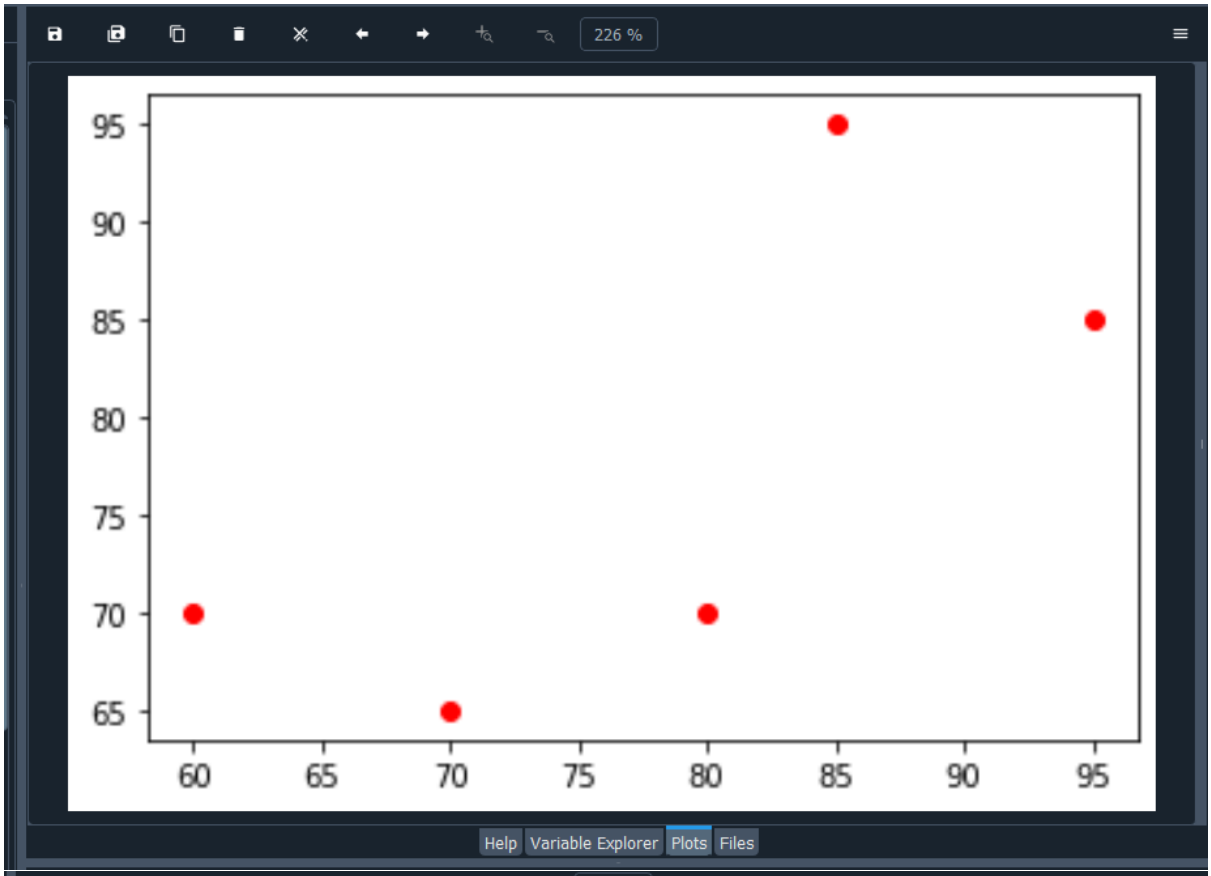
CODE :

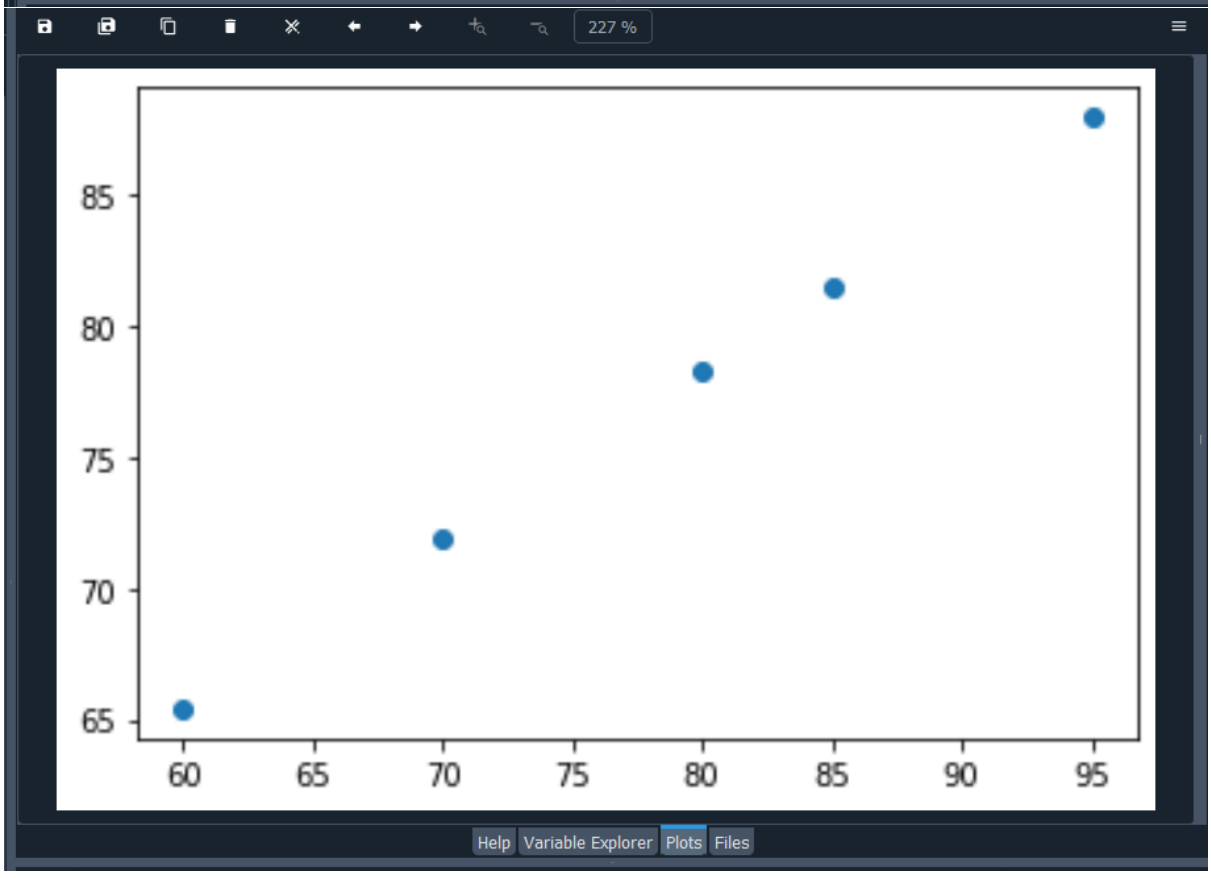
```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt


data=pd.read_csv(r"C:\Users\LP Lab\Documents\DSBDA
datasets\housing.csv")

print(data)


x=np.array([95,85,80,70,60])

y=np.array([85,95,70,65,70])

model=np.polyfit(x,y,1)

model

predict=np.poly1d(model)

predict(65)

y_pred=predict(x)

y_pred


from sklearn.metrics import r2_score

r2_score(y,y_pred)


y_line=model[1]+model[0]*x

plt.plot(x,y_line,c='r')

plt.scatter(x,y_pred)
```

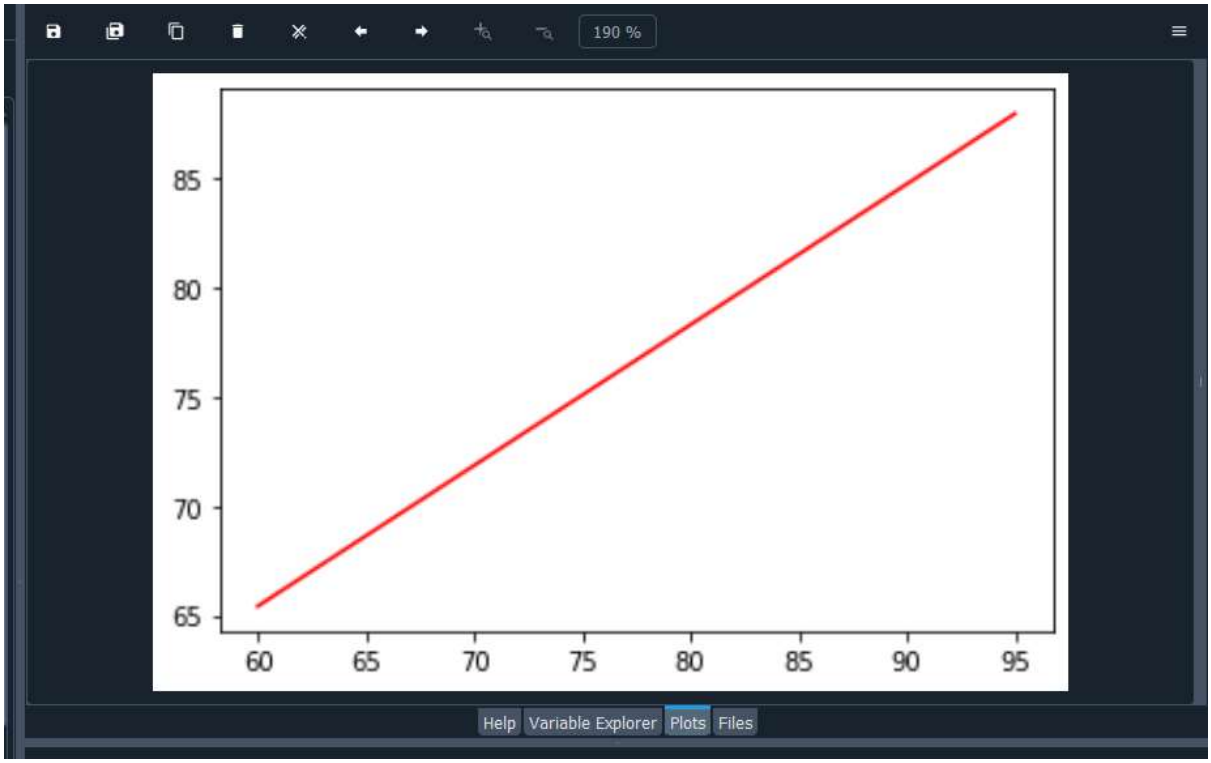
```
plt.scatter(x, y,c='r')

from sklearn.datasets import fetch_california_housing
housing=fetch_california_housing()
data=pd.DataFrame(housing.data)
data.columns=housing.feature_names
data.head()
data['PRICE']=housing.target
data.isnull().sum()
x=data.drop(['PRICE'],axis=1)
y=data['PRICE']

from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2,random_state=0)


from sklearn.linear_model import LinearRegression
lm=LinearRegression()
model=lm.fit(xtrain, ytrain)

ytrain_pred=lm.predict(xtrain)
ytest_pred=lm.predict(xtest)

df=pd.DataFrame(ytrain_pred,ytrain)
df=pd.DataFrame(ytest_pred,ytest)
```

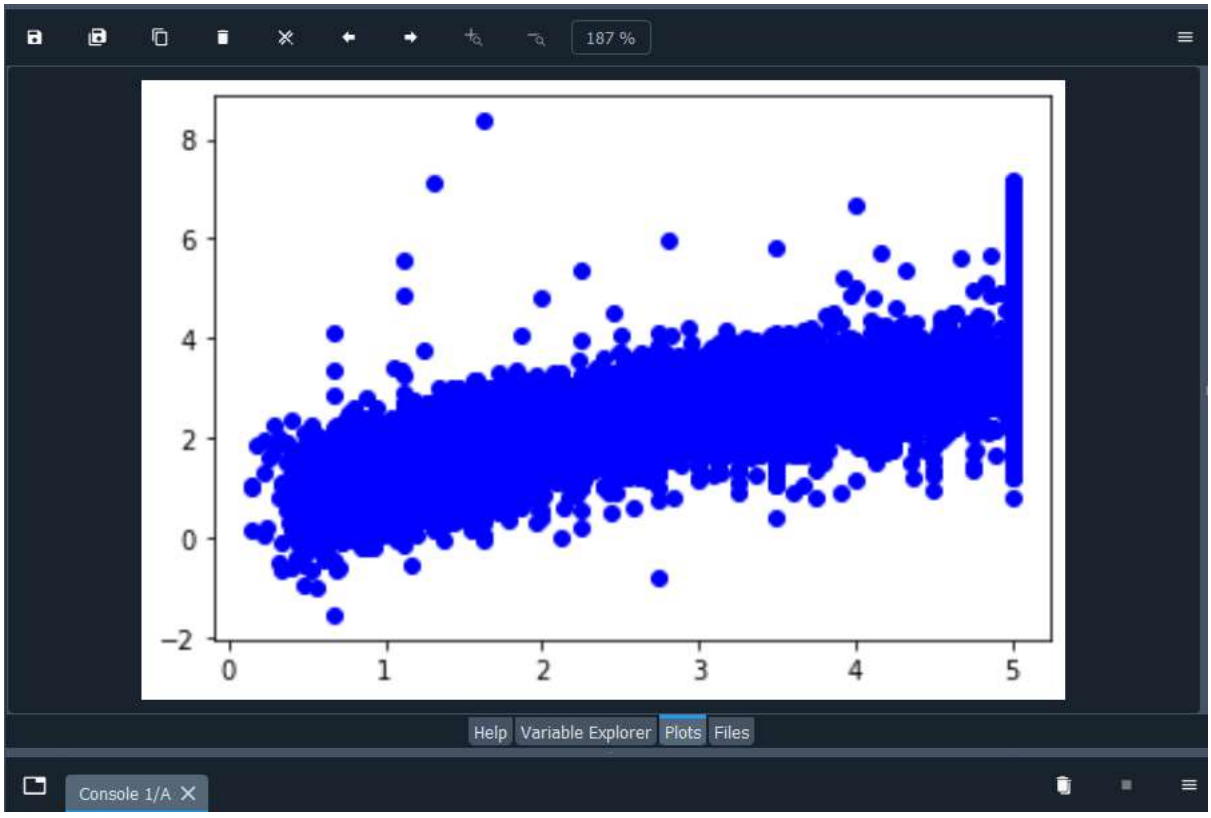
```
from sklearn.metrics import mean_squared_error,r2_score
mse=mean_squared_error(ytest, ytest_pred)
print(mse)
mse=mean_squared_error(ytrain_pred,ytrain)
print(mse)
mse=mean_squared_error(ytest,ytest_pred)
print(mse)

plt.scatter(ytrain,ytrain_pred,c='blue',marker='o',label='Training data')
plt.scatter(ytest,ytest_pred,c='lightgreen',marker='s',label='Test data')
plt.xlabel('True values')
plt.ylabel('Predicted')
plt.title('True value vs Predicted value')
plt.legend(loc='upper left')
#plt.hlines(y=0,xmin=0,xmax=50)
plt.plot()
plt.show()
```

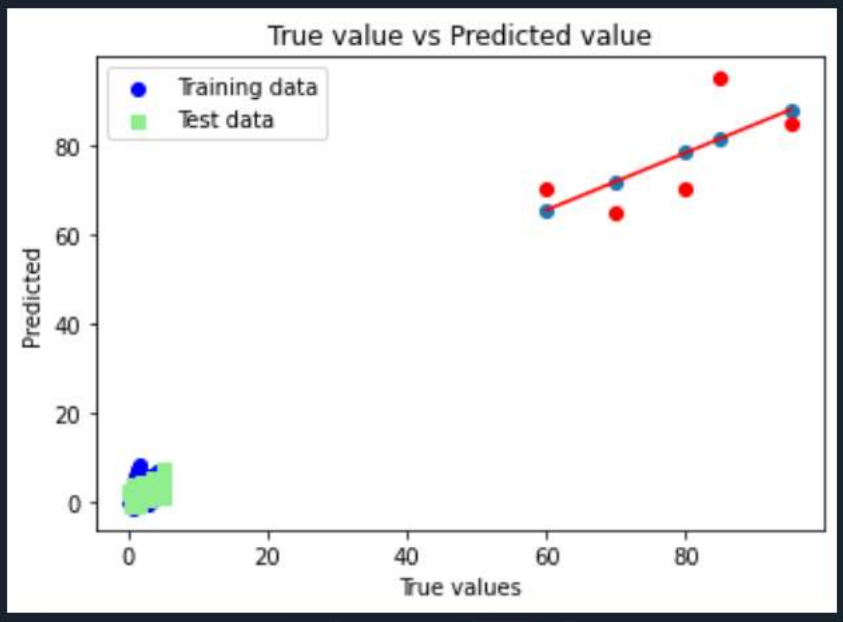
OUTPUT :


226 %
95
90
85
80
75
70
65
60
65
70
75
80
85
90
95
Help
Variable Explorer
Plots
Files


227 %
85
80
75
70
65
60
65
70
75
80
85
90
95
Help
Variable Explorer
Plots
Files


190 %
85
80
75
70
65
60
65
70
75
80
85
90
95
Help
Variable Explorer
Plots
Files


187 %
8
6
4
2
0
−2
0
1
2
3
4
5
Help
Variable Explorer
Plots
Files
Console 1/A



Help | Variable Explorer | Plots | Files

Console 1/A

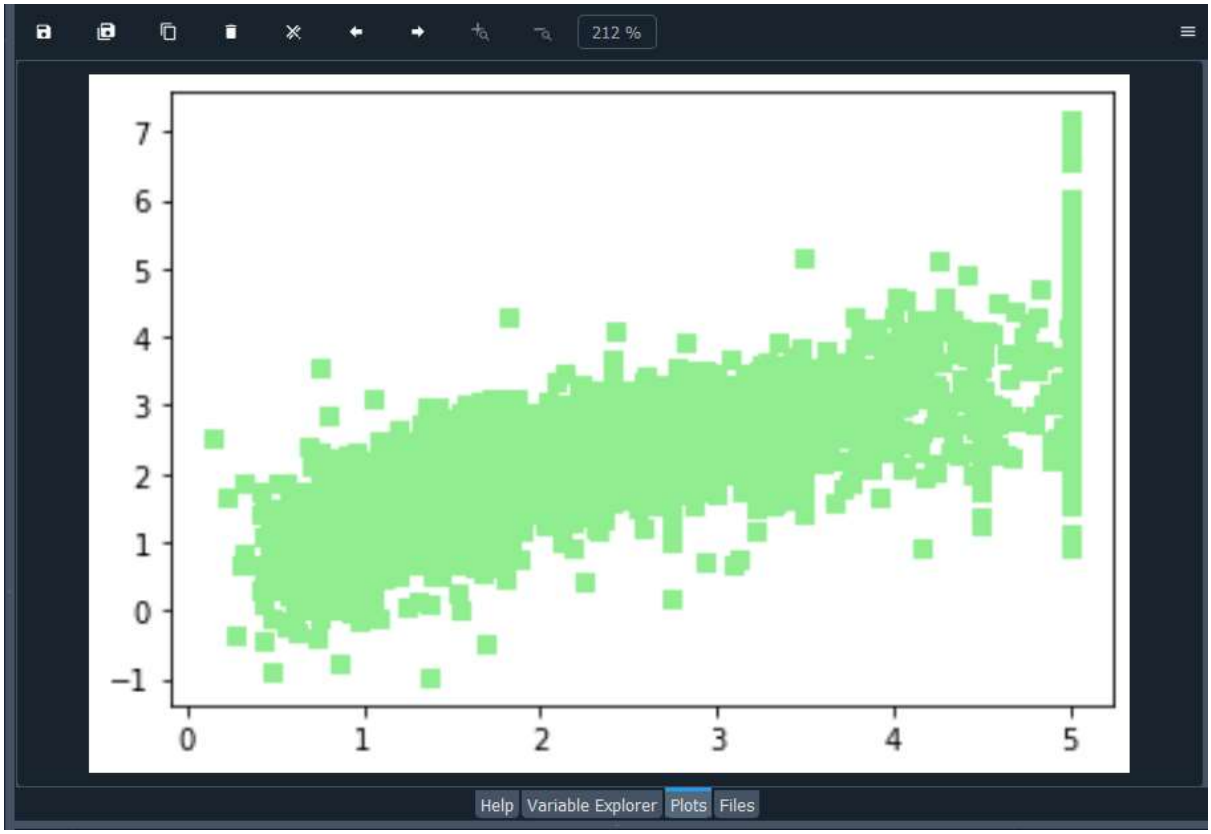
```
Type "copyright", "credits" or "license" for more information.

IPython 8.15.0 -- An enhanced Interactive Python.

In [1]: runfile('E:/DSBDA/dsbdapr4.py', wdir='E:/DSBDA')
       longitude  latitude  ...  median_house_value  ocean_proximity
0        -122.23     37.88  ...            452600.0         NEAR BAY
1        -122.22     37.86  ...            358500.0         NEAR BAY
2        -122.24     37.85  ...            352100.0         NEAR BAY
3        -122.25     37.85  ...            341300.0         NEAR BAY
4        -122.25     37.85  ...            342200.0         NEAR BAY
...          ...       ...  ...                 ...              ...
20635    -121.09     39.48  ...             78100.0           INLAND
20636    -121.21     39.49  ...             77100.0           INLAND
20637    -121.22     39.43  ...             92300.0           INLAND
20638    -121.32     39.43  ...             84700.0           INLAND
20639    -121.24     39.37  ...             89400.0           INLAND

[20640 rows x 10 columns]
0.5289841670367221
0.5234413607125449
0.5289841670367221
```

Important

Figures are displayed in the Plots pane by default. To make them also appear inline in the console, you need to uncheck "Mute inline plotting" under the options menu of Plots.

```
In [2]:
```

IPython Console | History


212 %
7
6
5
4
3
2
1
0
−1
0
1
2
3
4
5
Help
Variable Explorer
Plots
Files