import pandas as pd  
import numpy as np  
df=pd.read\_csv('C:\\Users\\rajkalyan\\Desktop\\py\\student\\Student\_dataML1.csv')  
import matplotlib.pyplot as plt  
print(df.head())  
#df.drop(['S.No','Roll num','Name'], axis=1)  
#df=df.drop(columns="Name")  
#print(df.head())  
  
##X = df.iloc[:, :-1].values  
#y = df.iloc[:, -1].values  
  
X = df.iloc[:, :-1].values  
#X=df.drop(['TOT'], axis=1)  
#y=df.drop(['IP','PCOL', 'PCOG','PJ'], axis=1)  
y = df.iloc[:, -1].values  
  
#  
#states=pd.get\_dummies(X['State'],drop\_first=True)  
  
# Drop the state coulmn  
#X=X.drop('State',axis=1)  
  
# concat the dummy variables  
#X=pd.concat([X,states],axis=1)  
  
  
#from sklearn.model\_selection import train\_test\_split  
#X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 1/3, random\_state = 0)  
  
# Fitting simple linear regression to training set  
#from sklearn.linear\_model import LinearRegression  
#regressor = LinearRegression()  
#regressor.fit(X\_train, y\_train)  
#LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=1, normalize=False)  
# Predicting the test set results  
#y\_pred = regressor.predict(X\_test)  
# Visualizing the training set  
#plt.scatter(X\_train, y\_train, color = 'red')  
#plt.plot(X\_train, regressor.predict(X\_train), color = 'blue')  
#plt.title('Subject marks vs Total marks')  
#plt.xlabel('Subject marks')  
#plt.ylabel('Total marks')  
#plt.show()  
  
  
  
  
# Splitting the dataset into the Training set and Test set  
from sklearn.model\_selection import train\_test\_split  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 0)  
  
# Fitting Multiple Linear Regression to the Training set  
from sklearn.linear\_model import LinearRegression  
regressor = LinearRegression()  
regressor.fit(X\_train, y\_train)  
print("Intercept value: ", regressor.intercept\_)  
print("coefficient value: ", regressor.coef\_)  
# Predicting the Test set results  
y\_pred = regressor.predict(X\_test)  
  
from sklearn.metrics import r2\_score  
score=r2\_score(y\_test,y\_pred)  
print(score)  
  
from sklearn import metrics  
explained\_variance=metrics.explained\_variance\_score(y\_test, y\_pred)  
mean\_absolute\_error=metrics.mean\_absolute\_error(y\_test, y\_pred)  
mse=metrics.mean\_absolute\_error(y\_test, y\_pred)  
mean\_squared\_log\_error=metrics.mean\_squared\_log\_error(y\_test, y\_pred)  
median\_absolute\_error=metrics.mean\_absolute\_error(y\_test, y\_pred)  
r2=metrics.r2\_score(y\_test, y\_pred)  
print('Explained\_variance: ', round(explained\_variance,2))  
print('Mean\_Squared\_Log\_Error: ',  
round(mean\_squared\_log\_error,2))  
print('R-squared: ', round(r2,4))  
print('Mean Absolute Error(MAE): ',  
round(mean\_absolute\_error,2))  
print('Mean Squared Error (MSE): ', round(mse,2))  
print('Root Mean Squared Error (RMSE): ',  
round(np.sqrt(mse),2))  
  
  
from statsmodels.api import OLS  
import statsmodels.api as sm  
X = sm.add\_constant(X)  
summ = OLS(y, X).fit().summary()  
print("Summary of the dataset: \n",summ)  
# plot  
#plt.scatter(X\_test, y\_test,color='black')  
#plt.plot (X\_test, y\_pred,color='blue', linewidth=3)  
#plt.xticks(())  
#plt.yticks(())  
#plt.title('SUBJECT MARKS versus TOT')  
#plt.xlabel('SUBJECT MARKS')  
  
#plt.ylabel('TOT')  
#plt.show()

"C:\Users\rajkalyan\PycharmProjects\python project 1\venv\Scripts\python.exe" C:/Users/rajkalyan/AppData/Roaming/JetBrains/PyCharmCE2022.1/scratches/Multiple.py

IP PCOL PCOG PJ TOT

0 20 22 22 22 86

1 20 22 23 22 87

2 20 22 21 21 84

3 17 19 22 20 78

4 9 13 12 13 47

Intercept value: -5.684341886080802e-14

coefficient value: [1. 1. 1. 1.]

1.0

Explained\_variance: 1.0

Mean\_Squared\_Log\_Error: 0.0

R-squared: 1.0

Mean Absolute Error(MAE): 0.0

Mean Squared Error (MSE): 0.0

Root Mean Squared Error (RMSE): 0.0

Summary of the dataset:

OLS Regression Results

==============================================================================

Dep. Variable: y R-squared: 1.000

Model: OLS Adj. R-squared: 1.000

Method: Least Squares F-statistic: 1.399e+30

Date: Tue, 06 Dec 2022 Prob (F-statistic): 0.00

Time: 10:38:18 Log-Likelihood: 3015.1

No. Observations: 103 AIC: -6020.

Df Residuals: 98 BIC: -6007.

Df Model: 4

Covariance Type: nonrobust

==============================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const 9.77e-15 3.85e-14 0.253 0.800 -6.67e-14 8.63e-14

x1 1.0000 2.96e-15 3.38e+14 0.000 1.000 1.000

x2 1.0000 4.04e-15 2.47e+14 0.000 1.000 1.000

x3 1.0000 3.97e-15 2.52e+14 0.000 1.000 1.000

x4 1.0000 4.02e-15 2.49e+14 0.000 1.000 1.000

==============================================================================

Omnibus: 5.411 Durbin-Watson: 0.088

Prob(Omnibus): 0.067 Jarque-Bera (JB): 5.404

Skew: -0.523 Prob(JB): 0.0671

Kurtosis: 2.593 Cond. No. 327.

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.