**DEPARTMENT OF INFORMATION TECHNOLOGY**

**COURSE CODE:** DJ19ITL406

**COURSE NAME:** Programing Laboratory 2 (Python)  **CLASS:** SYBTECH

**EXPERIMENT NO. 9**

**CO/LO: CO1, CO2,CO4.**

**AIM / OBJECTIVE:**

Write a Python program to implement data analysis using scipy and scikit learn

**DESCRIPTION OF EXPERIMENT:**

Describe the functions in scipy and scikit learn package

1. describe(), gmean(), hmean(), mode()
2. Liner regression of data

**QUESTIONS:**

1. Create sample data for student marks and process for the following :
   1. describe(), gmean(), hmean(), mode()

**SOURCE CODE:**

import scipy.stats as sp

marks=[15,19,25,25,24,25]

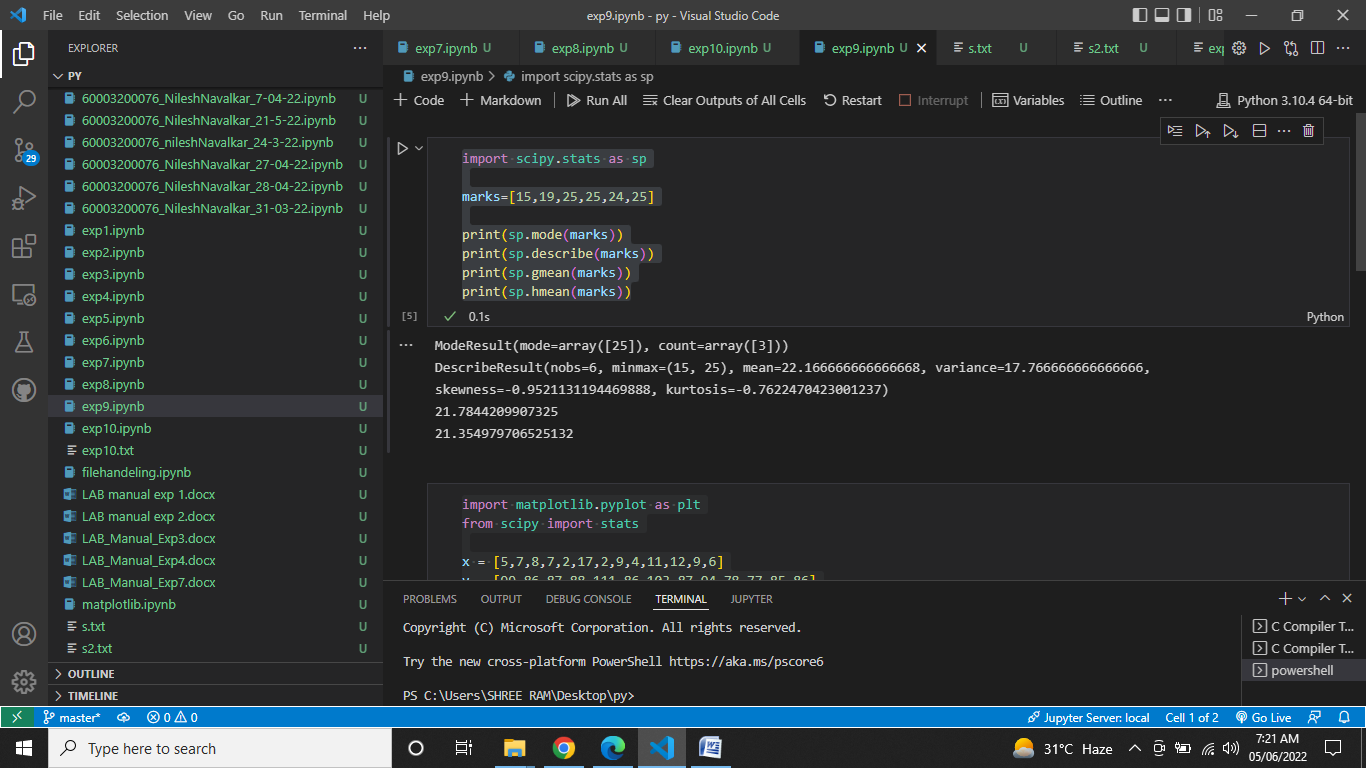
print(sp.mode(marks))

print(sp.describe(marks))

print(sp.gmean(marks))

print(sp.hmean(marks))

**OUTPUT:**

****

1. Perform Linear regression on any suitable data.

**SOURCE CODE:**

import matplotlib.pyplot as plt

from scipy import stats

x = [5,7,8,7,2,17,2,9,4,11,12,9,6]

y = [99,86,87,88,111,86,103,87,94,78,77,85,86]

slope, intercept, r, p, std\_err = stats.linregress(x, y)

def myfunc(x):

  return slope \* x + intercept

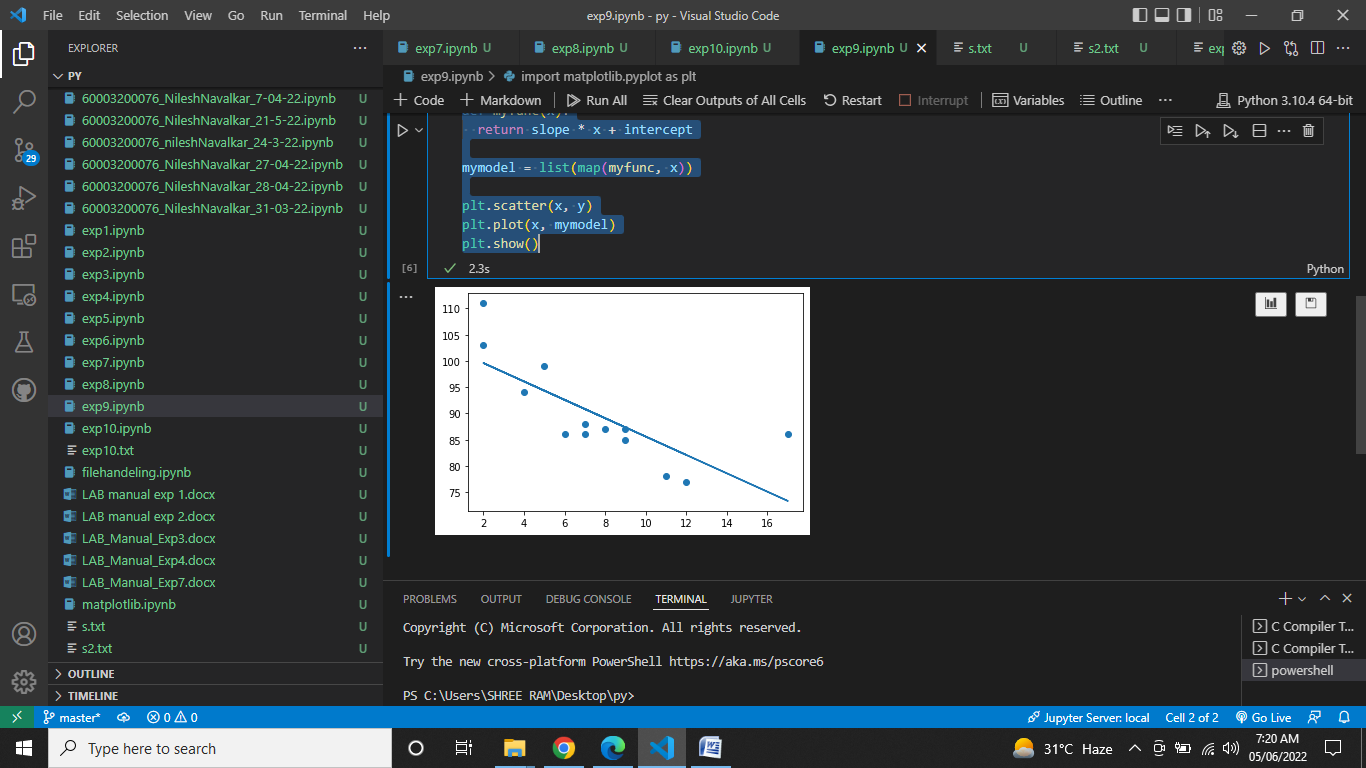
mymodel = list(map(myfunc, x))

plt.scatter(x, y)

plt.plot(x, mymodel)

plt.show()

**OUTPUT:**



**OBSERVATIONS / DISCUSSION OF RESULT:**

After performing the experiments we observed that,

1)The describe() function is used to get the general information abou the data presented in the form of a table

2)The mode function tells about the value which appears in the dataframe the most

**CONCLUSION:**

Hence, we successfully implemented

1. describe(), gmean(), hmean(), mode()
2. Liner regression of data

**REFERENCES:**

**Website References:​**

[1] https://www.w3schools.com/python