**Experiment – 3**

**Student Name: Vivek Kumar UID: 21BCS8129**

**Branch: BE-CSE(LEET) Section/Group: WM-20BCS-616/A**

**Semester: 5th Date of Performance: 31/08/2022**

**Subject Name: Machine Learning Lab Subject Code: 20CSP-317**

**1. Aim/Overview of the practical:**

Implement Linear Regression on any data set.

**2. Task to be done/ Which logistics used:**

Implement Linear Regression on any data set using sklearn.

**3. Algorithm/Flowchart (For programming-based labs):**

**4. Steps for experiment/practical/Code:**

from google.colab import drive

drive.mount('/content/drive')

import pandas as pd

import numpy as np

data=pd.read\_csv('/content/drive/MyDrive/Data/Students\_data.csv')

data.head()

data.isnull()

data.isnull().sum()

import matplotlib.pyplot as plt

import seaborn as sn

import sklearn

data.plot.scatter('race','Algebra')

x=data[['GPA',  'Algebra',  'Calculus1',    'Calculus2',    'Statistics'    ,'Probability', 'Measure'   ,'Functional\_analysis', 'from4',    'y']]

y=data['race']

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

from sklearn.linear\_model import LinearRegression

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.4)

lm=LinearRegression()

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

predictions=lm.predict(x\_test)

plt.scatter(y\_test,predictions)

plt.xlabel('y test')

plt.ylabel('predicted y')

from sklearn import metrics

print('MAE:',metrics.mean\_absolute\_error(y\_test,predictions))

print('MSE:',metrics.mean\_squared\_error(y\_test,predictions))

print('RMSE:',np.sqrt(metrics.mean\_squared\_error(y\_test,predictions)))

sn.distplot((y\_test-predictions),bins=50)

coefficient=pd.DataFrame(lm.coef\_,x.columns)

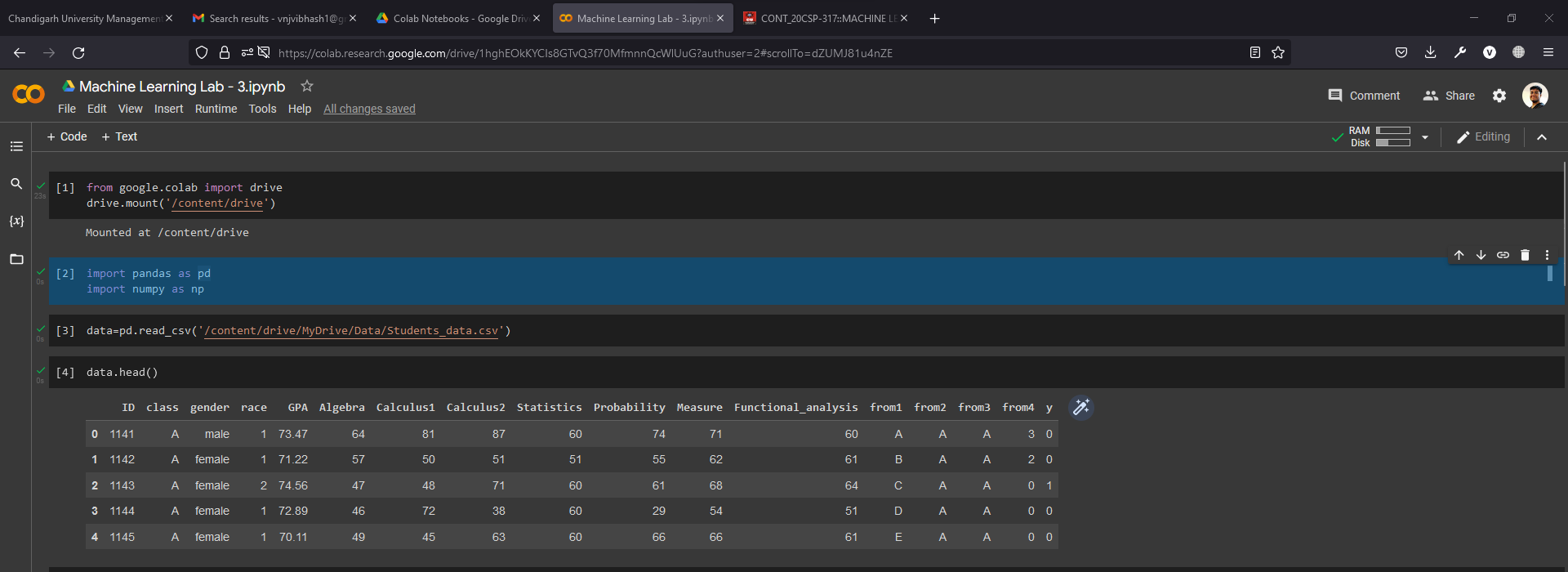
coefficient.columns=['coefficient']

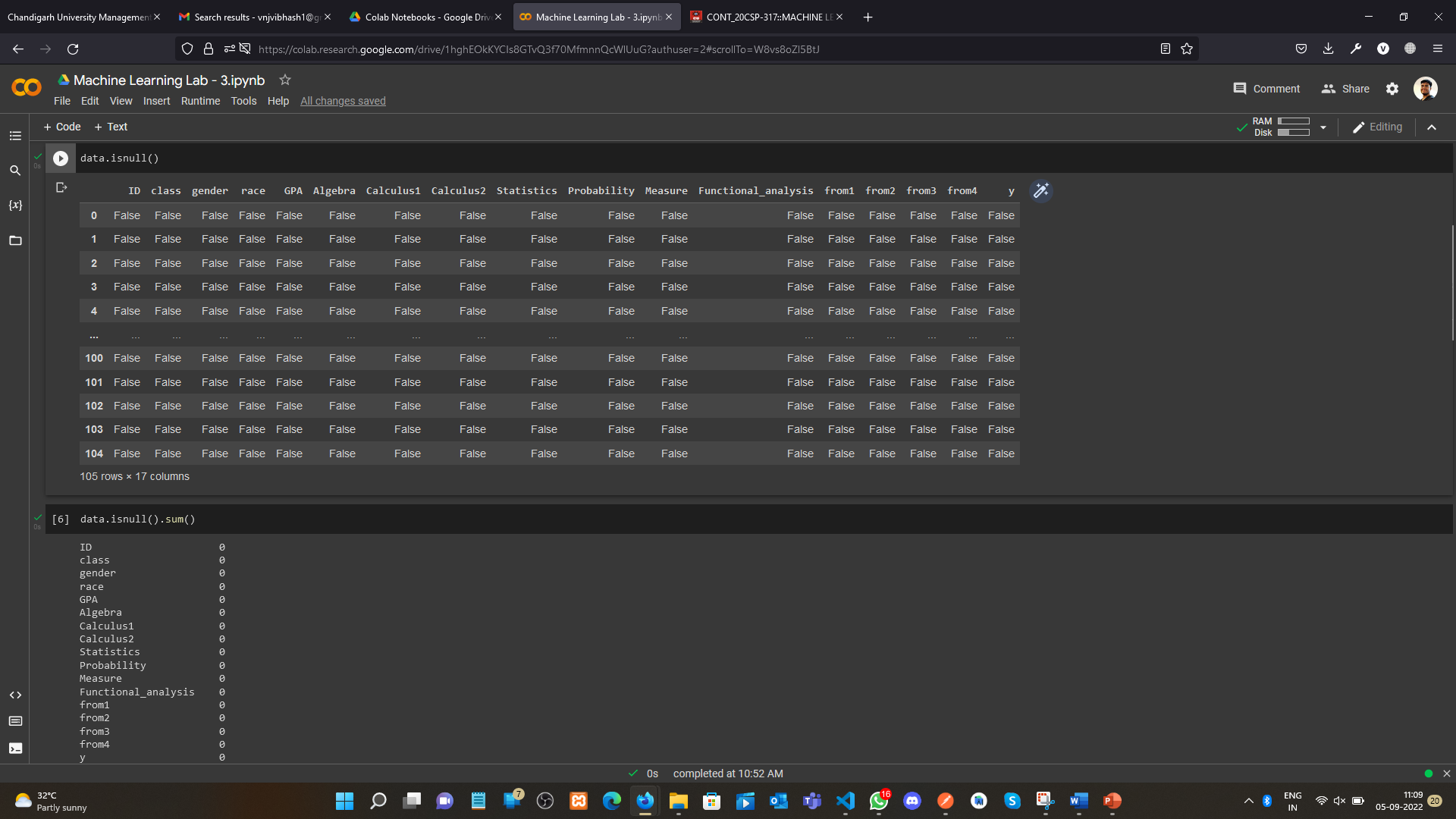
coefficient

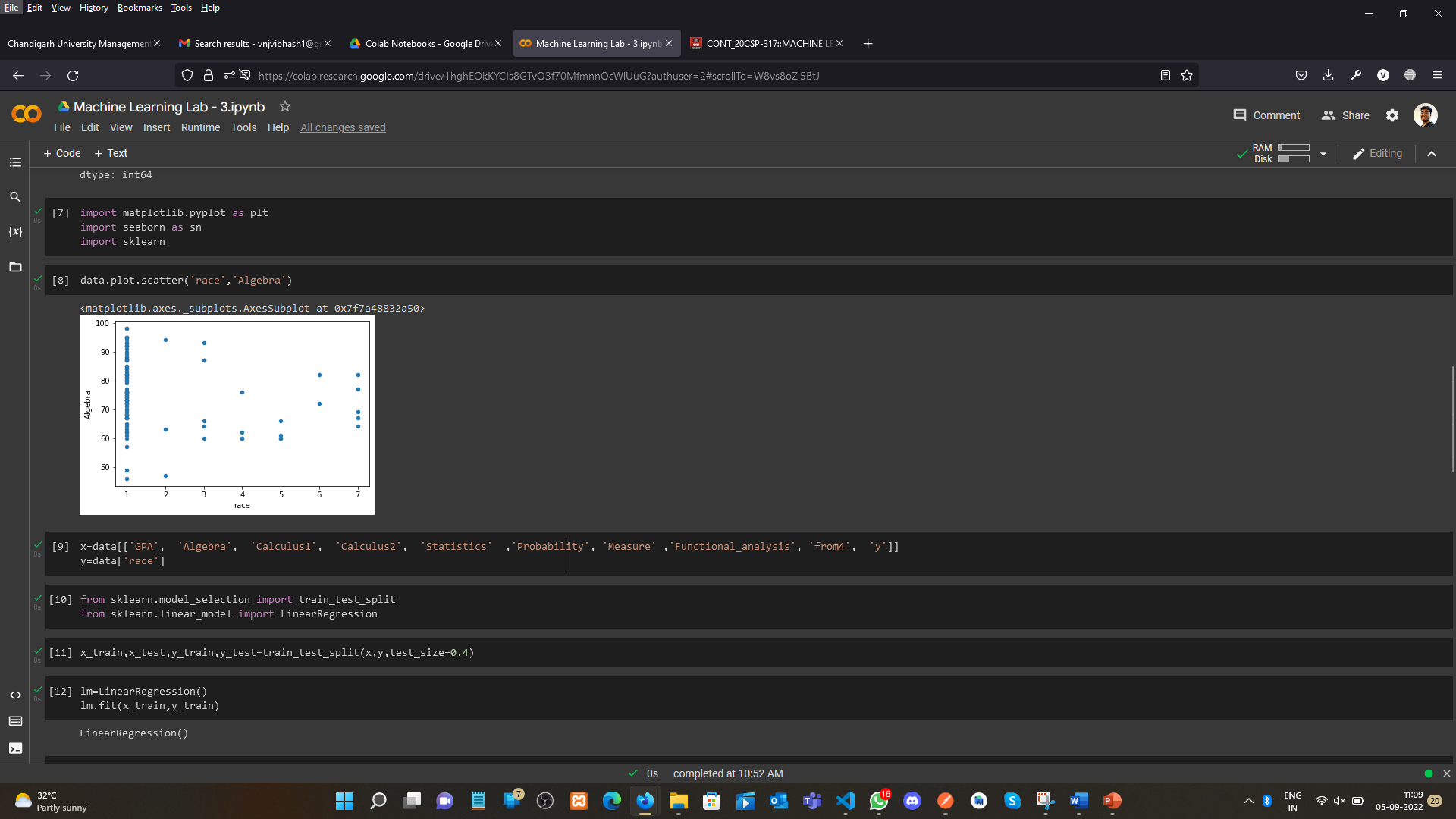
**5. Observations/Discussions/ Complexity Analysis:**

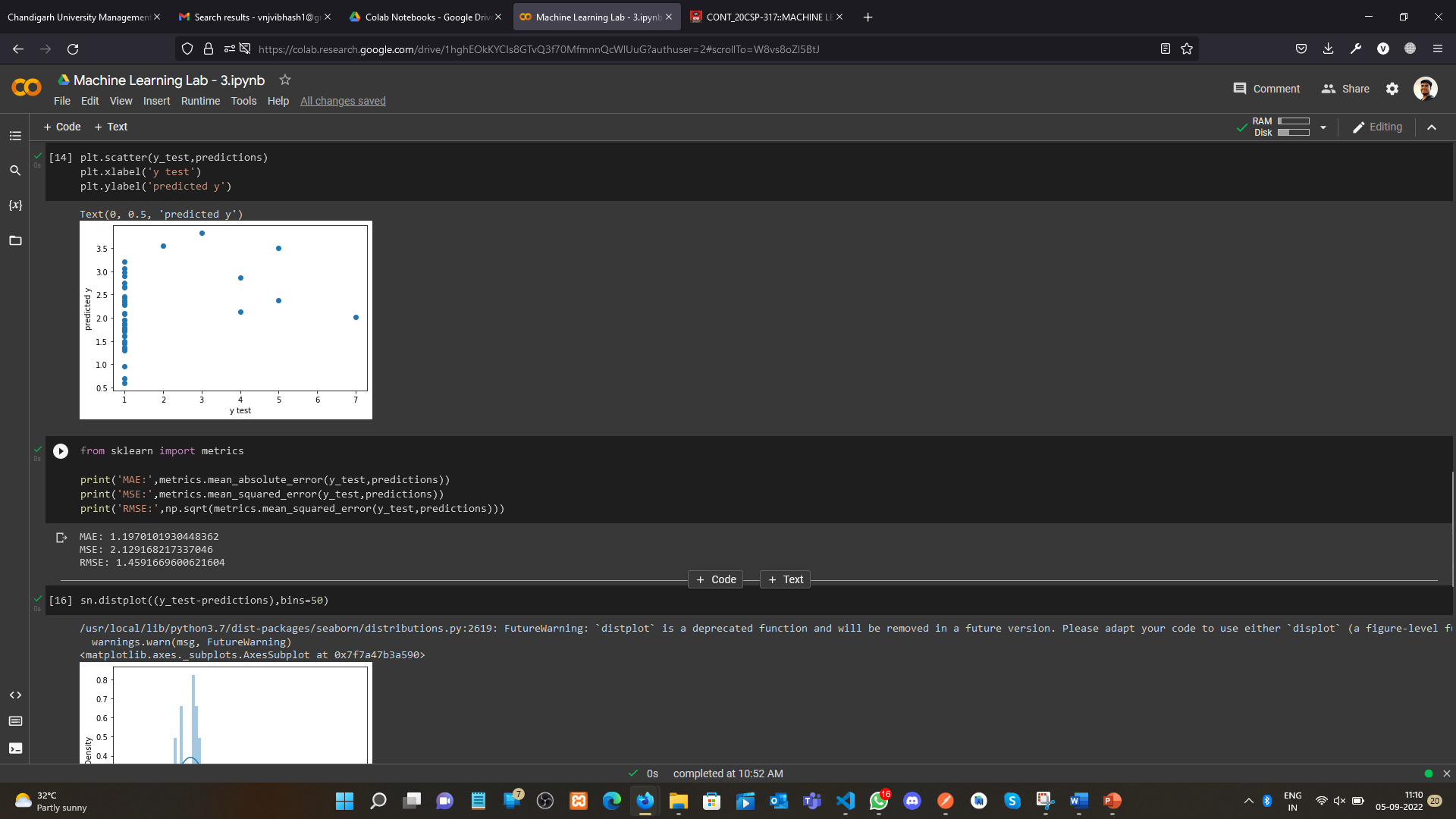
In this I have implemented simple Liner regression based on the data by splitting into two parts such as training and testing as well as I have displayed the plotted the graph and displayed coefficient of all the related columns.

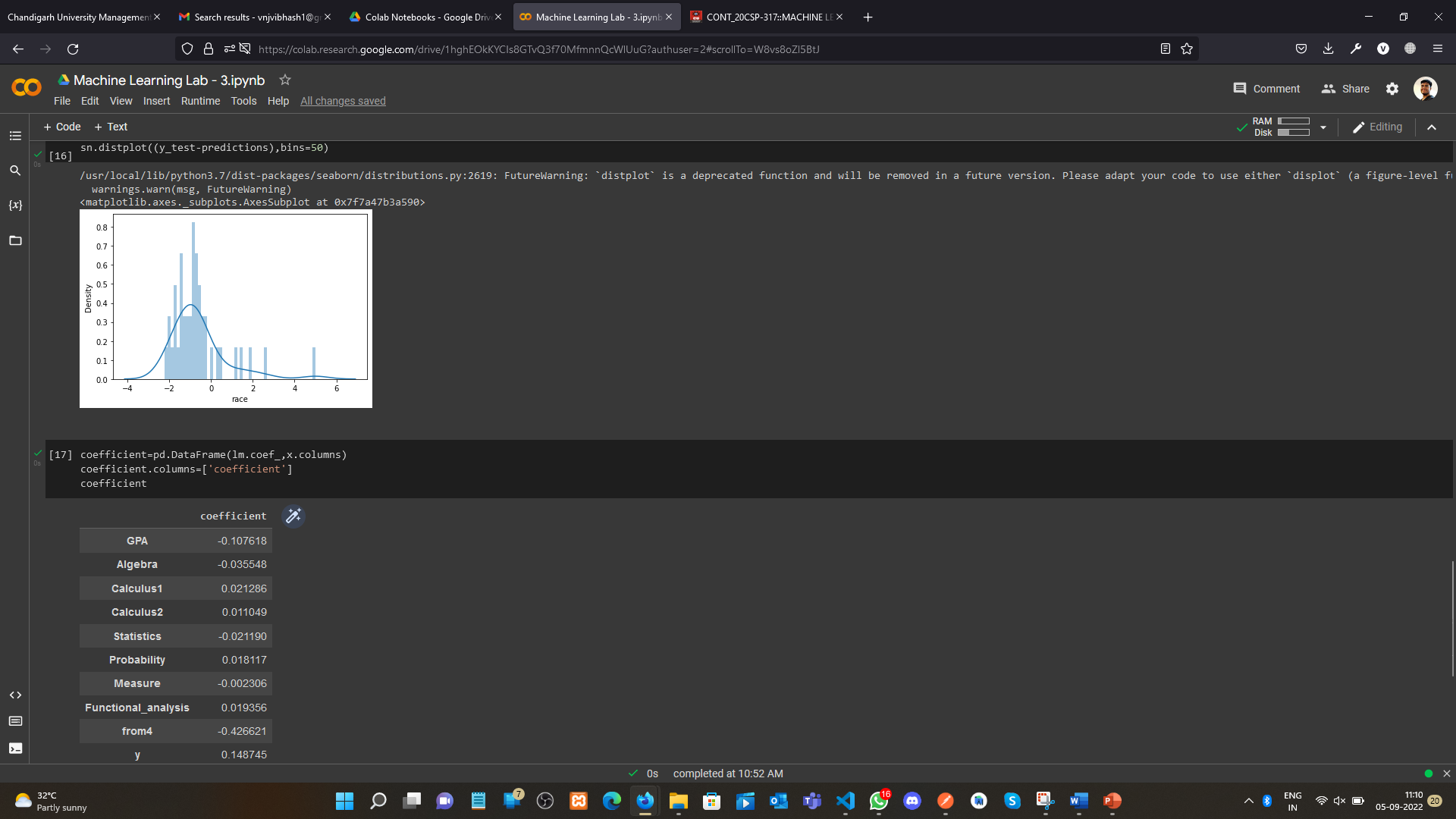
**6. Result/Output/Writing Summary:**

****









**Learning outcomes (What I have learnt):**

1. Data Splitting
2. Linear regression concept
3. Plotting the prediction.

**Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Parameters | Marks Obtained | Maximum Marks |
| 1. |  |  |  |
| 2. |  |  |  |
| 3. |  |  |  |
|  |  |  |  |