Experiment - 1

Student Name: Vivek Kumar **UID: 21BCS8129**
Branch: BE-CSE(LEET) **Section/Group: WM-20BCS-616/A**
Semester: 5th **Date of Performance: 16/08/2022**
Subject Name: Machine Learning Lab **Subject Code: 20CSP-317**

1. Aim/Overview of the practical:

Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.

2. Task to be done/ Which logistics used:

Data Manipulation using the Pandas and Seaborn library

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

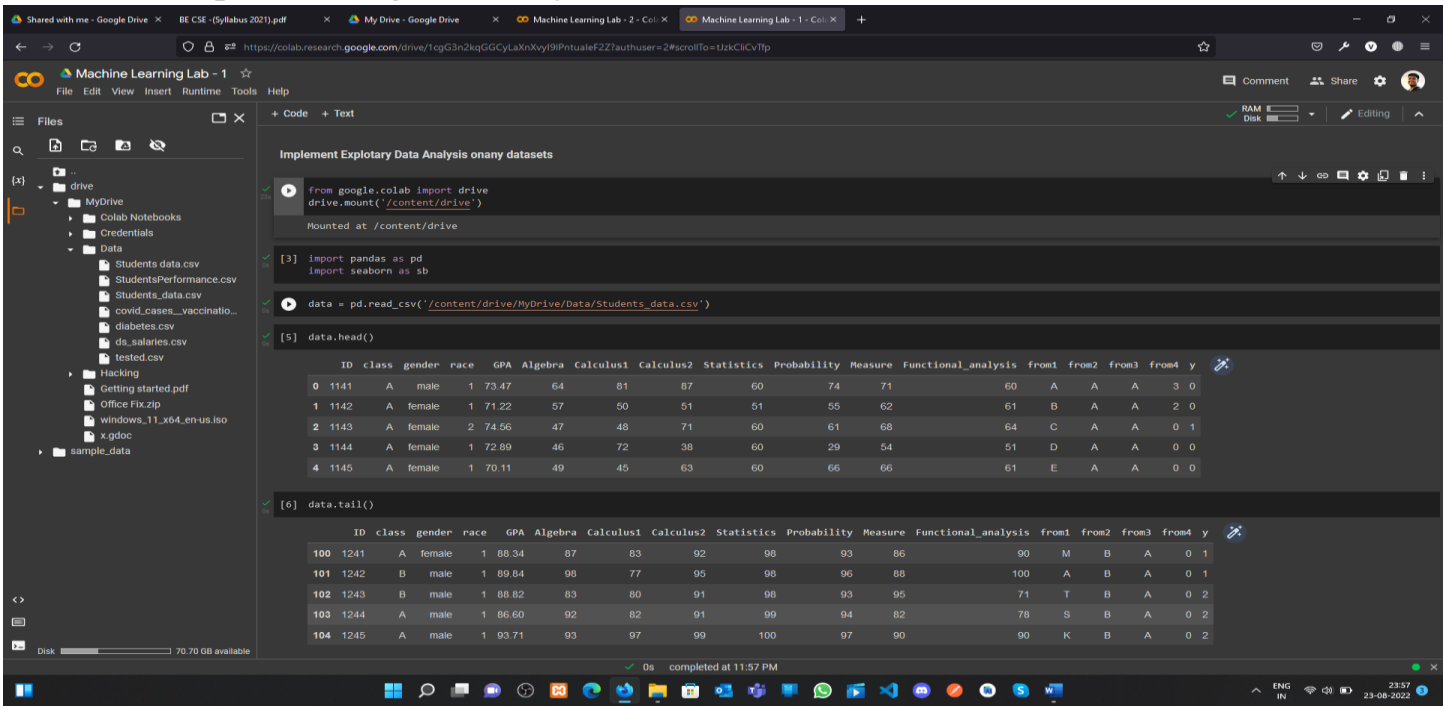
4. Steps for experiment/practical/Code:

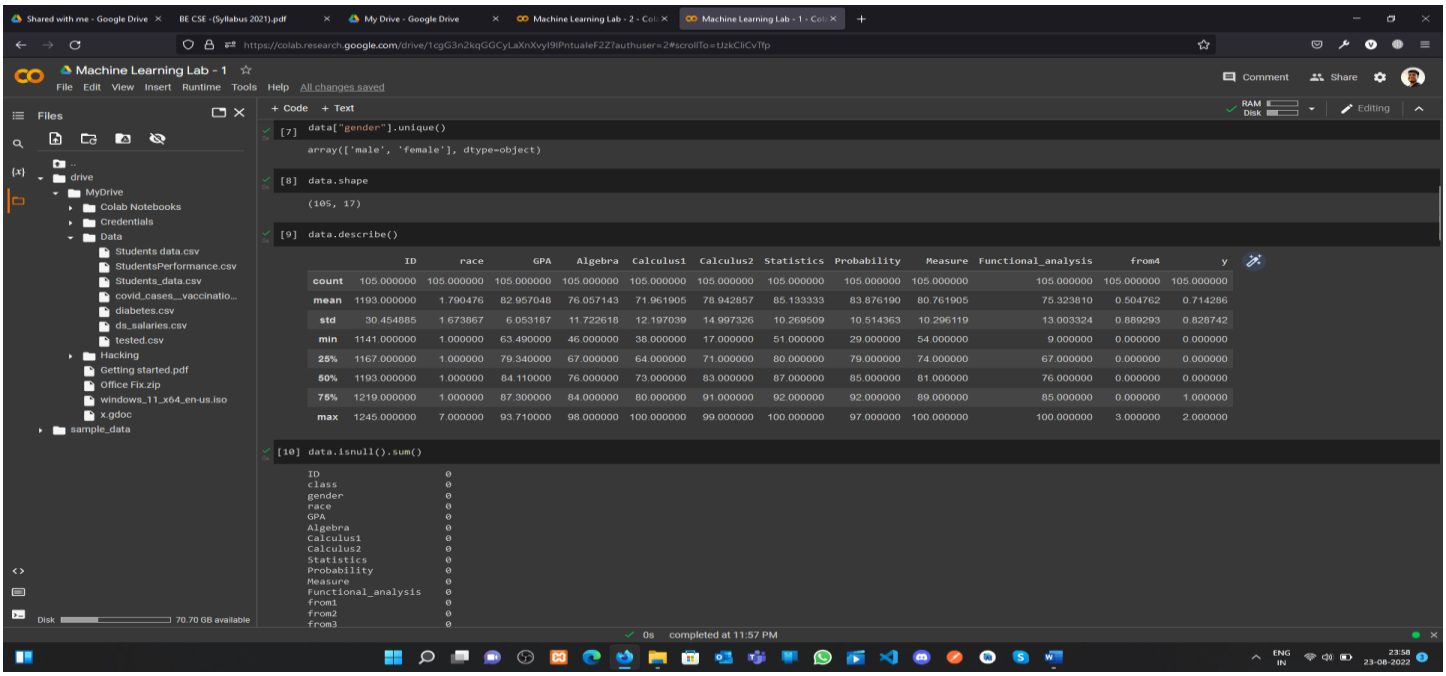
```
import pandas as pd
import seaborn as sb
data = pd.read_csv('/content/drive/MyDrive/Data/Students_data.csv')
data.head()
data.tail()
data["gender"].unique()
data.shape
data.describe()
data.isnull().sum()
data.iloc[:5, 0]
data.iloc[:5, 1]
data['GPA']
data.groupby('race').agg({'GPA': 'count'})
data.groupby('race').agg({'GPA': 'median'})
data.columns
new_data = data.drop('Calculus1', axis=1)
new_data.head
corelation = new_data.corr()
sb.heatmap(corelation, xticklabels=corelation.columns, yticklabels=corelation.columns
, annot=True)
sb.pairplot(new_data)
sb.relplot(x='GPA',y='Algebra',hue='race',data=data)
sb.distplot(data['Algebra'])
sb.catplot(x='GPA',kind='box',data=data)
```

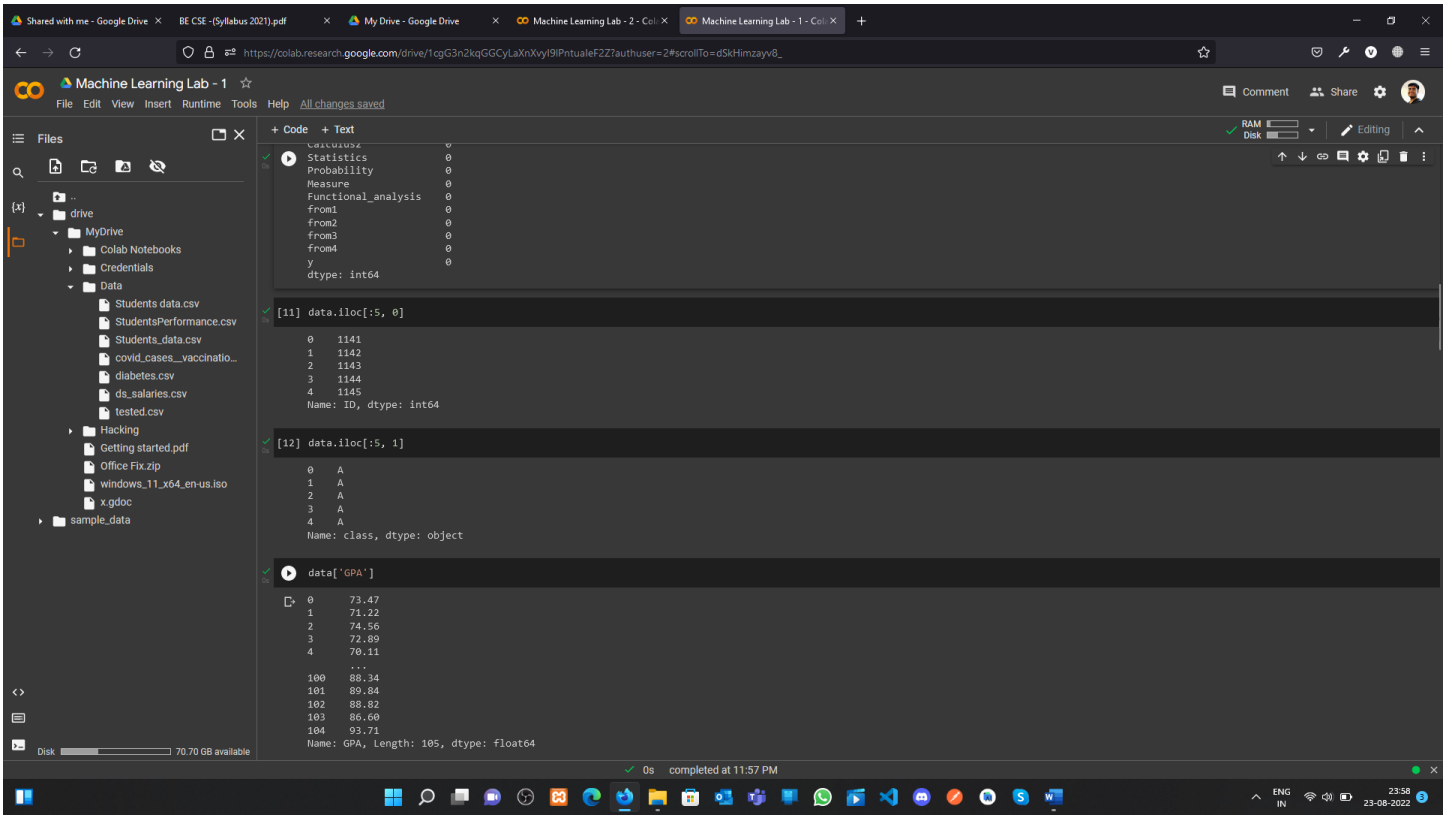
5. Observations/Discussions/ Complexity Analysis:

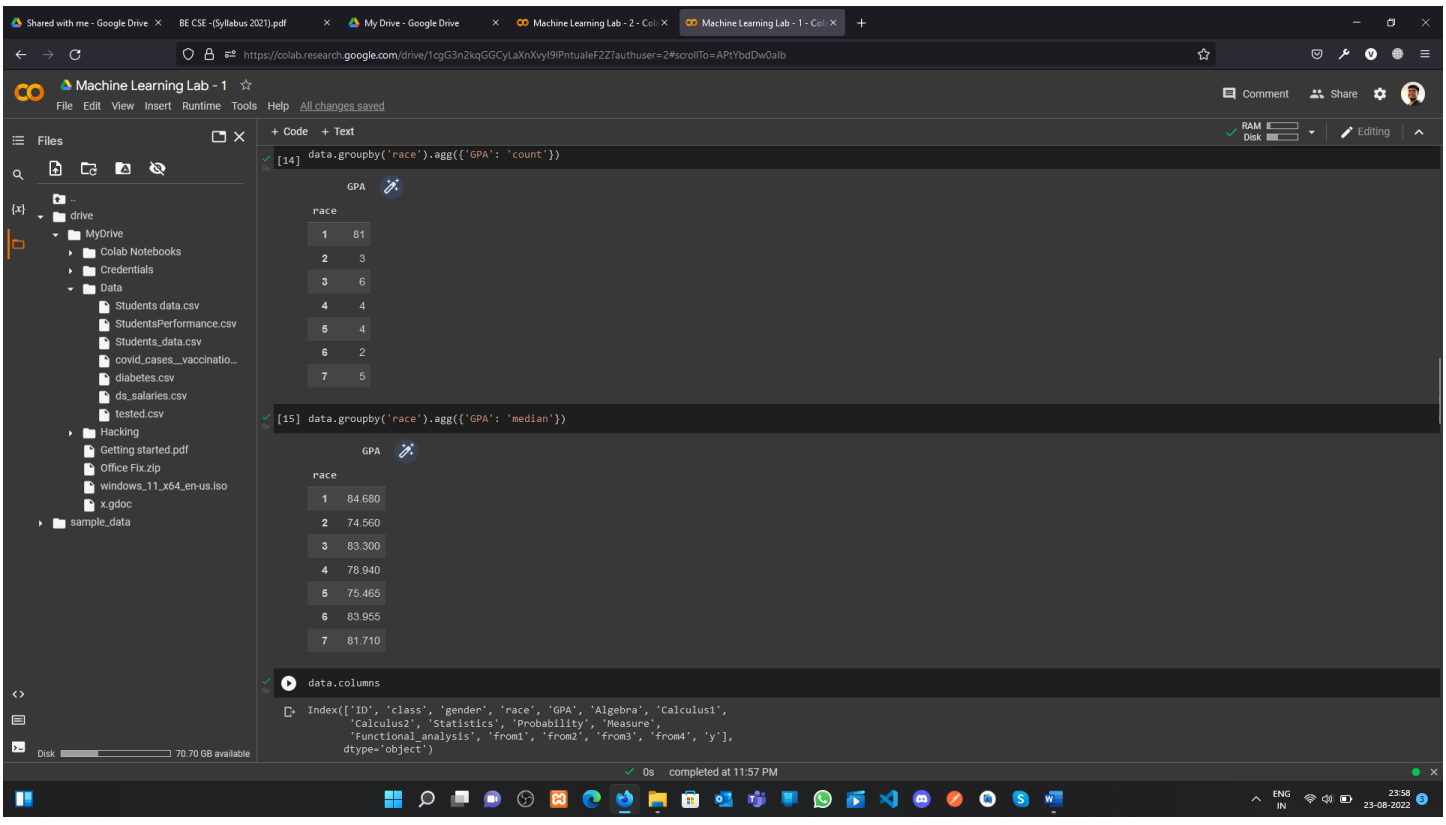
In this Experiment we have done the Data manipulation in such way where we have found the Unique, count of data, head, Tail, shape, and Descriptive data analysis, iloc, groupby, column, and core relation of the data. Moreover, we have Plotted the graph using seaborn library such as heatmap, pairplot, relplot, distplot and catplot.

6. Result/Output/Writing Summary:






Machine Learning Lab - 1
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk Editing
Files
drive
MyDrive
Colab Notebooks
Credentials
Data
Students data.csv
StudentsPerformance.csv
Students_data.csv
covid_cases_vaccinatio...
diabetes.csv
ds_salaries.csv
tested.csv
Hacking
Getting started.pdf
Office Fix.zip
windows_11_x64_en-us.iso
x.gdoc
sample_data
Disk 70.70 GB available
Statistics 0
Probability 0
Measure 0
Functional_analysis 0
from1 0
from2 0
from3 0
from4 0
y 0
dtype: int64
[11] data.iloc[:5, 0]
0 1141
1 1142
2 1143
3 1144
4 1145
Name: ID, dtype: int64
[12] data.iloc[:5, 1]
0 A
1 A
2 A
3 A
4 A
Name: class, dtype: object
data['GPA']
0 73.47
1 71.22
2 74.56
3 72.89
4 70.11
100 88.34
101 89.84
102 88.82
103 86.60
104 93.71
Name: GPA, Length: 105, dtype: float64
0s completed at 11:57 PM


Machine Learning Lab - 1
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk Editing
Files
drive
MyDrive
Colab Notebooks
Credentials
Data
Students data.csv
StudentsPerformance.csv
Students_data.csv
covid_cases_vaccinatio...
diabetes.csv
ds_salaries.csv
tested.csv
Hacking
Getting started.pdf
Office Fix.zip
windows_11_x64_en-us.iso
x.gdoc
sample_data
Disk 70.70 GB available
[14] data.groupby('race').agg({'GPA': 'count'})
GPA
race
1 81
2 3
3 6
4 4
5 4
6 2
7 5
[15] data.groupby('race').agg({'GPA': 'median'})
GPA
race
1 84.680
2 74.560
3 83.300
4 78.940
5 75.465
6 83.955
7 81.710
data.columns
Index(['ID', 'class', 'gender', 'race', 'GPA', 'Algebra', 'Calculus1',
'Calculus2', 'Statistics', 'Probability', 'Measure',
'Functional_analysis', 'from1', 'from2', 'from3', 'from4', 'y'],
dtype='object')
0s completed at 11:57 PM

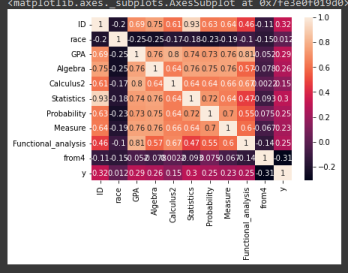
Machine Learning Lab - 1

```
[38] new_data = data.drop('Calculus1', axis=1)
     new_data.head()
```

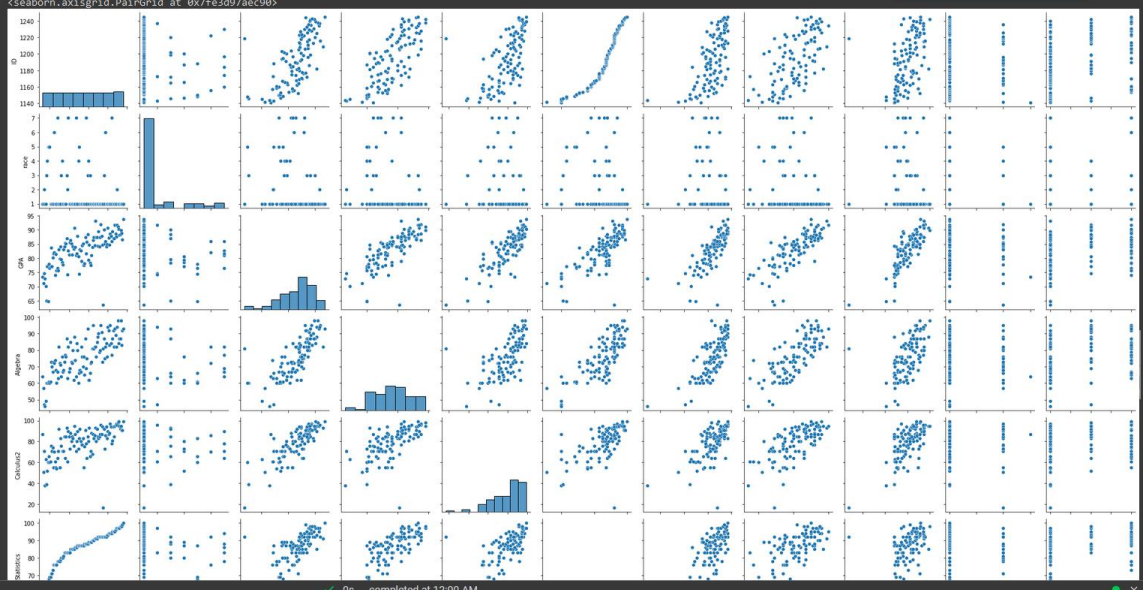
	ID	class	gender	race	GPA	Algebra	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y
0	1141	A	male	1	73.47	64	87	60	74	71	60	A	A	A	3	0
1	1142	A	female	1	71.22	57	51	51	55	62	61	B	A	A	2	0
2	1143	A	female	2	74.56	47	71	60	61	68	64	C	A	A	0	1
3	1144	A	female	1	72.89	46	38	60	29	54	51	D	A	A	0	0
4	1145	A	female	1	70.11	49	63	60	66	66	61	E	A	A	0	0

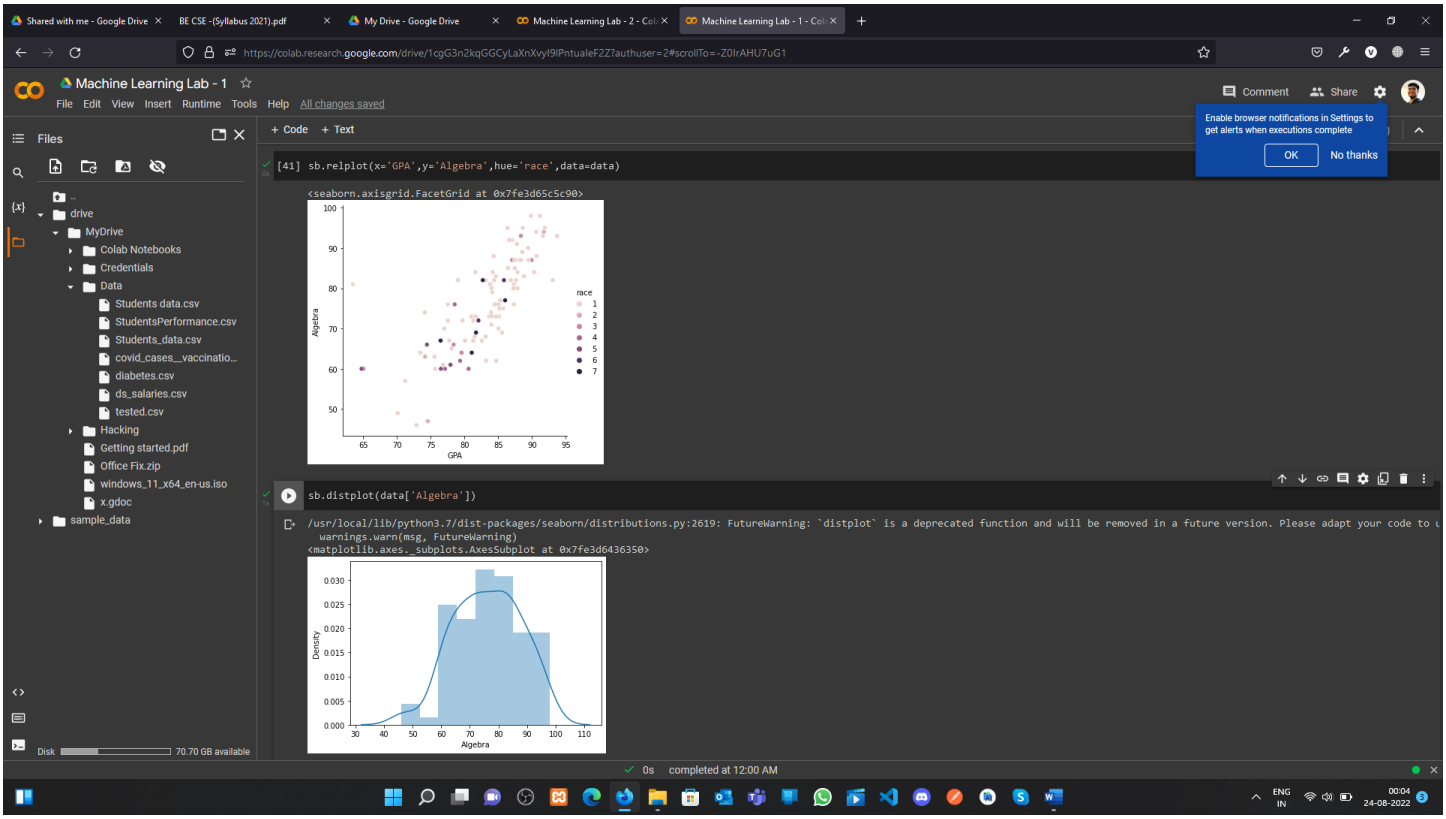
Relationship Analysis

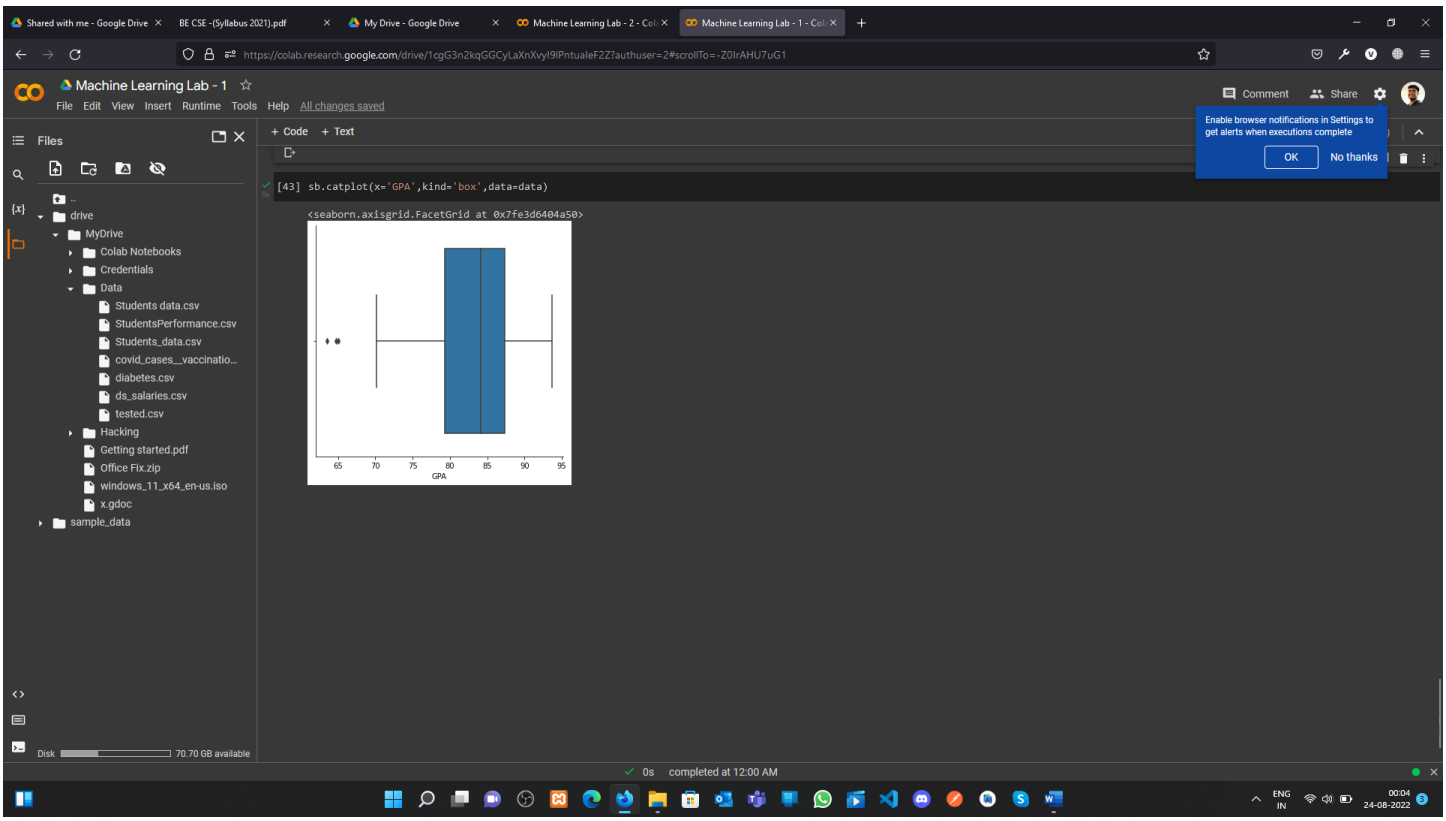
```
corelation = new_data.corr()
sb.heatmap(corelation, xticklabels=corelation.columns, yticklabels=corelation.columns, annot=True)
```




```
[40] sb.pairplot(new_data)
```




Machine Learning Lab - 1
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Enable browser notifications in Settings to get alerts when executions complete
OK No thanks
+ Code + Text
[41] sb.relplot(x='GPA',y='Algebra',hue='race',data=data)
<seaborn.axisgrid.FacetGrid at 0x7fe3d65c5c90>
race
sb.distplot(data['Algebra'])
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to u
warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fe3d6436350>
Density
Algebra
Files
drive
MyDrive
Colab Notebooks
Credentials
Data
Students data.csv
StudentsPerformance.csv
Students_data.csv
covid_cases_vaccinatio...
diabetes.csv
ds_salaries.csv
tested.csv
Hacking
Getting started.pdf
Office Fix.zip
windows_11_x64_en-us.iso
x.gdoc
sample_data
Disk 70.70 GB available
0s completed at 12:00 AM


Machine Learning Lab - 1
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Enable browser notifications in Settings to get alerts when executions complete
OK No thanks
+ Code + Text
[43] sb.catplot(x='GPA',kind='box',data=data)
<seaborn.axisgrid.FacetGrid at 0x7fe3d6404a50>
GPA
Files
drive
MyDrive
Colab Notebooks
Credentials
Data
Students data.csv
StudentsPerformance.csv
Students_data.csv
covid_cases_vaccinatio...
diabetes.csv
ds_salaries.csv
tested.csv
Hacking
Getting started.pdf
Office Fix.zip
windows_11_x64_en-us.iso
x.gdoc
sample_data
Disk 70.70 GB available
0s completed at 12:00 AM

Learning outcomes (What I have learnt):

1. Data manipulation using padas library

2. Data plotting using seaborn library

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.			