

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):

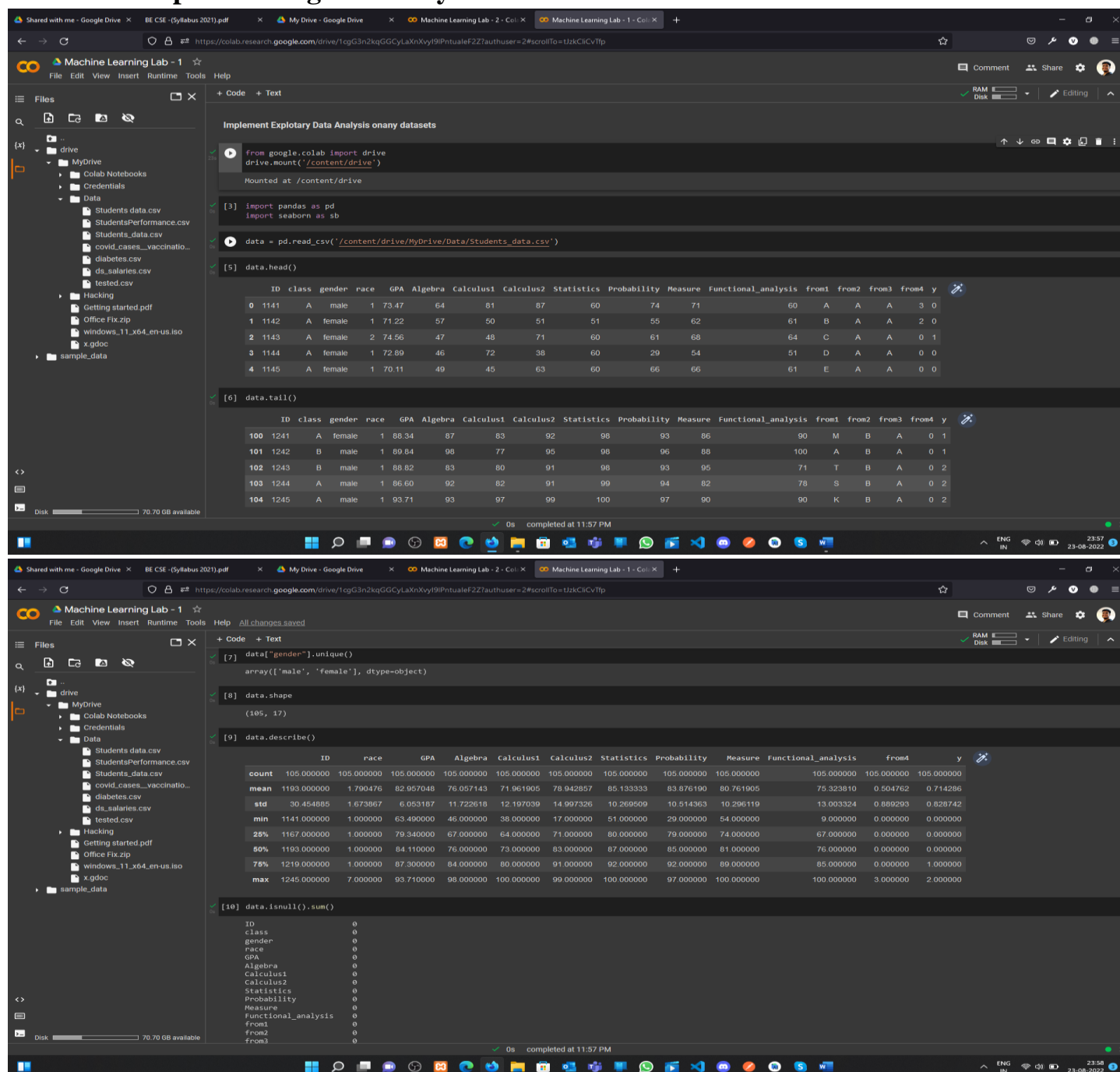
4. Steps for experiment/practical/Code:

```
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
correlation = new_data.corr()
sb.heatmap(correlation, xticklabels=correlation.columns, yticklabels=correlation.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:



The screenshot displays a Google Colab environment with the following code and output:

```

from google.colab import drive
drive.mount('/content/drive')

[3] import pandas as pd
import seaborn as sb

[5] data = pd.read_csv('/content/drive/MyDrive/Data/Students_data.csv')

[5] data.head()

```

ID	class	gender	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y	
0	1141	A	male	1	73.47	64	81	87	60	74	71	60	A	A	A	3	0
1	1142	A	female	1	71.22	57	50	51	51	55	62	61	B	A	A	2	0
2	1143	A	female	2	74.56	47	48	71	60	61	68	64	C	A	A	0	1
3	1144	A	female	1	72.89	46	72	38	60	29	54	51	D	A	A	0	0
4	1145	A	female	1	70.11	49	45	63	60	66	66	61	E	A	A	0	0

```

[6] data.tail()

```

ID	class	gender	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y	
100	1241	A	female	1	88.34	87	83	92	98	93	86	90	M	B	A	0	1
101	1242	B	male	1	89.84	98	77	95	96	96	88	100	A	B	A	0	1
102	1243	B	male	1	88.82	83	80	91	96	93	95	71	T	B	A	0	2
103	1244	A	male	1	86.60	92	82	91	99	94	82	78	S	B	A	0	2
104	1245	A	male	1	93.71	93	97	99	100	97	90	90	K	B	A	0	2

```

[7] data["gender"].unique()
array(['male', 'female'], dtype=object)

[8] data.shape
(105, 17)

[9] data.describe()

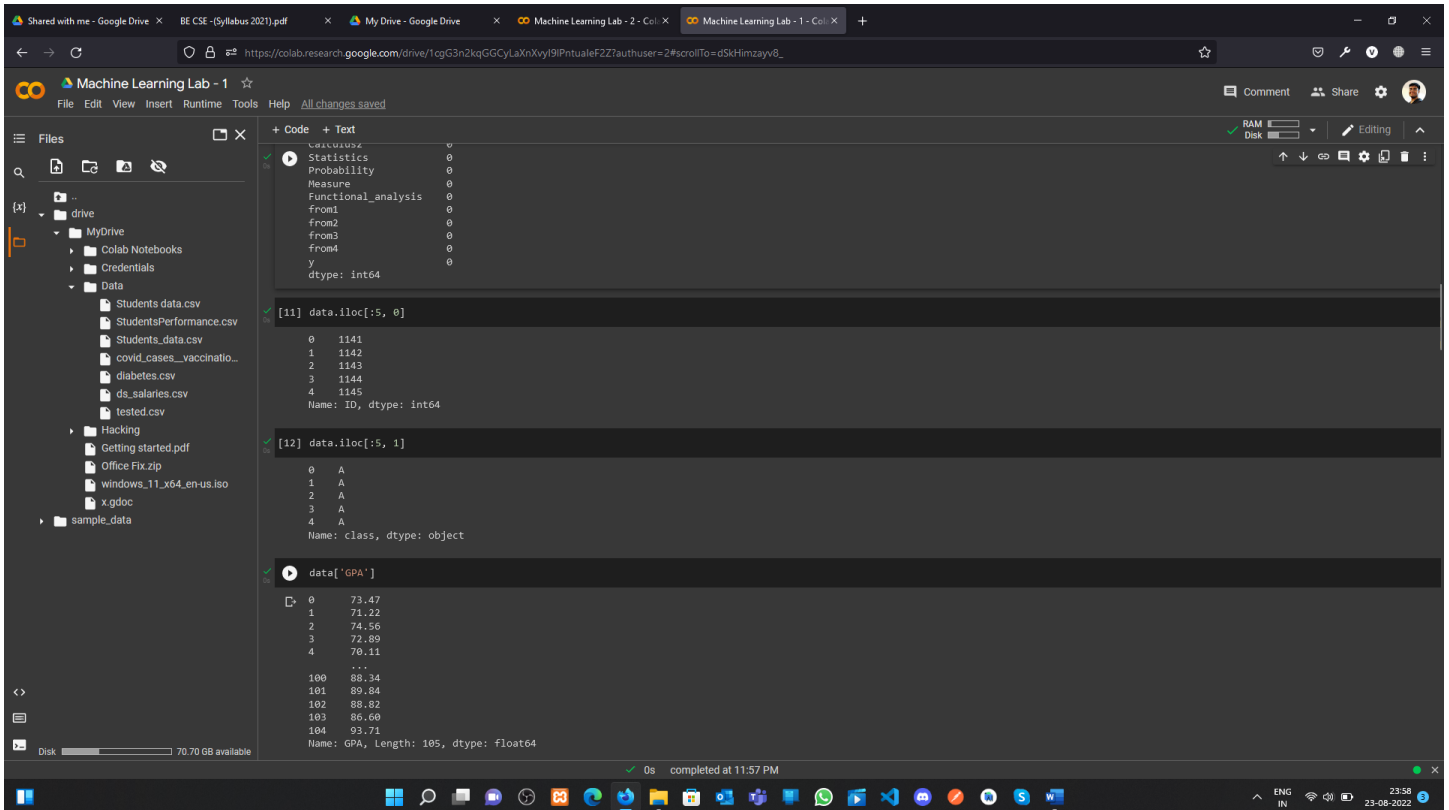
```

	ID	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from4	y
count	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000	105.000000
mean	1193.000000	1.790476	82.957048	76.057143	71.961905	78.942857	85.133333	83.876190	80.761905	75.323810	0.504762	0.714286
std	30.454885	1.673867	6.053187	11.722618	12.197039	14.997326	10.269509	10.514363	10.296119	13.003324	0.889203	0.826742
min	1141.000000	1.000000	63.490000	46.000000	38.000000	17.000000	51.000000	29.000000	54.000000	9.000000	0.000000	0.000000
25%	1167.000000	1.000000	79.340000	67.000000	64.000000	71.000000	80.000000	79.000000	74.000000	67.000000	0.000000	0.000000
50%	1193.000000	1.000000	84.110000	76.000000	73.000000	83.000000	87.000000	85.000000	81.000000	76.000000	0.000000	0.000000
75%	1219.000000	1.000000	87.300000	84.000000	80.000000	91.000000	92.000000	92.000000	89.000000	85.000000	0.000000	1.000000
max	1245.000000	7.000000	93.710000	98.000000	100.000000	99.000000	100.000000	97.000000	100.000000	100.000000	3.000000	2.000000

```

[10] data.isnull().sum()
ID          0
class       0
gender      0
race        0
GPA         0
Algebra     0
Calculus1   0
Calculus2   0
Statistics   0
Probability  0
Measure     0
Functional_analysis  0
from1       0
from2       0
from3       0

```



Machine Learning Lab - 1

Files: drive, MyDrive, Colab Notebooks, Credentials, Data, Hacking, Getting started.pdf, Office Fix.zip, windows_11_x64_en-us.iso, x.gdoc, sample_data

```

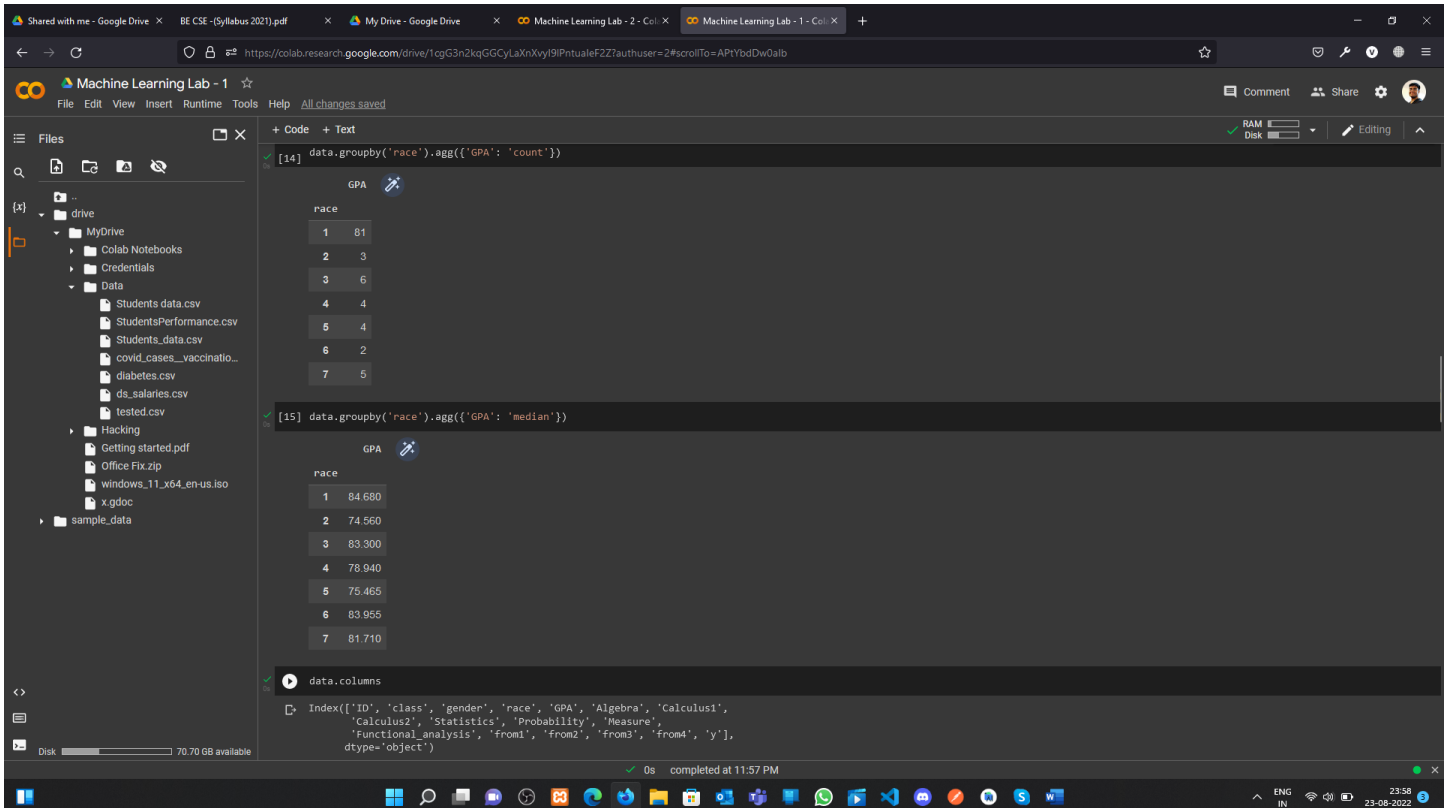
+ Code + Text
1. Calculus1 0
2. Statistics 0
3. Probability 0
4. Measure 0
5. Functional_analysis 0
6. from1 0
7. from2 0
8. from3 0
9. from4 0
10. y 0
dtype: int64

[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

Files: drive, MyDrive, Colab Notebooks, Credentials, Data, Hacking, Getting started.pdf, Office Fix.zip, windows_11_x64_en-us.iso, x.gdoc, sample_data

```

+ Code + Text
[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

Shared with me - Google Drive × BE CSE - (Syllabus 2021).pdf × My Drive - Google Drive × Machine Learning Lab - 2 - Col × Machine Learning Lab - 1 - Col × +

https://colab.research.google.com/drive/1cgG3n2kqGGCyLaXnXy9lPntualeF2Z2?authuser=2#scrollTo=Spk3_Tktwy8A

Machine Learning Lab - 1 ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Comment Share Settings

Enable browser notifications in Settings to get alerts when executions complete

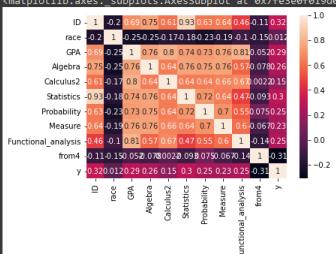
OK No thanks

```
[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

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



```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe3e0f0d0>
```

sb.pairplot(new_data)

0s completed at 12:00 AM

Shared with me - Google Drive × BE CSE - (Syllabus 2021).pdf × My Drive - Google Drive × Machine Learning Lab - 2 - Col × Machine Learning Lab - 1 - Col × +

https://colab.research.google.com/drive/1cgG3n2kqGGCyLaXnXy9lPntualeF2Z2?authuser=2#scrollTo=EpdI88UL1G-z

Machine Learning Lab - 1 ☆

File Edit View Insert Runtime Tools Help All changes saved

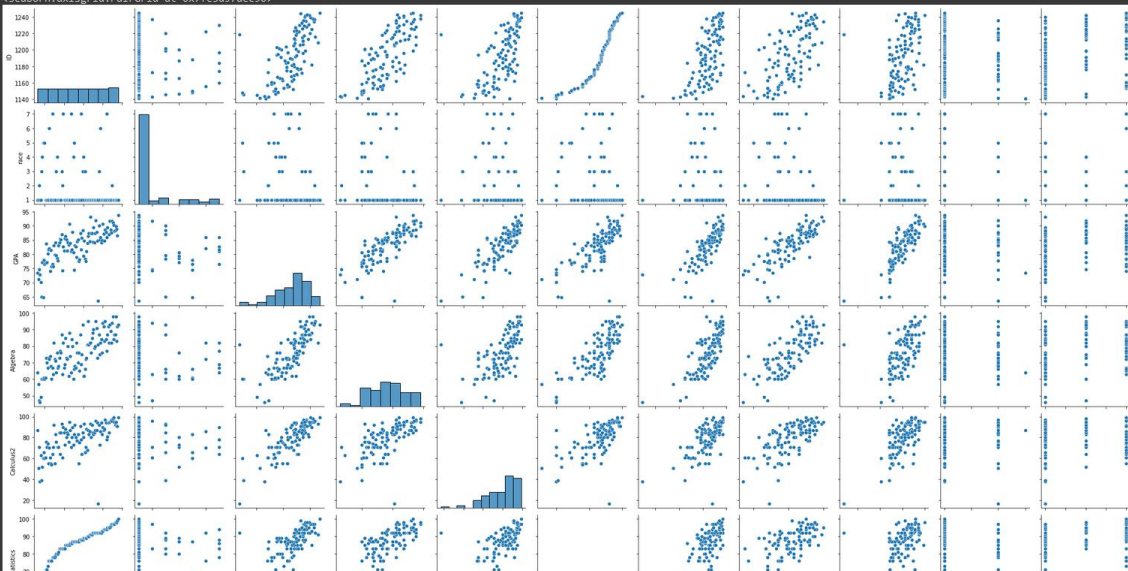
+ Code + Text

Comment Share Settings

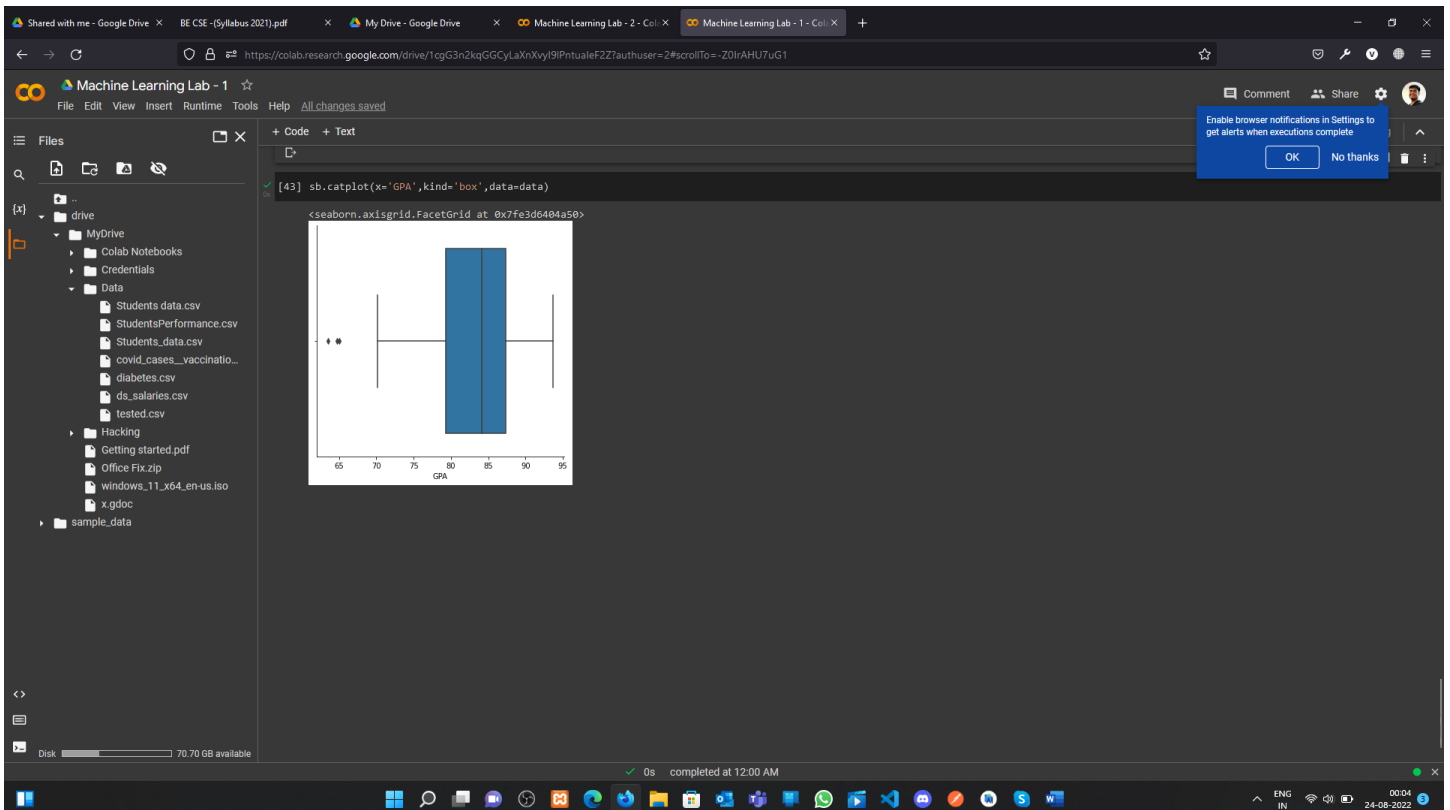
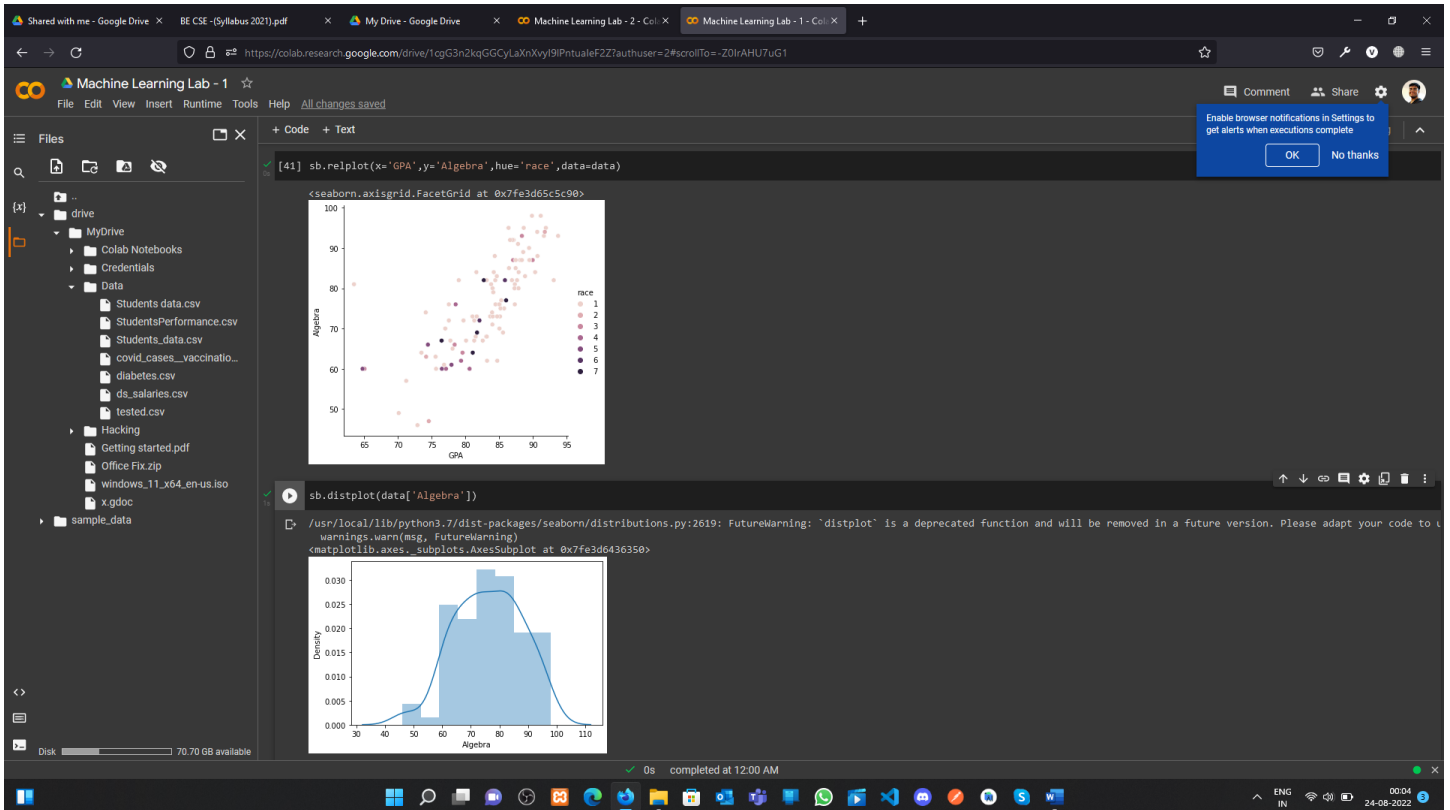
Enable browser notifications in Settings to get alerts when executions complete

OK No thanks

```
sb.pairplot(new_data)
```



0s completed at 12:00 AM



Learning outcomes (What I have learnt):

1. Data manipulation using pandas 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.			