PROJECT DEVELOPMENT PHASE

SPRINT –1

DOWNLOAD AND ANALYZE THE DATASET:

IMPORT LIBRARIES

```
import pandas as pd
import numpy as np
```

LOAD THE DATASET

```
df=pd.read_csv("Admission_Predict (1).csv")
df
df.shape
df.info()
```

PERFORM VISUALIZATION

i)UNIVARIATE ANALYSIS

```
sns.distplot(df.CGPA)
sns.histplot(df['GRE Score'])
sns.histplot(df.SOP)
```

ii)BI-VARIATE ANALYSIS

```
sns.scatterplot(df['SOP'],df['LOR '])
sns.scatterplot(df['CGPA'],df['SOP'])
```

iii)MULTI-VARIATE ANALYSIS

```
sns.pairplot(df)
sns.heatmap(df.corr(), annot=True)
```

PERFORM DESCRIPTIVE STATISTICS OF THE DATASET

df.describe()

df.Research.value_counts()

HANDLE THE MISSING VALUES

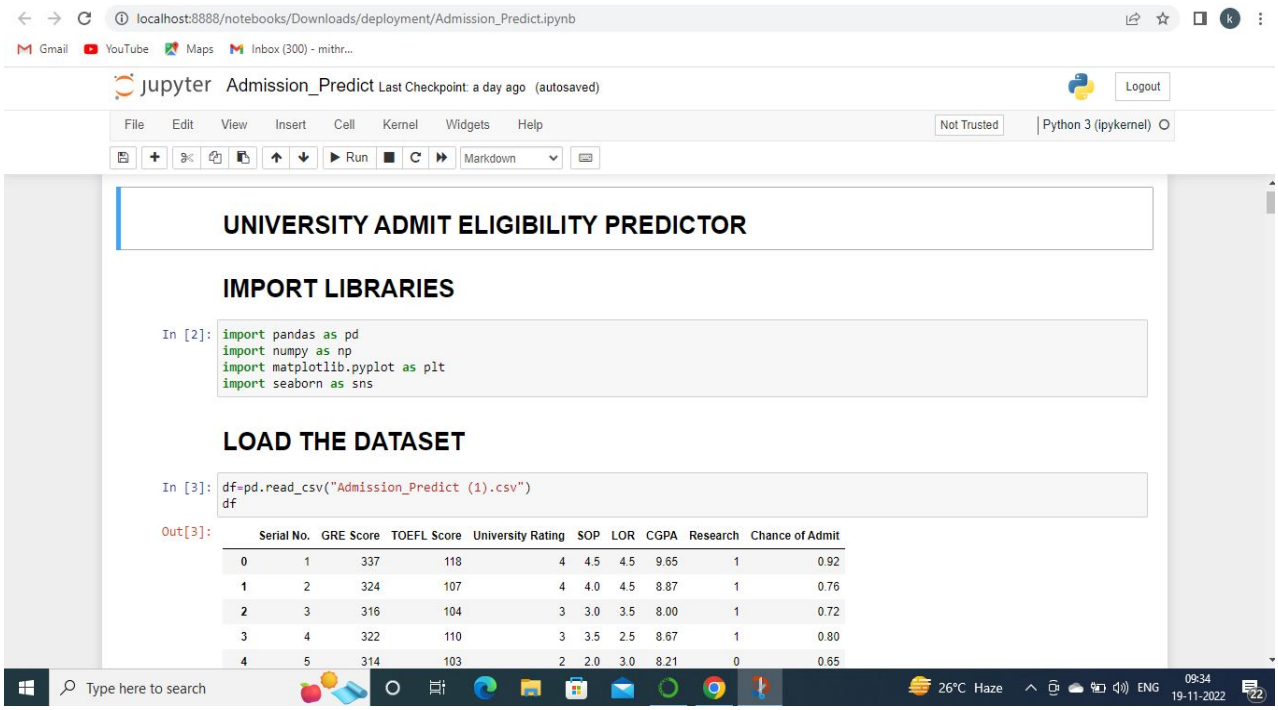
df.isnull().any()

df.isnull().sum()

OUTLIERS

sns.boxplot(df['GRE Score'])

df.shape



jupyter Admission_Predict Last Checkpoint: 2 minutes ago (autosaved)

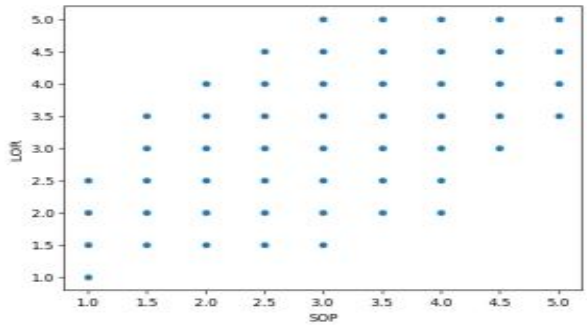
File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run Code

In [9]:
```
sns.scatterplot(df['SOP'],df['LOR '])
```

```
C:\Users\mithr\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword ar
gs: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

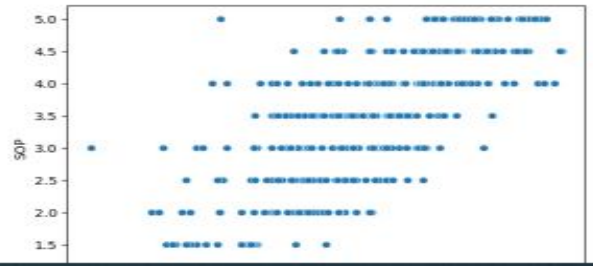
Out[9]: <AxesSubplot:xlabel='SOP', ylabel='LOR '>



In [10]:
```
sns.scatterplot(df['CGPA'],df['SOP'])
```

```
C:\Users\mithr\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword ar
gs: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit
keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[10]: <AxesSubplot:xlabel='CGPA', ylabel='SOP'>



jupyter Admission_Predict Last Checkpoint: 4 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

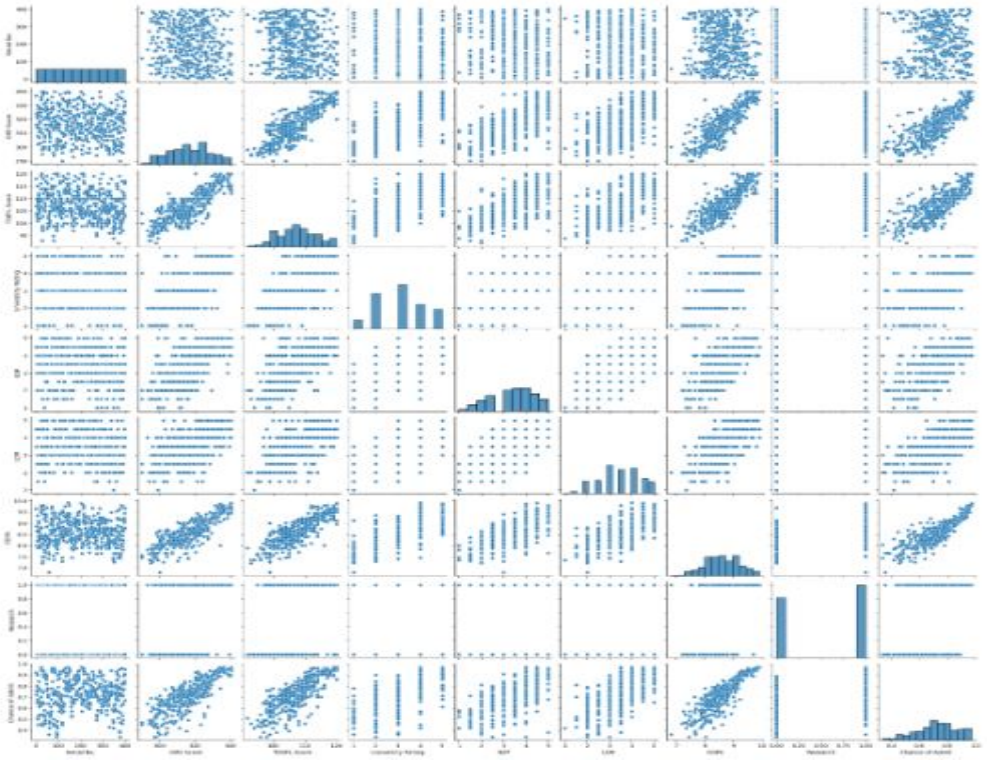
Not Trusted | Python 3 (ipykernel)

Run Code

MULTI-VARIATE ANALYSIS

In [11]:
```
sns.pairplot(df)
```

Out[11]:
```
<seaborn.axisgrid.PairGrid at 0x2e3a17849a0>
```



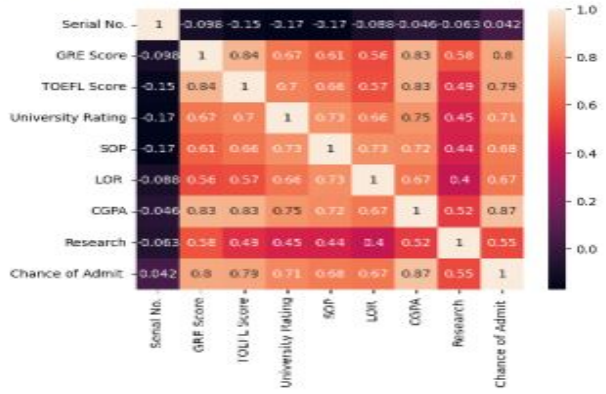
jupyter Admission_Predict Last Checkpoint: 5 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

```
In [12]: sns.heatmap(df.corr(), annot=True)
```

```
Out[12]: <AxesSubplot:>
```




```
In [13]: df.Research.value_counts()
```

```
Out[13]: 1    219
         0    181
         Name: Research, dtype: int64
```

PERFORM DESCRIPTIVE STATISTICS ON THE DATASET

```
In [14]: df.describe()
```

```
Out[14]:
```

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
count	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000	400.000000
mean	200.500000	316.807500	107.410000	3.087500	3.400000	3.452500	8.598925	0.547500	0.724350
std	115.614301	11.473646	6.069514	1.143728	1.006869	0.898478	0.596317	0.498362	0.142609
min	1.000000	290.000000	92.000000	1.000000	1.000000	1.000000	6.800000	0.000000	0.340000
25%	100.750000	308.000000	103.000000	2.000000	2.500000	3.000000	8.170000	0.000000	0.640000
50%	200.500000	317.000000	107.000000	3.000000	3.500000	3.500000	8.610000	1.000000	0.730000
75%	300.250000	325.000000	112.000000	4.000000	4.000000	4.000000	9.062500	1.000000	0.830000
max	400.000000	340.000000	120.000000	5.000000	5.000000	5.000000	9.920000	1.000000	0.970000

HANDLE THE MISSING VALUES

jupyter Admission_Predict Last Checkpoint: 6 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

HANDLE THE MISSING VALUES

In [15]:
```
df.isnull().any()
```

Out[15]:
```
Serial No.           False
GRE Score            False
TOEFL Score          False
University Rating    False
SOP                  False
LOR                  False
CGPA                 False
Research             False
Chance of Admit      False
dtype: bool
```

In [16]:
```
df.isnull().sum()
```

Out[16]:
```
Serial No.           0
GRE Score            0
TOEFL Score          0
University Rating    0
SOP                  0
LOR                  0
CGPA                 0
Research             0
Chance of Admit      0
dtype: int64
```

OUTLIERS

In [17]:
```
sns.boxplot(df['GRE Score'])
df.shape
```

```
C:\Users\mithr\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword a
rg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit ke
yword will result in an error or misinterpretation.
  warnings.warn(
```

Out[17]:
```
(400, 9)
```