Importing Libraries

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

Loading the Data set

In [2]:
```

```

Out[2]:

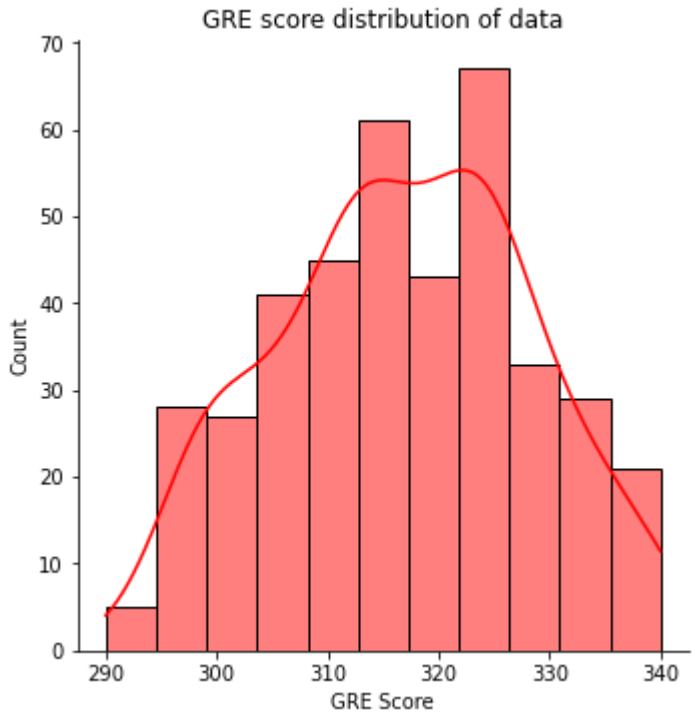
	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...	...
395	396	324	110	3	3.5	3.5	9.04	1	0.82
396	397	325	107	3	3.0	3.5	9.11	1	0.84
397	398	330	116	4	5.0	4.5	9.45	1	0.91
398	399	312	103	3	3.5	4.0	8.78	0	0.67
399	400	333	117	4	5.0	4.0	9.66	1	0.95

400 rows × 9 columns

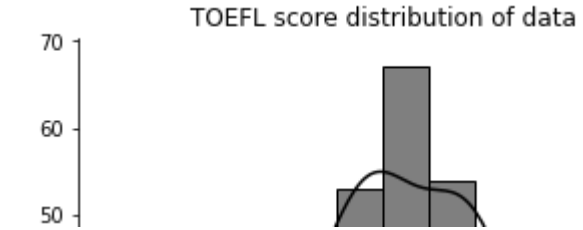
Data visualization

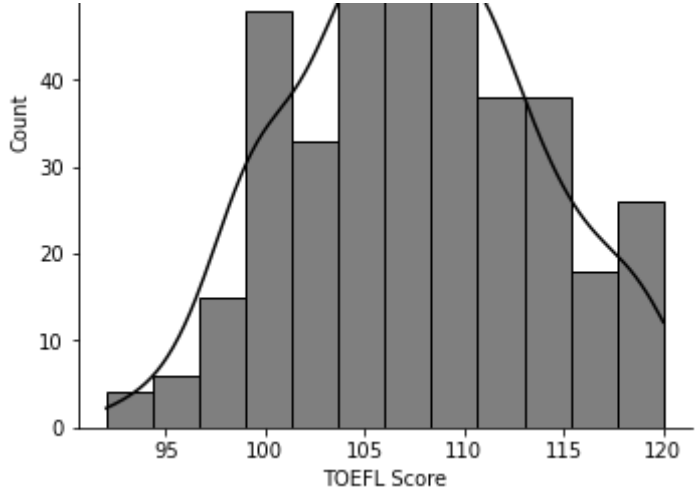
(i) Univariate Analysis

In [3]:
```
sns.displot(x=data["GRE  Score"],  kde=True,  color='Red')
plt.title("GRE score distribution of data");
```



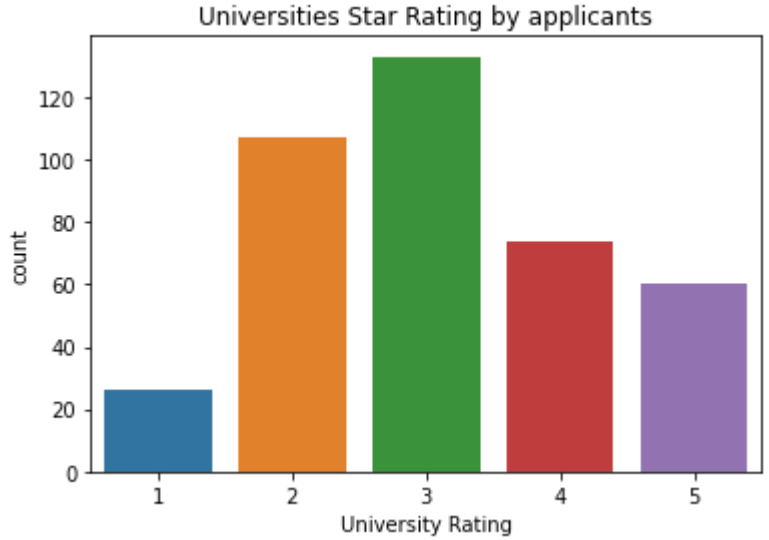
In [4]:
```
sns.displot(x=data["TOEFL  Score"],  kde=True,  color='Black')
plt.title("TOEFL score distribution of data");
```





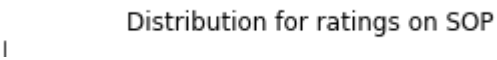
In [5]:

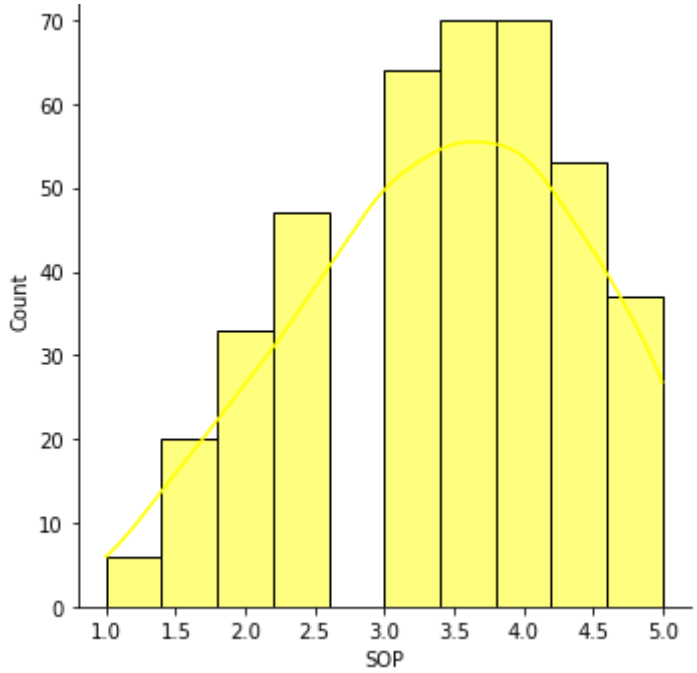
```
sns.countplot(x=data["University  Rating"]);
plt.title("Universities Star Rating by applicants");
```



In [6]:

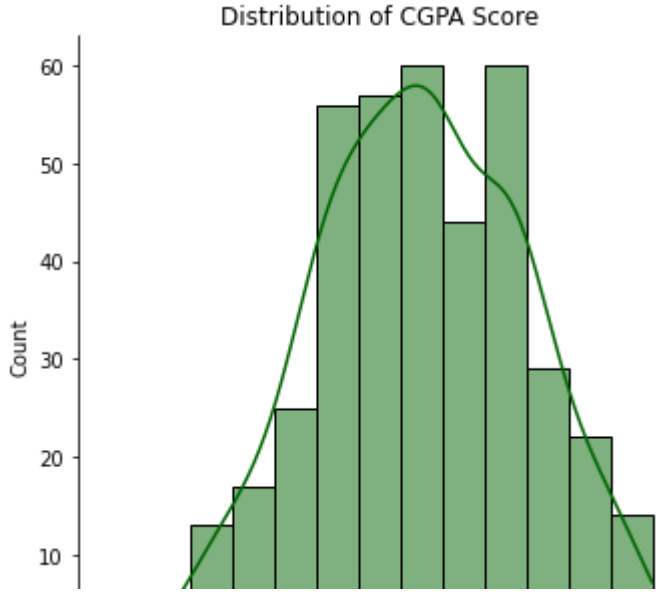
```
sns.displot(x=data["SOP"],  kde=True,  color='yellow');
plt.title("Distribution for ratings on SOP");
```

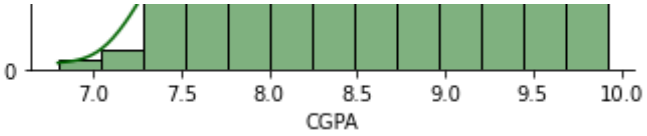




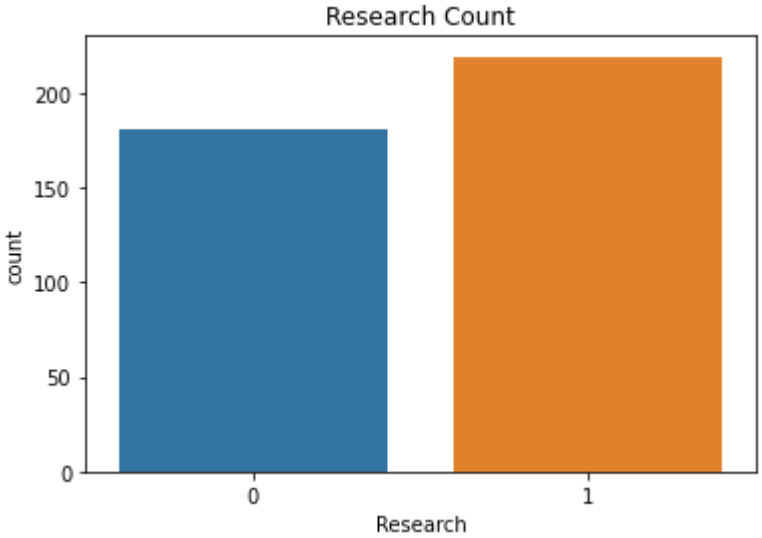
In [7]:

```
sns.displot(x=data["CGPA"], kde=True, color='Darkgreen');
plt.title("Distribution of CGPA Score");
```





In [8]:
```
sns.countplot(x=data["Research"]);
plt.title("Research Count");
```



(ii) Bivariate Analysis

In [9]:
```
cols = data.columns
features = [i for i in data.columns if i != 'Chance of Admit']
label = 'Chance of Admit'
features
```

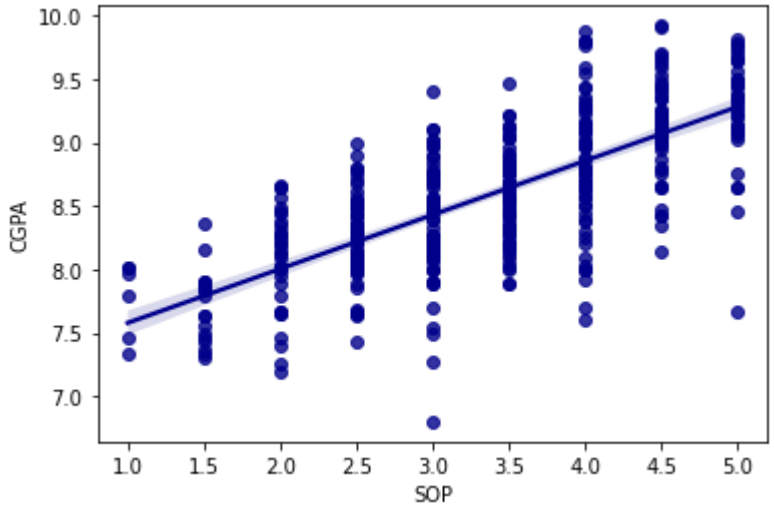
Out[9]:
```
['Serial No.',
 'GRE Score',
 'TOEFL Score',
 'University Rating',
 'SOP',
 'LOR ',
 'CGPA',
```

```
 'Research',
 'Chance of Admit ']
```

In [10]:
```
sns.regplot(data['SOP'],data['CGPA'],color='Darkblue')
```

```
C:\Users\AMMU\anaconda3\New folder\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: 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(
```

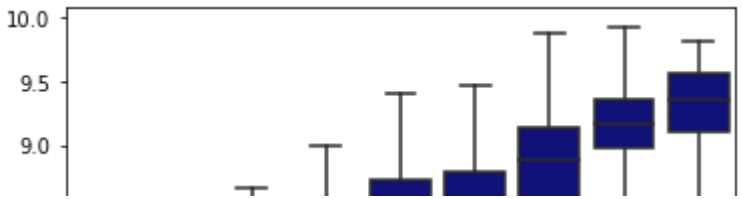
Out[10]:

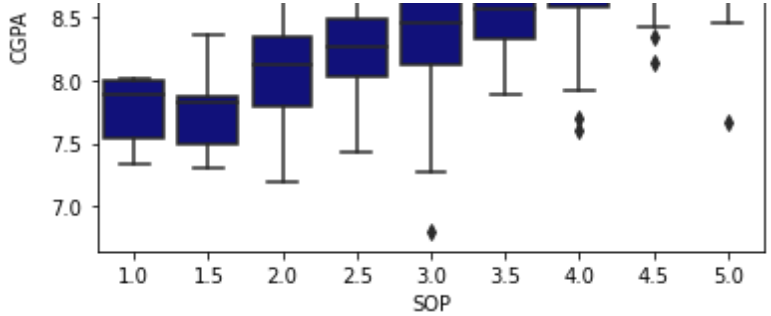


In [11]:
```
sns.boxplot(data['SOP'],data['CGPA'],color='Darkblue')
```

```
C:\Users\AMMU\anaconda3\New folder\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: 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[11]:

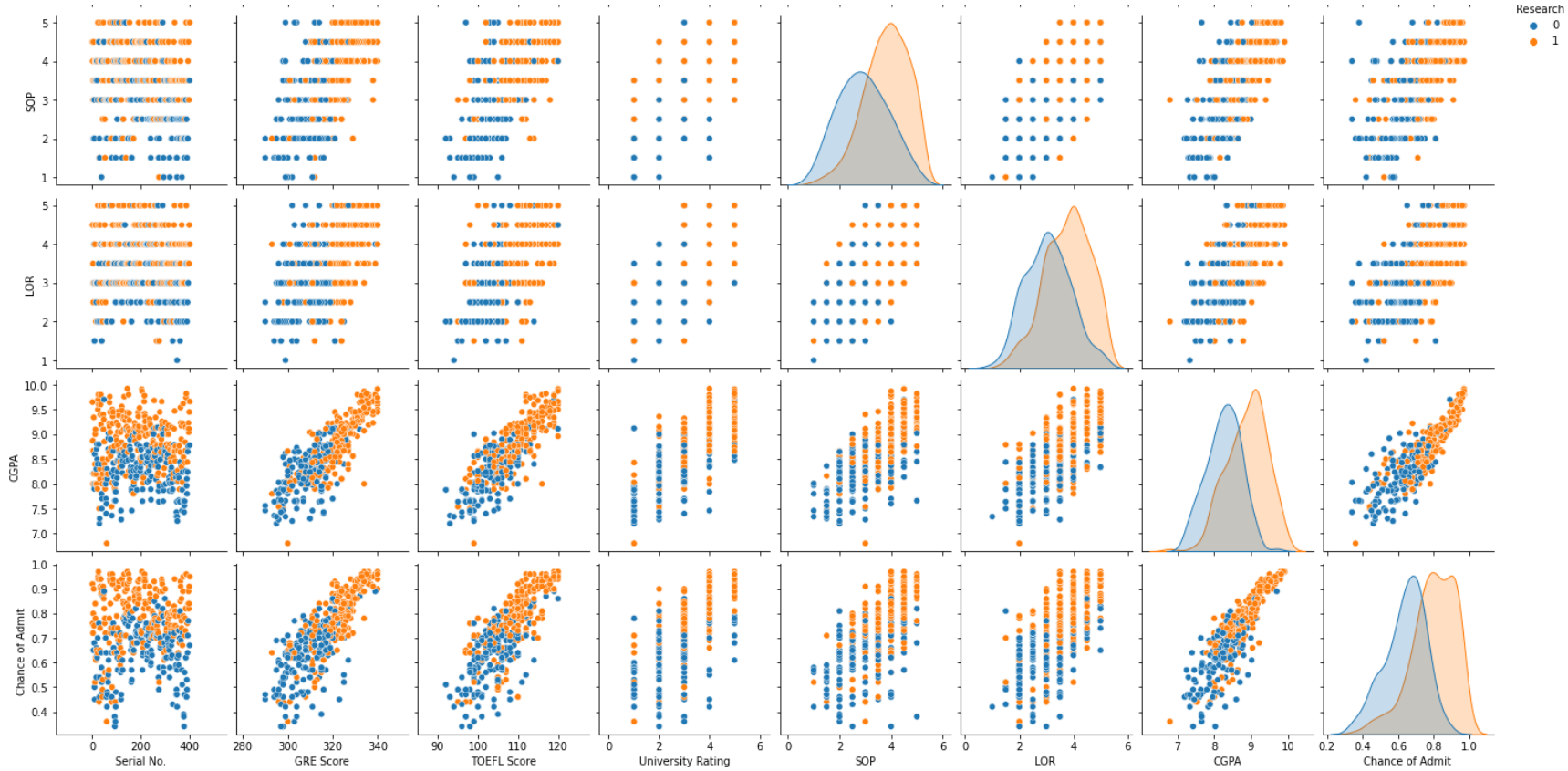




(iii) Multivariate Analysis

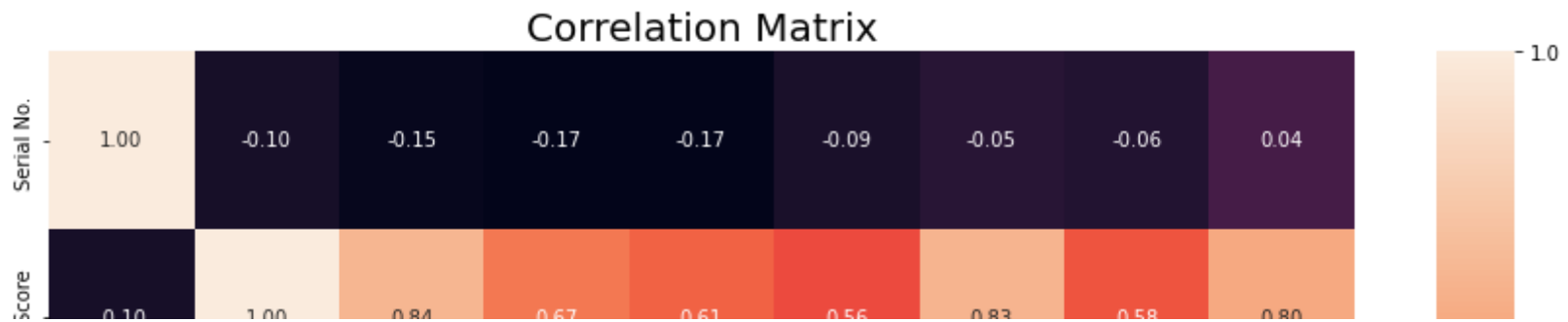
In [12]:
```
sns.pairplot(data,hue='Research')
```

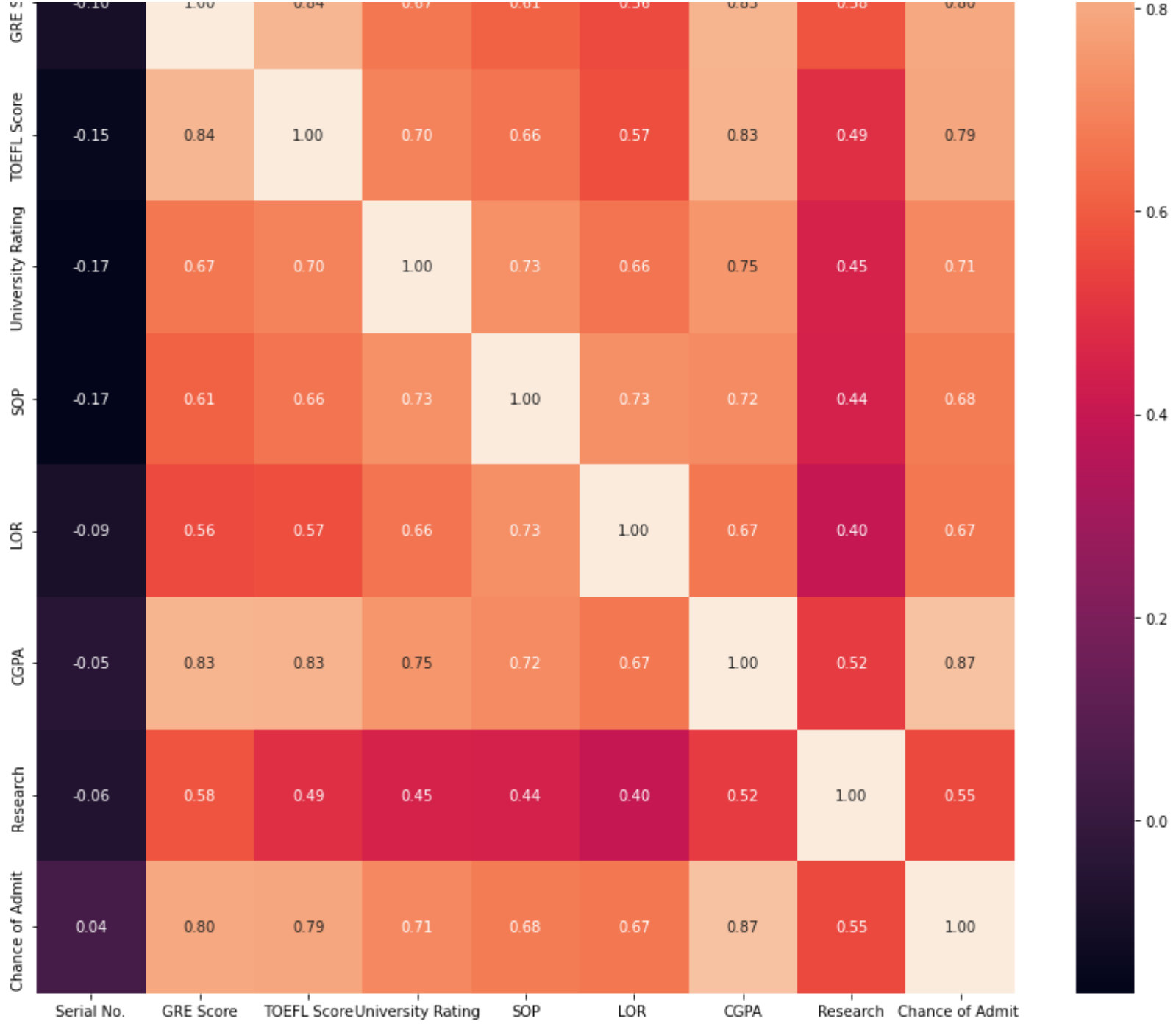
Out[12]:



In [13]:

```
corr_matrix = data.corr()
plt.figure(figsize = (15, 15))
sns.heatmap(corr_matrix,annot=True,fmt='0.2f')
plt.title("Correlation Matrix", fontsize = 20)
plt.show()
```

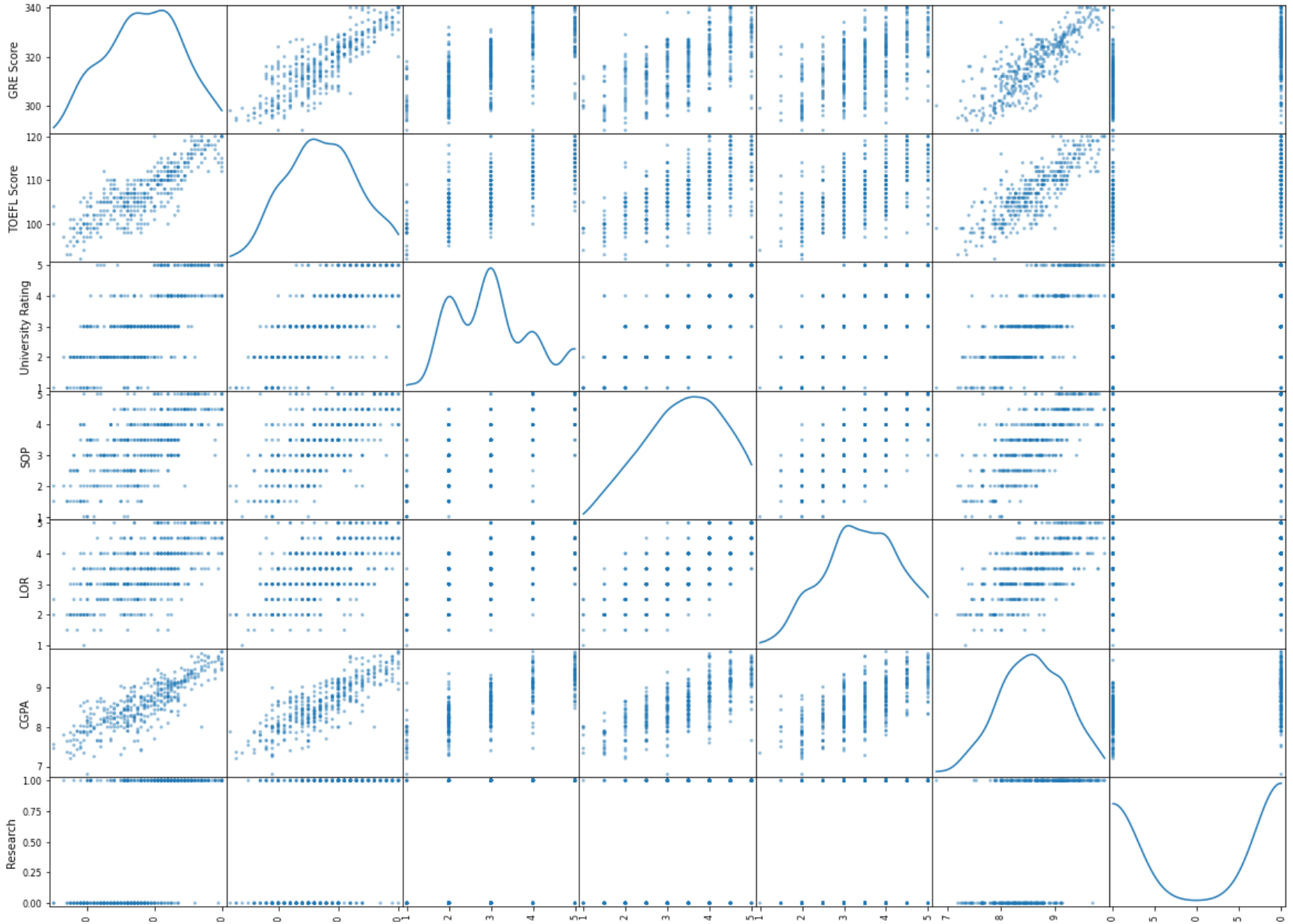



GRE S
TOEFL Score
University Rating
SOP
LOR
CGPA
Research
Chance of Admit
-0.15 0.84 1.00 0.70 0.66 0.57 0.83 0.49 0.79
-0.17 0.67 0.70 1.00 0.73 0.66 0.75 0.45 0.71
-0.17 0.61 0.66 0.73 1.00 0.73 0.72 0.44 0.68
-0.09 0.56 0.57 0.66 0.73 1.00 0.67 0.40 0.67
-0.05 0.83 0.83 0.75 0.72 0.67 1.00 0.52 0.87
-0.06 0.58 0.49 0.45 0.44 0.40 0.52 1.00 0.55
0.04 0.80 0.79 0.71 0.68 0.67 0.87 0.55 1.00
Serial No.
GRE Score
TOEFL Score
University Rating
SOP
LOR
CGPA
Research
Chance of Admit
0.8
0.6
0.4
0.2
0.0

In [14]:

```
pd.plotting.scatter_matrix(data.loc[:,"GRE Score":"Research"],diagonal="kde",figsize=(20,15))
plt.show
```

Out[14]:



GRE Score
TOEFL Score
University Rating
SOP
LOR
CGPA
Research

In [ ]: