DAV Assignment

Graduate Students Admission Analysis (Indian Perspective)

BY:- NIHAL RAMTEKE

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

In [21]:
```
Grad_add = pd.read_csv('Graduate Admissions.csv')
```

In [18]:
```
Grad_add.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 500 entries, 1 to 500
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   GRE Score          500 non-null    int64  
 1   TOEFL Score        500 non-null    int64  
 2   University Rating  500 non-null    int64  
 3   SOP                500 non-null    float64
 4   LOR                500 non-null    float64
 5   CGPA               500 non-null    float64
 6   Research           500 non-null    int64  
 7   Chance of Admit    500 non-null    float64
dtypes: float64(4), int64(4)
memory usage: 35.2 KB
```

In [20]:
```
Grad_add
```

Out[20]:

Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
1	337	118	4	4.5	4.5	9.65	1	0.92
2	324	107	4	4.0	4.5	8.87	1	0.76
3	316	104	3	3.0	3.5	8.00	1	0.72
4	322	110	3	3.5	2.5	8.67	1	0.80
5	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...
496	332	108	5	4.5	4.0	9.02	1	0.87
497	337	117	5	5.0	5.0	9.87	1	0.96
498	330	120	5	4.5	5.0	9.56	1	0.93
499	312	103	4	4.0	5.0	8.43	0	0.73
500	327	113	4	4.5	4.5	9.04	0	0.84

500 rows × 8 columns

In [16]:
```
print(Grad_add.columns.tolist())
```

```
['GRE Score', 'TOEFL Score', 'University Rating', 'SOP', 'LOR ', 'CGPA', 'Research', 'Chance of Admit ']
```

Question 1

Do GRE & TOEFL scores influence the chance of getting admitted?

In [7]:
```
Corr=Grad_add.corr()
Corr
```

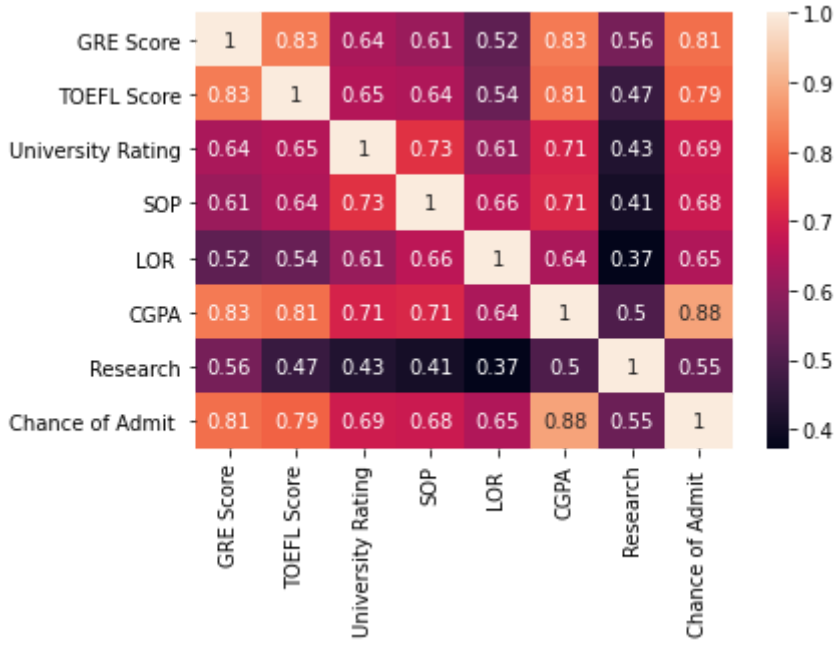
Out[7]:

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
GRE Score	1.000000	0.827200	0.635376	0.613498	0.524679	0.825878	0.563398	0.810351
TOEFL Score	0.827200	1.000000	0.649799	0.644410	0.541563	0.810574	0.467012	0.792228
University Rating	0.635376	0.649799	1.000000	0.728024	0.608651	0.705254	0.427047	0.690132
SOP	0.613498	0.644410	0.728024	1.000000	0.663707	0.712154	0.408116	0.684137
LOR	0.524679	0.541563	0.608651	0.663707	1.000000	0.637469	0.372526	0.645365
CGPA	0.825878	0.810574	0.705254	0.712154	0.637469	1.000000	0.501311	0.882413
Research	0.563398	0.467012	0.427047	0.408116	0.372526	0.501311	1.000000	0.545871
Chance of Admit	0.810351	0.792228	0.690132	0.684137	0.645365	0.882413	0.545871	1.000000

Heatmap shows how strong is the linear relationship between two parameters

In [10]:
```
sns.heatmap(Corr, annot=True)
```

Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x19f70370>



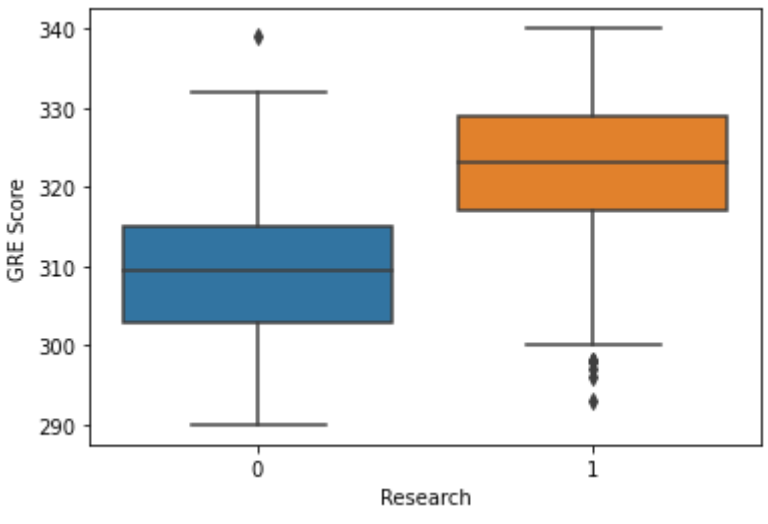
The heatmap shows us that correlation between Chance of Admit and TOEFL Score, GRE Score is very high to be precise its 0.79 and 0.81 respectively

Question 2

Does having Research experience makes you score well in GRE & TOEFL?

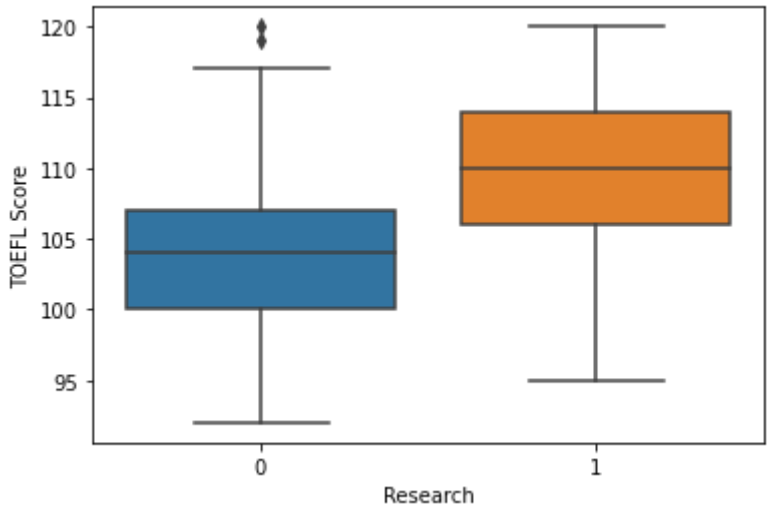
In [54]:
```
sns.boxplot(x='Research', y = 'GRE Score', data=Grad_add)
```

Out[54]: <matplotlib.axes._subplots.AxesSubplot at 0x21e8ae80>



In [57]:
```
sns.boxplot(x='Research', y = 'TOEFL Score', data=Grad_add)
```

Out[57]: <matplotlib.axes._subplots.AxesSubplot at 0x223bd988>



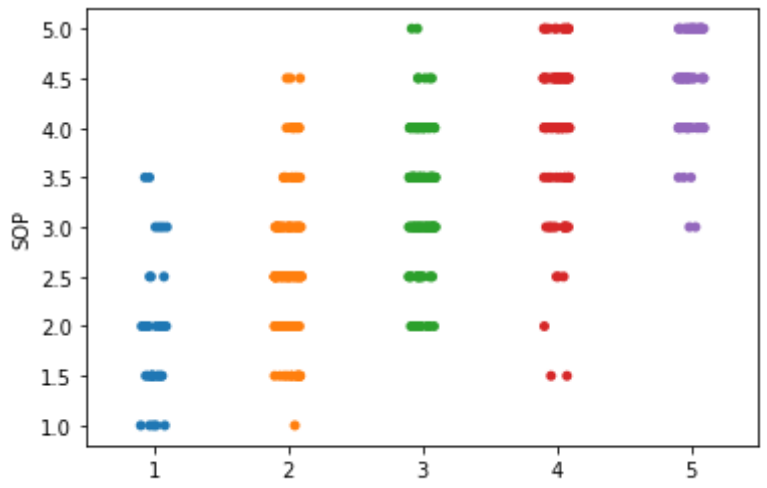
Although both of this Box plot show that students with research background have higher marks. But if we infere to heatmap we can see that the correlation between Research and TOEL Score, GRE Score is lowest among all other parameters. Therefore it will be a little ambiguous to say that students with research background have performed well.

Question 3

Does the University Rating influence my SOP and LOR Rating?

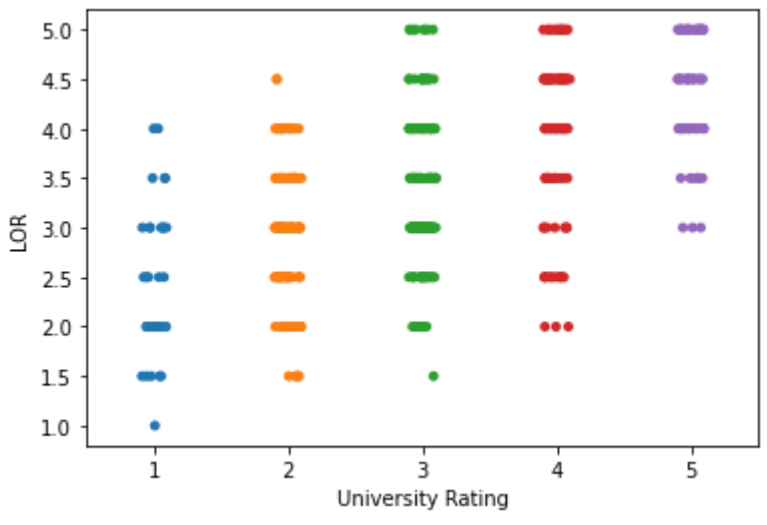
In [76]:
```
sns.stripplot(x='University Rating', y ='SOP', data = Grad_add)
```

Out[76]: <matplotlib.axes._subplots.AxesSubplot at 0x225bbcb8>



In [13]:
```
sns.stripplot(x='University Rating', y ='LOR ', data = Grad_add)
```

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1b65be08>



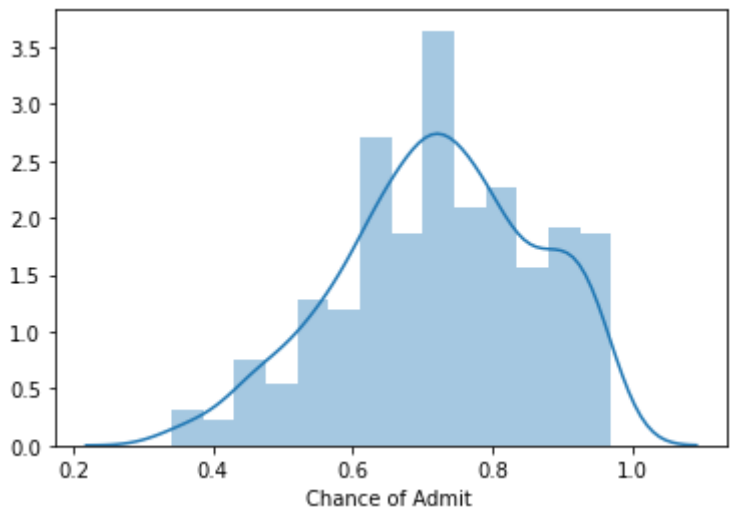
This two plots show a clear trend that students belonging to higher rated universities have higher rating for their LOR and SOP

Question 4

What is the average SOP & LOR Rating of students who got admitted?

In [21]:
```
sns.distplot(Grad_add['Chance of Admit '])
```

Out[21]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1d1bebc8>
```



Distribution plot shows that Chance of Admit is negatively skewed. That is most of the students have chance of admit greater than 60%(approx). Therefore it is a good approximation to assume that whoever has a chance of Admit greater than 0.6 will get the Admission. Therefore avg LOR and SOP rating for admitted students is the following

In [35]:
```
g=Grad_add[Grad_add['Chance of Admit ']>0.6]
g['SOP'].mean()
```

Out[35]:
```
3.606699751861042
```

In [36]:
```
g['LOR '].mean()
```

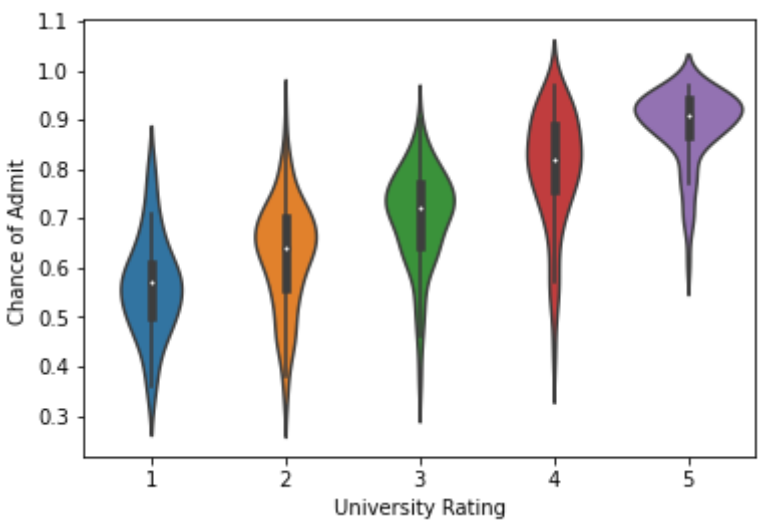
Out[36]:
```
3.6799007444168734
```

Question 5

How does the University Rating improve the chance of getting admitted?

In [39]:
```
sns.violinplot(x='University Rating', y ='Chance of Admit ', data = Grad_add)
```

Out[39]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1d0b5a18>
```



According to the plot students belonging to higher rated universities have better chance of admittance. For most of the ratings its a normal distribution but as the rating is increasing standard deviation is decreasing, that is more and more students have higher chance of admittance

Question 6

What should be your Scores for 0.9 (90%) Chance of Admission?

In [40]:
```
Grad_add[Grad_add['Chance of Admit ']==0.9]
```

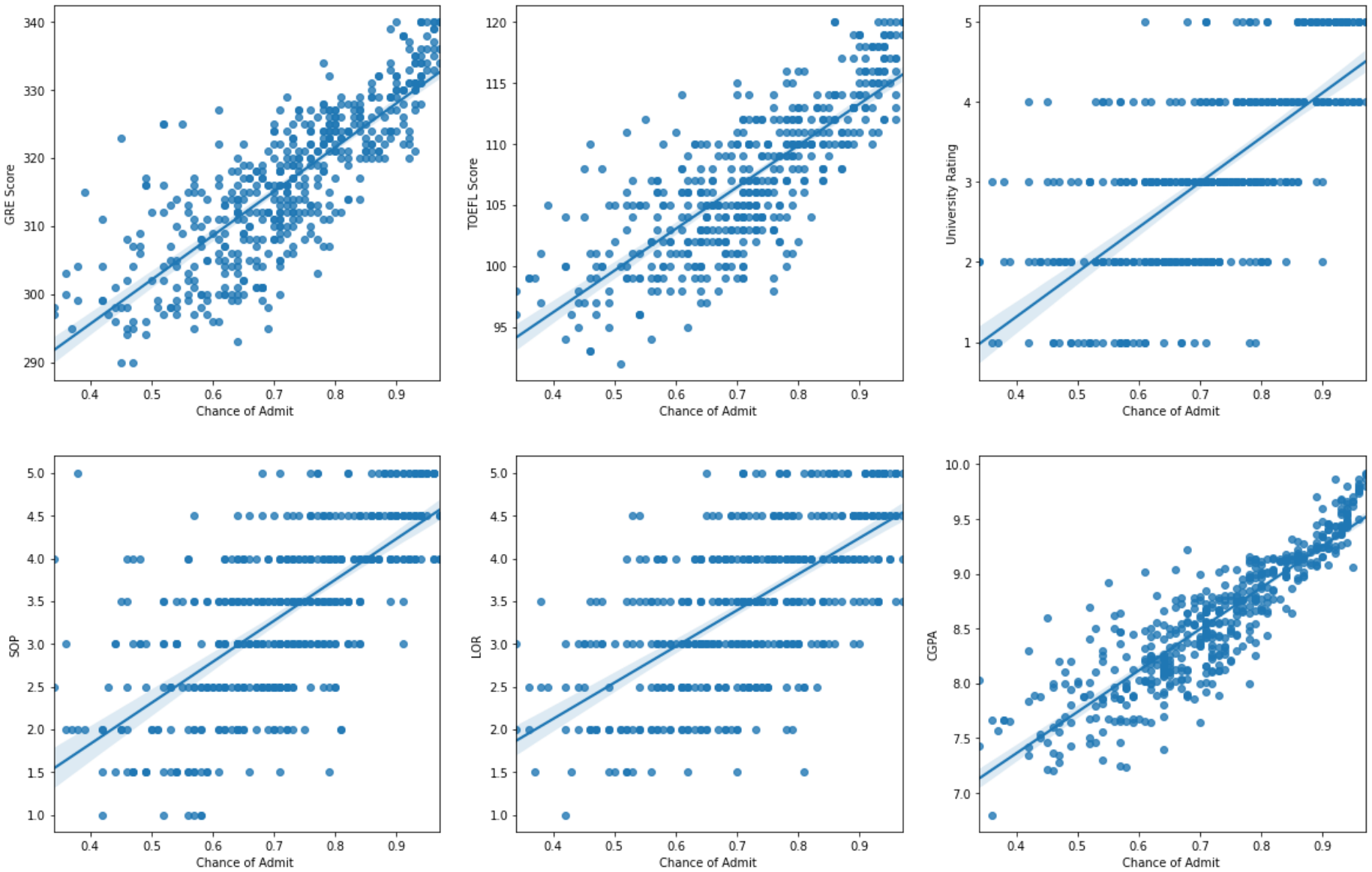
Out[40]:

Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
6	330	115	5	4.5	3.0	9.34	1	0.9
34	340	114	5	4.0	4.0	9.60	1	0.9
99	332	119	4	5.0	4.5	9.24	1	0.9
142	332	118	2	4.5	3.5	9.36	1	0.9
177	329	119	4	4.5	4.5	9.16	1	0.9
191	324	111	5	4.5	4.0	9.16	1	0.9
260	331	119	4	5.0	4.5	9.34	1	0.9
299	330	114	3	4.5	4.5	9.24	1	0.9
473	327	116	4	4.0	4.5	9.48	1	0.9

SELF QUESTIONS

In [28]:
```
plt.figure(figsize = (20, 20))
plot = 1

for col in Grad_add.columns:
    if col != 'Chance of Admit ':
        plt.subplot(3,3, plot)
        sns.regplot(
            x = Grad_add['Chance of Admit '],
            y = Grad_add[col]
        )
        plot += 1
```



Question 7

Which among the present parameters has most influence on the chance of admittance?

It has to be CGPA. It is evident from two plots:-

1. Heatmap --- the correlation factor between Chance of Admit and CGPA is 0.88 which is highest among all the parameters. It can also be proved from the regplot
2. regplot between CGPA and Chance of Admit--- shows how the datapoints fit very conveniently with linear plot.

Question 8

Is there any peculiar relation between Research and LOR parameter?

In [44]:
```
Corr[Corr['Research']==Corr['Research'].min()].drop('Chance of Admit ', axis=1)
```

Out[44]:

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research
LOR	0.524679	0.541563	0.608651	0.663707	1.0	0.637469	0.372526

We can see that Research and LOR have the weakest correlation factor. Which is kind of weird. Because mostly students take LOR from the proffessors under whom they have done their research. Therefore the LOR will always reflect the kind of work student has done in his research.

Question 9

Is it fair to say students with research background have performed better in everyway?

In [60]:
```
Grad_add.groupby('Research').mean()
```

Out[60]:

Research	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Chance of Admit
0	309.300000	103.990909	2.563636	2.918182	3.095455	8.234727	0.634909
1	322.107143	109.707143	3.546429	3.732143	3.789286	8.844929	0.789964

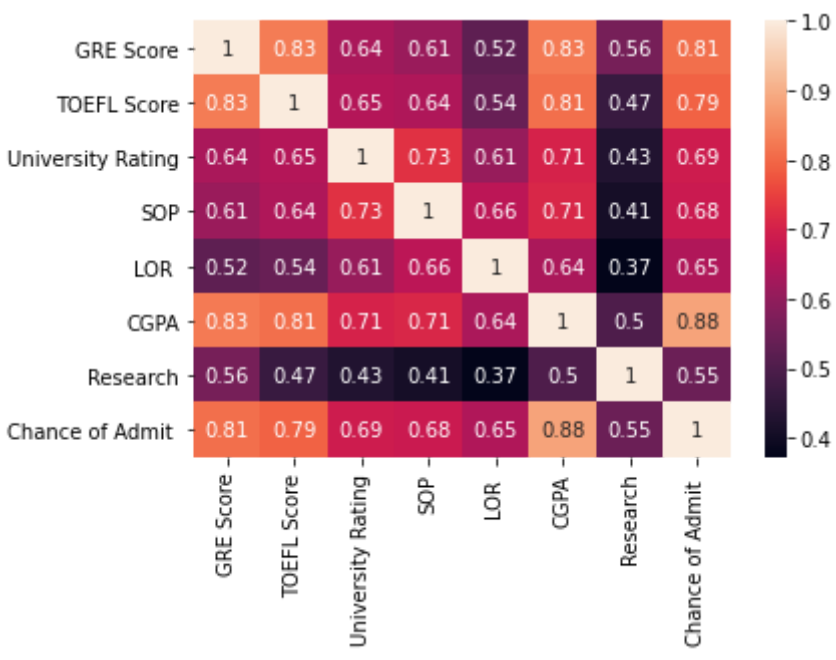
On an average students with research background have better numbers. Therefore we can say that students with research background have performed better.

Question 10

If answer to previous question is yes then, Why does Chance of Admit has weakest corelation with Research?

In [8]:
```
sns.heatmap(Corr, annot=True)
```

Out[8]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a6d5400>
```



Its because of the binary nature of the Research data. Because of which it higly deviates from the regession fit, which can ' aslo be seen in the 'Research' vs 'Chance of Admit ' regplot. The Research parameter shows fairly strong correlation with all other parameters and thats the reason why we have seen(9th question) that students with research background have scored well. Therefore its logical to expect that Research will have a strong correlation with Chance of Admit but our analysis contradicts our expectations.

Conclusion

Analysis of Graduate Admission Dataset has shown that, in Indian context the exam score are given the top most priority during admission process. It was highly unexpected of me that the research background would have such a low influence on students chances of admission. Now this could be due to data type of research parameter that we have talked about or its a flaw in our admission system. Apart from that the dataset is quite consistent.