Рубежный контроль № 1
Мельников К.И. ИУ5-24М

1. Разведочный анализ и подготовка данных

2. Вариант № 2

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sb
```

In [51]:
```
data = pd.read_csv('./Admission_Predict_Ver1.1.csv')
```

In [52]:
```
data.shape
```

Out[52]:
```
(500, 9)
```

In [59]:
```
data.dtypes
```

Out[59]:
```
Serial No.             int64
GRE Score              int64
TOEFL Score            int64
University Rating      int64
SOP                  float64
LOR                  float64
CGPA                 float64
Research               int64
Chance of Admit      float64
dtype: object
```

In [53]:
```
data.head()
```

Out[53]:

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA
0	1	337	118	4	4.5	4.5	9.65
1	2	324	107	4	4.0	4.5	8.87
2	3	316	104	3	3.0	3.5	8.00
3	4	322	110	3	3.5	2.5	8.67
4	5	314	103	2	2.0	3.0	8.21

	Research	Chance of Admit
0	1	0.92
1	1	0.76
2	1	0.72
3	1	0.80
4	0	0.65

In [54]:
```
data.columns
```

Out[54]:
```
Index(['Serial No.', 'GRE Score', 'TOEFL Score', 'University Rating', 'SO
       'LOR ', 'CGPA', 'Research', 'Chance of Admit '],
      dtype='object')
```

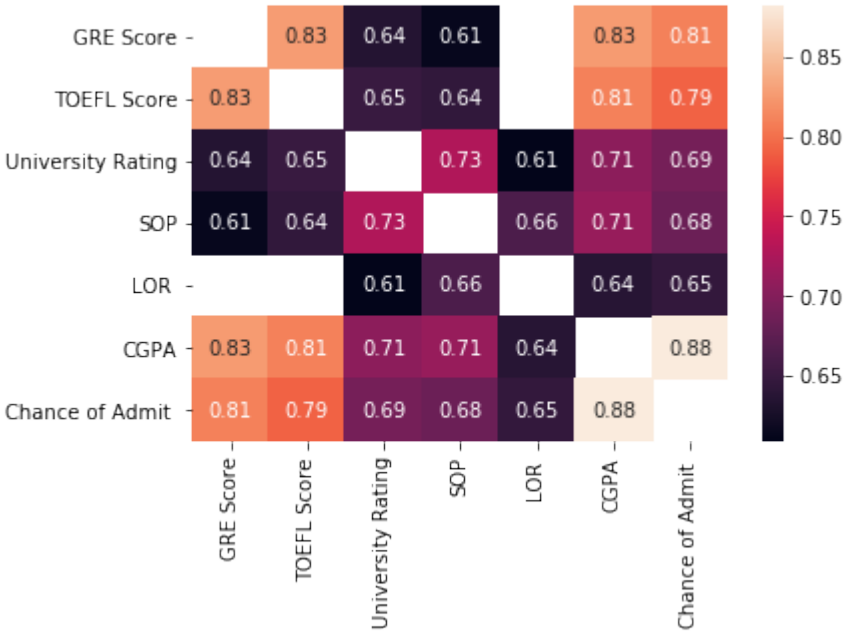
```
In [55]: for col in data.columns:
             df = data[pd.isnull(data[col])]
         df
```

```
Out[55]: Empty DataFrame
         Columns: [Serial No., GRE Score, TOEFL Score, University Rating, SOP, LOR
         Index: []
```

```
In [66]: corr = data.corr()
         corr = corr[(corr > 0.6) & (corr != 1.)].dropna(axis=1, how='all').dropna
```

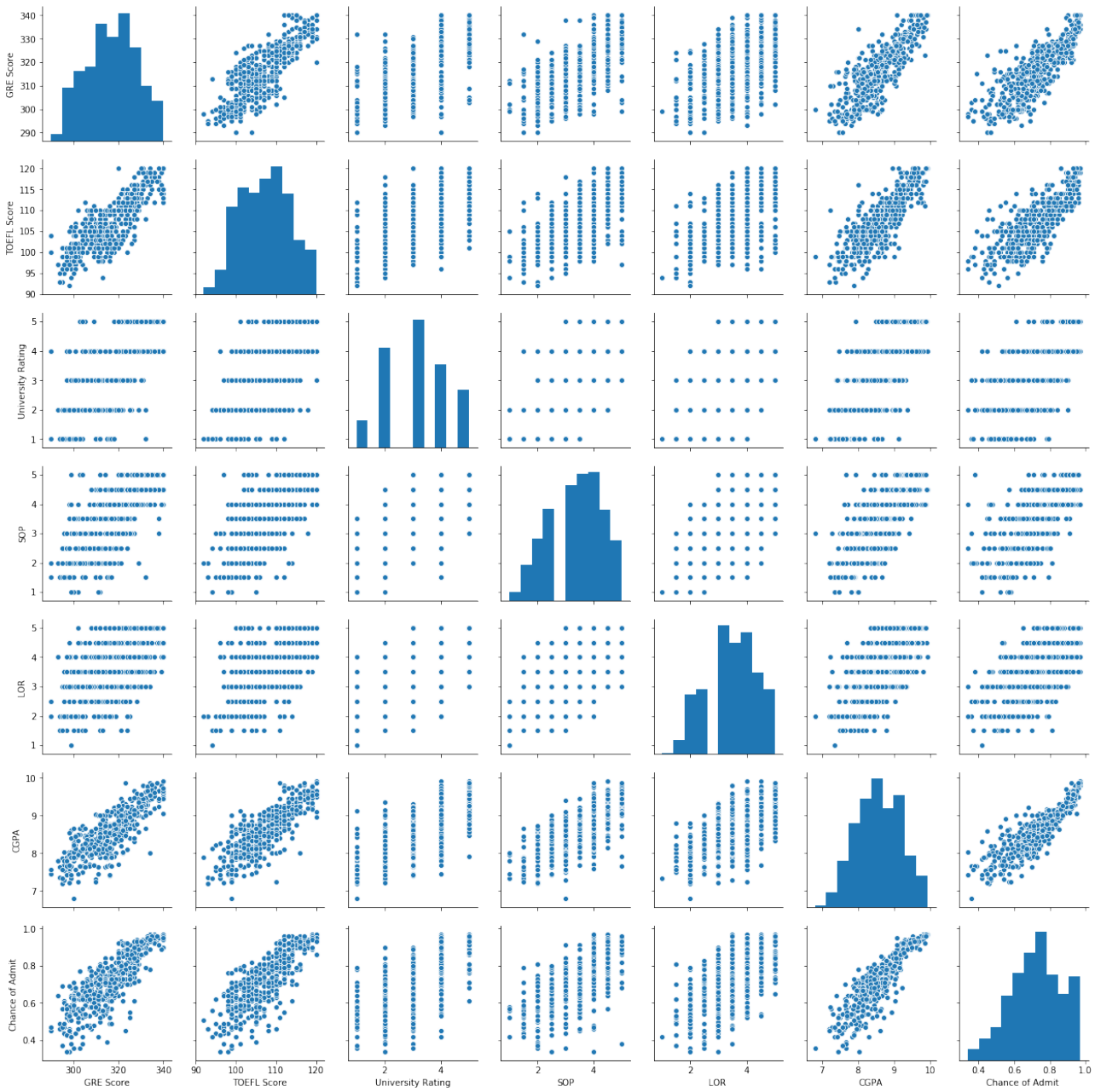
```
In [67]: sb.heatmap(corr, annot=True, fmt=".2f")
```

```
Out[67]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb2bf311940>
```




```
In [68]: sb.pairplot(data[corr.columns])
```

```
Out[68]: <seaborn.axisgrid.PairGrid at 0x7fb2c0cf3f60>
```


GRE Score
TOEFL Score
University Rating
SOP
LOR
CGPA
Chance of Admit