

## ПК №1

## Вариант 7

## Задача №1.

Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

Наборы данных: <https://www.kaggle.com/mohansacharya/graduate-admissions> (файл Admission\_Predict\_Ver1.1.csv)

Доп задание: для набора данных построить "парные диаграммы".

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(style="ticks")
```

[5]:

```
data = pd.read_csv('../input/graduate-admissions/Admission_Predict_Ver1.1.csv', sep=",")
```

[6]:

```
data.shape
```

[6]: (500, 9)

**Проверка на наличие пропусков в данных (isnull):**

```
[7]: data.dtypes
```

```
[7]: 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
```

[8]:

```
data.isnull().sum()
```

```
[8]: 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
```

```
[9]: data.head()
```

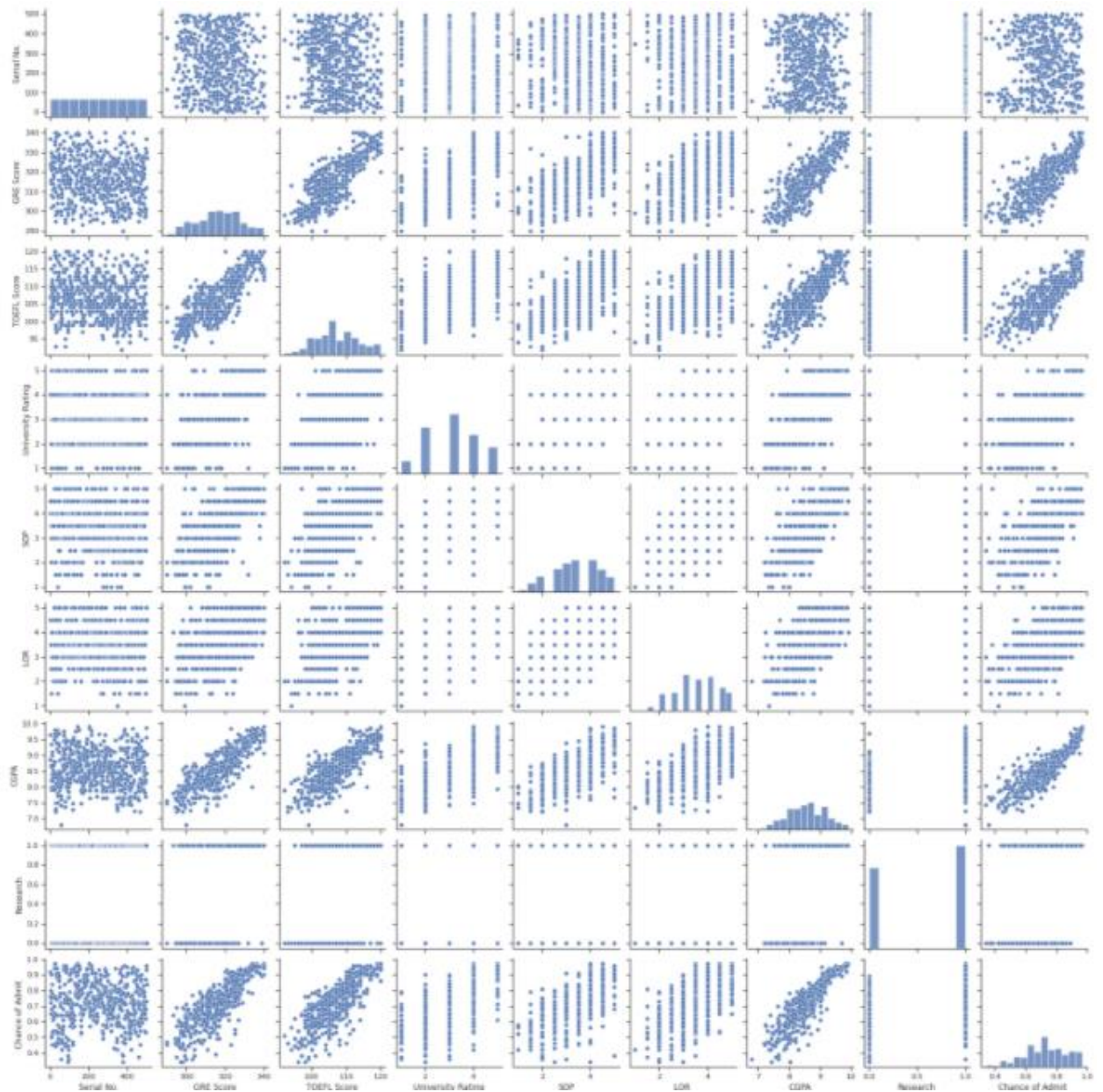
```
[9]:
```

	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

## Парные диаграммы:

```
▷ sns.pairplot(data)
```

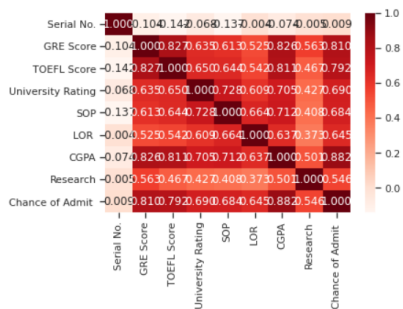
```
[10]: <seaborn.axisgrid.PairGrid at 0x7fd848197050>
```



## Корреляционный анализ:

```
[16]: sns.heatmap(data.corr(), cmap='Reds', annot=True, fmt='.3f')
```

[16]: <AxesSubplot:>



+ Code

+ Markdown

```
[14]: data.corr()
```

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
Serial No.	1.000000	-0.103839	-0.141696	-0.067641	-0.137352	-0.003694	-0.074289	-0.005332	0.008505
GRE Score	-0.103839	1.000000	0.827200	0.635376	0.613498	0.524679	0.825878	0.563398	0.810351
TOEFL Score	-0.141696	0.827200	1.000000	0.649799	0.644410	0.541563	0.810574	0.467012	0.792228
University Rating	-0.067641	0.635376	0.649799	1.000000	0.728024	0.608651	0.705254	0.427047	0.690132
SOP	-0.137352	0.613498	0.644410	0.728024	1.000000	0.663707	0.712154	0.408116	0.684137
LOR	-0.003694	0.524679	0.541563	0.608651	0.663707	1.000000	0.637469	0.372526	0.645365
CGPA	-0.074289	0.825878	0.810574	0.705254	0.712154	0.637469	1.000000	0.501311	0.882413
Research	-0.005332	0.563398	0.467012	0.427047	0.408116	0.372526	0.501311	1.000000	0.545871
Chance of Admit	0.008505	0.810351	0.792228	0.690132	0.684137	0.645365	0.882413	0.545871	1.000000

+ Code

+ Markdown

## Вывод на основе корреляционной матрицы:

Вероятнее всего не удастся построить нормальную модель МО, т.к. очень слабая корреляция между всеми параметрами