

Импорты



```
[1]: import pandas as pd
import seaborn as sns
```

Рубежный контроль №1

Обухов Антон Александрович ИУ5-65Б Вариант 13

Тема: Технологии разведочного анализа и обработки данных.

Варианты заданий.

Номер варианта - 13
Номер задачи - 2
Номер набора данных, указанного в задаче - 5

Для студентов группы ИУ5-65Б - для набора данных построить "парные диаграммы".

Задача №2.

Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака.

Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали?

Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

Выполнение:

```
[2]: df = pd.read_csv('./Admission_Predict.csv', sep = ',')
```

```
[3]: df.head()
```

```
[3]:
```

	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

```
[4]: df.dtypes
```

```
[4]: 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
```

```
[5]: df.isnull().sum()
```

```
[5]: 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
```

Так как у нет пропусков, запустим в холостую.

```
[6]: df.shape
```

```
[6]: (400, 9)
```

```
[7]: df = df.dropna(subset=['TOEFL Score', 'GRE Score'])
```

```
[8]: df.shape
```

```
[8]: (400, 9)
```

```
[9]: df.isnull().sum()
```

```
[9]: 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
```

Как мы видим, пропусков не появилось :)

Как мы видим, пропусков не появилось :)

Далее, если мы используем признаки GRE Score, TOEFL Score, University Rating, SOP, LOR, CGPA, Research построить модели, рассчитывающие шанс допуска к обучению в аспирантуре.

```
[10]: df = df.drop(columns=['Serial No.', 'GRE Score', 'TOEFL Score', 'University Rating', 'SOP', 'CGPA', 'Research'])
```

```
[11]: df.isnull().sum()
```

```
[11]: LOR          0  
Chance of Admit  0  
dtype: int64
```

```
[12]: sns.pairplot(df)
```

```
[12]: <seaborn.axisgrid.PairGrid at 0x1a7159374d0>
```

