In [2]:

```
!pip install jupyterthemes
!jt -t chesterish
```

```
Requirement already satisfied: jupyterthemes in c:\programdata\anaconda3\lib\site-packages (0.20.0)
Requirement already satisfied: notebook>=5.6.0 in c:\programdata\anaconda3\lib\site-packages (from jupyterthemes) (6.1.4)
Requirement already satisfied: ipython>=5.4.1 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from jupyterthemes) (7.22.0)
Requirement already satisfied: matplotlib>=1.4.3 in c:\programdata\anaconda3\lib\site-packages (from jupyterthemes) (3.3.2)
Requirement already satisfied: jupyter-core in c:\users\nikita\appdata\roaming\python\python38\site-packages (from jupyterthemes) (4.7.1)
Requirement already satisfied: lesscpy>=0.11.2 in c:\programdata\anaconda3\lib\site-packages (from jupyterthemes) (0.14.0)
Requirement already satisfied: ipykernel in c:\users\nikita\appdata\roaming\python\python38\site-packages (from notebook>=5.6.0->jupyterthemes) (5.5.0)
Requirement already satisfied: pyzmq>=17 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from notebook>=5.6.0->jupyterthemes) (22.0.3)
Requirement already satisfied: ipython-genutils in c:\users\nikita\appdata\roaming\python\python38\site-packages (from notebook>=5.6.0->jupyterthemes) (0.2.0)
Requirement already satisfied: prometheus-client in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (0.8.0)
Requirement already satisfied: argon2-cffi in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (20.1.0)
Requirement already satisfied: nbformat in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (5.0.8)
Requirement already satisfied: Send2Trash in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (1.5.0)
Requirement already satisfied: tornado>=5.0 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from notebook>=5.6.0->jupyterthemes) (6.1)
Requirement already satisfied: jupyter-client>=5.3.4 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from notebook>=5.6.0->jupyterthemes) (6.1.12)
Requirement already satisfied: nbconvert in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (6.0.7)
Requirement already satisfied: jinja2 in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (2.11.2)
Requirement already satisfied: traitlets>=4.2.1 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from notebook>=5.6.0->jupyterthemes) (5.0.5)
Requirement already satisfied: terminado>=0.8.3 in c:\programdata\anaconda3\lib\site-packages (from notebook>=5.6.0->jupyterthemes) (0.9.1)
Requirement already satisfied: setuptools>=18.5 in c:\programdata\anaconda3\lib\site-packages (from ipython>=5.4.1->jupyterthemes) (50.3.1.post20201107)
Requirement already satisfied: decorator in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (4.4.2)
Requirement already satisfied: colorama; sys_platform == "win32" in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (0.4.3)
Requirement already satisfied: backcall in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (0.2.0)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (3.0.18)
Requirement already satisfied: pygments in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (2.8.1)
Requirement already satisfied: jedi>=0.16 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (0.18.0)
Requirement already satisfied: pickleshare in c:\users\nikita\appdata\roaming\python\python38\site-packages (from ipython>=5.4.1->jupyterthemes) (0.7.5)
Requirement already satisfied: pillow>=6.2.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=1.4.3->jupyterthemes) (8.0.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=1.4.3->jupyterthemes) (2.4.7)
Requirement already satisfied: numpy>=1.15 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=1.4.3->jupyterthemes) (1.19.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=1.4.3->jupyterthemes) (1.3.0)
Requirement already satisfied: certifi>=2020.06.20 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=1.4.3->jupyterthemes) (2020.6.20)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from matplotlib>=1.4.3->jupyterthemes) (2.8.1)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=1.4.3->jupyterthemes) (0.10.0)
Requirement already satisfied: pywin32>=1.0; sys_platform == "win32" in c:\users\nikita\appdata\roaming\python\python38\site-packages (from jupyter-core->jupyterthemes) (300)
Requirement already satisfied: six in c:\users\nikita\appdata\roaming\python\python38\site-packages (from lesscpy>=0.11.2->jupyterthemes) (1.14.0)
Requirement already satisfied: ply in c:\programdata\anaconda3\lib\site-packages (from lesscpy>=0.11.2->jupyterthemes) (3.11)
Requirement already satisfied: cffi>=1.0.0 in c:\programdata\anaconda3\lib\site-packages (from argon2-cffi->notebook>=5.6.0->jupyterthemes) (1.14.3)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\programdata\anaconda3\lib\site-packages (from nbformat->notebook>=5.6.0->jupyterthemes) (3.2.0)
Requirement already satisfied: jupyterlab-pygments in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (0.1.2)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (1.4.3)
Requirement already satisfied: entrypoints>=0.2.2 in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (0.3)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (0.5.1)
Requirement already satisfied: testpath in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (0.4.4)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (0.8.4)
Requirement already satisfied: bleach in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (3.2.1)
Requirement already satisfied: defusedxml in c:\programdata\anaconda3\lib\site-packages (from nbconvert->notebook>=5.6.0->jupyterthemes) (0.6.0)
Requirement already satisfied: MarkupSafe>=0.23 in c:\programdata\anaconda3\lib\site-packages (from jinja2->notebook>=5.6.0->jupyterthemes) (1.1.1)
Requirement already satisfied: pywinpty>=0.5 in c:\programdata\anaconda3\lib\site-packages (from terminado>=0.8.3->notebook>=5.6.0->jupyterthemes) (0.5.7)
Requirement already satisfied: wcwidth in c:\programdata\anaconda3\lib\site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython>=5.4.1->jupyterthemes) (0.2.5)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\nikita\appdata\roaming\python\python38\site-packages (from jedi>=0.16->ipython>=5.4.1->jupyterthemes) (0.8.1)
Requirement already satisfied: pycparser in c:\programdata\anaconda3\lib\site-packages (from cffi>=1.0.0->argon2-cffi->notebook>=5.6.0->jupyterthemes) (2.20)
Requirement already satisfied: attrs>=17.4.0 in c:\programdata\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat->notebook>=5.6.0->jupyterthemes) (20.3.0)
Requirement already satisfied: pyrsistent>=0.14.0 in c:\programdata\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat->notebook>=5.6.0->jupyterthemes) (0.17.3)
Requirement already satisfied: async-generator in c:\programdata\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->notebook>=5.6.0->jupyterthemes) (1.10)
Requirement already satisfied: nest-asyncio in c:\programdata\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->notebook>=5.6.0->jupyterthemes) (1.4.2)
Requirement already satisfied: webencodings in c:\programdata\anaconda3\lib\site-packages (from bleach->nbconvert->notebook>=5.6.0->jupyterthemes) (0.5.1)
Requirement already satisfied: packaging in c:\programdata\anaconda3\lib\site-packages (from bleach->nbconvert->notebook>=5.6.0->jupyterthemes) (20.4)
```

РК 1 по ТМО

Выполнил:
Студент группы ИУ5-62Б
__Демирев Н.К.__
Преподаватель:
__Гапанюк. Ю.Е.__

Москва 2020 г.

1. Постановка задачи

Для набора данных Admission_Predict_Ver1.1.csv провести корреляционный анализ. В случае наличия пропусков в данных удалить строки или колонки, содержащие пропуски. Сделать выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

2. Обработка пропусков в данных

2.1. Импорт библиотек для анализа

Импортируем библиотеки с помощью команды import. Будем подключать все библиотеки последовательно, по мере их использования.

In [3]:

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

2.2. Импорт датасета и его анализ на пропуски

Импортируем наш датасет Admission_Predict_Ver1.1.csv:

In [4]:

```
data = pd.read_csv('dataset/Admission_Predict_Ver1.1.csv', sep = ",")
```

In [5]:

```
# Размер набора данных
print(f'Строк - {data.shape[0]}\nСтолбцов - {data.shape[1]}')
```

```
Строк - 500
Столбцов - 9
```

In [6]:

```
# типы колонок
data.dtypes
```

Out[6]:

```
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 [15]:

```
# проверим есть ли пропущенные значения
data.isnull().sum()
```

Out[15]:

```
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
```

In [16]:

```
data.head(6)
```

Out[16]:

	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
5	6	330	115	5	4.5	3.0	9.34	1	0.90

Как мы видим, наши данные не содержат пропусков, поэтому можно переходить к корреляционному анализу.

3. Корреляционный анализ данных

3.1. Корреляционная матрица

Корреляционную матрицу будем строить на основе коэффициента Спирмена

In [8]:

```
data.corr(method = "spearman")
```

Out[8]:

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
Serial No.	1.000000	-0.099592	-0.142607	-0.055424	-0.144249	0.004220	-0.075126	-0.005332	-0.001733
GRE Score	-0.099592	1.000000	0.823853	0.643423	0.620688	0.514352	0.829251	0.578487	0.822201
TOEFL Score	-0.142607	0.823853	1.000000	0.645533	0.644715	0.523434	0.809485	0.474540	0.793634
University Rating	-0.055424	0.643423	0.645533	1.000000	0.729399	0.602319	0.703333	0.435351	0.703742
SOP	-0.144249	0.620688	0.644715	0.729399	1.000000	0.662653	0.717384	0.409088	0.702799
LOR	0.004220	0.514352	0.523434	0.602319	0.662653	1.000000	0.639563	0.376166	0.643627
CGPA	-0.075126	0.829251	0.809485	0.703333	0.717384	0.639563	1.000000	0.509264	0.888786
Research	-0.005332	0.578487	0.474540	0.435351	0.409088	0.376166	0.509264	1.000000	0.565715
Chance of Admit	-0.001733	0.822201	0.793634	0.703742	0.702799	0.643627	0.888786	0.565715	1.000000

В качестве целевого признака выберем шанс поступления, так как студентам важно знать за счет чего повышается сам шанс поступления в тот или иной университет.

3.2. Визуализация корреляционной матрицы

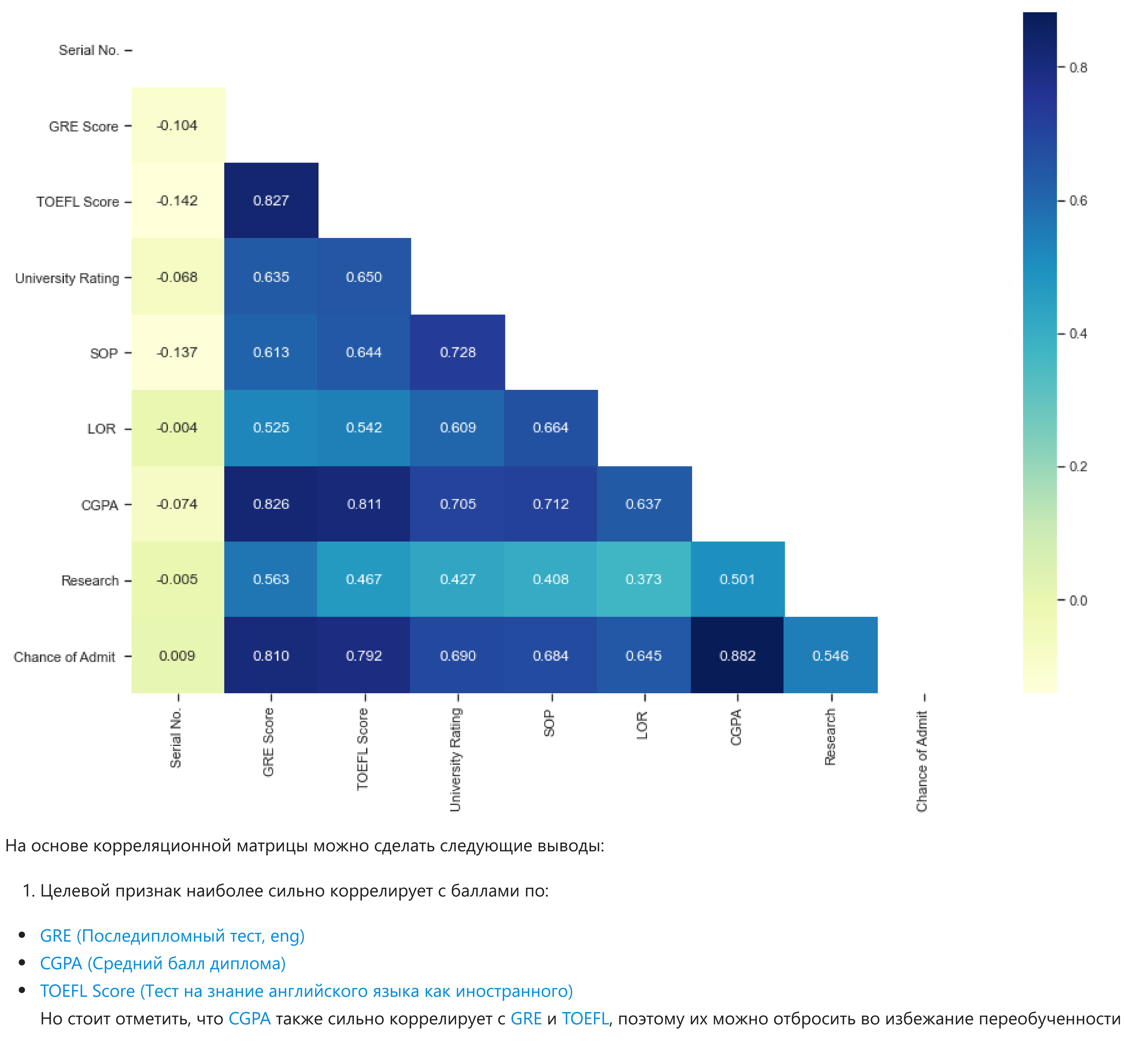
Для визуализации корреляционной матрицы будем использовать "тепловую карту" heatmap которая показывает степень корреляции различными цветами.

In [9]:

```
fig, ax = plt.subplots(figsize = (15,10))
mask = np.zeros_like(data.corr(), dtype=np.bool)
mask[np.triu_indices_from(mask)] = True
sns.heatmap(data.corr(), cmap='YlGnBu', mask=mask, annot=True, fmt='.3f')
```

Out[9]:

```
<AxesSubplot:>
```



На основе корреляционной матрицы можно сделать следующие выводы:

1. Целевой признак наиболее сильно коррелирует с баллами по:

- GRE (Последипломный тест, eng)
- CGPA (Средний балл диплома)
- TOEFL Score (Тест на знание английского языка как иностранного)
  Но стоит отметить, что CGPA также сильно коррелирует с GRE и TOEFL, поэтому их можно отбросить во избежание переобученности модели.

1. В меньшей степени целевой признак коррелирует с:

- University Rating (Рейтинг универа)
- SOP (Statement of purpose)
- LOR (Letters of Recommendation)
  Эти признаки можем оставить.

1. В наименьшей степени целевой признак коррелирует с Serial No. так что его можно откинуть. Также не имеет смысла включать в модель Research Experience так как он слабо коррелирует с целевым признаком по сравнению с 1 и 2 пунктами.

In [14]:

```
data1=data.drop(['TOEFL Score','GRE Score','Serial No.','Research'], axis=1)
data1.head(6)
```

Out[14]:

	University Rating	SOP	LOR	CGPA	Chance of Admit
0	4	4.5	4.5	9.65	0.92
1	4	4.0	4.5	8.87	0.76
2	3	3.0	3.5	8.00	0.72
3	3	3.5	2.5	8.67	0.80
4	2	2.0	3.0	8.21	0.65
5	5	4.5	3.0	9.34	0.90

In [ ]: