

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

data = pd.read_csv('Admission_Predict_Ver1.1.csv', sep=",")
data.shape

(500, 9)

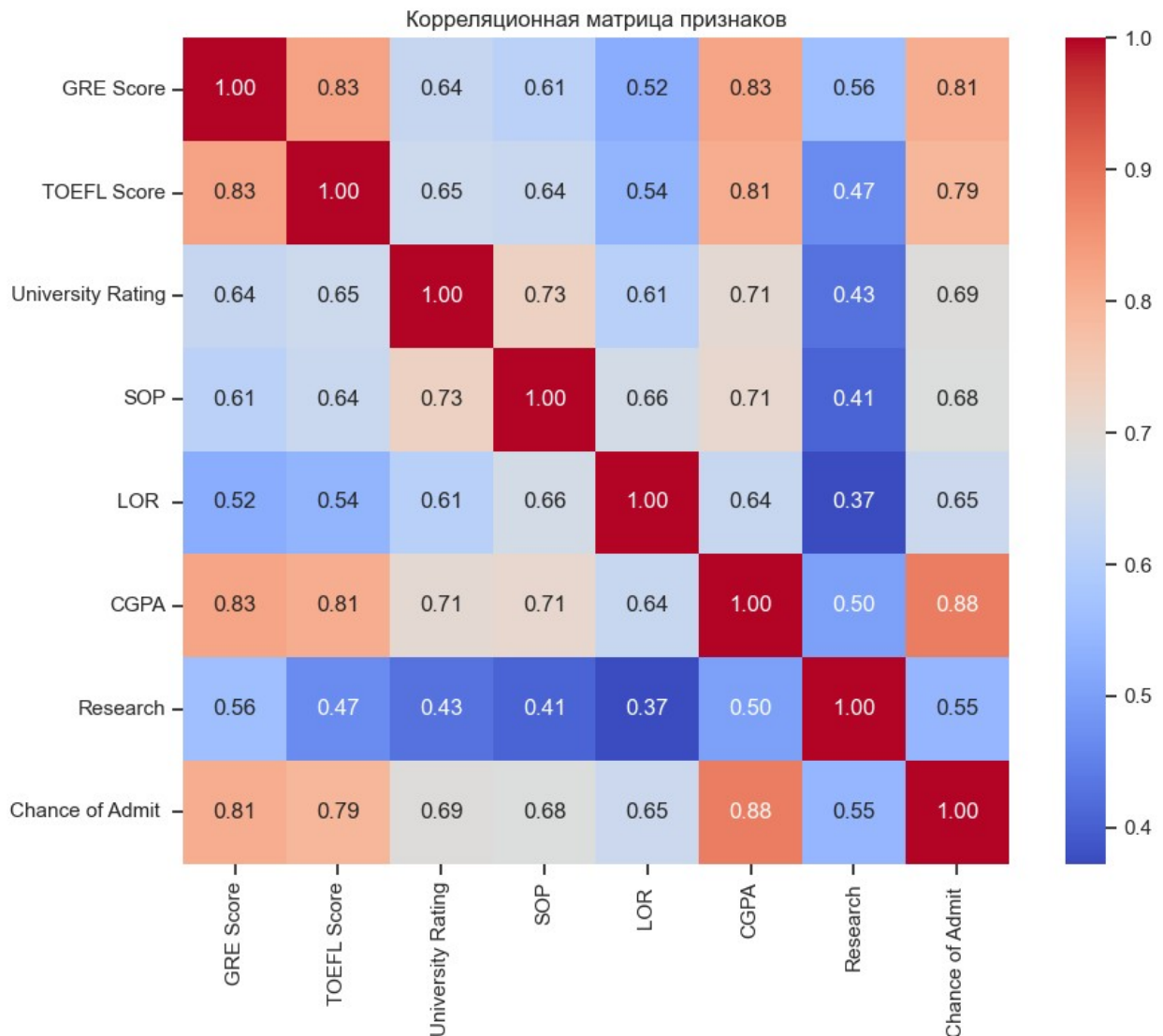
data.isnull().sum()

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

#удалим колонку, которая не влияет на корреляцию
data_clean = data.drop(columns=["Serial No."])

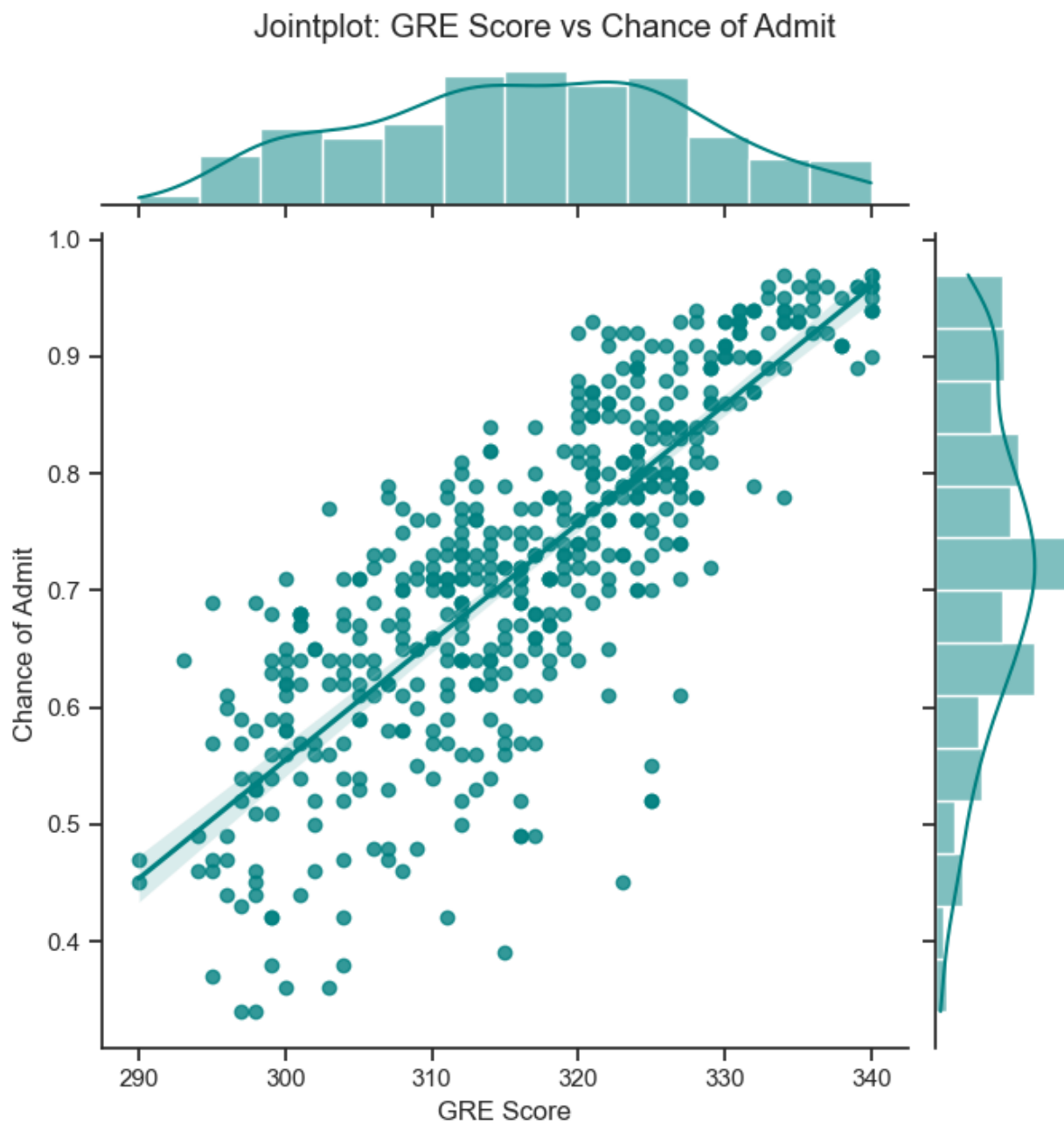
#построим корреляционную матрицу
correlation_matrix = data_clean.corr(numeric_only=True)

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm",
fmt=".2f", square=True)
plt.title("Корреляционная матрица признаков")
plt.tight_layout()
plt.show()
```



Наибольшую зависимость с шансами на поступление имеют признаки "CGPA", "GRE Score" и "TOEFL Score"

```
# Построим jointplot для GRE Score и Chance of Admit
sns.jointplot(data=data_clean, x="GRE Score", y="Chance of Admit ",
kind="reg", height=7, color="teal")
plt.suptitle("Jointplot: GRE Score vs Chance of Admit", y=1.02)
plt.show()
```



Видно, что между этими переменными есть положительная линейная зависимость — чем выше балл GRE, тем выше шанс поступления.