

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский университет ИТМО»

Практическая работа № 4

по дисциплине:

Визуализация и моделирование

Автор: Редичкина Анна Максимовна

Группа: К3242

Факультет: ИКТ

Преподаватель: Чернышева А.В

Санкт-Петербург

2020

Доказательство гипотез

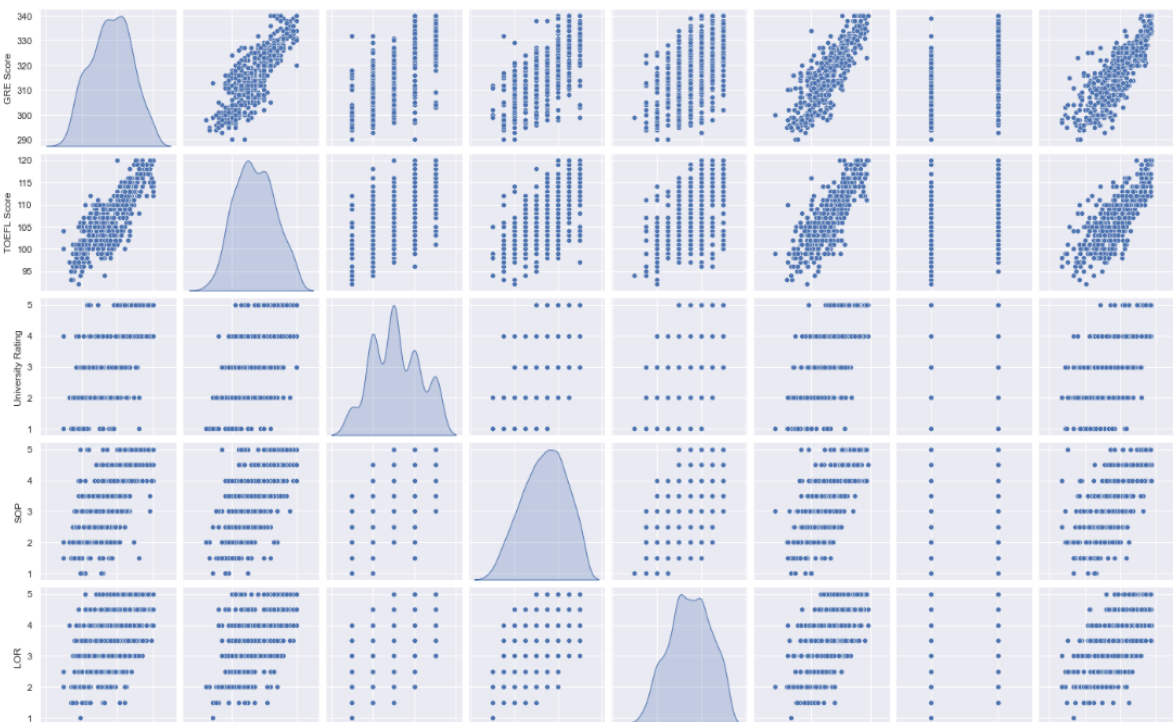
1. GRE и CGPA лучше других показателей определяют шансы на поступление

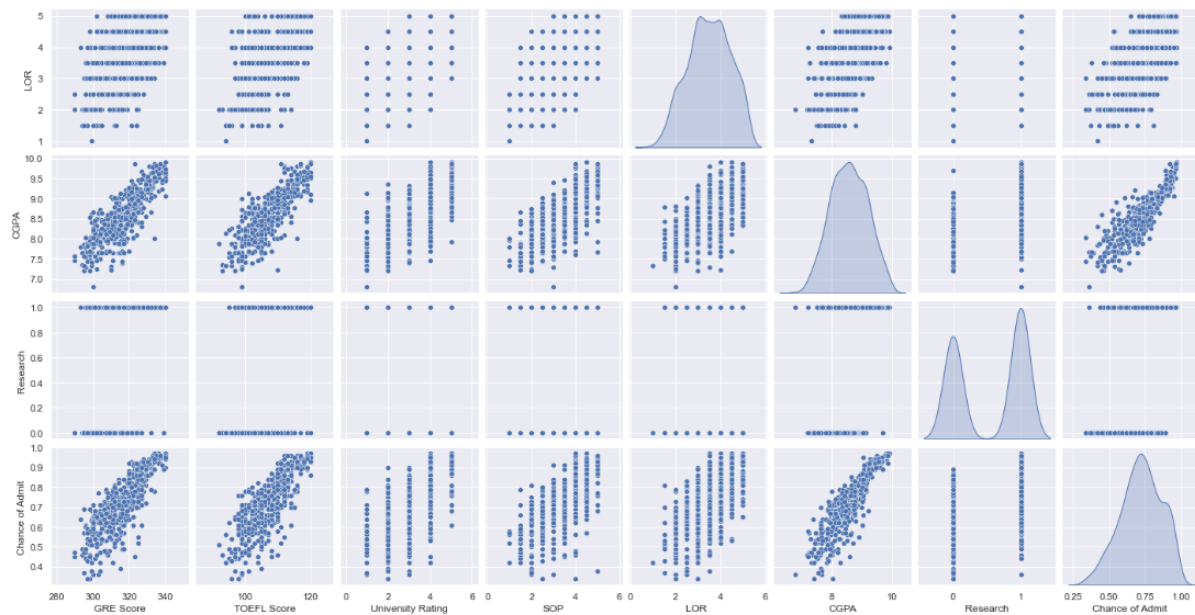
Посмотрим, как взаимосвязаны все показатели с шансом на поступление.

Код:

```
grad_num=df[['GRE Score', 'TOEFL Score', 'University Rating', 'SOP', 'LOR ', 'CGPA',  
            'Research', 'Chance of Admit ']]  
  
sns.pairplot(grad_num, diag_kind='kde')  
  
plt.show()
```

Результат выполнения кода:





Вывод: Между 'GRE Score', 'TOEFL Score', 'CGPA' & 'Chance of Admit' есть линейная связь, значит, чем больше баллы за каждый из этих критериев, тем выше шанс на поступление в университет, поэтому предположительно, 'Chance of Admit'

2. Опыт в исследовательской деятельности важен для поступления

Сначала посмотрим каково распределения абитуриентов с опытом в исследованиях и без по каждому университету.

Код:

```
sns.set_theme(style="darkgrid")
```

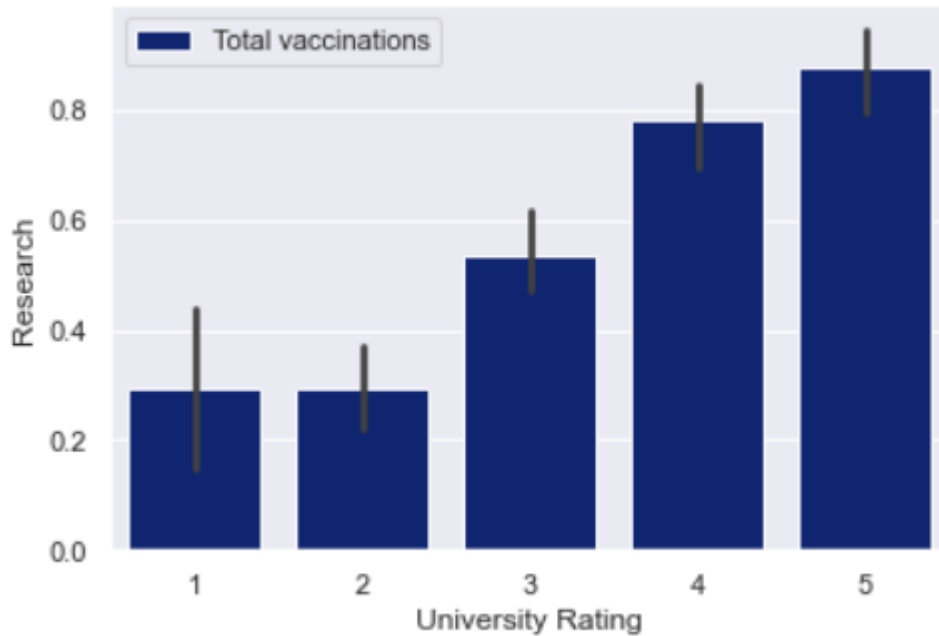
```
plt.plot()
```

```
sns.set_color_codes("dark")
```

```
sns.barplot(x="University Rating", y="Research", data=df, label="Total vaccinations",
color="b");
```

```
plt.legend(ncol=1);
```

Результат выполнения кода:



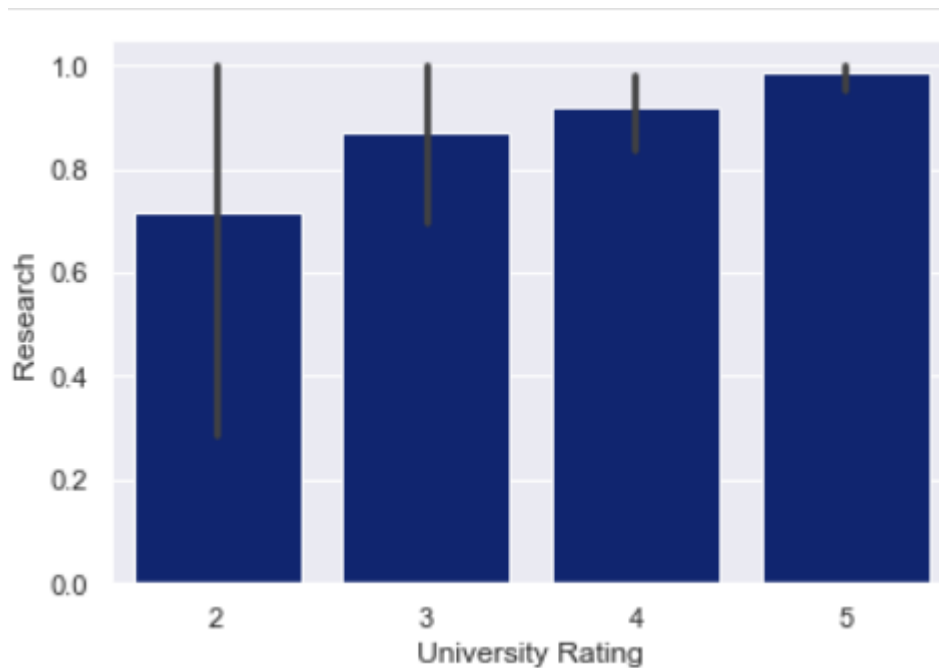
По графику видно, что университет, у которого процент абитуриентов с исследовательским опытом наибольший, занимает пятую позицию в рейтинге, а в университет на первой позиции большинство абитуриентов идут без опыта. Это может быть связано с отличающимися требованиями для поступления, направленностью учреждения (технический, гуманитарный и т.д), а со также степенью образования (чаще при поступлении в магистратуру смотрят на опыт в исследованиях).

Посмотрим, какой процент людей с высокими шансами (больше 80%) на попадание в каждый из университетов имели опыт в исследованиях.

Код:

```
res = df[df['Chance of Admit '] >= 0.8]
sns.barplot(x="University Rating", y="Research", data=res, label="Total vaccinations",
color="b");
```

Результат выполнения кода:



По графику можно сказать, что чем ниже позиция университета тем больше процент людей с исследовательским опытом. Абитуриенты с высокими шансами, поступающие в университет на первой позиции, не имеют опыта совсем.

Вывод: Из графиков выше можно сделать вывод, что важность 'Research' разнится для каждого университета.

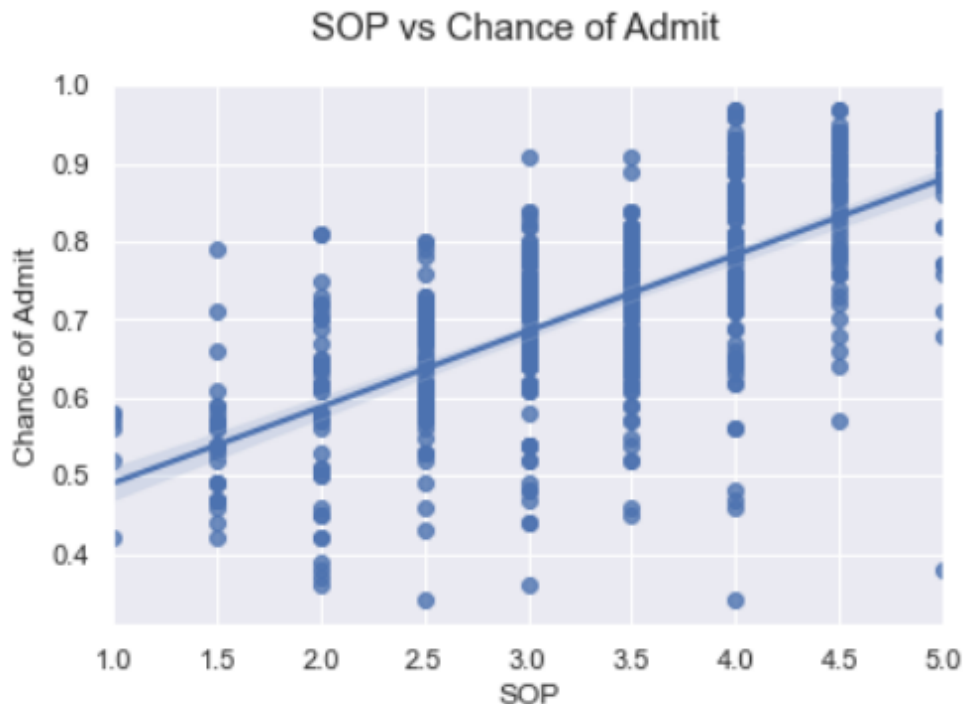
3. Оценка за заявление о намерениях (SOP) мало влияет на поступление

Посмотрим, какая зависимость между SOP и Chance of Admit.

Код:

```
fig = plt.figure()
sns.regplot(df['SOP'], df['Chance of Admit '])
fig.suptitle('SOP vs Chance of Admit', fontsize = 15)      # Plot heading
plt.xlabel('SOP', fontsize = 12)                            # X-label
plt.ylabel('Chance of Admit', fontsize = 12)
plt.show()
```

Результат выполнения кода:



Вывод: По графику можно сказать, что связь между двумя показателями больше случайная чем линейная. Абитуриент с низкой оценкой за SOP может иметь большие шансы на поступление и наоборот, поэтому критерий SOP мало влияет на шансы поступления.

4. Оценка за SOP и LOR меньше всего связаны с другими показателями абитуриентов

Код:

```
import numpy as np

corr = df.corr()

fig, ax = plt.subplots(figsize=(8, 8))

colormap = sns.diverging_palette(250, 15, s=75, l=40, n=9, center="dark")

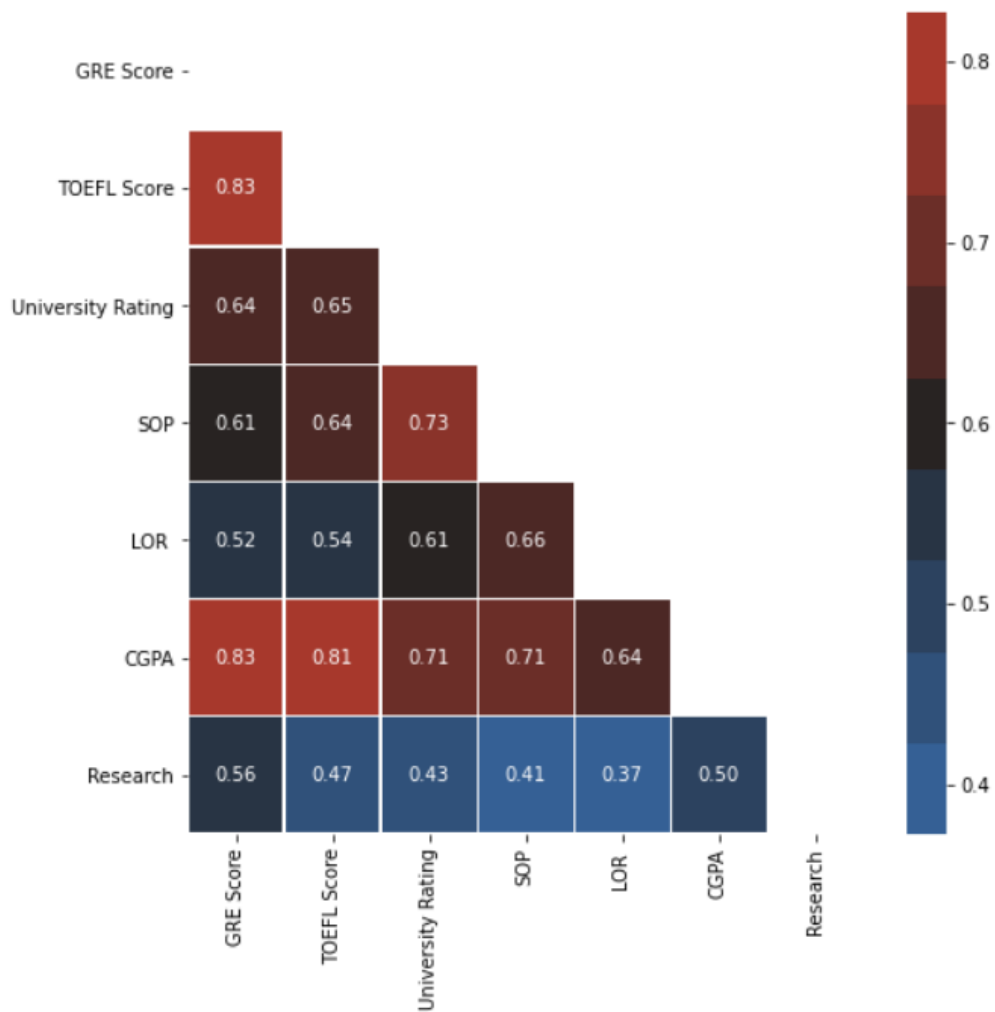
dropSelf = np.zeros_like(corr)

dropSelf[np.triu_indices_from(dropSelf)] = True

sns.heatmap(corr, cmap=colormap, linewidths=.5, annot=True, fmt=".2f",
            mask=dropSelf)

plt.show()
```

Результат выполнения кода:



Вывод:

'Research' меньше всего коррелирует с другими показателями.

5. Знание английского языка - один из самых важных показателей при приеме в университет

Построим график линейной зависимости между TOEFL Score и Chance of Admit.

Код:

```
fig = plt.figure()

sns.regplot(df['TOEFL Score'], df['Chance of Admit '])

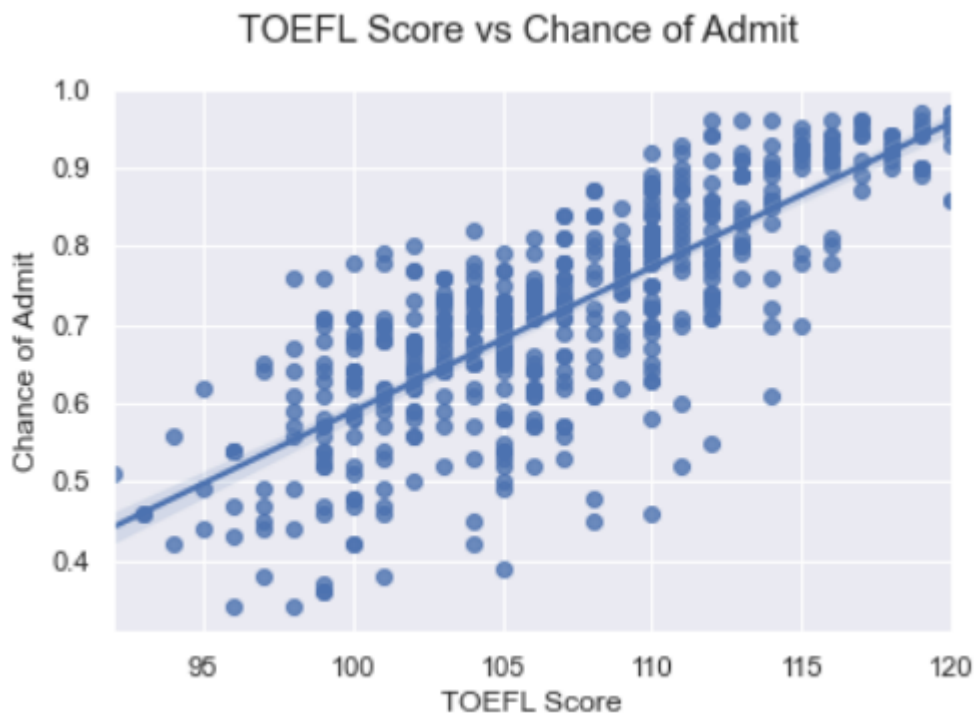
fig.suptitle('TOEFL Score vs Chance of Admit', fontsize = 15)      # Plot heading

plt.xlabel('TOEFL Score', fontsize = 12)                          # X-label

plt.ylabel('Chance of Admit', fontsize = 12)
```

```
plt.show()
```

Результат выполнения кода:



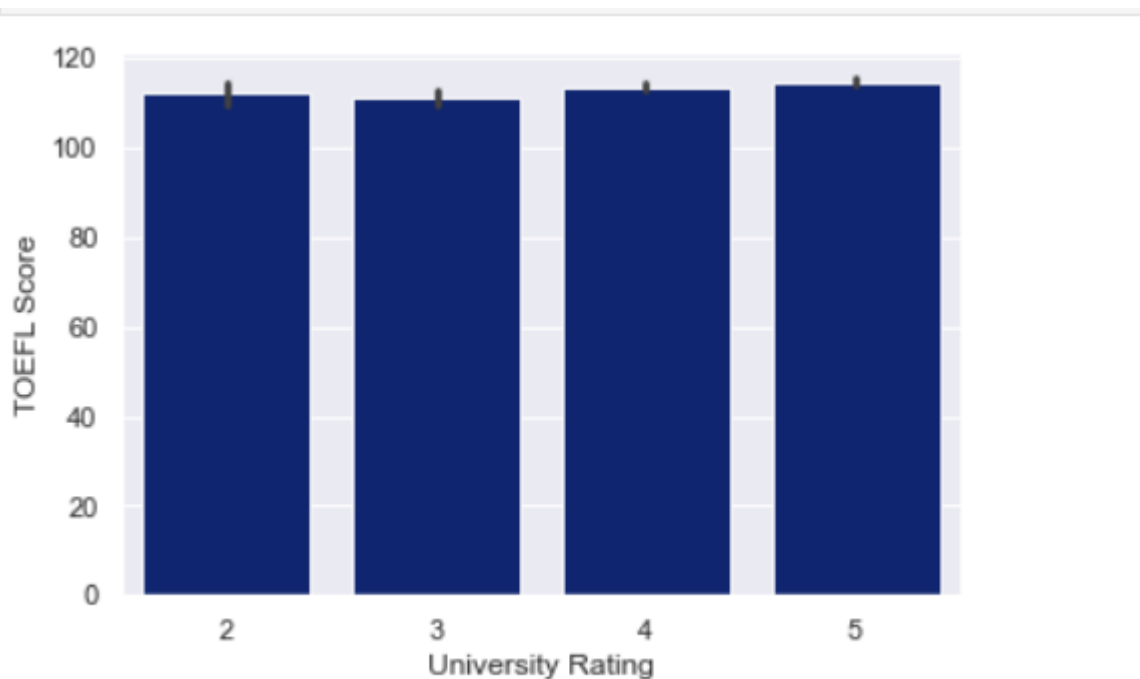
Данный график и график из пункта №1 показывает, что результат за экзамен TOEFL находится в линейной зависимости с Chance of Admit наряду с некоторыми другими показателями (GRE, CGMAT), поэтому его можно считать одним из самых важных критериев, определяющие шансы на поступление.

Также посмотрим, какой процент людей с высокими шансами на попадание в университет имеют хороший балл за экзамен

Код:

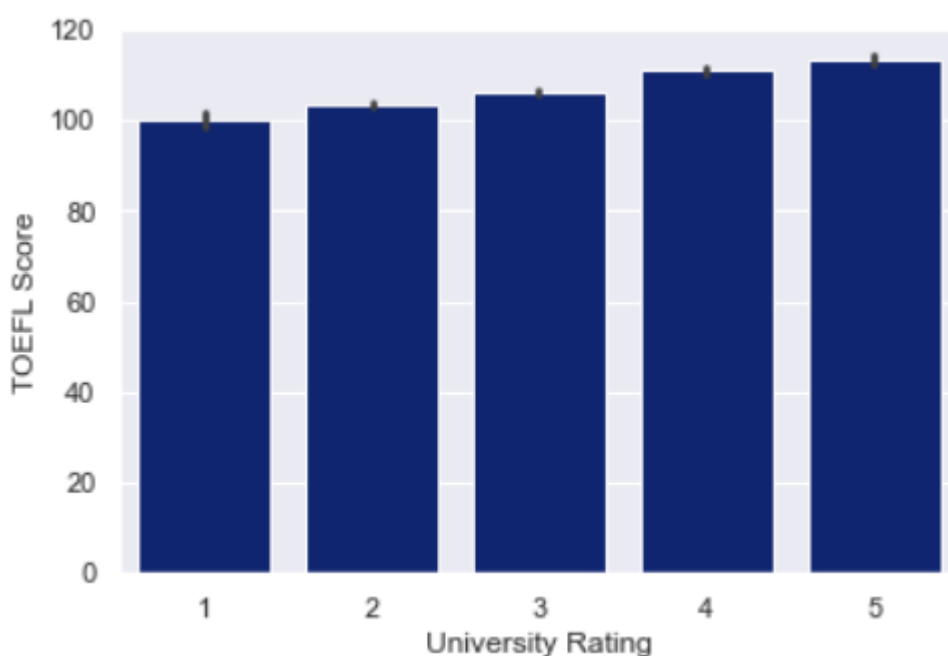
```
sns.barplot(x="University Rating", y="TOEFL Score", data=res, label="Total  
vaccinations", color="b");
```

Результат выполнения кода:



Мы видим, что у всех абитуриентов с наибольшим Chance of Admit хорошие результаты за экзамен TOEFL.

И построим график по всем абитуриентам:



По графику видно, что разница в баллах по университетам не слишком велика, а значит значимость знания английского при поступлении для каждого университета примерно одинаковая.

Вывод: Знание английского языка можно считать одним из самых важных критериев, определяющие шансы на поступление.