DAV Team Assignment

– Pranjal Gupta
190010053
pranjalgupta2507@gmail.com
9022416702

Assumptions –

- The university rating referred to in the dataset is the university **from which the student has graduated.**
- All the 500 students in the dataset are applying to the top **100 QS ranked universities,** because otherwise the concept of 'Chance of Admit' becomes illogical**.**

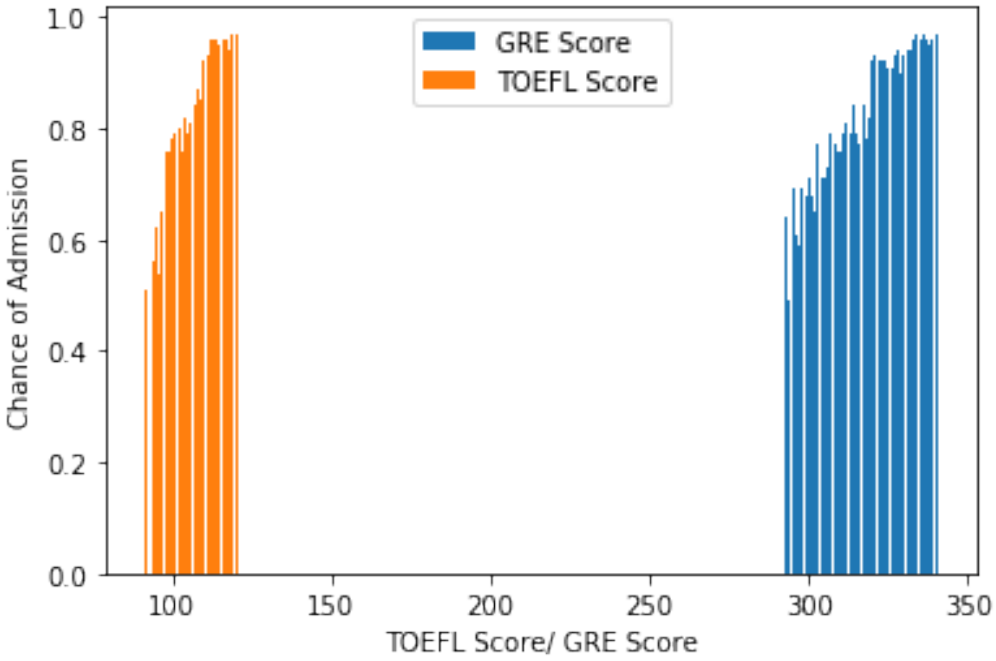
All the plots are made in Python using the **Pandas, Numpy** and the **Matplotlib** libraries**.**

Question 1

Do GRE & TOEFL scores influence the chance of getting admitted?

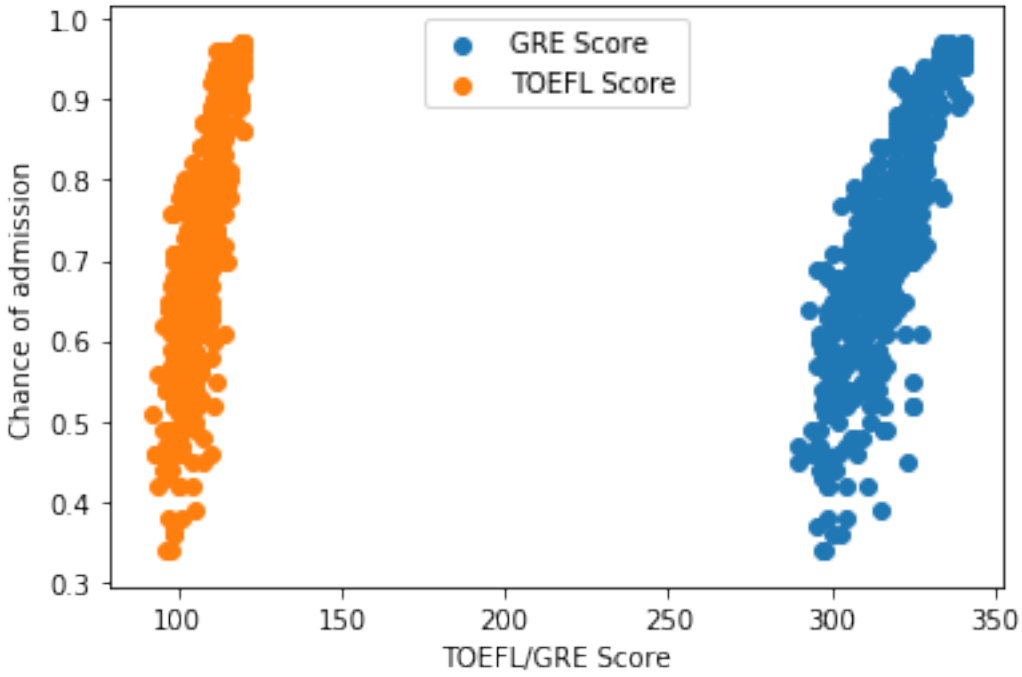
Yes, GRE Scores and TOEFL Scores influence the chance of getting admitted**.** It is the basic requirement and holds a major weightage in one's candidature when applying in Universities abroad for a Master's Program**.**

It can also be seen by analysing the data from the given dataset –



As you can see from the plot, as the TOEFL or GRE Score increase, the chance of admission also increases.

We can also visualise the correlation using a scatter plot –



We can also calculate the correlation between the two columns in the dataset using the '**corr()**' function.

The correlation is a number between –1 and 1 which shows how two variables are related with each other. **If corr() > 0, it means that they are positively correlated. corr() < 0 means they are negatively correlated and corr() = 0 means the variables are not related at all.**

The correlation between the GRE Score and Chance of Admit = **0.81**

Similarly,

The correlation between the TOEFL Score and Chance of Admit = **0.79**

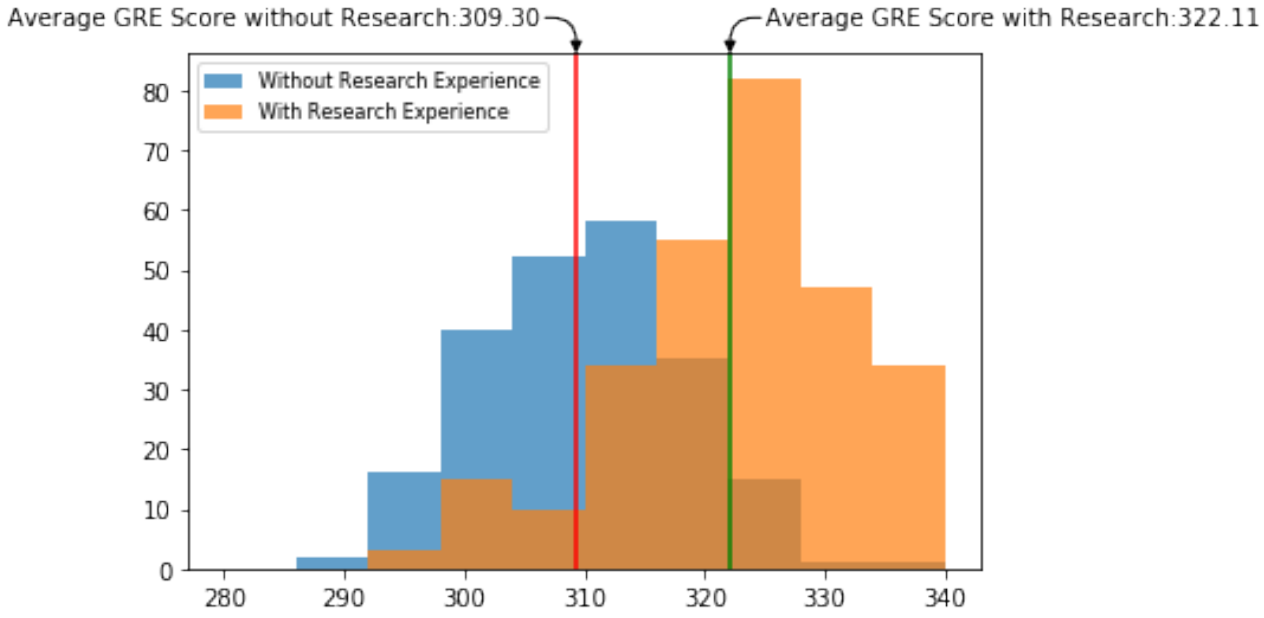
As we can see the correlation between the scores and the chances of admission are highly correlated (its not equal to one because there are other factors which come into play like SOP and LOR rating, Research experience, etc. But it is a major contributor in deciding one's probability of admission.

Question 2 –

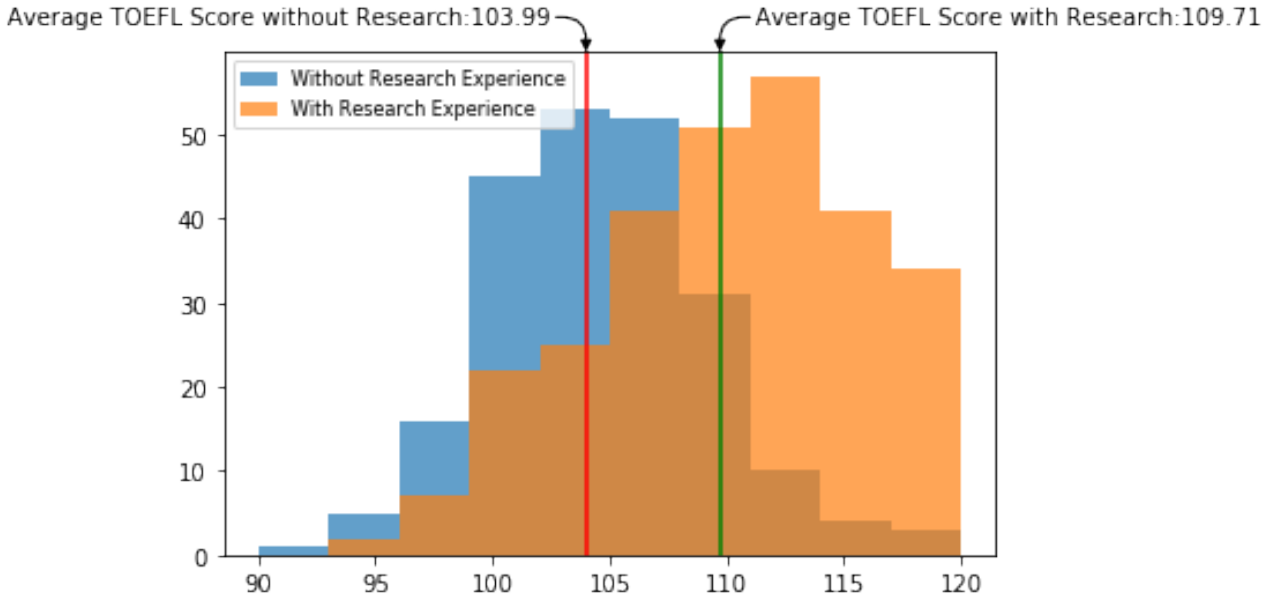
Does having Research experience makes you score well in GRE & TOEFL?

If we analyse the dataset, by dividing the dataset into two parts based on Research = 0 and Research = 1 and then plot a histogram of the GRE and TOEFL scores of both the datasets in one plot, we can see the difference between the average scores of both the datasets.

Plot for the GRE Scores –



Plot for the TOEFL Scores –



From the analysis, we can infer that a person on average will perform better in GRE and TOEFL if he/she has some Research experience.

Logical Explanation – **the knowledge one gains in Research may not necessarily help one in scoring better in GRE or TOEFL, but rather, while doing Research, one develops the required analytical skills, the Determination and the Discipline** with which anyone can easily score good in any exam and not just particularly GRE or TOEFL. So the analysis results make sense.

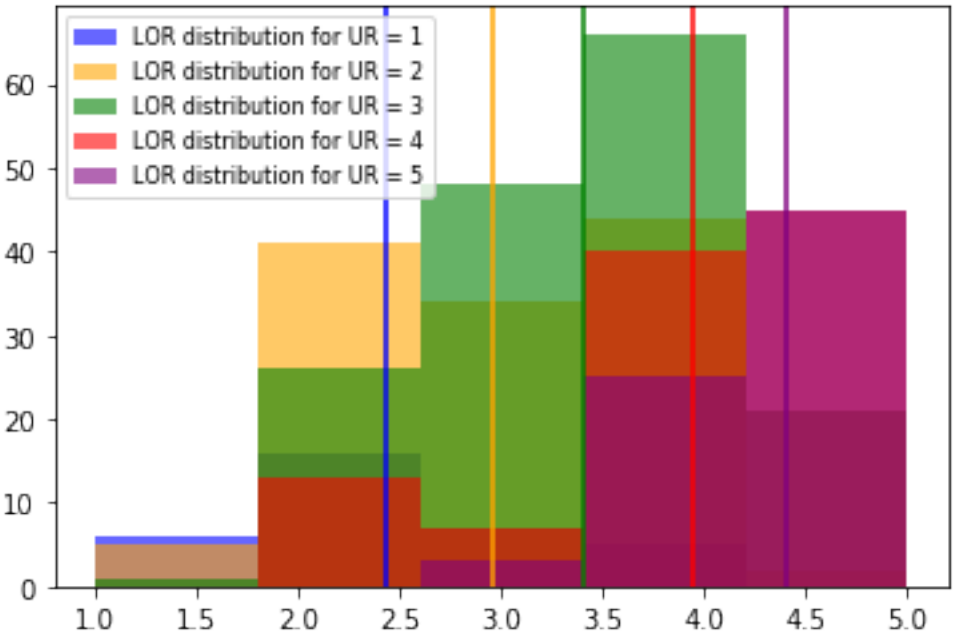
Question 3 –

Does the University Rating influence my SOP and LOR Rating?

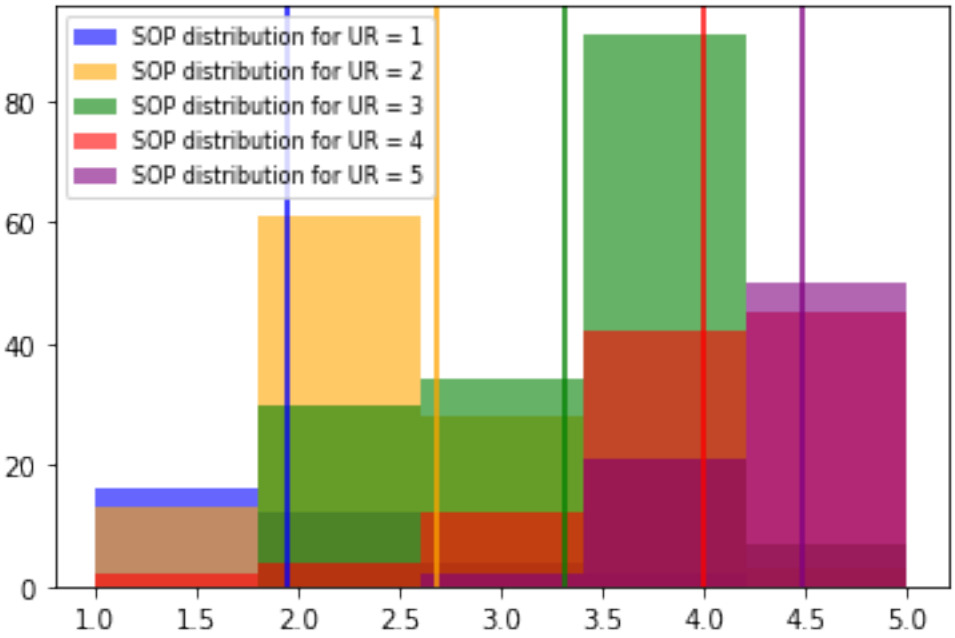
If we analyse the data from the dataset, we can see that there is an increasing trend followed in the SOP and LOR strengths as the University Rating increases.

This can be seen in the following plots –

Plot for the SOP Rating –



Plot for the LOR Rating –



Now, in the above plots, the **vertical lines** denote the **mean** for the histogram of the same colour. The legend has been provided in the plot for easy understanding.

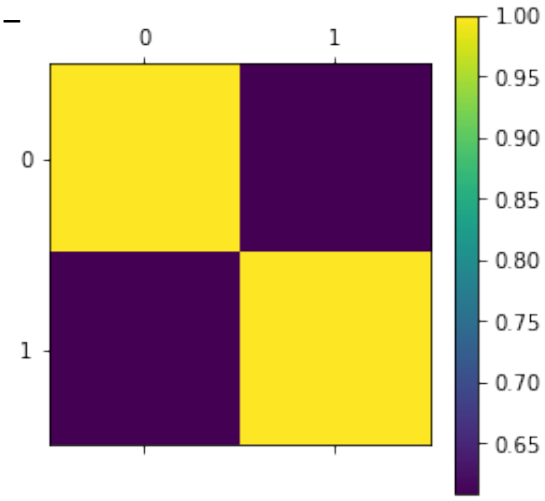
As can be seen from the plots, the Average SOP/LOR rating increases as the University Rating increases.

Logical Explanation – This makes sense logically, as the higher rated university will have a better culture and in general better applicants compared to a lower rated university. So a better applicant will have a better SOP and LOR as compared to others. Therefore the inference from the analysis can be justified.

In addition to a histogram distribution, a Correlation Matrix can also be plotted.(using **matshow()** function in Matplotlib library)

Correlation for University Rating and SOP rating = **0.73**

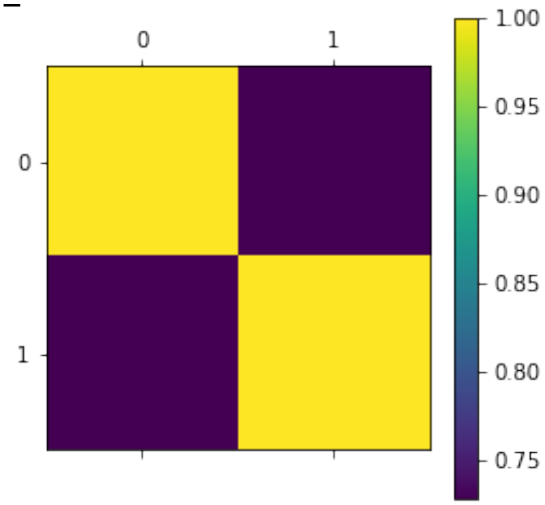
Correlation Matrix –



Here 0 refers to ‘University Rating’ and 1 refers to ‘SOP Rating’

Correlation for University Rating and SOP rating = **0.61**

Correlation Matrix –



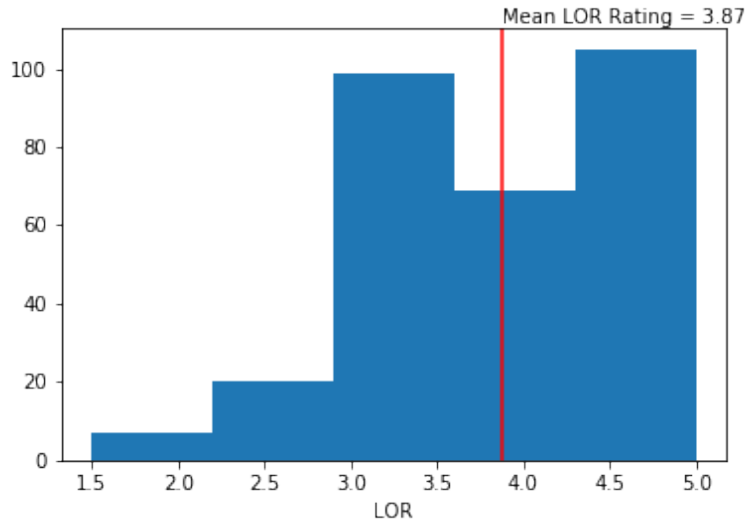
Question 4 –

What is the average SOP & LOR Rating of students who got admitted?

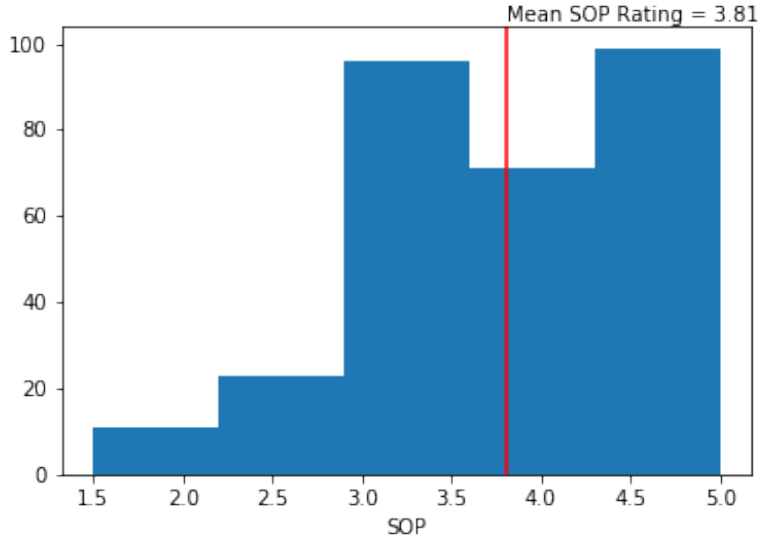
The minimum threshold of the Chance of Admission probability for getting admitted is kept at 70%. As previously mentioned, it is assumed that all applications are to the Top 100 QS Ranking universities and thus keeping the threshold probability at 50% does not make sense.

These were the results from the analysis –

The average LOR Rating for Chance of Admission greater than 70% – **3.87**



The average SOP Rating for Chance of Admission greater than 70% – **3.81**

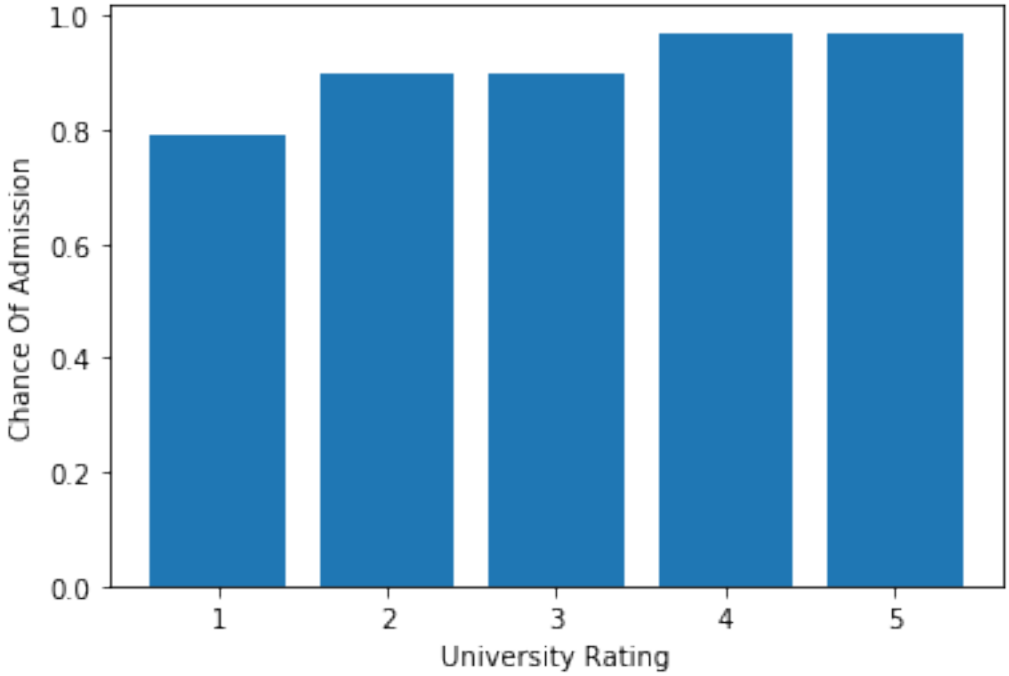


Question 5 –

How does the University Rating improve the chance of getting admitted?

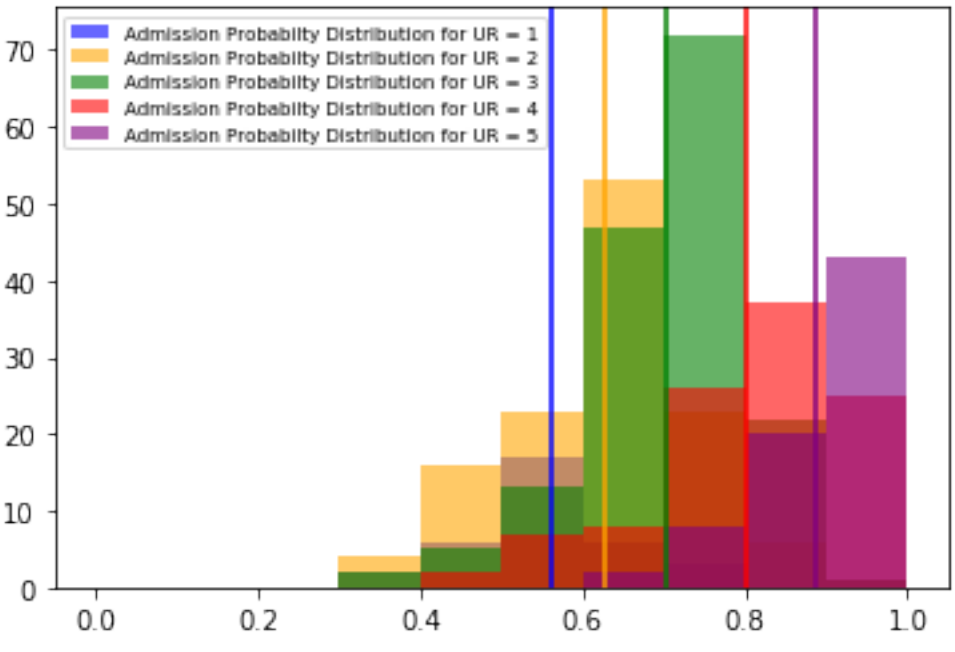
For this question, I created different subsets of the dataset based on the University ranking. Then I plotted the histogram of the 'Chance of Admit' variable for all the 5 sub-datasets. These are my results –

Bar plot – Chance of Admission vs University Rating –



From the bar plot we can see that the maximum chance of acceptance increases as the university rating increases. A better visualisation is shown below –

Histogram distribution of 'Chance of Admit' values in the 5 datasets –



In this plot, the 5 vertical lines represent the mean of the 5 different histogram distributions (of the same colour).

As expected, the average chance of admission sees an increasing trend as the University rating increases.

Correlation between University rating and Chance of Admit = **0.69**

Logical explanation – This trend also makes sense logically because **higher the rating of the university, better are the graduates in terms of SOP, LOR, CGPA and Research and thus the Chance of them getting accepted is more as compared to a lower rated university.**

Question 6 –

What should be your Scores for 0.9 % Chance of Admission?

For this question I am giving the analysis in two ways –

1. Range of all attributes
2. Average of all attributes

Also, it is important to give information about all the attributes because only one or two factors cannot decide the chance of admission.

These were my results –

Range –

1. GRE Score – **324 to 340**
2. TOEFL Score – **111 to 119**
3. SOP Rating – **4.0 to 5.0**
4. LOR Rating – **3.0 to 4.5**
5. CGPA – **9.16 to 9.6**

Averages –

1. GRE Score – **330.56**
2. TOEFL Score – **116.11**
3. SOP Rating – **4.50**
4. LOR Rating – **4.11**
5. CGPA – **9.32**

Question 7 –

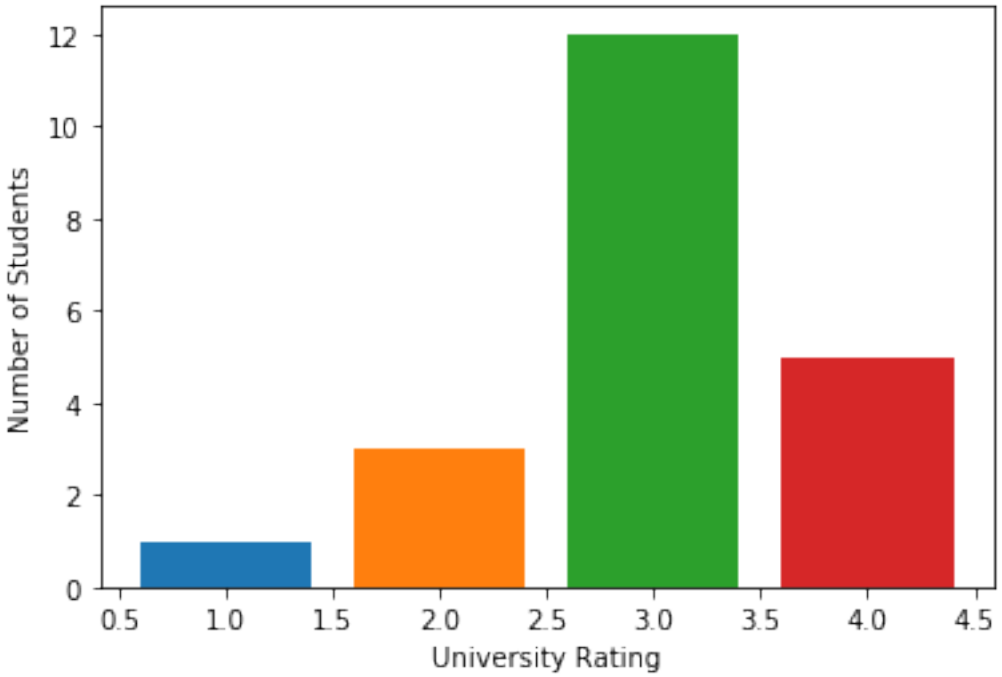
There are many students with more than average GRE and TOEFL scores but still their chance of admission is less than the average (~70%). What can be the reasons for this observation?

This is the case where the importance of all the factors comes into play to decide the chance of admit. Even though one may have good GRE or TOEFL scores, one may be rejected by the university. And this can happen with other attributes as well.

I analysed this problem by first finding all the entries where the conditions specified above(more than average GRE and TOEFL score and less than average chance of admit) from the dataset.

Then I plotted different attributes to analyse in what range they fall in. These are the plots –

a. Universities in which these students were enrolled –

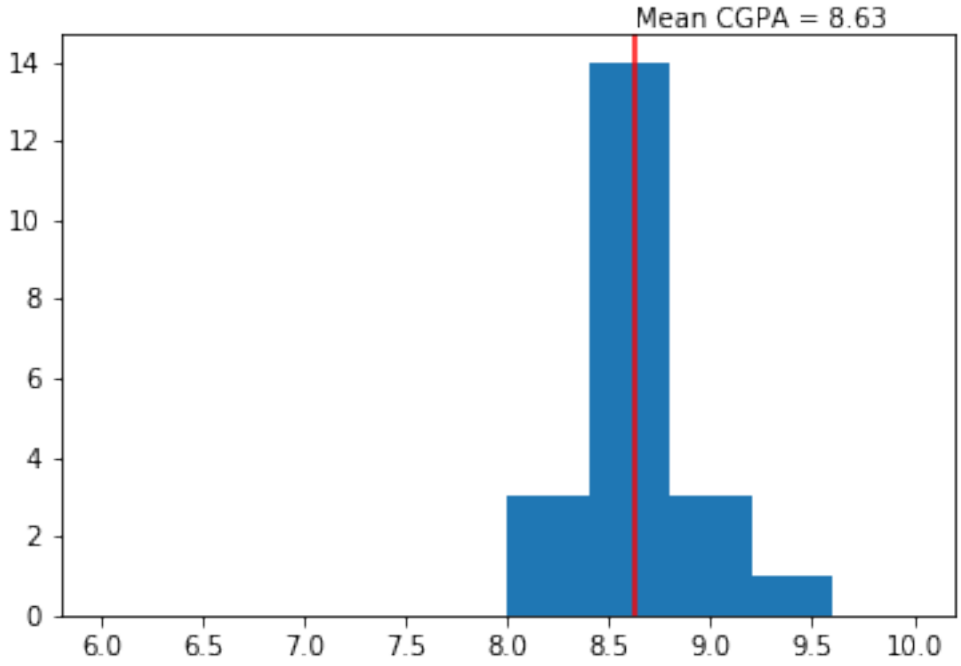


From this plot, you can see none of the students were enrolled in a 5 rated university. And the maximum of them were enrolled in 3 rated universities. The number from the 1 and 2 rated are low, because student from these universities generally apply to lower ranked universities and therefore seldom fall into this category.

b. Ratio of the students with Research experience to without Research Experience = 0.75

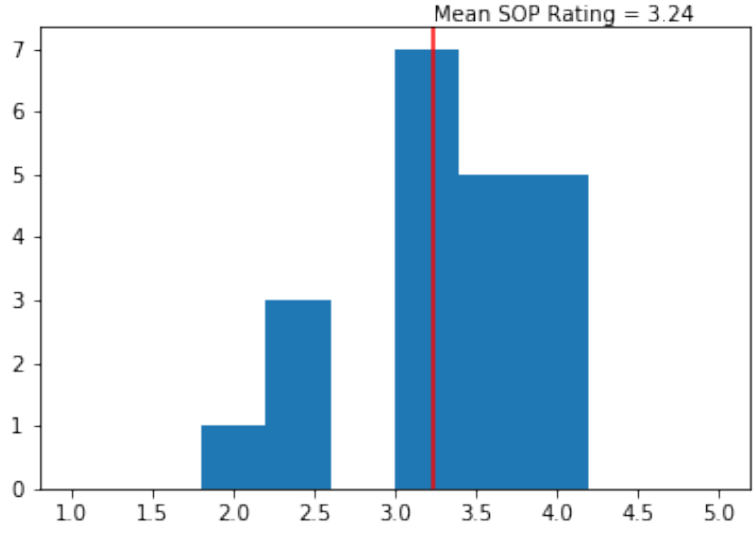
The ratio being less than one shows that more number of students are without research experience than the ones with research experience. Hence, this can be on of the reasons for the students' rejection.

c. Mean CGPA of these students = 8.63



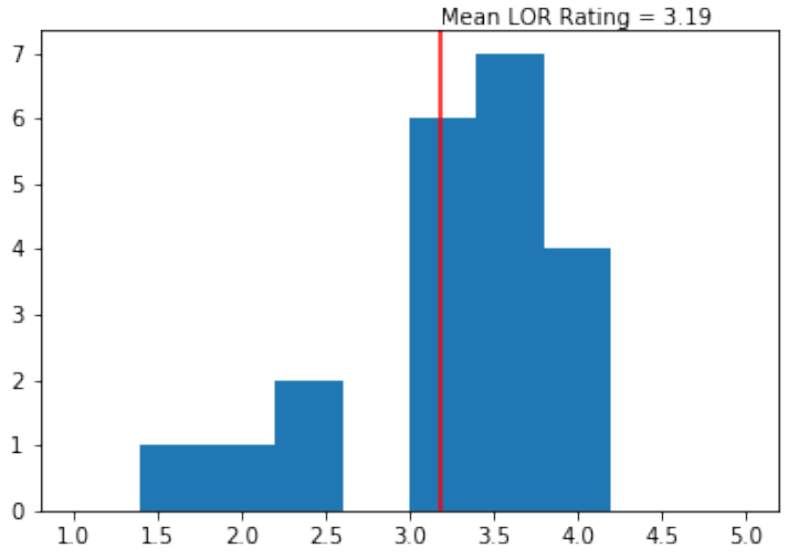
This bar plot shows the average CGPA of the students who were rejected in spite of good GRE and TOEFL scores and thus **CGPA can be a major factor in resulting an acceptance/rejection of the student.**

d. Mean SOP rating of the applications of these students = 3.24



This bar plot shows the average SOP Rating of the students who were rejected in spite of good GRE and TOEFL scores and thus **SOP strength can be a major factor in resulting an acceptance/ rejection of the student.**

e. Mean LOR Rating of the application of these students = 3.19

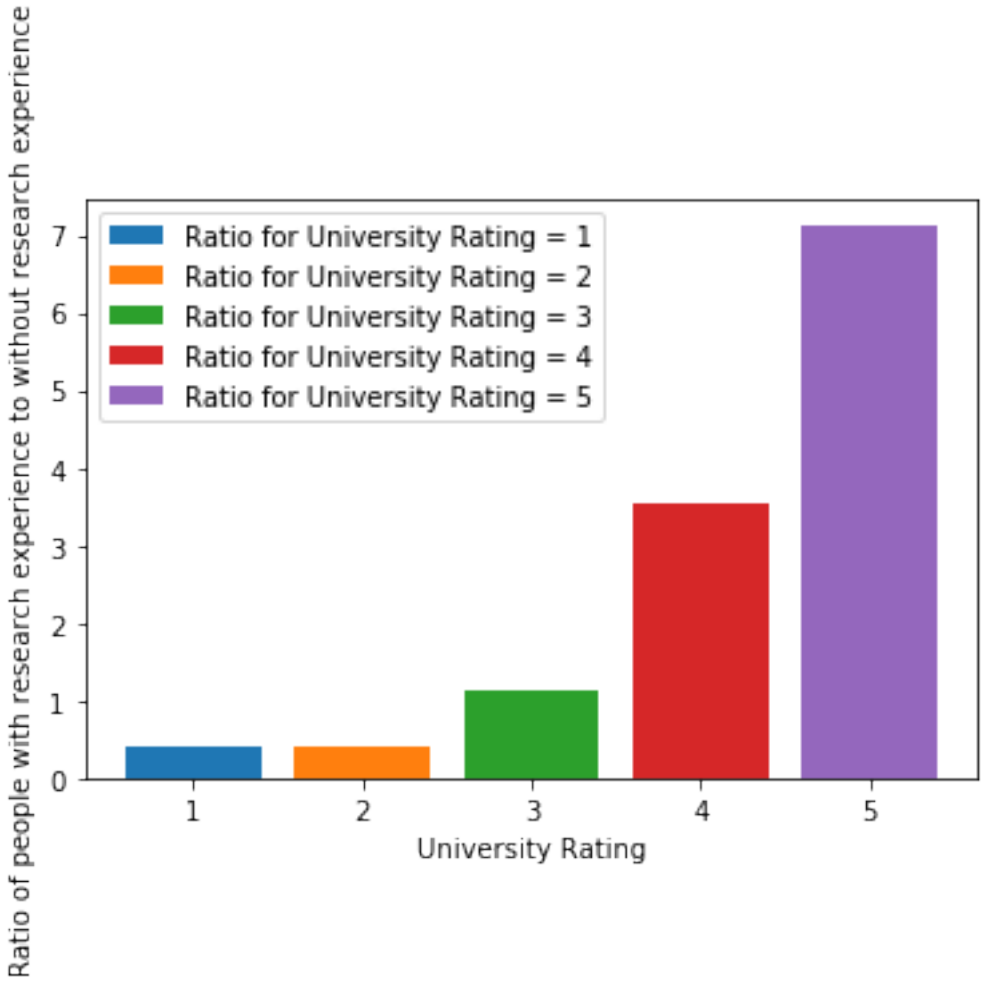


This bar plot shows the average LOR Rating of the students who were rejected in spite of good GRE and TOEFL scores and thus **LOR strength can be a major factor in resulting an acceptance/ rejection of the student.**

Question 8 –

How does the ratio of students with research experience to without research experience vary with respect to University ratings?

For this analysis, I used the previous 5 datasets which were obtained from the main dataset based on the university ratings. Then using the value_counts() attribute of the pandas, I calculated the number of people with and without research experience and then calculated the ratio. After getting this ratio for each of the 5 datasets, I plotted my readings on a bar ploy. This is my analysis –



As you can see from the bar plot, the ratio of the number of people with research experience to people without research experience keeps on increasing with the increase in the university ratings.

Final Inference –

This analysis makes logical sense**. Universities with higher rating are more reputed in terms of the research carried out in the university. Also the funds allotted for Research are much more in such reputed universities. So the opportunity and incentive for research in these universities is higher and thus more people can be found with research experience in higher rated universities.**

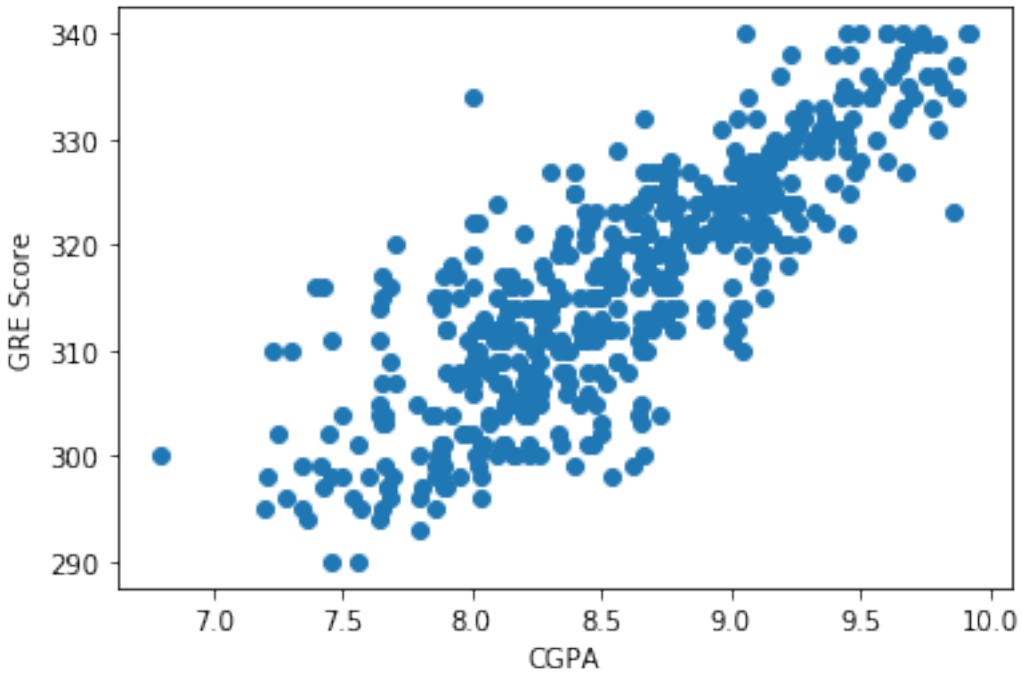
Question 9 –

Does a student's CGPA affect his/her GRE or TOEFL Scores?

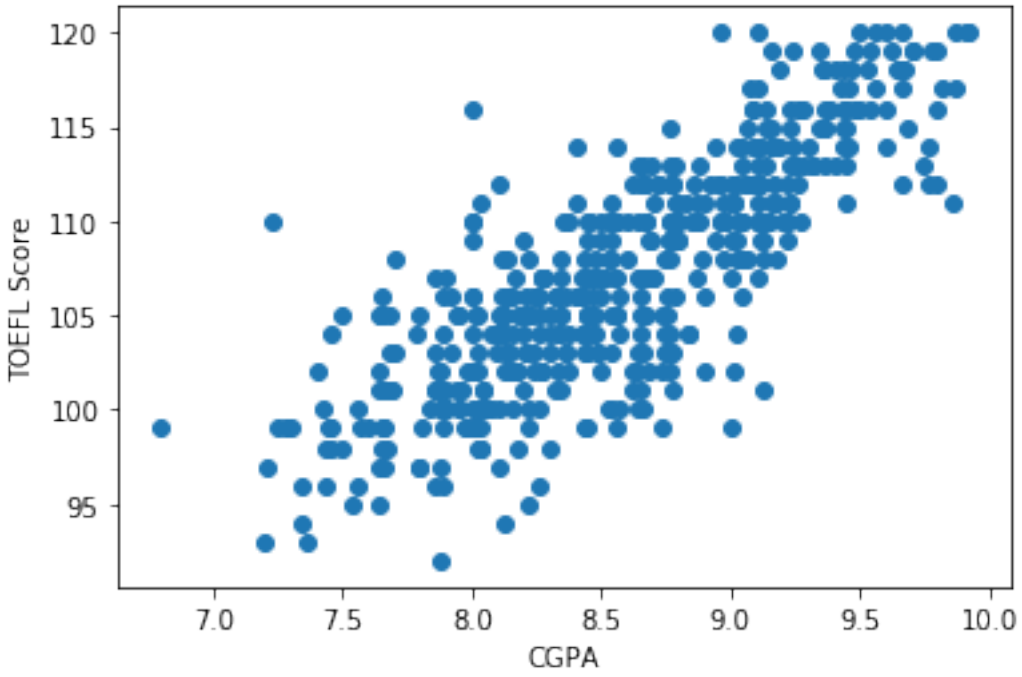
Here is the analysis for this question. This is again a bivariate problem analysed using Scatter plots and correlation matrix.

Scatter Plots –

GRE Score vs CGPA –

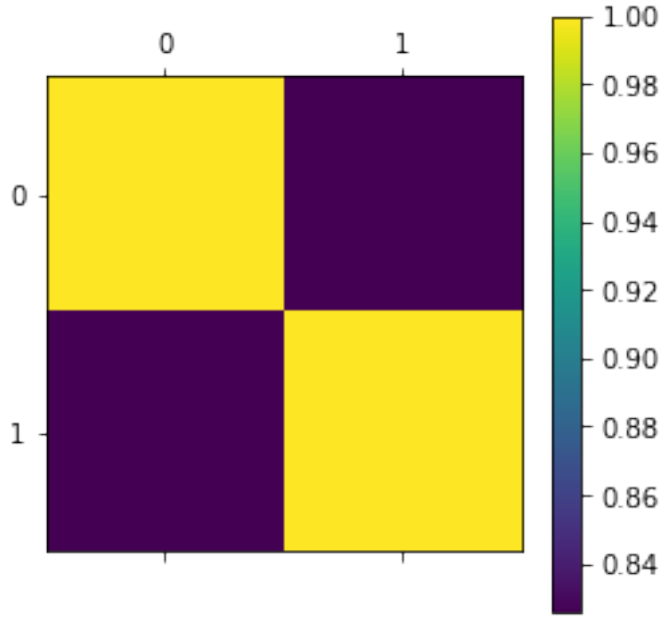


TOEFL Score vs CGPA –

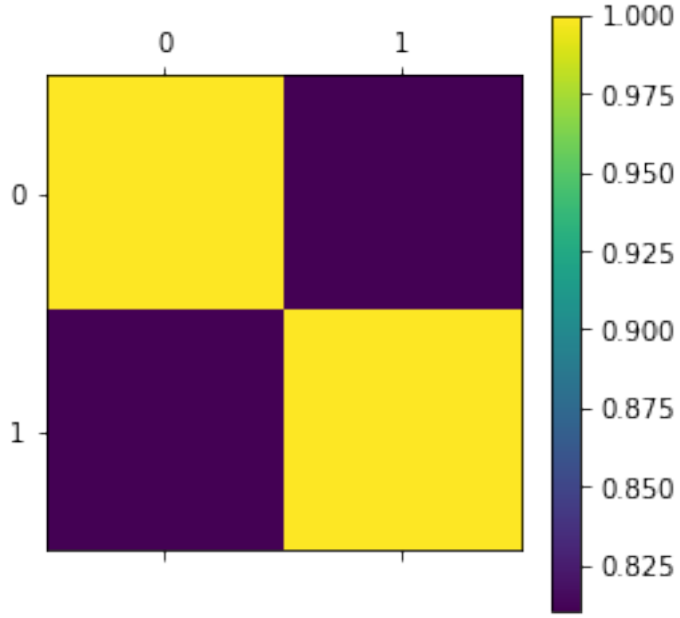


Correlation Matrix –

CGPA vs GRE Score –



CGPA vs TOEFL Score –



Correlation between CGPA and GRE Score = **0.82**

Correlation between CGPA and TOEFL Score = **0.81**

Final Inference/Logical Explanation –

In this question, the results from the analysis show that that CGPA and GRE/TOEFL scores have a high positive correlation. The reason for this observation is that **if one is able to maintain and score a high CGPA in his/her university, this means that he/she has the required determination, discipline and hard working ability to score well in any other exam and not just GRE/TOEFL.** This analysis definitely **does NOT show that the knowledge required to score a good CGPA is the same knowledge required for scoring good in GRE/TOEFL.** This is true in real life as GRE/TOEFL mainly consists questions based on Maths(easy) and English which is not the main focus of study of an undergraduate in his/her university.

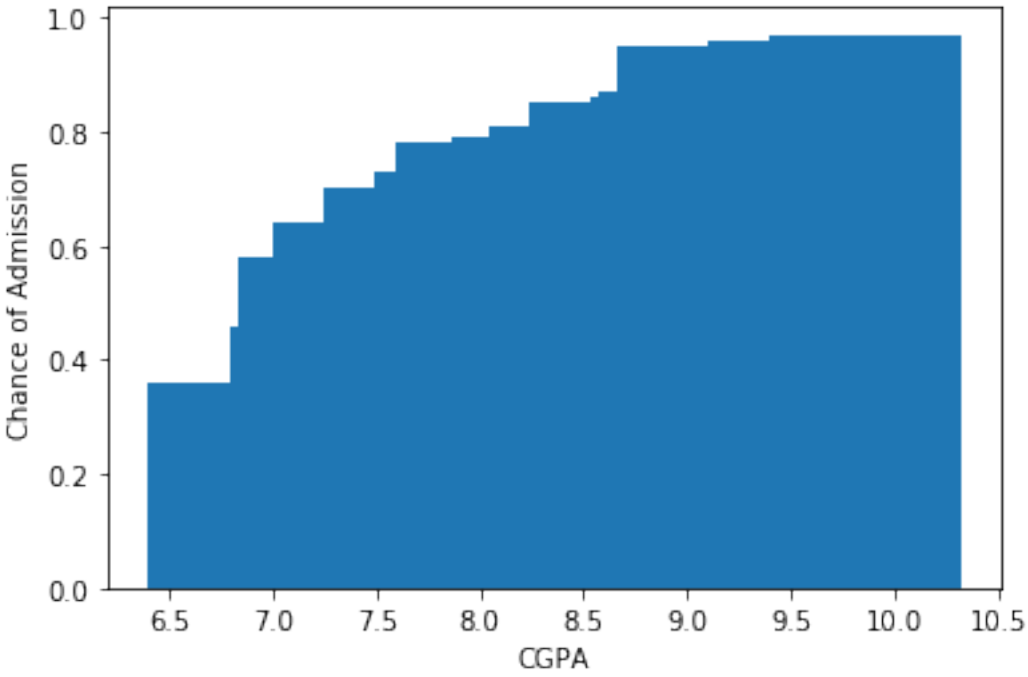
Question 10 –

How does the CGPA influence the Chance Of Admission?

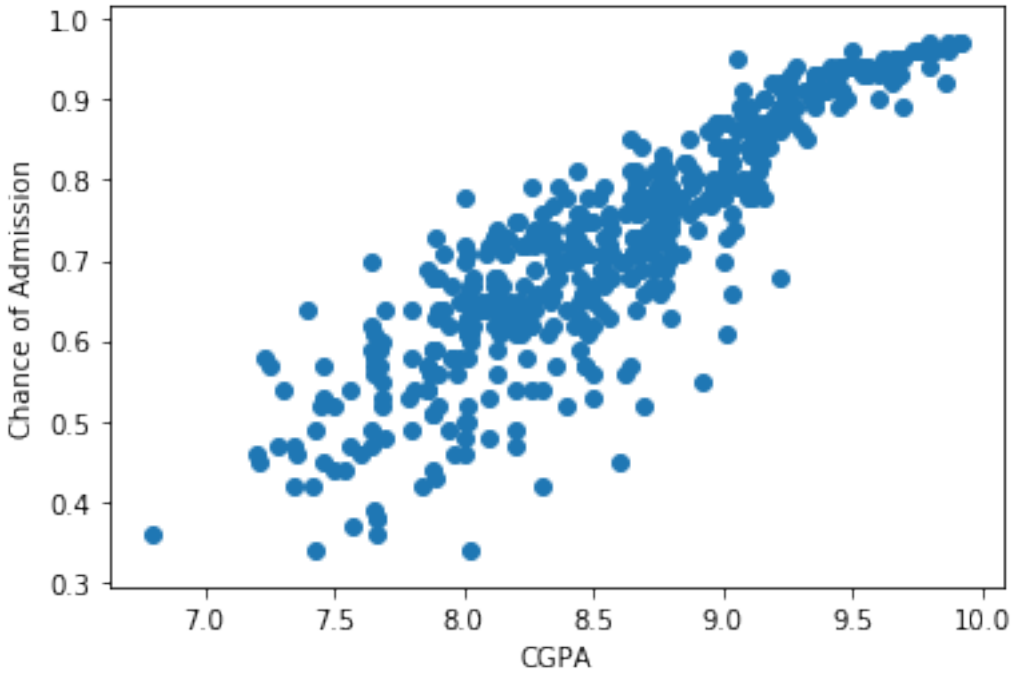
For this question, I performed bivariate analysis through scatter plots, bar plots and the Correlation matrix.

All the plots are shown below –

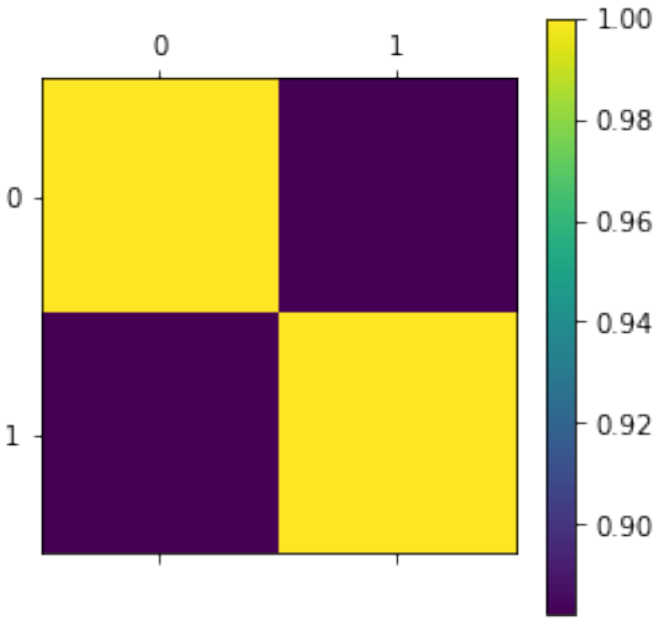
Bar plot –



Scatter Plot –



Correlation Matrix –



Here 0 stands for CGPA and 1 stands for Chance of Admit.

Correlation between CGPA and Chance of Admit = **0.88**

Logical Explanation – As you can see from the correlation coefficient between the two factors is close to 1, this means that both the factors are almost linearly related suggesting that CGPA plays a major role in admission apart from other factors like Research, SOP, LOR, GRE/TOEFL score, which is logically self explanatory.