Introduction

Every country has various ways in which universities approach the admission process for higher studies. Students that come to the United States from India for higher education face a very lengthy process of admissions at universities. To make the process easier, they pay huge chunks of money to educational consultancies which try and give them assurances of getting admitted to a certain university of their liking. When I had decided to come to the US for higher education, I had started doing my research of courses and universities that I was targeting only to find out that consultancies already had that information mapped out. Majority of students go through this process, however due to money constraints I had decided to do the admissions process on my own.

My motivation behind choosing this project is to help others see the relationship between chance of getting an admission based on features like CGPA, TOEFL Score, GRE Score, Letters of Recommendation
(LOR), Statement of Purpose (SOP), and University Ranking. This project will help students in shortlisting universities based on their profiles and give them a fair idea of what their chances of getting into the university of their choice.

1. Information about the dataset.

The admissions predict data set is taken form the UCLA Graduation data set. This data set is owned by "Mohan S Acharya". (Citation included at the end of paper). This data set contains the following parameters which are considered important during application of master's programs:

- GRE Scores (out of 340)
- TOEFL Scores (out of 120)
- CGPA (out of 10)
- University Rating (from (1,2,3,4,5))
- Chance of Admit (ranging from 0 to 1)
- Research Experience (either 0 or 1)
- SOP and LOR Strength (from (1,2,3,4,5))

In total the dataset has 500 observations and nine columns. The question I pose is according to me only GRE and CGPA scores matters the most even though there is a feeling all variables would matter together.

Data Description

The following summary of data shows that all the columns of are numerical data type.

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
MEAN	316.47	107.19	3.11	3.37	3.48	8.57	0.56	0.72
STD.	11.29	6.08	1.14	0.99	0.92	0.60	0.49	0.14
MIN.	290.00	92.00	1.00	1.00	1.00	6.80	0	0.34
25%	308.00	103.00	2.00	2.50	3.00	8.12	0	0.63
50%	317.00	107.00	3.00	3.50	3.50	8.56	1.00	0.72
75%	325.00	112.00	4.00	4.00	4.00	9.04	1.00	0.82
MAX.	340.00	120.00	5.00	5.00	5.00	9.92	1.00	0.97
MEDIAN	317.00	107.00	3.00	3.50	3.50	8.56	1.00	0.72

I also calculated based on how rating affects chance of admission.

RATING	GRE	TOEFL	SOP	LOR	CGPA	RESEARCH	CHANCE
1	304.91	100.20	1.94	2.42	7.79	0.29	0.56
2	309.13	103.44	2.68	2.95	8.17	0.29	0.62
3	315.03	106.31	3.30	3.40	8.50	0.53	0.70
4	323.30	110.96	4.00	3.94	8.93	0.78	0.80
5	327.89	113.43	4.47	4.40	9.27	0.87	0.88

This table infers that the average chance of admission of applicants that applied to the program with rating 1 is less than that of applicants that applied with higher ratings. We can now improve our research by calculating data of more than 82% chance of admissions to better understand the relationship.

Below is the short version of the table when we take mean based on 82% or more chance.

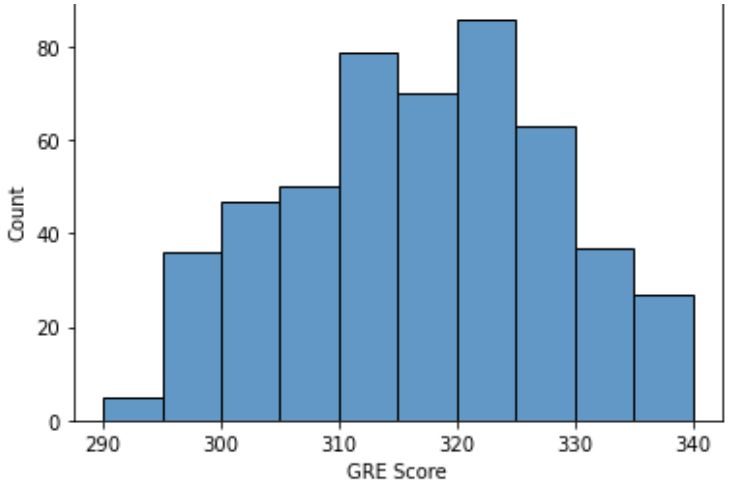
CHANCE	GRE	TOEFL	RATING	SOP	LOR	CGPA	RESEARCH
0.83	326.50	112.75	3.75	3.87	3.75	9.03	0.75
0.84	323.90	109.63	3.45	3.81	3.77	9.03	0.90
0.85	322.00	11.50	3.66	4.08	4.16	9.04	0.83
0.86	325.40	114.40	4.20	4.30	4.30	9.12	0.90
0.87	325.62	111.12	4.62	4.37	4.18	9.10	0.87

Observations from the above table are as follows:

- The data average CGPA is 8.57 from the summary table. Here we see that the average CGPAs with higher level of chance to admit is greater than 9.
- Similarly, SOP, TOEFL, GRE AND LOR scores with higher chance of admit are greater than the data average.
- The average research experience is 83% from the summary, but the above table shows 75% so there will be chances to reject the application in this case. So, we would need to look at greater than 83%.

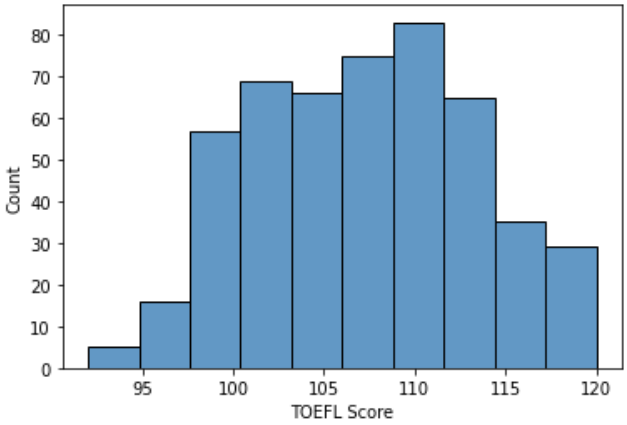
Now I will be showing some figures that will help us out in interpretation of the data in easier terms.

1. **Histogram of GRE Scores**



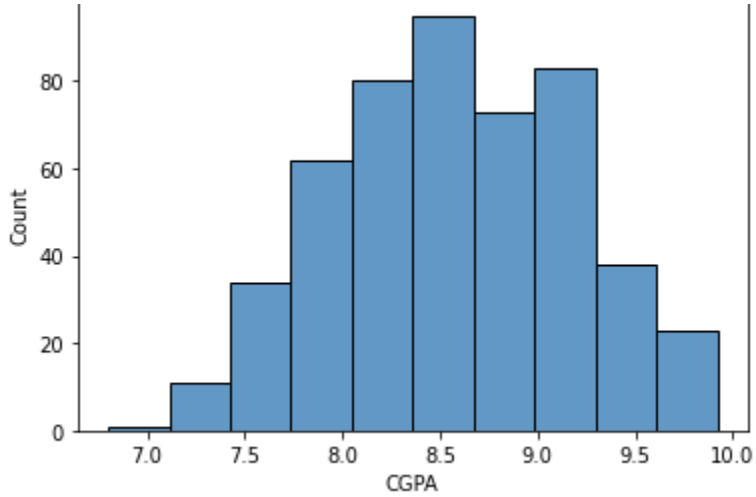
We can see that the GRE scores of most of the applicants are between 310 and 330. Very few have score more than 330 and similarly less than 310.

2. **Histogram of TOEFL Scores**



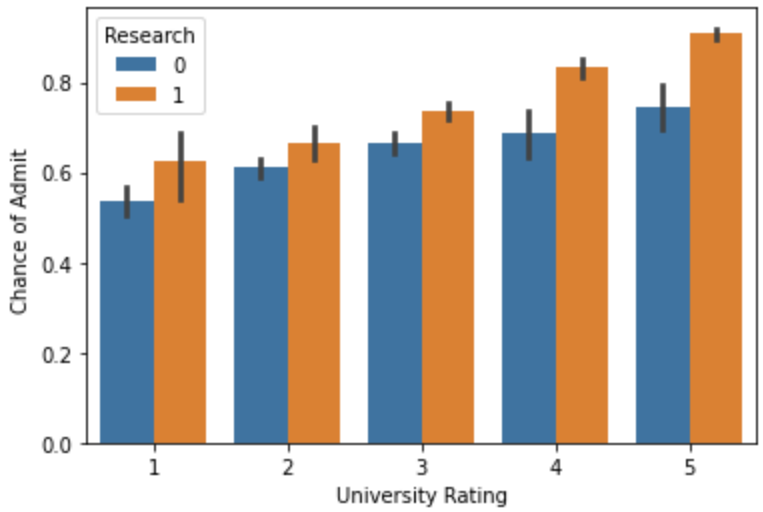
The above histogram shows that the score of most of the applicants fall in the range of 98 to 115. Similar as before, very few have score higher than 115 and lower than 95.

3. Histogram of CGPA of Applicants



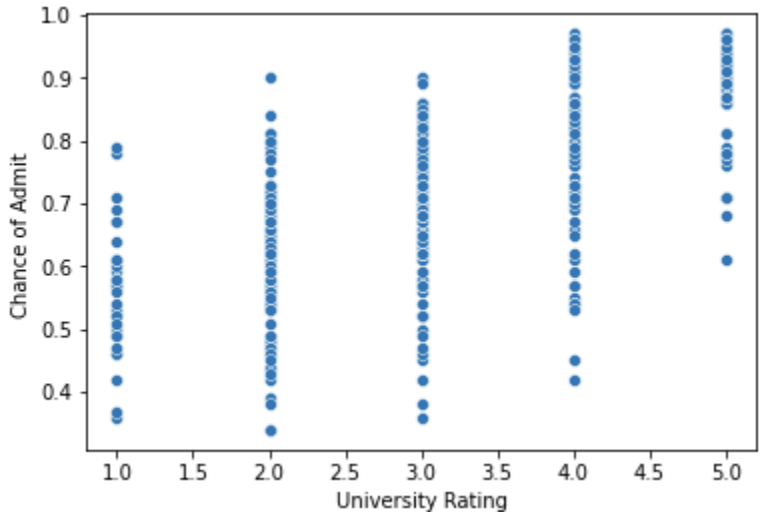
The plot shows that most of the applicants have CGPA in the range of about 8.00 to 9.25. Very few score below 8.00 and greater than 9.25 CGPA.

4. Influence of Research in Chance of Admit



This graph helps us to infer that the applicants that have conducted research in their undergraduate has a higher chance of admit in all the universities compared to when they did not conduct research. As we also see in the summary table above that the mean of research is 0.52 which tells us that around 250 applicants that is 50% have conducted research but same amount has not conducted a research project.

5. **Scatterplot of University Rating and Chance of Admit**

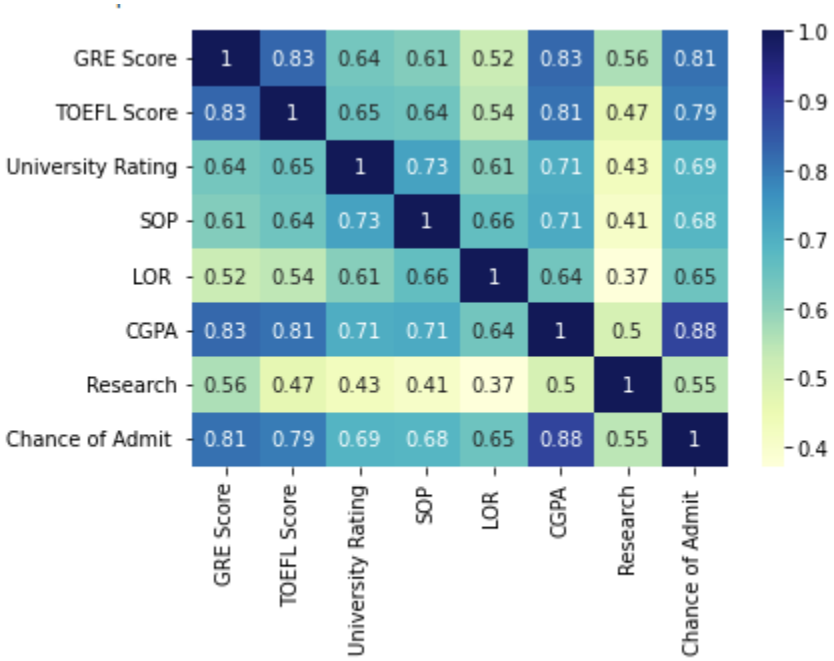


Correlation

According to me, it's very important to see correlation among the variables to give us a clear view in empirical analysis of data. The below table shows the correlation between all the variables in the data.

	GRE	TOEFL	RATING	SOP	LOR	CGPA	RESEARCH	CHANCE OF ADMIT
GRE	1.00	0.82	0.63	0.61	0.52	0.82	0.56	0.81
TOEFL	0.82	1.00	0.64	0.64	0.54	0.81	0.46	0.79
RATING	0.63	0.64	1.00	0.72	0.60	0.70	0.42	0.69
SOP	0.61	0.64	0.72	1.00	0.66	0.71	0.40	0.68
LOR	0.52	0.54	0.60	0.66	1.00	0.63	0.37	0.64
CGPA	0.82	0.81	0.70	0.71	0.63	1.00	0.50	0.88
RESEARCH	0.56	0.46	0.42	0.40	0.37	0.50	1.00	0.54
CHANCE OF ADMIT	0.81	0.79	0.69	0.68	0.64	0.88	0.54	1.00

Further, I made a heatmap which would help us to easily interpret the data.



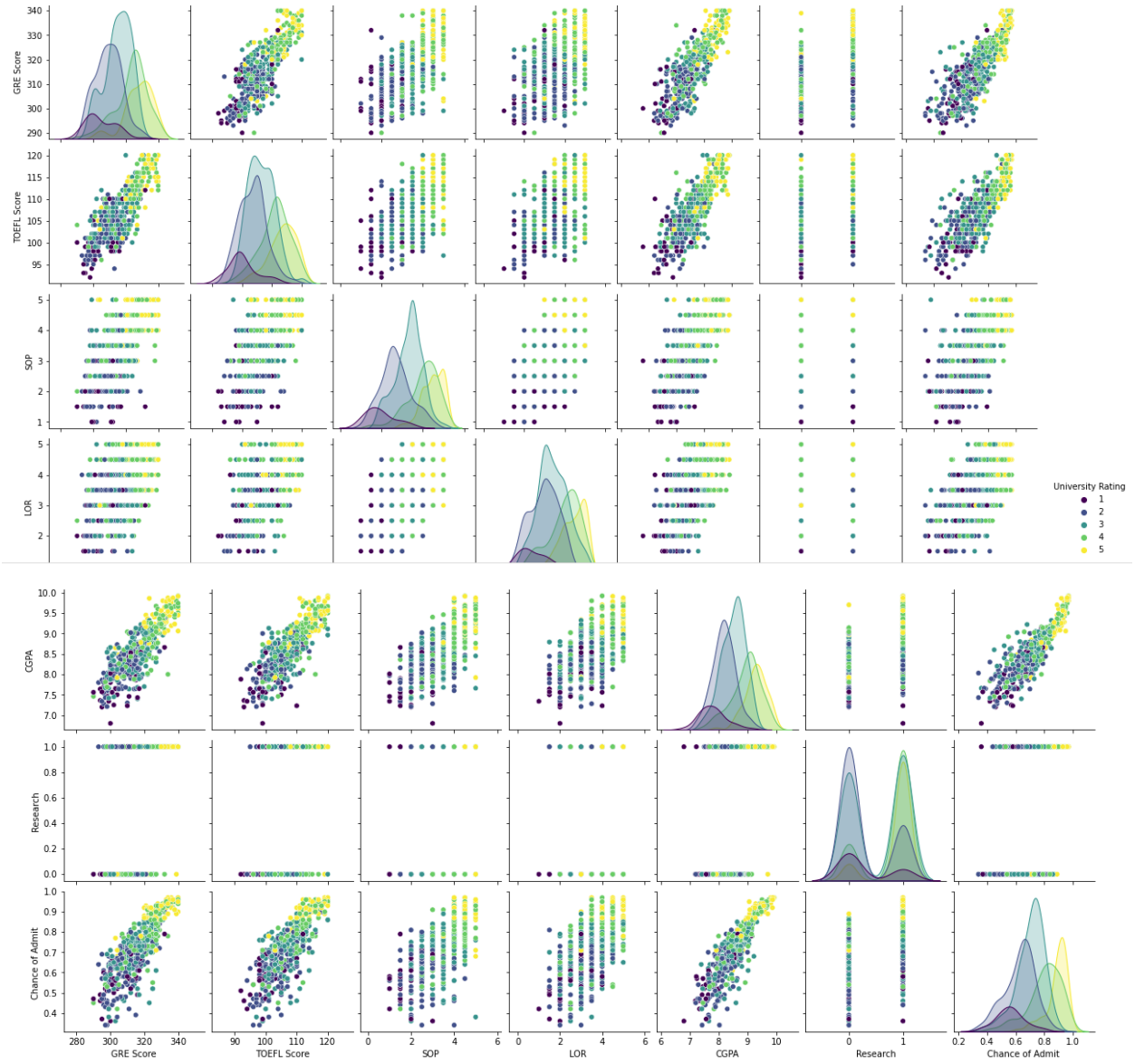
We can infer that:

- Chance of admit is weakly correlated with Research
- Chance of Admit is moderately correlated with University rating, SOP and LOR
- Chance of Admit is highly correlated with GRE, TOEFL and CGPA scores
- GRE Score is highly positively correlated with TOEFL score.

- CGPA is highly correlated with GRE and TOEFL score.

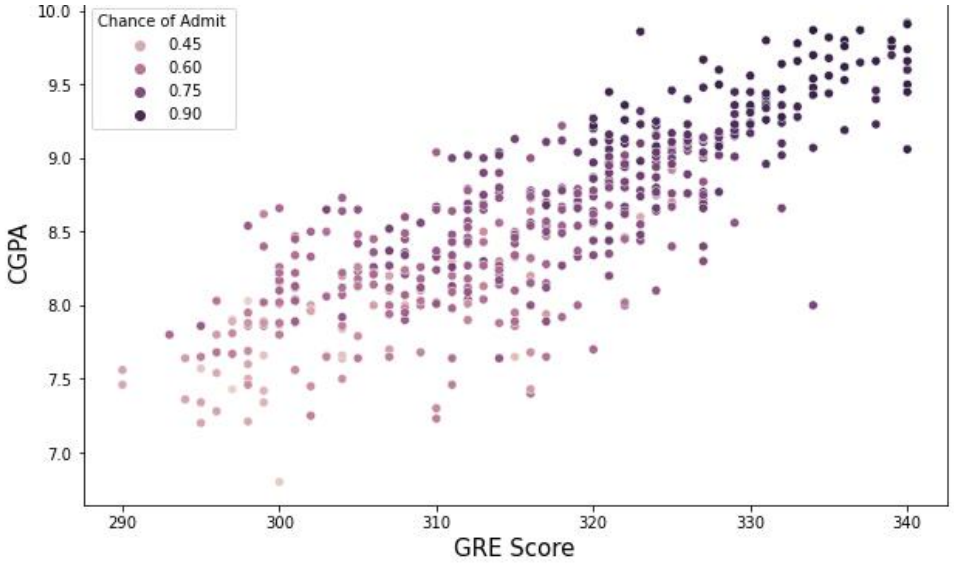
We can see that all 6 variables seem to have a linear relationship with the response variable. Most obvious ones are the relationship between the response and the CGPA, GRE and TOEFL scores. There is a strong correlation between CGPA and GRE with TOEFL score and the GRE and TOEFL score. Doing correlation helps in choosing the linear regression model next. We will see different graphs and make the best model.

Linear Regression



Here when we look at the pair plots, we can clearly see the relationship of chance of admit with the other variables.

Relationship between CGPA and GRE Score



This plot shows us that applicants with high GRE scores tend to get higher CGPA. When we compare it with the chance of admit, it shows that getting admit is higher when we have greater CGPA and GRE scores.

As clearly seen from the correlation between the variable and the dependent variable "Chance of Admit" is closer to one I considered all the variables together for my linear regression after conducting separate regression tests. To also improve the results, I also conducted an outlier test and removed the top 3 smallest columns that are (9, 65, 92).

My regression equation is as follows:

Chance Of Admit (Dependent Variable) = -1.301 + 0.002*GRE + 0.003*TOEFL + 0.006*University Rating + 0.004*SOP + 0.016*LOR + 0.118*CGPA + 0.019*Research

	Coefficient	**Std error**	**T**	**P value**
Intercept	-1.30	0.100	-13.049	0
GRE Score	0.0020	0	4.170	0
TOEFL	0.0026	0.001	3.155	0.002

University Rating	0.0057	0.004	1.571	0.117
SOP	0.0036	0.004	0.817	0.414
LOR	0.0157	0.004	3.985	0
CGPA	0.1185	0.009	12.853	0
Research	0.0194	0.006	3.077	0.002

R^2 = 0.836, adjusted R^2 = 0.834, F-Statistic = 356.4

I went ahead and did a simple interpretation of the results. As you see our r-squared shows that our results are almost 83% efficient. F-statistic is very high which shows that our model is correct and P values closer to 0 means we can reject the null hypothesis. As far as the coefficients are concerned because all the variables lead to increased factor of chance of admission. The positive value of coefficients interprets that the independent variables value increases than the mean of the dependent variable also tends to increase.

Conclusion

Personally, this data set was very helpful in working through a process of an econometrics analysis. The models contained a lot of heteroskedasticity, and I was able to work through various regressions to find the best relationship. Initially my question was that only GRE and CGPA mattered but as we went through the analysis strong multicollinearity amongst the variables showed that all the independent variables were important. When I had applied to the University of San Francisco for my sport management graduate program, I had given great attention to detail to all these variables to get into a very competitive program. Some of the limitations that I feel are that some data related to gender, which university they did undergrad from, and better input of SOP and variables would help in making a great analytical model for bigger research. However, the answer that I come to is that the better your scores and having a research thesis highly improves your chances of admissions with CGPA and GRE scores being the most relevant.

Citation

Acharya, M. (2019). Graduate Admissions. Retrieved April 26,2019, from Kaggle website: https//Kaggle.com/mohansacharya/graduate-admissions