Graduate Admission Prediction

Team 4:
Dhiab Aldhaheri
Dhruv Talwar
Sagar Bansal

Introduction of Data

- Data Source:
  https://www.kaggle.com/mohansacharya/graduate-admissions#Admission_Predict_Ver1.1.csv
- Independent Variables:
  - Quantitative variables: GRE score, TOEFL score, SOP strength, LOR strength and CGPA.
  - Qualitative variables: Research Experience, University rating
- Dependent variable: Chances of admit
- Data Characteristics:
  - GRE Score - ranges from 260 to 340, integer data type.
  - TOEFL Score - ranges from 0 to 120, integer data type.
  - SOP strength - ranges from 1 to 5, decimal data type.
  - LOR strength - ranges from 1 to 5, decimal data type.
  - CGPA - maximum 10, decimal data type.
  - Research Experience – 0 or 1.
  - University Rating - ranges from 1 to 5.
  - Number of observations: 400
  - Total features: 7

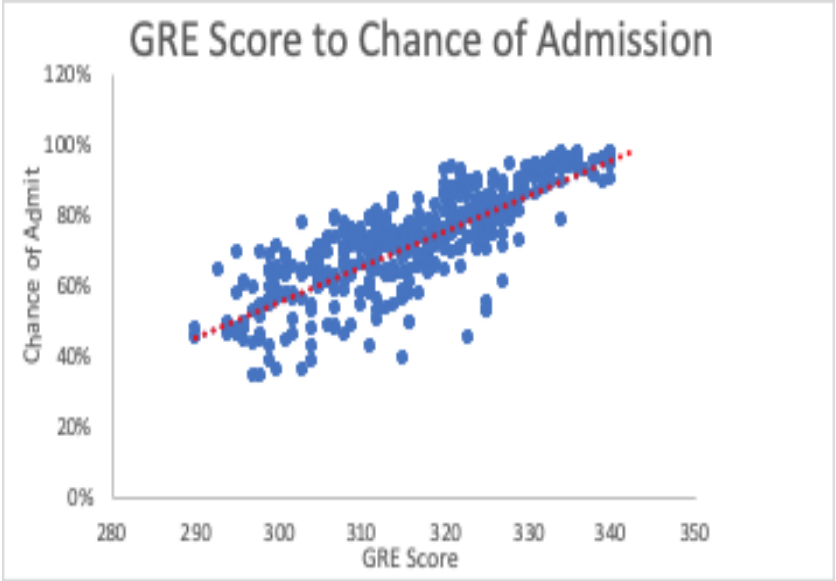
Model Prediction Precision

Prediction = lm(`Chance of Admit`~ 0 + LOR + CGPA +log(`GRE Score`*`University Rating`)).

- **Multiple variable Regression Analysis**
- **Regression through the origin**
- **Our model was able to predict 98% of the observed variation in the chances of admission of students**

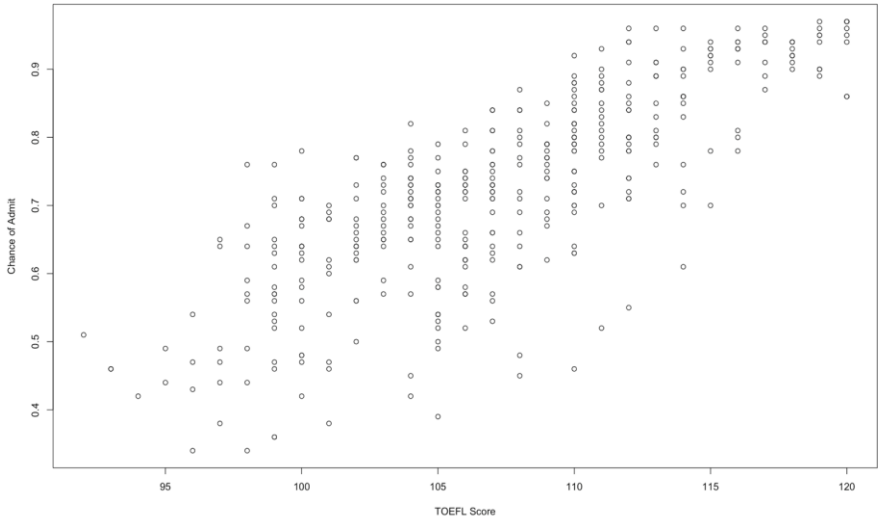
Building The Model

- Clear linear relation between GRE SCORE and Chance of Admission
- Correlation of 0.8026105
- Most Graduate School require GRE to admit students which justifies the correlation we see.



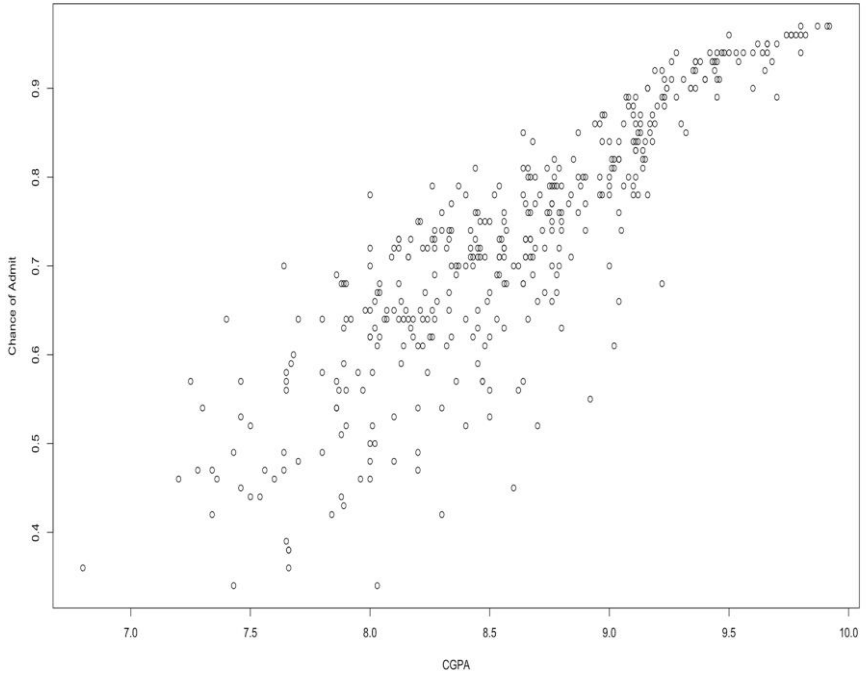
Building the Model

- Clear linear relation between the variables.
- Correlation of 0.79
- Drop this variable ? Why?



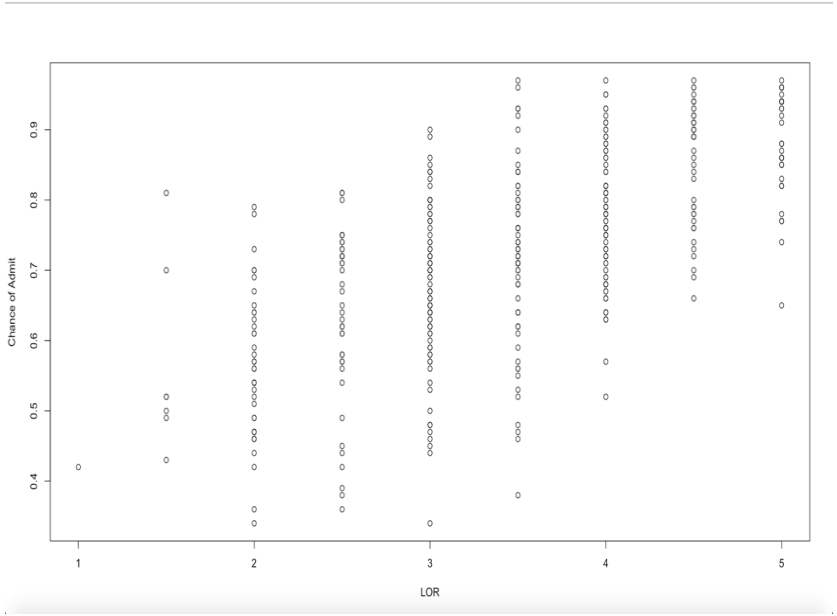
Building the Model

- Highest correlated variable to the Chances of Admission 0.88
- Clear Linear Relation
- We see a trend in the higher CGPA students
- For example, “We require students to have a CGPA > 9.0



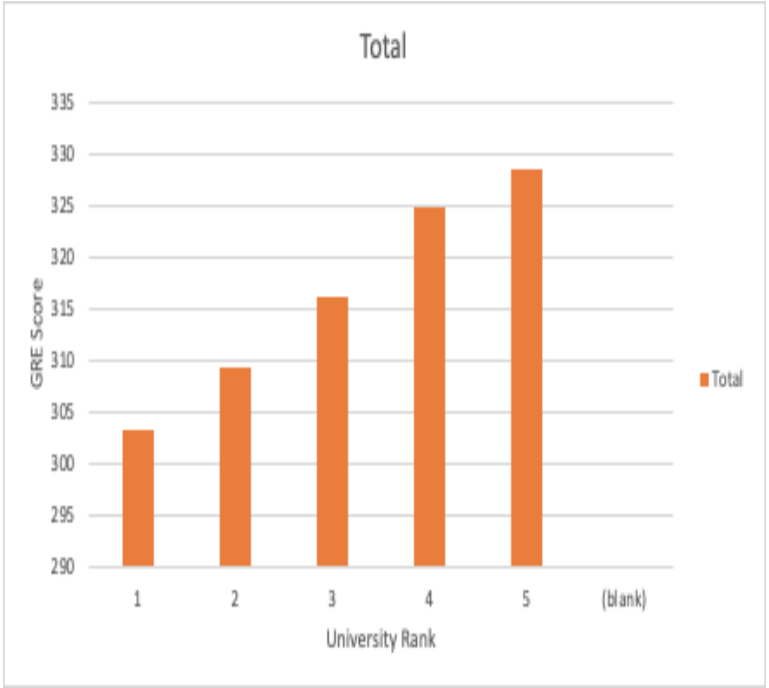
Building the Model

- LOR letter of Recommendation
- 0.67 correlation
- Highly useful predictor for our model



Interaction Term

- The higher the University Rating the higher the required GRE Score to achieve higher rates of Admission.
- Therefore, it made sense for us to add an interaction term.
- (University Rating * GRE Score)



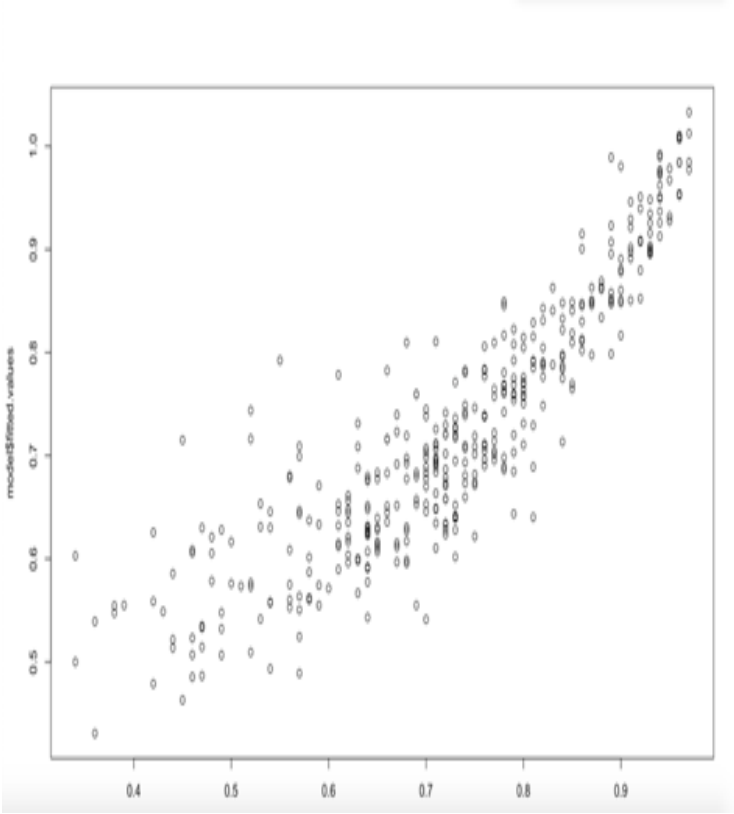
Dropping TOEFL Score

- Since this is graduate admissions, from our experience of applying to GRAD school we were not required to submit a TOEFL Score
- Moreover, Hypothesis testing proved it not as a useful predictor of our model.
- H0: $\beta_4 = 0$
- HA: $\beta_4 \neq 0$

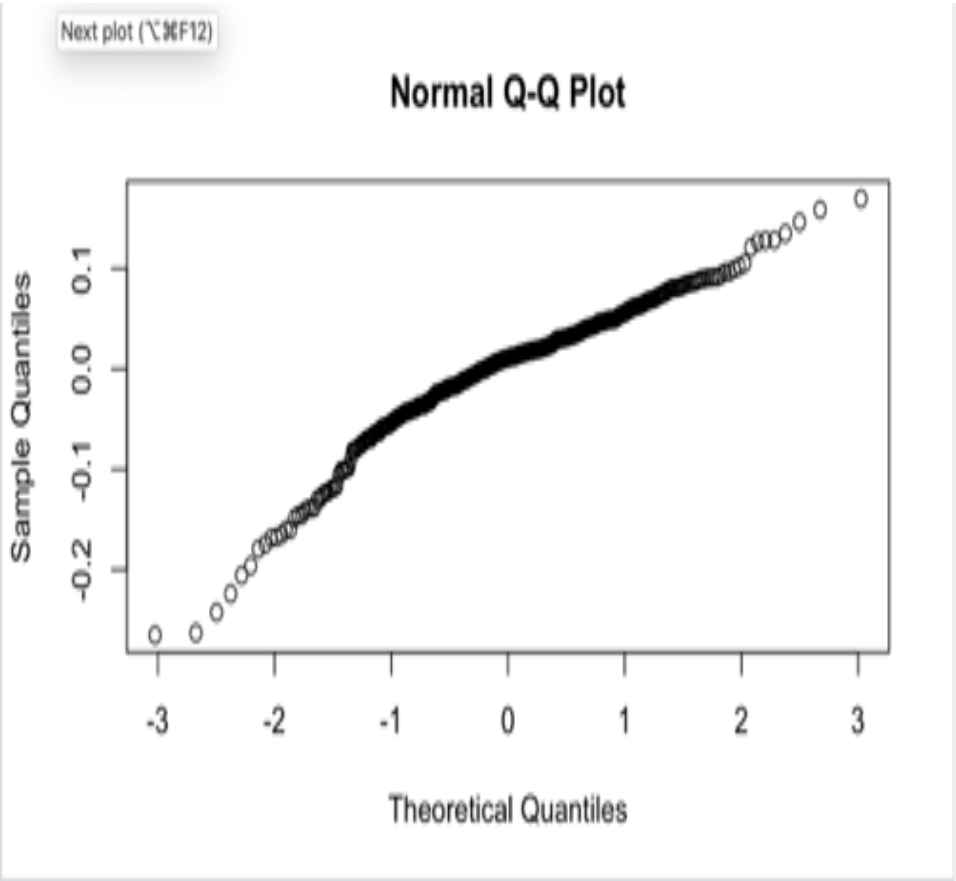
```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
LOR                                       0.064219   0.005926  10.836  < 2e-16 ***
CGPA                                      0.132731   0.014810   8.962  < 2e-16 ***
log(`GRE Score` * `University Rating`) -0.064326   0.015017  -4.284 2.31e-05 ***
`TOEFL Score`                             -0.001846   0.001188  -1.554    0.121
---
```

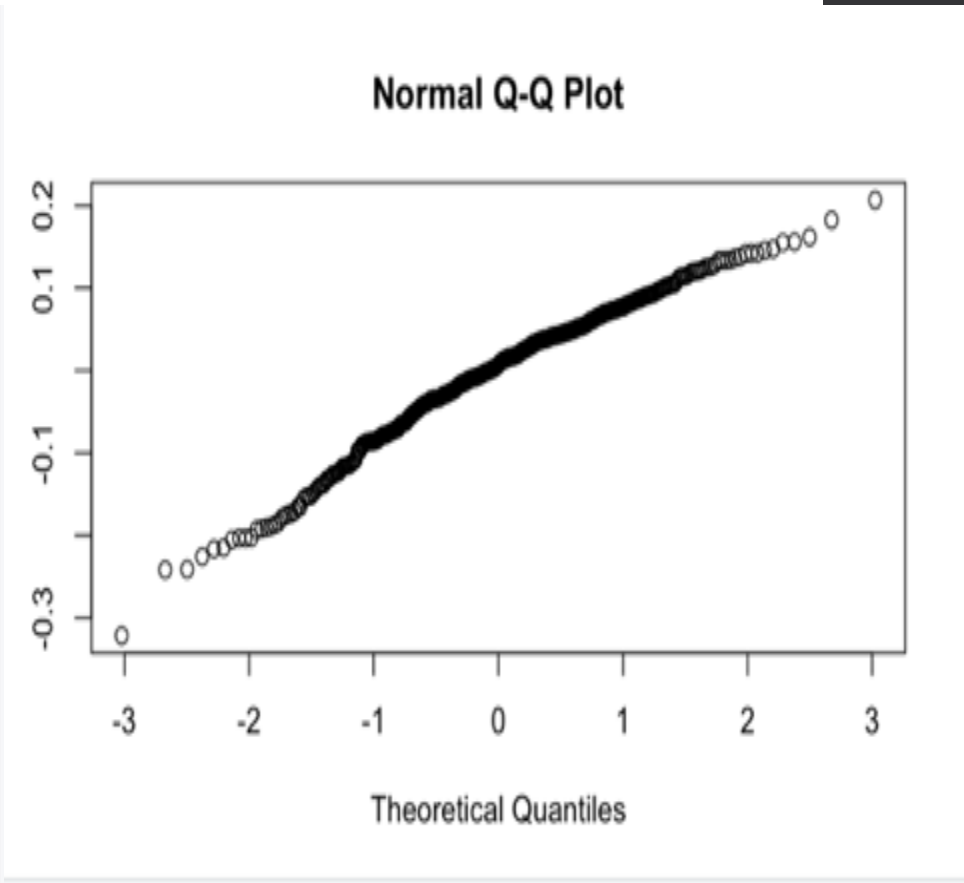
Heteroscedasticity (Model Assumptions)

- Model precisely predicts for higher Chances of Admission 80% and higher.
- Predicted errors have higher impacts on lower chances of admission
- How ambitious the student seems to the recruiter
- Uncaptured factors that contribute highly to the chances of admission
- Admission has a threshold for high scoring students
- Students with lower scores will be assessed on factors only the recruiter knows. eg: Legacy students, Donating Families.
- Recruiters decision differ from one another



Distribution of errors (log term)





Let us hypothesize the Interaction?

- Let us create a new term University Rating*GRE to conclude that when higher the GRE, better the chances of admit to a higher ranking university.
- We will proceed with the Hypothesis Testing to reach a conclusion
- For Hypothesis Testing, we will first define a Null & Alternate Hypothesis

H_0 = Interaction between UR & GRE <= 0
H_a = Interaction between UR & GRE > 0

For this hypothesis, we will use α = 0.05 = Significance Level

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	194508483	1	194508483	2580.00463	2.971E-252	3.85313813
Within Groups	60161818.1	798	75390.7495			
Total	254670301	799				

ANOVA!

H_0 = Interaction between UR & GRE <= 0

H_a = Interaction between UR & GRE > 0

For this hypothesis, we will use α = 0.05 = Significance Level

- Since p-value < α, we will reject the null hypothesis.
- This implies that when it comes to model the 'Chance of Admission', there exists some interaction between University Rating & GRE Score

- Which further explains that higher the GRE Score, more will be the 'Chance of Admission' to a higher ranking university.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	194508483	1	194508483	2580.00463	2.971E-252	3.85313813
Within Groups	60161818.1	798	75390.7495			
Total	254670301	799				

Exploring the Variables!

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change
1	.873[a]	.762	.762	.06964	.762
2	.884[b]	.781	.780	.06690	.019
3	.891[c]	.794	.792	.06499	.013
4	.894[d]	.799	.797	.06431	.005

a. Predictors: (Constant), CGPA

b. Predictors: (Constant), CGPA, GREScore

c. Predictors: (Constant), CGPA, GREScore, LOR

d. Predictors: (Constant), CGPA, GREScore, LOR, Research

- Stepwise analysis is a process of fitting regression model in which the choice of predictive model is carried out by an automatic process.
- CGPA, GRE Score, LOR, Research came out to be the most important independent variables which we can use to create a model.
- These 4 factors together explain ~80% of the variability in model.

Exploring the Variables!

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-1.331	.118		-11.307	.000
	GREScore	.003	.001	.207	4.816	.000
	LOR	.023	.005	.145	4.751	.000
	CGPA	.134	.011	.559	12.255	.000
	Research	.025	.008	.086	3.080	.002

a. Dependent Variable: ChanceofAdmit

- All the four independent factors have a positive relationship with Chances of Admission
- CGPA remains the most important factor to decide on the chances of admission followed by GRE Score, LOR and Research respectively.
- **Chance of Admit = 0.134*CGPA + 0.003*GRE + 0.023*LOR + 0.025*Research**

Confidence Interval

Table 5: Summary statistics of Graduate Admission

Features	Mean	Margin of error	Confidence Interval with 95% confidence level
GRE Score	316.8	1.12	[315.68, 317.92]
LOR	3.5	0.09	[3.41, 3.59]
CGPA	8.6	0.06	[8.54, 8.66]
Research	0.5	0.05	[0.45, 0.55]
Chance of Admit	72	1.37	[70.63, 73.37]

Source: Tab "Table 5" in the attached excel file

Interpretations:

- **GRE Score:**
  At 5% significance level, we can say that the average population value of the GRE Score will lie in between 315.68 and 317.92.

- **LOR:**
  At 5% significance level, we can say that the average population value of the LOR will lie in between 3.41 and 3.59.

- **CGPA:**
  At 5% significance level, we can say that the average population value of the CGPA will lie in between 8.54 and 8.66.

- **Research:**
  At 5% significance level, we can say that the average population value of the Research will lie in between 0.45 and 0.55.

- **Chance of Admit:**
  At 5% significance level, we can say that the average population value of our target variable (Chance of Admit) will lie in between 70.63% and 73.37%.

We noted that on average the mean of chance of admit varies between **70.63% and 73.37%**. This is **very important** because the range is **significantly far from 50%** i.e., many more than half of the students have a good chance to get an admit from the university they apply and they actually get it. This information can also be used by the future applicants as a **motivation** to apply and expect to get an admit.

Descriptive Statistics:

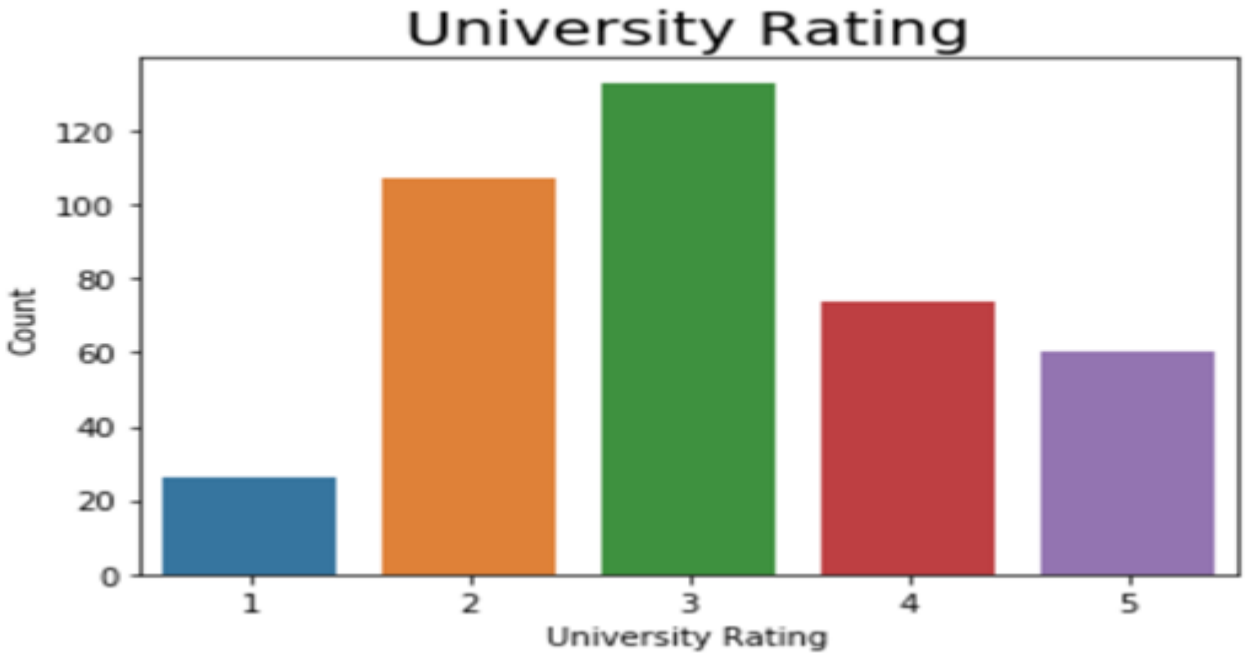
Table 1: Summary statistics of Graduate Admission

Measure	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
Average	316.8	107.4	3.1	3.4	3.5	8.6	0.5	72%
Median	317.0	107.0	3.0	3.5	3.5	8.6	1.0	73%
Standard Deviation	11.5	6.1	1.1	1.0	0.9	0.6	0.5	14%

Source: Tab "Table 1" in the attached excel file

- We can see that the average of the GRE, TOEFL score and CGPA is considerably more than the possible mid value i.e., 300, 60 and 5.0, respectively.
- The median and mean values are very close to each other for all of the features except Research.
- Standard deviation is relatively more for GRE and TOEFL score compared to all other features.

Histogram: University Rating



Universities with rating three gets the maximum number of student applications. We will explore more into the three and five rated universities and the typical student profiles needed.

Table 2: Average profiles with highest and lowest chance of admits for 3 and 5 rated universities.

University rating	Chance of Admit	Average of TOEFL Score	Average of CGPA	Average of LOR	Average of SOP	Average of GRE Score	Average of Research
3	90%	114	9	5	5	330	1
3	36%	99	8	3	2	303	0
5	97%	120	10	4	4	337	1
5	90%	113	9	4	4	331	1
5	61%	108	8	3	3	305	0

Source: Tab "Table 2" in the attached excel file

- Typical student profiles for 90% chance of admit matches for 3 and 5 rated universities.
- When the student has no prior research experience, the chance of admit gets hugely affected.

Conclusion:

Based on the above findings, or the analysis that we have done/presented, we can write the infer the below mentioned info:

- We are 95% confident that the mean of all chance of admits lies between 70.63% and 73.37% i.e, many more than half of the students who apply get an admit.
- Using ANOVA, we concluded that higher the GRE Score, more are the chances of a student getting admitted to a high university ranking!
- CGPA, GRE Score, LOR, and Research comes out to be the most important factors on which the chances of admission depend.
- Since all of these have a positive relationship with the Chance of Admission, higher the independent factors, more will be the Chance of Admission.

Time
For
Some Questions