

Module 4 Presentation

Will you get into UCLA?



// FLATIRON
SCHOOL

Project Scope

Develop a predictive model of chance of admission into a graduate program at the University of California Los Angeles (UCLA)

Background

UCLA is one of the nation's leading public universities with 121 programs

Objective

Implement an easy-to-use model to determine a student's chance of admissions at UCLA

Approach

Predictive linear regression of admissions based on UCLA admission data

Data overview

The dataset contains *Chance of Admission* and 8 student metric variables

Descriptive stats:

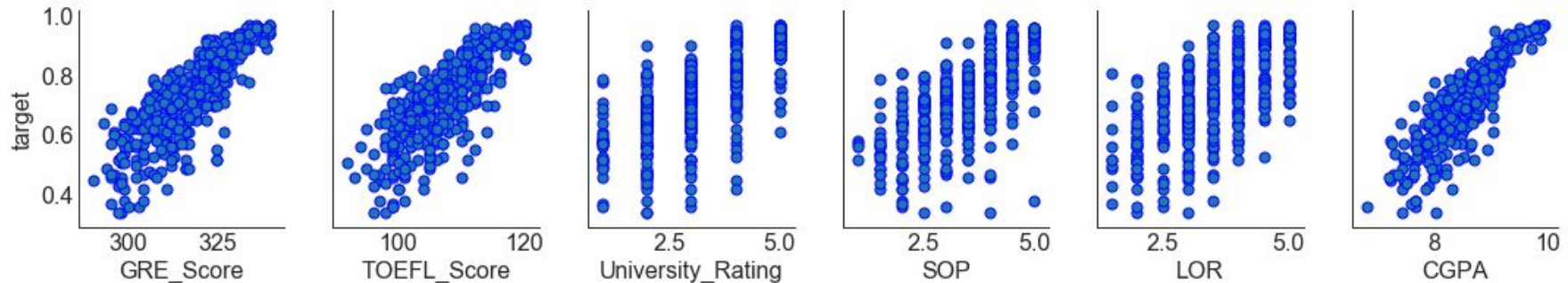
- Data from Mohan Et al on UCLA admissions available on [Kaggle](#)
- Dimensions: 500 rows, 9 columns
- Available Parameters:
 - Chance of Admission (target)
 - Cumulative GPA (CGPA)
 - TOEFL score
 - GRE score
 - Statement of Purpose quality
 - Letter of Recommendation quality
 - Research (binary)

	count	mean	std	min	25%	50%	75%	max
GRE_Score	500.0	316.47200	11.295148	290.00	308.0000	317.00	325.00	340.00
TOEFL_Score	500.0	107.19200	6.081868	92.00	103.0000	107.00	112.00	120.00
University_Rating	500.0	3.11400	1.143512	1.00	2.0000	3.00	4.00	5.00
SOP	500.0	3.37400	0.991004	1.00	2.5000	3.50	4.00	5.00
LOR	500.0	3.48400	0.925450	1.00	3.0000	3.50	4.00	5.00
CGPA	500.0	8.57644	0.604813	6.80	8.1275	8.56	9.04	9.92
Research	500.0	0.56000	0.496884	0.00	0.0000	1.00	1.00	1.00
Chance_of_Admit	500.0	0.72174	0.141140	0.34	0.6300	0.72	0.82	0.97

Data Overview

Student metrics include 3 continuous variables and 3 categorical variables. All student metrics are correlated with the target.

Scatterplot of Student Metrics against *Chance of Admission*

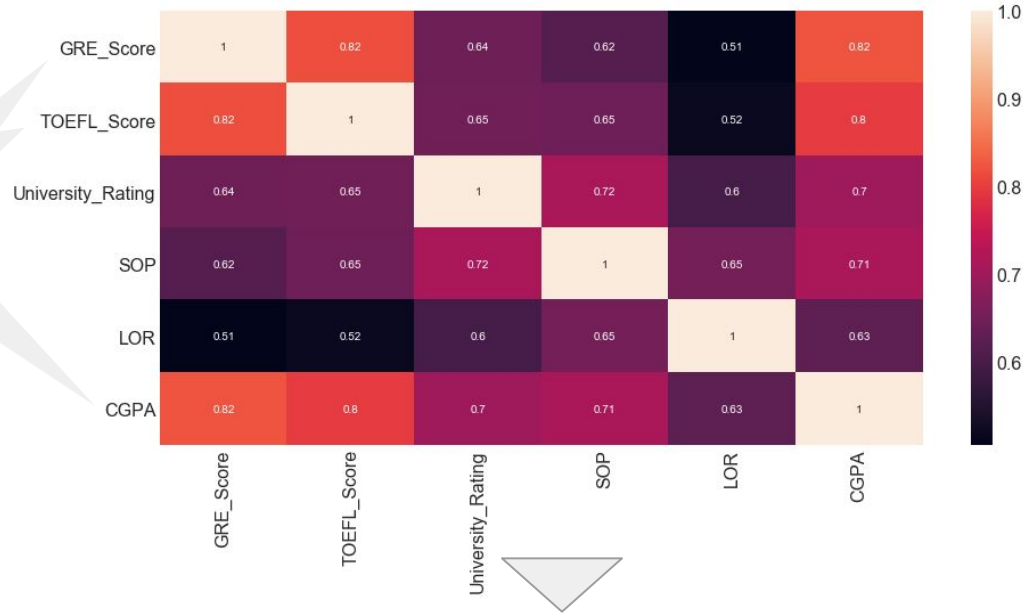


How do these student metrics interact with each other?

Data Overview

3 variables - GCPA, TOEFL, and GRE are closely correlated with each other however it may still to be worthwhile to keep these features

GRE, TOEFL and GCPA are closely correlated with one another



We should explore the value of these three variables to our model

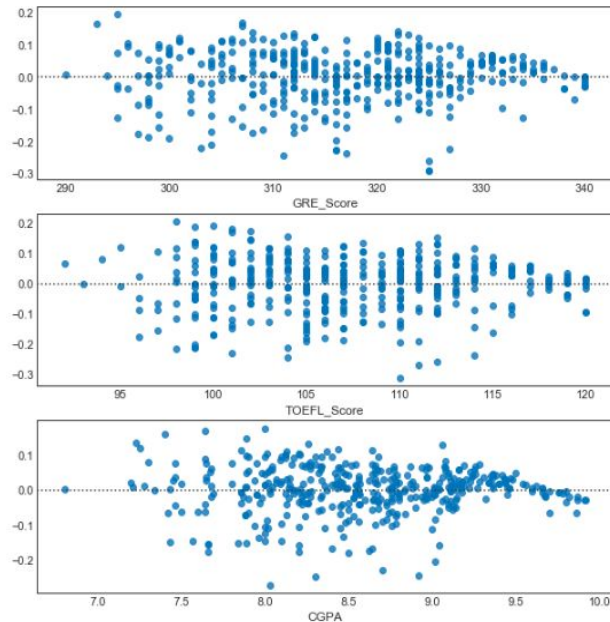
Model

Base model has an R^2 of 0.82 indicating strength in prediction however the residual distribution suggests that further variance can be explained

```
explained_variance: 0.8163  
mean_squared_log_error: 0.0014  
r2: 0.8163  
MAE: 0.0431  
MSE: 0.0036  
RMSE: 0.0602
```



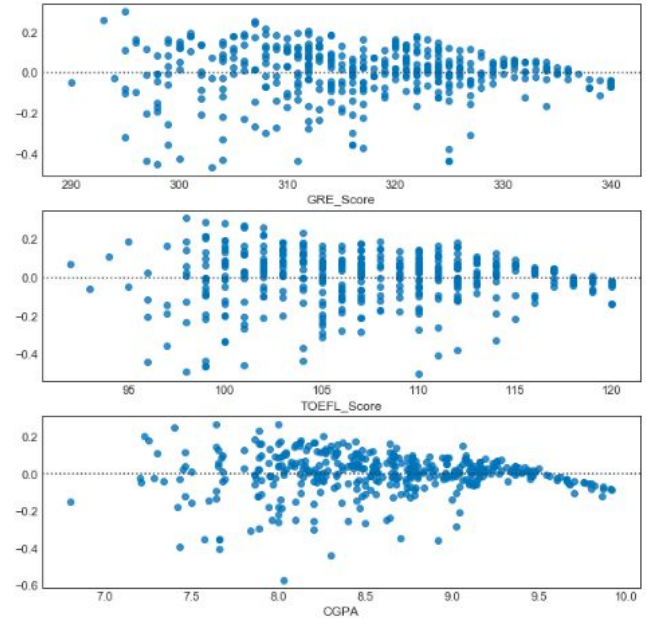
What transformations will allow further explanation of variance?



Model

We tried inverse transforming the target variable led to a lower R^2

```
explained_variance: 0.7591  
r2: 0.7591  
MAE: 0.071  
MSE: 0.0107  
RMSE: 0.1032
```

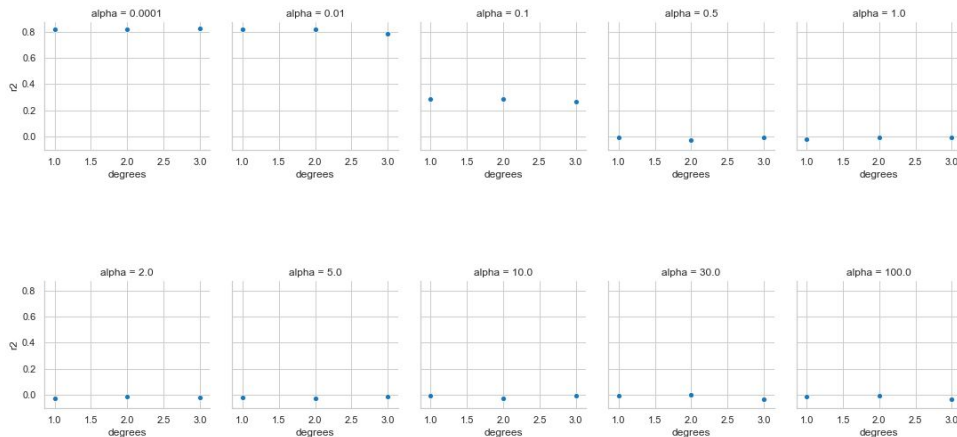


Log transforming the target variable...

Model

Using higher order features does not substantially improve the estimation of Chance of Admission

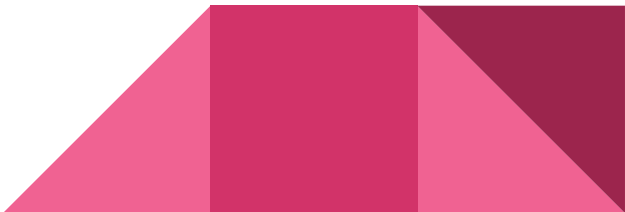
degrees	alpha	r2
3	0.0001	0.827762
2	0.0100	0.820106
1	0.0100	0.817674
2	0.0001	0.817327
1	0.0001	0.816461



We may want to explore other transformation

Next steps

Further areas of study

- Diversity: Look at alternate models for student assessment to find niche students not captured in the traditional model.
 - Pedagogy: What best predicts students' outcomes on the key student metrics? How can you improve GPA, GRE and TOEFL?
 - Outcomes: Is UCLA relying on the right student metrics? Do these metrics predict lifelong success?
- 

Model

Check out your chances of
gaining admission to UCLA!

Will you get into UCLA?

Student Admission Weights	
Metric	Weight
Base score	-1.587
GRE	0.00302
TOEFL	0.00286
CGPA	0.11269
SOP	0.00204
LOR	0.01732

Data Overview

Students with high CGPA also have high TOEFL and GRE Scores. This may be a question for further study.

