Allyson Busch

5/29/2019

DSC 530

Catherine Williams

DSC 530 Final Project

**Hypothesis:**

If a prospective student wants to be admitted to a graduate school program, then they should focus on raising their GRE score because this has the highest effect on chance of admittance.

**Outcome of EDA:**

A student’s GRE score does not have the highest correlation when comparing all the variables of the study and it is not the largest coefficient in our regression analysis. The correlation coefficient is 0.80, which is a high correlation coefficient, but there was one value that was higher. GPA, with a correlation coefficient of 0.87, had the highest value in the analysis.

Because of the correlation coefficient evaluation, one can conclude that the GRE score is not the most important variable in getting into a graduate school program. Instead, the GPA of a student has the highest correlation and the largest coefficient in our regression, meaning a student’s GPA has the highest effect on a student’s ability to get into a graduate program. GPA had the largest coefficient in the regression analysis by far; GPA had a coefficient of 0.143109 whereas GRE had a coefficient of -0.000118.

**What was Missed/Variables that Could Have Helped:**

I feel the research may be more accurate if it included data from students in a wider geographic region. The dataset included only students from India, and it was a relatively small dataset. A more diverse group of students and a larger collection would likely yield better results, and make the data more useful as a tool to predict application outcomes. As it is, the model may not be entirely accurate as international admissions to graduate schools in the United States can be different from domestic admissions.

Additionally, I would have liked variables that explained the quality of a student’s research experience, if a student had professional experience, the age of a candidate, and the university rating of the student’s undergraduate school. While I am not sure these would have large effects on the chance of admittance of a student, I would be interested in seeing their correlations and it could result in a more accurate model of prediction.

**Challenges:**

I did not run into many problems with the calculations for the data, other than having to research how to do several of them in Python rather than R. I feel some of the EDA would have been easier to do in R, but completing them in Python was a good exercise to complete.