Limitations

Benjamin Lordi

9/14/2025

Modern Software Concepts in Python

There are a number of inherent challenges to analyzing anonymously submitted data. One of the largest is that user submitted data is extremely prone to selection bias. The type of person that submits detailed data to a website, in the case of the GradCafe, application data, can be very different from the rest of the population. The data that comes from these sorts of submission can be heavily biased and not generalize well to the overall population. In the case of the GradCafe, people who submit data may have higher scores they want to show off, be more tech literate, or have a very positive or negative experience with their applications. If trying to generalize this data to the population of all applicants, these effects may skew the data to be unrepresentative and lead to false conclusions. Another challenge is data quality. There is no way to verify the accuracy of the information users are submitting, and no way to filter out fabricated data. User submitted data is also prone to input mistakes or omitted data. Data quality issues can make it difficult to clean and analyze the data, and if a high proportion of the data is fabricated or entered incorrectly, it might be impossible to create meaningful analysis.

The results from the analysis of the GradCafe are surprising to me. The GPA and GRE numbers are exceptionally high, much higher than I would have suspected before doing this project. My guess is that those numbers are inflated by the users of the GradCafe. I would expect that people that post to the GradCafe have higher than average scores compared with people who do not. The incentive to post a bad score is low, even if it did lead to an accepted application. If any part of the reason for posting on the GradCafe is to show off, then higher scores are to be expected. I would also guess that the type of person who is so invested in post bachelor's education that they are aware of the GradCafe and are an active user, is also the type of person who takes their test scores very seriously and would strive to score very high on any GRE.