

=====

Modern Concepts in Python: Spring 2026

Module 3 Assignment: Database Queries & Analysis Dashboard

Author: Eric Rying

File: limitations.pdf

=====

Assignment: Having carried out this assignment, please write two paragraphs about the inherent limitations of carrying out analytics over anonymously submitted data items. Did the analytic responses surprise you? How does this differ from standards? For example, the average GRE quantitative reasoning score was 157 for 2023-2023 and was nearly 165 for grad school entries submitted (see sample output). Why do you think that is? What might cause this to occur? Please place your essay into a file called limitations.pdf

Answer:

Analytics over anonymously submitted data is inherently constrained by the quality, consistency, and completeness of the information people choose to provide. Because submissions are voluntary and unverified, key fields such as GPA, GRE scores, citizenship, or admission outcomes may be missing, self-reported inaccurately, or formatted inconsistently. This creates gaps that limit the reliability of aggregate statistics and can skew averages or percentages when only a small subset of users provide usable data. Without a mechanism to validate entries or enforce required fields, the dataset will naturally contain noise, outliers, and contradictory values that reduce confidence in any conclusions drawn from it.

I was not surprised by the lack of GPA and GRE data submitted by folks using the GradCafe site as I imagine they either didn't want to share this private info or simply did not take the time to enter them. This contributes to substantial sampling bias. I think the relative difference between 157 vs 165 is not that significant, one would need to consider the error bars, population size and perhaps perform a statistical ANOVA analysis to determine whether such a difference is statistically significant and therefore meaningful. There could also be some intrinsic bias in the grad students submitting scores to the site

that may drive the average up, if indeed it is statistically difference. For example, users may be more likely to report high GPAs or acceptances to prestigious universities, while others may avoid reporting rejections or weaker profiles for emotional reasons such as pride. This self-selection bias means the dataset cannot be assumed to reflect true admissions trends, acceptance rates, or applicant demographics. As a result, analytics performed on anonymously submitted data should be interpreted as informal, community-generated insights rather than statistically rigorous or generalizable findings.