Analyzing anonymously submitted data introduces several inherent limitations that directly affect the reliability and interpretability of the results. Because contributors' self-report their information, there is no mechanism to verify accuracy, completeness, or consistency across entries. Applicants may omit fields, exaggerate their credentials, or use nonstandard terminology, and these inconsistencies propagate into the dataset. Even with LLM-generated normalization fields, subtle variations in program names, university names, or degree types can lead to undercounting or misclassification when performing exact-match queries. Anonymous datasets also tend to suffer from selection bias: only a subset of individuals choose to report their outcomes, and those who do may not be representative of the broader applicant pool. As a result, the analytics reflect the behavior of the reporting population and not the true population.

These limitations become especially visible when comparing the dataset's statistics to official standards. For example, the average GRE Quantitative score for the 2023–2024 applicant pool is around 157 nationally, yet the GradCafe submissions in this dataset cluster nearly 204. This discrepancy is not surprising once we consider who is likely to self-report: applicants with exceptionally strong scores, applicants targeting highly competitive programs, and applicants who are more engaged in online admissions communities. Lower-scoring applicants are less likely to publicly share their results, creating an upward bias in the reported averages. Additionally, GradCafe disproportionately attracts applicants in quantitative fields like computer science, engineering, and mathematics, where GRE scores tend to be higher than the general test-taking population. Together, these factors create a dataset that is informative for understanding trends among self-selected contributors but not representative of the full graduate admissions landscape.