

### Limitations

Analytics performed on anonymously self-submitted data sources such as Grad Café are typically going to be exposed to several inherent limitations that affect reliability and representativeness. Submissions are voluntary, the dataset will naturally suffer from selection bias. Applicants who choose to report their results are not necessarily representative of the broader applicant population. People with ‘better’ results, like higher GPAs, or successful admissions, may be more likely to submit their data, and those with rejections or lower GPAs may choose not to. Also, self-reported values such as GPA, GRE scores, and citizenship status are not verified. This introduces potential inaccuracies, exaggeration, or inconsistent formatting. Missing data is also a major issue, not all entries include every metric we evaluate, and so metrics like GRE were largely absent which limit the conclusions that can be drawn from those variables.

Some of the analytic results did differ from official statistics, as expected and shows these limitations in effect. For example, standardized reports indicate that the average GRE quantitative reasoning score in 2023 was 157, but Grad Café submissions report 165. This discrepancy likely reflects the selection bias/reporting bias as mentioned above. Applicants with stronger profiles are probably more motivated to share their results publicly. Also, Grad Café aggregates applicants across institutions, disciplines, and applicant pools without normalizing the data. This is unlike official reporting sources that enforce standardized definitions and sampling methods. Overall, all of these combined factors show us that data that stems Grad Café, or any form of self-submitted platforms, should not be interpreted as a statistically representative view of graduate admissions outcomes.