<u>Data Limitations</u>

When conducting analytics on anonymously submitted data items, there are unavoidable limitations that affect the accuracy and representativeness of the results. Anonymous submissions mean that the identity of the contributor cannot be verified, making it impossible to confirm whether the data is authentic, duplicated, or even fabricated. Important demographic details such as age, country of origin, or academic background may be incomplete or inconsistently reported, which further reduces the reliability of conclusions. This lack of verifiable structure creates challenges in ensuring that the dataset accurately reflects the real applicant pool, since what is collected is a self-selected sample of those motivated to post, not a controlled or audited population.

The analytic responses, while interesting, must therefore be interpreted cautiously. For example, it is not surprising that the average GRE Quantitative Reasoning score reported on Grad Café entries is higher than national averages published by ETS. This discrepancy likely arises because applicants with especially strong profiles may be more willing to share their scores publicly, while those with lower scores may withhold them. In other words, the dataset skews upward due to self-selection bias. This is a significant departure from official standards, which are based on all test-takers rather than a subset. The observed differences, such as an average GRE Q score of nearly 165 in Grad Café entries compared to the national average of 157, illustrate how voluntary, anonymous reporting can distort results and highlight the need for cautious interpretation.