

# Jon Robert N. Boland – Module 3

## SQL Data Analysis

The primary limitation of anonymous datasets is the skewed results. People posting anonymously are more likely to want to share good news and high scores, which may explain why the grad school entries have higher average scores than the overall population. Posters may also use the anonymous platform to settle a score, potentially by putting in bad data, or writing a comment discouraging other applicants from pursuing a program that rejected them.

Another limitation is the lack of standardization across the data set. Gradcafe doesn't require many fields outside of school, semester, and program, so many applicants may choose to leave out certain data points. This leads to inconsistent data, which is one of the reasons we were given this data set to clean up! Two analytic responses surprised me: the average GPA of 3.74 and the number of JHU Computer Science applicants (13). The average GPA seems high to me, and I thought the JHU applicant numbers would be higher!