There are certainly strong limitations with conducting analytics on a dataset of user-provided information. The two strongest limitations are in the form of poor data quality and biases. First off, if the data of interest is collected in the form of text fields where there is no limitation on the data value, it dramatically increases the chances of typos, misspellings, missing data, etc. For example, the module_2 data set from Gradcafe could contain 20 unique strings for the 'university' field when they were all referring to the same university. There are also cases of entries where the data is fake or misleading. For example, there was one such case where someone put in a fake university name and results just to leave a comment complaining about the types of comments people were making. This leads to greater variance in the aggregate data, making it much more difficult to derive meaning from the data analysis. As a solution, when collecting data from user input, forms should be designed to limit the variance in the input (ex: use a dropdown list of universities rather than have the user enter the name).

Another limitation with deriving meaning from user-entered information is that it can skew data in favor of a bias that a third party would not have. When reporting their own results, people are more likely to inflate their scores and acceptances. This gets amplified when they are able to remain anonymous. An anonymous user is far more likely to post incorrect or inflated information since there is no way for someone to fact check them. Also, users that perform better are more likely to report their results to others when compared to those who did not do well. This is the most likely reason why the average GRE score was 165 for applicants on GradCafe, but the actual average GRE score for 2023 was 157.