**Having carried out this assignment, please write two paragraphs about the inherent limitations of carrying out analytics over anonymously submitted data items.**

I found that the data quality is variable and doesn't properly confirm to a typed schema which enables me to perform analytics out of the box. I needed to create handlers for incorrect formats, formatting errors etc.

Anonymity also provides users with the opportunity to introduce bias to skew the results and there isn't a good way of determining which records are legitimate.

**Did the analytic responses surprise you? How does this different from standards? For example, the average GRE quantitative reasoning score was 157 for 2023-2023 and was nearly 165 for grad school entries submitted (see sample output). Why do you think that is? What might cause this to occur?**

The scores appeared higher than I would have expected and is also higher than the average that I looked at when doing a quick internet search. I think this may be due to the nature of self-reporting, caused by people being less likely to post lower scores. It may also be that certain years have fundamentally different application pools to analyse and thus it is more difficult to compare longitudinally. Grad school entries may also skew toward STEM fields where quantitative reasoning scores are higher, so that might be a contributing cause as well.