

# STA363 Client Challenge 1

Alex Zhang

Jan 2024

## Part 1

I will not recommend just removing 71 students from the data and proceed. The first thing is that by deleting these students, we also reduced our sample size by about 11%. Generally we should keep every student because everyone in this sample could mean certain pattern in the population. The second thing is that we may lose the representative nature of our sample. We may not be able to capture more patterns in this sample to represent the whole population. The third thing is that I think this data is missing for a reason. Some teachers prohibit ChatGPT for homework but some students may still use it anyway. They will intentionally leave the question blank because they do not want to be caught. We cannot just delete these students because maybe they actually use ChatGPT a lot. Deleting the data will underestimate the number of times students have used ChatGPT.

## Part 2