We expect the reports that answer our questions for the data analysis project to be in the "*AFYD*" format, which comprises 4 sections:

D - Do
Y - Why/Reasoning
F - Find
A - Answer

So please respond to each question by stating the following 4 points explicitly:
1) **D**: What did you do - what assumptions/choices did you make, if any (e.g. re independence, type of data, etc.)? [Doesn't have to be an essay - one high quality sentence is sufficient for everything you did and every choice you]
2) **Y**: Why did you do this - what was the reasoning for doing this (over all other possible things you could have done)? [Again, pithy is key, but when you make an assumption or choice above, it would be great to know what you were thinking when doing so]
3) **F**: What did you find doing that - at least one number (e.g. a p-value or an effect size), better a number *and* a figure. Context is better – what does the number represent, how was it arrived at (what was the test statistic and degrees of freedom?)
4) **A**: Given all of that, how do you answer the question, sth like "Given that my p value is x, I conclude that y]. If you made any questionable assumptions above, this is also where you might want to state limitations to the scope of your answer, if there are any, and you are aware of them. It's fine to have limitations, if you spell them out (e.g. concerns about alpha inflation, independence, distribution of data, etc - if there are any).

Some notes on this approach:

- If you are at all in doubt, add your code as an appendix at the end, so we can assess if a discrepancy is due to a bug or a deeper problem.
- We will grade for what makes sense. Because that is what the job will be, in Data Science. In other words, there are many approaches that are obviously wrong (e.g. using an obviously wrong test, using totally wrong df, etc.), but also many choices that are reasonable, given certain assumptions. So be reasonable. And spell out your reasoning. So other people can assess if you were reasonable.
- A lot of students find this approach cumbersome, but it has to be done like that. Why? Because the data doesn't speak for itself. Neither do the results. You have to interpret them in light of the (usually many) choices and decisions you made as to how to process/analyze the data. That's what infuses the numbers with meaning and leads to actionable insights. That's not easy to do actually. We're aware of that. But that's what the job is – to a large extent.
- Many former students have confirmed that making these (many) choices in a reasonable way, justifying them and then interpreting the findings in light of these analysis choices is indeed a big part of the job. Because this is so hard, few people can do it well. And that is why Data Scientists can command such high salaries - high demand for a skillset in combination with very limited supply leads to a high value. Calling the actual functions in Python is a very small part (and frankly not very differentiating) part of the job. Millions of people can do that. So that's not very valuable by itself. Coding is necessary, but ultimately just a means to an end. The end? Meaning.
- If you need an additional page to make your case, that's fine too.