

1 Project Description for DSCI401

You must find a sufficiently interesting dataset(s) and then do something sufficiently interesting with it using skills that you have learned in this class. You can do this in EITHER python or R, but you do NOT need to do this in both languages. Results should be presented in a **well-written** report that includes an introduction, a description of the data, the results, and all of the code to generate the results (i.e. everything should be reproducible).

Generally this should be 4-5 pages with the following,

- Describe the data well and the context of the dataset, setup your questions for the dataset
- Give a good exploration of the dataset, tables/summary statistics/etc.
- At least 1 very good graphic, more are possible
- Some added spice analysis, something a little extra beyond exploring the dataset

2 Format of Report and Rubric

1. (10 points): Introduction: Describe the question you are interested in answering and the data set.
2. (25 points) Methods: Here you describe the details of the methods that you are going to use to solve your problem. If there are formulas involved in your problem, this is where they usually go. You could also choose to put your code in this section (though I generally prefer to put the code in an Appendix).
3. (25 points) Results: Describe all of the interesting results that you have found. This is also where you would put tables and graphs with your results. (Tables are OK, but, imho, figures are better!). Make sure you describe in detail what is in each table and figure.
4. (10 points) Conclusions / Future Work: Briefly restate what you did and then describe the results. It is also good to mention in this section other work that you would do in the future if you had time. This section is almost exclusively text.
5. (30 points) Appendix: This is where your code should go. I should be able to take this code and run it and get ALL the results that appear in your report. Remember to use good style when writing your code (<http://adv-r.had.co.nz/Style.html>).

3 Timeline

1. **October 6th 2024** - Turn in dataset(s) for approval with any information and potential ideas for what to use the dataset for. (25pts)
2. **November 17th 2024** - Detailed project write up with plans for specific analysis or points of interest to investigate in the dataset(s). (25pts)
3. **December 12th 2024** - Written final report and analysis are due (100pts)