How do I Complete this Project?

This project is connected with the **Introduction to Data Analysis** course, but depending on your background knowledge, you may not need to take the whole class to complete this project.

Introduction

For the final project, you will conduct your own data analysis and create a file to share that documents your findings. You should start by taking a look at your dataset and brainstorming what questions you could answer using it. Then you should use pandas and NumPy to answer the questions you are most interested in, and create a report sharing the answers. You will not be required to use inferential statistics or machine learning to complete this project, but you should make it clear in your communications that your findings are tentative. This project is open-ended in that we are not looking for one right answer.

Step One - Choose Your Data Set

Click **this link** (available in a Google doc **here**) to open a document with links and information about data sets that you can investigate for this project. You **must** choose one of these datasets to complete the project.

Step Two - Get Organized

Eventually you'll want to submit your project (and share it with friends, family, and employers). Get organized before you begin. We recommend creating a single folder that will eventually contain:

- The **report** communicating your findings
- Any **Python code** you wrote as part of your analysis
- The **data set** you used (which you will not need to submit)

You may wish to use a Jupyter notebook, in which case you can submit both the code you wrote and the report of your findings in the same document. Otherwise, you will need to submit your report and code separately. If you would like a **notebook template** to help organize your investigation, you can click **here**. Or there may be a page in the project here called *Project Workspace: Complete and Submit Project*, where you can do all your work and submit the project.

Step Three - Analyze Your Data

Brainstorm some questions you could answer using the data set you chose, then start answering those questions. You can find some questions in the **data set options** to help you get started.

Try and suggest questions that promote looking at relationships between multiple variables. You should aim to analyze at least one dependent variable and three independent variables in your investigation. Make sure you use NumPy and pandas where they are appropriate!

Step Four - Share Your Findings

Once you have finished analyzing the data, create a report that shares the findings you found most interesting. If you use a Jupyter notebook, share your findings alongside the code you used to perform the analysis. Make sure that your report text is contained in Markdown cells to clearly distinguish your comments and findings from your code work. You should also feel free to use other tools and software to craft your final report, but make sure that you can submit your report as an HTML or PDF file so that it can be opened easily.