SI 618 PROJECT 2 REPORT

This document describes how the final reports will be graded. First and foremost, following the individual original work policy clearly stated at the start of the course, the topic and questions you ask in your project must be of your own invention.

# If you used ideas from a particular web site or previous project, or did your project as part of an existing research collaboration, you must identify your sources and/or collaborators and provide links and citation(s) where appropriate.

As a guide, the report should probably be not much more than 5-10 pages depending on space used for any visualizations, tables, etc. Examples from past years are on Canvas. The format of the report is flexible in that you can include additional information, but at a minimum it should have the following sections listed below.

* **Motivation** (4 points):
  + (a) Briefly state the nature of your research question and why you chose it.
  + (b) State THREE specific questions that you decided to explore for this dataset. Note that I asked you to come up with eight different potential analysis for your project mid-point report. You might find that some relate to the same overall question or maybe that some turned out to be more interesting than others. Choosing 3 questions allows you to demonstrate the best of what you found and/or combine analysis to make a stronger case.
* **Data Source** (6 points): Describe the dataset(s) you used, including the following minimal information:
  + The specific URL where the dataset or API resource was located (or provide a suitable sample),
  + The data format and important variables and their types,
  + How many records are included or retrieved (if using an API), and
  + What time periods they covered (if there is a time element)
* **Methods** (30 points): For each of the three questions, describe

1. How did you manipulate the data to prepare it for analysis?
2. How did you handle missing, incomplete, or noisy data?
3. What challenges did you encounter and how did you solve them?

* **Analysis and Results** (60 points): For each of the three questions, provide the following:

1. How did you perform data analysis in code, i.e. briefly describe the workflow of your source code
2. Summarize, in writing, the interesting result, relationship or insight (or maybe lack thereof) that you found in answer to that question using the analysis you described in the Methods section. Negative results, where your didn't end up being able to answer the question, are perfectly acceptable as long as you justify the methods
3. Include at least one accompanying data visualization that does an excellent job supporting the results of your analysis for that question (created in ).

Note that if you want to mention more than 3 questions, it is fine. It is also probably a good idea to look at a specific question using various analyses so multiple analyses can be listed for the same question.

# WHAT TO SUBMIT:

Please submit everything used for your project in the usual manner, by including it in a ZIP file project\_2\_report\_***youruniqname.zip***

Here are the files you need to submit:

* Report for part-2 named *si618-project-2-youruniqname.pdf*
* Source code for part-2 under folder *youruniqname-2*
* A folder containing data files and/or links to datafiles: For each data file used, include one of the following: 1) the data file itself if it is small, 2) a sample file containing the first 1000 records, or 3) working URLS that point to the data. If your data is sensitive (not publicly available), create a synthetic dataset that can allow us to test your code if necessary. If your data is sensitive and you cannot share, please contact me and confirm. In that case, you can include a readme file in this folder stating that.

***As part of the grading the instructor and/or GSI may attempt to reproduce your results using your code and data, and you are expected to assist with this if we request it.***

***Also note that the grade breakdown listed is assuming that the rest of the submission is proper (e.g. have working code that uses python).***