This page describes how the final report 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 be around 4-5 pages depending on space used for any visualizations, tables, etc. I've put some examples of past reports in the Resources/Example Projects folder.

The format of the report is semi-flexible - you can include additional information, but at a minimum it should have the following sections:

1. Motivation (5 points):  (a) Briefly state the nature of your project and why you chose it. (b)  What specific question or goal did you try to address?

2. Data Sources (16 points):  Describe the properties of the two dataset(s) or API services you used. Be specific. Your information at a minimum should include but not be limited to:

* where the datasets or API resources were located,
* what formats they returned/used,
* what were the important variables contained in them,
* how many records you used or retrieved (if using an API), and
* what time periods they covered (if there is a time element)

For example, if you downloaded data or used API services, you should state the specific URLs to those files or resources. It should require zero effort on my part to find and access the exact resources you used if I need to do so.

3. Data Manipulation Methods (24 points):  For each of your two sources, describe how you manipulated the data.  For example:

* How specifically did you need to manipulate the data?
* How did you handle missing, incomplete, or incorrect data?
* How did you perform conversion or processing steps?
* What variables and steps did you use to join the two data resources to perform your data analysis?
* Briefly describe the workflow of your source code and what the main parts do.
* What challenges did you encounter and how did you solve them?

4. Analysis and Visualization (20 points):

* A key goal of this project was bringing together two different data resources to answer an interesting question or find a new insight that could not have been answered with either data resource alone (which you summarized in part 1).  Now describe the analysis steps you performed on your combined dataset to address that goal/question. Be specific, and include references to key functions or parts of your code.
* What interesting relationships or insights did you get from your analysis?
* What didn't work, and why?
* To summarize your findings, include at least one visualization (chart, plot, tagcloud, map or other graphic) that summarizes your analysis.

*At the discretion of the instructor up to 10 bonus points will be awarded for especially high-quality, creative or insightful projects.*

Please submit the following files:

* Your project report, as a single PDF or Word document.
* All source code files/scripts (Python, or any other code) used for your project.
* Working URLs that points to either (a) the actual data/API resources you used or (b) if the datafile is over 10 Mb or not available in file form, create a sample file containing the first 1000 records.

***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.***

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