SI 618 WN 2020 Project Report Guidelines

Last revised: March 16, 2020

The individual project report in this course is worth 400 (of 1600) points.This document 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 citations where appropriate.

**NOTE: YOU CANNOT USE YOUR “FREE” LATE DAYS FOR THE PROJECT REPORT.**

As a guide, the report should probably be not much more than 5-6 pages depending on space used for any visualizations, tables, etc.

The format of the report is flexible in that you can include additional information, but at a minimum it should have the following sections that answer the following questions.

1. Motivation (30 points):  (a) Briefly state the nature of your project and why you chose it. (b) State FOUR specific questions that you decided to explore for this dataset.

2. Data Source (50 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)

3. Methods (80 points):  For each of the four questions, describe:

(a) how did you manipulate the data to prepare it for analysis?

(b) How did you handle missing, incomplete, or noisy data?

(c) How did you perform data analysis in code, i.e. Briefly describe the workflow of your source code

(d) What challenges did you encounter and how did you solve them?

4. Analysis and Results (200 points):  For each of the four questions, provide the following:

(a) 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 you didn't end up being able to answer the question, are perfectly acceptable as long as you justify the methods.

(b) include at least one accompanying data visualization that does an excellent job supporting the results of your analysis for that question.

5. Grammar, spelling, punctuation, and the inclusion of your name on the first page (20 points).

Here are the files you need to submit as part of the ZIP file described below:

* Your project report must be a \*self-contained\* file in PDF format.
* All source code (Python, or any other code) used for your project.
* A working URL that points to either (a) the actual data file you used to create your plots (this may be a derived file from a bigger, original dataset, e.g. suitably anonymized or sampled 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 (not .rar or other archive format) si618wn2019\_project\_report\_***youruniqname.zip***