Final Project

Rubric

Note: The Final Project is 20% of the total grade for this course.

Name:

<table>
<thead>
<tr>
<th>Category</th>
<th># Points Earned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposal</td>
<td>10</td>
</tr>
<tr>
<td>Design Elements</td>
<td>25</td>
</tr>
<tr>
<td>Execution</td>
<td>35</td>
</tr>
<tr>
<td>Final Write-up</td>
<td>15</td>
</tr>
<tr>
<td>Presentation</td>
<td>15</td>
</tr>
<tr>
<td>Extra Credit</td>
<td>10</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
</tr>
</tbody>
</table>
Proposal (10 points)

_____ (2 points) A hypothesis is stated that can be tested using data and computation. It is specific enough that you can reasonably evaluate it within the time frame for the project. Background context for the problem is given.

_____ (1 point) Briefly describe the data to be used in the project. item (1 point) What are the possible biases with the data?

_____ (1 point) Where is the data coming from (e.g., API, Web site crawling, manually downloading a dataset)?

_____ (1 point) What is the format of the data (e.g., CSV files, plain text, XML files, etc)?

_____ (2 points) The main steps of the program are numbered, specific, and manageable.

_____ (1 point) There is a description of how you will visualize your results (e.g., tables, charts, maps)? Spreadsheet software isn’t a valid option. Must use a graph in Plotly.

_____ (1 point) The major roadblocks are listed. What could go wrong with the steps you outlined, and what is your backup plan?

_____ Total

Code Design Elements (25 points)

_____ (6 points) Data is cleaned-up / pre-processed appropriately. For example, stemming isn’t always necessary or appropriate for text data, but you should ensure punctuation doesn’t negatively impact your results. If you’re using numerical data from CSVs, you shouldn’t have erroneous cells with incorrect formatting.

_____ (5 points) Is your approach to answering your question reasonable? For example, if you want to measure media bias, does your method to calculate such actually make sense, or is it likely be heavily biased or not a good approximation?

_____ (8 points) Structure. Given your approach to do the calculation you described in the previous item, did you make good design choices to breakdown your program into meaningful, coherent chunks in order to actually calculate such? You should use functions and variables to break the problem down into meaningful pieces that are pretty independent from one another.

_____ (6 points) Usage of external libraries when appropriate. If your input data is numerical data, you need to use Numpy and/or Pandas. Or, if your input data is textual, you must use a regular expression somewhere in your program. If you feel Regular Expressions are unnecessary for your text-based input data, you must make a compelling case for such in your write-up.

_____ Total
Code Execution (30 points)

A *bug* is a problem with the code, whereas an *error* might be an issue with the data or the user inputs.

- (4 points) Data Typing: do you use appropriate data types. For example, do your numerical calculations use a float when appropriate, or do you incorrectly cast them to Integers, causing rounding errors. Do you use Boolean values when appropriate (you may not need to use any)?

- (6 points) Data Structures: do you use appropriate data structures? For example, do you store items in a List when a Set would be more appropriate? Use a dictionary when necessary?

- (4 points) Style. Is your code easy-to-read? Commented well, and all functions have docstrings; no lines exceed 80 characters; good names for functions and variables; indented properly, etc. Your partner will have to interact with your code, so make it easy for them, our TAs, and your future self when you later look at the code.

- (4 points) Efficiency: We haven’t covered specific things to make your code faster. That’s okay, here, we are merely ensuring you do not wastefully compute the same thing *multiple* times. For example, you should only read from your *original* input data source *once*, while storing the elements you care about into data structures or immediately processing the relevant parts.

- (8 points) Are your computations correct? Results aren’t due to improper calculations?

- (4 points) Proper usage of control flow. Do you use if-statements, loops, and functions correctly and well?

  **NOTE:** The expectation is that your code runs without errors. You do not earn points for such. Yet, if it does not run without errors, we will deduct 10 points, as this denotes a serious oversight and carelessness.

- Total

Visualization (15 points)

- (4 points) Choice of output is an appropriately effective, compelling medium (not allowed to use Spreadsheet software). For example, does a bar chart best convey your data, or maybe a x-y scatter plot makes more sense?

- (4 points) Do you correctly display it? The above element checks if your approach made sense, this element evaluates if you properly executed such.

- (7 points) Is it easily digestible? Are all axis labelled (if you have axis)? You must include a title.

- Total
Write-up (15 points)

(4 points) Describe the background, hypothesis, data in well-written English. Someone completely unfamiliar with your project should be able to understand everything you’re doing, why you’re doing it, how it’s a valid approach, and how to interpret your results with respect to your area of inquiry.

(3 points) Methods, analysis, and results are presented well (writing is concise, tables/plots/screenshots are used, etc.).

(3 points) Overall document presentation (organized well, no spelling mistakes, no dead hyperlinks, files included in your final submission).

(2 points) Discussion of expected and unexpected findings.

(2 points) Describe what each person was responsible for, including who wrote which functions in the code (as a reminder, each person in a 2-person project receives the same score.)

(2 points) Reflection on approach, what went well and what did not. What would you change if you could do it all over again?

**NOTE:** Our baseline expectation is that we can easily follow what your project is about. If it’s truly unclear and hard to follow, we will deduct more than just the 4 points that we’re assigning to this element, for it would be very difficult to evaluate your project.

Total

Presentation (15 points)

You will give a presentation about your project lasting no longer than 6 minutes (if solo). If you have a partner, you may use upwards of 8-10 if needed, but it is not expected. Your presentation should include engaging visuals. Practice your talk several times on your own before class time. It will be recorded on lecture capture just like the regular lectures. Make sure you cover the following points:

(2 points) The problem and hypothesis; why is it interesting to explore?

(2 points) Your approach and the tools you used. Provide a justification for why your approach was suitable.

(1 points) What data did you analyze? Provide a description of relevant details (sampled vs comprehensive, collected or found, populations represented in the data)

(1 points) Visualization of results. Any tables, plots or maps are adequately labeled and easily digestible by the audience

(2 points) Discussion of your results. Was your hypothesis valid, why or why not?
(1 point) What are the limitations of your approach?

(1 point) Future directions for research. If you or someone else were to continue this line of research, what interesting problems were revealed by this project?

(1 point) Questions posed by audience were clearly answered.

(4 points) Did you attend the two days of presentations for which you are not presenting?

Total

**Extra Credit (max 10 points)**

(max 5 points) Use of sizeable dataset (1 point per 1,000,000 data points/words)

(max 5 points) Use of new methods or libraries

(max 10 points) Well-designed visualizations using elements/graphical types beyond what was demonstrated in class

(max 10 points) Additional elements not listed above

**NOTE:** If you feel one of the extra credit elements pertains to your project, please mention such in your write-up