

DSCI 510: Principles of Programming for Data Science

Fall 2022

Final Project Report

Due: 12/14/2022 4:59 PM PT

Task:

Please submit a zip package of:

- A **README.md** (a PDF, Markdown, or Jupyter notebook) file answering the following 5 questions. 15 points.
- Subfolder **code**. 5 points.
- Subfolder **data**. 5 points. A small sample is enough, less than 10MB.
- Subfolder **result**. 5 points. All your analysis / visualization results.

1. What's the name of your final project? Please describe it as a research question and provide a short description [3 points]
2. How to run your code? [3 points]
 - a. Please specify the dependency and version via requirements.txt or environment.yml. Add a note if you use some special libraries.
 - b. Please specify how to re-produce your result.
 - i. Please make it only one script if possible
 - ii. If your results need multiple steps/files to run, specify each of them.
 - iii. If your code is a python script, specify command arguments.
 - iv. If your code is a Jupyter notebook, ensure it could be run using Cell->Run All.
 - c. Provides a link to your github repository. If it's a private repository, you need to add mail@gleb.com and yuzhongh@usc.edu as your collaborator in the repo settings.
3. What data did you collect? How did you collect them? How many data samples did you collect? [3 points]
 - a. Specify exact data sources and your approach.
 - b. Describe what has been changed from your original plan, what challenges you encountered or resolved.

4. What kind of analyses or visualizations did you do? [3 points]
 - a. Describe the figure you made. Explain its setup, meaning of each element.
 - b. Describe your observations and conclusion. Describe the impact of your findings.
 - c. Describe what has been changed from your original plan, what challenges you encountered or resolved.
5. Future Work [3 points]
 - a. Given more time, what direction would you take to improve your project?

Rubric:

1. The final project is an individual project. No collaboration between students.
2. Please spend a decent amount of time on the report. Your report is the first file we read. We won't know how great your project is if you can't explain it clearly.
3. If you think your data collection part is easy, you could show your work by doing more in the analysis part.
4. For extra credit, you could do whatever you want to show your effort, including knowledge from other classes.