DS2001 Computer Science Practicum
Fall 2021
Final Project

Milestone	Date	Notes
Group reporting, initial topics	Friday, October 22nd (end of week 7)	Complete on Gradescope by **9pm**
Proposal deadline	Tuesday, November 2nd (beginning of week 9)	Submit proposal (PDF) on Gradescope by **9pm**
Project deadline	Tuesday, November 30th (beginning of week 13)	Submit written report (PDF) and personal reflections on Gradescope by **9pm**
Presentations	December 1st, 2nd, and 8th (weeks 13 and 14)	Presentations to take place during practicum. Felix will send a form to coordinate scheduling of presentations in November. You might not necessarily present during your normal practicum meeting time so that we can accommodate all groups. You are only required to attend practicum once in the last two weeks of the semester.
Peer Reviews	December 8th	Completion of a peer review is done during your presentation slot and counts for part of your final project grade

The goal of this project is to gain hands-on experience with finding, importing, analyzing, visualizing, and presenting a dataset of your choosing. The idea is to perform an end-to-end data science project on a realistic task, using Python.

You'll work solo or in a small group of 2-3 members. If you work with a group, you are responsible for arranging your collaboration and meetings within the university's social distancing and covid guidelines. All team members must be present for the project presentation.

If you are looking for team members, we suggest you post on the course discussion board or use the provided class time to exchange contact information with your classmates.

You may work with anyone from any of Felix's sections of DS 2001, but you may not work with anyone from a section of DS 2001 that is not taught by Felix.

Project Requirements

You must identify a real dataset that is both substantial, and of interest to you (e.g., https://www.data.gov, https://data.boston.gov/, https://www.kaggle.com/datasets, https://toolbox.google.com/datasetsearch, https://data.noaa.gov/dataset/, https://healthdata.gov/, are just *a few* places to look).

Excluding cleaning the data, your project must include at least 3 Python functions (we'll be covering functions during weeks 5 - 7 of DS 2000/2001) that operate on the data: each should be motivated by a question you have about the data and should be accompanied by textual discussion/analysis. Each function must take parameters and use them in an appropriate way. You must create at least two visualizations to include in your written report and in your presentation.

You are welcome to use Python modules not covered in class. Provide links to all resources/tutorials used.

You must submit your Python code, all datasets you used, and a PDF of your final report.

Your group will present your work in-person during practicum. Your presentation will be graded on substance, clarity, and narrative. We'll go over presentation guidelines during week 11 in DS 2001.

Proposal Requirements
You must submit a written proposal as a PDF. This proposal should outline:

- A high-level statement of the problem you intend to address or investigate.
- The data source(s) you intend to use.
    - Link to the data source, either on the web or uploaded to google drive/dropbox/another location that course staff can access it from.
    - You should include a description of the format of the data. What format is it available in? What information does it include? Do you identify any ethical issues around the manner in which the data was collected?
- The goals of your analysis, ideally in the form of a testable hypothesis, or via well-defined success metrics.
- The products you plan to build, including at least three ideas for specific visualization that may help you investigate your problem.
    - Make sure to propose specific visualizations that you will work to create. Even if you later find that there are better visualizations than the ones that you propose, this will give you a place to start from.

Clearly specify the names and emails of all team members on the proposal. Proposals are not guaranteed to be accepted -- you may have to revise and resubmit. We have designated some time during practicum to work on your proposals and project ideas; use this time to discuss with us before submitting your proposal.

Organize your proposal clearly so that we can easily identify all components that we've asked for.

Proposals will be submitted in PDF form, on Gradescope.

No late proposals will be accepted.

Report Requirements
Your project report is the formal description of your project. The report should be 4-8 pages in length, double-spaced, and submitted as a PDF. The report will be graded on content and clarity. Your report must include the following sections, clearly labeled:

Problem Statement and Background
Give a clear and complete statement of the problem. Don't describe methods or tools yet. Where does the data come from; what are its characteristics?

Introduction and Description of the Data
Describe your motivations for choosing this particular project and related dataset(s). Where did you obtain the data? Why is this problem important?

Include background material as appropriate: who cares about this problem, what impact it has, what implications better solutions might have.

Methods
Describe the methods you explored (algorithms or data cleaning approaches). Justify your methods in terms of the problem statement. Do not explain your code line-by-line to us. Imagine that you are explaining to a peer who has not learned to code what your methodology was. The goal here is for you to explain the overall strategies that you used to conduct your analysis.

Results, Conclusions and Future Work
Give a detailed summary of the results of your work. Summarize the strengths and shortcomings of your project, and speculate on how you might address these shortcomings given more time.

Please use visualizations whenever possible, both in your report and in your project presentation. Make sure all visualizations are clearly labeled and explained. All group members' names and email addresses must be included in the written report.

No late reports will be accepted.

Individual Reflections
The project will be submitted by the entire team, and a grade will be assessed for all team members. Additionally, each team member will submit their own individual reflection.

In the reflection, you'll be asked to describe your own experience in the group project as well as the contributions of yourself and your teammates.

These reflections are serious statements and will be used to re-distribute individual grades on the project. If one person went above and beyond for the project, their grade could be adjusted up. If one person contributed significantly less than others, their grade could be adjusted down.

We will take these reflections seriously. One group member's grade could be impacted as much as 50% as a result of reflections of their teammates.

Your individual reflection should be 1 page, single-spaced, submitted as a PDF, and should answer the following questions (label these clearly):

1. *What's the most important thing you learned during this project?*
2. *Does this project reflect your best work? What would you have liked to have spent more time on or done differently?*
3. *For each team member, how did they contribute to the project and work with the group? (You must answer this question for everyone on the team, including yourself.)*

Project Presentations
Your presentation must be 8-10 minutes long, and will take place during the final two weeks of the course. All team members must be active participants in your group's presentation.

We will set up the presentation schedule towards the end of November. **In addition to your own presentation, you must also reviews of other presentations during your practicum sessions; this will count towards your project grade.**

You must use slides (Powerpoint or Google Slides or Keynote). Your presentation must include at least one visualization you made from your data. Your presentation will be evaluated for substance, clarity, and narrative.

Grading
Your entire group will receive a grade based on your proposal, project, written report, and presentation. Your individual grade may be adjusted up or down based on your group members' individual reflections.

Group Grade

Factor	**Weight**	**Description**
Group + Initial Topics	2%	● You must submit on gradescope confirming your group members and initial topic ideas ● This form must be submitted on time.
Proposal	10%	● Proposal must satisfy requirements specified in this document. ● Proposal must be submitted on time.
Project + Report	60%	● Project must satisfy data/documentation/function/ visualization/submission requirements ● Code used in the project must be high quality, well-documented, and clear. ● Written report must cover all sections enumerated in this document. ● All sources and datasets must be cited. ● Projects must be submitted on time. Late projects will not be accepted.
Presentation	25%	● Presentation must establish the connection between the data and your project motivation/questions and functions/visualizations. ● Presentation length must meet requirements. ● All group members must be present. ● Presentation must be clear, well-rehearsed, and professional.
Peer Reviews	3%	

Individual Grade

- Your grade will be the same as your group's overall grade with a possible adjustment based on the individual reflections.
- The individual reflection document itself will not receive a grade, but failure to submit the document and/or reflect on ALL teammates, including yourself, will have a negative impact on your grade