

DS2001 - CS Practicum

Spring 2023

Project #2

Milestone	Date	Notes
Group reporting, initial topics	Mar. 16	<p>Tell us your teammates:</p> <p>https://docs.google.com/spreadsheets/d/15Y7puYQw0SVdkNU1RH6uJtQdo5F-nykFWxkMLDcSkw8/edit?usp=sharing</p> <p>If you prefer to work alone, you can request:</p> <p>https://forms.gle/xk5yum7r37Wf3LUG7</p> <p>If you do not fill out either form by March 16th, you'll be assigned to a team.</p>
Proposal deadline	Mar. 16	Submit proposal (PDF) on gradescope by 8pm. You will have time in practicum to work on your proposal.
Project deadline	Apr. 11	Submit written report (PDF), presentation slides, and personal reflections on gradescope by 8pm
Presentations	Apr. 12, 13, 19	Presentations during practicum
Peer Review	Apr. 19	<p>Completion of a peer review is one of your practicum assignments.</p> <p>https://forms.gle/JtiqbDGh2UZ11V158</p>

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

You'll work in a team of 2-4. We expect a substantive project that incorporates the contributions of all teammates. You may choose your own teammates; let us know who's on your team by the deadline listed above. If you're looking for teammates, try posting on our DS2001 Piazza or chatting with the folks in your DS2001 section. Your teammates can be in **any** DS2001 section taught by Prof. Mosca.

You may choose to work alone if you let us know by the deadline listed above. If you have not told us your teammates or that you're working solo by then, you'll be assigned to a team.

All team members are all responsible for arranging your collaboration and meetings. All team members must be present and participate in the project presentation.

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://toolbox.google.com/datasetsearch>, <https://data.noaa.gov/dataset/>, <https://healthdata.gov/>, etc).

Your analysis of the data should be motivated by some overarching theme, context, or question, and must include at least 3 specific analysis questions that fall under that umbrella. For example, if I chose the bike data from our previous practicums my overarching question might be: What does bicycle commuting in Cambridge look like?. My specific analysis questions might be: (1) In which season is bike commuting most popular? (2) What time of day do most people bike commute? (3) Is there a difference in how many people travel westbound vs eastbound for the morning commute?

Excluding cleaning the data, your project must include at least 3 Python functions that operate on the data. You will also need to generate at least three visualizations; one to help answer each analysis question you asked.

You are welcome to use Python modules not covered in class as long as you provide links to 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.

Proposal Requirements

You must submit a written proposal, submitted 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.
 - You should include a description of the format of the data. What format is it available in? What information does it include? Identify any ethical issues around the manner in which the data was collected, or the information in the data.
- The goals of your analysis, ideally in the form of testable hypothesis, or via well-defined success metrics.
- The products you plan to build, including a minimum of three visualizations.
 - Make sure to propose specific visualizations that you will work to create. Even if you later find that there is a better visualization than the one you proposed, 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 a practicum for proposal work; use this time to discuss with us before submitting your proposal.

Proposals will be submitted in PDF form, on the same Gradescope where you submit your practicum assignments.

No late proposals will be accepted.

Report Requirements

Your project report is the formal description of your project. The report should be 2-4 pages in length, formatted cleanly, and submitted as a PDF. The report will be graded on content and clarity. Your report must include the following sections:

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; do you have any concerns about ethical issues around how the data was collected and shared or what is in the data?

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 (data science algorithms or other approaches to analyzing the data). Justify your methods in terms of the problem statement.

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. (And remember, your report must include at least 3 visualizations that answer each of your analysis questions.) 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 will be 1-2 pages, formatted cleanly, submitted as a PDF, and will answer the following questions:

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. It will take place during practicum as specified above. You may sign up to present in any practicum. All team members must be active participants in your group's presentation.

We will set up the presentation schedule as we approach the end of the semester. **In addition to your own presentation, you must also attend another group's presentation and complete a review; this will count as one of your practicum assignments.**

You must use slides (Powerpoint or Google Slides or Keynote). Your presentation must include at least three visualizations you made from your data that answer your three analysis questions. 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
Proposal	10%	<ul style="list-style-type: none">• Proposal must satisfy requirements specified in this document.• Proposal must be submitted on time.
Project + Report	60%	<ul style="list-style-type: none">• 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	30%	<ul style="list-style-type: none">• 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.

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 be graded, but failure to submit the document and/or reflect on ALL teammates, including yourself, will have a negative impact on your grade.