

Applied Data Analysis

Winter 2022

Project Guidelines

Project Type and Scope

The project is an important component of this course. It is an opportunity for you to work on a data analysis project of your choice. Also keep in mind that course projects are one of the favorite topics during a job interview.

Your project should significantly advance your skills beyond what is covered in class.

For some examples of data sets you could use, look [here](#). There is also an Ed Discussion thread for sharing interesting data sets.

There are many topics of active research, and you will likely find research articles containing sophisticated solutions for these problems. Given the available time, however, you cannot expect to develop a state-of-the-art solution during this course project. Use this opportunity to delve deep into a problem of your choice, apply what you know so far in data analysis, study the literature and learn new techniques relevant to your problem, develop a serious solution, evaluate its performance, and present your experience and results. Along the way, look for some insights that future interviewers may find unexpected or interesting—“Out-of-state students have higher graduation rates than in-state students” or “Grade Point Average is a poor predictor of career success.”

Please reach out for feedback during the quarter, if you have questions as your project evolves.

Time Commitment

The project counts for 60% of the grade, and you should plan to work accordingly. Plan to spend a minimum of 50 hours on your project during the quarter.

Most projects will require a combination of, and several iterations of, reading up on the existing literature, thinking on the problem, learning how to use relevant software libraries, writing and executing code, and preparing a presentation and report. Depending on the project you will spend more time in one activity than another.

Deliverables

There will be five deliverables:

- i. Project Proposal (10% of total course grade)
- ii. Mid-Quarter Presentation (5%)
- iii. Final Report (20%)
- iv. Final Presentation (15%)
- v. Peer Review (10%)

The mid-quarter presentations will be in front of the class. If classes are remote, they will be recorded and posted on Canvas. You should offer constructive feedback to other students on their proposals and presentations.

Project Proposal

Submit your proposal on Canvas by 9:00 a.m., Saturday, February 19th. It should be 3-4 pages long and include the following:

- Title.
- Brief description of what you want to do, including why it is useful, the data and software you will use, and the software you will write, if any.

- **A detailed description of the related work.** You should search for research papers and projects that solve the same or a similar problem. For the closest two or three such papers, you should describe what methods they used for obtaining the data, preprocessing the data, learning models, choosing metrics, and evaluating their results. You should also report their results, and what implications their work has on your project. (Learn as much as you can from such related work as it will give you an idea of what you need to teach yourself—beyond the material covered in class—for your project.)
- Brief plan of action, including any insights you have, the various steps of the project, the software libraries or packages you will be using, and the software you will be developing on your own. Also include what you plan to demonstrate for the mid-quarter and for the final presentation.
- Brief description of how you will evaluate your work. There may be existing techniques or results you could compare to, or you could test how well your solutions performs on test data, or how well it models the available data.

Mid-Quarter Presentation

The mid-quarter presentation should be about 5 minutes describing the project objectives, work done so far, and challenges/obstacles, including any demonstrations of software. You will present your project to the rest of the class on Saturday, February 19th. If classes are remote, you will need to submit your video on Canvas by 9:00 a.m., Saturday, February 19th. Note that this could be a narration over slides, rather than video of you presenting.

Final Report

Submit your final report on Canvas by 9:00 a.m., Saturday, March 19th. The report should resemble a professional research paper and include—

- An executive summary describing the work you have done.
- An introduction describing the problem and its significance.
- A detailed description of the related work.
- A detailed description of your solution and the work you have done.
- Include both what worked and what didn't, particularly if you have insights into the reason why.
- A detailed presentation of the results you have obtained and its analysis.
- Suggestions for future work along the same lines.
- Description of your effort, including the relative effort on different activities. What did you have to learn for the project? What skills did you already possess? Did you need to learn how to use particular libraries? Did you spend more time reading research papers, or fine tuning your parameters?
- Bibliography.

Please write your report in clear, concise English, and include clearly captioned, helpful figures where appropriate. Unclear or sloppy reports will affect your grade.

Also upload all your source code and data to your GitLab repo for the course. In your repo, include a readme file containing—

- A description of the directory structure, and a brief description of the purpose of each file and the number of lines of code in it.
- A list of files you are using that contain code not written by you, or not written as part of this project (maybe it was part of another project).
- A description of any data files you use for your work.

- Any other relevant information about your code.

Your code should be documented such that one can get a reasonable idea of how it works. For example, each function should have a document string in the recommended style of the programming language.

Final Presentation

During our final exam time, Saturday, March 19th, 9:00 a.m. - Noon, you will deliver a 10-15 minute presentation to the class describing the motivation/usefulness of your project, data used, your approach, unexpected challenges and modifications, and any results or conclusions you have reached. If classes are remote, you will need to submit your video on Canvas by 9:00 a.m., Saturday, March 19th. Note that this could be a narration over slides, rather than video of you presenting.

Peer Review

Submit your peer reviews on Canvas by 11:59 pm, Monday, March 21st. You will have the reports and presentations of three other students shared with you Saturday at 9:00 a.m. For each report & presentation, prepare a review (about one page) of the project. This should not be a summary of their project. This should be your views on what was done well or poorly, whether or not the conclusions are accurate or meaningful, possible improvements to the existing project, and possible extensions of the project for future consideration. This will not affect the grade of the author of the report and presentation, but will impact the grade of the reviewer. I am looking for constructive, insightful feedback, that can be returned to the author.