

JSC370 2026: Final Project

Final Project

Due Date and Grading Scheme

- **Due date April 26th, 2026**
- A 5-minute recorded presentation (10%).
- A GitHub website (10%).
- A written report on your website and downloadable from the website (80%) (see below for more details). The report length should be a about 6–7 pages including single spaced text, figures, and tables.

Details

Learning Objective: The main objective is to apply the skills learned in JSC370 to explore a research question(s) and/or hypothesis(es) by analyzing and interpreting a dataset of your choice.

Narrative: Through this project you will showcase a data science project on your personal GitHub website. Using the dataset from your midterm (if you chose to switch topics from your midterm, please let me know), formulate clear and concise question(s)/hypothesis(es) to answer. You will apply the skills learned throughout the semester to answer the question(s)/test the hypothesis(es).

The final project should include modeling (predictive and/or inferential) and focus on the regression and classification methods covered in this course — such as XGBoost, Random Forest, or generalized additive models — or unsupervised learning such as LDA topic modeling. For prediction methods, include feature engineering (variable creation), variable importance, and training/testing splits. If you wish to include a method learned elsewhere (e.g., deep learning, k-means, PCA), that is fine as long as you also include and compare to a method covered in this course.

Deliverables:

1. A 5-minute presentation where you walk through your website and main findings. Upload the video file/link to Quercus or include the link in your website.
2. A GitHub repo and GitHub website for your project. On the website, provide a summary of the project on the main page (i.e., `index.html`). You may embed the interactive visualizations (homework 5) in your report or put them on a separate page (e.g., main report on `index.html`, interactive viz on `viz.html`). Please upload the link to your GitHub repo and/or website to Quercus.
3. A written report with embedded tables and figures, submitted as a downloadable link on your website.

Note: if you wish to have your full report on your website, that is fine — please also create a PDF for download, but it can be simpler and indicate that the full report is on the website (please link to it).

The report should have the following sections (elaborate/expand from what was written in your midterm):

Introduction provide background on the topic and data, and a formulated question(s) and/or hypothesis(es). This does not require references to journal articles, but should include a narrative of the topic you studied.

Methods include how and where the data were acquired, how you cleaned and wrangled the data, what tools you used for data exploration, description of what models and/or statistical tests, and training/testing you used. Please include some statistical details on the model(s) (e.g. if you used XGBoost, briefly describe the model, hyperparameters, training/validation/testing) and test statistics (e.g. R^2 , RMSE, MAE, ROC/AUC, etc).

Results provide final, publication ready tables and figures from your analysis, refer to your website if needed (for the interactive plots).

Conclusions and Summary interpret your findings in a bigger picture way and point out any limitations.

In your report, please do not include code chunks or unformatted output.

Summarize data and model output in well captioned and numbered tables and figures.

Checklist for Final Project

1. **Create a website.** It should feature:
 - a. A description of the project (alternatively you may put your full project on the website).

- b. Interactive visualizations (graded via homework 5), each with a caption that describes the plot. This may be plotly figures on their own or through a dashboard, but the dashboard is not required. The interactive visualizations may be embedded in your report or on a separate page.
 - c. A link to the PDF version of the report (i.e. a link to “Download the report.”).
 - d. A link to your github repo.
2. Make sure that the **website** actually **works**, i.e., figures and interactive visualizations are properly rendered when visiting the website.
3. Upload **all** source code, data, website files, and the PDF report to your GitHub repository. Make sure your project is reproducible!
4. **Have a README.md file** in the upper level of your project github repository. The **README.md file contains a link to your website** e.g. <https://username.github.io/JSC370-finalproject> and provides a *brief* description of the project, like title, link(s) to the API and any other sources of data.
5. If you are using saved data, store the data in a **“data” folder**. If you are not providing a dataset, the API instructions should be clear in your code and please provide a link to the API in the README.md of your project repo.
6. Again, your report should be **fully reproducible**, meaning you don’t have paths to files not shared in the repository.