

# Final Project: Part 1

Data 100/200A: Principles and Techniques of Data Science

Fall 2021

The purpose of this project is to put into practice what you have learned in this course through the design and implementation of a typical data science workflow, including data cleaning, visualization, exploratory data analysis, feature selection, and modeling.

Your group will have the choice to select one of three different datasets which we will provide. Each dataset will have a different set of guiding questions, but will more or less be graded similarly.

This is Part 1 of the project, where you will load and clean the data and produce some exploratory data analysis both guided and on your own. After a review of your work, we will then release the second part. Both these parts combined will make up your final report and submission.

## Project Guidelines

The project involves carrying through the following steps.

### 1. Load and Clean Data

- Guided questions to read in the given the dataset(s)
- Guided questions to clean the dataset(s) to effectively complete the next steps in the project.

### 2. Exploratory Data Analysis (Learning set only.)

- Guided EDA questions
- Open Ended Exploration of the data

### 3. Design Review

- Write a report on your data exploration and your plan for analysis/modeling for part 2.
- Present your ideas to your TA in discussion section.

## Timeline

| Date (by EOD at 11:59pm) | Event / Deliverable            | Relevant Links                       |
|--------------------------|--------------------------------|--------------------------------------|
| 11/3                     | Project Part 1 Released        |                                      |
| 11/5                     | Project Group Form Submitted   | <a href="#">Project Group Form</a>   |
| 11/8                     | Project Dataset Form Submitted | <a href="#">Project Dataset Form</a> |
| 11/17                    | Design Document Due            |                                      |
| 11/19-23                 | Design Document Review         |                                      |
| 11/24                    | Project Part 2 Released        |                                      |
| 12/13                    | Final Deliverable Due          |                                      |

# Report Format and Submission

The project submission will be the autograder-generated zip file as well as the PDF of the design document.

1. **Code.** Use the provided starter notebooks to complete the following aspects of the project. Each dataset will have their own starter notebook.

- (a) Loading the Data
- (b) Cleaning the Data
- (c) Exploratory Data Analysis

Note: We will run the notebooks when grading, so please account for that.

2. **Design Document Proposal.** This typed portion of the notebook should summarize your workflow and what you have learned. You should discuss your EDA, along with any questions that you have come up with regarding the data. Additionally, you should have a proposal for the modeling portion of the project.

- Describe the data.
- Explain what exploratory data analysis (EDA) you conducted on your own and provide presentable data visualizations.
- Propose a problem that you will address with modeling. Perhaps this is a problem you discovered while conducting EDA. Some potential questions to answer: What is the problem? Why is it relevant/intriguing? How will your model address this problem?
- What sort of modeling do you plan on conducting? Carefully describe the methods you plan on using and why they would be appropriate for the question to be answered.

## Grading.

Part 1 of the project will be graded based on your code, design document, and design document review.

### Grading Breakdown

- **Part 1:** 50%

| Project Component      | %  |
|------------------------|----|
| Guided Cleaning        | 7  |
| Guided EDA             | 10 |
| Open Ended EDA         | 5  |
| Design Document        | 18 |
| Design Document Review | 10 |

[Point breakdown by question](#)

- **Part 2:** 50%

### Team work.

You must complete the project together with two other classmates. You will be graded equally. Your group must consist of other students from your assigned discussion section, and you must have submitted the Project Group form.