Final Project

The Final Project will give you the chance to explore a topic of your choice and to expand your analytical skills. By working with real data of your choosing you can examine questions of particular interest to you.

The broad objectives for the project are to:

- Identify the problems and goals of a real situation and dataset.
- Choose an appropriate approach for formalizing and testing the problems and goals and be able to articulate the reasoning for that selection.
- Implement your analysis choices on the dataset.
- Interpret the results of the analyses.
- Work effectively to manage a project as part of a team.

To accomplish this, you will work in teams of 3 to 4 students to conceive of and carry out an analysis project. We will form your groups randomly based on the number of students enrolled in the class. Everyone who will be part of some group*.* You will find in your future careers the need to work on projects in groups frequently (even if you really, really don't want to). Larger Goal is to give your more hands-on practice to the concepts you will learn in the lessons and can apply that knowledge to new challenges.

The Project will be completed in multiple parts due separately:

Part I: Propose a dataset and purpose of analysis due Friday, February 10, 2023, before midnight.

Part II: EDA, Insights and Modelling – Friday, May 05, 2023

Part I:

Propose a data analysis project. This can be almost anything that you choose. You might select a project from your field of study, your extracurricular interests, government, or public policy, or elsewhere. We strongly encourage you to discuss potential project ideas with your TAs and IAs, and/or with Professor. This will give us a chance to make sure you're on the right track even before you submit your draft.

The main purpose of the proposal is for us to give feedback on whether the scope of the project is in the range of what we're expecting. On average we expect proposals to be about 1-2 pages long, though we know the lengths will vary. Please create a document containing the following **two** parts.

1. **Dataset**
   - State what data you plan to use?
   - Describe the data. As part of this, please include the total size of the dataset (e.g., number of rows) and a small sample of the data.
   - Include a link to the source of the data and discuss any difficulties you anticipate getting the data ready for analysis.
2. **Goals**
   - Formulate a specific set of research questions you want to answer, points you want to make, or issues you wish to explore through the data. Be as concrete as possible.

- Give background and context of your topic and question in a few paragraphs.
    1. Why is this question of interest to your group?
    2. What background information led you to your hypothesis.
    3. Why is this important?

What To Turn In

Your proposal should be in a pdf document named **group#_proposal.pdf**. Include clearly at the top of the document the **name**(s) and **SUID**(s) of all the group members submitting the proposal, then include the two parts of the proposal specified above. Upload the pdf document to Canvas.

Part II:

In this part we want you to do Exploratory data analysis (EDA), which allows you to uncover new information in the data that is not self-evident (i.e., plotting, slice and dice the data in different ways, create new variables, or join separate data frames to create new summary information). You will provide your findings in the form of plots and tables. Make sure Graph(s) are carefully tuned for desired purposes and are properly formatted (plot, axis titles, legends (if necessary), scale is appropriate)

Insights obtained from the analysis are thoroughly, yet succinctly, explained. Easy to see and understand the interesting findings that you uncovered.

Use any regression/ML model required to address the problem (research questions) you have posed in the part 1 of this project.

- Summarize the problem(s) you are addressing in this project. Explain clearly how you addressed this problem.
- Summarize the interesting insights that your analysis provided.
- Discuss the limitations of your analysis and how you, or someone else could improve or build on it.

What To Turn In:

- Your report as a PDF. **(Do not turn in a Word document or plain text).**
- Your Python file(s) and any other supporting files you need.
- Your README file explaining how to set up your project and how to run your code to reproduce your results. (If required)
- You do not need to submit your dataset for this part, but there should be clear instructions in your README for the TAs/IAs to download the data so they can run your program.