MATH113/CAIS105: Intro to Data Science

Fall 2023

Final Project

For your final project you will design, run, and communicate an analysis related to an area of interest to you. Your analysis can be related to anything you'd like as long as you meet the requirements below.

Proposal – 10 points

For your project proposal you will identify your project group, your topic of interest, the major building blocks to your analysis, and potential roadblocks. Type up a document that answers the following questions. Your document should be 1 – 2 pages long and well formatted. Submit as a group on PLATO.

1. Group
   a. With whom do you plan to work? Groups must be 2 – 3 members.
   b. Talk about your schedules. How will you coordinate your work?
2. Topic
   a. What topic area will you work with?
   b. Find data to use in your analysis. Provide a link to the data.
3. Building Blocks
   a. Broadly speaking, what is the overarching research question for your analysis?
   b. On a fine-grained level, what are *five analysis questions* you will ask and answer?
      - Each should relate to your overarching research question but should be specific and distinct.
      - What variables in your dataset will you use to answer each question?
   c. We have covered many core data science programming structures and techniques in this class. Your program must include at least two of the ones listed below. Which two structures/techniques will your code include?
      - Joins
      - Pivots
      - Functions
4. Roadblocks
   a. What roadblocks do you anticipate as you perform your analysis?
   b. What is your plan for overcoming these roadblocks or pivoting around them?

In-progress – 24 points

You will turn in an in-progress version of your analysis that will include a working (**but not complete**) version of your code, and a README section. Submit via an R Markdown generated PDF, with a README section at the end, as a group on PLATO.

1. Your file should:
    a. Include a header with group names, date, and project title
    b. Include an introduction to your project with your research question, 5 specific analysis questions, and a brief explanation of why your analysis is interesting and important
    c. Include an overview of your data explaining where you found the data, who collected it, and any biases or ethical issues with the data
    d. Be well structured into reasonable code blocks and text blocks
    e. Have code with appropriate variable names and function names that is well commented
    f. Have text that explains any data cleaning you performed, outputs of analyses, and answers analysis questions
    g. Include citations for any sources referenced
2. The README section at the end of your file should include:
    a. An update on the progress of your analysis – how many questions have you answered so far and what is left to answer?
    b. A plan for finishing the analysis. Do you anticipate having to pivot anywhere? If so, how?

Report – 80 points

Your final submission will include a PDF report generated from R Markdown, a working version of your code in a .Rmd file, an in-class presentation of your work, and an individual reflection. You will submit your PDF and .Rmd on PLATO, as a group, and your individual reflection on PLATO individually.

1. Your final report should include:
    a. Interspersed text and code chunks
    b. Text chunks should:
        i. Be well formatted
        ii. Be proofread
        iii. Include an introduction (as specified above)
        iv. Include a data overview (as specified above)
        v. Walk the reader through the analysis. What specific analysis question does each code chunk relate to? What do outputs indicate?
        vi. Include a summarization of the analysis. What did you find with respect to your research question?
        vii. Include citations for any sources referenced

    c. Code chunks should:
        i. Be well documented with comments
        ii. Include appropriate function and variable names
        iii. (at least two should) Include the programming structures / techniques you indicated you would use in your proposal
        iv. Run without errors

2. Your presentation should include:
    a. An introduction to your analysis
    b. An overview of your data
    c. Analysis questions and answers
    d. At least 3 appropriate visualizations
    e. A summary of what your analysis indicates with respect to your research question
    f. Contributions from all group members (each person must talk)
    g. **Do not include code in your presentation** You should explain at a high-level how you performed your analysis but should not walk the audience through code

3. Your individual reflection should include:
    a. Your specific contributions to the project
    b. Your teammates contributions to the project
    c. Whether you navigated any conflict or discrepancy in workloads with your teammates
    d. How you navigated those conflicts or redistributed work

** Your final grade will be a combination of your grades for all parts of this project and may be adjusted based on individual reflection feedback.

Submission

All work will be submitted through PLATO as a group with the exception of your individual reflection.