Overview

If you feel comfortable with using Github, we would warmly recommend doing so when working on the project. Not only is it a good idea to have a "save point" of sorts throughout the process of writing code. If you are not comfortable with Github, something that has a version history (e.g., Google Docs/Drive, Overleaf, etc.) would also work.

Generally speaking, it is a good idea to document workload and track of project progress. In the event that you make a concession request, one of the first things we will ask for is proof of progress so it's a good idea to save your things somewhere that will allow you to document this.

If you are having group collaboration issues, please let your project mentor and the instructor know as soon as possible!

Deliverables

1. **All** the Python file(s) used in the project. If you have other code files, include them too.

   At minimum, your Python files should consist of:
   a. The Python code you used to dynamically generate your INSERT statements that are in your .sql file
   b. The code that you used to pull data from the database and the subsequent handling of that data to tabulate your results and create the figures used in the final research paper.
   c. The Python file should contain the SQL queries that you have written to pull data from the database.
      i. The SQL queries should be as specific to your needs as possible. That is, do not write queries to generically pull all the data to your database and then process the data after. Your queries should be as specific as possible for your needs.

2. A README file that contains information someone would need to know if they were to run your code. The README should also list your research questions and the SQL queries that were used to answer the questions. Please include line numbers to make it easier for your TA to find the particular queries in your code.

   Please list out each of your files and provide a short description of each so your TA can quickly reference the files they need to look at when grading.

3. Your research paper which includes the following sections. You are free to reuse text from your proposal/midway checkpoint.

    a. Introduction
        i. Give a bit of background on the topic to give context to your research question. Imagine that this paper is being read by someone totally new to your project (e.g., someone outside of the class). What would they have to know in order to understand your research topic/research questions?

        ii. State your research questions.

            WARNING: If you are in a situation where your project TA has concerns about the suitability/scope/feasibility of your project, we expect that you have taken steps to resolve those issues **prior** to the submission of your research paper **and** that you have double checked your changes with your TA prior to submission. We warmly recommend that you take advantage of the weekly meetings with your TA to ensure that your research questions (and project) meet expectations around scope and suitability.

            In the event that you do not resolve the outstanding concerns, we reserve the right to apply a 60% penalty to your project. The penalty will be calculated based on the full value of the deliverable.

        iii. What is the impact of knowing this information?

    b. Related work/work that has been done in this area by others. Don't forget to cite!
        i. Based on prior work that has happened in the area of interest, what is the background information that suggests that your research question is important/valid?
        ii. Discuss some other related work/projects people have done in the area.

    c. Data
        i. What datasets are you using?
        ii. Why did you choose these particular datasets?
        iii. What data cleaning steps did you perform and why do you think these steps were sufficient.
        iv. What are some things people need to be aware of about your data? This is where you would discuss issues around outliers or missing data or any other significant findings from your EDA process.

d. Methodology
    i. What methodology did you employ to answer your research questions?
    ii. Why did you choose your particular methodology?

e. Results and Discussion
    i. What were your results?
    ii. For each of your research questions, produce a graph that depicts your results.
    iii. Validity refers to how generalizable your results are to the overall population/situation. Comment on the following:
        1. Internal validity: How well does your measure reflect what your research question is measuring? How do you know?

        2. External validity: How well does your measure reflect the truth in real life? How do you know?

    Remember that it's fine if this is not the most perfectly designed study. Even published research studies often have shortcomings - sometimes due to the design but often due to constraints outside of their control.

    We are looking to see that you can critically reflect upon your work to understand what is good about it and what shortcomings it has.

f. Discussion of results
    i. Explain and interpret your results by putting it in context based on your research questions.
        1. In cases where you do not have clear/meaningful results, explain why this might have been the case.
        2. What impact does this have on the interested community
    ii. If there are other limitations to your results that you have not discussed in the validity section, discuss them here.

g. Future work
    i. How can people use your results to benefit them?
    ii. What are some things that you would like to explore but did not have time to? Another way to think about this question is if you had an extra four months to work on the project, what else would you do? This could be extensions to your original research question or a follow up on your results or using another method to analyze the data.

h. References
    i. References should be given in ACM format

4. Declaration of AI tool use

   If you have used an AI tool to help refine your work, state which tool and what prompts were given. You can also provide a URL of the chat log instead (please make sure the URL is accessible).

   If you have not used an AI tool to help refine your work, please explicitly state so.

Congratulations!

You've completed the research project!