

Philosophy

The point of the midway checkpoint is to ensure that you are on track for your project and that progress is being made. As you have started working more on the project, some of your plans may have changed and we would like to check in with you to ensure that the changes will not impede your ability to meet the rest of the project requirements.

Overview

When writing your research paper, we warmly recommend that you choose a tool that has the ability to track revision history. Examples of this are Google Drive, Overleaf (particularly useful for those of you who want to use LaTeX), or Git. If a disagreement arises over workload or participation, the revision history will be useful in demonstrating who produced what work.

If things are not going well in your project group, please reach out to your project mentor or the instructor right away! The sooner you speak up, the more options that are available. Generally speaking, your project mentor should respond within two **business** days. If they have not, then please let the instructor know.

Deliverables

All of the deliverables should be combined into a single PDF file. Please be sure to clearly denote which parts of the document are meant to answer what question.

1. Provide a short 2-3 sentence summary of your project. Your TAs are managing many projects at once so this will help jog their memory.
2. State your research questions. You may have tweaked your research question(s) based on the proposal feedback. If this is the case, please state what the old research question was, what the new one is, and why you changed it. If you have not changed your research questions, please explicitly state so.

In cases where you have not incorporated your TA's feedback, please briefly describe why. This will help us ensure that everyone is on the same page with regards to the intent and direction of your project.

3. Describe the data cleaning steps that you had to undertake before/after the datasets were combined. In the case you did not have to perform any data cleaning, describe the steps that you undertook in order to come to the conclusion that no data cleaning was necessary.
4. Exploratory data analysis (EDA) is often the first step in a data science project. This involves looking through your data to identify outliers and/or missing values as well as any interesting trends. The process of exploratory data analysis is often done through visualization (e.g., creating a scatterplot).

Create one visualization to try and spot a trend in your data. You must create the visualization by using Python. You can use libraries such as matplotlib, Seaborn, or Vega-Altair. If you want to use another library, check in with your project mentor first.

- a. Submit the visualization
 - b. Submit the code used to generate the visualization as well as the data files. In other words, if your TA were to take your data files and run your code, they should be able to recreate your visualization.
 - c. Explain what this visualization is meant to examine and why it is suitable for your given data.
 - d. Explain the follow up work that needs to be done given your EDA results.
5. Explain how the results from your EDA process will affect the way you handle your data. If there is no effect, explain why you believe that this is the case.
6. Produce a SQL script that will load data into your database. The SQL script you produce will be similar to the [bookbiz.sql file found on the CPSC 368 website](#). You do not have load data that your research questions do not need.
7. The schema for the data that has been loaded into the database. Be sure to indicate the primary, candidate, and foreign keys. You must have at least two tables where one table contains a foreign key referencing the other. All other constraints such as candidate keys and NOT NULL constraints have been included.
8. 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.