

Weekly Progress Report

(Methods + Data + Implementation)

Overview

This week, you will (1) expand the **Methods** section from last week's report, (2) identify and load a relevant dataset for your capstone project, and (3) implement and test your methods on real data. The goal is to move from planning to execution: you should have a dataset in hand, working code, and evidence that your analysis pipeline runs end-to-end.

Task Breakdown

1. Expand the Methods Section (from last week)

Update and expand the **Methods** section of your previous report. At minimum, include:

- **Problem definition:** what you are predicting/estimating/explaining and what the inputs/outputs are.
- **Data description (initial):** what data you plan to use, key variables/features, and expected sample size.
- **Preprocessing plan:** cleaning steps, missing data handling, encoding, normalization/scaling, train/test split strategy, etc.
- **Modeling/analysis plan:** algorithms/methods you will use and why they fit the problem.
- **Evaluation plan:** metrics (e.g., accuracy, RMSE, AUC), baselines, and how you will validate results.

2. Find a Relevant Dataset and Verify You Can Load It

Find a dataset relevant to your capstone project. This can be:

- a dataset from a **published paper** (recommended if you plan to reproduce results), or
- a public dataset you are interested in analyzing.

Requirements:

- Provide the dataset source (paper link, repository link, or dataset citation).
- Confirm the dataset is accessible (no paywall you cannot access).

- Demonstrate that you can **load the data into Python/R** and perform basic checks (dimensions, variable names/types, missing values).

3. Implement the Code/Methods and Test on Real Data

You must implement your proposed methods and test them in one of the following ways:

- **Option A (recommended):** reproduce at least one key result from **one peer-reviewed paper**.
- **Option B:** test your code on your chosen dataset and report preliminary results (even if the paper you chose does not provide a direct benchmark).

Your report must include:

- what you implemented (brief description of the pipeline),
- what you tested (which dataset, which subset, which variables),
- what worked and what did not (bugs, limitations, unexpected issues),
- preliminary outputs (tables/figures/metrics).

4. Advisor-Led Projects (If You Work With Another Professor)

If you are working with another faculty advisor (e.g., Dr. Seals or Dr. Cohen), you should:

- follow your advisor's instructions for the week, and
- use this report to document your dataset, analysis progress, and next steps.

Formatting and Submission

- Submit a single PDF on Canvas.
- You may write your report in Overleaf or on your GitHub Website and export to PDF.
- Your report should be clear and readable. Use figures/tables where appropriate.
- Include a short code appendix or provide a GitHub link to your code (if applicable).

Suggested Report Structure

Use the following headings (you may add more if needed):

- 1. Project Goal and Research Question
- 2. Expanded Methods (updated from last week)
- 3. Dataset and Access
- 4. Implementation and Experiments
- 5. Results (preliminary)
- 6. Issues / Limitations
- 7. Next Steps

Grading Metric

Your submission will be evaluated using the following criteria:

- **Expanded Methods Section (30%)**
Clear, detailed, and correct methodology. Appropriate evaluation plan.
- **Dataset Identification and Loading (25%)**
Dataset relevance to the capstone, proper citation/link, and proof you can load and inspect it.
- **Implementation and Testing (35%)**
Working code and meaningful testing (paper reproduction or real-data experiments). Evidence via outputs/metrics/figures.
- **Clarity and Organization (10%)**
Well-structured writing, readable figures/tables, and clear next steps.

Please reach out if you have questions or run into access issues with datasets.