Project Instructions — Stage 6: Data Preprocessing

Today's Project Contribution:
Today you'll complete a piece of your full data project.
This task aligns with the Data Preprocessing stage, where you will:
• Clean, transform, and prepare the dataset for modeling.
• Add to your existing project repo (or update prior files)

By the end of this assignment, your project should include the elements listed below.

Deliverable Options

- Load and structure the raw data.
- Develop reusable data cleaning/preprocessing functions.
- Store these functions in `src/cleaning.py`.
- Save this script in the `/src/` folder to keep code modular and organized.
- Document assumptions made during cleaning and their rationale.
- Include a notebook demonstrating the preprocessing transformations.
- Add documentation in the notebook or a dedicated section in the project README.
- Save the preprocessed dataset.

How This Fits Into Your Final Project

Your work today builds toward a complete, end-to-end project.

In your homework, you produced reusable cleaning functions and applied them to a dataset.
Now, you will adapt those functions to clean and prepare your actual project dataset.

Before next class:

- Save your files in the appropriate folders (`/data/`, `/src/`, `/notebooks/`)
- Commit and push your changes to your GitHub repo
- Review any assumptions, risks, or notes — these will carry across your stages

By the end of the course, your full project will follow the complete lifecycle.