Project Instructions — Stage 05: Data Storage (preview)

Today's Project Contribution:
Today you'll complete a piece of your full data project.
This task aligns with the **Data Storage** lifecycle stage, where you will:
• Save and reload raw data reproducibly using environment-driven paths.
• Add storage conventions and documentation to your repo.

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

Deliverable Options (from Stage Scaffold)

Required:

- `data/raw/` and `data/processed/` folder structure.
- At least one saved raw data file (e.g., `.csv` or `.parquet`).
- Add a **Data Storage** section to `README.md` explaining:
    - data folder structure,
    - file formats used, and
    - how your code reads from these files.
- Use of `.env` + secure file path handling (e.g., via `os.getenv(...)`) — especially if loading from protected locations or environment variables set earlier.

How This Fits Into Your Final Project:
Your work today ensures anyone can **recreate your data state** by running your IO code, unlocking reliable preprocessing and modeling later.

Before next class:

- Save files in `/data/raw/` and `/data/processed/` as appropriate.
- Commit and push changes to your GitHub repo (never commit `.env`).
- Review assumptions, risks, and validation notes in your notebook/README.

Explicit Chain:
In your **homework**, you produced **CSV + Parquet saves, reloads, and IO utilities with env-driven paths**. **Now, you will adapt that work** to **formalize your project's storage layer and documentation** so teammates can run it without path edits.

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