Project Instructions — Stage 04: Data Acquisition and Ingestion

Today's Project Contribution:
Today you'll complete a piece of your full data project.
This task aligns with the **Data Acquisition/Ingestion** lifecycle stage, where you will:
• Acquire data programmatically using an API and/or scraping a permitted public table.
• Add raw data and ingestion scripts/notebooks to your existing project repo.

By the end of this assignment, your project should include the elements listed below.
Deliverable Options (from Stage Scaffold):

Required:

- `.env` file storing API keys, tokens, or file paths (**not committed**); include `.env.example` in the repo.
- Raw data saved in `data/raw/` folder (CSV, JSON, etc.).

Optional (Choose One):

- API pull script.
- Web scraping notebook.
- Pre-collected dataset with short written justification.

How This Fits Into Your Final Project:
Your work today builds toward a complete, end-to-end project by establishing a **trusted raw data foundation** for later EDA and modeling.

Before next class:

- Save files to appropriate folders (`/data/raw/`, `/src/` if scripting, `/notebooks/`).
- Commit and push changes to GitHub. Ensure `.env` is ignored.
- Document sources, parameters, and validations in your README or notebook.

Explicit Chain:
In your **homework**, you produced an **API pull and a scraped table with validations**.
Now, you will adapt that work to **formalize your project's ingestion layer**, ensuring raw files are saved with reproducible filenames, secrets are managed via `.env`, and documentation is updated.

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