

Project Instructions — Stage 02: Tooling Setup

Today's Project Contribution:

Today you'll complete a piece of your full data project.

This task aligns with the **Tooling Setup** lifecycle stage, where you will:

- Set up your project environment and file structure.
- Initialize version control and connect your GitHub repository.
- Create the scaffolding necessary to support future work in your pipeline. By the end of this assignment, your project should include the elements listed below.

Deliverable Options (from Stage Scaffold):

Required:

- A GitHub repository initialized with the following structure:

```
/data/  
/data/raw/  
/notebooks/  
/src/  
/docs/  
README.md
```

- A `.gitignore` file that excludes sensitive files (e.g. `.env`) and system clutter.
- A `README.md` file explaining the project purpose and structure.

Optional (Choose One):

- A `requirements.txt` or `environment.yml` file for reproducible environment setup.
- A `Makefile` or setup script for automating tasks.
- A `pyproject.toml` or `setup.py` file as preparation for packaging code.

How This Fits Into Your Final Project:

Your work today builds toward a complete, end-to-end project by establishing a **maintainable and organized technical foundation**. Good structure, version control, and automation will support all future stages of your project lifecycle.

Before next class:

- Save and organize your files using the designated folder structure.
- Commit and push your initial setup to GitHub.
- Ensure `.gitignore` and `README.md` are included and up to date.

Explicit Chain:

In your **homework**, you planned your repo setup and connected version control.

Now, you will adapt that work to scaffold your live project repo, preparing it to support data ingestion, preprocessing, modeling, and reporting.

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