Project Instructions — Stage 7: Outlier Analysis

Today's Project Contribution: Today you'll complete a piece of your full data project. This task aligns with the **Outlier Analysis** lifecycle stage, where you will: • Define outliers and assess their impact on the analysis and model results. • Add to your existing project repo (or update prior files).

Deliverables

Required

- A function to detect, remove, or flag outliers (place in `src/`, e.g., `src/outliers.py`). The function must be reusable and documented.

Optional (choose one or more)

- A sensitivity analysis comparing results with and without outliers (e.g., a notebook `notebooks/sensitivity_outliers.ipynb` with a comparison table).
- A visual comparison of data or model with and without outliers (e.g., boxplots, regression fits).
- An explanation of outlier assumptions, including your definition and potential risks to results (add to README or `docs/outliers.md`).

How This Fits Into Your Final Project

Your work today builds toward a complete, end-to-end project by solidifying your **data cleaning & preprocessing** stage with explicit outlier management. It ensures later modeling and reporting rest on documented assumptions.

Before next class

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

Chain: In your homework, you produced reusable outlier detection/handling code and a sensitivity comparison. Now, you will adapt that code into your **project repository**, integrate it with your pipeline, and document your outlier assumptions where teammates and reviewers can find them.