

# Project Instructions — Stage 10a

---

## Today's Project Contribution

Today you'll complete a piece of your full data project. This task aligns with the **Modeling** stage, where you will:

- Select and fit an appropriate **model type** for your problem
  - You only need to choose one: regression, classification or time series
- You may choose to use classification or time series rather than regression.
- If you use regression,
  - Do a train-test split and
  - Find appropriate features and engineered features to make your regression not violate the regression assumptions.
  - Use diagnostic plots.
  - Evaluate coefficients and
  - Document the chosen features and transformations to create those features.
  - Automate the modeling process so you can auto-try the model with some variations.
  - Document what the coefficients are telling you and why you chose those features.
- Evaluate your model with **appropriate metrics** and include a short, risk-aware interpretation.

## Deliverable Options

Choose **at least one** modeling track:

- Regression model
- Classification model (you may choose this - coming next)
- Time Series (you may choose this - coming next)

## Required

- Code that fits the model(s)
- Save your modeling notebook in the [/notebooks/](#) folder
- Residual or prediction error analysis
- Explanation of key modeling assumptions

## How This Fits Into Your Final Project

Your work today builds toward a complete, end-to-end project by establishing a first, reproducible modeling baseline with proper time-aware logic and metrics.

## Before Next Class

- Save files in [/notebooks/](#) and any helpers in [/src/](#)
- Commit and push your changes to GitHub
- Review assumptions, risks, and notes — these carry across stages

## Chain Statement

**In your homework, you produced** lag/rolling features and an initial pipeline on your dataset. **Now, you will adapt** that work to finalize a project-ready modeling notebook with evaluation and assumptions documented.