

# Homework 0: Data Engineering with Class Data

August 29, 2025

You may use **R** or **Python** for this assignment.

## Tasks

### 1. Data Cleaning & Preparation

- Carefully inspect every column in the dataset.
- Clean entries so that values are consistent and interpretable.
- Ensure that missing values remain only where data was originally missing (no unintended information loss).
- Convert each column into its appropriate data type (e.g., numeric, categorical, text, date).

### 2. Univariate Exploration

- Create clear visualizations to explore the distribution of each column individually.
- Choose appropriate graphs (e.g., histograms, bar plots, box plots, word clouds, etc.) depending on the type of data.

### 3. Bivariate & Multivariate Exploration

- Investigate relationships between two variables at a time (bivariate).
- Extend this to explore interactions among three or more variables (multivariate).
- Use visualizations (e.g., scatter plots, grouped bar charts, heatmaps, pair plots) to highlight trends, correlations, or patterns.

## Deliverables

- A cleaned version of the dataset.
- A set of visualizations with captions or notes that describe what you observe.
- A short summary (1–2 paragraphs) of your key findings from the univariate, bivariate, and multivariate analyses.