

Project Analysis Framework

Data Preparation and Preprocessing

1. Loading Data (15 points)

- Load the provided Public Profit and Emission Database and OWID CO2 Dataset. Use these to perform a firm-level analysis involving profitability and emissions.
- Load the CarbonPlan CDR Database to incorporate data on carbon dioxide removal (CDR) costs.
- You must use at least two datasets, and all processing (e.g., merging, reshaping) must be done within the .qmd file.

2. Handling Missing Data and Cleaning (15 points)

- Identify missing values for each dataset and conduct necessary preprocessing. For instance, identify variables with many missing values and explore why they occur.
- After handling missing values, categorize each variable using Altair's syntax (e.g., Quantitative, Nominal).

3. Data Merging and Structuring (10 points)

- Merge carbon emissions and profitability data to create a foundational data frame for analysis.
- Identify and address any potential issues, such as duplicate entries or inconsistencies, during the merge process.

Visualization and Analysis

1. Creating Static Plots (25 points)

- Create at least two static plots using Altair or GeoPandas. For example:
 - Generate a scatter plot showing the **relationship between firm profitability and emissions**.
 - Create a bar chart to **compare emission efficiencies across industries**.
 - Visually represent emission distributions to show which sectors contribute the most.

2. Developing a Shiny Dashboard (30 points)

- Develop a Shiny dashboard based on your selected analysis topic, allowing users to dynamically explore specific industries or firms' emissions data.
- Include at least one dynamic plot in the dashboard to make the analysis more interactive and understandable.

3. Modeling CDR Costs (20 points)

- Use CDR cost data to model its economic impacts on firms and industries.
- Utilize Pandas and Altair to compare CDR costs to firm profitability and simulate the economic implications.

Additional Components and Writeup

1. Reproducibility (10 points)

- Share all code on GitHub and provide the dataset links alongside the .qmd file so that TAs can reproduce your analysis.
- The project files and structure should allow TAs to clone the repository, knit the .qmd file, and get the same results.

2. Git Usage (10 points)

- Use multiple branches during the project to separate different phases of analysis.
- Ensure the final repository contains one main branch, and that commit history shows equal contributions from all group members.

3. Extra Credit: Text Analysis (up to 10 points)

- Introduce a text analysis using natural language processing techniques to add additional depth to your analysis. For example, analyze corporate sustainability reports to see how carbon emissions are discussed in relation to profitability.