DECEMBER 5TH, 2025

GROUP PROJECT

Progress Report

FOUNDATIONS OF DATA SCIENCE, DATA 5100

GITHUB REPOSITORY: GROUP-PROJECT

This document outlines the original project problem statement that formulates the question to be answered, data sources utilized to research into the problem being answered, **outlines a progress update that reflects the true state of our project right now,** and calls out any risks to successful completion.

PROBLEM STATEMENT

"Is a higher share of renewable energy in total energy supply associated with lower per-capita CO_2 emissions?"

Global carbon dioxide (CO_2) emissions have continued to rise despite record levels of investment in renewable energy technologies. According to the International Energy Agency (IEA, 2025), energy-related CO_2 emissions reached an all-time high in 2024, driven by growth in total energy demand. Previous research shows that renewable energy expansion can help reduce emissions intensity, but results vary by region and economic structure (Wang & Zhao, 2023; Ozturk & Acaravci, 2013). Other studies note that investment alone may not lead to immediate reductions, as infrastructure transitions and fossil fuel dependencies can delay impacts (Bölük & Mert, 2014).

However, there remains limited research examining the relationship between share of renewable energy investments and national levels of carbon emissions. This project aims to address that gap using cross-country data from 2010-2024 to explore the relationship between renewable energy shares and CO_2 outcomes. We will investigate whether countries with a higher share of renewable energy in their electricity mix emit less carbon overall or per capita. This question matters because decarbonizing national energy systems is central to limiting global warming. Renewable deployment is expanding rapidly, but we anticipate the extent to which it translates into measurable emission reductions will vary across regions. Quantifying this relationship provides evidence of the real-world effectiveness of renewable energy transitions.

DATA SOURCES

Final data sources used in analysis include:

1. **Global Carbon Atlas**. (2025). Global Carbon Atlas Database: CO_2 Emissions by Country. https://globalcarbonatlas.org: Provides annual country-level carbon dioxide emissions (total, per capita, and per GDP) spanning several decades.
2. **World Bank**. (2025). World Development Indicators (WDI). The World Bank Group. https://databank.worldbank.org/source/world-development-indicators: Supplies complementary data on renewable electricity share (% of total electricity generation), GDP, and population.
3. **Energy Institute**. (2025). Statistical Review of World Energy 2025. Energy Institute. https://www.energyinst.org/statistical-review: Provided by the Energy Institute, this can offer us some insight into the demand for energy supply by fuel, region, etc.

We are confident that these sources are sufficient as they provide consistent data for many countries over several decades. Preprocessing has been completed —including merging datasets by **year and country**; handling missing values; and normalizing variable scales by **converting emissions to metric tons per capita**. All the data sources we have identified so far are CSV files available via direct download here: data_processing.ipynb.

ANALYTICAL APPROACH & TECHNIQUES

We have conducted a comparative observational study using quantitative, correlational analysis to examine the relationship between **carbon emissions** and the **share of renewable energy** in national electricity generation.

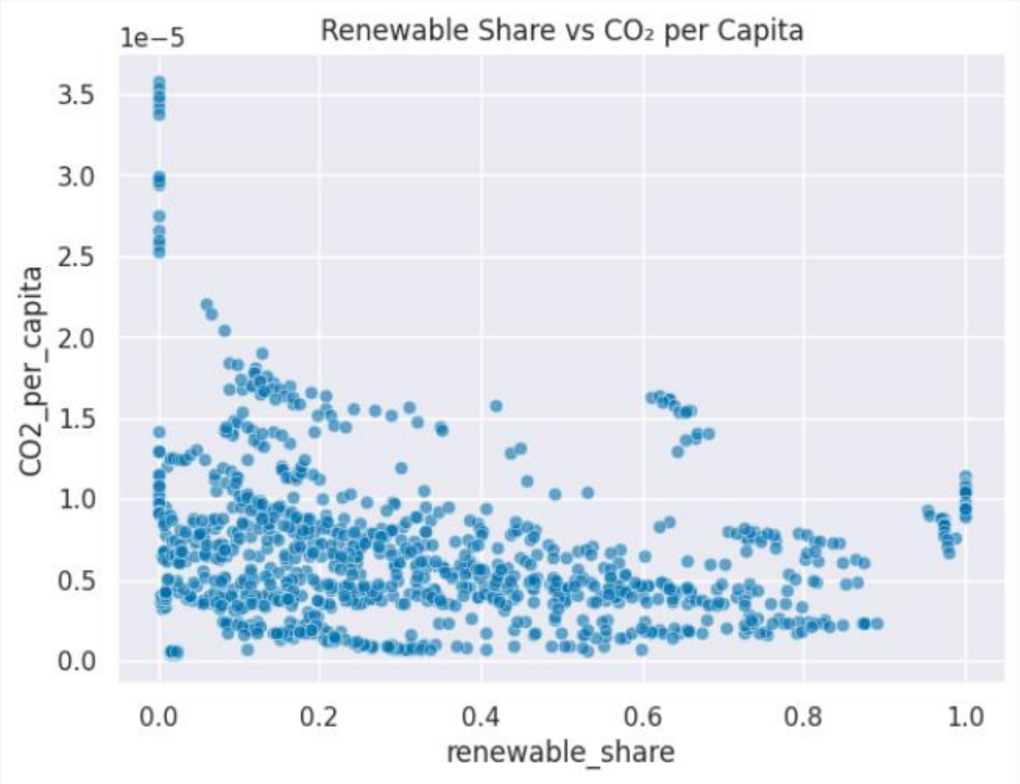
Our analysis uses a country-year panel dataset, containing annual observations from multiple countries, to examine how renewable energy shares and emissions change over time.

Features include analysis across the following variables:

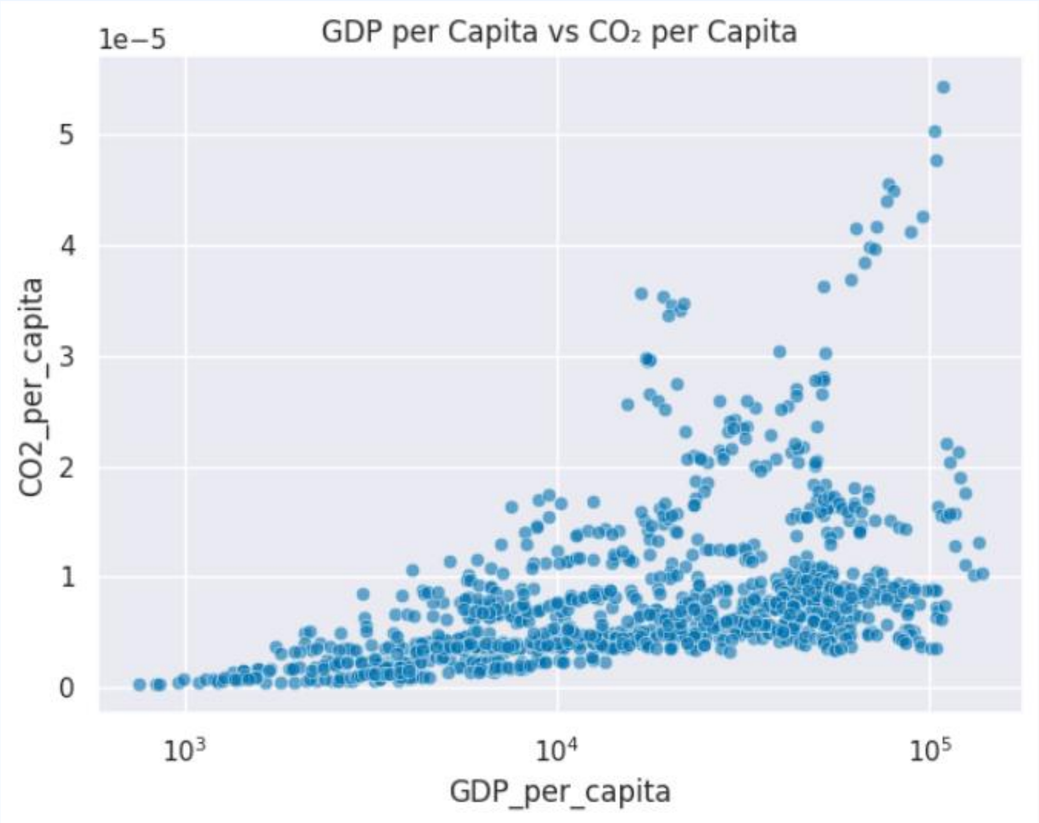
- Country/Regional identifiers
- Country/Regional population
- CO_2 emissions
- GDP per capita
- Renewable Energy Share

Initial exploratory analysis involves traditional descriptive statistics and visualization—scatter plots, time trends, and correlation matrices— to explore global patterns. Please see some examples below.

Analyzing the relationship between ***renewable energy shares*** and ***CO_2 emissions*** *per capita*:



Analyzing the relationship between ***economic output (GDP)*** and ***CO_2 emissions*** *per capita*:



Further analyses include:

- Analyzing global time trends of renewable shares for each year in 2010-2024.
- Comparing global renewable share levels across regions: Top 10 and Bottom 10 Renewable Countries in 2010-2024.
- Modeling a panel regression model to analyze the relationships between renewable energy share per capita, CO_2 emissions per capita, and GDP per capita in 2010-2024.
- Analyzing the delayed effects of "lagging" renewable energy shares in 2010-2024 (i.e. analyzing whether renewable energy reduces carbon emissions immediately, or whether the benefits show up over time).
- Executing a robustness test to check whether our main findings are reliable and don't disappear when we change assumptions, model structure, or variable definitions.

SOLUTIONS TECHNOLOGY

Processing and exploratory analyses has been done in Python as the primary language, specifically: pandas and NumPy for data wrangling and merging; matplotlib and seaborn for visualization; stats models and scikit-learn for regression and exploratory modeling; and Data Processing Notebook in Google Colab for reproducible analysis and documentation.

PLAN FOR COMPLETION

We are at the stage in our analysis where our main result is clear (there is a correlation between higher renewable energy shares and lower CO_2 emissions per capita) and we feel confident in answering the problem statement with the visuals and test results we have now.

To continue building our results, we want to look deeper into **how** renewable energy reduces emissions. To do this, we will finalize our analysis by answering the following expanding questions:

1. Does renewable energy only begin to significantly reduce CO_2 emissions *after* a country reaches a certain renewable share threshold?
2. Do renewable energy shares reduce emissions even while economies grow?

Steps needed to do this include testing for nonlinear effects by adding a squared term for renewable share in our model. If this squared term is significant, it will tell us whether renewable energy becomes more impactful after a threshold or whether the impact changes at high renewable levels. We will also create a simple visualization to show whether emissions start decreasing faster after renewables make up a larger part of the energy mix. This will help us determine if there is a "tipping point" where renewable energy starts making a bigger difference.

We also want to check whether renewable energy helps reduce emissions even when economies are growing. A country's CO_2 emissions can increase simply because people and businesses are using more energy as the economy gets bigger. So instead of only looking at CO_2 per capita, we will also look at **carbon intensity**, which measures CO_2 emissions relative to economic output. To do this, we will create a new variable that divides CO_2 emissions by GDP per person and re-run our model. If renewable energy still shows a strong negative relationship with carbon intensity, that means renewables are helping countries become more energy-efficient and environmentally sustainable, not just reducing emissions because of lower demand. This gives us a stronger result because it shows the benefits of renewables even when economic development continues.

In addition to further analyses, we will work on communicating the overall results by writing the final report and putting together our presentation.

CHALLENGES, UNCERTAINTIES, IMPEDIMENTS

Interpreting results responsibly will be essential. Even statistically significant findings may be shaped by factors outside our dataset, and visualizations or results can be easily misinterpreted if not contextualized. Careful framing and transparent discussion of assumptions will therefore be an essential part of the project, which be emphasis in our presentation and the further analysis.

Another uncertainty is assuring that our audience has a clear understanding of the problem statement we are intending to investigate. There was a lot of analysis done across many variables and its importatnt that we use our further analysis without expanding too broadly into a realm of analysis that is still ambiguous.

RELEVANT CITATIONS

When studying renewable energy and carbon emissions, researchers frequently mention total energy supply (TES) or total primary energy supply (TPES) because it provides context for how big a country's renewable energy share actually is.

Sources that may be used for reference during our project are as follows:

1. **Intergovernmental Panel on Climate Change (IPCC)**. (2011). Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change (SRREN). Cambridge University Press. https://www.ipcc.ch/report/renewable-energy-sources-and-climate-change-mitigation/This is something we can investigate before and after we analyze the data to guide us through foundational discussion of renewables, mitigation potential, and challenges.
2. **Wang, Q., & Zhao, M**. (2023). Can renewable energy investment reduce carbon dioxide emissions? Evidence from scale and structure. Energy Economics, 112, 106215. https://www.sciencedirect.com/science/article/abs/pii/S0140988322003334A theoretical paper that addresses one of our core questions by linking the renewable energy investment to CO_2 emissions while decomposing effects into scale, technique, and structure.
3. **Liu, Y., Zhang, H., & Li, X**. (2023). The effect of clean energy investment on CO_2 emissions. Energy Economics, 123, 107048. https://www.sciencedirect.com/science/article/pii/S014098832300498XA recent study published by Science Diet, diving into quantifying the elasticity between clean energy investment and emissions.
4. **Chen, Z., Li, F., & Wu, S**. (2022). The effect of renewable energy on carbon emissions through globalization. Frontiers in Environmental Science, 10, 960795. https://www.frontiersin.org/articles/10.3389/fenvs.2022.960795/fullPublished by PMC Central, this report assesses direct and indirect (mediated) effects of renewable energy on emissions
5. **Bölük, G., & Mert, M**. (2014). Fossil and renewable energy consumption, economic growth, and carbon emissions: An empirical analysis. Energy Policy, 74, 471–479. https://ideas.repec.org/a/eee/energy/v74y2014icp439-446.htmlProvides us with panel data for the period of 1990-2008.
6. **Ozturk, I., & Acaravci, A**. (2013). The long-run and causal analysis of energy, growth, openness, and financial development on carbon emissions in Turkey. Energy Economics, 36, 262–267. https://econpapers.repec.org/RePEc:eee:eneeco:v:36:y:2013:i:c:p:262-267

Note: Not all citations will be referenced in our final report, and there may be additional sources identified throughout the process.