1. Introduction

The rapid rise in global CO_2 emissions due to fossil fuel consumption has become one of the most pressing environmental challenges of our time. As greenhouse gases like carbon dioxide continue to accumulate in the atmosphere, they contribute to global warming, causing widespread climate change that affects ecosystems, weather patterns, and human health. Understanding the trends in CO_2 emissions across different countries and their respective energy sources is critical for shaping effective policies to mitigate climate impacts. Therefore, our primary question is: **How can we better understand the trends and sources of CO_2 emissions globally, and which countries or sources are contributing the most to this issue?**

To address this problem, we will leverage data on fossil CO_2 emissions across multiple countries from 1750 to 2022. By analysing how emissions from different sources—such as coal, oil, gas and cement have changed over time, we can identify trends, investigate the impact of different energy sources, and determine per capita emissions, which help in assessing the relative contribution of each country.

The goal of this project is to develop data-driven insights that can inform policy-makers, environmentalists, and businesses on emission reduction strategies. To achieve this, we will refine the problem into several tangible objectives:
1. **Explore and visualise emission trends** across countries and time to identify key patterns.
2. **Analyse the impact of various emission sources** on total CO_2 output, identifying which energy sources are the most critical drivers of emissions.

The final data product will include:
- **A data pipeline** that processes and cleans the emissions data, ensuring it is ready for analysis.
- **A set of visualisations and statistical models** that provide insights into the trends and drivers of CO_2 emissions.

Through these techniques, this project will contribute to a deeper understanding of global CO_2 emissions and inform actionable strategies to reduce the human impact on climate change.

2. Data Science Techniques

2.1 Exploratory Data Analysis (EDA):

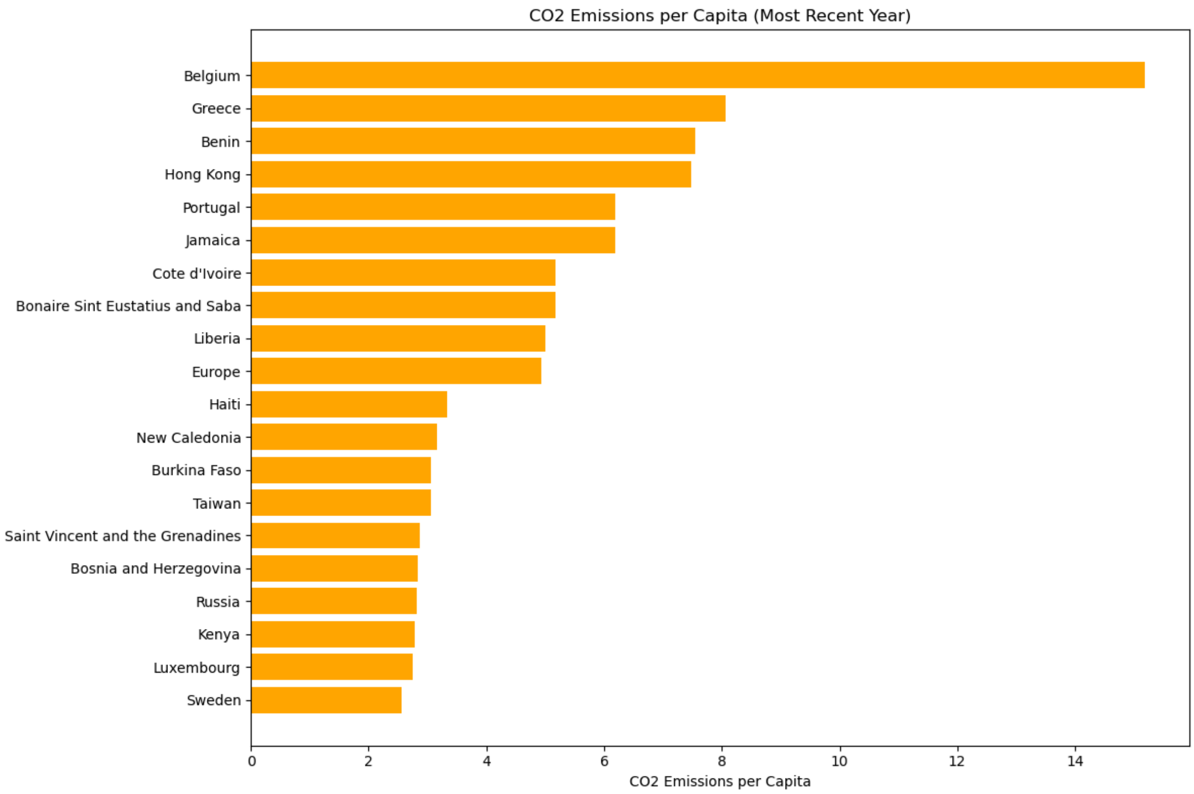
1. Data Visualization:

Use visualisations to uncover patterns in CO_2 emissions over time. Techniques such as:

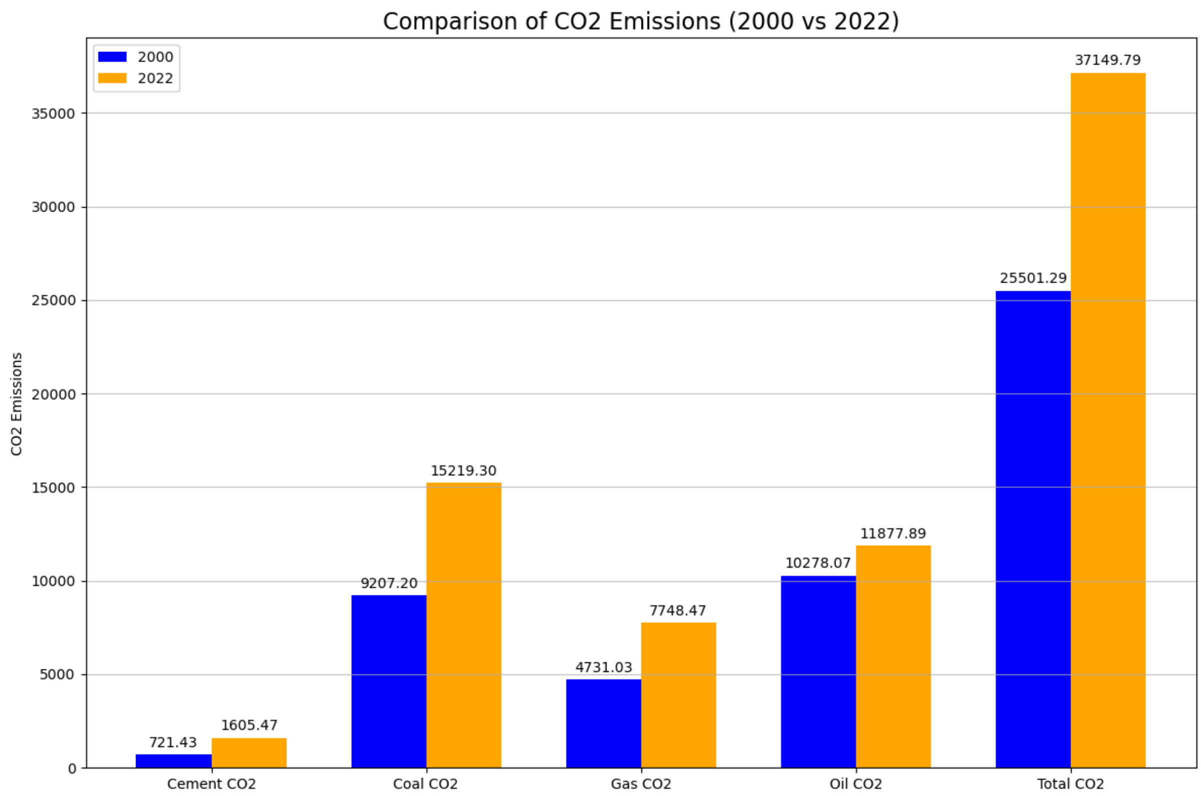
1. Line plots to show the trends of CO_2 emissions (total and by category) for individual countries.
2. Heatmaps to display the emissions per capita across different countries.
3. Bar charts for comparing emissions between countries.
4. Correlations: Investigate how different emission sources (coal, oil, etc.) correlate with total emissions or per capita emissions. This helps determine which energy sources are most responsible for the rise in emissions.

Some visualisation below:

1. **Bar chart of CO2 emissions per capita**



2. **Comparison of CO2 Emissions (2000 vs 2022)**



2. Regression Models:

1. **Linear Regression** can help understand the relationship between population and emissions. We predict total CO_2 emissions based on energy source contributions, or CO_2 per capita as a function of economic indicators.
2. **Multivariate Regression:** Use multiple independent variables (coal, oil, gas, etc.) to predict total emissions and analyse which sources are most impactful.

3. Data Suitability & Potential Challenges

3.1 Suitability:

The dataset is appropriate for regression models given the continuous variables (emissions, per capita emissions) and multiple independent features (coal, oil, gas, etc.).

3.2 Challenges:

1. **Missing Data:** Emissions data may be missing for certain countries or years. This could necessitate imputation techniques or careful analysis of missingness patterns.
2. **Skewed Data:** Emissions data might be heavily skewed, with a few countries contributing disproportionately (e.g., the USA, China). Techniques like log transformation might help normalise the data for analysis.

3. **Multicollinearity:** Some of the emissions variables (coal, oil, gas) might be highly correlated. This can be a challenge for regression models, where multicollinearity can distort results.

4. Timeline

Date	Task/ Milestone	Status
4 September, 2024	Team work contract, Code of conduct	Done
4 September, 2024	Setup github	Done
15 September, 2024	Choose dataset, Explore the dataset	Done
28 September, 2024	Project proposal	In progress
30 October, 2024	Create the analysis model	NA
18-19 November, 2024	Final presentation	NA
25 November, 2024	Project report	NA

5. References

1. International Energy Agency (IEA). (2023). *CO2 Emissions in 2022*. IEA. Paris.
2. Borowiec, J., & Papież, M. (2024). Convergence of CO2 emissions in countries at different stages of development: Do globalization and environmental policies matter? *Energy Policy*, 184, 113866.
3. Yang, S., Wang, X., Ge, Z., Dong, G., Ma, M., & Han, X. (2023). Global Per Capita CO2 Emission Trends. *Atmosphere*, 14(12), 1797.
4. Liu, Z., Deng, Z., Davis, S., et al. (2023). Monitoring global carbon emissions in 2022. *Nature Reviews Earth & Environment*, 4, 205–206.
5. Ahmed, M. U., Karim, M. A., Tahsin, M. S., Rahman, Y., & Tafannum, F. (2022). Analyzing CO2 Emission in Developing Countries Using Auto Regression and Auto Regression Walk Forward: A Time Series Approach. In *2022 IEEE Delhi Section Conference (DELCON)* (pp. 1-5). New Delhi, India.