Analysing Global CO_2 Emission Trends and Sources for Climate Action

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 Data description:

1. **co2**: Annual CO_2 emissions (in million tonnes).
2. **co2_per_capita**: Annual CO_2 emissions per capita (tonnes per person).
3. **co2_growth_abs**: Annual CO_2 emissions growth in absolute terms (million tonnes).
4. **co2_growth_prct**: Annual CO_2 emissions growth in percentage.
5. **co2_per_gdp**: Annual CO_2 emissions per GDP (kg per international dollar).
6. **gdp**: Gross domestic product (adjusted for inflation and cost of living differences).
7. **population**: Population size (number of people).
8. **primary_energy_consumption**: Primary energy consumption (terawatt-hours).
9. **energy_per_capita**: Primary energy consumption per capita (kilowatt-hours per person).
10. **energy_per_gdp**: Primary energy consumption per unit of GDP (kilowatt-hours per international dollar).
11. **cement_co2**: Annual CO_2 emissions from cement production (million tonnes).
12. **coal_co2**: Annual CO_2 emissions from coal (million tonnes).
13. **oil_co2**: Annual CO_2 emissions from oil (million tonnes).
14. **gas_co2**: Annual CO_2 emissions from gas (million tonnes).
15. **flaring_co2**: Annual CO_2 emissions from flaring (million tonnes).
16. **land_use_change_co2**: Annual CO_2 emissions from land-use change (million tonnes).
17. **cumulative_co2**: Cumulative CO_2 emissions (in million tonnes).
18. **cumulative_co2_including_luc**: Cumulative CO_2 emissions including land-use change (in million tonnes).
19. **methane**: Total methane emissions (in million tonnes).
20. **nitrous_oxide**: Total nitrous oxide emissions (in million tonnes).

Note: The dataset has 80 columns, the above is a list of columns we think are important for the analysis.

2.2 Exploratory Data Analysis (EDA):

1. Data Visualization:

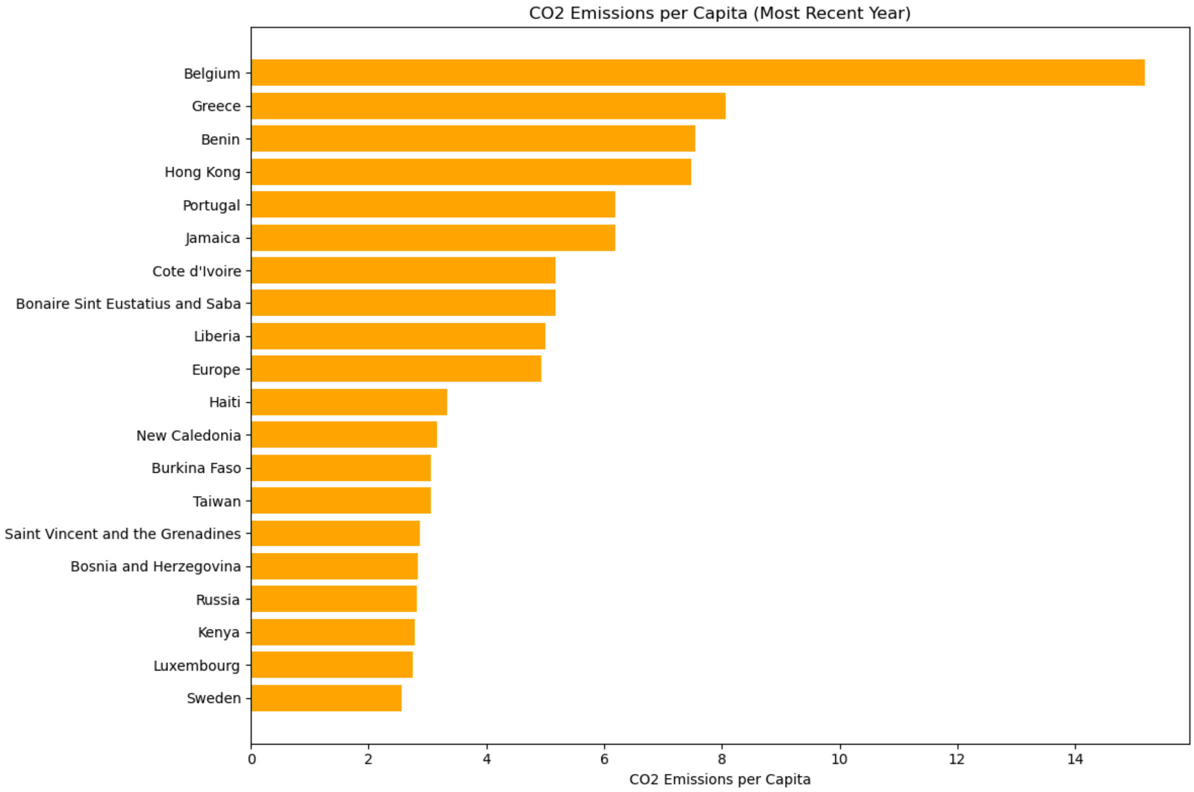
Use visualizations to uncover patterns in CO2 emissions over time. Techniques such as:

1. Line plots to show the trends of CO2 emissions (total and by category) for individual countries.
2. Heat maps are used to display the emissions per capita across different countries.
3. Bar charts for comparing emissions between countries.

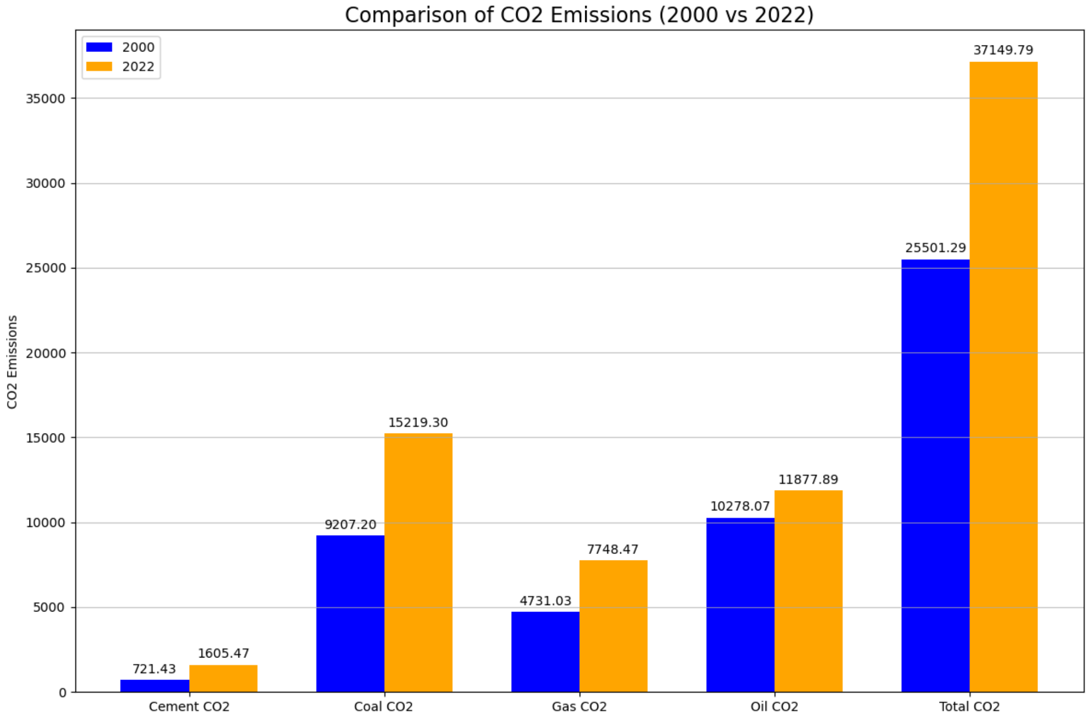
4. Correlations: Investigate how different emission sources (coal, oil, etc.) correlate with total emissions or per capita emissions. This helps determine which energy sources are most responsible for the rise in emissions.

Some visualizations are below:

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



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



2.3 Regression Models:

1. **Linear Regression** can help understand the relationship between population and emissions. We predict total CO2 emissions based on energy source contributions, or CO2 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.

2.4 Success Criteria:

Linear Regression	
R^2	High value (above 0.7) indicates a strong fit between population/economic indicators and emissions.
P-values	Low p-values (< 0.05) suggest that the independent variables significantly influence CO_2 emissions.
MAE & RMSE	Lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) reflect better model performance.

Multivariate Regression	
Adjusted R^2	High adjusted R^2 indicates that multiple factors (coal, oil, gas, etc.) explain a significant part of emissions.
Multicollinearity (VIF)	Low Variance Inflation Factor (VIF) ensures independent variables are not highly correlated.
P-values	Low p-values confirm that energy sources (independent variables) significantly impact emissions.
Feature Importance	Analysis of feature importance helps identify which variables contribute most to emissions.

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. Our World in Data. CO2 Emissions Data. Retrieved from Our World in Data GitHub Repository https://github.com/owid/co2-data.

2. International Energy Agency (IEA). (2023). *CO2 Emissions in 2022*. IEA. Paris.
3. 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.
4. Yang, S., Wang, X., Ge, Z., Dong, G., Ma, M., & Han, X. (2023). Global Per Capita CO2 Emission Trends. *Atmosphere*, 14(12), 1797.
5. Liu, Z., Deng, Z., Davis, S., et al. (2023). Monitoring global carbon emissions in 2022. *Nature Reviews Earth & Environment*, 4, 205–206.
6. 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.