Global Carbon Emission Patterns

(COMP3125 Individual Project)

School of Computing and Data Science

***Abstract* – Nowadays, carbon emissions is one of the most talked global concerns, which cause climate change, influence environmental policies and economic sustainability. This project analyzes carbon emissions across various countries using datasets that contain carbon emissions and GDP figures. The goal of this project is to analyze patterns between economic growth and emissions, and categorize countries based on their carbon footprint.**

With the help of Python, we will use statistical analysis, data visualization and machine learning techniques such as linear regression and K-means clustering to analyze results. The study highlights a big difference in emission per capita, the relationship between GDP and CO_2 emissions. The results will offer great source of knowledge for policymakers, researchers and economists.

Keywords: CO_2 emissions, GDP, climate change, data analysis, machine learning

Introduction (Heading 1)

The Industrial revolution changed everything in the world. With the invention of internal combustion engines, as wood, peat and coal were not efficient enough, the need for more efficient and versatile fuel types arose. As petroleum was more efficient and easily transportable, it became the main fuel type all over the world very soon. With the oil boom countries with huge oil reserves started to extract the oil in huge amounts. However, even though it was very efficient as a fuel type, it was very dangerous for nature. This project aims to find answers for 3 questions:

1. “Is there a relationship between a country’s GDP and its CO_2 emissions?”
2. “How does GDP influence CO_2 emissions across countries, and does this relationship change when considering per capita values?”
3. “Do wealthier countries emit more CO_2 per person compared to developing countries?”.

This individual project’s goal is to analyze the relationship between the biggest carbon emitters and their GDP values.

Datasets

A. Source of dataset(Heading 2)

In this project I used 2 datasets. I obtained CO_2 emissions dataset from Kaggle, and GDP per country dataset from MarketWatch. Both of these sources are credible and widely used. Here is more information:

1. Carbon Emissions by country – From 1990 to 2019
2. Gross Domestic Product by country – From 1960 to 2023

B. Characteristics of datasets

Both datasets were downloaded in .csv format. GDP by country dataset needed some preprocessing.

Dataset	**Variables**	**Units**	**Processing steps**
CO_2 emissions by country	Country, Region, Date	Kilotons of CO_2, Metric Tons Per Capita	None
GDP per country	Country name and code, Indicator name and code, years	Dollar amount	Skipped first 4 rows

Used Pandas to merge CO_2 emissions dataset with the GDP dataset on the Country and Date columns. I also used Seaborn hue for visualization. I did not perform any unit conversions, as all the values were used in their original form. The data cleaning process involved cleaning metadata rows from GDP dataset where I skipped first four rows and I also handled missing values ensuring consistency across the dataset before analysis.

Methodology

A. K-Means Clustering Analysis

K-Means clustering was used to identify patterns in **CO_2 emissions across countries**. This method helps categorize countries into distinct groups based on their average emissions.

Implementation:

```
X_cluster = df.groupby('Country')[['Kilotons of Co2']].mean()

kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
```

```
kmeans.fit(X_cluster)

X_cluster['Cluster'] = kmeans.labels_
```

The **advantages** of K-Means clustering include its **simplicity, efficiency, and ability to uncover patterns in emissions data**. However, a **key limitation** is that the number of clusters must be specified in advance, rather than being determined automatically.

B. Correlation Analysis

This analysis examines the relationship between **GDP and CO_2 emissions** to understand how economic activity influences environmental impact.

Implementation:

```
merged_df = df.merge(df_gdp, on=['Country', 'Date'], how='inner')

X = merged_df[['GDP']]

y = merged_df['Kilotons of Co2']

model = LinearRegression()

model.fit(X, y)
```

The **advantages** of this method include its **ease of implementation and clear visualization of economic-environmental trends**. However, its **limitations** include detecting only **linear relationships** and being **sensitive to outliers** in the data.

C. Data visualization

1. **Line Plot**

   Illustrated total CO_2 emissions across different regions over time

2. **Scatter Plot**

   Explored relationship between GDP and CO_2 emissions

3. **Bar Charts**
   Visually illustrated top and bottom countries by CO_2 emissions by capita

4. **Heatmaps**
   To identify correlations between GDP, total CO_2 emissions, and CO_2 emissions per capita

Results

A. Global CO_2 Emission Patterns

Time series analysis of global CO_2 emissions revealed clear regional patterns and evolving trends over the years.

Key Observations:

- Overall increase in CO_2 emissions over time, particularly in regions undergoing industrialization.
- While some regions like Asia showed a steep increase, others as Europe had plateaus or slight declines because of the environmental policies and adoption of renewable energy sources.
- The visualization helped identify which regions contribute most significantly to global carbon emissions.

B. Economic Correlations

Correlation and regression analysis showed the relationship between CO_2 emission, GDP and per capita emission levels.

1. **GDP vs CO_2 Emissions**
    - Revealed a strong positive correlation (0.81)
    - This proves that countries with higher GDPs tend to emit more CO_2 to the atmosphere. This confirms the connection between economic activity and environmental impact.
    - Regression line in the scatter plot also indicated this trend

    Some Notes:

    - Some high GDP countries emitted relatively less CO_2. The reason is these countries try to use cleaner energy sources as renewable energy sources and nuclear power.
    - There were also some countries that had high emissions, but lower GDP. The reason is because these countries are heavy industry reliant.

2. **GDP vs CO_2 emission per capita**

- The graph revealed a weaker correlation. This means high GDP did not always mean higher GDP per capita.
- The reason behind this could be several factors as population size, energy efficiency and policies behind determining per capita impact.

C. Country Comparison and Emission Inequality

Bar charts illustrated top and bottom 10 countries based on CO_2 emission per capita. The graph shows:

Top Emitters Per Capita

- Countries such as Qatar, Kuwait and UAE ranked among the highest
- Even though, these countries have smaller populations, they have high fossil fuel usage and energy exports.

Bottom Emitters Per Capita

- This list mostly consists of developing or underdeveloped countries of Africa and South Asia
- Low per capita emissions reflect limited industrialization, fewer vehicles and reduced electricity consumption. These countries are mostly poor countries.

Unexpected Findings

- Some wealthy nations maintained moderate per capita emissions. This indicates that these countries had effective sustainability strategies.
- There was a clear inequality in environmental impact when emissions were measured on a per-person basis. The reason being while some countries had high emissions they also had high population, so their emissions per capita was not as high.

Discussion

The deeper we dig into global carbon emission trends, the more it became clear that even though GDP is one of the main indicators of a country's carbon footprint, the full story is much more complex. Our research confirmed the expected – wealthier nations tend to emit more CO_2 , but

they also highlighted several different points that proved that the assumptions about economic development and environmental impact are oversimplified. This analysis was structured around three research questions:

1. **Is there a relationship between a country's GDP and its CO_2 emissions?**

Our findings proved there is a huge correlation between GDP and CO_2 emissions. Top CO_2 emitters all had high GDPs. This indicates that in general wealthier countries tend to emit more CO_2 due to higher levels of industrial activities, energy consumption and transportation demands. However, this relationship varied across regions. While most of the rich Asian countries had very high CO_2 emissions, most of the rich European countries had comparatively lower emissions. The reason behind this is European countries invest more in clean energy technologies and they adopt stricter governmental regulations.

2. **How does GDP influence CO_2 emissions across countries, and does this relationship change when considering per capita values?**

On the other hand, if we look at the per capita values, the relationship with GDP became less clear. While countries like Qatar and UAE displayed both high GDP and high per capita emissions due to fossil fuel dependency and energy intensive industries, countries like Japan and Germany demonstrated high GDP but comparatively lower per capita emissions. This suggests that, factors like energy efficiency, government policies, technological advancements and population size all contribute to variations in emissions intensity.

3. **Do wealthier countries emit more CO_2 per person compared to developing countries?**

Surprisingly, this was not always the case. Some wealthy countries, especially those in Western Europe, had moderate or even lower per capita emissions, which indicates successful implementation of sustainable development practices. On the other hand, smaller countries with rich oil reserves ranked among the highest per capita emitters even though they had smaller populations. We also saw that developing countries that did not have huge oil reserves had low GDPs and often very low per capita emissions. This underlines the environmental inequality problems. Even though they contribute the least they are the most vulnerable to climate change.

Conclusion

To sum up, this project provided valuable details about the relationship between economic growth and carbon emissions across the world. Thanks to the analysis of GDP and CO_2 records, we discovered that although economic development is frequently associated with increased emissions, this is not always the case. Even though countries with high GDPs generally contribute more to global emissions, several exceptions demonstrated how investing in clean and green policies and technological innovation can lower the emission levels while having the same or more economic developments.

The use of machine learning techniques like K-means clustering helped to sort countries based on their emission levels, illustrating inequalities in per capita emissions. While some nations emit large amounts of CO_2 due to industrialization and energy consumption, others contribute minimally, due to lower economic activity or population size.

The findings show us the importance of environmental policies that consider not just total emissions, but also per capita values and economic contexts. Moving forward, global cooperation and sustainable development strategies will be essential in balancing economic gains with environmental responsibilities. This study can serve as an important data for policymakers, economists, and environmentalists who seek data-driven approaches to combat climate change.

References

- Statista. (2023). *World carbon dioxide emissions by region*. Retrieved from https://www.statista.com/statistics/205966/world-carbon-dioxide-emissions-by-region/

- Our World in Data. (2023). *CO_2 emissions vs GDP*.

  https://ourworldindata.org/grapher/co2-emissions-vs-gdp

- International Energy Agency. (2023). *The relationship between growth in GDP and CO_2 has loosened – it needs to be cut completely*.

  https://www.iea.org/commentaries/the-relationship-between-growth-in-gdp-and-co2-has-loosened-it-needs-to-be-cut-completely

- World economies' progress in decoupling from CO2 emissions
  https://www.nature.com/articles/s41598-024-71101-2