Forecasting Carbon Emissions Across Continents

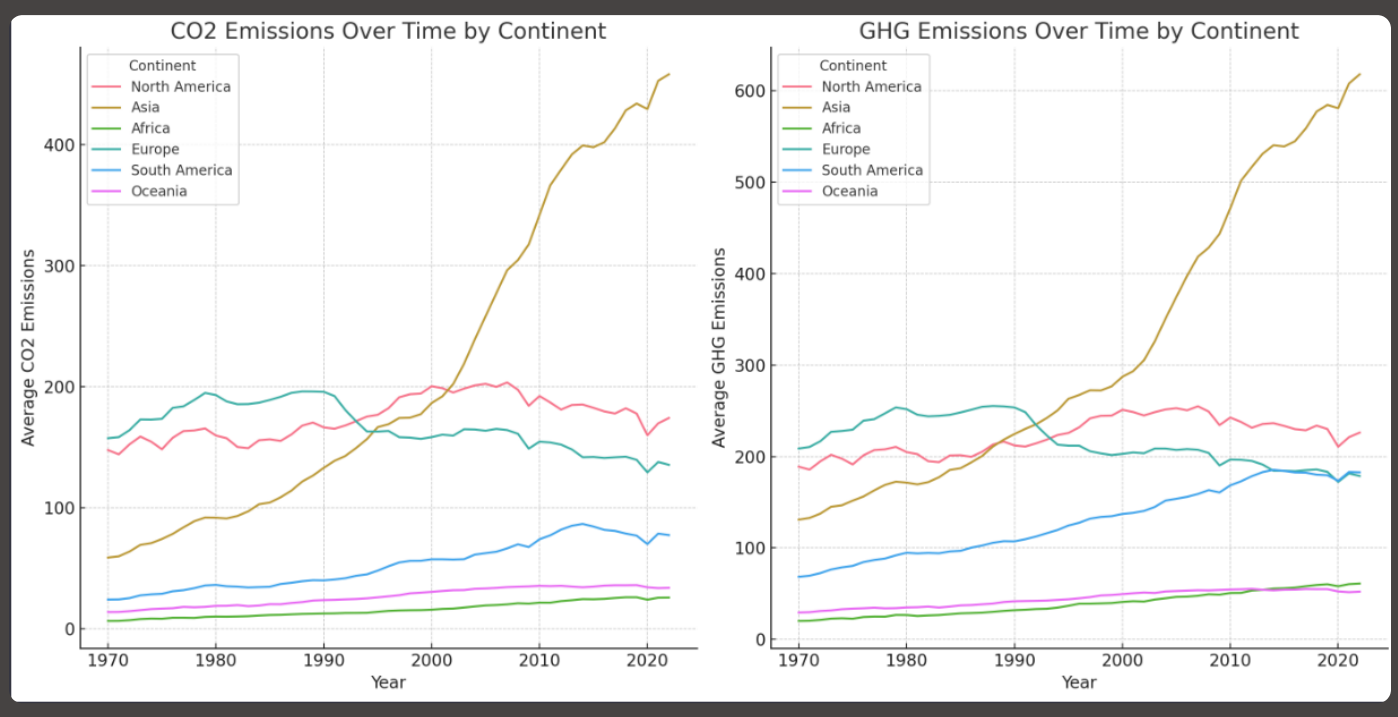
This report focusses on various aspects related to emissions and greenhouse gases. Through exploratory data analysis, we dive into key statistical measures, correlations, and temporal trends. This report provides valuable insights and highlights the significance of understanding these environmental factors. The report finally concludes with the results of a prediction model that estimates the CO2 and GHG emissions for the next three years.

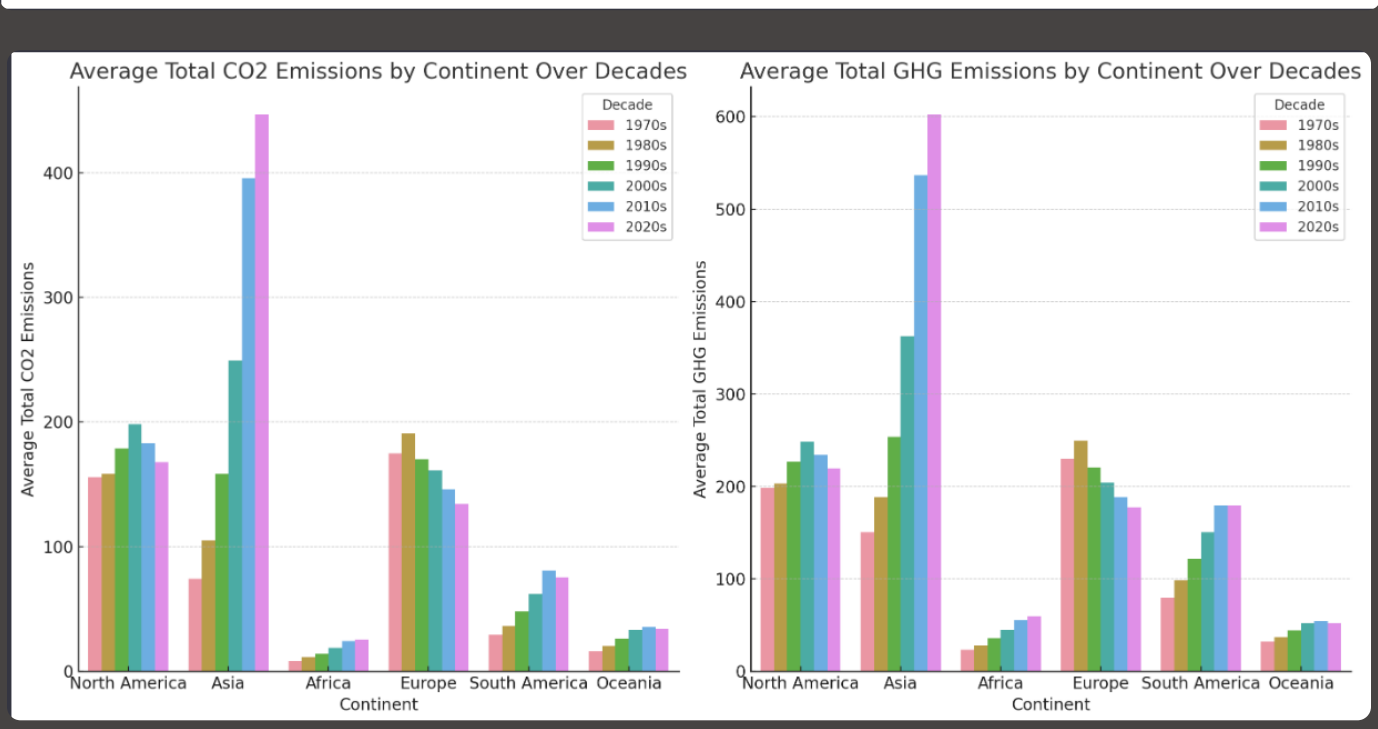
by Gagan Aryan

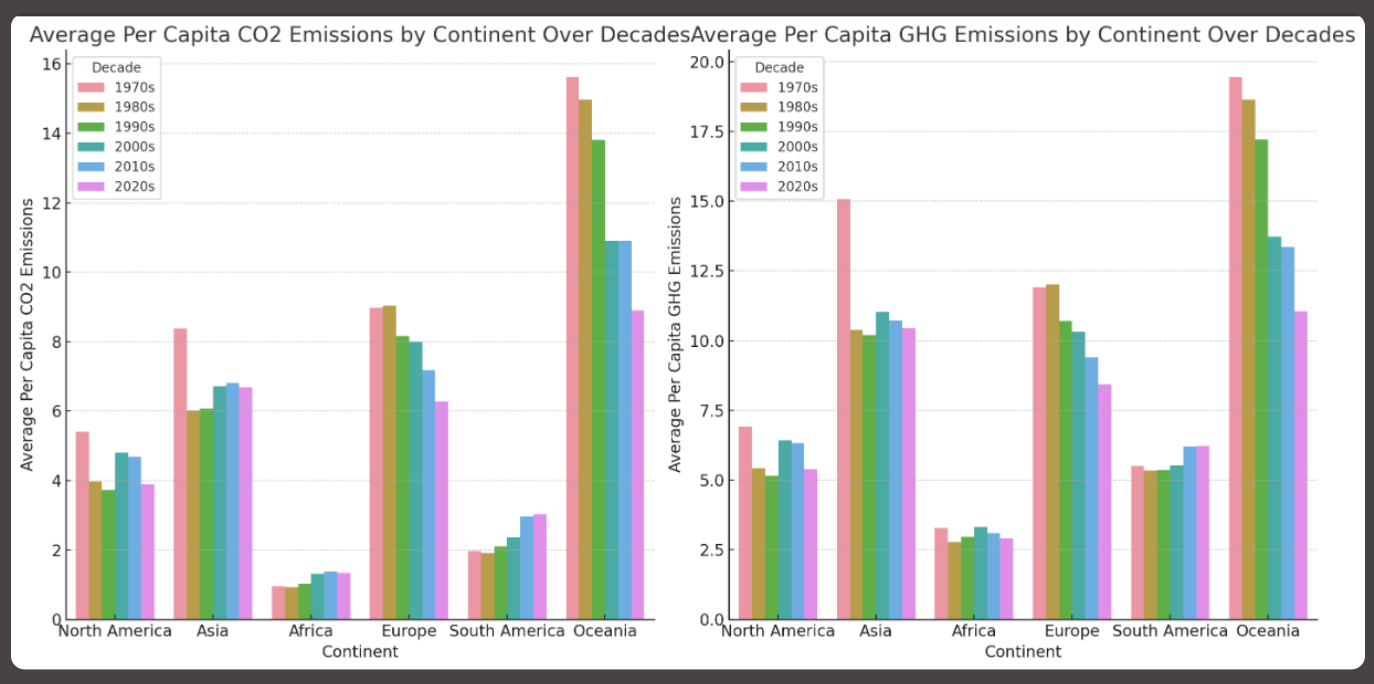
Exploratory Data Analysis

In this section, we perform in-depth exploratory data analysis (EDA) on a dataset spanning continents. By categorizing outcomes and insights based on continents, we uncover key statistical measures, patterns, and trends. We also discuss data cleaning and preprocessing approaches, emphasizing addressing missing values and outliers while ensuring data uniformity.

Time based comparative graph across continents



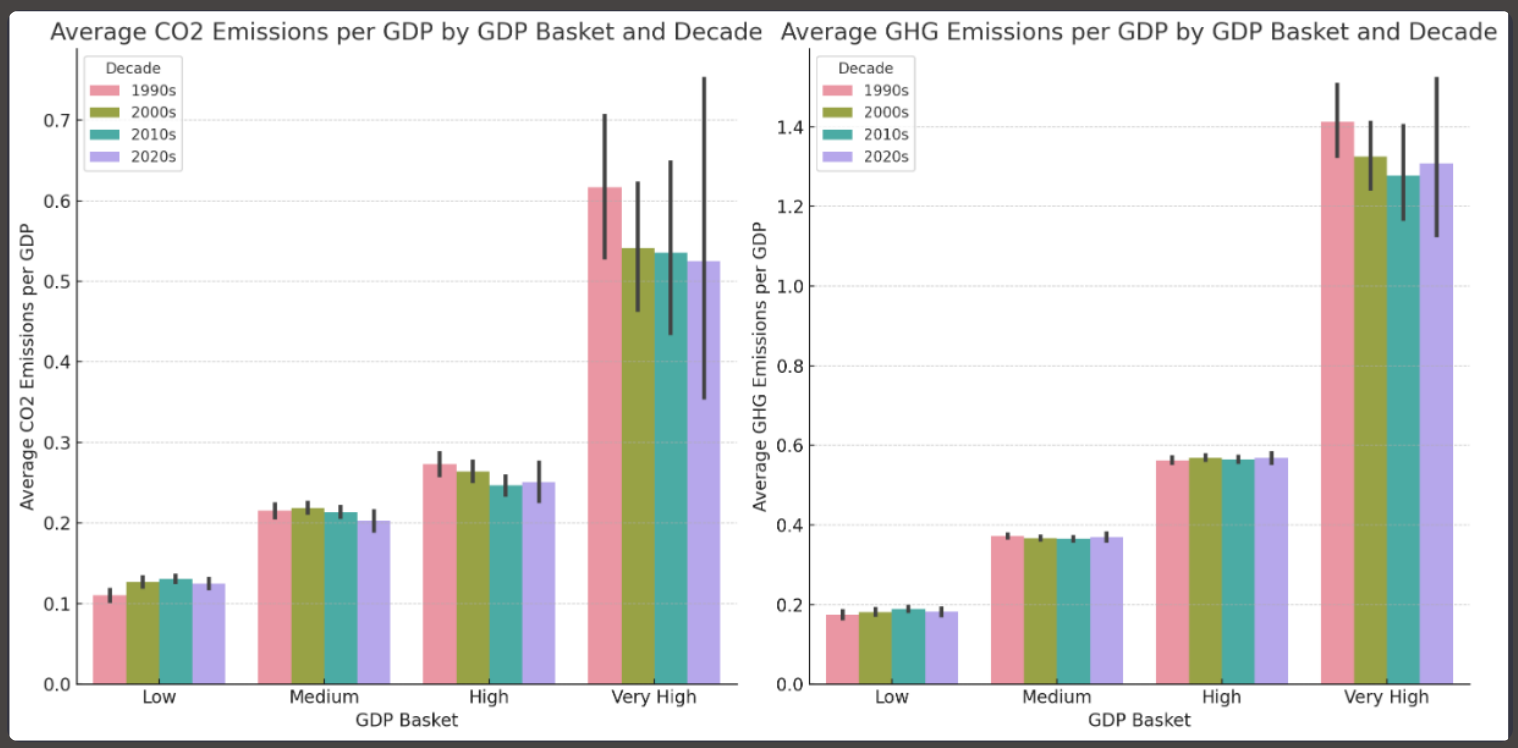


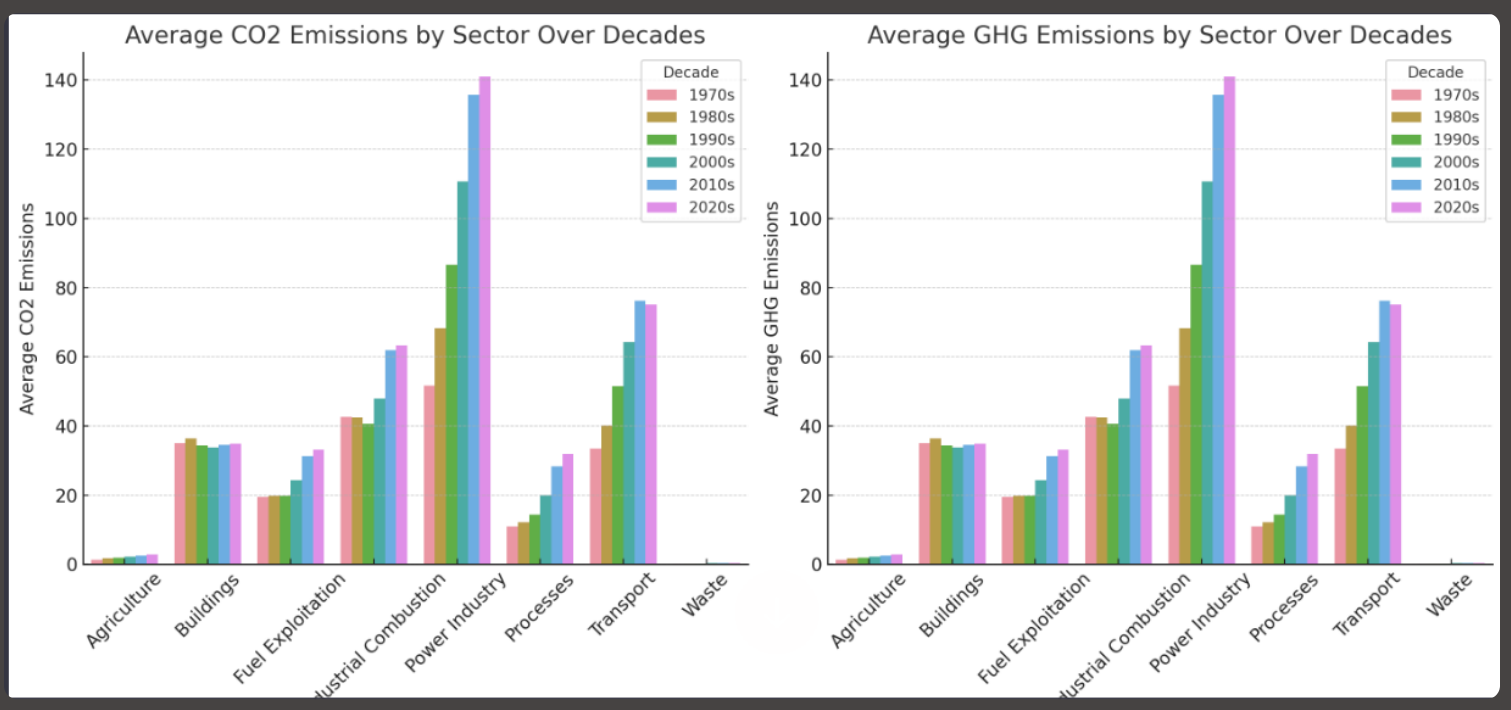


This graph depicts that the highest increase in the emissions for both CO_2 and GHG has been in the case of Asia over the years. Oceania, Europe and North America seems to have been seeing decrease in the emissions in the past couple of decades with the other continents increasing steadily.

- **Population** - This is one of the major factors for continent like Asia to show the major rise in in CO_2 and GHG emissions.
- **Industrialization and Emissions**: There's a correlation between the level of industrialization in a continent and its emissions.
- **Development and Emissions**: Developed continents like Europe and North America have high per capita and total emissions but are showing signs of efficiency improvements.
- **Emerging Economies**: Continents like Asia are showing increasing trends in emissions, correlating with their economic growth and industrialization.
- **Environmental Policies**: The data might reflect the impact of environmental policies and measures in different regions.

Comparison based on GDP and Sector





Emissions by Sector Analysis

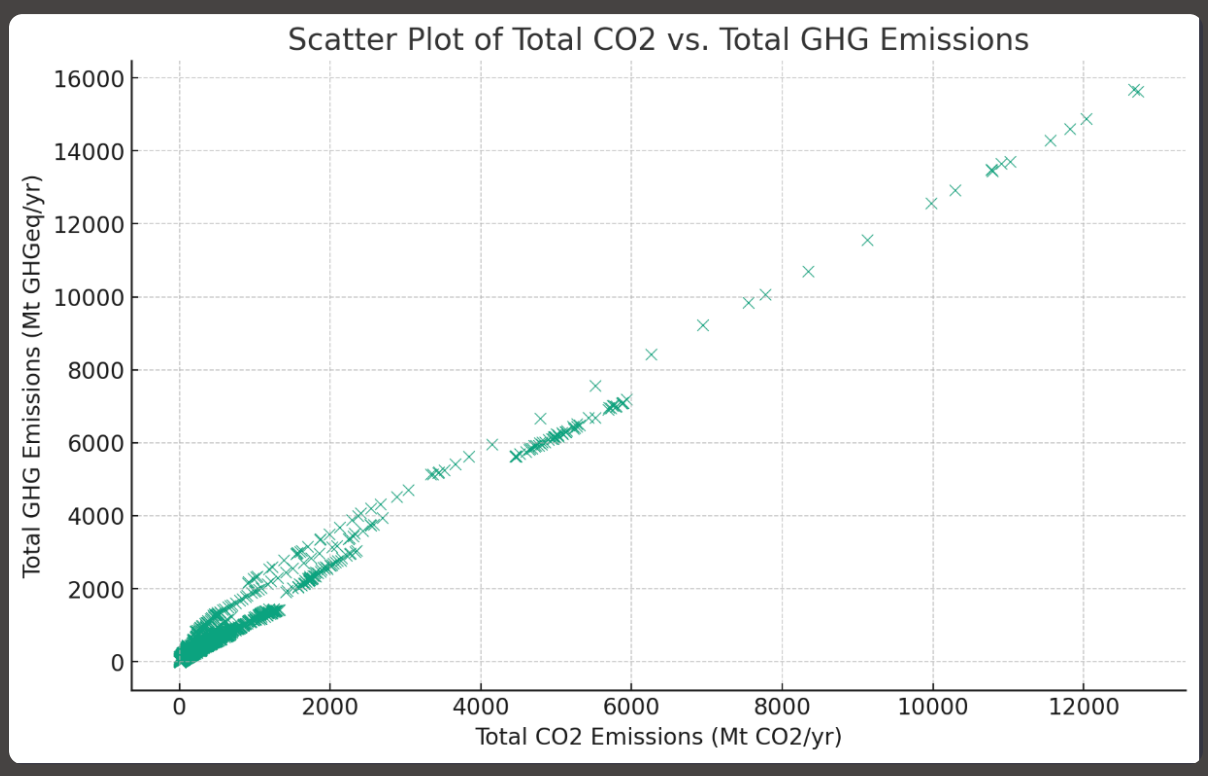
1. **Variability Across Sectors**:
    - Over the decades, the scale of emissions in each sector may have varied, reflecting changes in industrial practices, technological advancements, or shifts in economic focus. However its worth noting that this does not seem to be the case for buildings sector.
2. **Industrial and Energy Sectors**:
    - These sectors often emerge as major contributors to CO2 and GHG emissions. High emissions from these sectors can be attributed to reliance on fossil fuels, high energy demands, and industrial processes.
3. **Transportation Sector**:
    - This sector may show significant emissions, highlighting the impact of vehicular and aviation emissions on the environment. Changes over decades can reflect improvements in fuel efficiency, shifts to cleaner energy sources, or increased transportation demands.
4. **Agriculture and Waste Management**:
    - These sectors may show lower emissions compared to industrial sectors but are still crucial due to methane and other greenhouse gases. Variations over time can indicate changes in agricultural practices and waste management techniques.

Emissions by GDP Basket and Decade

1. **High GDP and Emissions**:
    - There may be a correlation between higher GDP and increased emissions, indicating that economically stronger nations or regions contribute more to emissions. This could be due to larger industrial bases, higher energy consumption, and greater overall economic activity.
2. **Economic Growth and Sustainability**:
    - The analysis can provide insights into how economic growth has been balanced with sustainability efforts. Regions showing a decrease in emissions per GDP over the decades are likely integrating environmental considerations into their economic growth models.

CO2 vs GHG Correlation

By exploring the correlation between country-specific total CO2 emissions and total greenhouse gas (GHG) emissions, we aim to understand the relationship between these two key factors. We identify any outliers in the dataset and explain their presence if observed, shedding light on the broader environmental impact.



Correlation between total CO2 and GHG emissions

- The correlation matrix indicates a very high correlation (approximately 0.994) between total CO2 and total GHG emissions. This strong correlation suggests that as CO2 emissions increase, GHG emissions also increase correspondingly, which is expected given that CO2 is a major component of greenhouse gases.

Scatter Plot Observations

- The scatter plot visually reinforces this strong correlation, displaying a clear linear relationship between CO2 and GHG emissions.
- **Outliers**: There are some points that stand out from the main cluster, indicating countries with disproportionately high CO2 or GHG emissions compared to the global average.

Identification of Outlier Instances

Using Interquartile Range (IQR) method, some of the notable outlier instances are -

Outliers in Total CO2 Emissions

- **United States**: Consistently appearing as an outlier with exceptionally high CO2 emissions (e.g., 4853.78 Mt CO2/yr in 2022).
- **China**: A significant outlier with high emissions, such as 910.08 Mt CO2/yr in 1970, reflecting its large industrial base.
- **Canada, Australia, Czechia**: These countries also appear as outliers in various years, indicating higher CO2 emissions relative to the global average.
- **South Africa and Vietnam (2022)**: Representing developing economies with rising CO2 emissions.

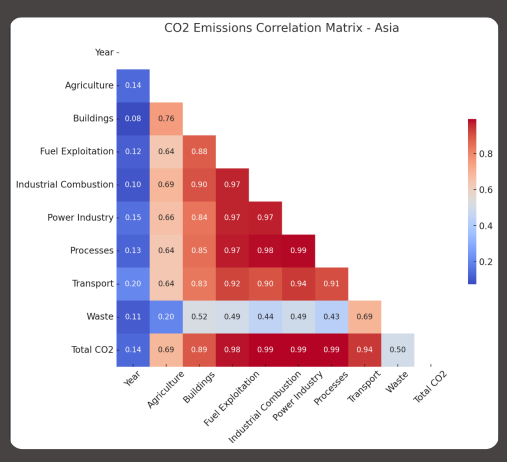
Outliers in Total GHG Emissions

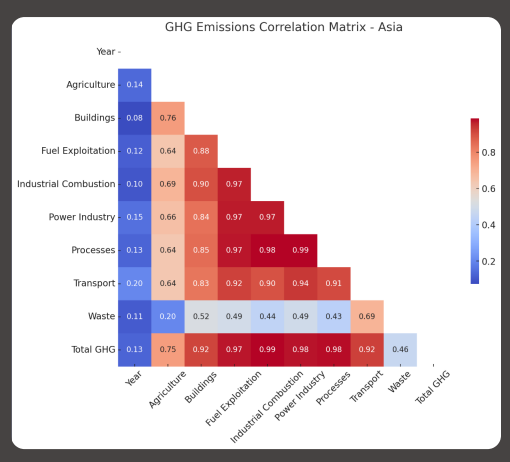
Similar to CO2 emissions, some countries stand out in terms of GHG emissions:

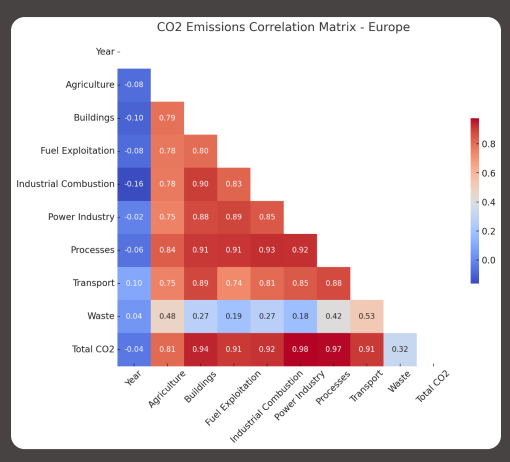
- **United States**: With very high GHG emissions, such as 6017.44 Mt GHGeq/yr in 2022.
- **China**: Showing high GHG emissions, for example, 2142.61 Mt GHGeq/yr in 1970.
- **Argentina, Brazil, Canada (1970)**: These countries also exhibit high GHG emissions, reflecting various factors like industrial activities, energy usage, and land use patterns.
- **Ukraine, Uzbekistan, Vietnam (2022)**: Highlighting growing GHG emissions in these regions.

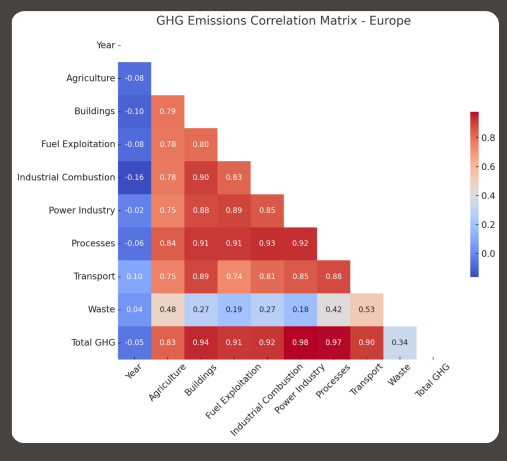
Sector Contribution Correlation

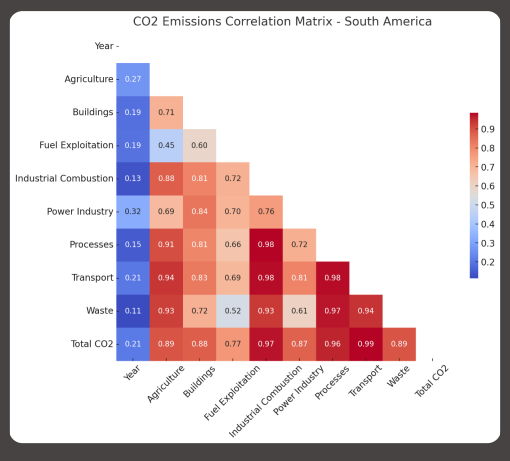
Investigating the correlation between different sectors and aggregate CO2 and GHG emissions, we gain insights into how various industries contribute to emissions. By organizing the results based on continents, we can identify patterns and trends specific to each region, enabling targeted interventions and policies.

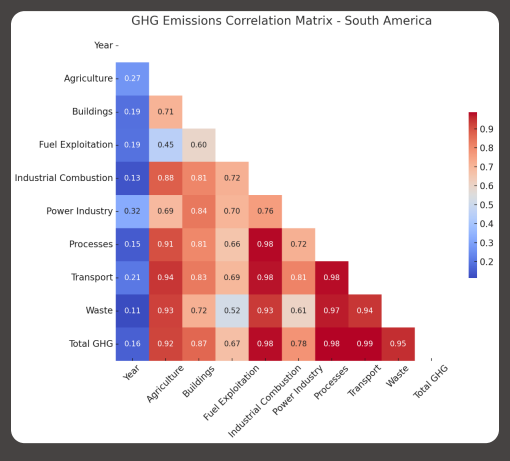


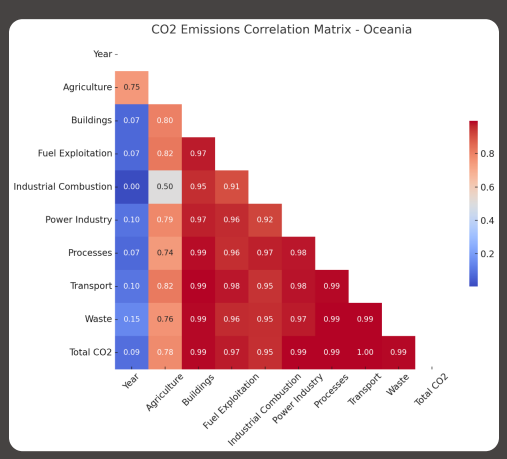


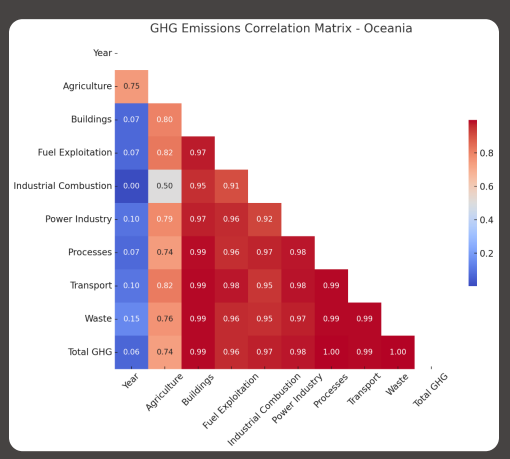


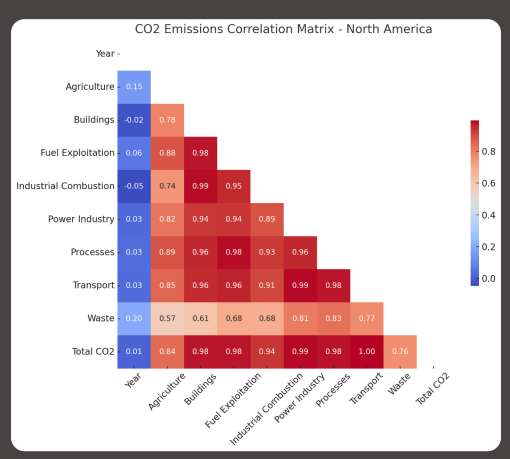


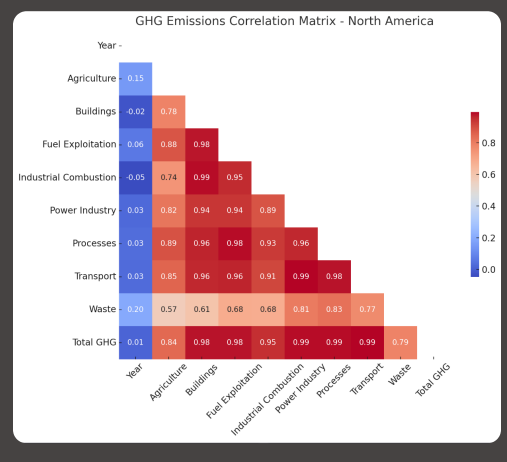


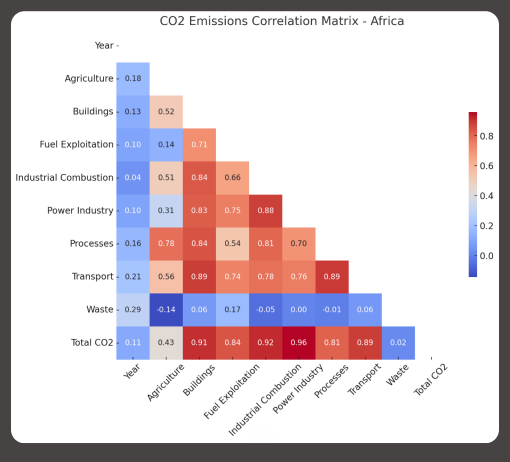


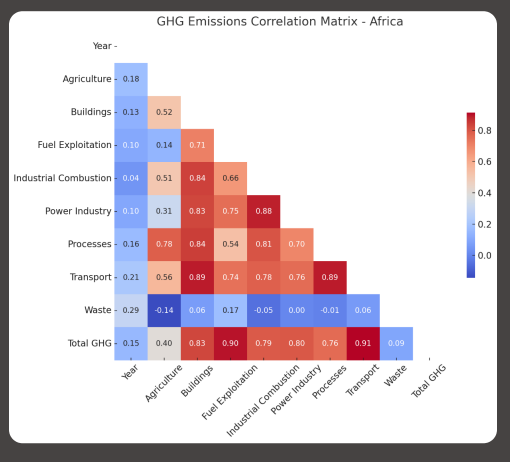












CO2 Emissions Correlation

- **North America**: High correlation in sectors like Transport (0.995) and Power Industry (0.988). Agriculture and Waste have lower correlations.
- **Asia**: Industrial Combustion (0.989) and Power Industry (0.992) show high correlations. Agriculture and Waste have relatively lower correlations.
- **Africa**: Power Industry (0.959) and Industrial Combustion (0.918) are highly correlated. Waste shows a very low correlation.
- **Europe**: Power Industry (0.977) and Processes (0.971) are highly correlated. Waste has a lower correlation.
- **South America**: Transport (0.985) and Processes (0.959) are highly correlated. Fuel Exploitation has a lower correlation.
- **Oceania**: Power Industry (0.994) and Transport (0.996) show very high correlations.

GHG Emissions Correlation

- **North America**: Similar trends as CO2, with Transport (0.994) and Power Industry (0.986) showing high correlations.
- **Asia**: Industrial Combustion (0.988) and Power Industry (0.977) are highly correlated. Waste has a lower correlation.
- **Africa**: Transport (0.913) and Fuel Exploitation (0.899) have relatively high correlations. Waste shows a very low correlation.
- **Europe**: Power Industry (0.977) and Processes (0.971) are highly correlated. Waste has a lower correlation.
- **South America**: Transport (0.988) and Processes (0.985) have high correlations. Fuel Exploitation has a lower correlation.
- **Oceania**: Processes (0.998) and Waste (0.997) show very high correlations.

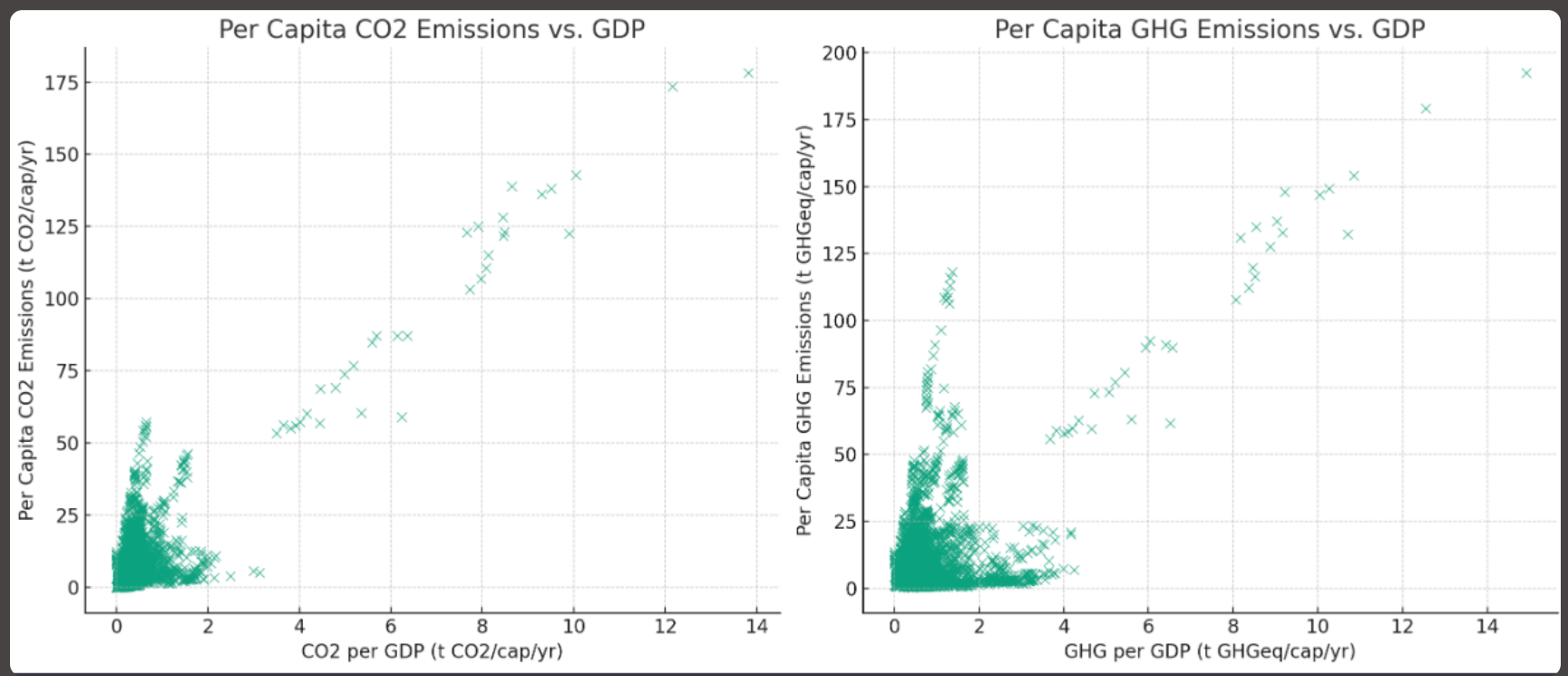
Key Insights

- **Industrial and Power Sectors**: These sectors generally show a strong correlation with total emissions in most continents, reflecting their significant contribution to CO2 and GHG emissions.
- **Variation by Continent**: Each continent shows a unique profile in terms of which sectors are most strongly correlated with emissions, reflecting their economic structure and energy usage patterns.
- **Agriculture and Waste**: These sectors show variable correlation across continents, indicating diverse practices and impacts in different regions.

Note: In emissions data, "Processes" refer to industrial and production activities that release greenhouse gases as byproducts. This includes chemical reactions in manufacturing, mineral processing, oil and gas refining, waste management, and certain agricultural practices. It encompasses a range of sectors where emissions are not directly from fuel combustion but from industrial operations.

Emissions vs GDP Correlation

Examining the correlation between per capita CO2 emissions and GDP, as well as per capita GHG emissions and GDP, we analyze the relationship between emissions and economic development. Through exploring data from both individual countries and continents, we uncover any outliers and provide explanations, contributing to sustainable development strategies.



Emissions vs GDP Correlation Analysis

Correlation Values

- **CO2 Emissions vs GDP**: There's a substantial correlation (approximately 0.79) between per capita CO2 emissions and GDP.
- **GHG Emissions vs GDP**: The correlation between per capita GHG emissions and GDP is lower (approximately 0.50) compared to CO2 emissions.

Insights from Scatter Plots

- The scatter plot for **CO2 vs. GDP** shows a clearer positive trend, indicating that higher GDP per capita tends to correlate with higher CO2 emissions per capita. This suggests that wealthier countries generally have higher CO2 emissions per person.
- The scatter plot for **GHG vs. GDP** shows a less distinct trend, suggesting a more complex relationship between wealth and GHG emissions per capita.

Outliers

- In both plots, there are points that deviate significantly from the general trend, representing countries with either unusually high emissions for their GDP level or high GDP with lower emissions.
- These outliers could be due to a variety of factors:
  - **Economic Structure**: Countries heavily reliant on industries with high emissions (like oil and gas production) might show higher emissions despite a lower GDP.
  - **Energy Sources**: Nations using renewable energy extensively might have lower emissions despite high GDP.
  - **Environmental Policies**: Countries with stringent environmental regulations might achieve lower emissions at higher GDP levels.

Palau - Specific Outlier Instance in Emissions vs. GDP

Per Capita CO2 Emissions vs. GDP Outliers

- **Palau (1994)**: Extremely high per capita CO2 emissions at 178.18 t CO2/cap/yr with a corresponding GDP value of 13.81 t CO2/cap/yr.
- **Palau (2012)**: Another year where Palau stands out with 173.59 t CO2/cap/yr in emissions and 12.15 t CO2/cap/yr in GDP.
- Additional years for **Palau** (1990, 1997, 1991) also show similarly high emission levels.

Per Capita GHG Emissions vs. GDP Outliers

- **Palau (1994)**: Leading in GHG emissions with 192.46 t GHGeq/cap/yr and a GDP value of 14.92 t GHGeq/cap/yr.
- **Palau (2012)**: High emissions recorded at 179.31 t GHGeq/cap/yr with a GDP value of 12.55 t GHGeq/cap/yr.
- Other years for **Palau** (1990, 1991, 1997) show high GHG emissions levels.

Explanation of Outliers

- **Palau's Unique Case**: The recurring appearance of Palau as an outlier in both CO_2 and GHG emissions can be attributed to its small population and unique economic structure. Being a small island nation, any significant industrial or developmental activity can lead to disproportionately high per capita emissions.
- **Economic Activities**: Palau's economic activities, possibly including tourism, construction, and transportation, may contribute to its high emissions relative to its GDP.
- **Small Population Effect**: In countries with small populations, per capita metrics can be significantly affected by relatively small changes in absolute numbers, leading to outlier appearances in such analyses.

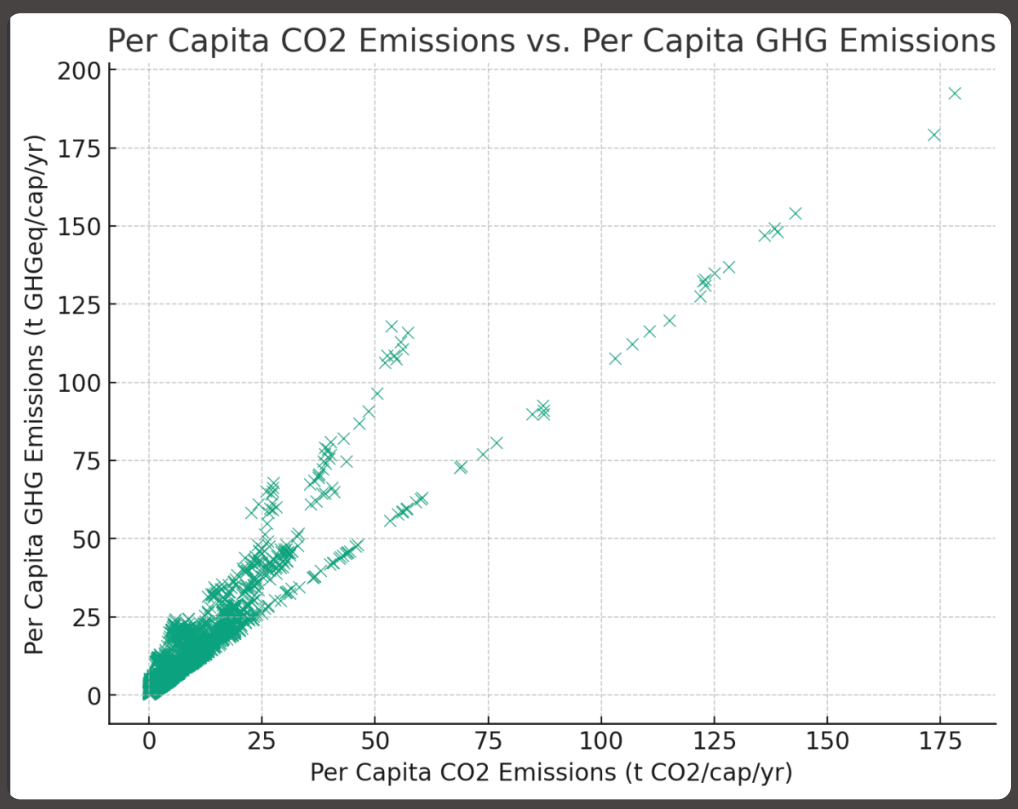
Per Capita CO2 Emissions vs Per Capital GHG Emissions

In addition to the predefined correlations, we have identified another relevant correlation within the datasets. Justified by its significance and potential insights, this correlation sheds light on an important aspect of the data context. By exploring this new relationship, we expand our understanding of the complex interplay between different factors influencing emissions and environmental impact.

Justification for Choice

- **Indicator of Environmental Impact**: Per capita metrics offer a more equitable way to assess and compare the environmental impact of different countries, accounting for population size.
- **Policy and Strategy Implications**: Understanding the relationship between CO2 and overall GHG emissions is crucial for formulating climate policies and strategies. It helps in identifying whether efforts to reduce CO2 emissions are aligned with the broader goal of reducing overall greenhouse gases.

Analysis of Per Capita CO2 Emissions vs. Per Capita GHG Emissions



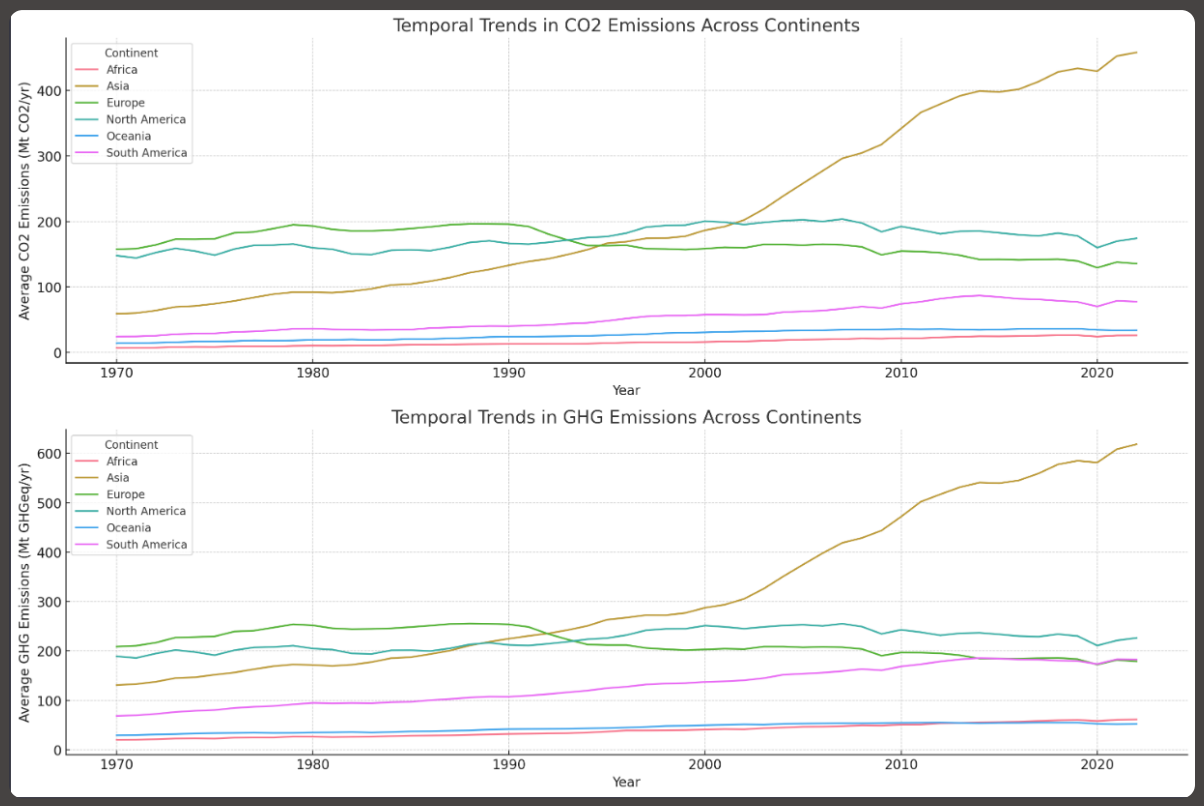
Correlation Results

- The correlation coefficient between per capita CO2 emissions and per capita GHG emissions is approximately 0.95. This indicates a very strong positive correlation between these two metrics.

Significance of This Correlation

- **Indicative of Emission Profiles**: This strong correlation underscores the significant role of CO2 emissions within the broader context of greenhouse gases. It suggests that countries with high per capita CO2 emissions are also likely to have high overall GHG emissions per capita.
- **Policy Implications**: For environmental policies and climate action strategies, this correlation highlights the importance of targeting CO2 emissions as a key area for reducing overall greenhouse gas emissions.
- **Global Climate Efforts**: Understanding this relationship is crucial for international climate agreements and efforts, as it points to CO2 emissions as a primary indicator for assessing and comparing the environmental impact of different nations.

Temporal Analysis



1. **Africa's Increase in CO_2 Emissions (1973 and 1976)**:
    - Africa's rapid increase in CO_2 emissions during the 1970s can be attributed to a combination of population growth and increased per capita energy use. As the continent embarked on a path of industrialization and urbanization, its energy consumption patterns changed significantly. Despite traditionally low CO_2 emissions, this period marked a substantial rise, driven by increased fossil fuel use in a few industrializing countries like South Africa, Algeria, Nigeria, Egypt, and Morocco. This trend of rising emissions continued due to the growing energy demands of a burgeoning population and efforts towards economic development.
2. **South America's Increase in CO_2 Emissions (2010 and 2021)**:
    - For South America, the increase in emissions in 2010 likely reflects the region's economic growth and corresponding energy needs. By 2021, following the initial impact of the COVID-19 pandemic, there was a resurgence in industrial and economic activities, leading to increased CO_2 emissions. The initial drop in emissions during 2020, predominantly due to the global economic slowdown caused by the pandemic, was followed by a rebound as countries sought to recover economically.
3. **Across Continents Decrease in CO_2 Emissions (2020)**:
    - North America's decrease in CO_2 emissions in 2020 is closely linked to the COVID-19 pandemic. The pandemic led to a significant reduction in transportation and industrial activities, especially during the lockdown phases. This temporary decline highlights how large-scale global events, such as pandemics or economic crises, can have a pronounced impact on emissions trends, albeit often temporarily.

Prediction Model

Using a random forest regressor that contains sector wise, gdp wise, total and per capital emissions, we get the following values as the RMSE and the predicted values for the next three years -

RMSE for CO_2 - 210.5933 Mt CO_2/yr

RMSE for GHG - 281.913 Mt GHGeq/yr

Predicted values for CO_2 - 50.598, 152.54, 274.58 CO_2/yr

Predicted values for GHG - 89.86, 207.371, 325.48 Mt GHGeq/yr

Link to the code and data - **https://github.com/encrypted-soul/curve-trading-volume-and-dynamics/tree/main/carbon_emission**