SUSTAINABILITY ANALYTICS PROJECT REPORT

CO2 EMISSION'S REPORT

GROUP 6

1.Gopi Krishna Sajja (GXS220009)

2.Preet Paras Desai (PXD210046)

3.Snehal S Khapte (SXK220022)

4.Tanmayee Ashok Dharam (TXD220006)

5.Venkat Keerthan Boddu (VXB210054)

THE UNIVERSITY OF TEXAS AT DALLAS
Naveen Jindal School of Management

Contents

Introduction

The objective of our project is to examine carbon emissions in different sectors like Waste, Industrial Combustion and Power Industry across multiple countries. The analysis tells us which countries have the highest and lowest Carbon emissions in those sectors and the reasons behind those. To conserve our environment and for sustainable development, it is very crucial to understand those trends and factors which lead to carbon emissions so that necessary measures can be taken to decrease those effects.

Carbon dioxide (CO_2) emissions have emerged as a major global issue, significantly affecting our environmental, economic, and social well-being. As one of the major greenhouse gases contributing to climate change, CO_2 emissions have received increasing attention worldwide. From industrial to transportation and energy production, human activity contributes significantly to CO_2 emissions. Understanding the dynamics of CO_2 emissions is essential to address the challenges posed by climate change and the transition to a more sustainable future in this context, our research examines several aspects of CO_2 emissions in 2010, and examine its sources, trends and implications in different communities and geographies. By unlocking the complexities of CO_2 emissions, we aim to clarify pathways for effective mitigation strategies and drive global cooperation towards a greener planet and it is very resistant.

Data Preparation

1. Data Sourcing

The data has been taken from EDGAR – Emissions Database for Global Atmospheric Research [1].

We've web scrapped the data using libraries like BeautifulSoup. The above URL has been fed into the code and relevant columns have been extracted accordingly. The website is static in nature. The code provided would work for dynamic websites. The code pulls the data every 10 minutes.

```
import requests
from bs4 import BeautifulSoup
import time

def scrape_data():
    # Send a GET request to the URL
    url = "https://edgar.jrc.ec.europa.eu/report_2023?vis=co2tot#emissions_table"
    response = requests.get(url)

    # Parse the HTML content
    soup = BeautifulSoup(response.text, 'html.parser')

    # Find the table containing the data
    table = soup.find('table', {'id': 'emissions_table'})

    # Check if the table was found
    if table:
        # Extract data from the table
        data = []
        for row in table.find_all('tr'):
            cols = row.find_all('td')
            cols = [col.text.strip() for col in cols]
            data.append(cols)

        # Print the extracted data
        for row in data:
            print(row)
    else:
        print("Table not found. Check if the HTML structure has changed.")

# Run the scraping function every 10 minutes
while True:
    scrape_data()
    time.sleep(600)  # 600 seconds = 10 minutes
```

Data set link for Global Temperature.

https://data.world/data-society/global-climate-change-data

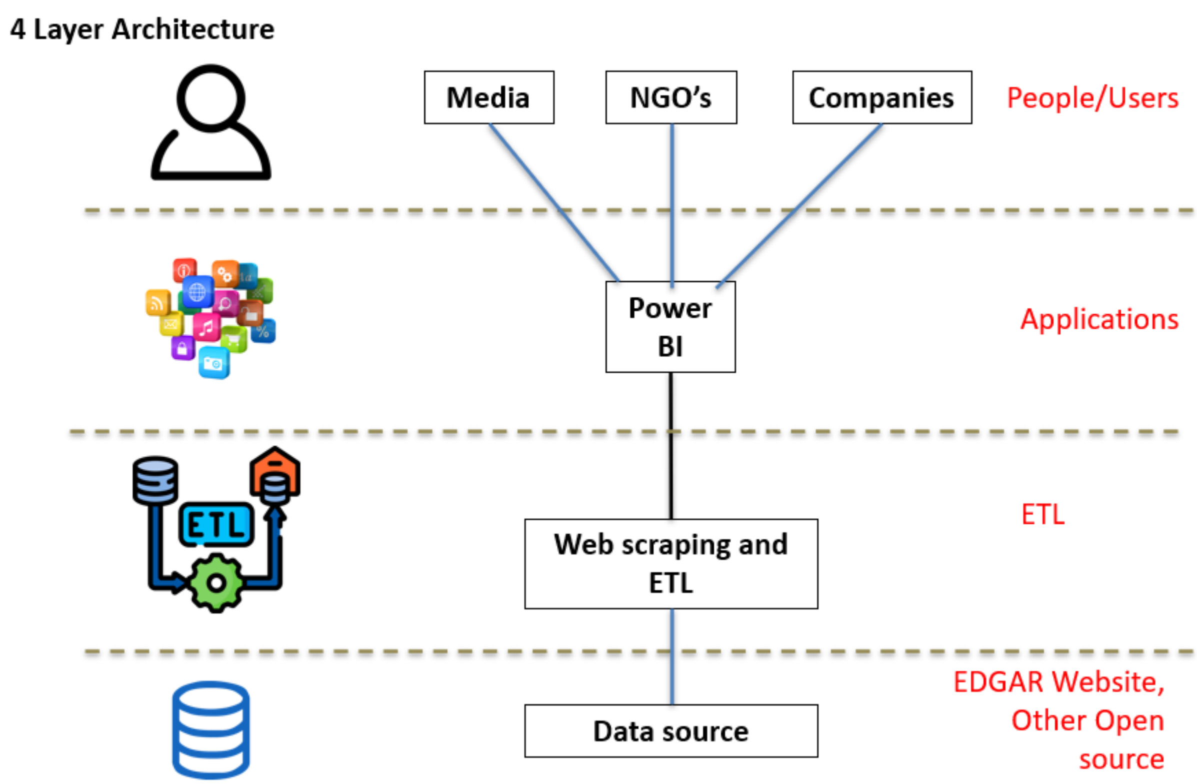
2. System Architecture

We use a four-layer architecture, the first layer is the Data source, where raw data is collected from various sources such as EDGAR databases, APIs, or files.

The second layer is ETL (Extract, Transform, Load), which processes the raw data, extracts context, and transforms it into usable form.

The third level, PowerBI, acts as a visualization and analytics tool, allowed us to create interactive reports and dashboards from processed data.

The fourth layer consists of end users who interact with reports and dashboards created in PowerBI. These users can be analysts, managers, or decision makers who rely on insights from data to make informed decisions and drive business results.



3. Data Cleaning

The dataset was processed by extracting the relevant sheets and filtering the dataset to include rows related to sectors for our analysis i.e. Waste, Industrial combustion and Power Industry. Missing values were handled through imputation with zero values and irrelevant rows were removed.

As we had data from various sources, we used the powerful relational database management system Postgresql to merge them together. Specifically, we used the JOIN function in Postgresql to combine the various datasets based on the common points of country and year. By doing so, we were able to create a more comprehensive and complete dataset that allowed us to perform more in-depth analysis and gain valuable insights.

Data Modelling

1. Data Analysis & Data Modelling

1) Industrial Combustion:

a) To investigate the effects of Industrial Combustion Co2 emissions on various countries, we have sorted the data (Link) in descending order to see the top five countries that are having highest Co2 emissions through Industrial Combustion.

b) The data has been fit into linear regression to forecast the emissions till 2030 for all the countries. Appropriate graphs have been shown in the insights section.

c) The r2 for linear regression turned out to be 0.13 which is quite low. So, we've tried polynomial function to fit the model. We used polyfit in Pandas with 7 d.o.f to fit the data and tried to extrapolate the results. The r2 for the polynomial function turned out to be 0.71, which is quite impressive.

d) Now, we have done similar modelling for Industrial Combustion Co2 emissions data for the world in total and tried to compare the top five countries and world's extrapolation.

2) Waste:

a. To investigate the effects of Carbon Emissions from the Waste sector of various countries, we performed Exploratory Data Analysis and prepared the data for modeling by calculating the Total Emissions of each country.

b. The modeling approach used was the Trend Analysis and visualized Carbon emissions over time for various countries. Out of those countries, we divided them into the highest and lowest carbon emitters in the Waste sector.
c. We investigated the trends and patterns of all the countries in emissions growth or decline and conducted Comparative Analysis among top countries to understand variations in emissions.

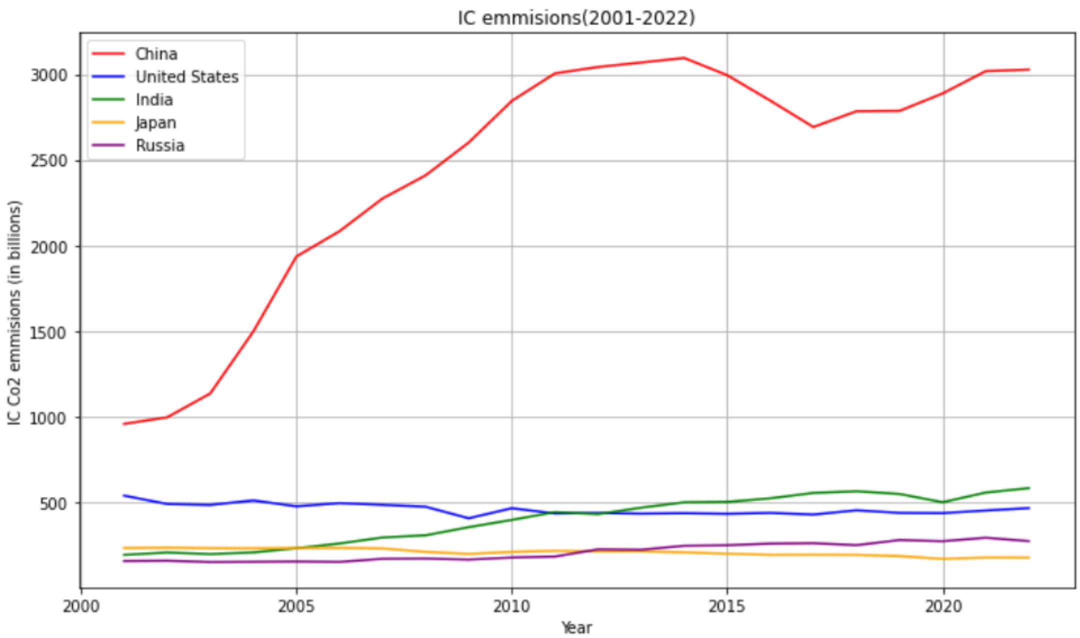
3) Global Temperature due to CO_2 Emissions:

a. CO_2 (Carbon Dioxide) is a greenhouse gas. CO_2 Emissions have a direct impact on our global temperature. These emissions trap heat in our atmosphere, which leads to rising temperatures, the visual evidence and observed impacts like melting of ice, extreme weather change, etc. further confirm this correlation.
b. To Investigate the rising temperature from CO_2 emissions, we performed EAD from the data which had the average temperatures for all the countries around the world from 1970 to 2020.
c. We have created a dashboard to show the average temperatures around the world, where we get a clear picture of the maximum and minimum temperatures, how it has changed over time, and how they are expected to change in the future.

Results / Findings

1. Industrial Combustion Co2 emissions:

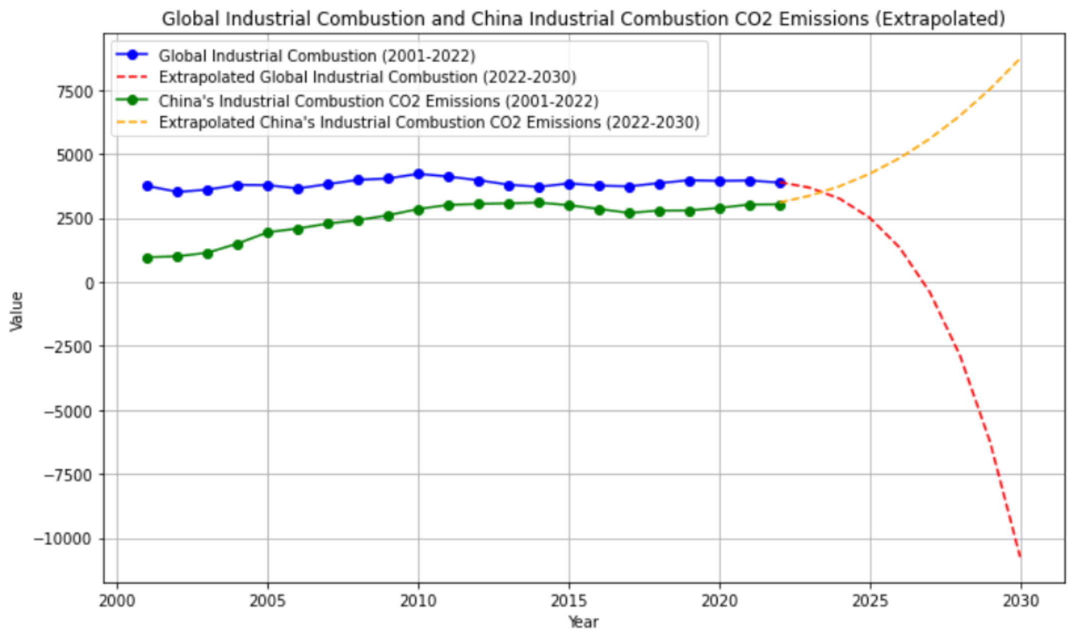
1. Co2 emissions by Industrial Combustion:



```
]: data = {
    'Year': list(range(2001, 2023)),
    'China': [958.838371, 996.753992, 1136.149084, 1497.944862, 1935.745671, 2083.310942, 2274.395752, 2409.719956, 2601.242236,
    'United States': [540.010835, 490.815980, 485.416016, 511.673771, 477.325302, 495.815665, 486.010588, 474.767057, 407.538226,
    'India': [193.979693, 207.670196, 198.424610, 208.255615, 232.618952, 260.307188, 295.258494, 308.178911, 355.584575, 398.015
    'Japan': [233.892141, 236.388654, 233.087055, 232.234308, 234.092679, 234.626958, 231.448840, 211.458056, 199.698312, 211.358
    'Russia': [157.877256, 159.651287, 151.782739, 153.212061, 154.876350, 152.633941, 170.987747, 172.377420, 166.226982, 178.42
}

df = pd.DataFrame(data)

# Plot line graphs for each country's GDP
plt.figure(figsize=(10, 6))

# Plot and set colors for each country
plt.plot(df['Year'], df['China'], label='China', color='red')
plt.plot(df['Year'], df['United States'], label='United States', color='blue')
plt.plot(df['Year'], df['India'], label='India', color='green')
plt.plot(df['Year'], df['Japan'], label='Japan', color='orange')
plt.plot(df['Year'], df['Russia'], label='Russia', color='purple')

# Add title and labels
plt.title('IC emmisions(2001-2022)')
plt.xlabel('Year')
plt.ylabel('IC Co2 emmisions (in billions)')
plt.legend()

# Show plot
plt.grid(True)
plt.tight_layout()
plt.show()
```

The above graph displays the top five countries that have the highest Co2 emissions through Industrial Combustion. It is evident that majority of the emissions are from China, US, Russia, India and Japan. Out of which China is the major contributor with almost 3000 Mt of Co2 released in 2022.

2. Extrapolated data for world emissions and China emissions:




```
# Extrapolate global industrial combustion data for the years 2022 to 2030
extrapolated_global_ic_poly_degree_7 = global_poly_func_degree_7(np.arange(2022, 2031))

# Extrapolate China's industrial combustion CO2 emissions data for the years 2022 to 2030
extrapolated_china_ic_poly_degree_7 = china_poly_func_degree_7(np.arange(2022, 2031))

# Plot the extrapolated data on the same graph
plt.figure(figsize=(10, 6))

# Plot Global Industrial Combustion
plt.plot(years_ic, industrial_combustion, label='Global Industrial Combustion (2001-2022)', marker='o', color='blue')
plt.plot(np.arange(2022, 2031), extrapolated_global_ic_poly_degree_7, label='Extrapolated Global Industrial Combustion (2022-203

# Plot China's Industrial Combustion CO2 Emissions
plt.plot(years_china, china_industrial_combustion_co2_array, label="China's Industrial Combustion CO2 Emissions (2001-2022)", mar
plt.plot(np.arange(2022, 2031), extrapolated_china_ic_poly_degree_7, label="Extrapolated China's Industrial Combustion CO2 Emissi

# Add title and labels
plt.title('Global Industrial Combustion and China Industrial Combustion CO2 Emissions (Extrapolated)')
plt.xlabel('Year')
plt.ylabel('Value')
plt.legend()
plt.grid(True)

# Show plot
plt.tight_layout()
plt.show()
```

The above graph shows the Co2 emissions in Mt for the world in total (Blue line) and for China (green line) and their extrapolations till 2030 using polyfit with 7 d.o.f. We can clearly see that almost 40% of the world's Co2 emissions through Industrial Combustion are from China.

2. Waste Co2 emissions:

The results of the analysis included the top countries with the highest and lowest carbon emissions, the trends observed for specific countries and insights like growth rates from the Comparative analysis.

```
In [6]: # Filtering the dataset to include only rows related to waste
        df_waste = df[df['Sector'] == 'Waste']

In [7]: # Handle missing values by imputation
        df_waste.fillna(0, inplace=True)

        /var/folders/2n/6xs598_96n9b7pqtqqdg_fyh0000gn/T/ipykernel_36424/2996640884.py:3: SettingWithCopyWarning:
        A value is trying to be set on a copy of a slice from a DataFrame

        See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returni
        ng-a-view-versus-a-copy
          df_waste.fillna(0, inplace=True)

In [8]: print(df_waste['Country'].unique())

        ['Albania' 'Argentina' 'Australia' 'Austria' 'Azerbaijan' 'Belgium'
         'Benin' 'Bulgaria' 'Bahamas' 'Bosnia and Herzegovina' 'Belarus' 'Bolivia'
         'Brazil' 'Brunei' 'Canada' 'Switzerland and Liechtenstein' 'Chile'
         'China' 'Congo' 'Colombia' 'Cuba' 'Czechia' 'Germany' 'Dominica'
         'Denmark' 'Dominican Republic' 'Algeria' 'Ecuador' 'Egypt'
         'Spain and Andorra' 'Estonia' 'Ethiopia' 'Falkland Islands'
         'France and Monaco' 'Faroes' 'Gabon' 'United Kingdom' 'Gibraltar'
         'Greece' 'Greenland' 'French Guiana' 'Guyana' 'Hong Kong' 'Honduras'
         'Croatia' 'Hungary' 'Indonesia' 'Ireland' 'Iceland'
         'Italy, San Marino and the Holy See' 'Jamaica' 'Jordan' 'Japan' 'Kenya'
         'Cambodia' 'South Korea' 'Kuwait' 'Laos' 'Lebanon' 'Saint Lucia'
         'Lithuania' 'Macao' 'Moldova' 'Maldives' 'Mexico' 'North Macedonia'
         'Mali' 'Malta' 'Myanmar/Burma' 'Mongolia' 'Mauritius' 'Malaysia'
         'Netherlands' 'Norway' 'New Zealand' 'Peru' 'Philippines' 'Palau'
         'Papua New Guinea' 'Poland' 'North Korea' 'Portugal' 'Paraguay' 'Qatar'
         'Romania' 'Serbia and Montenegro' 'Singapore' 'Saint Pierre and Miquelon'
```

Time Trend Analysis

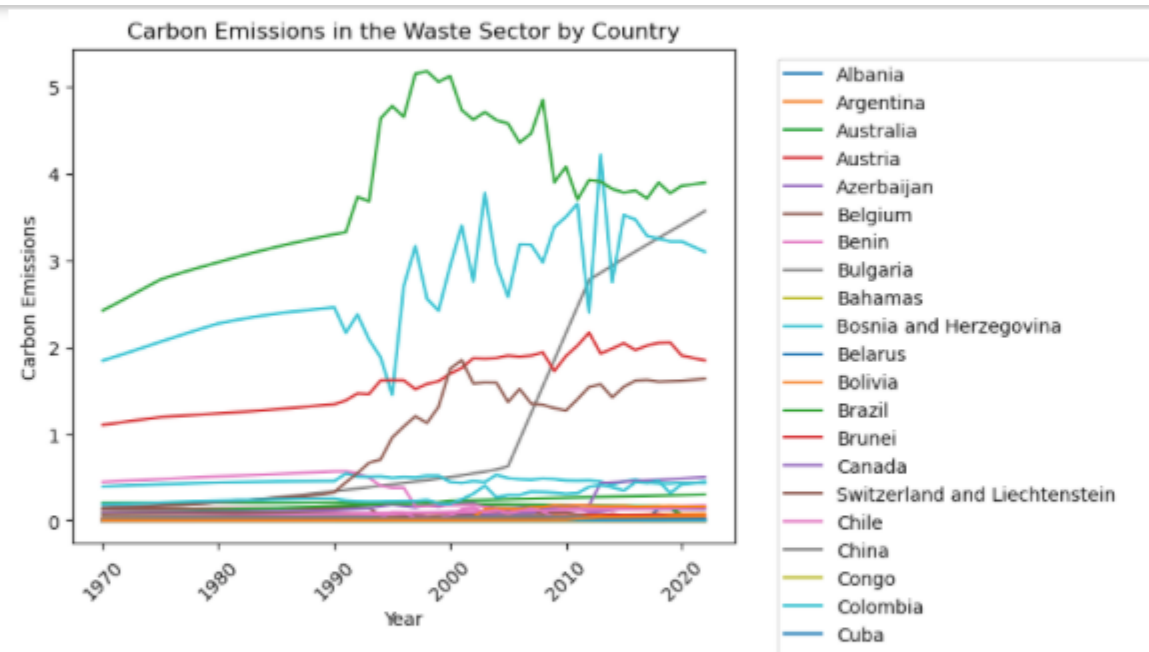
```
In [12]: import matplotlib.pyplot as plt

In [17]: # Select columns for plotting (years)
         years = df_waste.columns[4:]

In [18]: # Plot carbon emissions for each country
         for index, row in df_waste.iterrows():
             country = row['Country']
             emissions = row[4:]
             plt.plot(years, emissions, label=country)

         plt.xlabel('Year')
         plt.ylabel('Carbon Emissions')
         plt.title('Carbon Emissions in the Waste Sector by Country')
         plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
         plt.xticks(rotation=45)
         plt.tight_layout()
         plt.show()
```

In [20]:
```
# Calculate total carbon emissions for each country
df_waste['Total Emissions'] = df_waste.iloc[:, 4:].sum(axis=1)
```

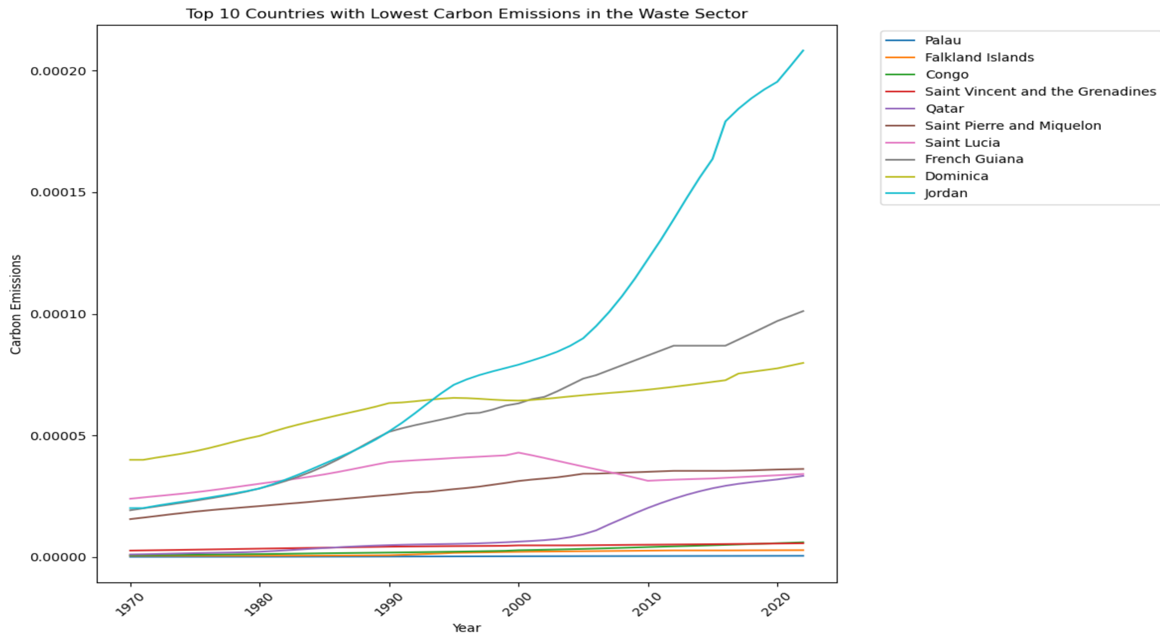
In [25]:
```
# Select top 10 countries with the highest total emissions
top_10_countries = df_waste.nlargest(10, 'Total Emissions')

years = top_10_countries.columns[4:-1]
```

In [32]:
```
# Plot carbon emissions for each country
plt.figure(figsize=(12, 8))
for index, row in top_10_countries.iterrows():
    country = row['Country']
    emissions = row[4:-1]
    plt.plot(years, emissions, label=country)

plt.xlabel('Year')
plt.ylabel('Carbon Emissions')
plt.title('Top 10 Countries with Highest Carbon Emissions in the Waste Sector')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

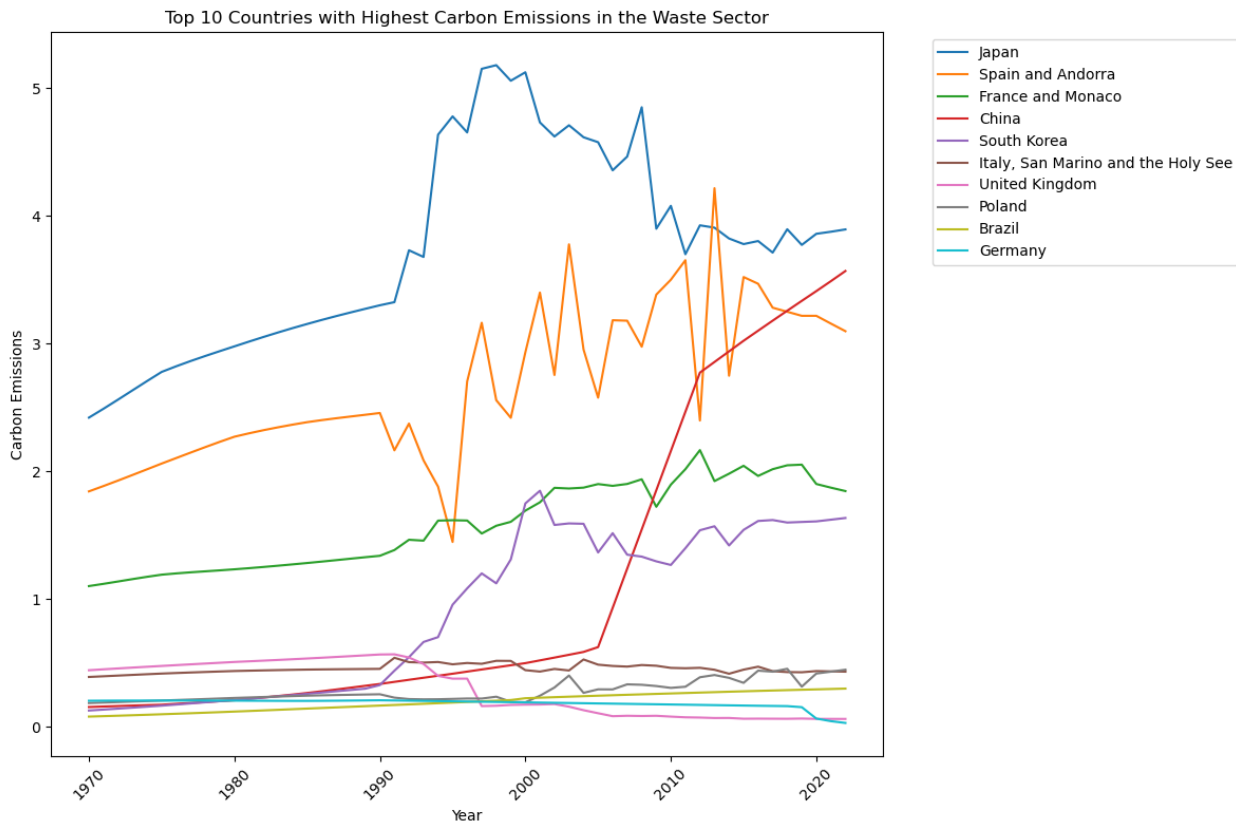
1. Co2 emissions by Waste



The above graph shows the carbon emissions for all the countries in the world.



The above graph shows the lowest carbon emitting countries. Evidently, trends in emissions varied among top and bottom countries.



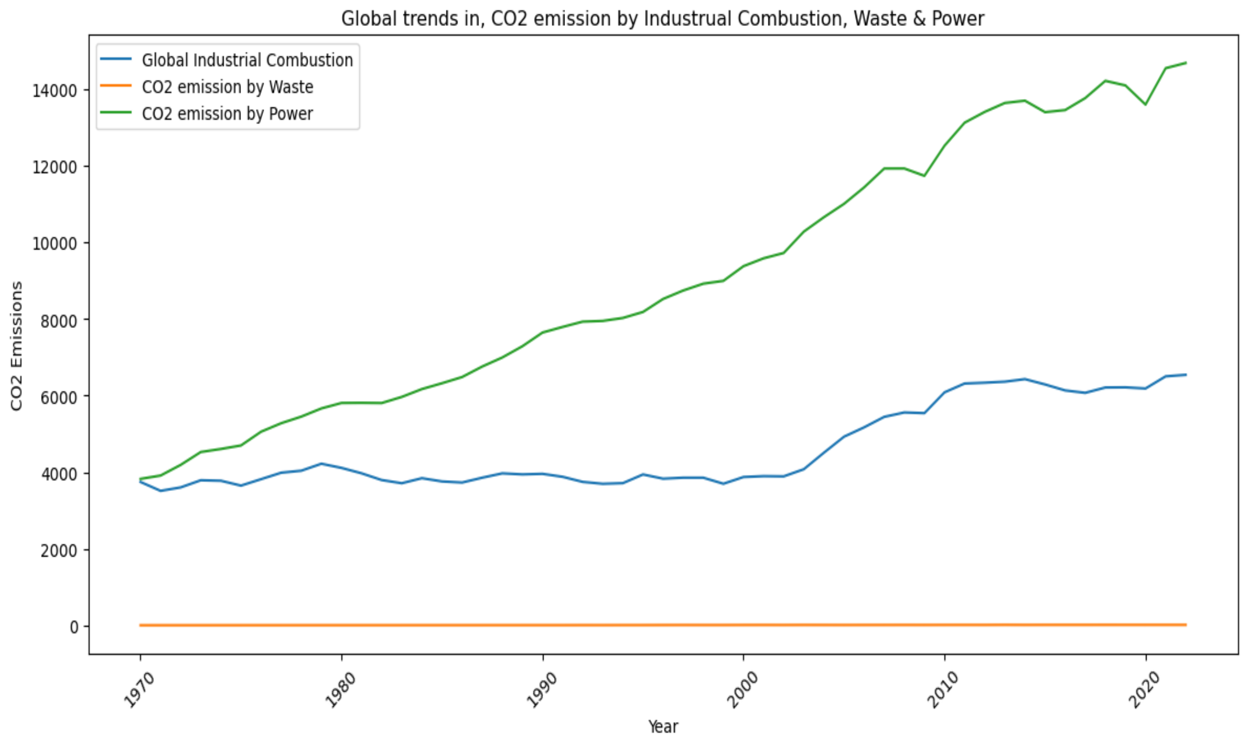
The above graph shows the carbon emissions for the top 10 countries. Here, we can see that Japan emerges as the highest emitter in the waste sector, followed by Spain and France. China also exhibited significant growth in emissions post-2005. Whereas the UK has shown a significant decrease in this sector. Despite Japan's reputation for efficient waste management practices, there are several factors which contributes to being a significant emitter of carbon emissions as below:

1. **Reliance on Incineration**: Japan relies heavily on waste incineration as a method of waste disposal due to limited landfill space. On one hand, it is effective in reducing waste volume. However, it generates a significant amount of carbon emissions if not managed properly.
1. **Energy Intensity:** Japan's industrial and manufacturing sectors are energy-intensive, leading to higher emissions associated with waste processing, production, and transportation.

2. **Limited Recycling Rates:** Despite Japan's reputation for advanced recycling practices, certain waste streams, such as plastic and electronic waste, pose challenges in recycling and often end up in incineration facilities, contributing to emissions.
3. **Population Density:** Japan's high population density necessitates efficient waste management solutions, but this also leads to greater overall waste generation and associated emissions.
4. **Technological Complexity:** Japan's advanced waste management infrastructure, including state-of-the-art incineration facilities, requires substantial energy inputs and maintenance, contributing to carbon emissions.

3. Trend Analysis:

Analyzing the Impact of Industrial Combustion, Waste Management, and Power Generation on Carbon Emissions:



Since 1970, there has been a consistent upward trend in CO2 emissions from all three sectors - industrial combustion, waste, and power. Among the three sectors, power generation has been identified as the highest contributor to the rising CO2 emissions. This trend is concerning, as it indicates the harmful impact of human activities on the environment. It is important for individuals and industries to take steps towards reducing their carbon footprint and promoting sustainable practices to mitigate the effects of climate change.CO2 emissions from all three sectors (industrial combustion, waste, and power) have been increasing since 1970, with power generation being the highest contributor.

Predictions on dataset: We first examined the correlations before delving into predictive analytics. The results indicated a strong positive correlation between CO2 emissions and global power, CO2 emissions and global industrial waste, and global power and global industrial waste.

To conduct further analysis, we used global data of CO2 emissions resulting from these factors over 5 years. The aim of this model is to make accurate CO2 emission predictions based on a few variables. After standardizing the data, we randomly split it into training and test sets. We decided to build both Linear Regression and random forest models to analyze and predict CO2 values based on factors.

```
In [ ]: # Separate features (X) and target variable (y)
        X = data[['Global_Industrial', 'Global_waste', 'Global_Power']]
        y = data['CO2']

In [ ]: scaler = StandardScaler()
        X_scaled = scaler.fit_transform(X)
        y_scaled = scaler.fit_transform(y.values.reshape(-1, 1))

In [ ]: X_train_scaled, X_test_scaled, y_train_scaled, y_test_scaled = train_test_split(X_scaled, y_scaled, test_size=0.2, random_state=42)
```

Linear Regression

```
In [ ]: model_scaled = LinearRegression()
        model_scaled.fit(X_train_scaled, y_train_scaled)

In [ ]: y_pred_scaled = model_scaled.predict(X_test_scaled)

In [ ]: mse_scaled = mean_squared_error(y_test_scaled, y_pred_scaled)
        print("Mean Squared Error (scaled):", mse_scaled)
        # Print coefficients of scaled model
        print("Coefficients (scaled):", model_scaled.coef_)
        r2_score(y_test_scaled, y_pred_scaled)
```

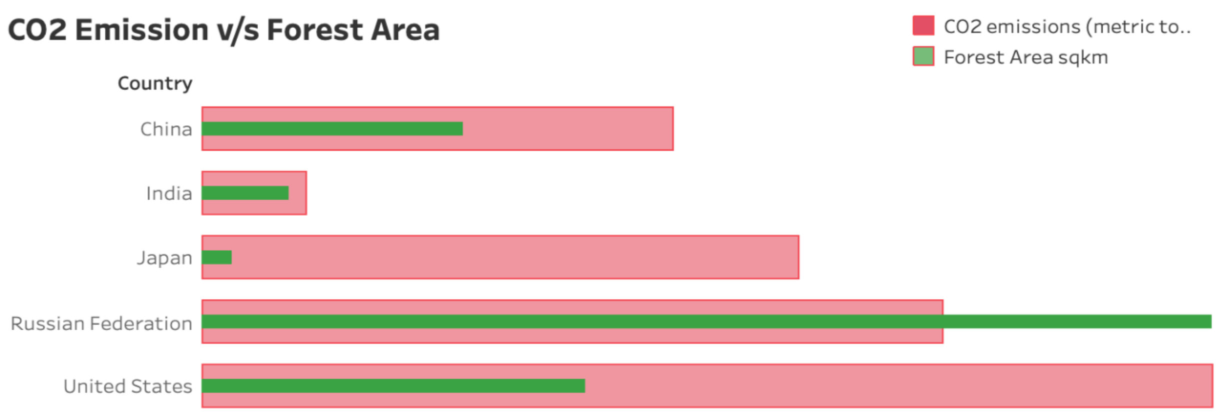
K fold validation

```
In [ ]: kf = KFold(n_splits=5, shuffle=True, random_state=42)

In [ ]: cv_scores = cross_val_score(model_scaled, X_train_scaled, y_train_scaled, cv=kf, scoring='r2')

In [ ]: # Evaluate results
        print("Cross-Validation Scores:", cv_scores)
        print("Mean R-squared:", cv_scores.mean())
        print("Standard Deviation of R-squared:", cv_scores.std())
```

Random Forest

```
In [ ]: rf_model = RandomForestRegressor(random_state=42)
        rf_model.fit(X_train_scaled, y_train_scaled)
        y_pred_rf = rf_model.predict(X_test_scaled)
```

After performing cross-validation, we found that both the linear regression and random forest models are effective in predicting CO2 levels. We also decided to focus our attention on countries with high average CO2 emissions over the past decade. We checked the total land occupied by forest areas in each country and discovered that those with more forest land area tend to have comparatively lower CO2 emissions (though not fully controlled).



From our analysis, we can conclude that increasing forest land can help control CO2 emissions. This application could be useful for policymakers looking to reduce CO2 emissions in their respective countries.

4. Global Temperature due to CO2 Emissions:

1. Dashboard for Global Temperatures

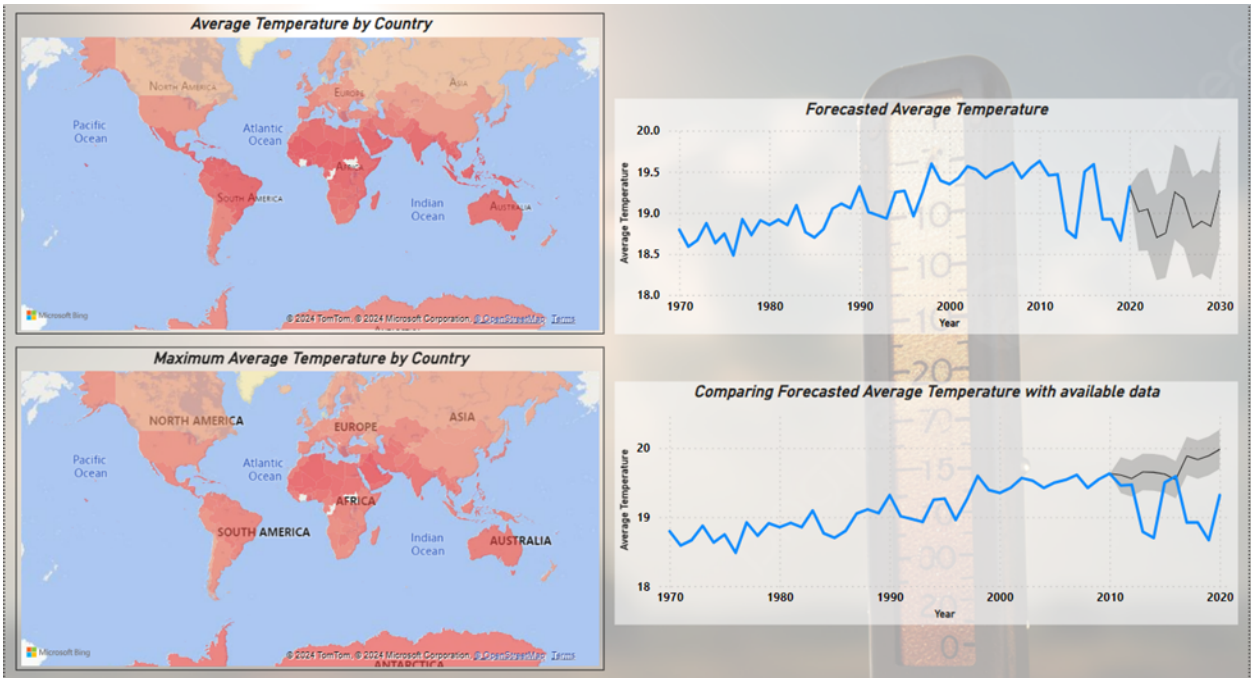


Fig 1: Dashboard for Global Temperatures

The above dashboard displays information about the average temperature across the world. We tried to focus on 3 main topics, they are as follows:

a. Average temperature by Country: We have used a world map to show the average temperature and Maximum average temperature for each country. There is a legend that indicates the average temperature range, cold countries with light red color and hot countries with Red.

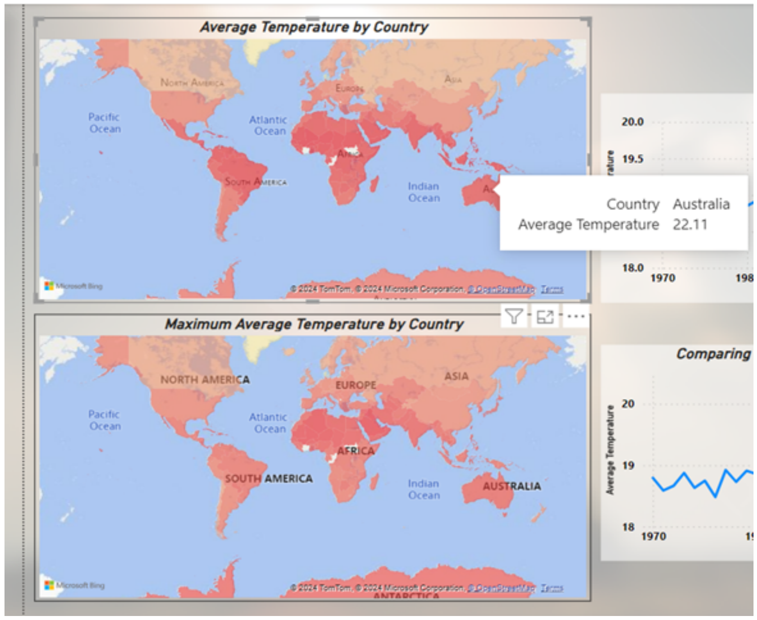


Fig 2: Average temperature for Australia

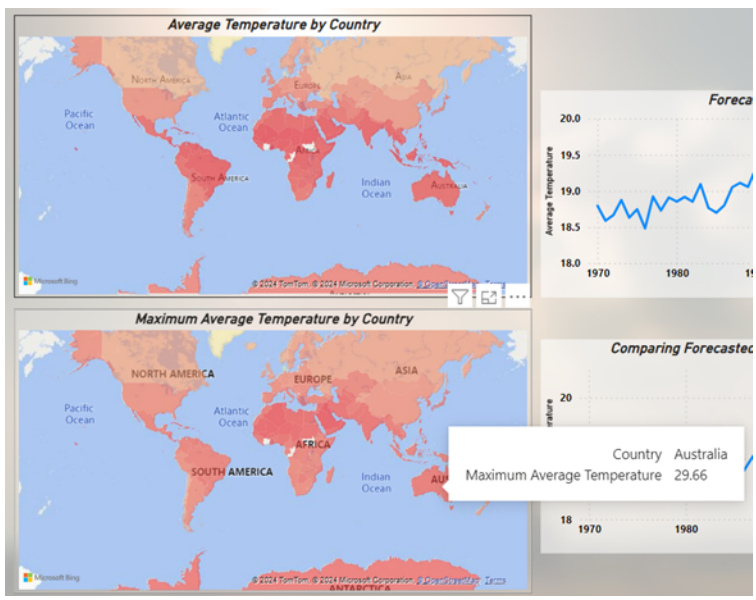


Fig 3: Maximum Average Temperature for Australia

b. Forecasted Average Temperature: We have used a line graph to show the forecasted average temperature over time for each country. It will predict the forecasted average temperature for that country. We can see the below example for Australia, it predicts the

Forecasted Temperature i.e. 21.19 Degrees Celsius. Here, Upper Bound represents the highest temperature and Lower Bound represents the lowest temperature for that country.

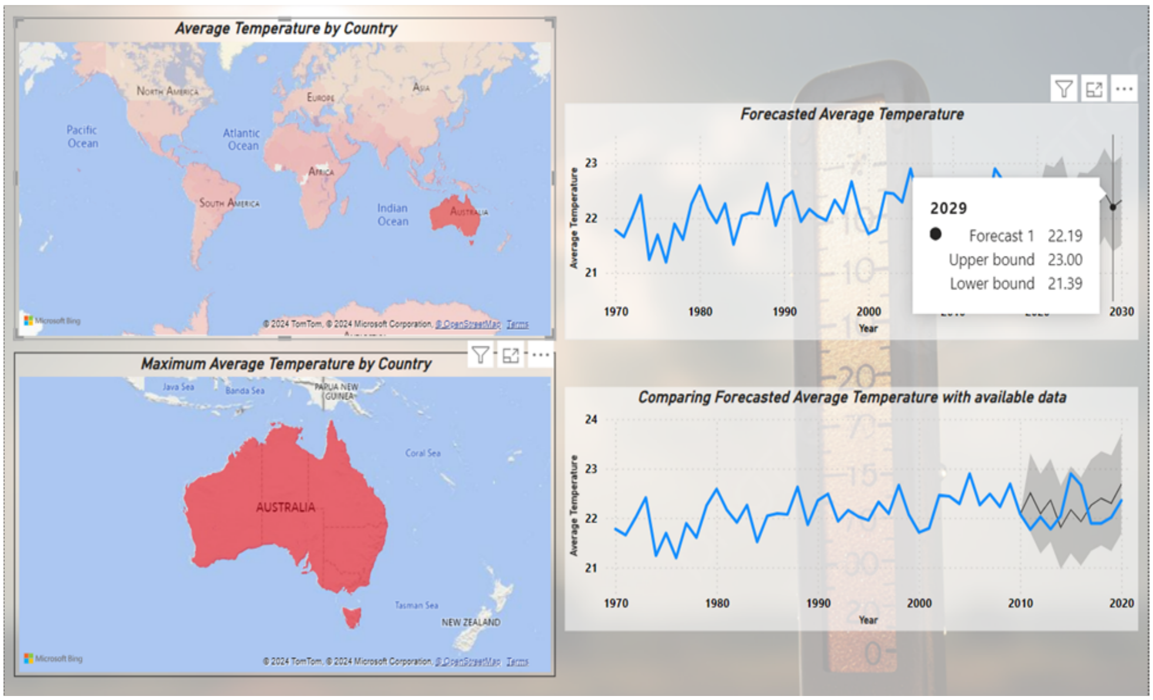


Fig 4: Forecasted Average Temperature

c. Comparing Forecasted Average Temperature with available data: Here we focused on comparing the previous 10 years of average temperature data to forecasted data to see the trend. The blue line shows the previous 10 years' data, whereas the black line shows the forecasted values. From Fig 1 We can see that the forecasted temperature line starts above the available data line and increases steadily over time. Data lines show variations over time, but also show a general increase in temperatures.

Conclusion

The Co2 emissions from Industrial Combustion have increased over the years, by 2022 around 3500 Mt of Co2 is being emitted from Industries worldwide. Out of the 3500 Mt around 2500 Mt is from China. This is due to variety of factors such as the Industrial boom in China which is attracting Industries from other countries to migrate to China. This is an alarming situation, and measures must be taken to reduce these emissions from China. The extrapolated results show that although the emissions worldwide are expected to decrease eventually but China's emissions tend to increase. This suggests that focusing on China alone and taking measures that reduce these emissions would bring down the Co2 emissions drastically worldwide.

The Co2 emissions from the Waste sector have decreased over the years from 114.9 MMT CO2e in 2011 to 103.3 MMT CO2e in 2021 which is approximately 10% reduction. Out of those, emissions form the waste sector in Japan amounted to 20.37 million t-CO2 eq. Even though it is known for its best techniques for waste disposal and incineration, Japan is among one of the highest carbon emitters in the Waste sector. Japan has revised its policy aiming to be Carbon Neutral by 2050. Understanding Japan's waste-related emissions underscores the importance of integrated waste management strategies that prioritize emissions reduction and waste reduction and resource recovery.

The analysis of global temperature from Co2 emissions confirms that there is a warning trend across the globe, with extreme rising temperatures due to emissions.

Recommendations

To reduce the Co2 emissions from Industrial Combustion from China, there are few methods:

- Transitioning to renewable energy sources like biomass to run industries. As of now coal is the major source to power the industries, which is also a great contributor to Co2 emissions.
- Establishing Green Financing (Loans and Grants to Industries that promote environmental-friendly activities) will encourage Industries to tune their processes and implement methods that reduce their overall Co2 emissions.

To reduce the Co2 emissions from Waste from Japan, there are few methods:

- Promoting waste reduction and recycling practices to minimize the amount of waste sent for incineration and to implement policies that prioritize waste reduction at the source and promote circular economy approach.
- Technologies such as carbon capture and storage (CCS) to capture CO2 emissions generated during nuclear power plant operations or wastewater treatment processes should be explored.

To keep global temperatures in control, there are 2 recommendations:

- To reduce the effect of rising temperatures and global warming, we should implement policies and practices, individual actions that reduce C2 emissions.
- Monitoring global temperatures and developing or getting used to some strategies which address the consequences of rising temperatures are essential.

References

[1] EDGAR (Emissions Database for Global Atmospheric Research) Community GHG Database, a collaboration between the European Commission, Joint Research Centre (JRC), the International Energy Agency (IEA), and comprising IEA-EDGAR CO2, EDGAR CH4, EDGAR N2O, EDGAR F-GASES version 8.0, (2023) European Commission, JRC (Datasets).

Data Source Link: https://data.world/data-society/global-climate-change-data