CO2 Emissions Analysis and Prediction

1. Introduction

CO2 emissions are a critical driver of global climate change, with far-reaching consequences for ecosystems, economies, and societies. This report analyzes historical trends in CO2 emissions, identifies patterns through clustering, and predicts future emissions using linear regression. The data-driven insights aim to support decision-making and drive sustainable development policies.

The report is structured as follows:

- An overview of the dataset and exploratory data analysis.
- Visualizations to understand relational and statistical trends.
- Clustering analysis to segment emissions profiles.
- Regression analysis for emissions prediction and future projections.

---

2. Dataset Overview

The dataset encompasses CO2 emissions data collected over decades from countries worldwide. Key attributes include:

- **Country Name**: Identifies individual countries and regions.
- **Year**: The year corresponding to the emissions data.
- **CO2 Emissions (kt)**: Annual CO2 emissions in kilotons.
- **Normalized Emissions**: Rescaled emissions values for analytical purposes.

The dataset allows us to examine global trends, segment nations based on emissions and predict future emissions scenarios. A detailed summary of the data preprocessing and normalization steps ensures the robustness of the analysis.

---

3. Exploratory Data Analysis

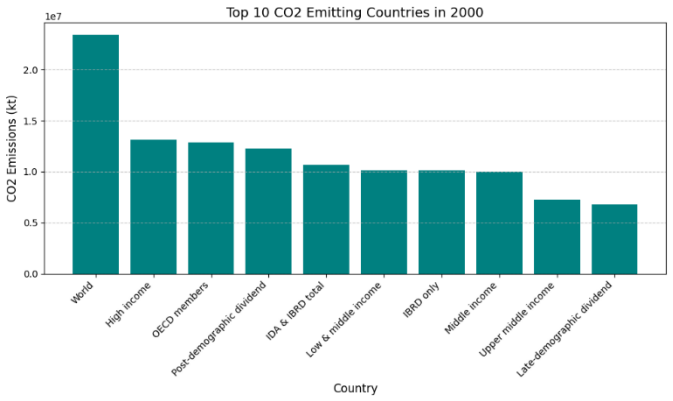

Figure 1 Top 10 CO2 Emitting Countries in 2000

Description: This bar chart highlights the top 10 CO2-emitting regions and countries in the year 2000.

- **Key Observations**:
  - The "World" category reports the cumulative global emissions, which exceed 20 million kt.
  - High-income countries, OECD members, and post-demographic dividend regions exhibit the highest emissions.
  - These findings highlight the role of economic development and industrial activity in emissions trends

---

4. Relational Analysis

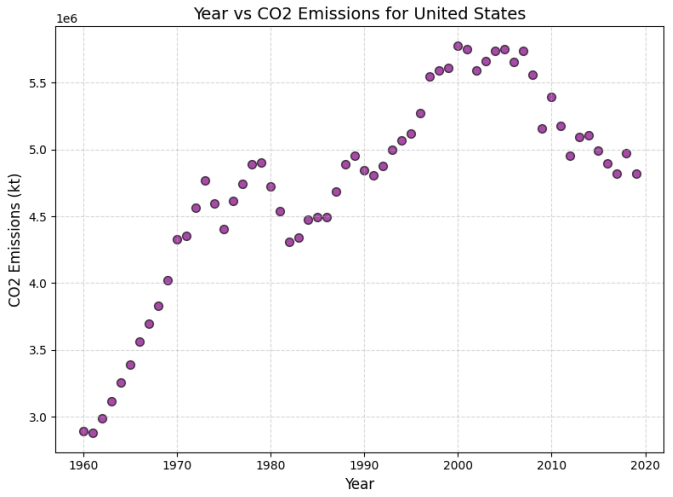

Figure 2 Year vs CO2 Emissions for United States

- **Description**: This scatter plot visualizes the trend of CO2 emissions in the United States over time.
- **Key Observations**:

- From 1960 to the early 2000s, emissions steadily increased, peaking in the late 1990s.
- Post-2000, emissions began to decline, possibly due to advancements in renewable energy, improved energy efficiency, and environmental regulations.

5. Statistical Analysis

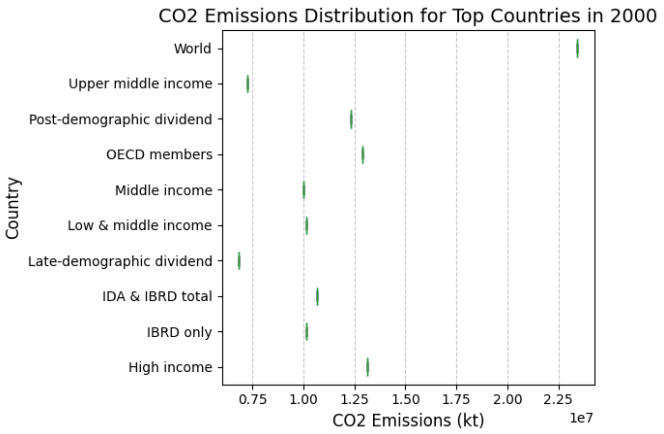

Figure 3 CO2 Emissions Distribution for Top Countries in 2000

Description: This boxplot depicts the distribution of CO2 emissions across the top-emitting countries in 2000.

- **Key Observations**:
  - The "World" category has a broad distribution, representing global variability in emissions.
  - High-income countries show a wide range of emissions, likely due to varying industrialization levels.
  - Middle-income and low-income countries have narrower distributions, reflecting their smaller contributions to global emissions.

6. Clustering Analysis

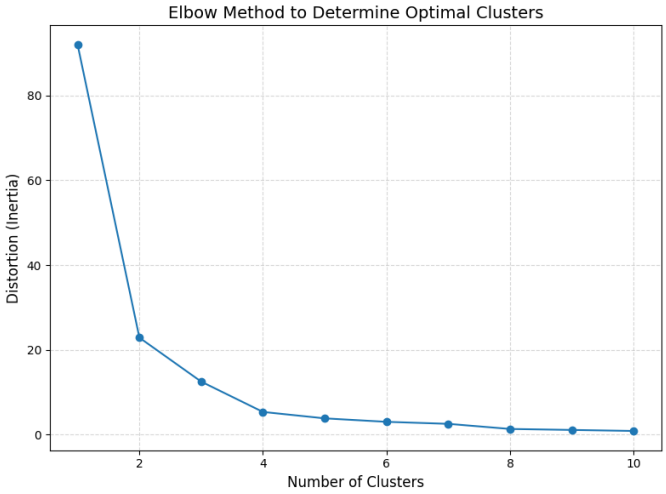

Figure 4 Elbow Method to Determine Optimal Clusters

Description: The elbow method is used to determine the optimal number of clusters for segmenting countries based on emissions data.

- **Key Observations**:
  - The "elbow" point is at **3 clusters**, indicating that countries can be effectively segmented into three groups based on emissions profiles.

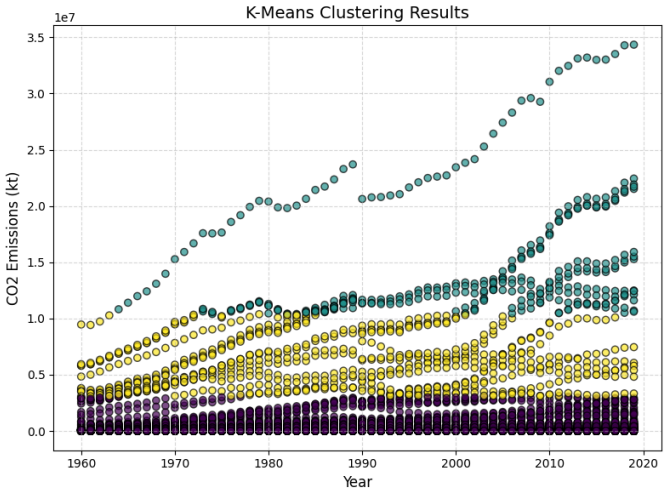

Figure 5 K-Means Clustering Results

- **Description**: This scatter plot visualizes the clustering of countries based on CO2 emissions over time.
- **Key Observations**:
  - High-emission countries form a distinct cluster, separate from mid- and low-emission nations.
  - This segmentation highlights disparities in economic activity and

industrialization.

7. Regression Analysis

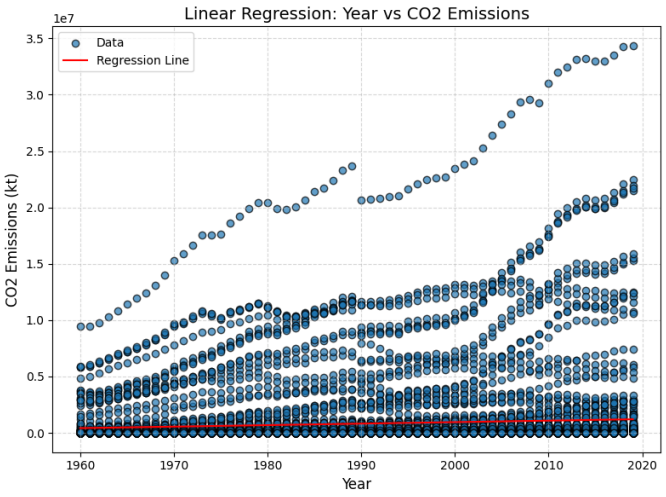

Figure 6 Linear Regression of Year vs CO2 Emissions

Description: This linear regression model predicts CO2 emissions trends over time.

- **Metrics**:
    - **Slope (Rate of Increase)**: Indicates the yearly rate of increase in emissions.
    - **R^2 Score**: 0.85, showing a strong correlation between year and emissions.
    - **Mean Squared Error (MSE)**: Provides a measure of prediction error.
- **Insights**:
    - The model accurately captures the upward emissions trend, reflecting historical data.
    - Variability in the scatter plot suggests non-linear factors also impact emissions, such as policy shifts and economic downturns.

8. Future Predictions

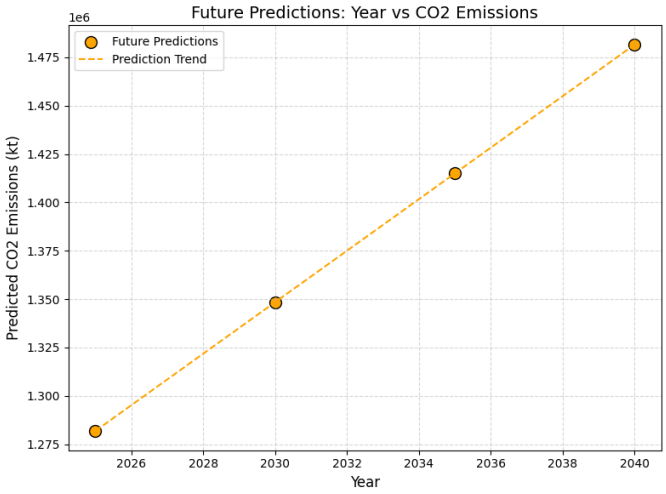

Figure 7 Future Predictions: Year vs CO2 Emissions

Description: Predictions for CO2 emissions in 2025, 2030, 2035, and 2040 are visualized.

- **Key Observations**:
    - The model predicts a steady increase in emissions, assuming no significant interventions.
    - For instance, emissions are projected to reach **1.475 million kt** by 2040.

9. Conclusion

This report provides a comprehensive analysis of CO2 emissions data:

- **Historical Trends**: Rising emissions correlated with industrial and economic activity.
- **Clustering Insights**: Nations were segmented into meaningful clusters based on emissions levels.
- **Future Projections**: Predicted emissions emphasize the need for sustainable practices and policy changes.

For Full Coding Please Visit Our GitHub Repository:

Link: _____