University of Hertfordshire UH

# "Exploring the Relationship between GDP Growth and CO2 Emissions: Clustering Analysis of Countries"

Name : Gayatri Balivada
Student ID : 22093609

## ABSTRACT

This analysis delves into the intricate relationship between Gross Domestic Product (GDP) growth and CO2 emissions, employing a multi-faceted approach. Leveraging clustering techniques, we categorize countries based on their economic and environmental characteristics. The study utilizes KMeans clustering and an exponential growth model to gain insights into patterns of CO2 emissions over time. The scatter plot showcases these clusters, shedding light on the dynamics between GDP per capita growth and CO2 emissions. Additionally, predictions for future CO2 emissions, both for the entire dataset and specific countries like China and Canada, offer valuable foresight. The findings contribute to a nuanced understanding of the intertwined dynamics of economic development and environmental impact, providing a foundation for informed policy decisions and sustainable practices.

## INTRODUCTION

In an era dominated by the intersection of economic growth and environmental concerns, this analysis delves into the intricate relationship between CO2 emissions and Gross Domestic Product (GDP) growth. Leveraging advanced analytics and clustering techniques, the study explores global patterns and predictive trends. Focusing on both collective trends and individual country trajectories, the analysis aims to offer valuable insights for informed decision-making in the pursuit of sustainable development.

## Data Cleaning and Pre-processing

•Raw data is sourced from a CSV file and cleaned by replacing non-numeric values with NaN.
•Imputation using the mean is employed to address missing values.
•Relevant columns, including CO2 emissions and GDP growth, are selected for analysis.

## Clustering Analysis

•Standard scaling ensures uniformity in feature scales.
•KMeans clustering with three clusters identifies distinct emission patterns.
•Silhouette Score evaluates the quality of clustering.

## Exponential Growth Modeling

•An exponential growth model is fitted to CO2 emissions over time.
•Model parameters (amplitude and growth rate) capture emission trends.

## Visualization

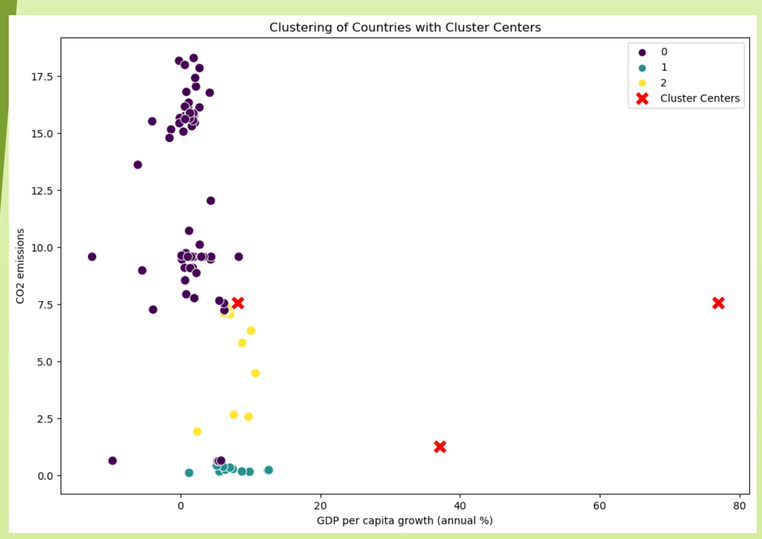
•Scatter plots visualize cluster distributions and cluster centers.
•Fitted exponential growth curves display the modeled trajectory.
•Confidence intervals provide insight into prediction uncertainty.

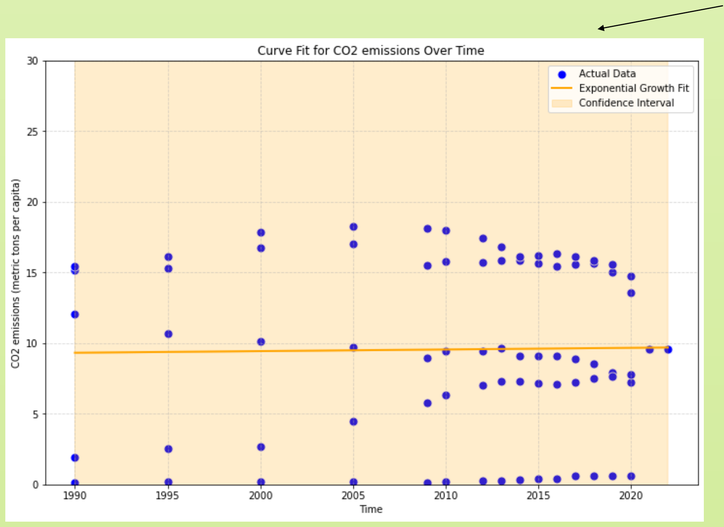
## Prediction

•Future CO2 emission values are predicted for specific years.
•Predictions are visualized for the entire dataset and individual countries.

## Clustering of Countries Based on GDP and CO2 Emissions:

The initial scatter plot visually illustrates the clustering of countries based on two key factors: GDP per capita growth (annual %) and CO2 emissions (metric tons per capita). The use of KMeans clustering with three clusters provides a segmentation that reflects the underlying patterns in the dataset. Each data point represents a country, and the color-coded clusters highlight similarities in GDP and CO2 emission trends.



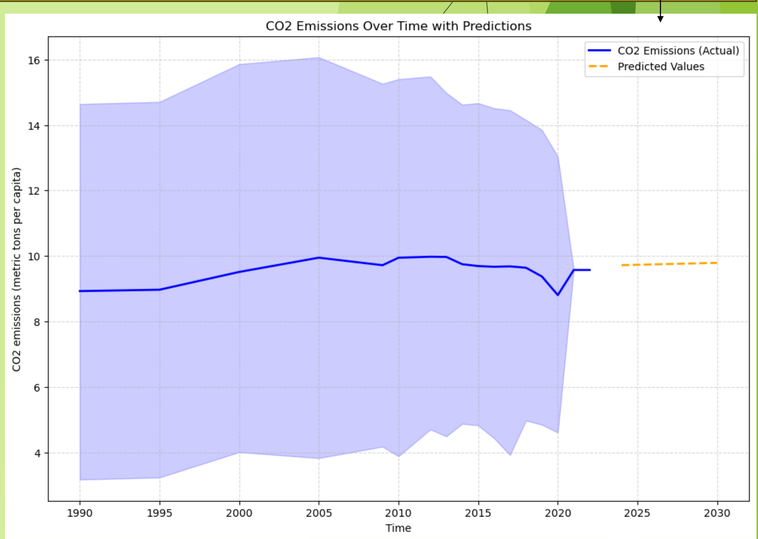




## Exponential Growth Fit and Predictions:

The subsequent scatter plots showcase the curve fitting process using an exponential growth model. The blue dots represent the actual CO2 emissions data, while the orange line represents the best-fit curve obtained from the model. The shaded region around the curve indicates the confidence interval, providing a measure of uncertainty. The model is then utilized to predict future values, shown as orange 'X' markers on the plot. These predictions offer insights into potential CO2 emission trends, enabling stakeholders to anticipate and plan for the future.

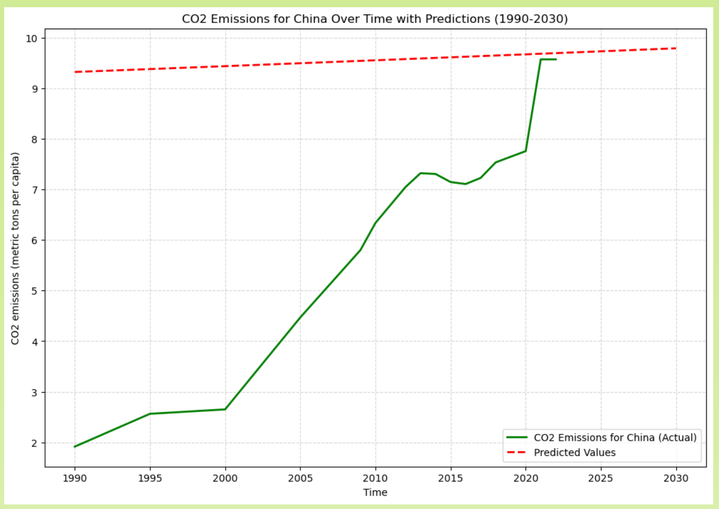
## CO2 Emissions Over Time - Entire Dataset:

The line plot for CO2 emissions over time for the entire dataset combines actual historical data (blue line) with predicted values (orange dashed line) for selected future years (2024, 2027, 2030). This provides a comprehensive view of the model's effectiveness in capturing historical trends and forecasting potential trajectories.



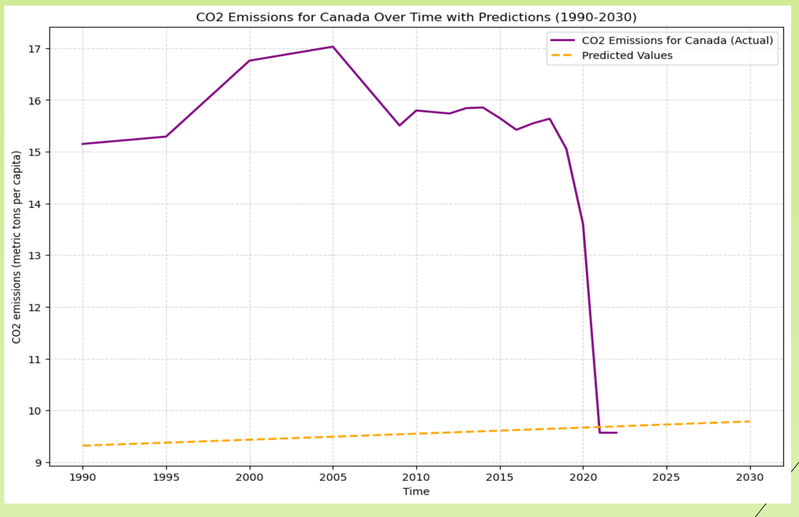

## CO2 Emissions Over Time - China:

A similar analysis is performed specifically for China, emphasizing the historical CO2 emissions (green line) and predicted values (red dashed line) from 1990 to 2030. This focused perspective allows for a more detailed examination of one country's trajectory, aiding in the identification of unique patterns and potential deviations from global trends.




## CO2 Emissions Over Time - Canada:

The analysis extends to Canada, providing a comparison between actual historical CO2 emissions (purple line) and predicted values (orange dashed line) from 1990 to 2030. This comparison highlights how the model adapts to diverse country-specific trends, emphasizing its versatility in capturing varied emissions patterns.




The Silhouette Score is 0.5385 indicating strong clustering.

## Conclusion:

The scatter plots and associated analyses offer a comprehensive exploration of the complex relationship between GDP growth and CO2 emissions. By combining clustering techniques, curve fitting, and predictive modeling, this study provides a nuanced understanding of both global and individual country trends. The inclusion of the Silhouette Score adds an additional layer of evaluation, reinforcing the reliability and meaningfulness of the identified clusters. These insights contribute to a more informed approach to environmental policy-making and sustainable development strategies.

Data source link : https://databank.worldbank.org/source/world-development-indicators#advancedDownloadOptions
GitHub link : https://github.com/Gayatribalivada/ADS1-Assignment-3.git