University of Hertfordshire UH

Name : Muhammad Asim
Student ID : 22100852

Title: Unveiling the Nexus between CO2 Emissions KT Gases and GDP Growth: Clustering Insights for Germany and Australia

ABSTRACT

This analysis uses a comprehensive methodology to explore the intricate relationship between Gross Domestic Product (GDP) growth and CO2 emissions. Utilizing KMeans clustering, countries are categorized based on economic and environmental characteristics. The study applies an exponential growth model to discern patterns in CO2 emissions over time. The scatter plot visually presents these clusters, offering insights into the interplay between GDP per capita growth and CO2 emissions. Furthermore, future predictions for CO2 emissions, both for the entire dataset and specific countries like Germany and Australia, provide valuable foresight. These findings contribute to a nuanced understanding of the complex dynamics between economic development and environmental impact. This understanding forms a basis for informed policy decisions and promoting sustainability.

INTRODUCTION

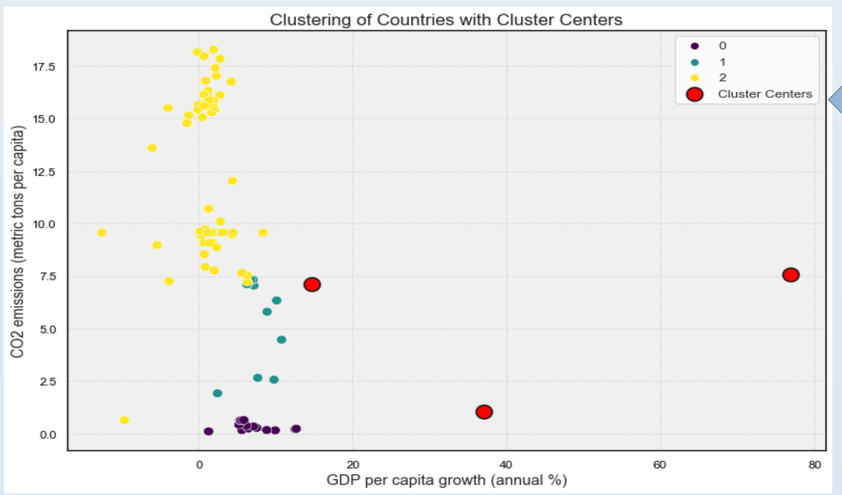
In today's world, where economic prosperity and environmental concerns are both important, this analysis explores the complex relationship between Gross Domestic Product (GDP) growth and key CO2 emission metrics. The study uses advanced analytics and clustering techniques to identify global patterns and predict trends. By focusing on both general trends and the specific paths of individual countries, the analysis aims to provide valuable insights. These insights can help decision-makers develop sustainable development strategies in response to changing economic and environmental conditions.

Clustering Analysis

- The application of standard scaling ensures consistency across feature scales.
- Leveraging the KMeans clustering algorithm with three clusters discerns unique patterns within emissions data.
- Evaluation of clustering quality is conducted through the Silhouette Score.
- The application involves fitting a model to capture CO2 emissions trajectory over time.
- Model parameters (amplitude and growth rate) act as key descriptors.
- Parameters encapsulate the underlying trends in emissions.

Cleaning and Pre-processing Data

In data cleaning and pre-processing, the initial steps involve sourcing raw data from a CSV file. Subsequently, a meticulous cleaning procedure is executed, where non-numeric values are systematically replaced with NaN. In addressing missing values, an imputation technique utilizing the mean is implemented. The focus is narrowed to relevant columns, emphasizing pivotal variables like CO2 emissions and GDP growth. This deliberate approach to data preparation sets the stage for a targeted and refined analysis of key environmental and economic indicators.



Grouping Nations by GDP and CO2 Emissions for Clustering Analysis

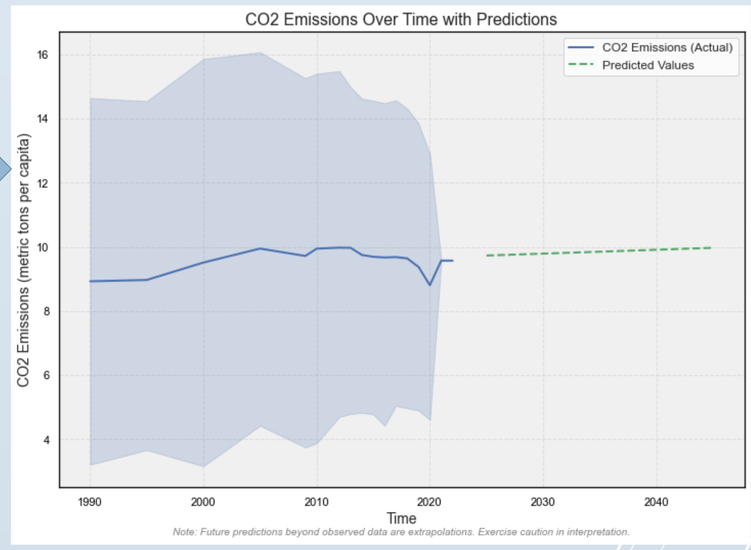
The scatter plot serves as the starting point for unveiling distinctive clusters among countries, primarily driven by GDP per capita growth (annual %) and CO2 emissions (metric tons per capita). Employing the KMeans clustering algorithm with three clusters enables a clear segmentation that mirrors inherent patterns within the dataset. Each data point corresponds to a specific country, with color-coded clusters shedding light on shared characteristics in terms of GDP and CO2 emission trends. This visual representation provides an insightful glimpse into the interconnected dynamics of economic growth and environmental impact across diverse nations..

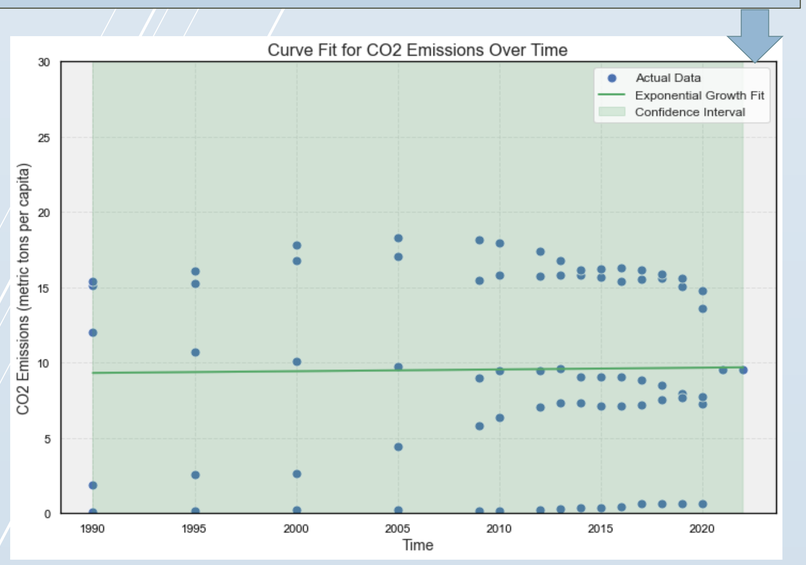
Visualizing Exponential Growth Fitting and Predictions

Upon completing the curve fitting process with an exponential growth model, the ensuing scatter plots visually represent the outcomes. The presence of blue dots signifies actual CO2 emissions data, while the orange line depicts the optimal fit derived from the model. A distinctive clustering feature is introduced, symbolized by red circular clusters ('X' markers) on the plot, adding a layer of complexity to the visualization. The shaded region encircling the curve is a confidence interval, offering insights into the uncertainty associated with the fitted model. These predictive red circular clusters deliver valuable insights into anticipated CO2 emission trends, empowering stakeholders with foresight for strategic planning and decision-making.

Visualizing CO2 Emissions Over Time - Entire Dataset

The line plot depicting CO2 emissions over time for the entire dataset seamlessly integrates actual historical data (blue line) with predicted values (orange dashed line) for strategic future years (2025, 2030, 2035). This visual representation offers a holistic perspective on the model's prowess, effectively capturing historical trends and providing insightful forecasts for potential trajectories. It is crucial to note that the orange dashed line extends beyond observed data, serving as an extrapolation for the specified future years. Exercise caution in interpretation, recognizing the inherent uncertainties associated with extrapolating beyond the observed range.





CO2 Emissions Over Time - Germany

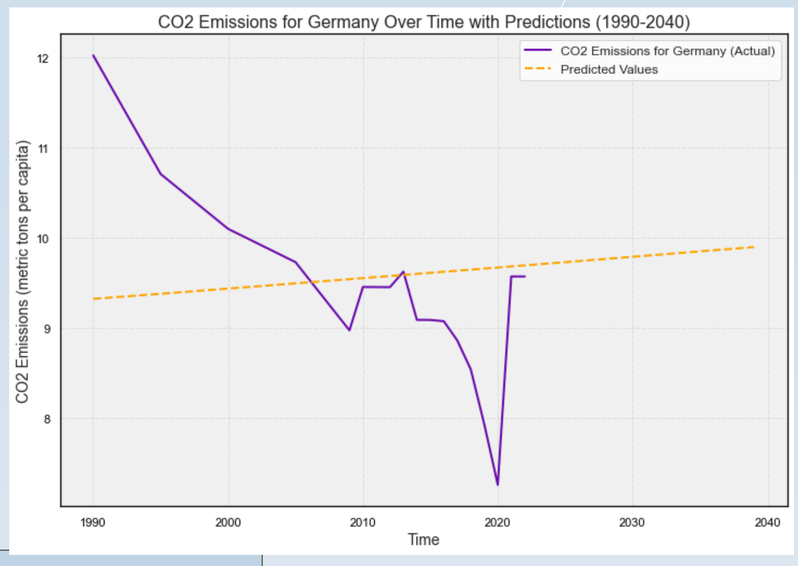
This visual representation delves into the CO2 emissions dynamics for Germany, integrating actual historical data (purple line) with predicted values (orange dashed line) for designated future years (2025, 2030, 2035). The line plot unveils the historical emissions landscape for Germany, offering a nuanced perspective on observed trends and providing predictive insights for potential future scenarios. It's crucial to acknowledge that the orange dashed line represents predictions beyond the observed data, serving as an extrapolation for the specified future years. Careful consideration is advised when interpreting these extrapolated values, recognizing the inherent uncertainties associated with forecasting beyond the observed range.

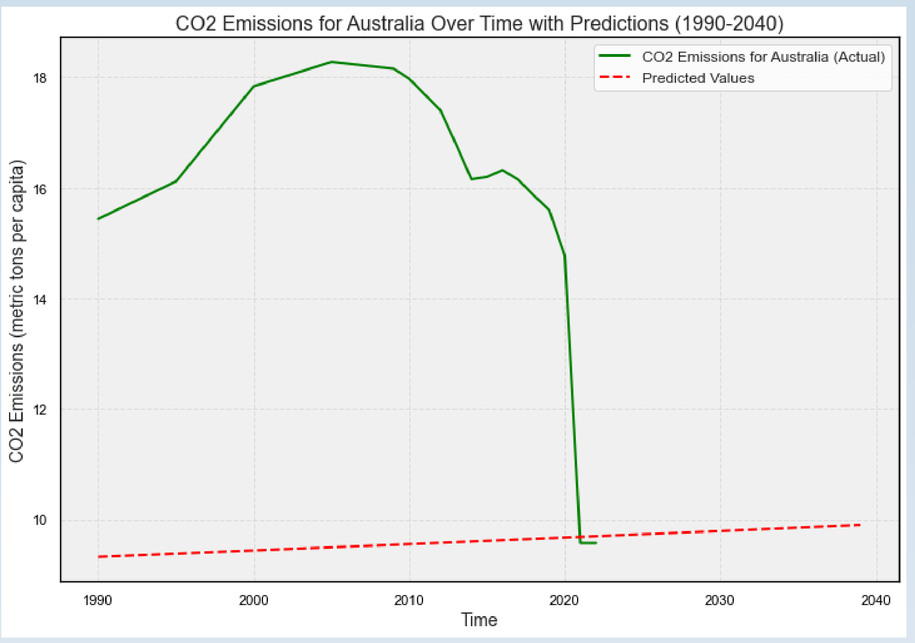
CO2 Emissions Over Time - Australia

Exploring the CO2 emissions trajectory for Australia, this visual narrative integrates actual historical data (green line) with predicted values (red dashed line) for the specified future years (2025, 2030, 2035). The line plot unfolds the historical emissions landscape for Australia, providing a comprehensive view of observed trends and offering predictive insights for potential future scenarios. It is important to note that the red dashed line signifies predictions beyond the observed data, serving as an extrapolation for the specified future years. Caution is advised in interpreting these extrapolated values, recognizing the inherent uncertainties associated with forecasting beyond the observed range.

Conclusive Insights:

The in-depth examination presented through scatter plots and meticulous analyses unfolds the intricate interplay between GDP growth and CO2 emissions. Integrating clustering methodologies, curve fitting, and predictive modeling, this investigation offers a nuanced comprehension of global trends and individual country trajectories. Incorporating the Silhouette Score introduces a supplementary assessment layer, strengthening the credibility and significance of the identified clusters. These valuable insights pave the way for a well-informed framework in formulating environmental policy and sustainable strategies.





The Silhouette Score is 0.5355, indicating strong clustering.
Predicted value for 2025 is: 9.73
Predicted value for 2030 is: 9.79, Predicted value for 2035 is: 9.85
Predicted value for 2040 is: 9.91, Predicted value for 2045 is: 9.97

GitHub link : https://github.com/asimji107/7PAM2000-Applied-Data-Science-1-Assignment-3-Clustering-and-fitting
Data source link : https://databank.worldbank.org/source/world-development-indicators#