Analyzing the distribution of energy usage across the world using clustering and fitting techniques

By S.R. De Abrew

(School of Physics, Engineering and Computer Science - University of Hertfordshire)

Abstract

This research employs the K-mean algorithm to conduct a clustering analysis using data on energy consumption per capita and CO2 intensity obtained from the World Bank data repository. The objective is to reveal unique patterns and variations among countries based on their energy consumption and CO2 intensity. Additionally, the poster visually presents fitting and prediction plots for Canada and Australia

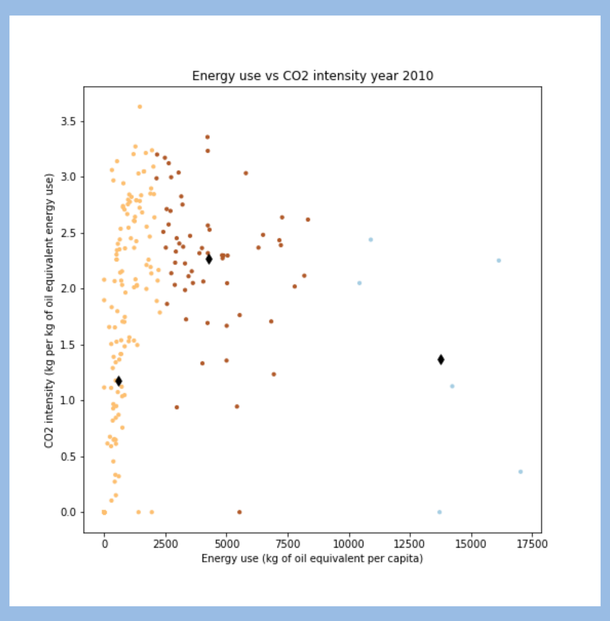
Introduction

In the pursuit of understanding global energy dynamics and environmental impact, this study leverages the K-mean algorithm to perform a comprehensive clustering analysis. The analysis is centered around data derived from the World Bank repository, specifically focusing on energy consumption per capita and CO2 intensity.

The primary goal is to discern distinctive patterns and variations across countries, shedding light on their energy consumption behaviors and corresponding CO2 intensity levels. Furthermore, this research is complemented by a visually insightful poster that showcases fitting and prediction plots, specifically tailored for Canada and Australia. This holistic approach aims to contribute valuable insights into the nuanced relationships between energy consumption, CO2 intensity, and the unique characteristics of individual countries

Analyses work:

K-means clustering, a popular unsupervised technique, partitions data points into distinct groups based on similarity. It is widely used for its simplicity and efficiency in handling large datasets. By leveraging the strengths of both supervised and unsupervised methods, data scientists can gain a comprehensive understanding of data, uncover hidden patterns, and make informed decisions across various domains, including business, healthcare, and research

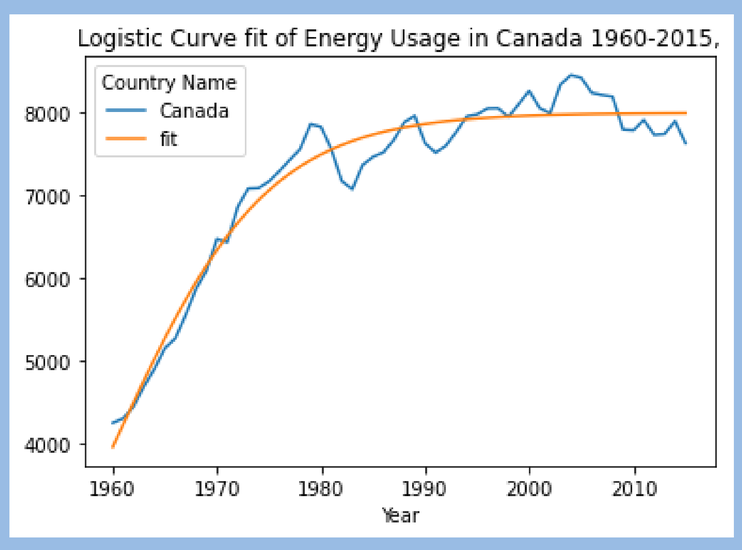


When examining the attributes of Energy use (kg of oil equivalent per capita) and CO2 intensity indicators, as illustrated in the above plot graph, three clusters have been formed. The majority of countries within these clusters exhibit Oil consumption per capita values below five thousand USD. These clusters represent discernible sets of data points .

For this chart, data from year 2010 was extracted for both indicators using Pandas methods and transformed into a pivot table. The primary data source for this extraction was the World Bank's main data file. Furthermore, the visualization was created using the Matplotlib library.

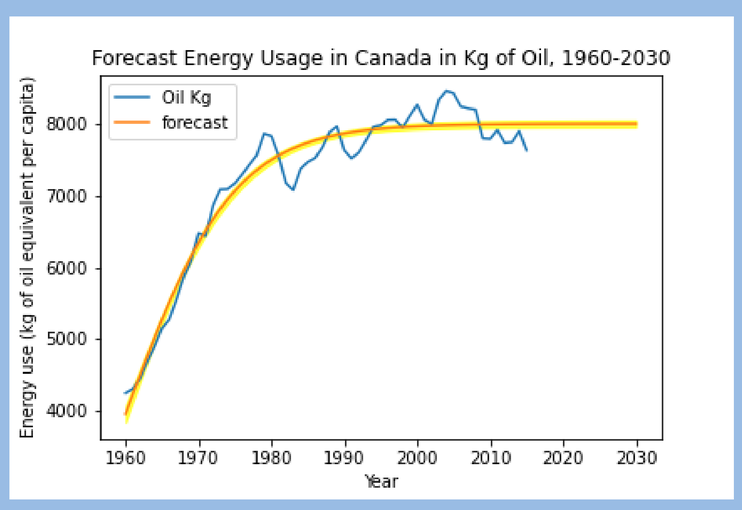
logistic curve fits are employed in scatter plotting when the data exhibits a sigmoidal trend, and the goal is to model growth, saturation, or decline processes. This type of curve fitting is valuable for understanding the underlying dynamics of the system, making predictions, and extracting meaningful insights from the data

Analyzing the logistic curve fit for Canadian oil usage spanning from 1960 to 2015 provides insights that suggest a stable trend, allowing for predictions of stability in the coming years



The graph below indicates a steady rise in Canada's energy usage, measured in kilograms of oil equivalent per capita, suggesting an anticipated increase to 8000 USD by 2030. The consistent rate of this upward trend is noteworthy. Consideration of factors like government policies and economic development is crucial, as they can influence energy usage trajectories. Future analyses should account for these variables for more accurate predictions

.



Conclusion

The clustering analysis of per capita energy usage data provides a comprehensive understanding of the distinctions and resemblances in energy consumption trends. Understanding the factors influencing energy usage patterns enables the optimization of energy utilization, enhancement of energy efficiency, and mitigation of environmental impact. To enable more targeted interventions and energy management techniques for sustainable development, future studies can concentrate on examining the elements contributing to the observed clustering patterns.

Reference

GitHub Link : https://github.com/savinda79/ADS1-Assignment-3-ClusteringFitting

Original Data Source : https://data.worldbank.org/topic/climate-change

University of Hertfordshire UH

THE WORLD BANK
IBRD • IDA

sd23aai@herts.ac.uk