CO2 EMISSION TRENDS IN DIFFERENT COUNTRIES
ADS-1 Clustering and fitting

MD REZAUL KARIM

Repo Link: https://github.com/rezaulcmt/ADS1_Assignment_Clustering-and-fitting.git

Abstract

The project conducts a comprehensive analysis of CO2 emission within selected countries by focusing to cover-up the patterns and new trends that are very important for understanding the international challenges related to the environment. The technique of generating clusters such as K-Means and analyzing modelling, this project explores the dynamics of CO2 emission by going through the key indicators involved in climate change. The data is collected from the World Bank which is a public platform for the research dataset. The analysis of the project spans the selection of pertinent indicators and different countries. It is done by normalization with the help of Standard Scaler, and employing a predictive model for calculating the rate. The technique of visualization including the heat maps and clustered data plots is used to visualize the findings. The uniqueness of the study is to address the change of climate by CO2 emission to the environment.

Introduction

The step-by-step global concerns related to the change of climate underscore the imperative to comprehensively figure out and understand the dynamics of CO2 emission. As an initial innovator of climate change CO2 emissions are intricately linked to the degradation of the environment. Also, the rising temperature, ecological imbalance and many more. This project report is determined to contribute to the ongoing disbalance by conducting a deep analysis of the emission of CO2. After finding out the intricacies between the activities of industry, the growth of population, and different policies the research is focused on unveiling the patterns and the trends in case of emission. This research considers different indicators of key and clustering techniques by normalization and predictive modelling. After going through this manifested exploration the report of the project endeavours to inform multiple discussions on favourable environmental practices and policies. It is also aimed to foster a good understanding of the data of the World Bank in the context of climate change mitigation.

CO2 Emission Analysis

The changing patterns of atmospheric carbon dioxide discharges across a couple of picked countries are inspected in the part on CO2 Emission Analysis. The analysis endeavours to recognize patterns and examples in emissions by using bunching strategies, normalized techniques, and prescient displaying. Visualization innovations contribute fundamental information for observing the climate and strategy improvement by working on the translation of complicated data (Dong *et al*. 2020). The various components of the carbon dioxide emissions specifically countries are analyzed in the following segment on the CO2 Emission Assessment. Utilizing vision conduct, normalization, and grouping strategies, the analysis means to pinpoint specific occasions and patterns in emissions. Visualization advancements work with the understanding of complex information and deal with fundamental bits of knowledge for ecological observing and vital preparation.

Data Acquisition and Cleaning

The validity and exhaustiveness of the CO2 emission study are guaranteed through the obtaining and purifying of information from the data set of the World Bank. The dataset, which is organized in CSV design, remembers information for CO2 emissions and related measurements for different countries. During the cleaning system, pointless or superfluous information was disposed of to make information consistent. In the beginning, different columns such as Country Code and Indicator Code have been removed from the dataset (Khan *et al*. 2020). As a solution standarized were maintained by renaming columns to make strong interpretability by simplifying into Country name and Country.

Selection of Indicators and Countries

The objective of this study is to catch different patterns in carbon dioxide (CO2) emissions, which has prompted an essential strategy in the determination of markers and countries. "Population development (yearly %)" and "CO2 emissions (kt)," the chosen measurements, were painstakingly decided to offer a piece of careful information on the intricate elements prompting ecological worries (Crippa *et al*. 2020). The previous assesses this present reality effect of CO2 emissions, while the last option goes about as a financial market, presenting factors associated with population.

Clustering Analysis

Heatmaps and clustered information plots are two instances of visuals that can make clustering discoveries more clear. These realistic guides upgrade comprehension of the examples found and give data on the links and contrasts in CO2 emissions between countries (Yoro and Daramola, 2020). In aiding in the evaluation and check of the subsequent gatherings, a graphical score is utilized, which offers a mathematical portrayal of the nature of the gathering..

Normalization and Predictive Modeling

By using a Standard Scaler to standardize the information, giving a uniform scale to CO2 emission markers among nations is conceivable. Using uniform frameworks, any predispositions coming about because of various estimating units are tended to and emissions are made more similar. Finding huge examples is supported by the normalization step, which makes sure that every country's cooperation to the analysis of clustering is determined on a comparable scale.

The bend fitting strategies are then used in prescient displaying to decide plausible patterns in CO2 emissions (Yao *et al*. 2020). The inborn development paces of every country's emission direction are caught by the remarkable development model. The exploration goes past previous information and gives projections to the impending a very long time by using bend fitting. The use of certainty stretches in mistake range calculations reinforces the legitimacy of these gauges. This reenactment strategy adds to an intensive knowledge of the ecological impact of the picked countries by giving sagacious data about conceivable future emission situations.

Visualization

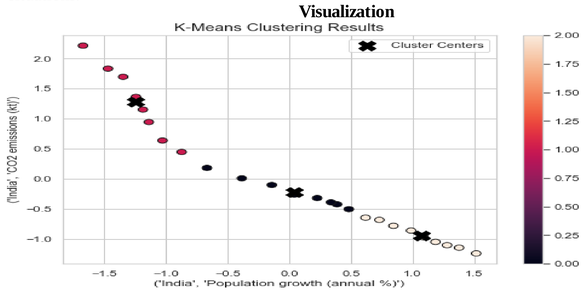


Figure 1: K-means Clustering Result for CO2 emission vs Population Growth in India

(Source: Generated using Python in Sypder by self)

A data frame is created with the CO2 emission and Population Growth data of India. The clustering of the data is generated using K Means with the number of clusters as 3. Then the Silhouette score is calculated using the score silhouette generation function. The Silhoutee score for CO2 emission vs Population Growth in India is 0.55 approx. Then using the plot clustered data function the scatter plot is generated to visualize the clustered data points. The centroids or central tendencies are marked as a black cross in the visualization. The legend of the visualization shows different colors for the different density levels (Rietmann *et al*. 2020). It helps to understand the density of each cluster point.

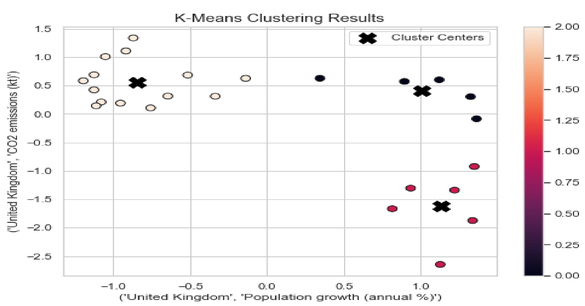


Figure 2: K-means Clustering Result for CO2 Emission vs Population Growth in the United Kingdom

(Source: Generated using Python in Sypder by self)

A data frame is created with the CO2 emission and Population Growth data of the United Kingdom. The clustering of the data is generated using K Means with the number of clusters as 3. Then the Silhouette score is calculated using the score silhouette generation function. The Silhoutee score for CO2 emission vs Population Growth in the United Kingdom is 0.62 approx. Then using the plot clustered data function the scatter plot is generated to visualize the clustered data points (Sovacool et al. 2020). The centroids or central tendencies are marked as a black cross in the visualization. The legend of the visualization shows different colors for the different density levels. It helps to understand the density of each cluster point.

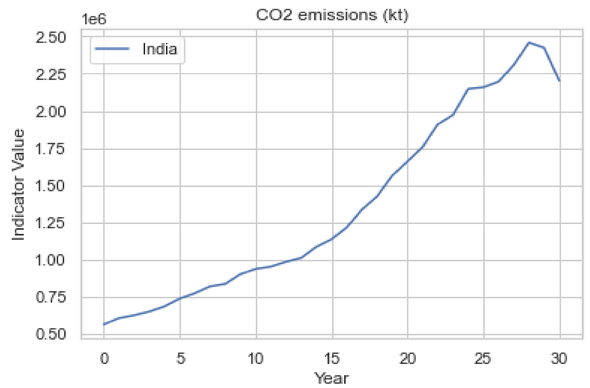


Figure 3: CO2 Emission in India

(Source: Generated using Python in Sypder by self)

The CO2 emission graph for India shows the historical CO2 emission data trends between the years 1990 to 2020. The curve plot is generated using the predict future function which shows the curve based on the CO2 emission. It mainly helps to understand the upcoming possible emissions in India. According to the graph, Carbon dioxide emissions increased gradually from 1990 to 2016. The highest CO2 emission occurs in India is close to 2400000 kt. Then after 2016, it decreased. So In the future, it can be assumed that CO2 emissions will decrease in India.

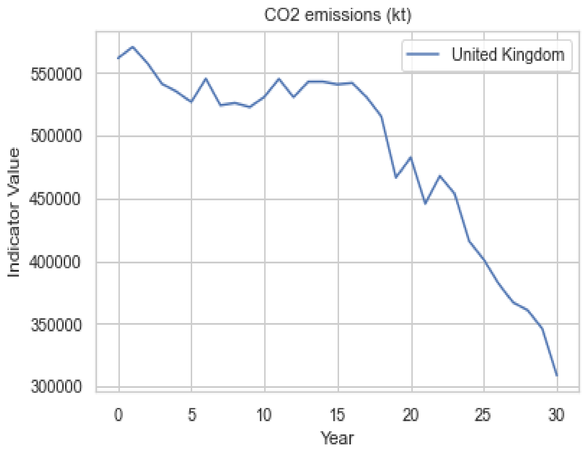
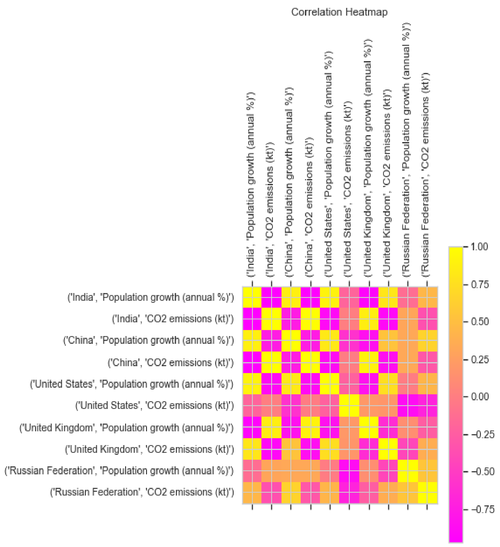


Figure 4: CO2 Emission in The United Kingdom

(Source: Generated using Python in Sypder by self)

The CO2 emission graph for the United Kingdom shows the historical CO2 emission data trends between the years 1990 to 2020. The curve plot is generated using the predict future function which shows the curve based on the CO2 emission. It mainly helps to understand the upcoming possible emissions in the United Kingdom. According to the graph, Carbon dioxide emissions have moderately decreased till now starting from 1990. The highest CO2 emission occurs in the United Kingdom is close to 550000 kt (Uddin *et al*. 2020). According to the visualization, In the future, it can be assumed that CO2 emissions will further decrease in the United Kingdom.

The Correlation Heatmap is generated based on the data frame consisting of the countries, CO2 emission rate and Population growth rate. The correlation values are represented using the colour gradient. The higher density of the colour represents a lower value of correlation and the lower density of the color represents a higher value of correlation (Alsarayreh *et al*. 2020). For better understanding, the values are also shown along with the color gradient.



- **Figure 5: Correlation Heatmap**
- (Source: Generated using Python in Sypder by self)

Conclusion

In conclusion the emission of CO2 analysis that happened in this project report opens about the insightful patterns and different trends, shedding light, on the critical dynamics of the emission throughout some selected countries. The analysis of different clustering methods facilitated by the K-means clustering successfully recognized unique groups by offering a nuanced understanding of different emissions. The process of normalization includes a Standard scaler has contributed to a comprehensive assessment by facilitating a comparison of different countries. Furthermore, the predictive modelling of the coming emission trends has provided valuable foresight by enabling a proactive approach to addressing the challenges related to the environment.

Reference List

Journals

Alsarayreh, M.M.M., AlSuwaidi, M.F., Al Sharif, R.A. and Kutty, A.A., 2020, April. The factors affecting CO 2 emission in the European union countries: A statistical approach to sustainability across the food industry. In *2020 IEEE 7th international conference on industrial engineering and applications (ICIEA)* (pp. 599-604). IEEE.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J.G.J. and Vignati, E., 2020. Fossil CO2 emissions of all world countries. *Luxembourg: European Commission*, pp.1-244.

Dong, K., Hochman, G. and Timilsina, G.R., 2020. Do drivers of CO2 emission growth alter overtime and by the stage of economic development?. *Energy Policy*, *140*, p.111420.

Khan, A.N., En, X., Raza, M.Y., Khan, N.A. and Ali, A., 2020. Sectorial study of technological progress and CO2 emission: Insights from a developing economy. *Technological Forecasting and Social Change*, *151*, p.119862.

Rietmann, N., Hügler, B. and Lieven, T., 2020. Forecasting the trajectory of electric vehicle sales and the consequences for worldwide CO2 emissions. *Journal of Cleaner Production*, *261*, p.121038.

Sovacool, B.K., Schmid, P., Stirling, A., Walter, G. and MacKerron, G., 2020. Differences in carbon emissions reduction between countries pursuing renewable electricity versus nuclear power. *Nature Energy*, *5*(11), pp.928-935.

Uddin, M.M., Mishra, V. and Smyth, R., 2020. Income inequality and CO2 emissions in the G7, 1870–2014: Evidence from non-parametric modelling. *Energy economics*, *88*, p.104780.

Yao, Y., Ivanovski, K., Inekwe, J. and Smyth, R., 2020. Human capital and CO2 emissions in the long run. *Energy economics*, *91*, p.104907.

Yoro, K.O. and Daramola, M.O., 2020. CO2 emission sources, greenhouse gases, and the global warming effect. In *Advances in carbon capture* (pp. 3-28). Woodhead Publishing.