Clustering and Fitting Analysis

Report: Trends in CO2 Emissions and Agricultural Land Use Over Time.

Student Name: Pranay Reddy Bandharapu

Student Number: 23028438

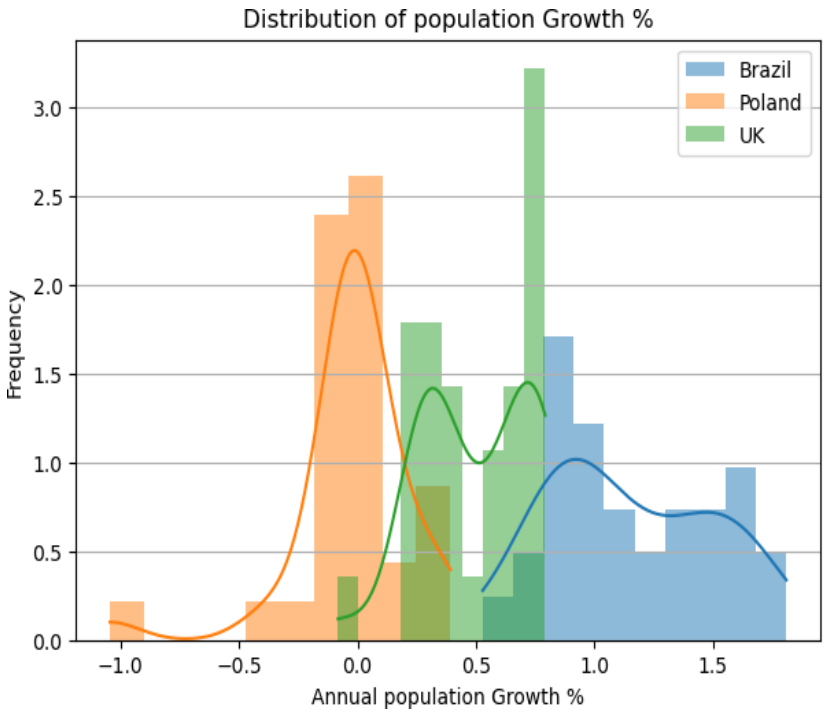
GitHub Repository: https://github.com/Pranayr5/pranay

Introduction:

In order to investigate clustering patterns and fitting relationships within the data, the paper gives an examination of several indicators of seven different countries during a given time period. The analysis consists of fitting analysis with linear regression, grouping with the k-means algorithm, and exploratory data visualization. The data collection includes indicators such as agricultural nitrous oxide emissions (measured in thousand metric tons of CO2 equivalent), GDP growth (annual percentage), population growth (annual percentage), population density (people per square kilometer of land area), and agricultural land (% of land area).

Data Analysis:

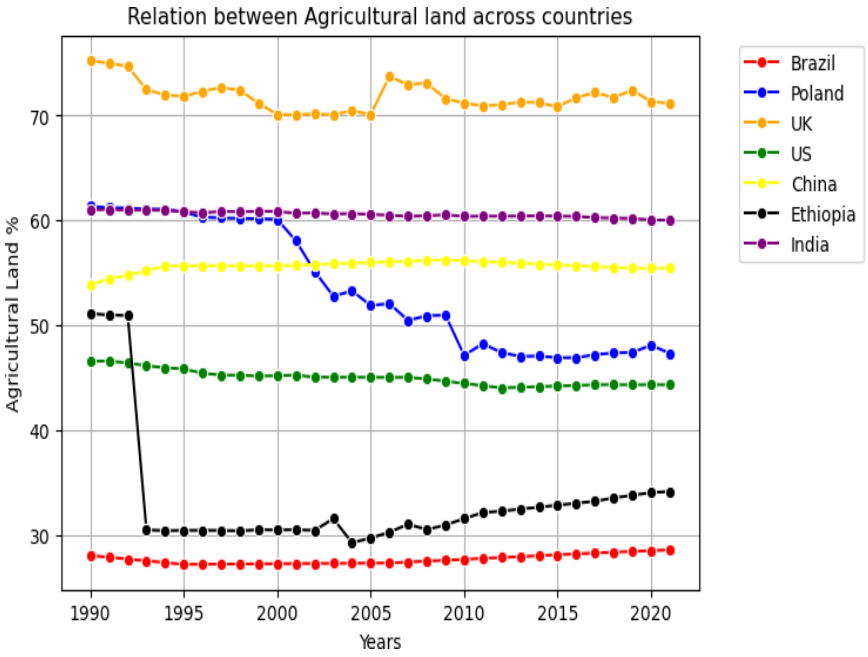
An examination of survey data analysis provides insight into the GDP growth percentage and shifts in farmland use across various nations over time. The data story was visualized using a variety of visual aids, such as scatter plots, heat maps, and histograms. In order to go even farther, we also performed correlation analysis and summary statistics, which helped to clarify data sets and quantify associations.



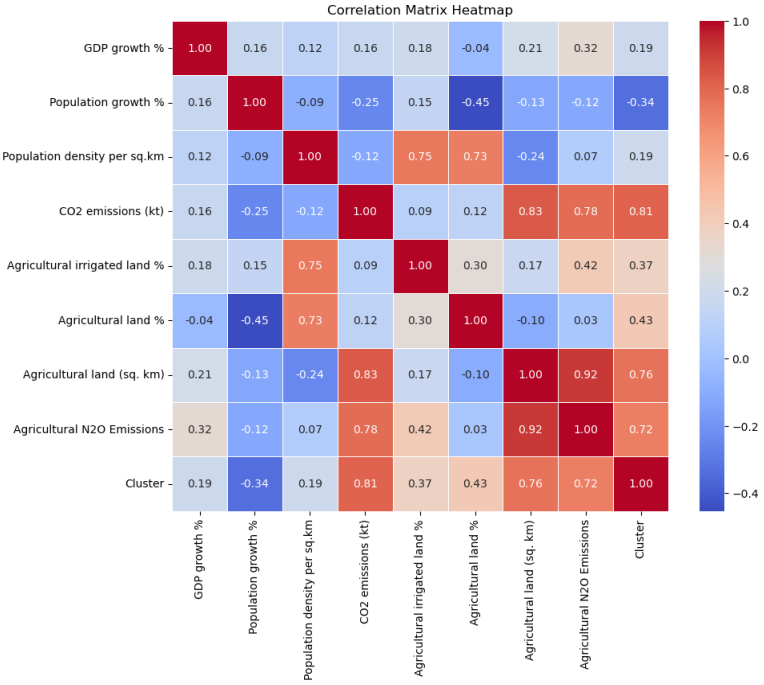
The following observations can be made based on this visualization:

Brazil's distribution is widely dispersed, suggesting a more variable pace of population growth over time. The distribution of Poland points toward the middle of the scale, indicating more steady rates of population growth. The distribution of the UK is very erratic, with several peaks, indicating large fluctuations in the rates of population growth.

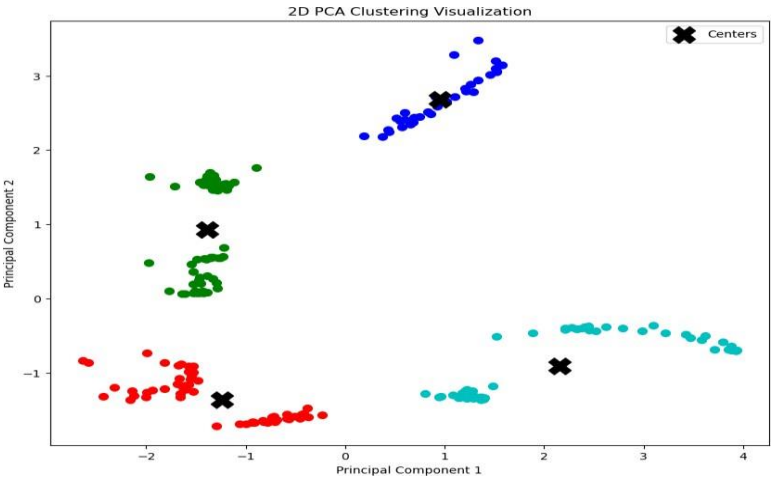
The need for necessities like food and water rises with population growth, which may result in more waste and environmental strain. Furthermore, urbanization tends to increase with population expansion, and this might result in decreased agricultural land, increased pollution, and habitat degradation.



Agricultural land has been comparatively steady throughout time in Brazil and India, with Brazil holding the largest percentage of the listed countries. Poland too has a steady proportion of its land area dedicated to agriculture, albeit smaller than in Brazil or India. The two countries with the lowest percentages during a 30-year period of time were the UK and the US. China begins, then spreads to the US and the UK. Roughly in a more noticeable downturn since 2005. Following an early decline, Ethiopia's share of agricultural land remains stable.



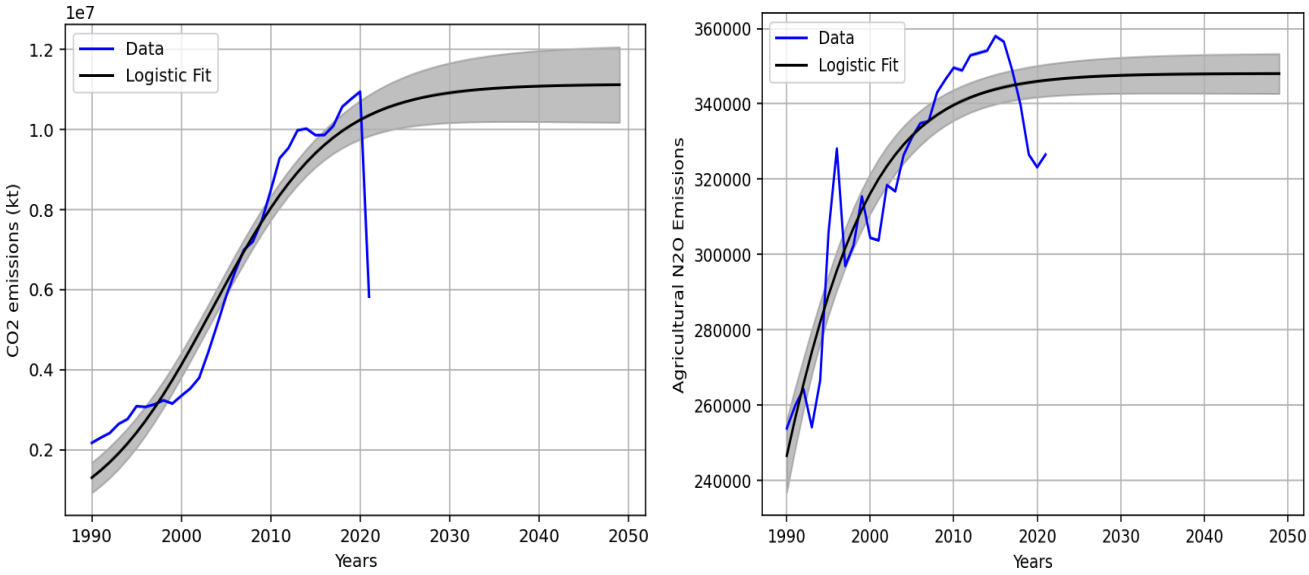
Population and CO2 emissions should rise somewhat when GDP rises, according to the positive downward connection between GDP growth percentage and population per sq km (0.18) and the weak positive correlation between GDP growth percentage and CO2 emissions (kt) (0.16). Agricultural N2O emissions (0.92) Agricultural land (sq km) the quantity of land utilized for agriculture and the greenhouse gas N2O linked to agricultural activities had a very strong positive correlation, showing a direct linkage.



Clustering Analysis

Data from irrigated agricultural land are clustered. For k-means clustering, the ideal number of clusters (k) in a dataset is determined using both elbow and Silhouette Score. The inertia diminishes more slowly after the elbow, which seems to be approximately 4 or 5. From two to four clusters, the silhouette score gets better, indicating that four clusters may be the ideal number. Following that, the score keeps rising but does so more slowly and with some volatility.

increase but at a slower pace and with some fluctuations. Given that this dataset strikes a balance between cluster compactness and separation, both the Elbow approach and the Silhouette score, selecting four clusters, may be appropriate. There are four unique clusters that each have their data points clustered together. In terms of characteristics, similarity between points within a cluster is indicated by their closeness. Two components, "Principal Component 1" and "Principal Component 2," which represent the majority of the variation in the data, are the PCA's output. After the data is reduced to two primary components, this graphic aids in understanding the grouping that the K-means algorithm identified and provides an intuitive understanding of the structure of the data.



The first figure shows CO2 emissions over time, the logistic fit from 1990 to 2050 shows that CO2 emissions are increasing and are expected to increase as the curve approaches the upper limit (L) at year 2050. Where the uncertainty becomes about 944,829 kilotons.

The presented figure shows historical and forecasted agricultural N2O (nitrous oxide) emissions from 1990 to later 2050. The model predicts that agricultural N2O emissions for the year 2050 will be approximately 348,044 units. With a small uncertainty of ±5,299.01 units

Regarding the adverse effects, elevated CO2 levels may result in noteworthy environmental transformations, such as sea level rise, global warming, extreme weather, and altered disease prevalence and distribution. Additionally, air quality may have an impact on public health. The expenses of resolving these effects on environmental health might be high from an economic standpoint. It is crucial to remember that projections for future emissions are dependent on a variety of variables, including shifts in the economy, technology, and policy.

Because N2O is a strong greenhouse gas, increasing emissions of this gas can have unfavorable side effects on the environment, such as contributing to global warming and ozone layer depletion. The usage of nitrogenous fertilizers and increased agricultural activity may be linked to the rise in emissions. Sustainable farming methods and long-term environmental regulations are essential for controlling these emissions.