

K-Means Clustering of Countries Based on Population Growth and Agricultural Land

Student Name: Anusha Vishumolakala

Student ID: 22033054

Report Link :<https://github.com/anushavishnumolakala/ADS-Assignment3>

Abstract: This project analyzes World Bank data on mortality rate and population growth for India, United States, Bangladesh, and China from 1990 to 2020. The KMeans clustering method was used to identify distinct clusters of countries with different trends in mortality rate and population growth. Simple curve models such as exponential growth, logistic function, and low order polynomials were also employed to fit the data sets and predict future values with confidence intervals.

Introduction: Population growth and child mortality are important indicators of a country's development and well-being. The World Bank collects and publishes data on these indicators for many countries around the world. In this project, we analyzed World Bank data on population growth and under-5 mortality rate for the period 1990 to 2020 in four selected countries: India, United States, Bangladesh, and China. Our goal was to identify interesting clusters of data and create simple models that fit the data sets using `curve_fit`.

Result: Our analysis of World Bank data on mortality rate and population growth for four selected countries (India, United States, Bangladesh, and China) revealed three distinct clusters with different trends. Exponential growth and logistic function models provided the best fit for the data sets and were used to predict future values with confidence intervals. These findings can inform policies aimed at addressing challenges related to population growth and child mortality rate.

Conclusion: Our analysis of World Bank data on population growth and child mortality rate for four selected countries showed that there were distinct clusters of countries with different trends in these indicators. We also found that simple curve models such as exponential growth and logistic function provided good fits for the data sets and could be used to predict future values with confidence intervals. These findings can provide valuable insights for policymakers and researchers interested in understanding and addressing issues related to population growth and child mortality.

