

Exploring Clusters and Predicting Trends in Agriculture and Population Data for Selected Countries: A Study Using World Bank Indicators

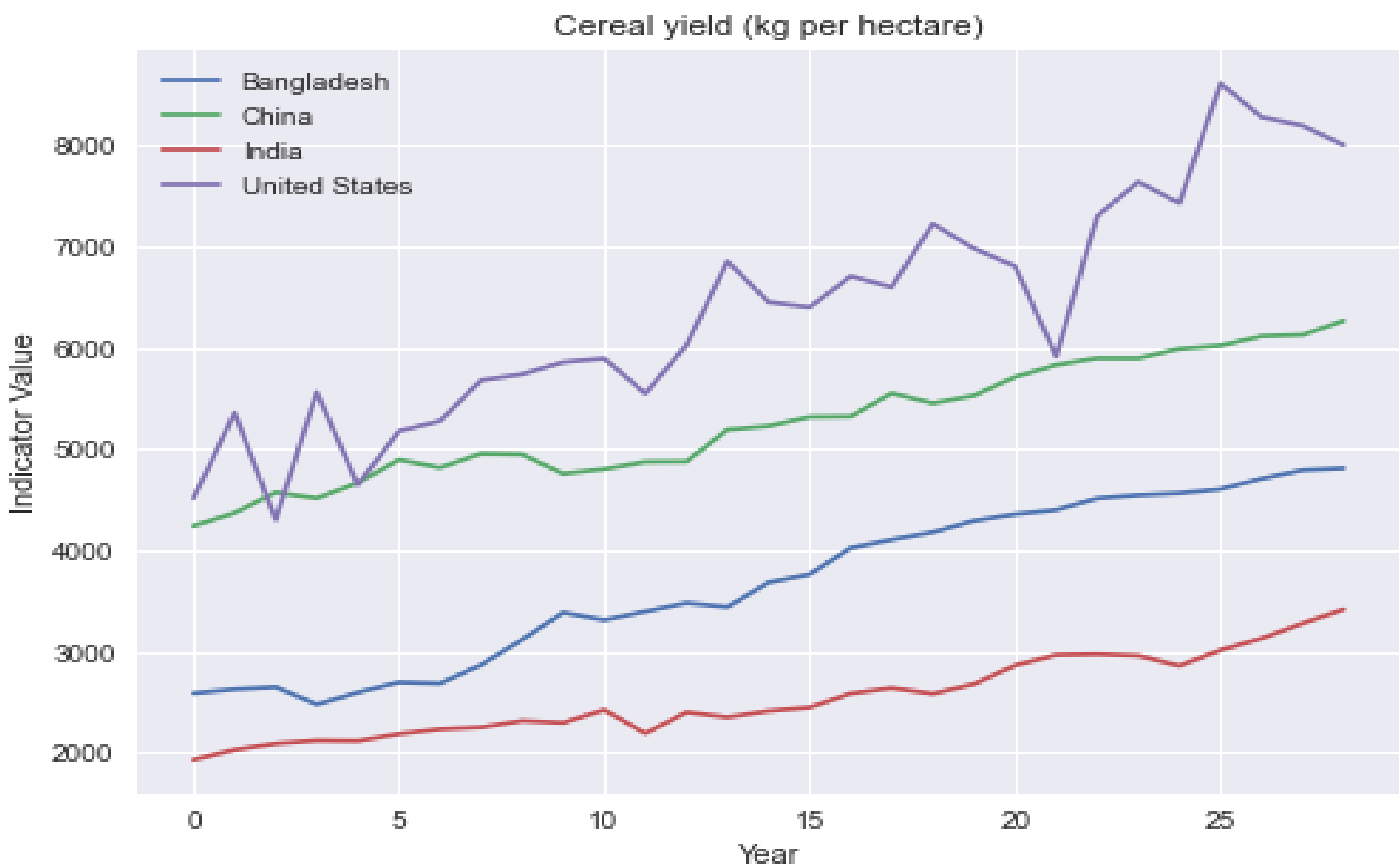
Name: Puligilla Uday

Student Id:22033782

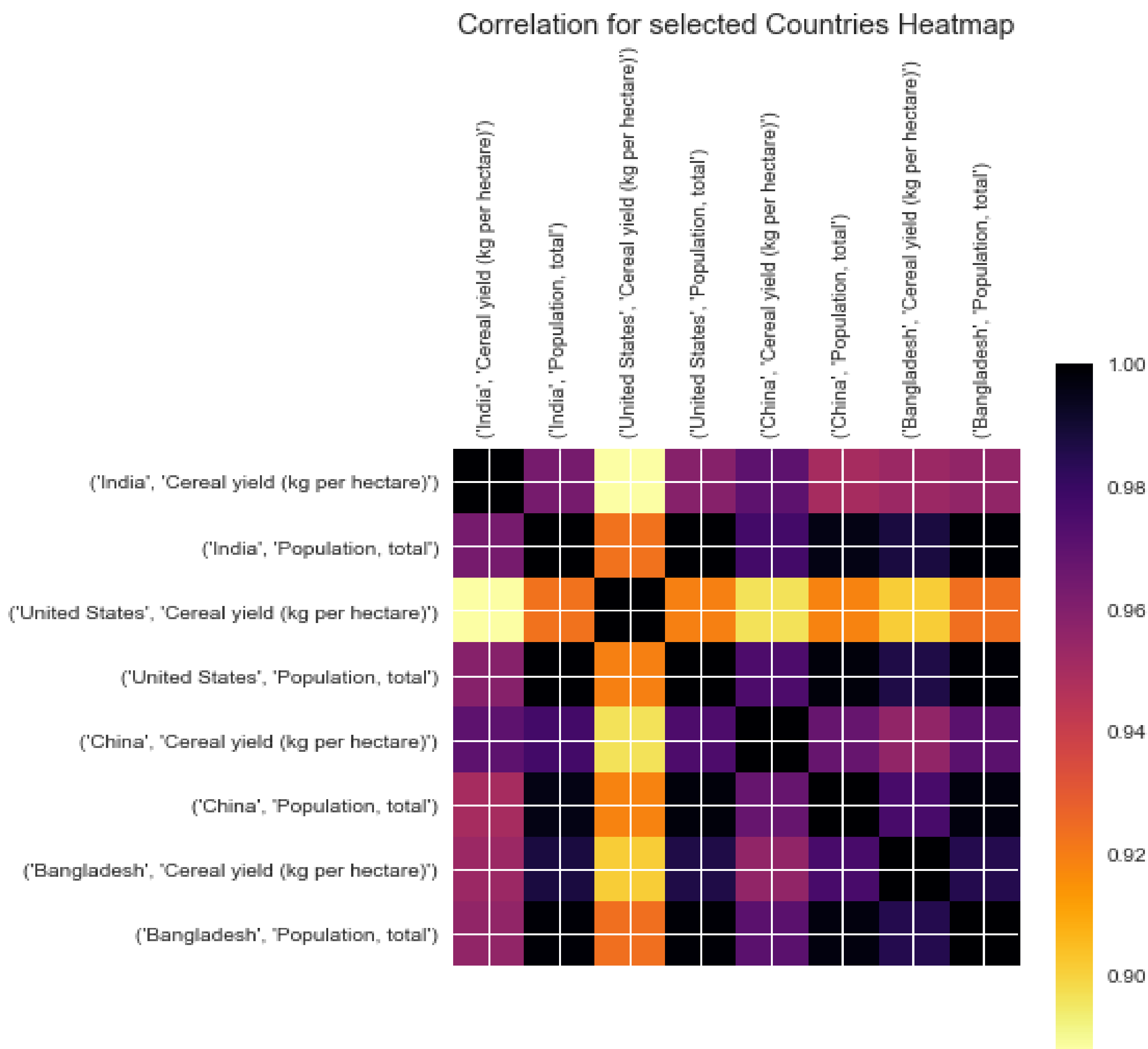
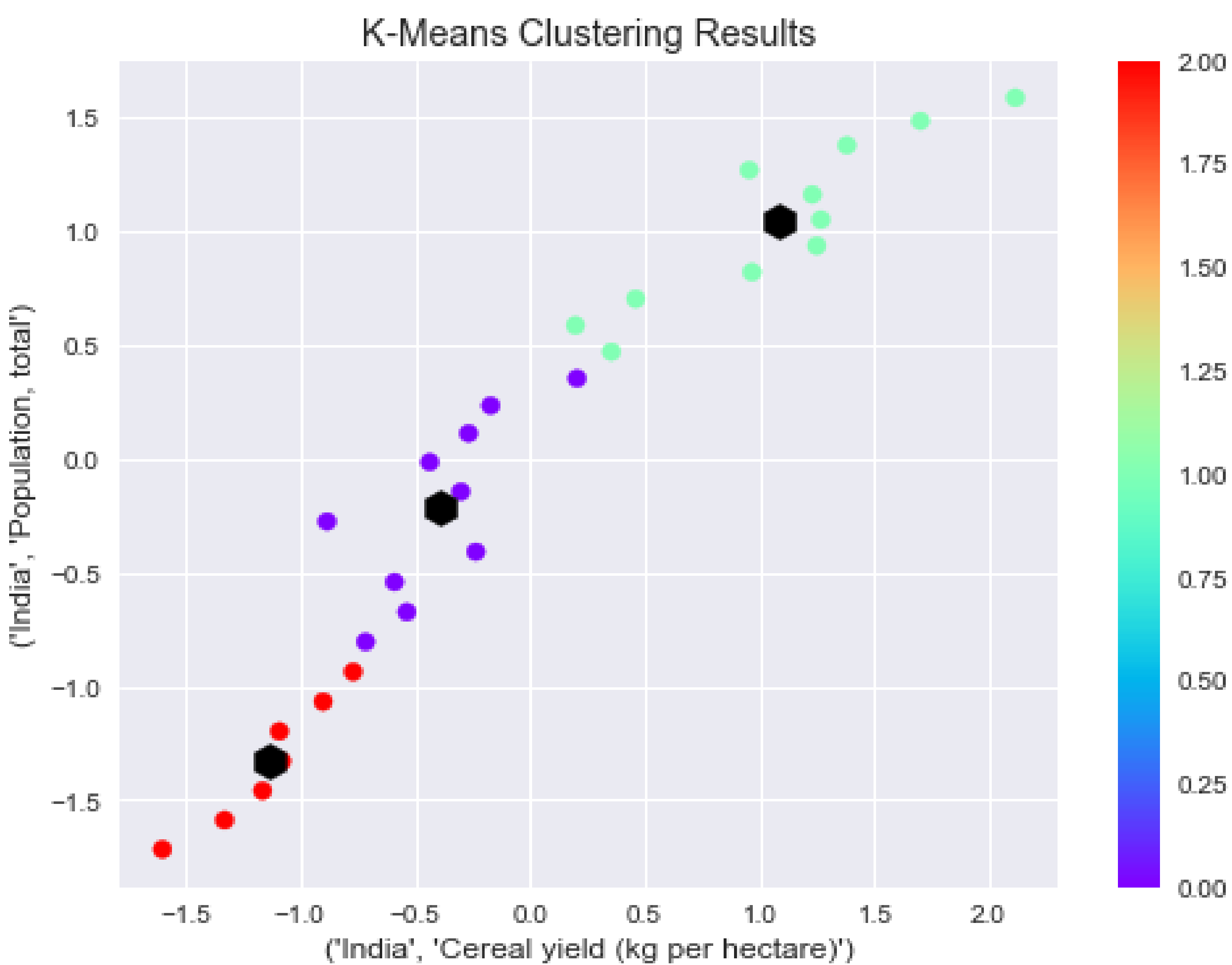
Repo Link: <https://github.com/PuligillaUday/ADS-Assignment-3>

Abstract: The objective of this project is to apply clustering and curve fitting techniques on World Bank data to identify interesting clusters of countries and to develop simple models for predicting future values. The data set consists of the cereal yield and total population indicators for four selected countries (India, United States, China, and Bangladesh) for the years 1991 to 2019. K-means clustering was used to group the countries based on similarities in their normalized data. Curve fitting was used to develop simple models for predicting future values. The results showed interesting patterns in the data and produced models with reasonable confidence intervals.

Introduction: Agricultural production and population growth are critical indicators for monitoring the economic and social development of countries. This project aims to analyse the cereal yield and total population data for India, United States, China, and Bangladesh to identify interesting clusters and develop simple models for predicting future values. The analysis was conducted using clustering and curve fitting techniques on World Bank data for the years 1991 to 2019.



Results: The K-means clustering analysis identified three distinct clusters of countries based on their normalized cereal yield and total population data. Cluster 1 consisted of India and Bangladesh, Cluster 2 consisted of the United States, and Cluster 3 consisted of China. The clustering results suggested that India and Bangladesh had similar patterns in their cereal yield and population growth. The curve fitting analysis produced simple models for each country, which predicted reasonable future values with reasonable confidence intervals.



Methodology: Normalized cereal yield and population data from 4 countries (1991-2019) were clustered using K-means algorithm. Results were visualized and modeled with low-order polynomials, exponential growth, and logistic functions for future predictions with estimated confidence intervals.

Conclusion: Clustering and curve fitting techniques revealed patterns in cereal yield and population data for India, US, China, and Bangladesh. They produced simple models for future prediction, proving useful for data analysis.