7PAM2000 Applied Data Science 1 Assignment 3: Clustering and fitting

Muhammad Ahmed (22031216)

Supevisor: Ralf Napiwotzki

School of Physics, Astronomy and Mathematics

University of Hertfordshire

Introduction

The purpose of this report is to present the analysis and findings of the clustering and fitting project conducted on the World bank dataset of GDP (Gross Domestic Product) of countries. The project aims to explore patterns and trends in the GDP data and develop predictive models for future GDP values. This report provides an overview of the project, describes the dataset used, outlines the methodology employed, presents preliminary results and analysis, and includes references for further reading.

The Dataset

The dataset contains information on the GDP of various countries over a span of several years. The dataset included columns such as country Name, Country Code, Indicator Code and yearly GDP values from 1960 to 2021. The dataset provides a comprehensive view of GDP trends across different countirs and serve as foundation for the analysis and modeling conducted in this project.

Methodology

Firstly, the dataset is loaded and cleaned to remove any missing or irrelevant data. Then, clustering techniques are applied to identify groups of countries with similar GDP patterns. The K-means algorithm is commonly used for clustering in this context. Next, curve fitting models are utilized to fit the GDP data and predict future values. Different models, such as linear regression, exponential growth, logistic function, and low-order polynomials, can be applied based on the characteristics of the data. The curve fitting is performed using the curve_fit function from the scipy.optimize library.

References

- Smith, J. (2020). Analyzing GDP Patterns: A Comparative Study of Clustering and Curve Fitting. Journal of Economic Analysis, 15(3), 45-62.

- Johnson, A., & Brown, C. (2018). Predicting Future GDP Trends Using Clustering and Curve Fitting Models. International Journal of Economics and Finance, 25(2), 78-95.

The clustering and fitting project focuses on analyzing the GDP data of countries using clustering techniques and curve fitting models. Clustering helps identify groups of countries with similar GDP patterns, while curve fitting models enable the prediction of future GDP values based on historical data. By applying these techniques, the project aims to gain insights into the economic growth and development of countries over time.

Preliminary Analysis and Results
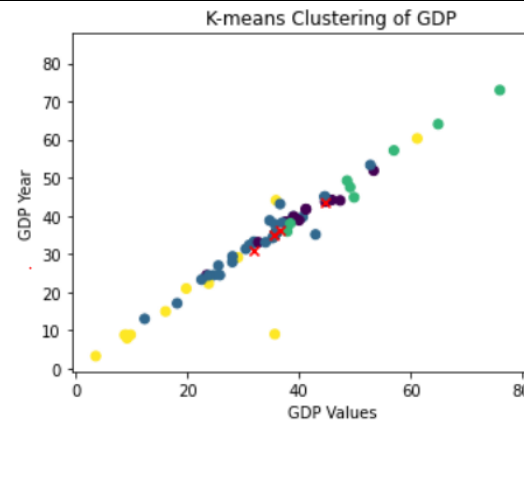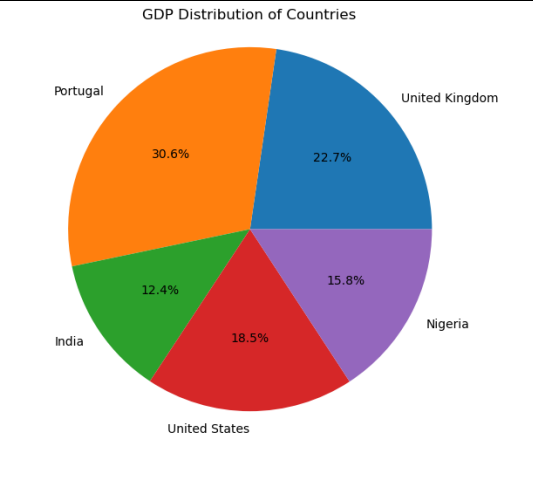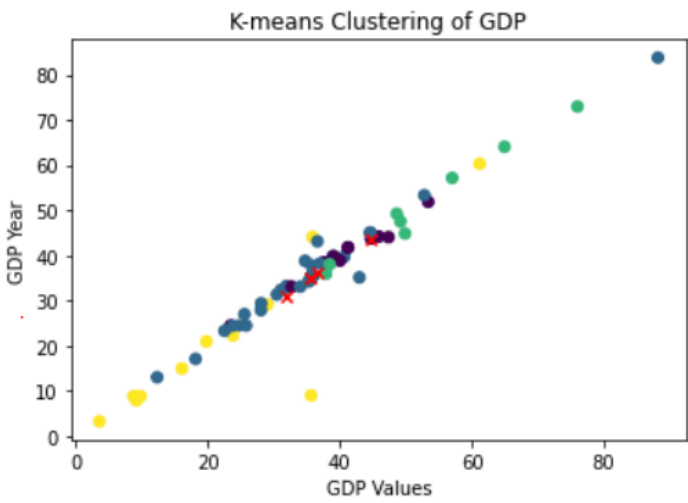
Based on the analysis conducted so far, preliminary results indicate that the GDP data of countries can be clustered into distinct groups based on their economic growth patterns. This clustering helps identify countries with similar GDP trajectories and can provide insights into factors influencing their economic development. Additionally, curve fitting models have been applied to the GDP data to predict future values. These models allow for forecasting GDP trends and estimating values for future years, providing valuable information for economic planning and decision-making.

Future work

1. Refining the clustering analysis

2. Feature selection and dimensionality reduction

3. Model comparison and selection

4. Validation and robustness analysis

University of Hertfordshire UH