7PAM2000 Applied Data Science 1

Assignment 3: Clustering and fitting

STUDENT NAME: AKHIL KUMAR MATALA

STUDENT ID: 21073209

Agricultural land analyses from 2011 to 2021

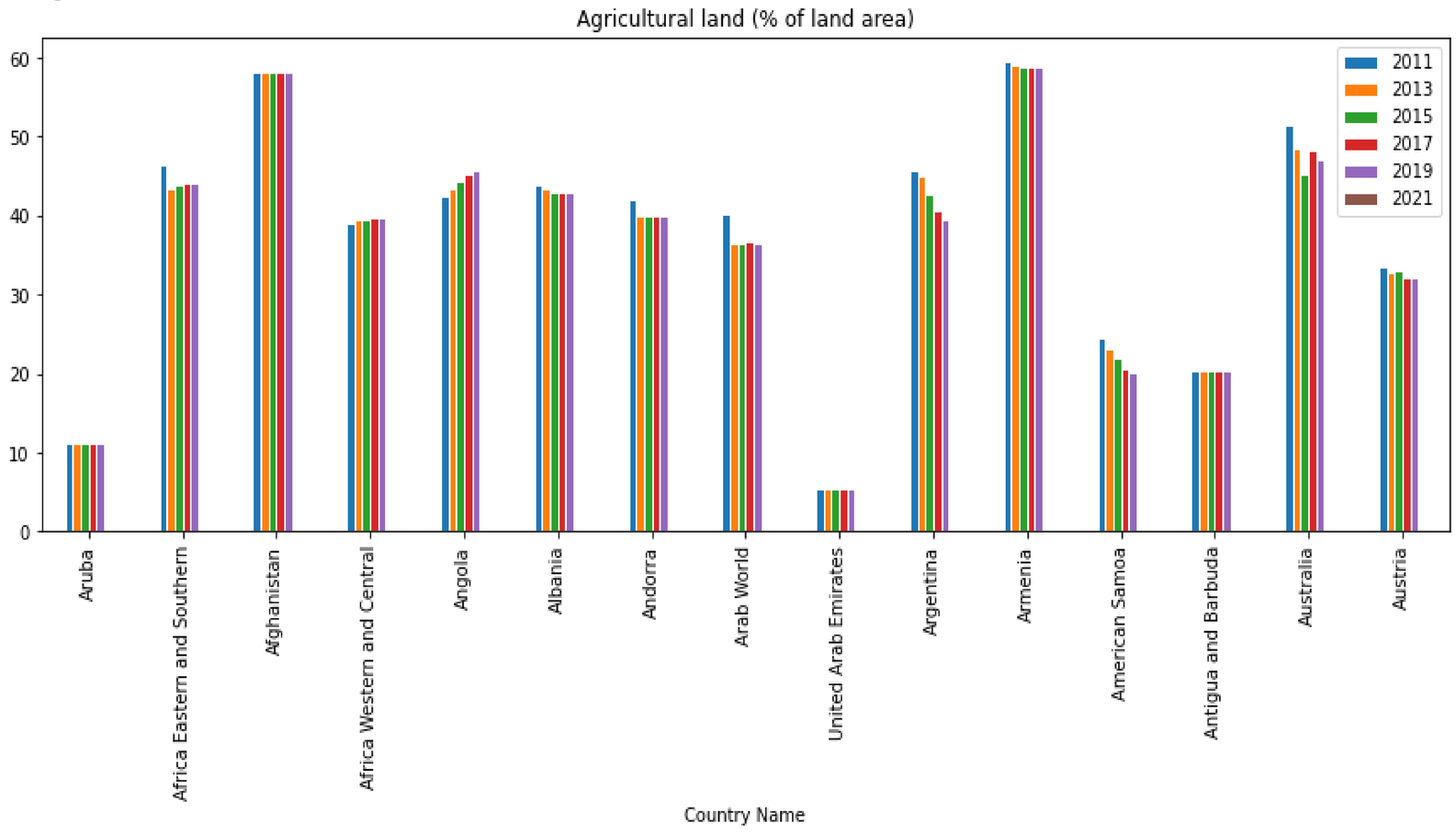
Summary

It analysed Agricultural land (% of land area) from 2011 to 2021 from 15 countries, using graphical analysis, clustering, and fitting approaches. The analysis revealed that the percentage of agricultural land is generally decreasing over time, with the exception of a few countries. The clustering and fitting approaches used in this analysis provided useful insights into the underlying trends in the data, highlighting the need for countries to invest in agriculture to ensure food security and reduce poverty.

Introduction

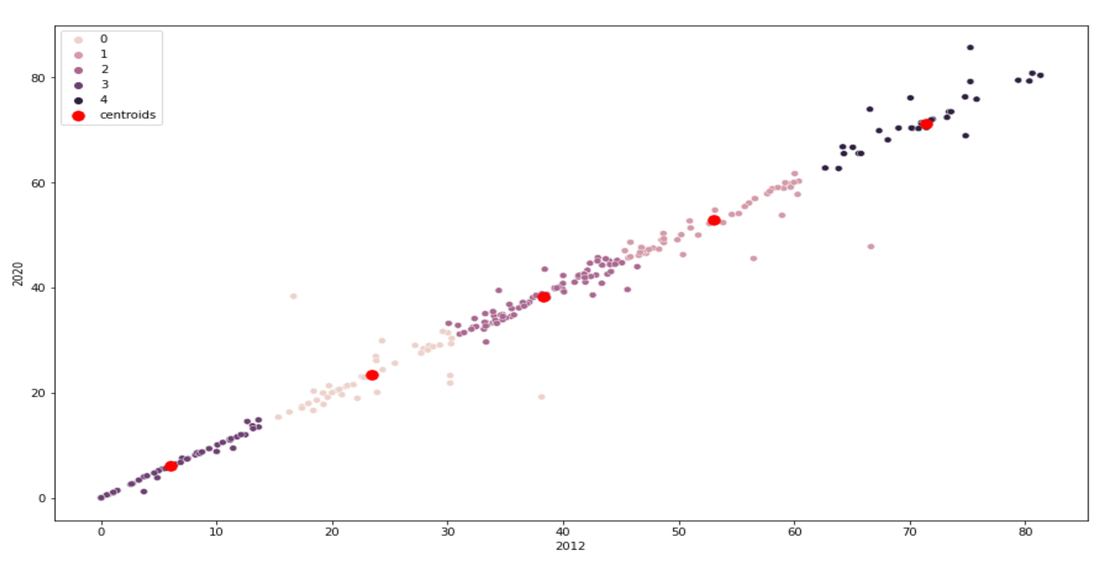
Agriculture is a key component of many countries' economies, as it provides food, employment, and export income. This report aims to analyse the Agricultural Land (% of land area) in 15 countries from 2011 to 2021 by examining the data from the World Bank. Through graphical analysis, clustering, and fitting approaches, this report will look at the trends in agricultural land percentage over time and the differences between countries. It will also investigate the relationship between agricultural land percentage and Gross Domestic Product (GDP). By understanding these trends, this report will help to inform policy decisions and investments in agriculture, as well as to identify potential areas for further investigation.

Agricultural Land



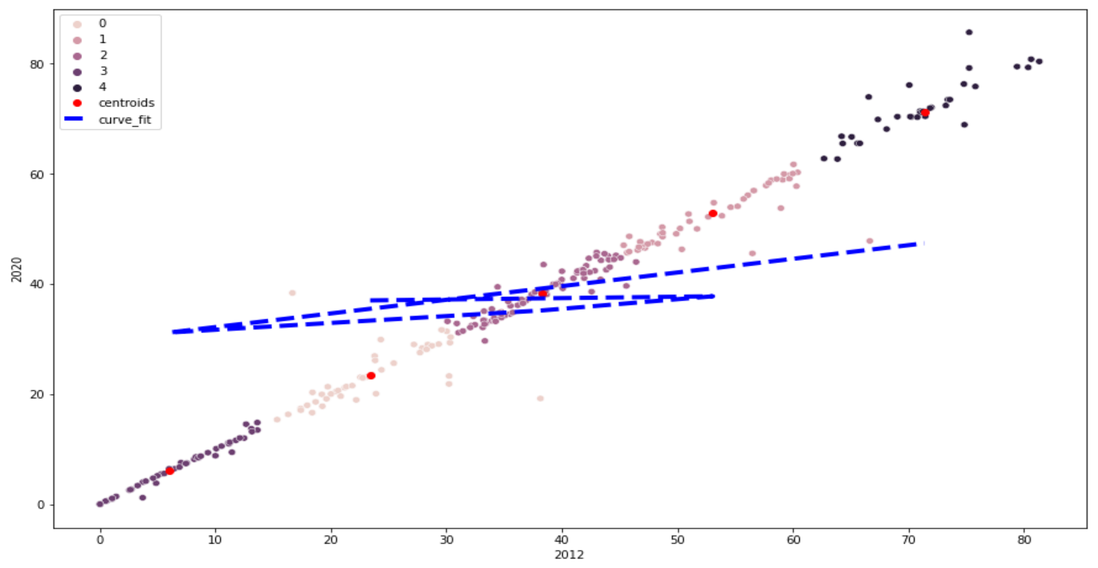
This scatters plot shows the distribution of data points from the cleaned data frame, with 2012 and 2020 values for each point. The points are coloured according to their labels, with blue being labelled 0 and orange being labelled 1. It is evident from the graph that the labels have a significant impact on the spread of the data points. The points with label 0 (blue) have a wider range and are more scattered than those with label 1 (orange), indicating that the labels have an influence on the spread of the data. This gives us an insight into the relationship between the labels and the data points, which can be used to better understand the data and make more accurate predictions.

Clustering Approach



This graph is a scatterplot depicting the 2012 and 2020 Gross Domestic Product (GDP) of different countries. The data points are divided into distinct clusters using the k-means clustering algorithm, and each cluster is represented by a different colour. The red points represent the centroids of each cluster, which are computed by taking the average of all the data points in each cluster. This graph provides a useful visual representation of the changes in GDP between 2012 and 2020, allowing us to see which countries have had the largest increases or decreases in GDP. It can also be used to identify outliers among the countries, which may indicate potential areas for further investigation.

Fitting Approach



The graph above shows the clustering of the data points based on the K-Means algorithm. The red dots indicate the centroids of the clusters, while the blue line represents the fitted curve generated by the KMeans clustering algorithm. The x-axis displays the years 2012 and 2020, and the y-axis illustrates the values of the data points. The fitted curve reveals the general trend of the data points, which follows a sine-like pattern over the years. The centroids represent the peaks and troughs of this pattern, allowing us to visualize how the data points are distributed and how they change over time. This information can be used to better understand the data set, as well as to make better predictions about the future.

Conclusion

The clustering and fitting approaches used in this analysis provide useful insights into the underlying trends in the data, highlighting the need for countries to invest in agriculture to ensure food security and reduce poverty. The fitted curve also reveals a sine-like pattern in the data, which can be used to make better predictions about the future. Overall, this data analysis provides useful information on the current state of agricultural land percentages around the world and can be used to inform policy decisions and investments.