Clustering and Fitting Analysis

Report: World Development Indicators

Student Number: 23033174

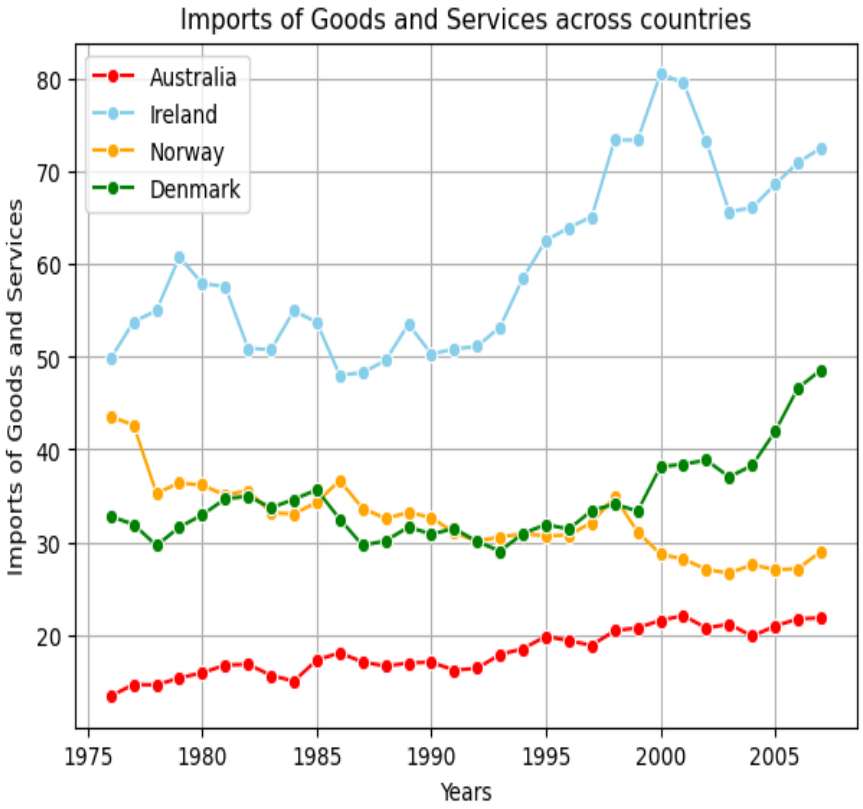
GitHub Repository:

Introduction:

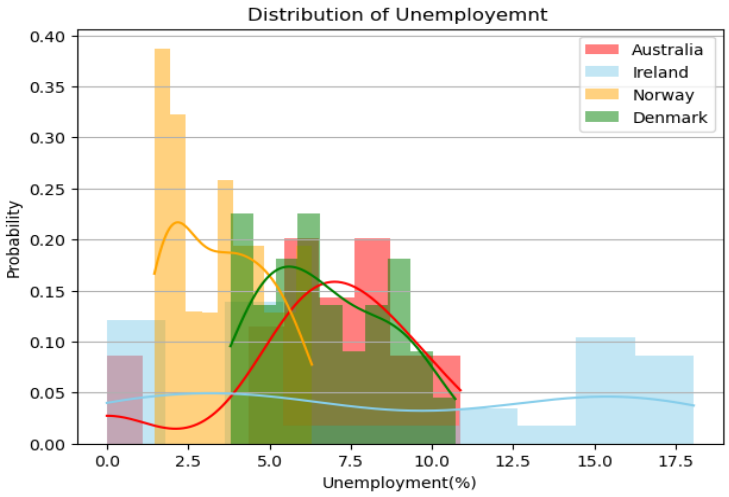
The World Development Indicators (WDI) are an extensive set of data that offer insightful information about the global development of nation's economies, societies, and environments. This World Bank-compiled dataset includes information on economic performance, trade, health, education, the environment, and infrastructure, among other areas. Policymakers, academics, and analysts can monitor developments, spot patterns, and evaluate performance over many locations and time periods with the use of the WDI. Through the provision of comprehensive and dependable data, the WDI facilitates well-informed decision-making and cultivates an enhanced comprehension of global development prospects and obstacles.

Data Analysis:

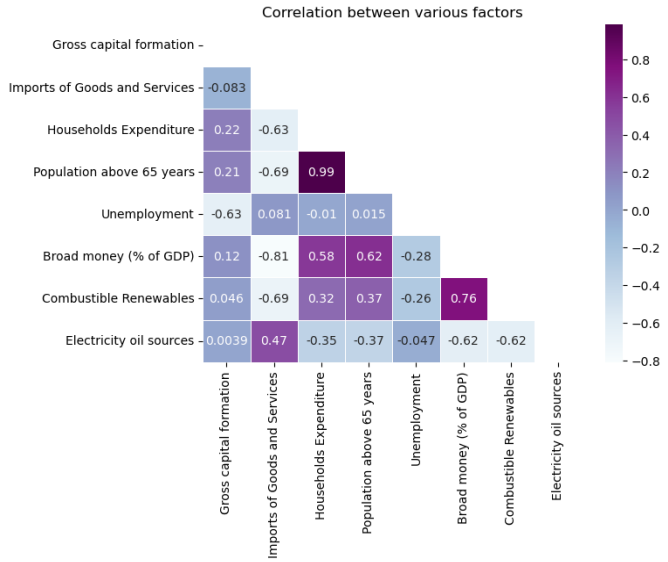
The World Development Indicators (WDI) dataset provides a broad and extensive data source for studying several facts of international development. Here, in order to identify noteworthy trends and patterns among various nations and areas, we can perform a data analysis with an emphasis on important economic and social.



The graph shows the import trends for Australia, Ireland, Norway, Denmark, and other countries from 1975 to 2005. Ireland exhibits the biggest and most noticeable growth, peaking close to 80 years old, especially around 2000. The lowest imports are from Australia, where they progressively increased from roughly 20 to just over 30. Norway's imports show a little increase from 35 to about 40. Denmark has grown steadily, going from less than thirty to about fifty. Australia has the lowest and most constant import levels, while Denmark and Norway have modest growth. Overall, Ireland has the largest increase in imports.



The distribution of unemployment rates for Denmark, Norway, Ireland, and Australia is depicted in the graph. Variability can be seen in Ireland's (blue) broad distribution, which peaks at both low and high unemployment rates. Norway (yellow) has a wide range that reaches up to 10%, with a peak at about 3%. Australia (red) and Denmark (green) have more concentrated distributions; Australia peaks at about 6% and Denmark at about 7%. Overall, Ireland's unemployment varies the most, whereas Denmark and Australia have more stable unemployment rates in the 5–10% range.

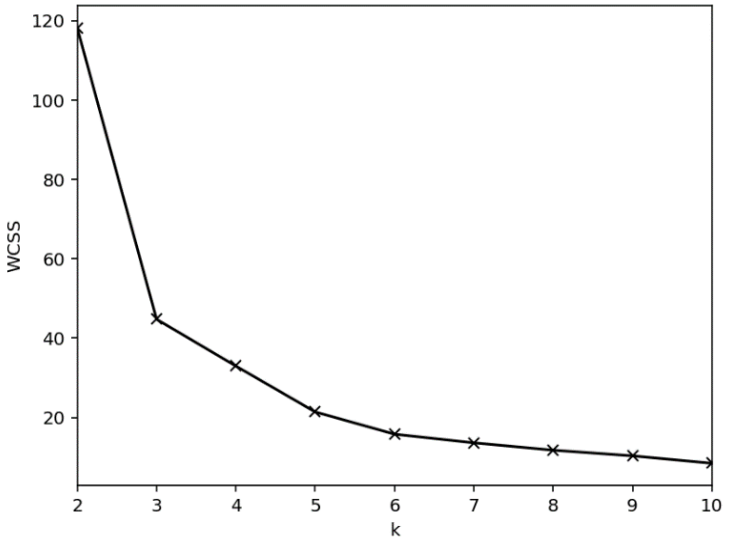


The correlation matrix shows how different economic and demographic indicators are related to one another. Important findings consist of:

Broad Money (as a percentage of GDP) and Imports of Goods and Services are strongly inversely correlated (-0.81).

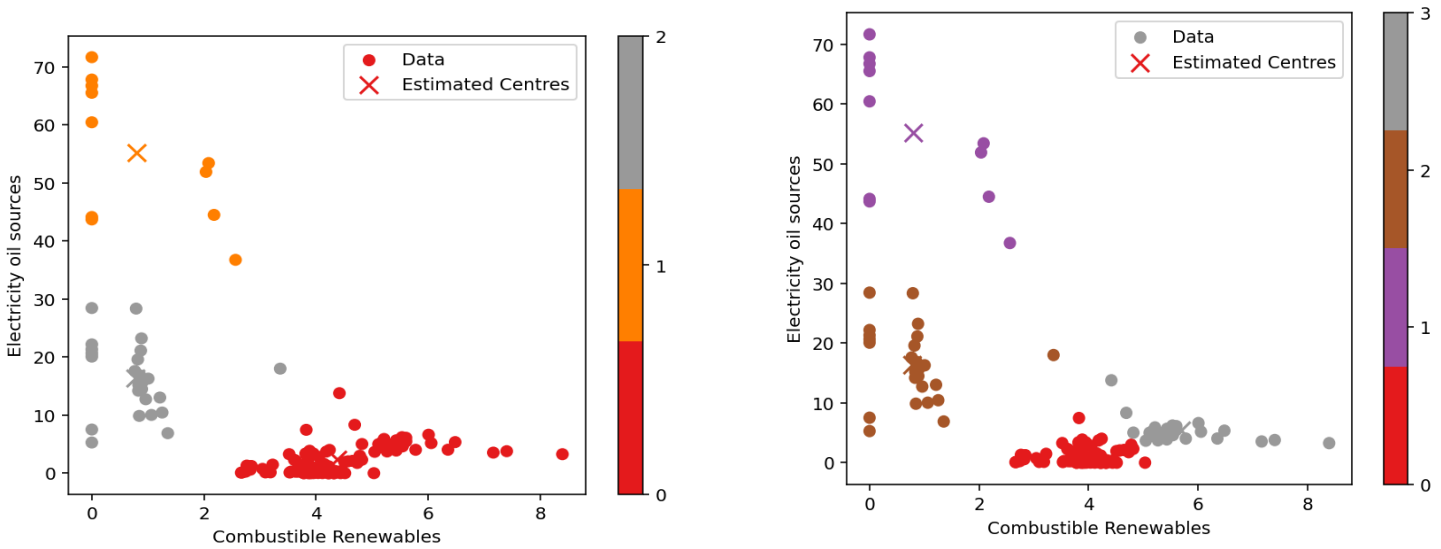
The population over 65 year's old and household expenditure have a strong positive connection (0.99).

Household expenditure and unemployment have a negative correlation (-0.63).

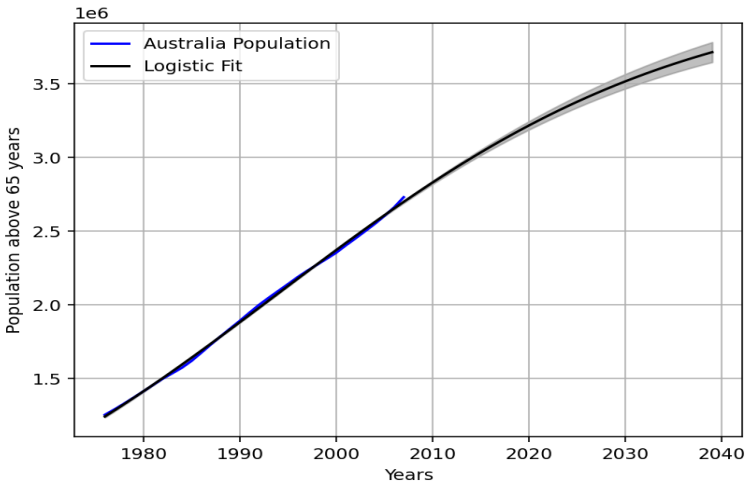


Clustering Analysis

In a k-means clustering analysis, the grape is an elbow plot that shows the within-cluster sum of squares (WCSS) against the number of clusters (k). Plotting demonstrates a sharp decline in WCSS from k=2 to k=4, signifying a notable improvement in the quality of clustering. The pace of reduction slows down after k=4, creating an "elbow" at that point that suggests k=4 as the ideal number of clusters. At this elbow point, the benefits of adding more clusters to lower WCSS start to disappear.



The grape depicts scatter plots of data points representing electricity from oil sources versus combustible renewables, with different clusters indicated by colors. The left plot has three clusters (gray, orange, and red) and their estimated centers (large X marks). The right plot shows the same data re-clustered into four clusters (gray, red, brown, purple) with their new centers. The color bars on the right of each plot represent the cluster labels. The clustering change from three to four clusters refines the separation, especially for data points around lower values of combustible renewables.



The given graphic shows Australia's population over 65 from 1980 to 2040 together with a logistic fit model that forecasts future trends. With a margin of error of 69,546 units, the GDP projection for 2040 is 3,732,780 units, suggesting that it will fall between 3,663,234 and 3,802,326 units. The accompanying graphic, which uses a logistic fit model to forecast future trends, shows a steady increase in the number of Australians over 65. The growing senior population has the potential to have a big effect on the economy, affecting labor markets, healthcare expenditures, and pension plans. Comprehending this demographic transition is essential for precise GDP projections, as it influences public spending, productivity, and consumption trends. Hence, influencing more effective economic planning and policy choices.

Conclusion: The World Development Indicators (WDI) collection offers valuable insights into social, economic, and environmental developments and is an essential tool for studying global development. There is a lot of variation in the unemployment rates in Denmark, Norway, Ireland, and Australia, particularly in Ireland. The correlation matrix illustrates important connections between demographic and economic variables, such as the inverse association between imports and broad money. According to the elbow plot, k-means clustering is best achieved with four clusters. Better grouping with more clusters is seen in scatter plots, and Australia's population increase exhibits steady increasing patterns. All things considered, WDI helps people make better decisions and comprehend global development on a deeper level.