**Clustering Interpretation:**

After iterating through K medoid, K means and hierarchical clustering methods, it is determined that k=2 is the appropriate number for clusters for this dataset and ‘Ward.D2’ is the best linkage methods of hierarchical clustering.

Cluster 1 has 45 countries and cluster 2 has 106 countries. Some countries belonging to cluster 1 are Afghanistan, Ethiopia, Burki Faso, South Africa, South Sudan, Central African Republic, Nigeria, Rwanda, Yemen, Uganda etc. and that belonging to cluster 2 are Russia, Australia, Oman, New Zealand, Denmark, Greece, Israel etc. There is a clear distinction that the cluster 1 countries are underdeveloped or developing countries and that in the cluster 2 are developed countries.

We interpret the characteristics of these clusters by checking their means. Let’s base the interpretation of our clustering analysis on the following three indicators:

1. Economy of countries
2. Quality of Life
3. Health of people.

**Economy of countries:**

The economy of countries is represented by the variables such as GDP, GDP growth, GDP per capita, Health expenditure per capita, Health expenditure - public and Inflation. The table below compares the means of these variables:

|  |  |  |
| --- | --- | --- |
| Variable | Mean of Cluster 1 | Mean of Cluster 2 |
| GDP | 21,480,832,920 | 235,212,552,272 |
| GDP growth | 5.87 | 3.41 |
| GDP per capita | 2,204.46 | 14098.43 |
| Health expenditure per capita | 83.54 | 1145.92 |
| Health expenditure - public | 2.90 | 4.39 |
| Inflation | 9.25 | 4.95 |

From the above table, we can see that the developed countries represented by cluster 2 has higher economy and health expenditures as compared to the underdeveloped countries represented by cluster 1. We can also observe that the inflation in underdeveloped countries is relatively more than the developed countries. One important factor to observe is that our clustering result tells us that the GDP growth in developed countries in lesser than the GDP growth of the under developing countries.

**Quality of life:**

The quality of life is observed by looking at Life expectancy, mortality rate and population growth. The average life expectancy of male and female in cluster 1 countries is 57.1 and 59.7 respectively which is lower as opposed to that in cluster 2 with their averages being 71.63 and 77.38 respectively.

The mortality or the death rate in developed countries for both males and females is very less than the underdeveloped countries with the values being 19.47 and 15.98 for cluster 2 and 92.15 and 81.03 for cluster 1.

The average population growth of the developed group is 0.94 whereas that for undeveloped group is 2.58. Therefore, the population is also under control in the developed countries.

**Health of people:**

We observe better indicators of health of people in developed group than the underdeveloped group. This can be seen by the lesser average birth rate of 16.90 in group 2 than 36.40 in group 1, higher immunization rate against diseases like DPT and measles with average values of 93.75 and 93.37 in group 2 as opposed to lesser immunization rates of 77.88 and 74.89 for underdeveloped group, and lesser women’s share of population living with HIV with an average of 29.88 as opposed to 55.04 in group 1.

**Clustering analysis with 3 clusters:**

To drill down further, we run a clustering again with K = 3 clusters. We obtain 3 clusters of size 45, 61 and 45 respectively.

Some typical countries in the 3 clusters when K = 3 are:

Cluster 1: Afghanistan, Ethiopia, Burki Faso, South Africa, South Sudan, Central African Republic, Nigeria, Rwanda, Yemen, Uganda etc.

Cluster 2: Thailand, Turkey, Vietnam, Nepal, Iran, Mexico, Malaysia, Bhutan, Bangladesh etc.

Cluster 3: Russia, Australia, Oman, New Zealand, Denmark, Greece, Israel, Germany, Sweden, United Kingdom etc

From above, we can observe that 45 underdeveloped countries earlier in cluster 1 (when k =2) remained the same in cluster 1 when k =3. The 106 developed countries earlier represented by cluster 2 were divided into two clusters of 61 and 45 respectively where cluster 2 represented slightly lesser developed and cluster 3 represents the highly developed / rich countries / superpowers.

We can observe similar trends in the economy, quality of life and people’s health in cluster 2 and cluster 3 (That is lesser developed and fully developed countries) when K = 3 as we observed earlier with K = 2. The highly developed countries of group 3 have higher average GDP, Per capita GDP, health expenditure per capita and lower inflation than the lesser developed countries of group 2. This trend can also be seen from the box plot below:

Diagram

Description automatically generated with medium confidence

Industrialization / Co2 Emission:

The development of a country is majorly governed by the industries thriving in that country. On one hand, industrialization brings about employment to the citizens of a country, they hugely contribute to the income, economy and GDP of a country, the flip side of more industrialization is their adverse biproducts that negatively affect the environment overall. Co2 emission and checking on the same becomes of interest while we are looking into the characteristics of the country.

In our analysis, when there were 2 clusters that is k=2, the average co2 emission by the developed countries was far greater than that of the underdeveloped group with their averages being 4.96 as opposed to 0.71 MT per capita respectively. When these countries are divided into three clusters, we can see a similar trend of relatively higher co2 emission as the countries are developing that is the most developed countries contribute to the high co2 emission to the environment.

Therefore, this factor becomes of interest. As the world bank on one side is helping countries towards development, it becomes necessary to curb the adverse effects of development, like the co2 emission to the environment.

Conclusions and Recommendations:

1. We observed the adverse effect of industrialization and a country’s development on the environmental sustainability. Efforts shall be undertaken by the world bank to check and curb the high CO2 emission from developed countries alongside the other goals.