A Project Report on,

# ‘Study On Different Country’s Development On various factors’

By

Mr. Ashish Subhash Parpolkar

## Contents

# Abstract

In this paper the dataset covers 167 countries and nine socio-economic and health factors, collected from the World Bank. The data pre-processing ensures completeness, consistency, and improved model interpretability. Principal Component Analysis (PCA) serves as an unsupervised pre-processing technique for dimensionality reduction, revealing critical insights into variable relationships. Clustering, assessed through the Hopkins score, K-Means, and Hierarchical methods, aims to categorize countries effectively. The final clusters, derived through K-Means clustering, identify countries in need of aid, focusing on child mortality, income, and GDP per capita. The analysis successfully identifies nine countries requiring immediate attention, namely Afghanistan, Comoros, Eritrea, Gambia, Liberia, Madagascar, Rwanda, Tanzania, and Uganda.


Keywords: country clustering, K-means, Hierarchical methods, PCA


# Introduction:

Attaining the Sustainable Development Goals requires building productive capacity and transforming economies by shifting resources to more productive and sustainable sectors and enhancing their competitiveness. Investment, trade and technology are important channels for achieving economic diversification and structural transformation. Trade and trade-related investment, combined with technology upgrading, could enable countries to improve productivity, develop productive capacities and climb regional and global value chains. To fully reap their associated benefits for development, which are not automatic nor evenly spread across countries, foreign direct investment (FDI) inflows should be directed towards activities and projects that lead to enhanced economic and social development, and do not affect the environment negatively.

Developing countries continue to face significant challenges in the formulation of development-oriented trade policy frameworks that are best fit to their national circumstances and development needs. Effectively done, trade policy can drive progress towards the Sustainable Development Goals. Linking trade policy and productive capacity and structural transformation is thus crucial for the achievement of SDGs (Sustainable Development Goals). Governments in developing countries need to take swift action to counter the infection and its socio-economic consequences, but pre-existing structural challenges risk hampering the crisis response, both in the short and long term. GIVE a similar idea of this above information in precise manner.

The countries face a range of development challenges,including economic disparities, social inequality, inadequate healthcare, environmental degradation, and limited access to quality education. These obstacles impede progress, perpetuating a cycle of poverty and hindering the country's potential for growth. To overcome these challenges, the HELP Foundation recognizes the importance of data-driven decision- making, which is where PCA and clustering play a vital role.

The HELP Foundation, founded in 2010, is a global initiative committed to addressing diverse challenges impeding progress in various countries. With a visionary approach, the foundation, guided by a team of dedicated experts, employs data-driven analysis, community engagement, and strategic partnerships to foster positive change.

## Literature Review:

Sethi's et al (1971) played a significant role in examining international marketing data, incorporating information on political conditions, trade, transportation, communications, and consumption.

Mishra et al. (2017) provide a foundational understanding of Principal Component Analysis (PCA) as an unsupervised pre-processing technique for dimensionality reduction in multivariate datasets.

Mahmood et al. (2021) contribute to the literature by proposing a hybrid clustering technique for country categorization, specifically for the HELP International foundation.

A critical aspect of achieving SDGs involves formulating development-oriented trade policy frameworks. Brass et al. (2018) highlight the challenges faced by developing countries in crafting trade policies tailored to their national circumstances.

## Objectives:

1] To categorize the countries using socio-economic and health factors that determine the overall development of the country.

2] To identify the ones which are in the direct need of aid so that immediate help can be given to them through the funding received by the NGO.

3] To understand the most important factors or attributes that affect the development of the countries so that we can primarily focus on improving them.

## Data:

This work focusses on 167 different and 9 socio economic and health factors that affect the country's developments. The dataset has been collected from the website of www.databank.worldbank. It enables us to get a clear idea on how each of the different factors affect the country's development and social status. The target is to be able to group the countries on the basis of the similarities between them and identify the ones that are in dire need of help. Here is a snippet of the first 5 countries along with the corresponding factors that describe their developments.

Again, on checking, it is found that none of the columns have inconsistent datatype, hence no transformation or conversion is required. To prepare the data for modelling, it has been segregated into categorical (country) and numerical (Child mortality rate, exports, health, imports, income, inflation rate, life expectancy, total fertility rate and GDPP) segments

# Research Methodology:

- Data pre-processing:

Before we can directly proceed to modelling, we need to engage in some data pre-processing. Data pre-processing is extremely crucial before moving on to the analysis for various different reasons:

- Data becomes completer and more trustable
- Complexity of the model can be reduced
- Interpretability and the model's performance and accuracy can be better

- Data Cleaning:

Some cleansing checks need to be done so that during modelling we can use the correct and complete data. If there is presence of any null values in the data, then it might lead to a loss of variation and also be a cause of bias during modelling.

However, on checking, it is found that none of the columns have any null values, hence no imputation is required.

- Data Reduction:

For data reduction, PCA has been used and the working has been explained.

## Principal Component Analysis (PCA):

PCA is an unsupervised pre-processing technique that is used for dimensionality reduction if we have large number of variables in our data. Reducing the dimension makes the computation and analysis easier and enables us to get a much clearer model and output.

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

## Goals of PCA:

The goals of PCA are to

1. extract the most important information from the data table.
2. compress the size of the data set by keeping only this important information.
3. simplify the description of the data set.

**Clustering:**

**Hopkins Score:-**

The Hopkins statistic, also known as the Hopkins score or the Hopkins coefficient is a measure used to assess the cluster tendency or the degree of the ability to form clusters of a given dataset. The Hopkins statistic measures the extent to which a dataset can be considered as being generated by a random process, compared to a clustered process. Therefore, it's always important to complement the Hopkins score with other clustering evaluation metrics and techniques.

**K-Means Clustering: (NON- Hierarchical)**

The algorithm seeks to partition a set of n data points into k clusters (where k is a pre-specified number) in which each point belongs to the cluster with the nearest mean, or centroid.

The K-means clustering algorithm follows these steps:

- Initialization: Choose k initial centroids randomly or based on some heuristic, where k is the number of clusters.
- Assignment: Assign each data point to the nearest centroid, based on the Euclidean distancebetween the point and the centroid.
- Update: Re-calculate the centroid for each cluster based on the mean of the points assigned to that cluster.
- Repeat: Repeat steps 2 and 3 until the centroids no longer move or a maximum number ofiterations is reached.
- Termination: The algorithm terminates when the centroids no longer move or the maximum number of iterations is reached.

The goal of the K-means algorithm is to minimize the sum of the squared distances between each data point and its assigned centroid, known as the Within-Cluster Sum of Squares (WCSS). The WCSS is used as a measure of how well the data points are clustered.

**Hierarchical Clustering:**

Hierarchical clustering is a method of cluster analysis in data mining that creates a hierarchical representation of the clusters in a dataset. The method starts by treating ach data point as a separate cluster and then iteratively combines the closest clusters until a stopping criterion is reached. The result of hierarchical clustering is a tree-like structure, called a dendrogram, which illustrates the hierarchical relationships among the clusters.

For performing hierarchical clustering, **agglomerative** hierarchical clustering is chosen.

Algorithm of Hierarchical Clustering:

1) Starts with N clusters each containing single entity and n x n symmetric matrix of distance (or similarities) $D=\{d_{ik}\}$
2) Search the distance matrix for the nearest (most similar) pair of cluster. Let the distance between most similar clusters u & v be $d_{uv}$
3) Merge clusters u & v labelled the newly formed cluster (uv) update the entries in the distance matrix by
    I) Deleting the rows & columns correspond the cluster u & v
    II) Adding rows and columns giving the distance between (uv) and the remaining cluster.
4) Repeat step 2 and 3 total number of (N-1) times.
5) Record the entities of cluster that are merged and the levels (distance and similarities) at which merge takes place.

- **Single linkage:**
  In statistics, single-linkage clustering is one of several methods of hierarchical clustering. It is based on grouping clusters in bottom-up fashion (agglomerative clustering), at each step combining two clusters that contain the closest pair of elements not yet belonging to the same cluster as each other.
  Single linkage can be computed by
  $d_{(uv)w} = \min(d_{(uw),}d_{(vw)})$
- **Complete Linkage:**
  Complete-linkage clustering is one of several methods of agglomerative hierarchical clustering. At the beginning of the process, each element is in a cluster of its own. The clusters are then sequentially combined into larger clusters until all elements end up being in the same cluster. The method is also known as **farthest neighbour clustering**.
  complete linkage can be computed by
  $d_{(uv)w} = \max(d_{(uw),}d_{(vw)})$

# Exploratory Data Analysis:

Data Description:

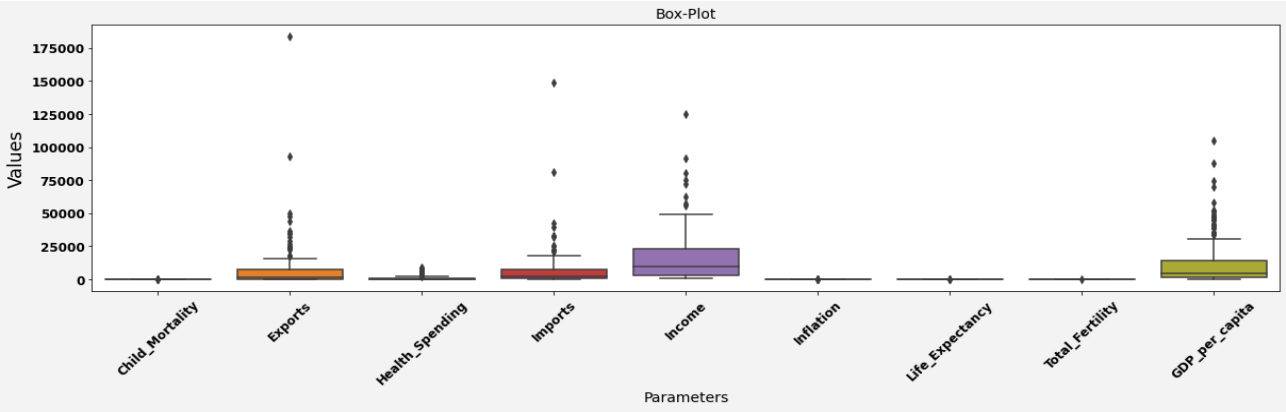
The data description is shown as follows. Apart from the count, minimum and maximum value, statistical measures like the mean, standard deviation, median, 1st and 3rd quartile values for each of the factors are also checked for better understanding of the data.

| | child_mort | exports | health | imports | income | inflation | life_expec | total_fer | gdpp |
|---|---|---|---|---|---|---|---|---|---|
| count | 167.0000 | 167.0000 | 167.0000 | 167.0000 | 167.0000 | 167.0000 | 167.0000 | 167.0000 | 167.0000 |
| mean | 38.2701 | 41.1090 | 6.8157 | 46.8902 | 17144.6886 | 7.7819 | 70.5557 | 2.9480 | 12964.1557 |
| std | 40.3289 | 27.4120 | 2.7468 | 24.2095 | 19278.0677 | 10.5706 | 8.8932 | 1.5138 | 18328.7048 |
| min | 2.6000 | 0.1100 | 1.8100 | 0.0700 | 609.0000 | -4.2100 | 32.1000 | 1.1500 | 231.0000 |
| 25% | 8.2500 | 23.8000 | 4.9200 | 30.2000 | 3355.0000 | 1.8100 | 65.3000 | 1.7950 | 1330.0000 |
| 50% | 19.3000 | 35.0000 | 6.3200 | 43.3000 | 9960.0000 | 5.3900 | 73.1000 | 2.4100 | 4660.0000 |
| 75% | 62.1000 | 51.3500 | 8.6000 | 58.7500 | 22800.0000 | 10.7500 | 76.8000 | 3.8800 | 14050.0000 |
| max | 208.0000 | 200.0000 | 17.9000 | 174.0000 | 125000.0000 | 104.0000 | 82.8000 | 7.4900 | 105000.0000 |

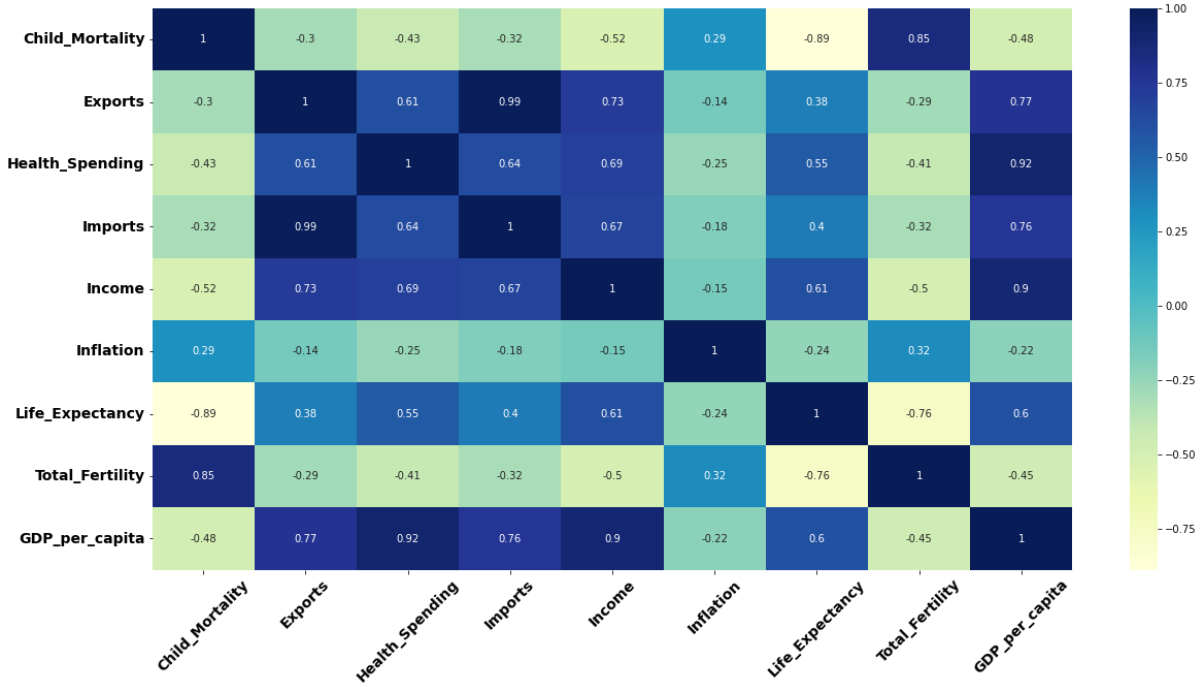
When we go for the data cleaning process, However on checking, it is found that none of the columns have any null values, hence no imputation is required.

## Outlier Analysis:

Outlier analysis is a process of identifying and handling the outliers in the dataset. One commonly used statistical test for outlier analysis is the Interquartile Range (IQR) test. The IQR is a measure of the spread of a dataset that represents the range between the first and third quartiles of the data. The IQR can be used to identify outliers by calculating the lower and upper bounds of the dataset.




From the box-plot graph, we observe that the outliers observed in the case of exports, health, imports, income, and gdpp,




1. Exports, imports, health spending, income are highly correlated with GDP per capita.
2. These numerical variables are also observed highly correlated with each other.

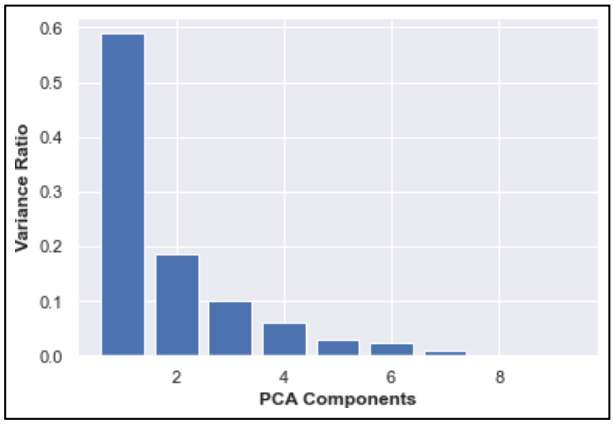
The imports, exports and health spending have been deduced from percentage values to actual values of their GDPP per capita as, the percentage values don't give a clear picture of the country. Since PCA is going to be used, hence rescaling the attributes is needed.

There are multiple ways to perform scaling, however, here standardization is used. It is important for all the numerical attributes to follow the same scale or range without changing the differences of range between them or losing any information. If 2 or more attributes are recorded in different scales, then it becomes difficult to interpret them as a result of which the model accuracy is compromised and misleading results are obtained due to the large differences in range.

# Analysis and Interpretation

### 1.Principal Component Analysis:

PCA has been performed on the standardized data. The principal components have been found and the variability explained by them has also been noted.

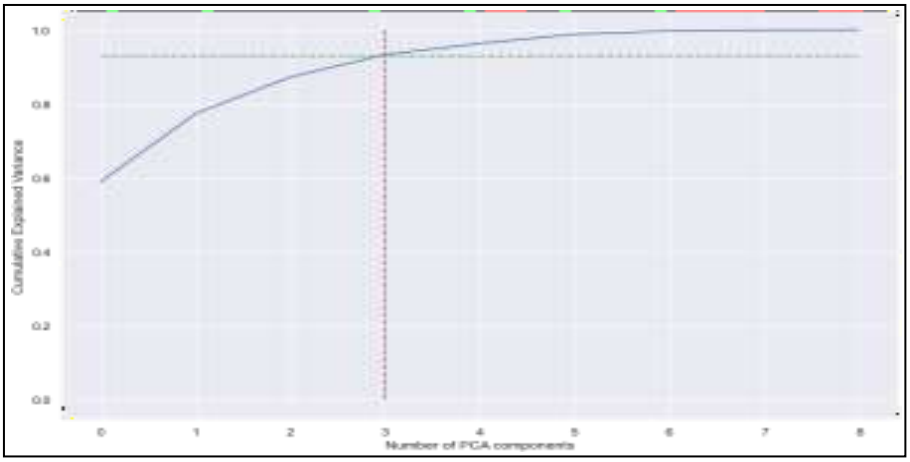



**BAR PLOT** – A bar plot of the variance ratio explained by the principal components vs the principal components has been plotted

## Interpretation:

It can be seen that the 1st principal component explains 60% of the total variation. With the 2nd principal component almost 20% of the total variation of the data can be explained.

**SCREE PLOT** – A scree plot also tells us how much variation each principal component captures from the data. It plots the eigen values, i.e., the variance explained by the principal components against the principal components or factors.



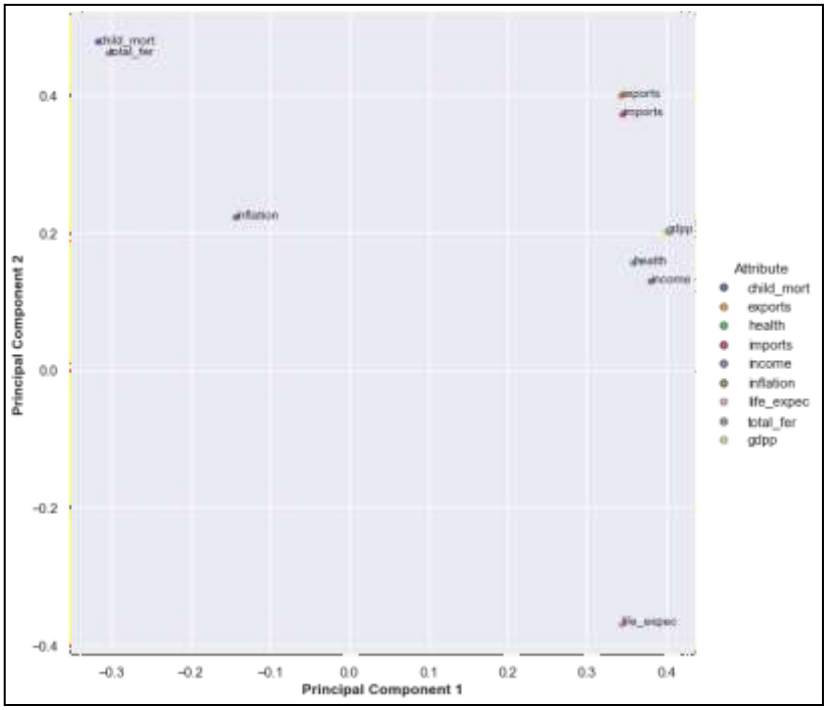

## Interpretation:

It is evident from the Screen plot that more than 90% variance is being explained by the first 3 principal components. Hence, we will proceed with only these 3 components.

1.2 Checking which attributes are well explained by the principal components.

First, all the attributes are plotted with principal component 1 (PC1) and principal component2 (PC2) for better visualization.

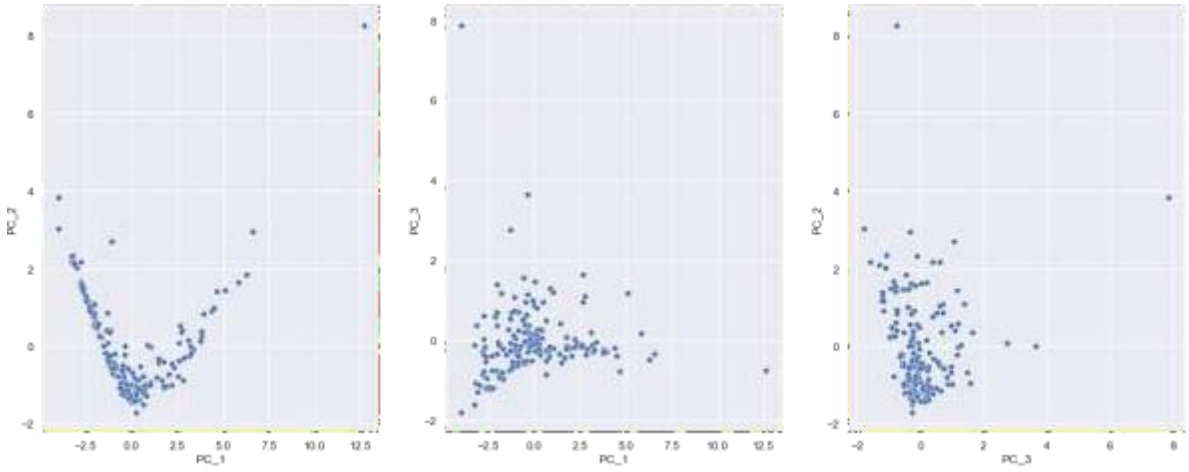
**Interpretation**:




- Life Expectancy, income and gdpp and health are very well explained by PC1.
- Imports and Exports are well explained by both the components PC1 and PC2.
- Child mortality and total Fertility are well explained by PC2.
- Inflation is neither explained by PC1 nor with PC2

Similarly, we can see from the plot that the attribute 'inflation' is well explained by PC3.

**1.3 SCATTER PLOT –** A scatterplot is plotted to check the spread of the data across theprincipal components.



## Interpretation:

Since 90% variance is explained by 3 principal components, let's build the final dataset using those 3 components.

| | country | PC_1 | PC_2 | PC_3 |
|---|---|---|---|---|
| 0 | Afghanistan | -2.6374 | 1.4690 | -0.5414 |
| 1 | Albania | -0.0223 | -1.4319 | -0.0207 |
| 2 | Algeria | -0.4576 | -0.6733 | 0.9619 |
| 3 | Angola | -2.7245 | 2.1746 | 0.6067 |
| 4 | Antigua and Barbuda | 0.6499 | -1.0244 | -0.2501 |

## 2.0 Clustering:

Hopkins Score: - The Hopkins score that we get is 0.83, which is good for Clustering.

**2.1 K-Means Clustering: –**

**Elbow method** – we seen the elbow curve then we can easily proceed with either 4 or 5 clusters

The **SILHOUTTE SCORE** for 2,3,4,5 and 6 clusters are listed below –

```
For n_clusters=2, the silhouette score is 0.48734222042380726
For n_clusters=3, the silhouette score is 0.4639787559241272
For n_clusters=4, the silhouette score is 0.3987406059038731
For n_clusters=5, the silhouette score is 0.3617086573187365
For n_clusters=6, the silhouette score is 0.36603889134181417
```

We begin with considering 4 clusters. Each country is assigned to a respective cluster. Here is a snippet of the countries assigned to their respective clusters.

The total number of countries present in each cluster is expressed as –

| | country | PC_1 | PC_2 | PC_3 | Cluster_Id4 |
|---|---|---|---|---|---|
| 0 | Afghanistan | -2.6374 | 1.4690 | -0.5414 | 3 |
| 1 | Algeria | -0.4576 | -0.6733 | 0.9619 | 0 |
| 2 | Antigua and Barbuda | 0.6499 | -1.0244 | -0.2501 | 2 |
| 3 | Armenia | -0.3327 | -1.2745 | 0.1766 | 0 |
| 4 | Australia | 3.1804 | -0.2508 | -0.1169 | 1 |

| Clusters (0 = 1st cluster, 1 = 2nd cluster, 2 = 3rd cluster, 3 = 4th cluster) | Number of countries present in each cluster |
|---|---|
| 0 | 48 |
| 1 | 19 |
| 2 | 29 |
| 3 | 23 |

## Interpretation:

It seems there are good number of countries in each cluster. It can be seen that cluster 0 (1st cluster) has the maximum number of countries present in it whereas cluster 1 (2nd cluster has the least). Cluster 2 and 3 (3rd and 4th cluster) consist of almost similar number of countries.

In an attempt to solve the problem faced during considering 4 clusters, this time 5 clusters areconsidered. Here is a snippet of the countries assigned to their respective clusters. The total number of countries present in each cluster is expressed as –

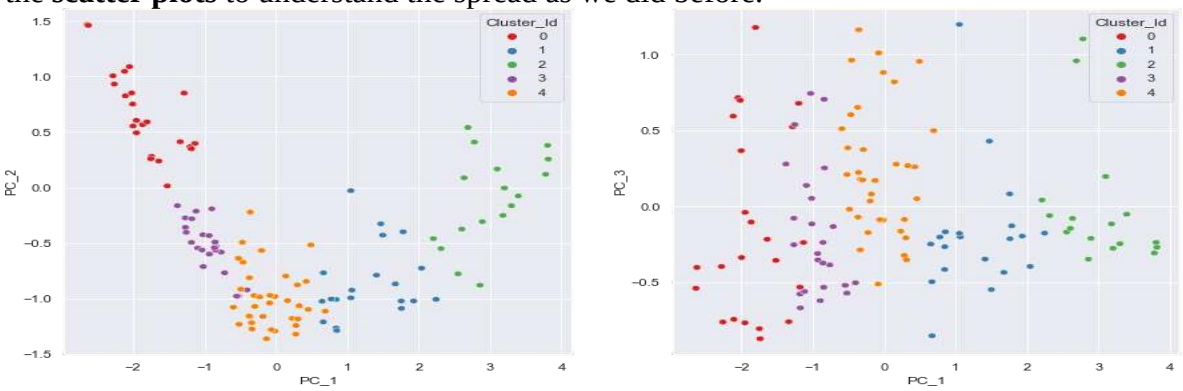
| | country | PC_1 | PC_2 | PC_3 | Cluster_Id |
|---|---|---|---|---|---|
| 0 | Afghanistan | -2.6374 | 1.4690 | -0.5414 | 0 |
| 1 | Algeria | -0.4576 | -0.6733 | 0.9619 | 4 |
| 2 | Antigua and Barbuda | 0.6499 | -1.0244 | -0.2501 | 1 |
| 3 | Armenia | -0.3327 | -1.2745 | 0.1766 | 4 |
| 4 | Australia | 3.1804 | -0.2508 | -0.1169 | 2 |

(a)

| Clusters (0 = 1st cluster, 1 = 2nd cluster, 2 = 3rd cluster, 3 = 4th cluster, 4 = 5th cluster) | Number of countries present in each cluster |
|---|---|
| 0 | 23 |
| 1 | 20 |
| 2 | 17 |
| 3 | 24 |
| 4 | 35 |

Interpretation:

Again, a fairly good number of clusters belong to each of the clusters. Let us plot and check the **scatter plots** to understand the spread as we did before.

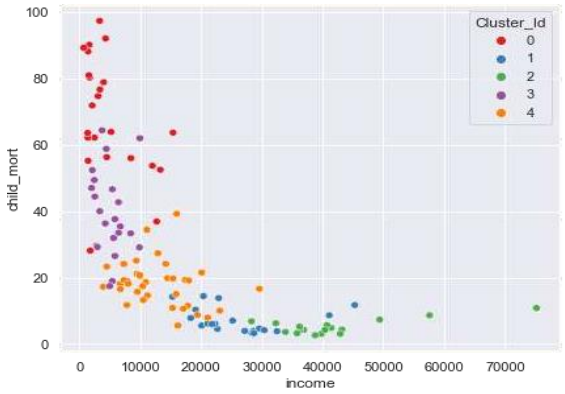


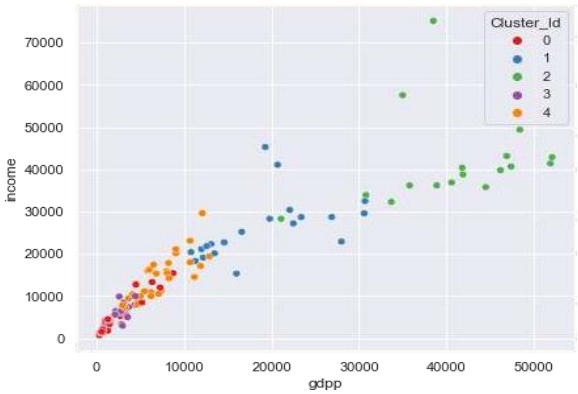

## Interpretation:

Here also we have got the same problem that we had faced with 4 clusters, however in this case, we have got a new segment, so we proceed with K means using 5 clusters.

From the respective average values of the attributes and from the business understanding it was seen that child Mortality, income and gdpp are some of the most important attributes that decide the development of any country. We have also cross checked with Principal components and found that these variables have good score in PCA. Hence, we will proceed with analyzing these 3 components to build clusters.

A **Scatterplot** has been plotted to check the spread of the factors – child mortality, incomeand gdpp, for each of the countries as per their respective clusters.

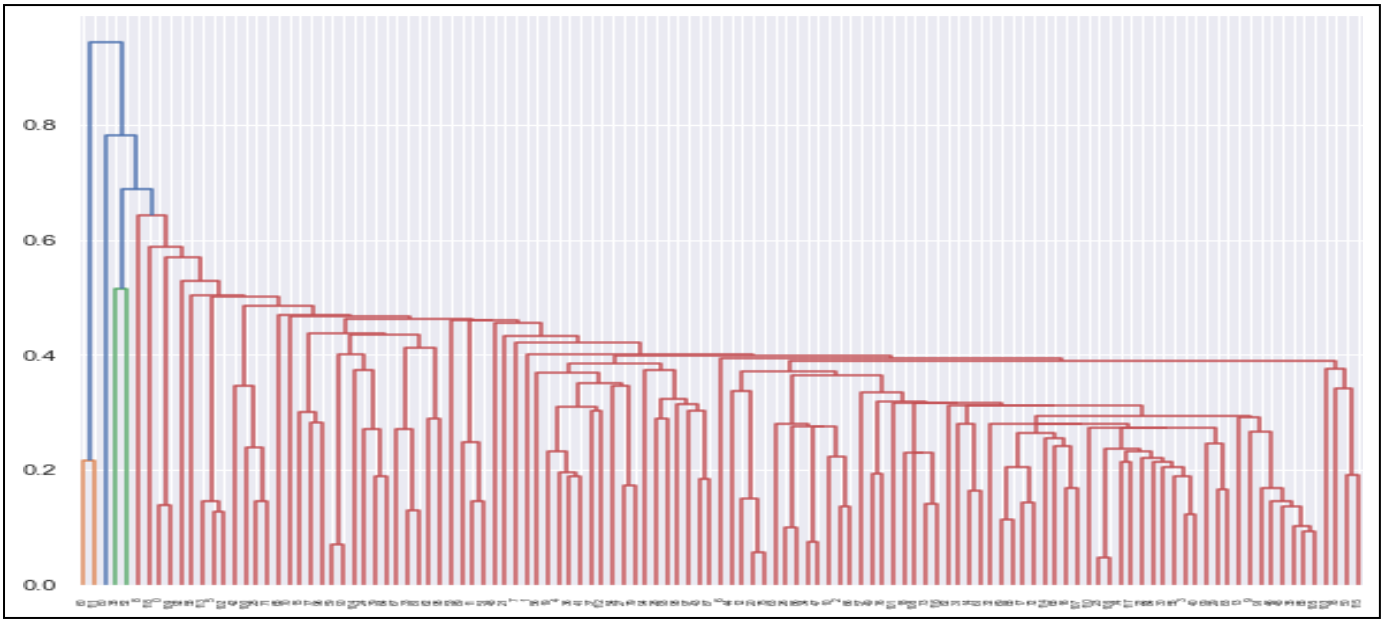


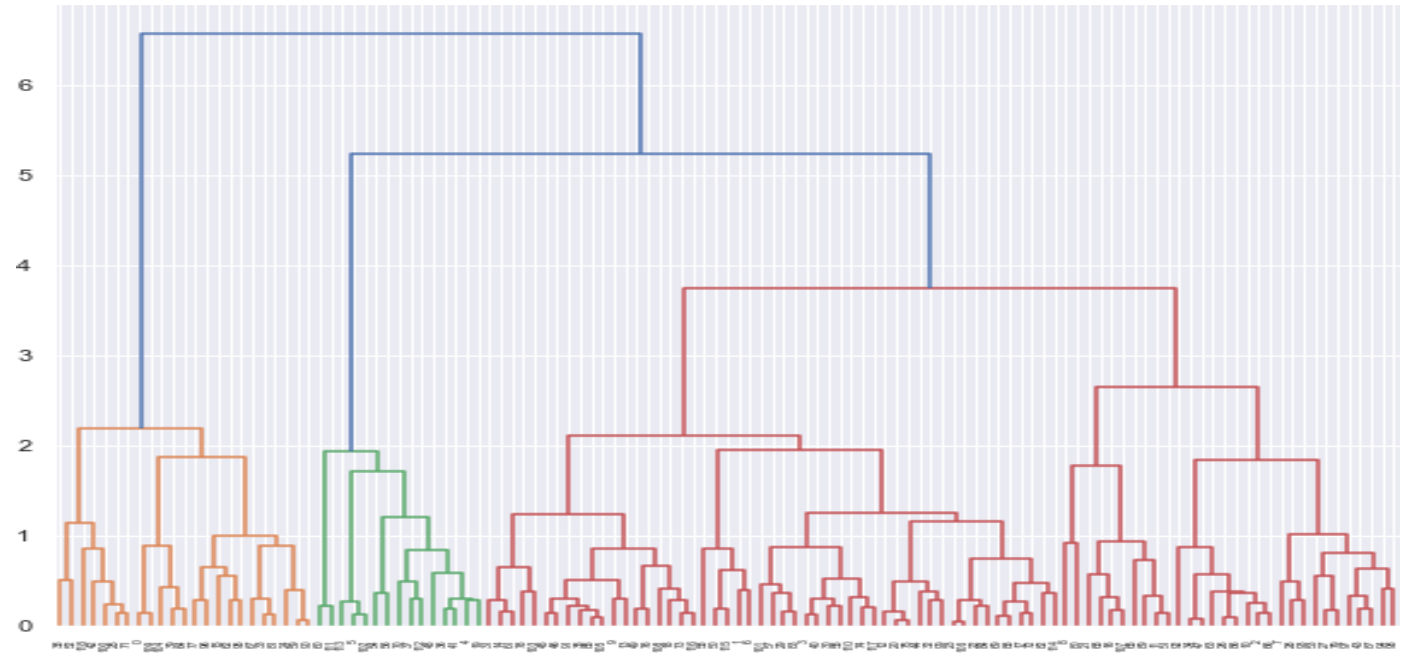

Now let's apply Hierarchical Clustering to see if any better clusters are formed or not.

**2.2 Hierarchical Clustering:-**

For performing hierarchical clustering, **agglomerative** hierarchical clustering is chosen. We begin with **Single linkage** hierarchical clustering. The dendrogram is shown below -




The single linkage does not show a very clear clustering. Hence **complete linkage hierarchical clustering** is considered.



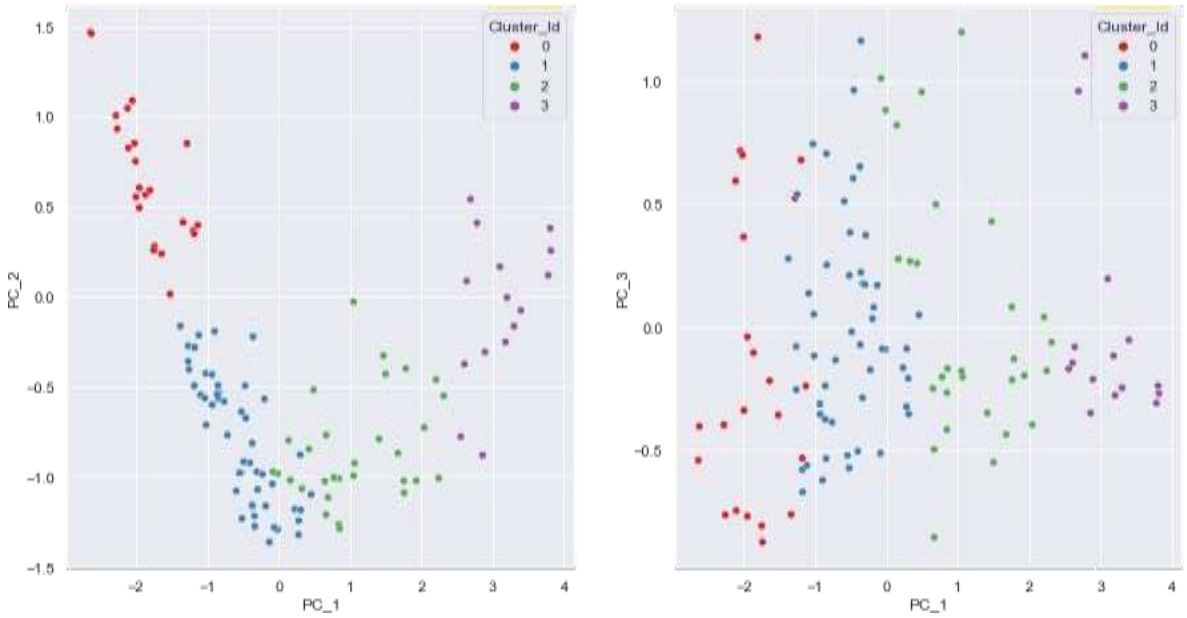
## Interpretation:

From both the dendrograms made, it can be seen that the tree should be cut at a height of approximately 3 to get 4 clusters. Hence, we take 4 clusters and divide each of the countries into their respective clusters.
Further, we proceed to see if any better cluster formation can be achieved from this or not. Here is a snippet of the countries assigned to their respective clusters.

**Scatterplots** are plotted to understand the spread of the country data points in their corresponding clusters, the distance between the country data points belonging to the same cluster as well as the distance between those belonging to different cluster

*(b)*

| | country | PC_1 | PC_2 | PC_3 | Cluster_Id |
|---|---|---|---|---|---|
| **0** | Afghanistan | -2.6374 | 1.4690 | -0.5414 | 0 |
| **1** | Algeria | -0.4576 | -0.6733 | 0.9619 | 1 |
| **2** | Antigua and Barbuda | 0.6499 | -1.0244 | -0.2501 | 2 |
| **3** | Armenia | -0.3327 | -1.2745 | 0.1766 | 1 |
| **4** | Australia | 3.1804 | -0.2508 | -0.1169 | 3 |

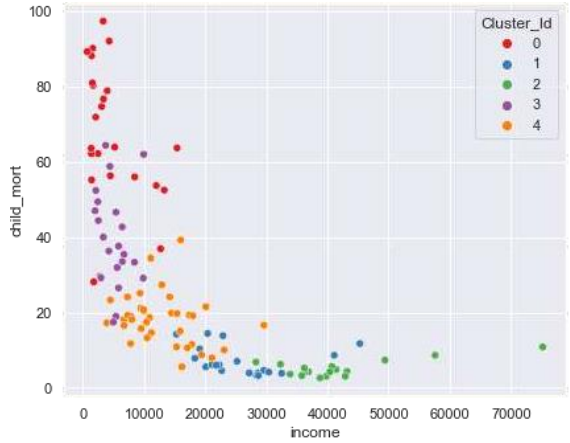


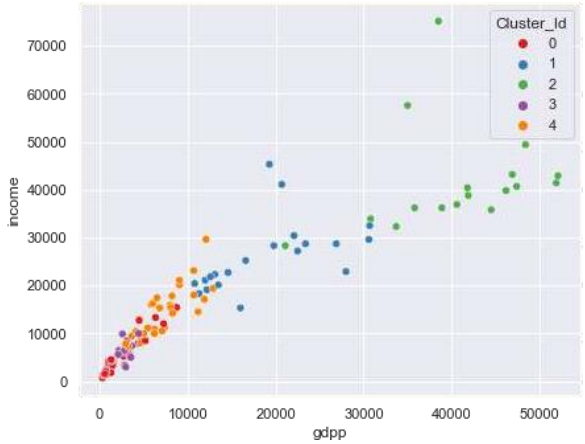

**Interpretation:**

- Cluster 3 (4th cluster) doesn't seem to be too precise from both the plots.
- The spread of the components as per their clusters is quite clear in the first plot butcomparatively less precise in the 2nd plot (PC3 vs PC1)

Again, just like it was done in the case of K- Means clustering, to understand the developments of the countries based on their respective factors or attributes, we take the mean of each of the attributes of as per the clusters to which they belong. A Scatterplot has been plotted to check

the spread of the factors – child mortality, income and gdpp, of the countries as per their respective clusters.




**Interpretation :**

We analysed both K-means and Hierarchical clustering and found that the clusters formed are not identical. So, we will proceed with the clusters formed by K-means Clustering and based on the information provided by the final clusters, we will proceed to deduce the final list of countries which are in need of aid.

## Final Analysis:

Since we are proceeding with the clusters formed by K-Means clustering, therefore the list of countries that are in need of aid are the countries belonging to Cluster 0 (1st cluster) and Cluster 3 (4th cluster).

| Cluster 0 | Cluster 0 | Cluster 3 | Cluster 3 |
|---|---|---|---|
| Afghanistan | Gambia | Bangladesh | Kyrgyz republic |
| Botswana | Ghana | Belize | Micronesia Fed. Sts |
| Comoros | Iraq | Bhutan | Morocco |
| Congo republic | Kenya | Bolivia | Myanmar |
| Eritrea | Lao | Cambodia | Nepal |
| Gabon | Liberia | Cape Verde | Philippines |
| Mauritania | Madagascar | Egypt | Samoa |
| Namibia | Sudan | Fiji | Tajikistan |
| Pakistan | Tanzania | Guatemala | Tonga |
| Rwanda | Uganda | Guyana | Turkmenistan |
| Solomon Islands | Yemen | India | Uzbekistan |
| South Africa | | Indonesia | Vanuatu |

Earlier we had made the inference of child mortality rate, income and gdpp being the 3 most important factors. So, now we need to classify the 47 needy countries as per the 3 attributes. We begin by calculating the average values of the 3 attributes so that we can find out all those countries that have,
child mortality rate > average child mortality rate, and /or income <= average income and/or gdpp <= average gdpp

The countries thus found will be the ones that need the most help from the NGO.
So, first, the average child mortality rate of the 47 countries is found.

- **The average child mortality rate is 52.8574 (~ 53)**

Therefore, all the countries among the 47 countries that have Child mortality rate > Average Child Mortality rate (i.e., 53) are

| | | | |
|---|---|---|---|
| Afghanistan | Eritrea | Ghana | Liberia |
| Comoros | Gabon | Kenya | Madagascar |
| Congo Rep. | Gambia | Lao | Mauritania |
| Namibia | South Africa | Uganda | Myanmar |
| Pakistan | Sudan | Yemen | Turkmenistan |
| Rwanda | Tanzania | India | |

All the above mentioned 23 countries have the average child mortality rate greater than 53. Further, the average income value of the above mentioned 23 countries is found.

- **The average income is 4228.6957 (~ 4229)**

Therefore, all the countries among the 23 countries that have

- Income <= Average Income (i.e., 4229) are

| | | | | |
|---|---|---|---|---|
| Afghanistan | Gambia | Lao | Mauritania | Tanzania |
| Comoros | Ghana | Liberia | Rwanda | Uganda |
| Eritrea | Kenya | Madagascar | Sudan | Myanmar |

All the above mentioned 15 countries have the average income less than or equal to 4229.Further, the average gdpp value of the above mentioned 15 countries is found.

- **The average gdpp value is 803.4000 (~ 803)**

Therefore, all the countries among the 15 countries that have

- gdpp <= Average gdpp (i.e., 803) are

| | | |
|---|---|---|
| Afghanistan | Gambia | Rwanda |
| Comoros | Liberia | Tanzania |
| Eritrea | Madagascar | Uganda |

It is clear that the 9 countries that we are left with surely has **child mortality rate > 53**, and/or **income value <= 4229**, and/or **the gdpp value <=803**.

## Closing Statement:

So, firstly we have used PCA to reduce the variables involved and then performed clustering of the countries based on the principal components that we received. Further we identified 3 of the most important factors like Child Mortality, income and gdpp which play a vital role in deciding the development status of the countries and built the clusters of the countries based on that. Later, based on those clusters, we identified the list of 9 countries which are in dire need of aid. The list of countries is definitely subject to change as it is based on the factors like Number of component Chosen, Number of cluster chosen, Clustering method used, etc. which we have used to build the model. However, from the point of view of this work, we have been successfully identify the countries that need the most help for the betterment of their development through the fund received by the NGO Foundation. Thus, we have been ableto successfully meet and fulfil the objective and aim of this project.

# Major Findings:

1] Child Mortality, income and gdpp are some of the most important attributes that decide the development of any country.

2] According to final analysis, there are 9 countries **like Afghanistan, Comoros, Eretria, Gambia, Liberia, Madagascar, Rwanda, Tanzania, Uganda** need the most aid for their development.

**BIBLIOGRAPHY & REFERENCES**

1] Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research*, *7*(5), 60-78.

2] Grein, A. F., Sethi, S. P., & Tatum, L. G. (2010). A dynamic analysis of country clusters, the role of corruption, and implications for global firms. *East-West Journal of Economics and Business*, *13*(2), 33-60.

3] Mahmood, M. A., Abd El-Aziz, A. A., Gasmi, K., & Hrizi, O. (2021). A Hybrid Clustering Technique to Propose the Countries for HELP International. *Indian Journal of Computer Science and Engineering*, *12*(1), 306-314.

4] Brass, J. N., Longhofer, W., Robinson, R. S., & Schnable, A. (2018). NGOs and international development: A review of thirty-five years of scholarship. *World Development*, *112*, 136-149.

5] Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research*, *7*(5), 60-78.

6] Aboul-Atta, T. A. L., & Rashed, R. H. (2021). Analyzing the relationship between sustainable development indicators and renewable energy consumption. *Journal of Engineering and Applied Science*, *68*(1), 1-16.