Clustering & PCA

Problem to be tackled:

After the recent funding programs we have been able to raise around **$ 10 million** but now we need to find and decide how to use this money **strategically** and **effectively**. The significant issues that come while making this decision are mostly related to choosing the countries that are in the direst need of aid.

We basically need to find which countries are in emergency of need and in which particular sector we can help them to maximize their recovery.

Approach for the analysis:

- After checking the dataset and performing some basic data preparation we can see that **our data consist highly correlated variables** and hence **PCA (principle component analysis)** can be easily applied on it.
- We will be moving forward with the unsupervised learning approach **"Clustering Analysis"** using PCA and **K-means clustering**
- Also we might use **Hierarchical clustering(both single and complete linkage)** on this dataset to create the clusters just check if there is any other kind of information we can squeeze out of this data.

Contd..

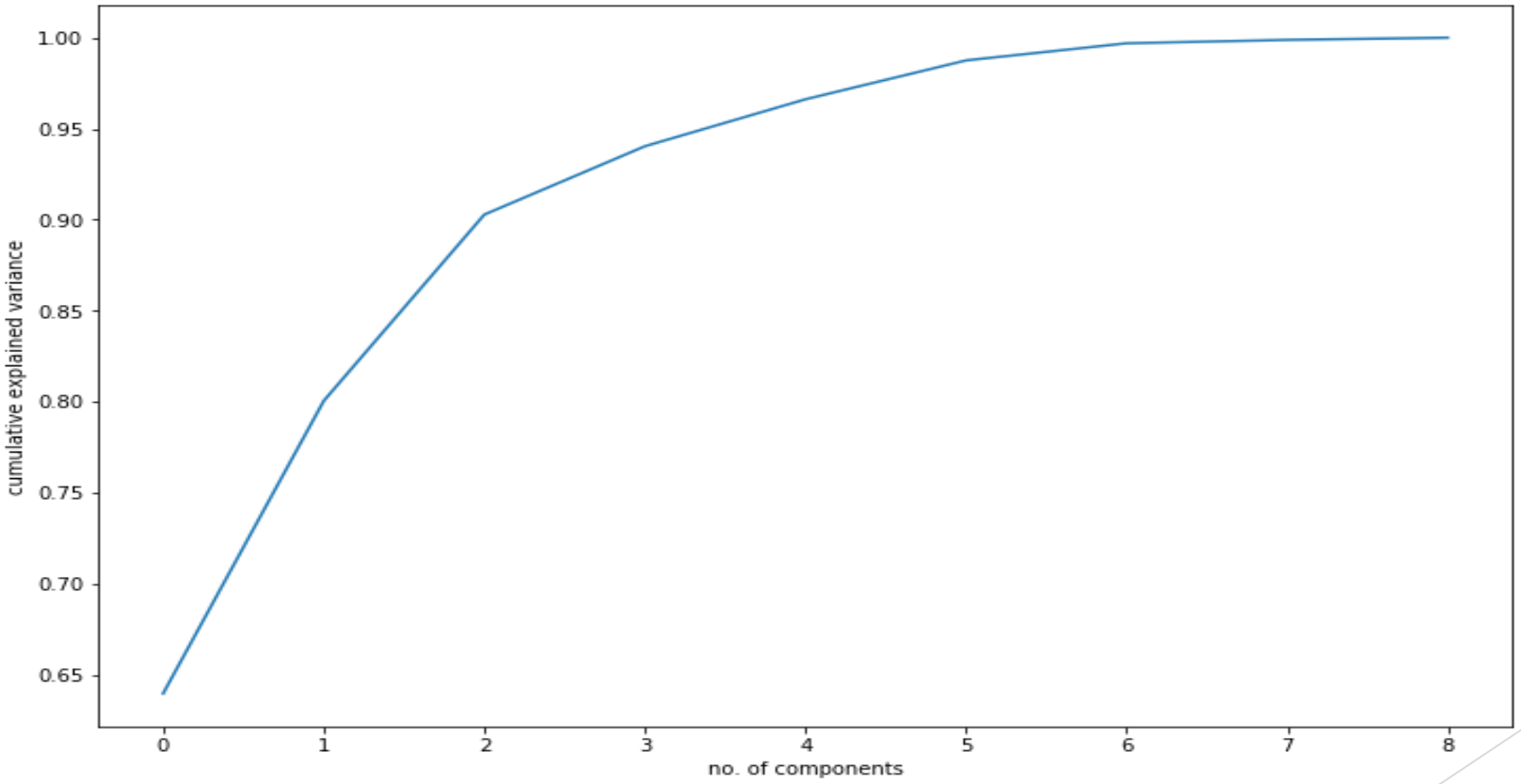
In the end we will Analyze the clusters and identify the ones which are in dire need of aid. we will be analyzing the clusters by comparing how these three variables –

- **GDP per capita**
- **Child mortality(**Death of children under 5 years of age per 1000 live births **)** and
- **Income**

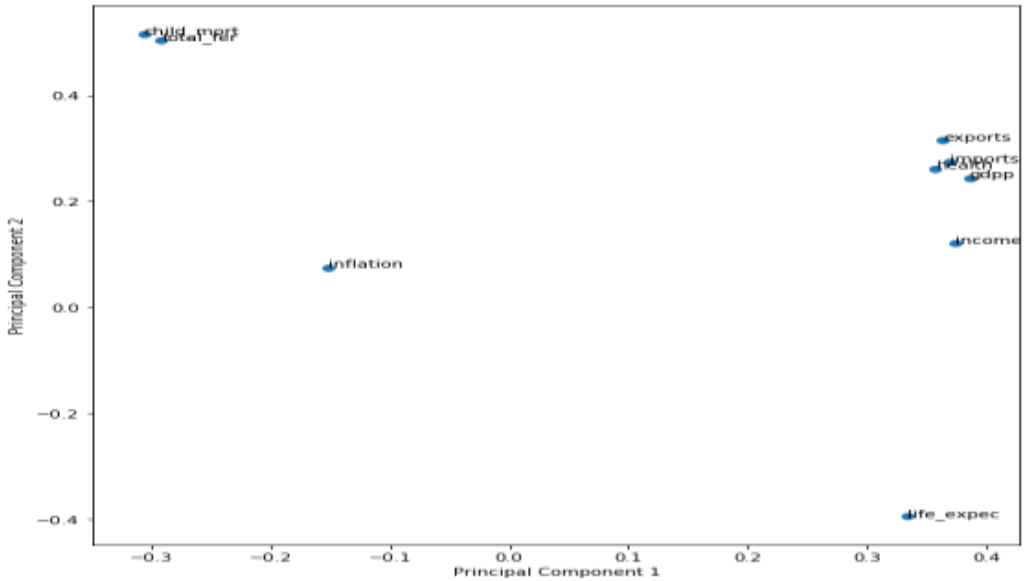
vary for each cluster of countries to recognize and differentiate the clusters of developed countries from the clusters of under-developed countries.

Analysis Observation:

Following are observations which are made after the PCA analysis

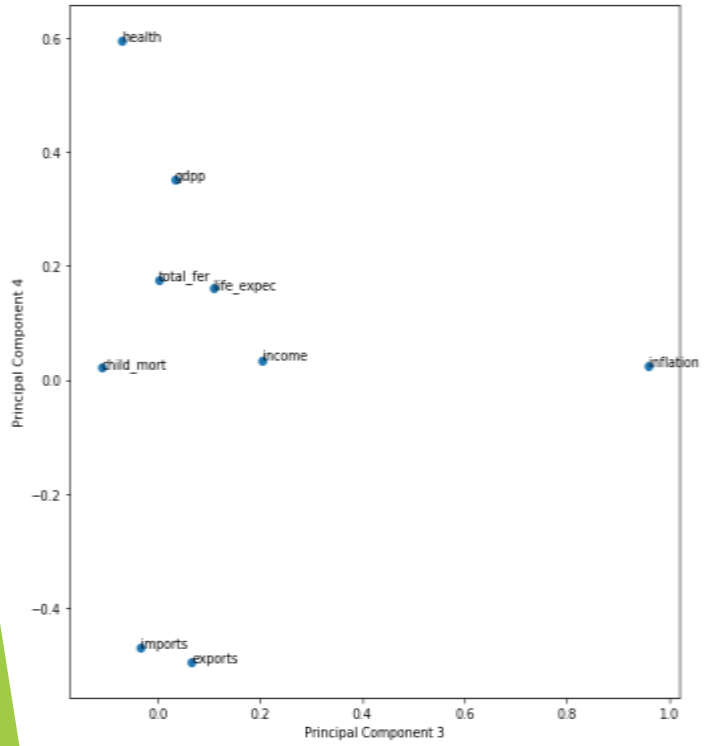


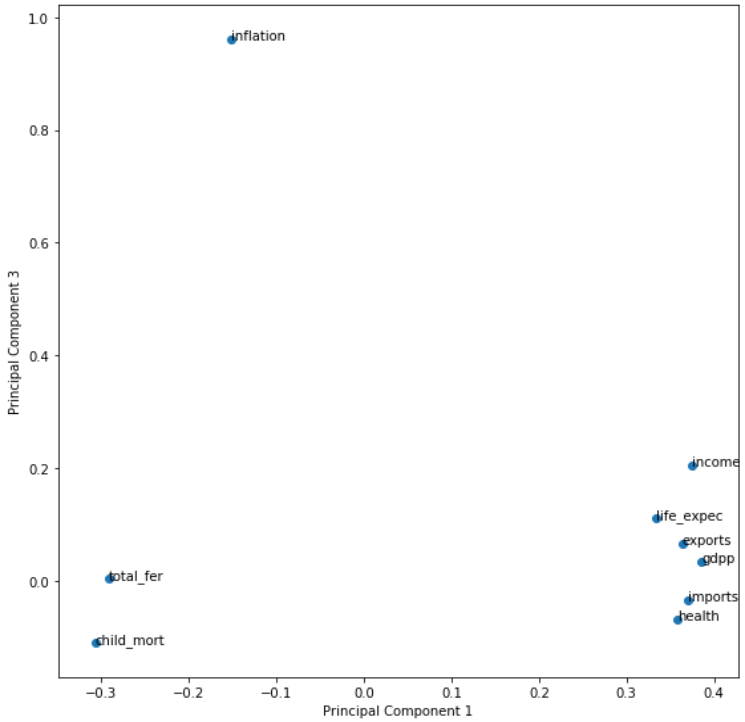
We took 4 principle components as over **95%** of the data is properly explained by the first **4 principal components.**

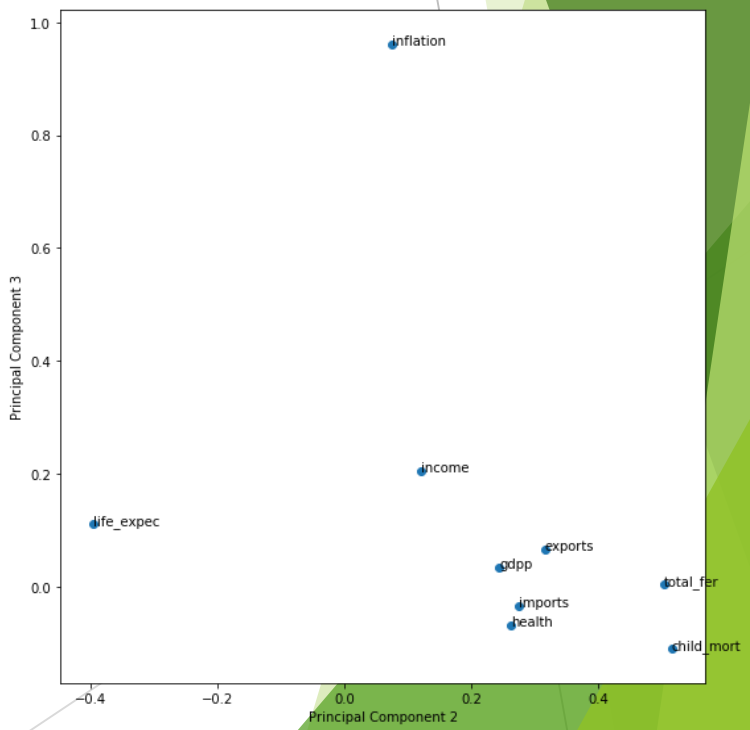


The following graphs tells us that variables such as

- **Life expectancy income GDP per capita** are well explained by **pc1**.
- **Child mortality** and **total fertility** by **pc2**.
- **Inflation** by **pc3**.
- **Health** by **pc4**.

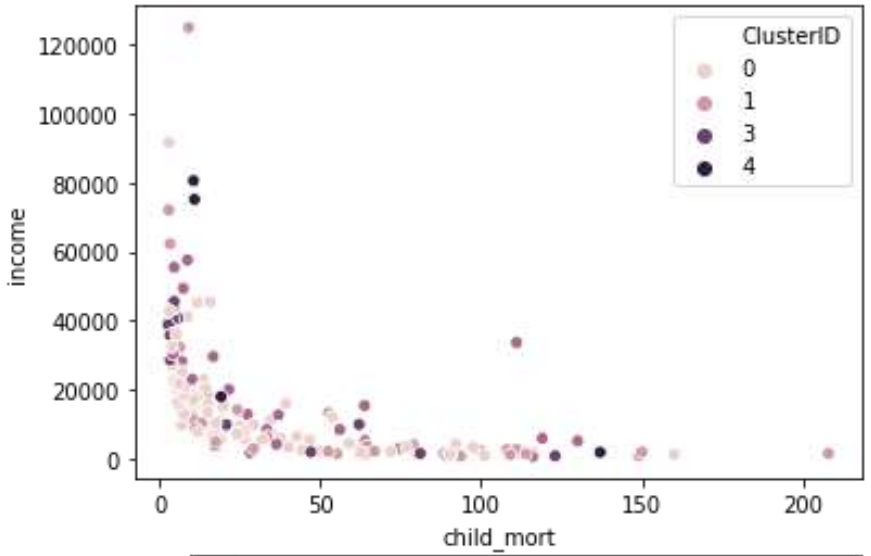




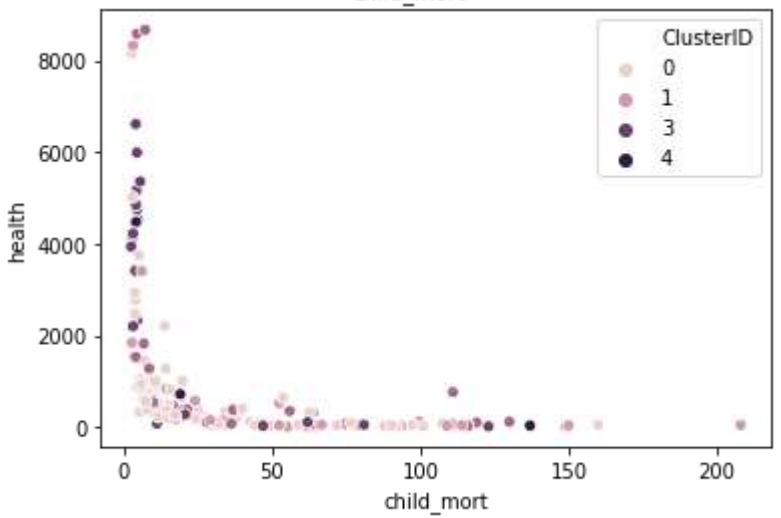


Cluster analysis:

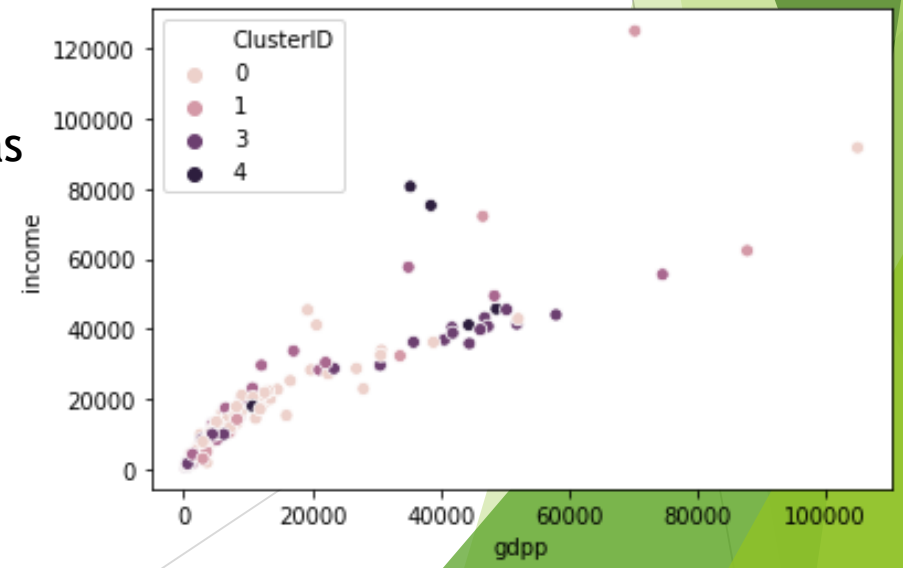
We made five clusters of countries from the dataset using k-means method and PCA.
Here are some visualizations of those clusters based on various variables



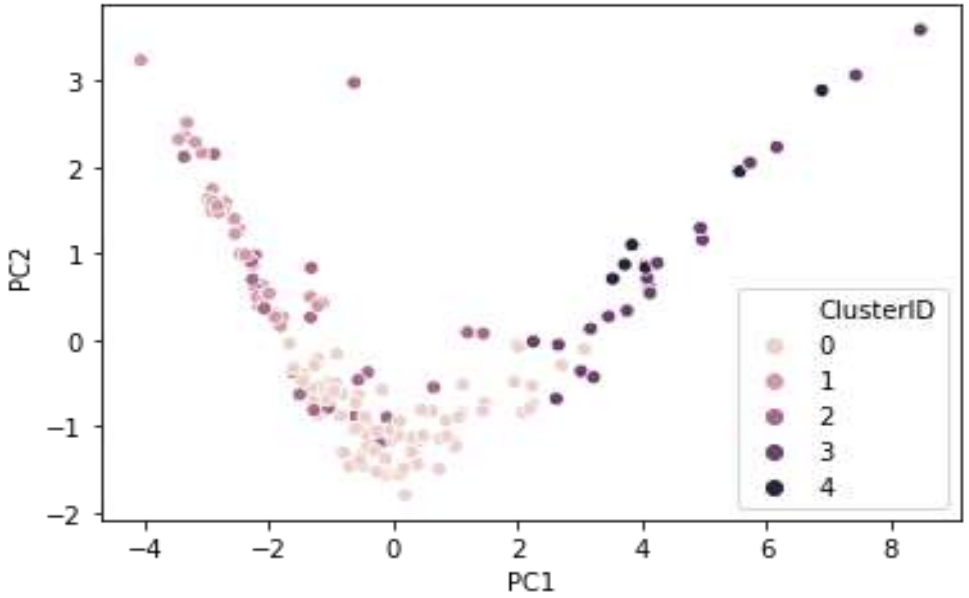
We can observe countries In **clusters 0 and 1** have **low mortality** and **clusters 0 and 1 low income** as compared to other clusters.

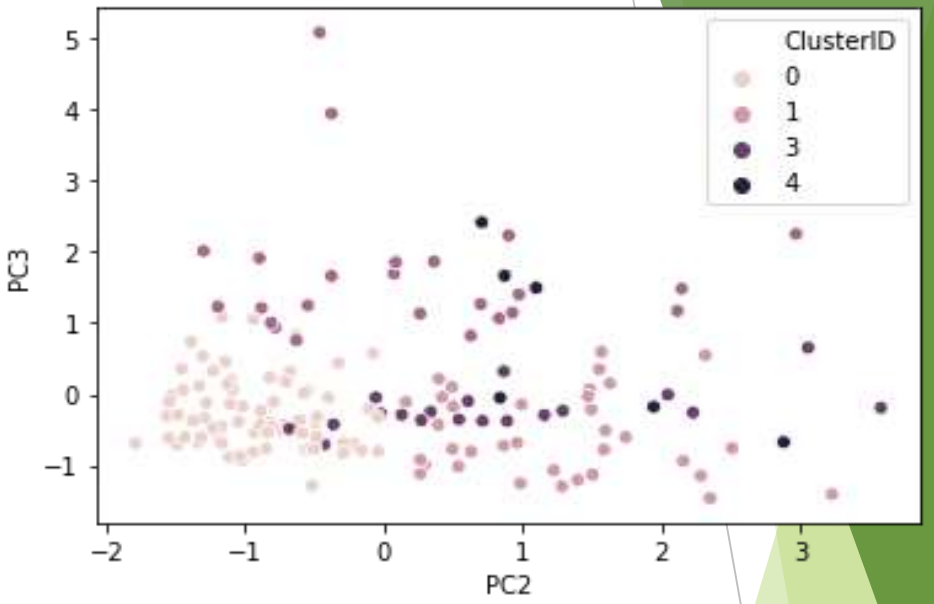


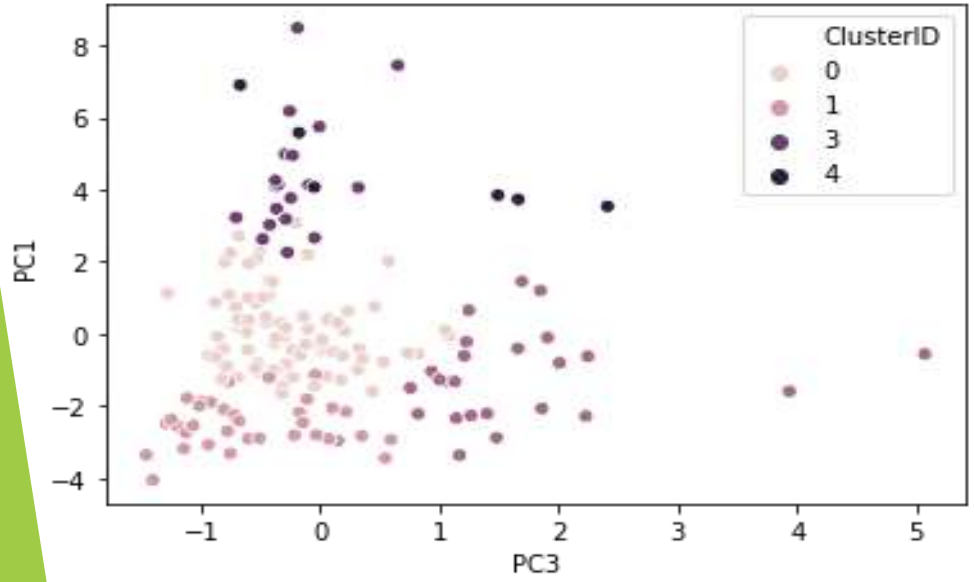
Also **clusters 0 and 1** has very **low gdp** as well as health facilities are also not looking great for these two.

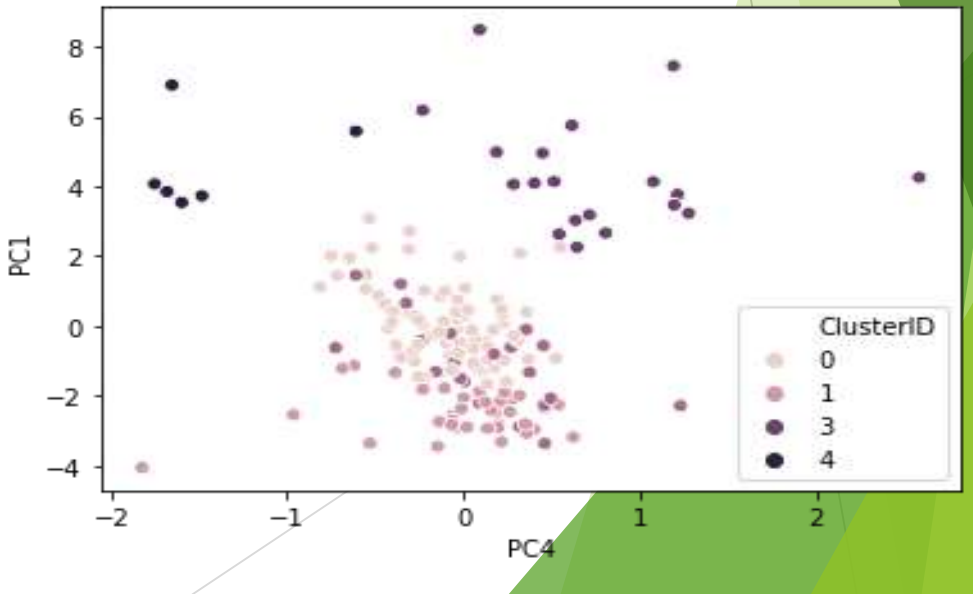


Clusters 0 and 1 showing low values on all indicator variables:

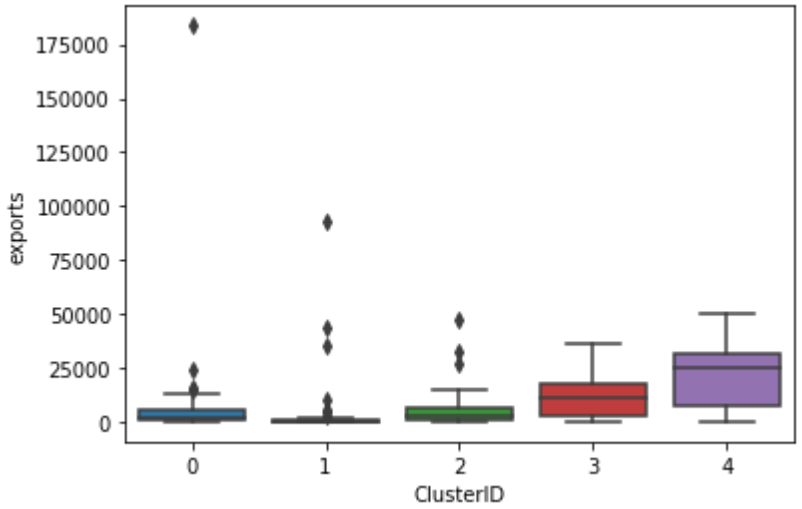




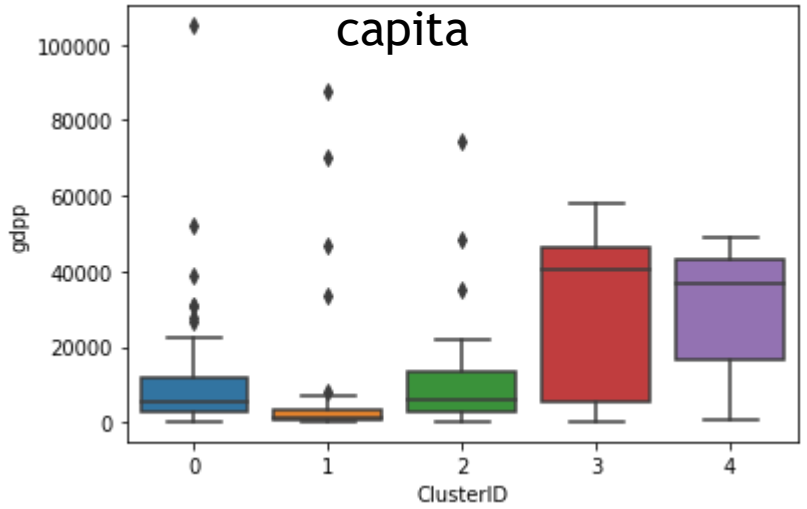




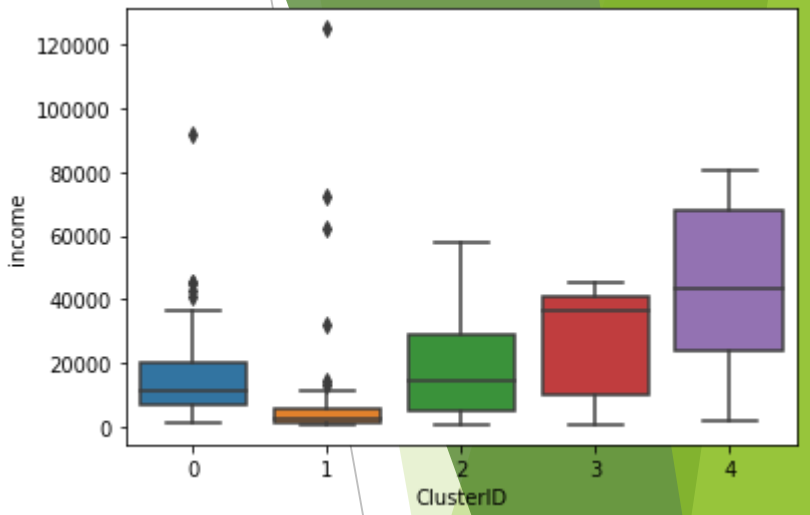
Exports



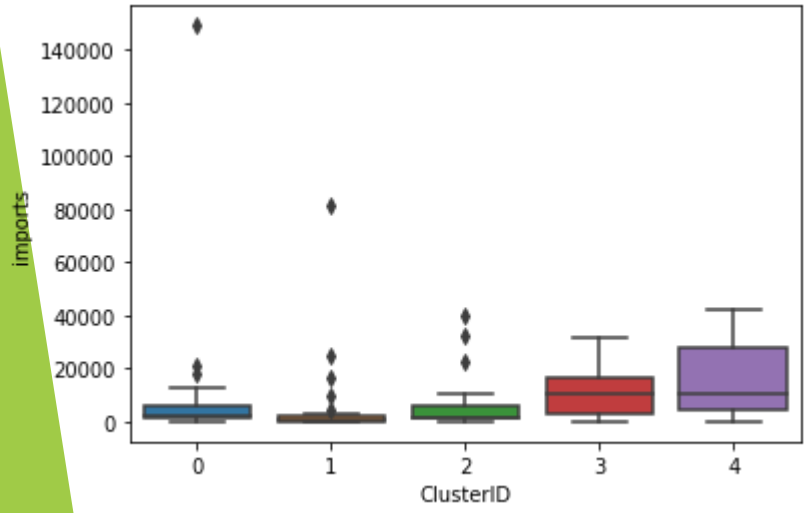
GDP per capita



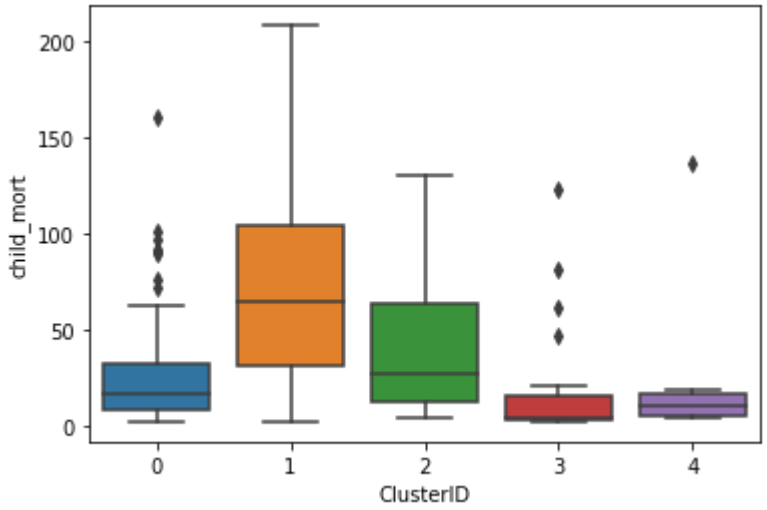
Income



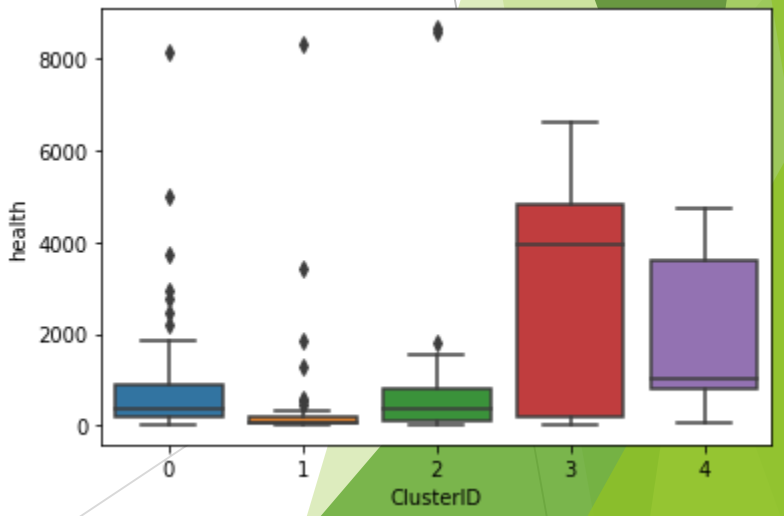
Imports



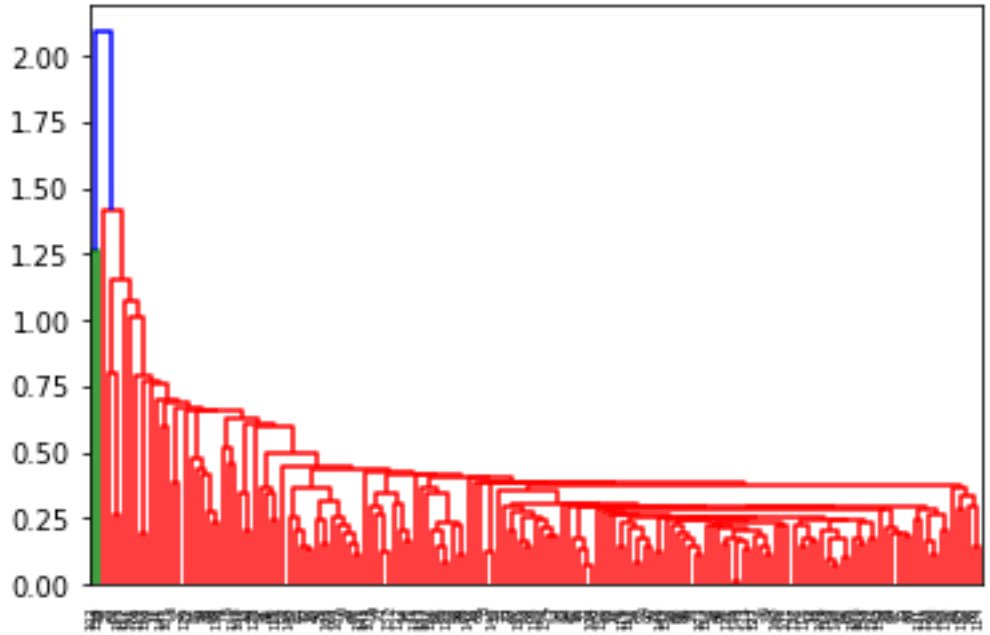
Child mortality



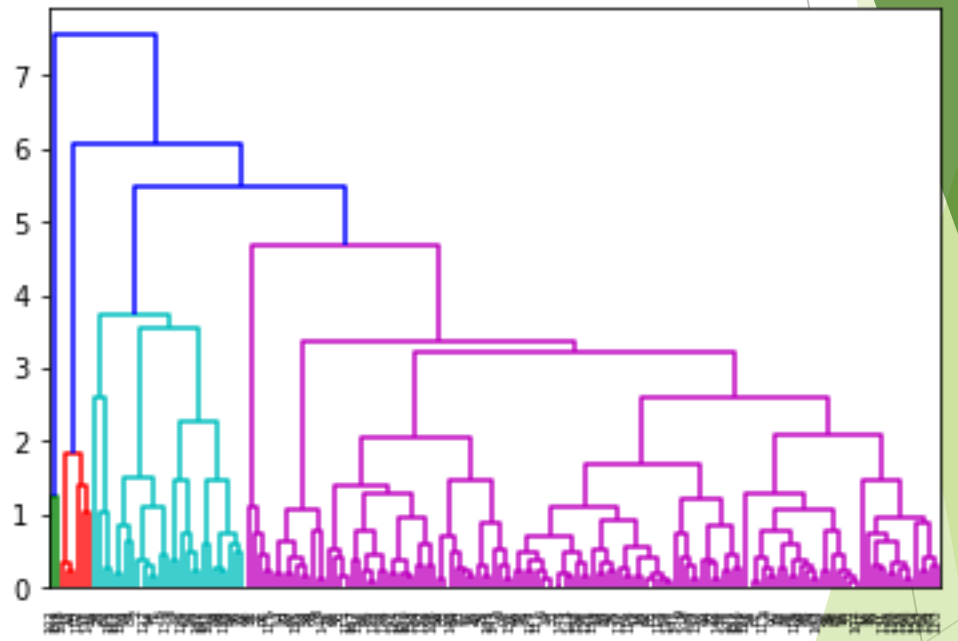
Health



Hierarchical clustering:



Single linkage



Complete linkage

As we can observe from hierarchical clustering no such useful insights can be drawn or evaluated properly hence we can just move forward with **k-means clustering method.**

Summary from the above visualizations:

From the above visualizations it can be very easily observed that countries in **clusters 0 and 1** have low mortality for small children low income and low GDP per capita.

They also have poor imports, exports and health facilities.

Hence countries in the these clusters need help.

We further narrow down our list to those countries in these clusters

who are in emergency for help by using binning method by setting limits which can be found by mean or median of the clusters data

we can use median to find that.

Important Results and Conclusions:

After all the filtering and binning of data we have arrived at the following countries which immediately requires help and are in urgent need of those funds we collected :

1. Afghanistan
2. Benin
3. Burkina Faso
4. Burundi
5. Central African Republic
6. Chad
7. Comoros
8. Congo, Dem. Rep.
9. Gambia
10. Guinea-Bissau
11. Haiti
12. Liberia
13. Malawi
14. Mali
15. Mozambique
16. Niger
17. Sierra Leone
18. Tanzania
19. Togo
20. Uganda