CLUSTERING AND PCA ASSIGNMENT

Done By : Sharath chandra vc

PROBLEM STATEMENT

HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.

After the recent funding programmes, they have been able to raise around $ 10 million. Now the CEO of the NGO needs to 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.

OBJECTIVE

The requisite is:

- To categorise the countries using some socio-economic and health factors that determine the overall development of the country.

- To suggest the countries which the CEO needs to focus on the most.

DATA PROCESSING

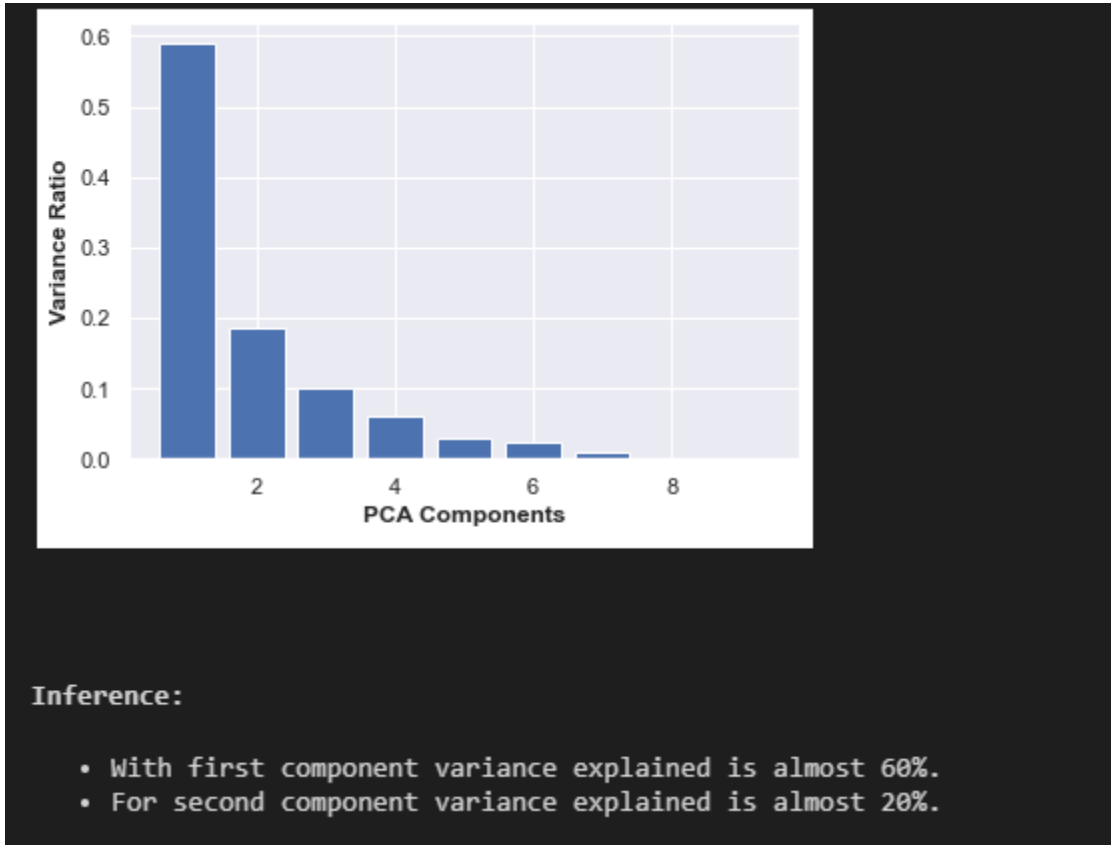
It was found that there were no null values

There were also no duplicate values for country

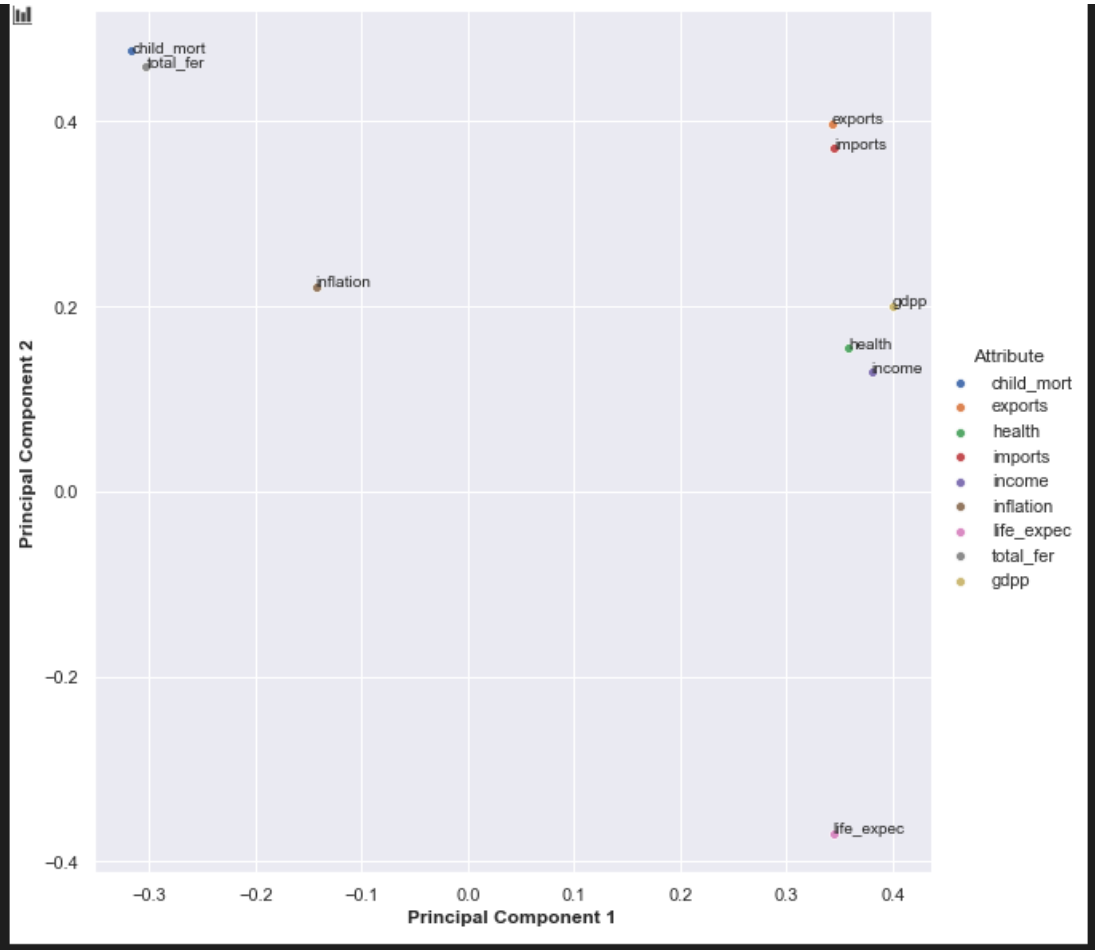
There were a few outliers and they were treated later on during PCA

- The data was standardized for Principal Component Analysis

PCA



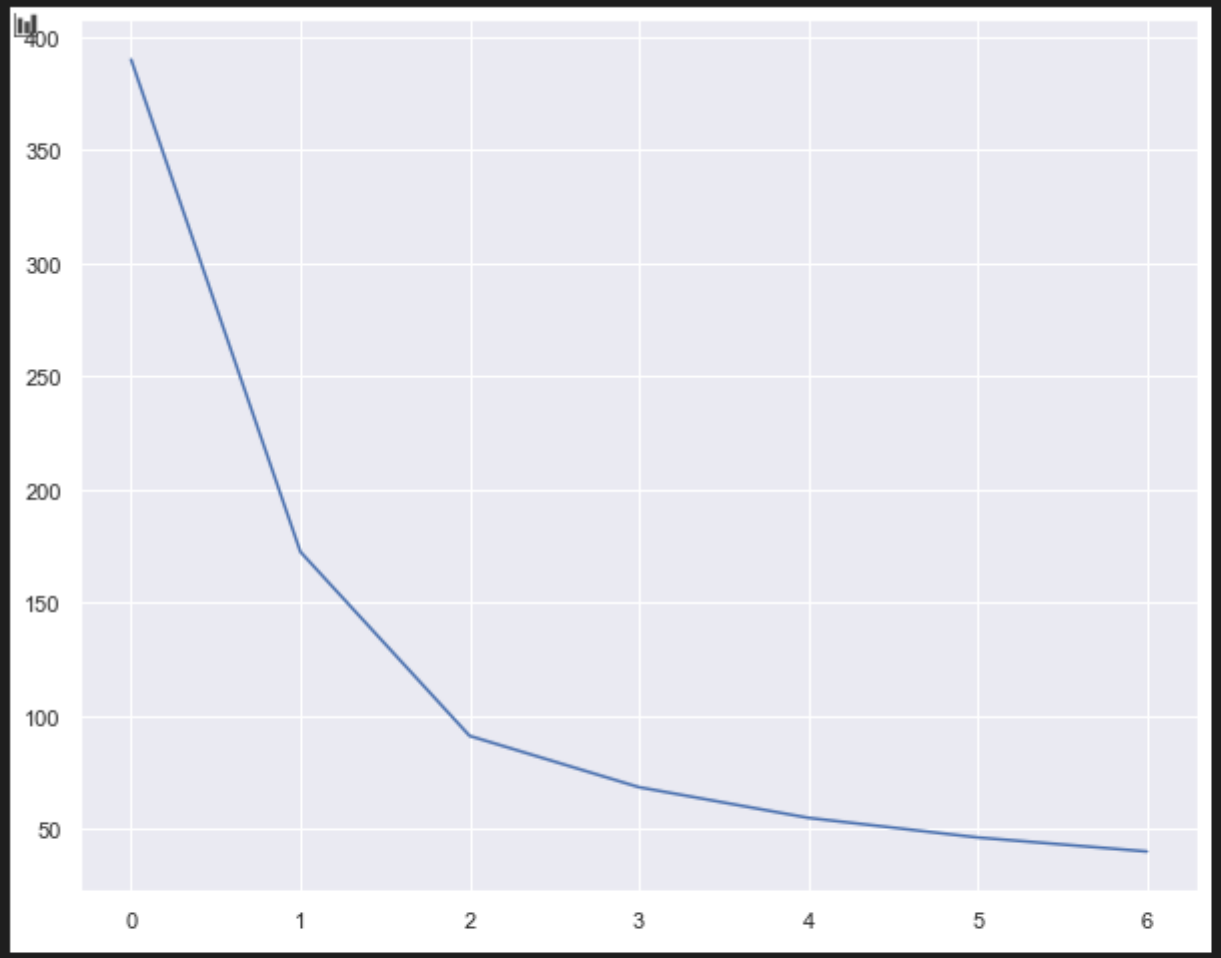
Variable Relation with 2 PCs



Inference:

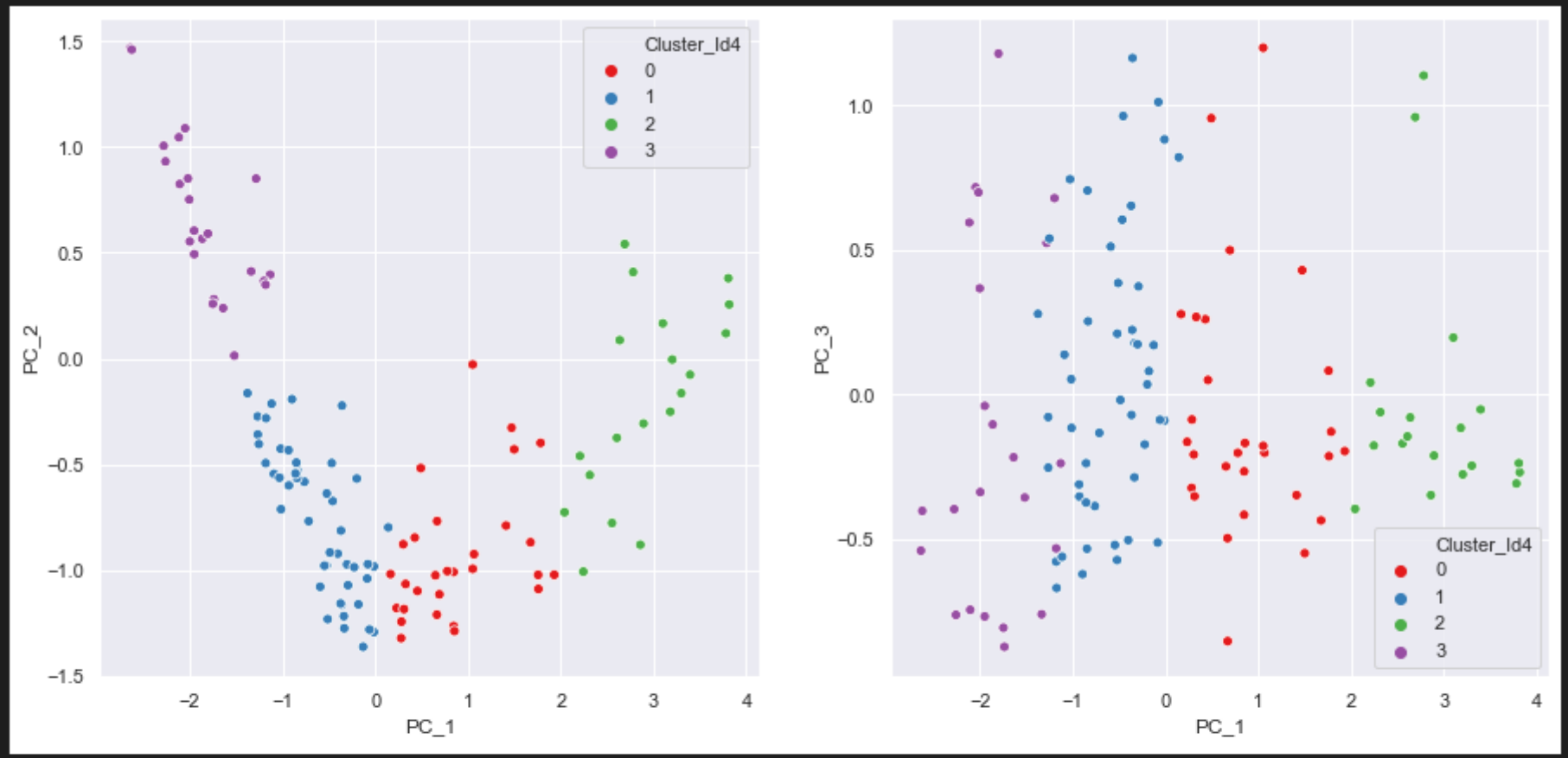
- life expectency, income, gdpp and health are very well explained by PC1.
- imports and exports are well explained by both the components PC1 and PC2.
- child mortality and total fertility are well explained by PC2.
- inflation is neither explained by PC1 nor with PC2

```
[<matplotlib.lines.Line2D at 0x2e2c9f7db50>]
```



Inference:

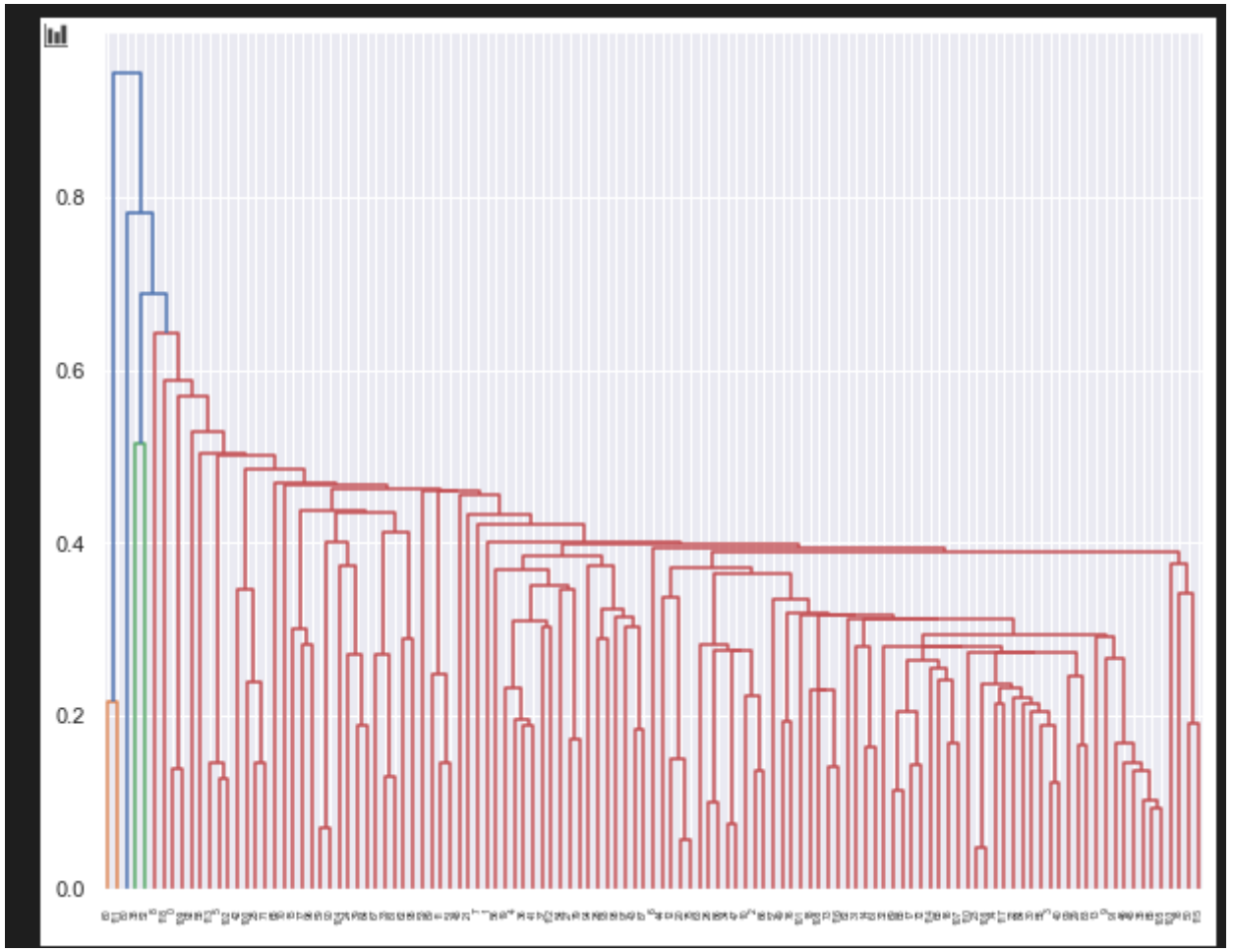
Looking at the above elbow curve it looks good to proceed with either 4 or 5 clusters.

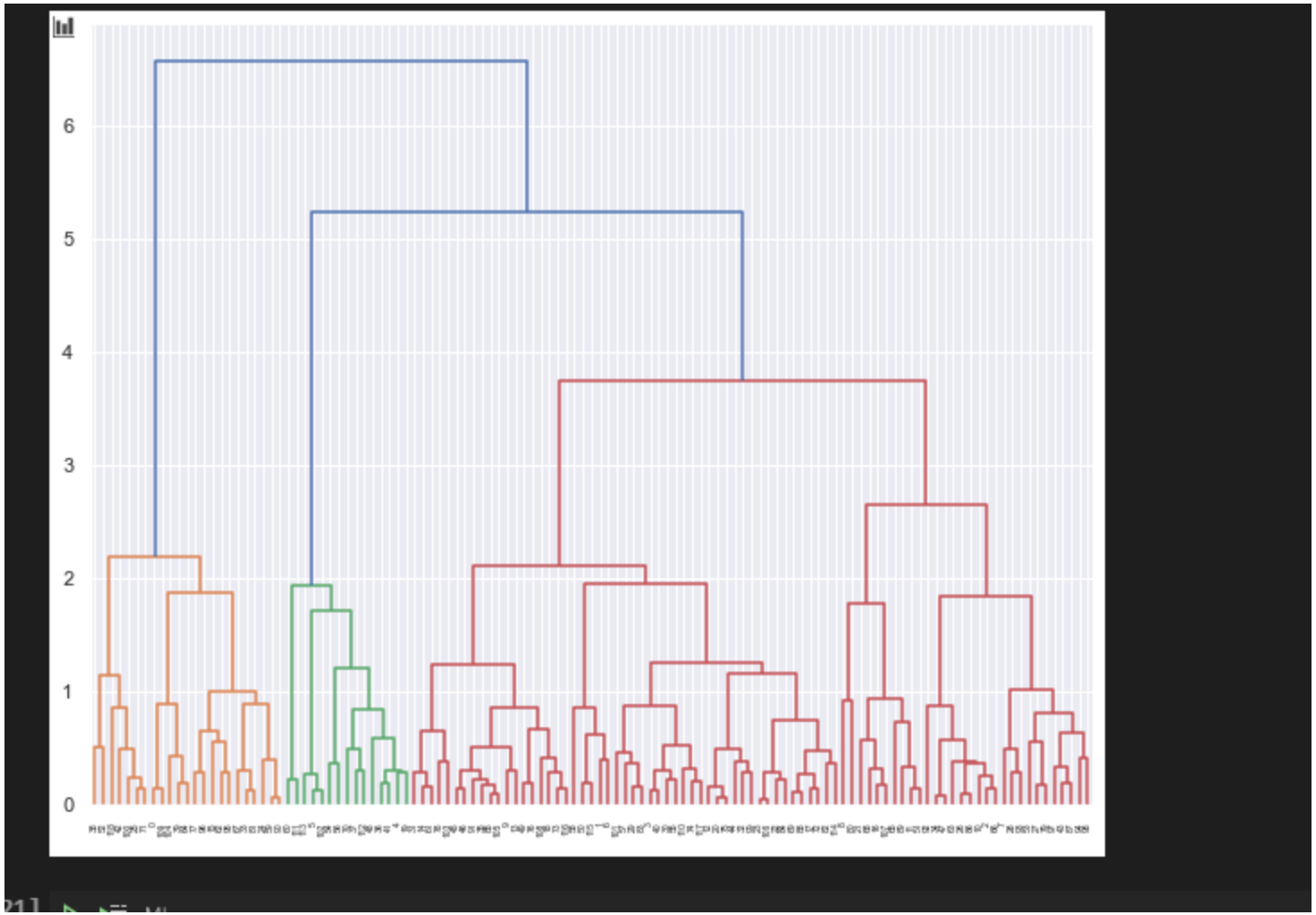


{}

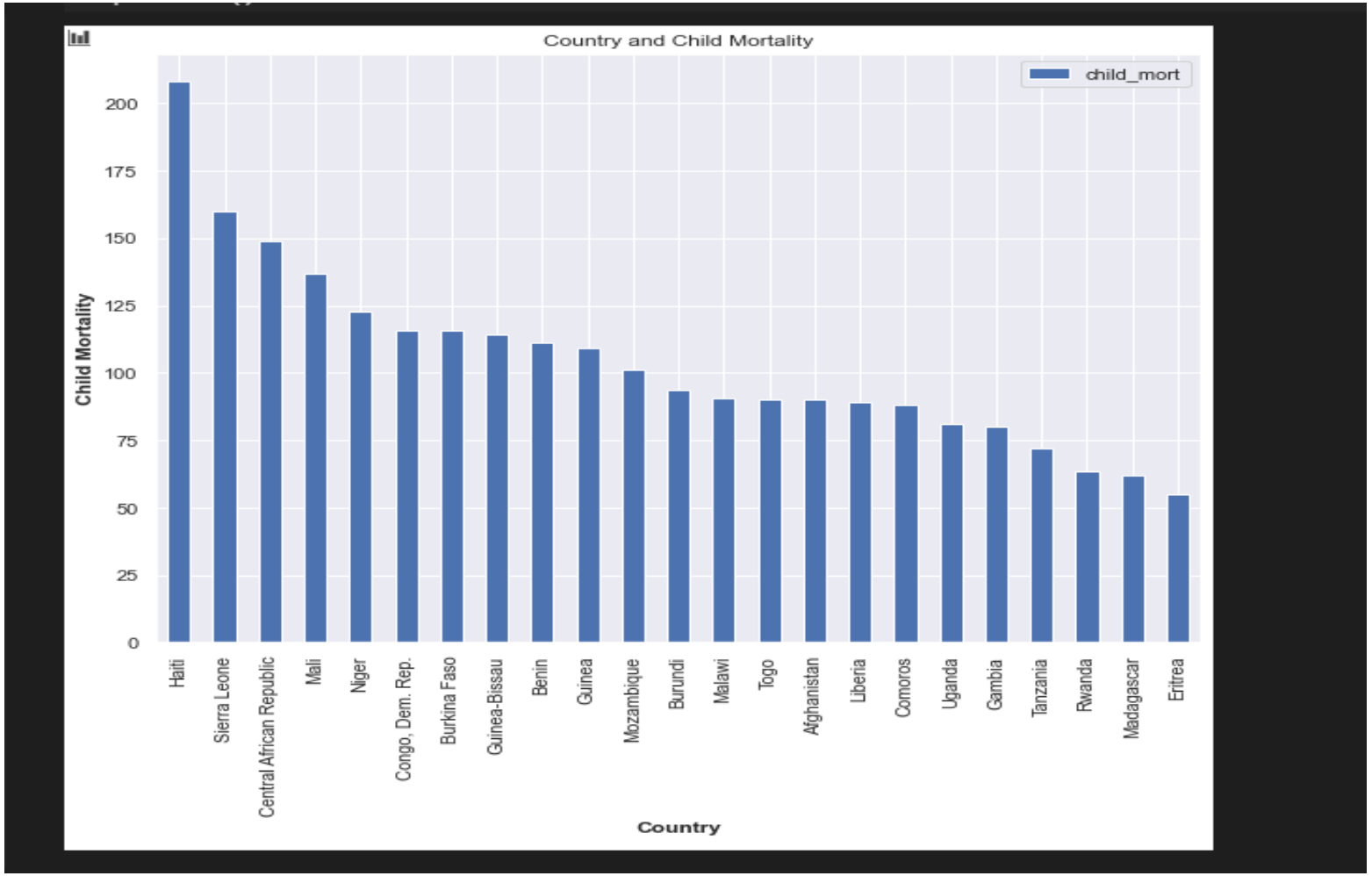
Inference:

- In plot 1, it seems lot of intra-distance between the cluster elements, which is not a good sign.

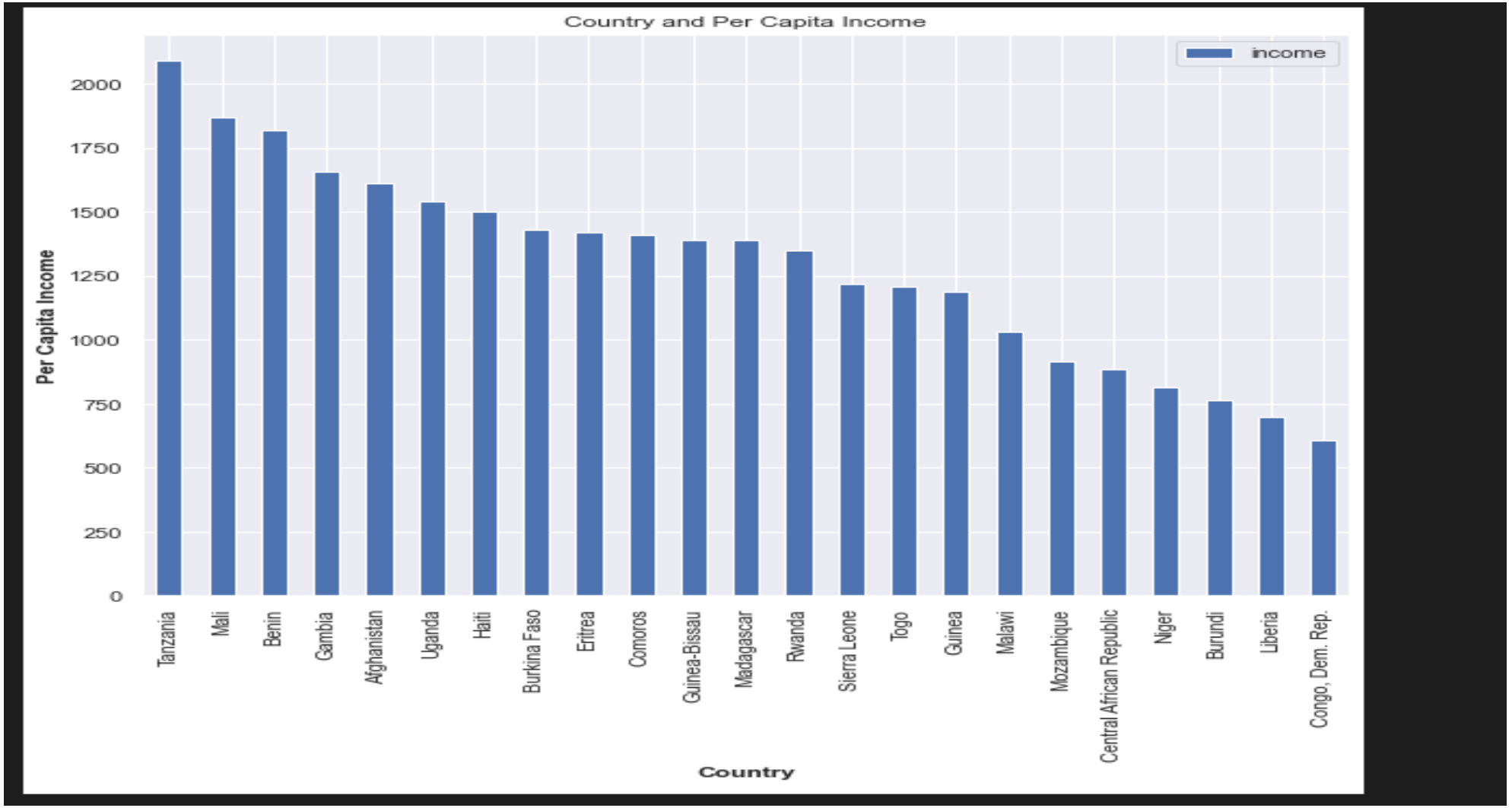

0.8
0.6
0.4
0.2
0.0


6
5
4
3
2
1
0

BarPlot for Child Mortality of countries which are in need of aid



BarPlot for Per Capita Income of countries which are in need of aid



CONCLUSION

The countries that require help the most are listed below:

Afghanistan, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Cote d'Ivoire, Equatorial Guinea, Eritrea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Haiti, Iraq, Kenya, Lao, Madagascar, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, Pakistan, Rwanda, Senegal, Sierra Leone, South Africa, Sudan, Tanzania, Timor-Leste, Togo, Uganda, Yemen and Zambia.

These countries have

- very low rate of net income per person, GDP per capita, average number of years a new born child would live, total health spending and imports of goods and services.

- very high rate of measurement of the annual growth rate, number of children that would be born and child mortality rate.

It is clear that these countries require very quick aid in terms of money, education and services.