Aniket Verma

Problem Statement

We are given a raw data pertaining to different countries across the world with respect to their socio-economic, health and other parameters.

The goal is to assist HELP International, an NGO to allow a fair distribution of funds to the nations which aren't developed which can be useful for them during the time of disasters or natural calamities.

Approach to Tackle the Problem

This is a classic clustering problem and we have used K Means and hierarchical clustering with complete linkage to solve the problem.

We have used Principal Component Analysis for achieving dimensionality reduction.

Key Takeaways / Findings of PCA and Clustering

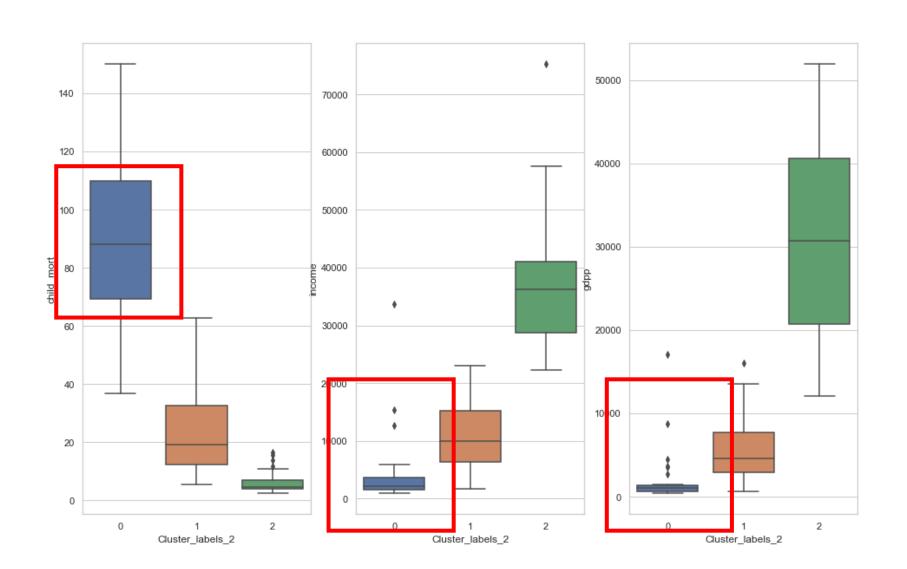
We were able to achieve dimensionality reduction using PCA and found that 3 principal components could be leveraged to explain more than 90% variance in the data set.

We then leveraged both silhouette score method and elbow curve to determine optimal number of clusters for K Means.

We also performed hierarchical clustering (complete linkage) and found that the results were very similar to the K Means algorithm.

- After analysing the raw data, we obtain 3 broad categories of countries, classified on the basis of the following parameters:
 - Child Mortality Rate
 - Net Income
 - GDP Per Capita

Category – I (Critical Area)



Category – I (Critical Area)

The first category belongs to the countries with the following features:

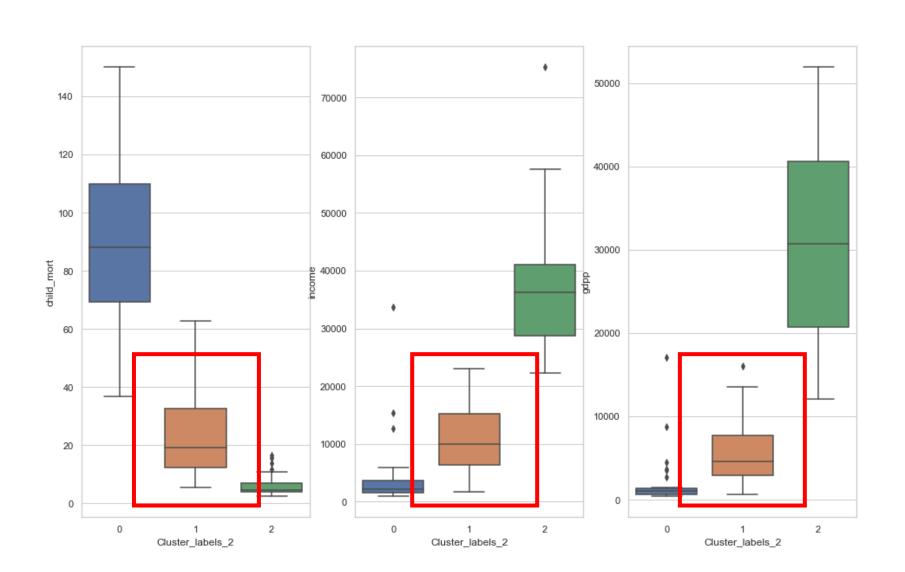
High Mortality Rate

Very Low Income

Very Low GDP per capita.

Observation: Clearly, this area demands the most attention and major chunk of the funds needs to be diverted to this area.

Category – I (Sub-Critical Area)



Category — II (Sub-Critical Area)

The first category belongs to the countries with the following features:

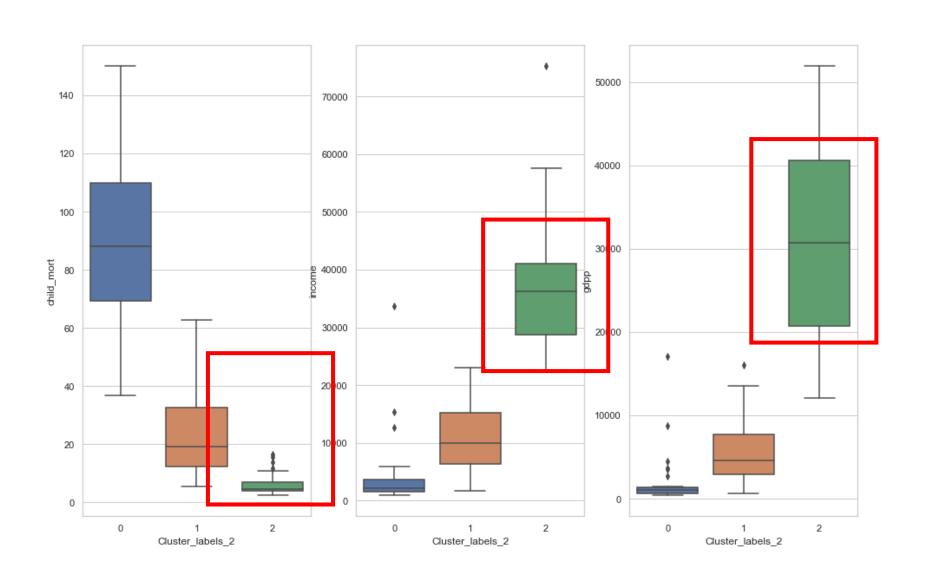
Low Mortality Rate

Low Income

Low GDP per capita.

Observation: This area might be lacking in surplus funds or appropriate income, however the health and wellness isn't compromised.

Category – III (Safe Area)



Category – III (Safe Area)

The first category belongs to the countries with the following features:

Low Mortality Rate

High Income

High GDP per capita.

Observation: This area is doing fine and as of now might not require much attention with respect to funds.

Based on out analysis we propose to diverge maximum funds to Critical (Category – I) areas and if needed, to a few Sub-Critical (Category – II) areas.

Let us look at which countries fall in the critical region.

Countries under Critical Area

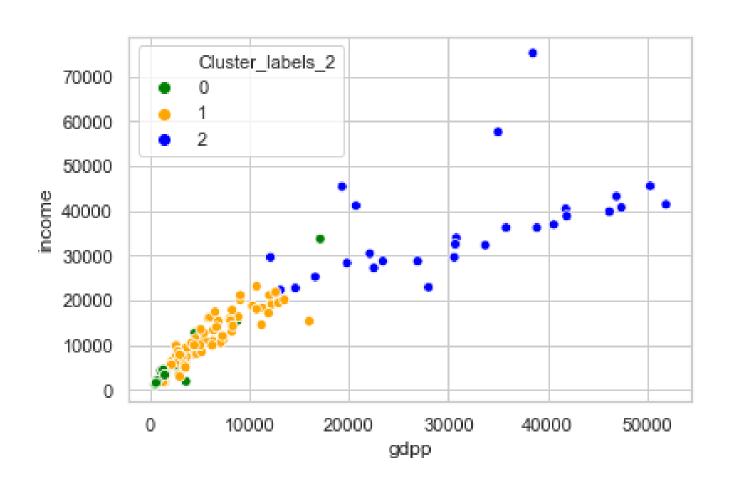
country	gdpp	child_mort	income
Afghanistan	553	90.2	1610
Angola	3530	119	5900
Benin	758	111	1820
Burkina Faso	575	116	1430
Cameroon	1310	108	2660
Chad	897	150	1930
Comoros	769	88.2	1410
Congo, Rep.	2740	63.9	5190
Cote d'Ivoire	1220	111	2690
Equatorial Guinea	17100	111	33700

country	gdpp	child_mort	income
Mauritania	1200	97.4	3320
Pakistan	1040	92.1	4280
Rwanda	563	63.6	1350
Senegal	1000	66.8	2180
Sudan	1480	76.7	3370
Tanzania	702	71.9	2090
Timor-Leste	3600	62.6	1850
Togo	488	90.3	1210
Uganda	595	81	1540
Yemen	1310	56.3	4480
Zambia	1460	83.1	3280

country	gdpp	child_mort	income
Gabon	8750	63.7	15400
Gambia	562	80.3	1660
Ghana	1310	74.7	3060
Guinea	648	109	1190
Guinea-Bissau	547	114	1390
Iraq	4500	36.9	12700
Kenya	967	62.2	2480
Lao	1140	78.9	3980
Malawi	459	90.5	1030
Mali	708	137	1870

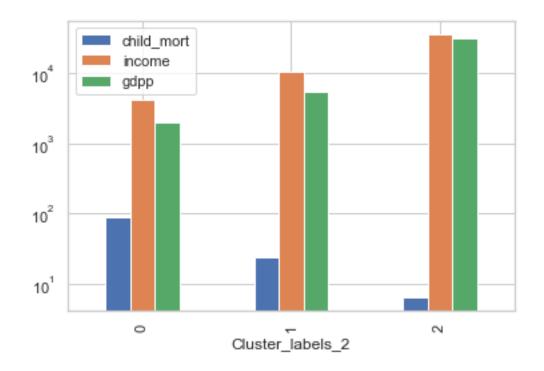
- There are a total of 31 countries in the critical area.
- Highest child mortality rate Chad (150).
- Lowest net income Guinea (1190)
- Lowest GDP Per Capita Malawi (459)
- Equatorial Guinea is doing fine with respect to GDP and net income compared to other countries however the mortality rate is still high.
- Gabon and Iraq are two other countries which are having high mortality rate despite comparatively better GDP and net income.

Category – I (Green Region)



Comparison of All Regions

• In the following graph it is clearly evident that the first region is liable to get the maximum financial aid.



Thank You