

HELP International - Investment plan

Submitted by – Soumik Banerjee

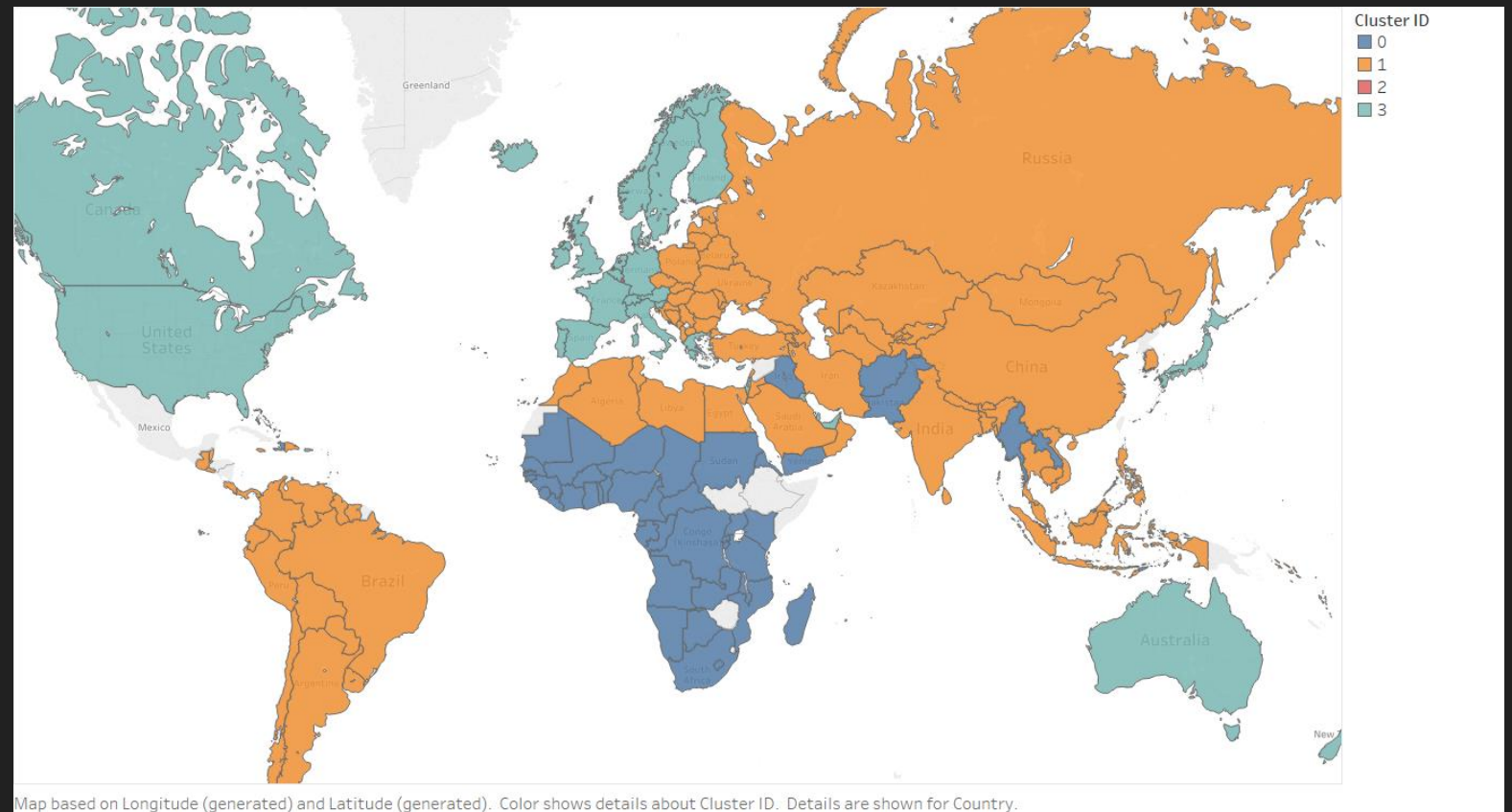
Resources and Available data

- Total available Corpus - \$ 10 million
- Country Related data explaining the following metrics for 167 total countries
 - Child mortality
 - exports
 - health
 - imports
 - Income
 - Inflation
 - Life expectancy
 - Total fertility
 - GDPP

Countries as per the clusters

Colour	Cluster Number
Blue	Cluster – 1
Orange	Cluster – 2
Red	Cluster – 3
Teal	Cluster – 4

Cluster #	# of Countries
CL - 1	48
CL - 2	87
CL - 3	3
CL - 4	29



Trends in the country dataset

Comparative study of all country data – results are in comparison to global average

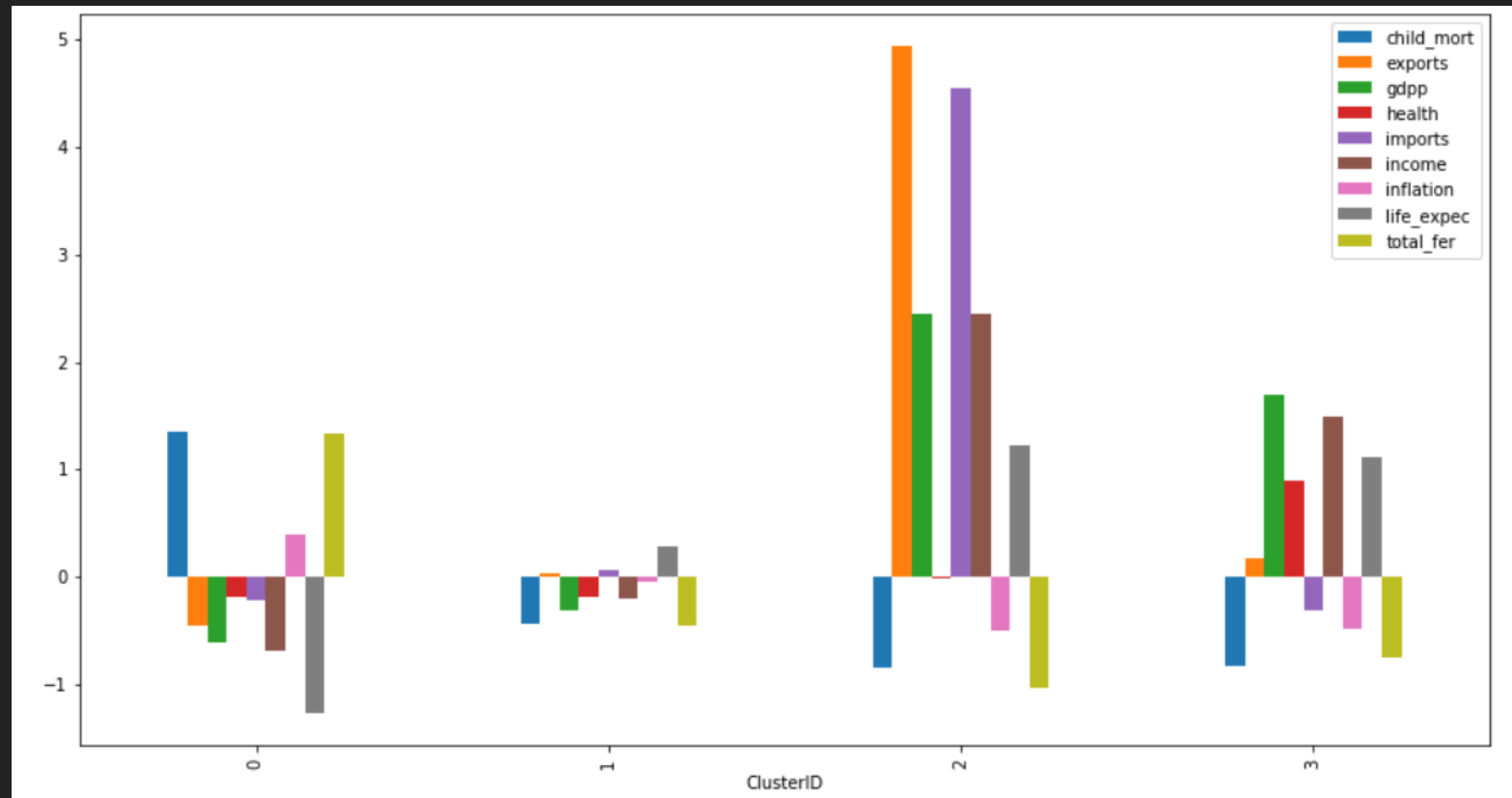
The countries are classified into 4 sets with the characteristics as explained in the table.

Cluster-1	Child mortality – high Exports – low Health expenses – comparable or lower Imports – low Income – inferior Inflation – moderately high Life expectancy – inferior Total fertility – high GDPP – inferior	Cluster-2	Child mortality – low Exports – comparable Health expenses – comparable or lower Imports – comparable or higher Income – low Inflation – comparable or lower Life expectancy – high Total fertility – high GDPP –low
Cluster-3	Child mortality – inferior Exports – high Health expenses – comparable Imports – high Income – high Inflation – inferior Life expectancy – high Total fertility – inferior GDPP – high	Cluster-4	Child mortality – inferior Exports – comparable or higher Health expenses – high Imports – low Income – high Inflation – inferior Life expectancy – high Total fertility – inferior GDPP – high

Trends in picture

The trends are documented in previous slide.

This image is a pictorial representation of cluster behavior across all parameters in the given dataset.



Technical Discussions – Steps to identification of the Clusters

- PCA analysis gave 5 components which were able to account for 94% variance in data.
- The PC analysis gave the following trend –
 - PC1 - 45% - child mortality/ income/ life expectancy/ total fertility/GDPP
 - PC2 - 18% - exports/imports
 - PC3 - 13% - health/inflation
- Silhouette score gave 4 and 5 as best candidates for number of clusters in K-means approach.
- 4 clusters were chosen as 5 clusters were having some groups with < 5% of total populations as members.
- Hierarchical Approach was also attempted on the dataset with multiple cut – levels.
- All resultant clusters had at least one group having < 5% of total population as members.
- Reducing the number of clusters, is resulting in loss of insight.

Principal Component Analysis

K- Means Clustering Process

Hierarchical Clustering Process

Recommendations

- **Cluster 1** countries are in a comparatively difficult situation and they are showing significant

- Child Mortality rates
- Low life expectancy
- Low per capita income and high inflation

The above, along with other factors, are resulting in a very low GDPP for them

- **Cluster 2** countries are also lacking in some parameters but have a better GDPP compared to Cluster 1 countries. Areas that can be improved are

- Per Capita Income
- Exports (as they are significantly lower than Cluster 3/4 countries)
- Total fertility rate