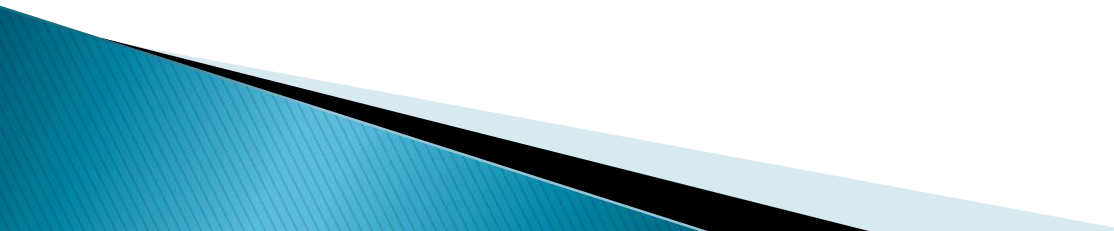


Clustering & PCA Assignment

Naga Sai Aditya B

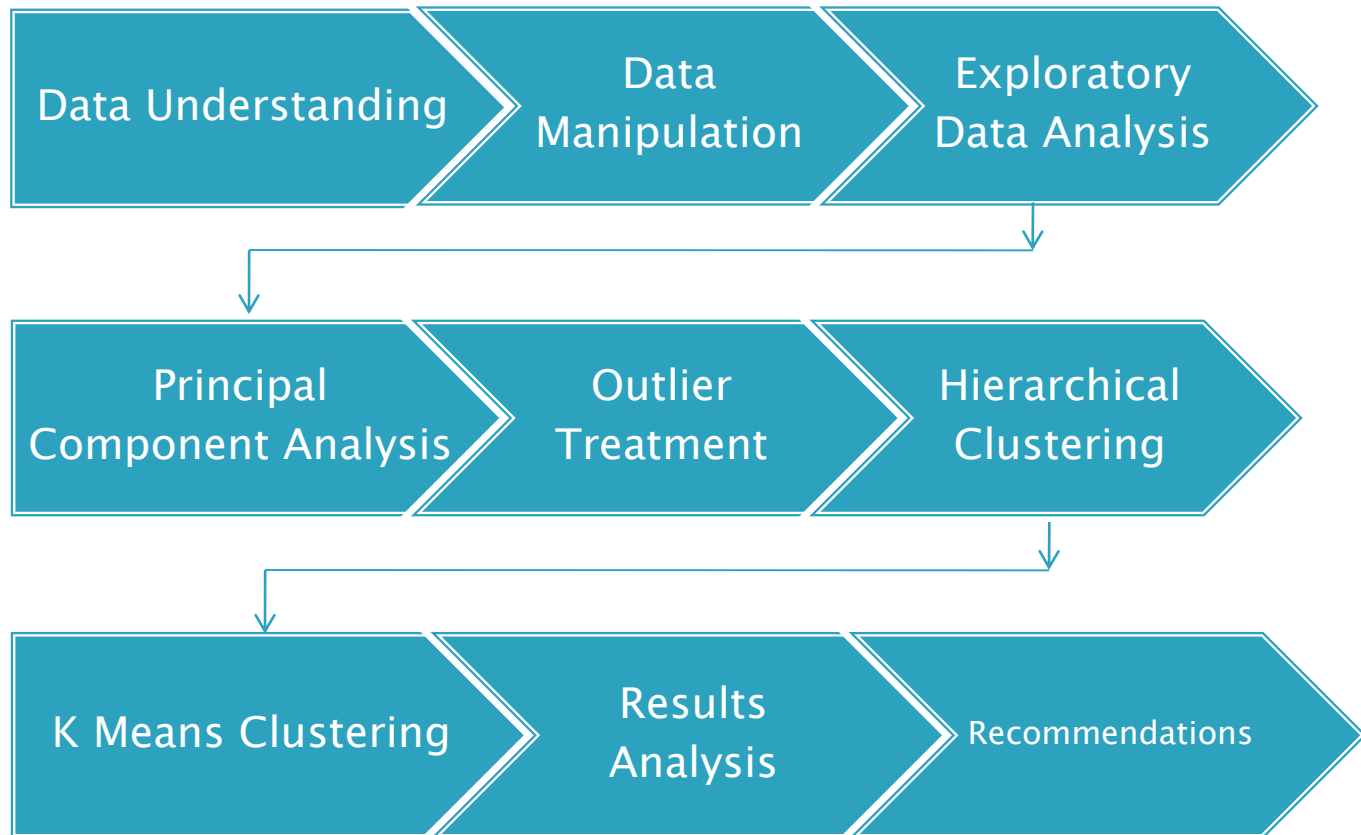


HELP International – is an international humanitarian NGO

- HELP International is an organization to fighting poverty and providing the basic amenities and relief to backward countries..
 - HELP International was founded in October 2003 as a merger of charities working in Zimbabwe and India from 1992
 - It's mission to assist adults with acquiring life skills that empower them to increase their self-sufficiency within their communities
- 

Objectives

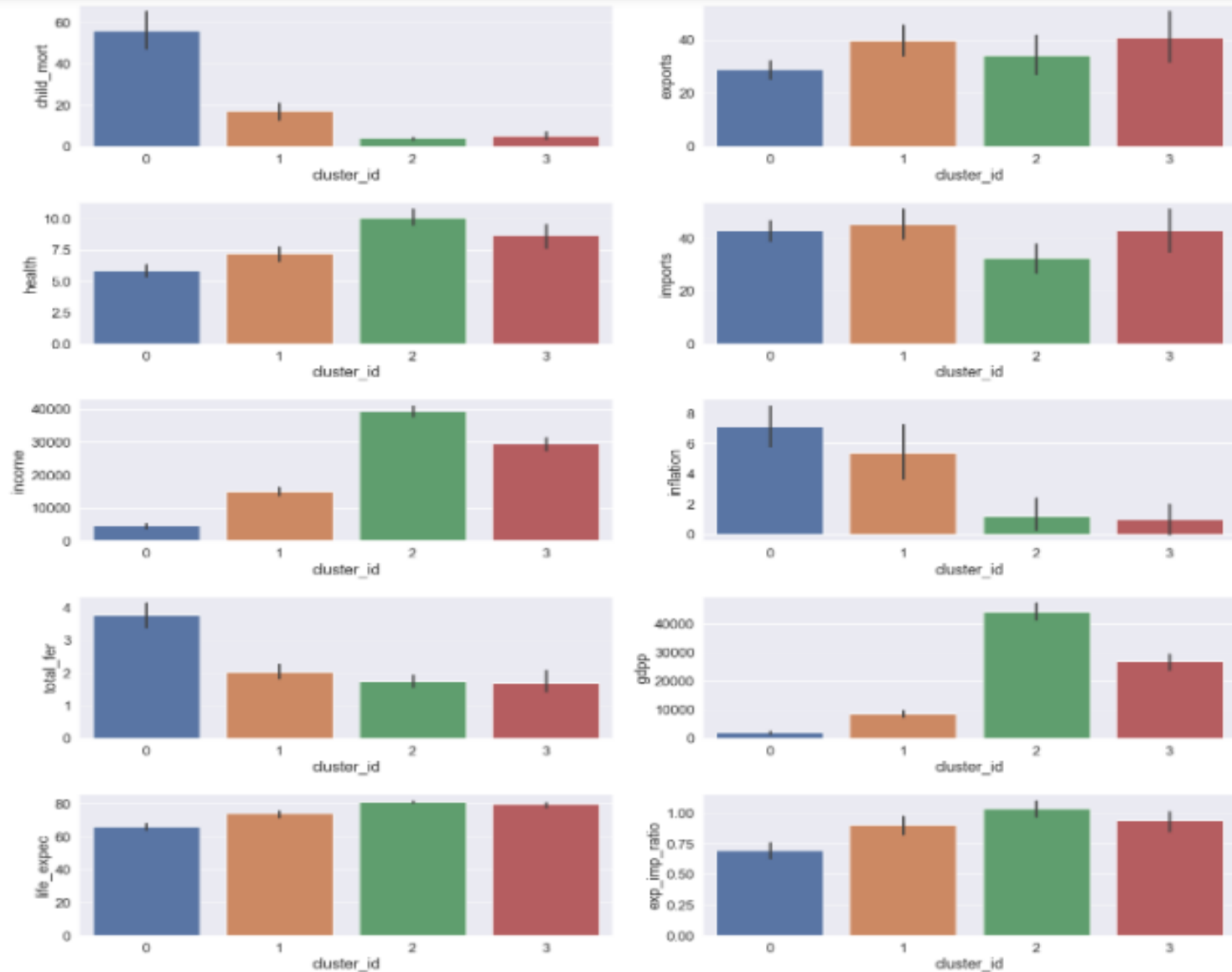
- Categorize using socio-economic and health factors that determine the overall development of the country
- Suggest list of Countries to CEO that they can focus more strategically and effectively

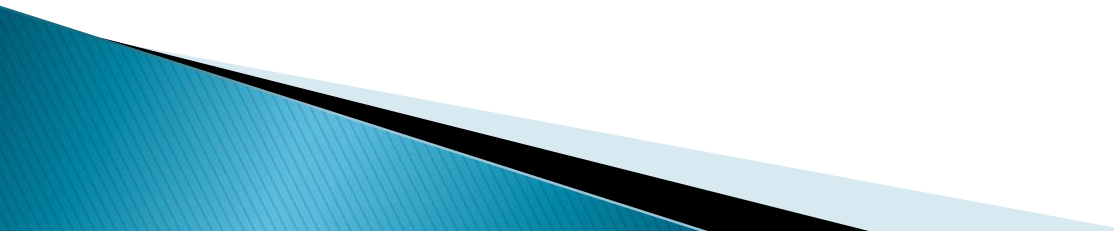


Technical Aspects in Brief

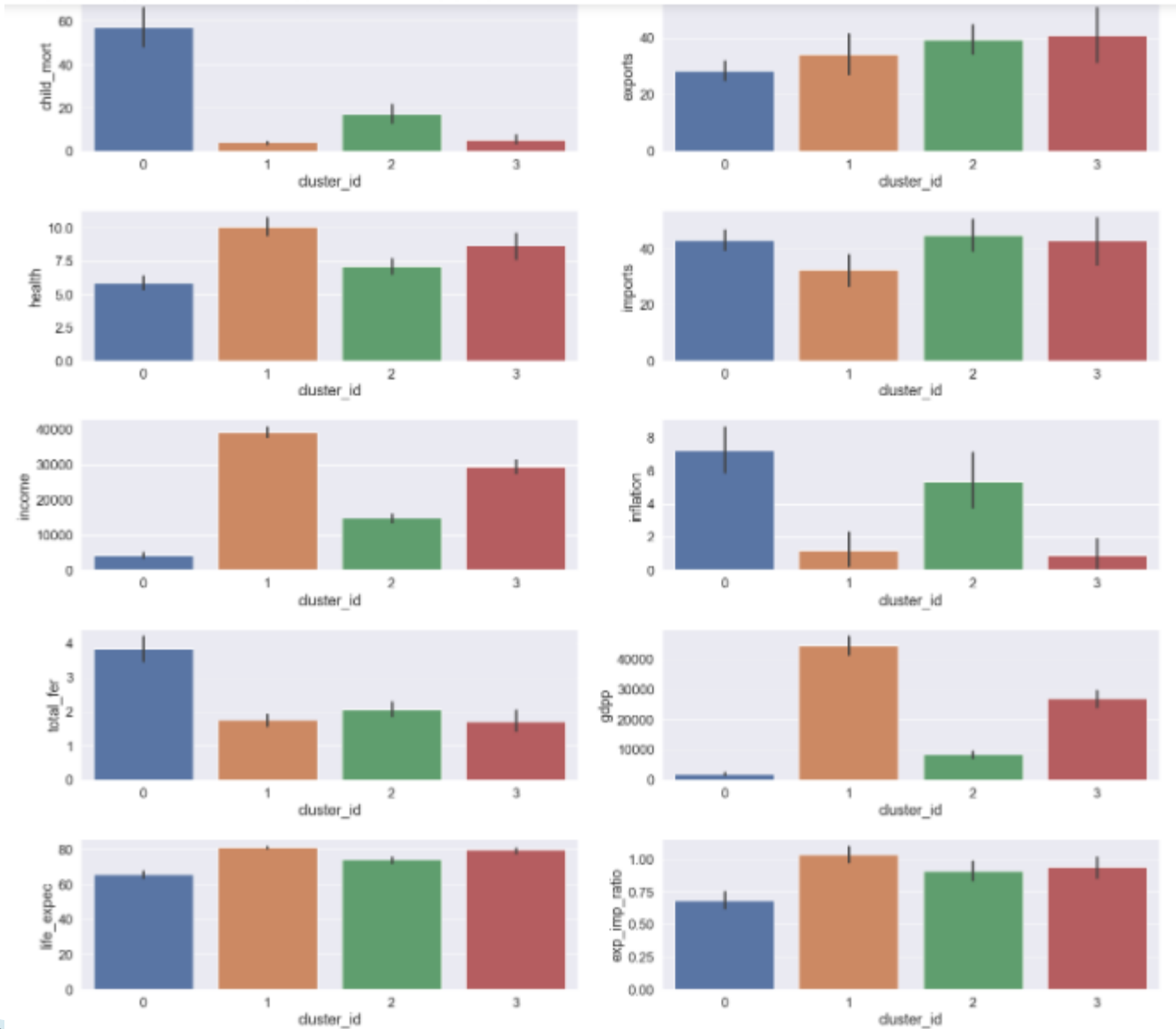
- ▶ Data was thoroughly understood
- ▶ Derived metric export to import ratio was added.
- ▶ Using Pair Plot the relations and patterns were understood better
- ▶ The data was normalized for further analysis
- ▶ Highly correlated metrics were dropped after plotting heatmap.
- ▶ PCA was applied without performing outlier treatment as doing so diminished the data set size by more than half
- ▶ Scree Plot showed that 4 principal components would show 95% of variance and it was taken
- ▶ Got very minimal maximum correlation of 1.058×10^{-16} and minimum correlation of -5.79×10^{-16}
- ▶ Outlier Treatment was done. 11% of data on the higher end was only treated.
- ▶ Hierarchical clustering was performed and 4 clusters were chosen after plotting dendrograms as it gave the optimal conditions for analysis
- ▶ K Means clustering was performed and 4 clusters were chosen from elbow curve as it gave the optimal conditions for analysis

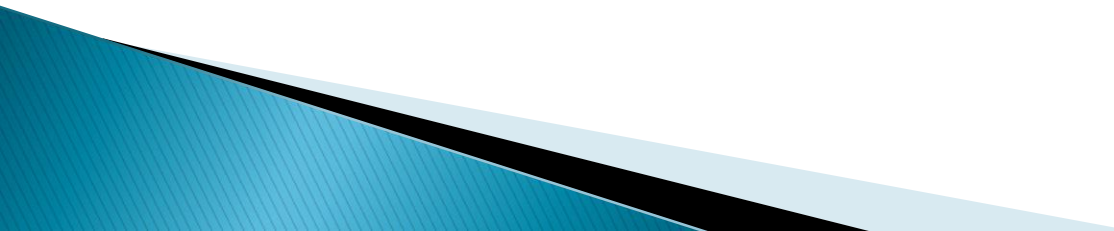
Hierarchical Analysis



- We can clearly see that countries in cluster 0 have highest child mortality rate, imports, inflation and least income, life expectancy, exports thus least GDP which clearly show that these are lagging behind in socio-economic and health terms.
 - On the other hand Cluster 3 and 4 have highest health, GDP, life expect, export-import ratio and least child mortality rate, inflation which shows that they are well ahead in the same.
 - And cluster 2 looking average on all aspects income, GDP, child mortality rate, exports etc which means these are better than lagged countries and worse than ahead countries.
- 

K-Means Analysis



- ▶ The K-Means clustering analysis also speaks the same language as that of Hierarchical clustering analysis.
 - ▶ We can clearly see that countries in cluster 0 have highest child mortality rate, imports, inflation and least income, life expectancy and exports thus least GDP which clearly show that these are lagging behind in socio-economic and health terms.
 - ▶ On the other hand Cluster 3 and 4 have highest health, GDP, life expectancy, export-import ratio and least child mortality rate, inflation which shows that they are well ahead in the same.
 - ▶ And cluster 2 looking average on all aspects income, GDP, child mortality rate, exports etc which means these are better than lagged countries and worse than ahead countries.
- 

Recommendations

It was observed that the K Means could not capture Peru and Tunisia which Hierarchal did. But all the rest of the countries are properly captured.

It will recommend the CEO of the Help International NGO to focus on the countries that belong to Cluster 0 as they were observed to be socio-economically and health wise backward when compared to the whole lot. The list is as follows:

'Afghanistan', 'Albania', 'Armenia', 'Bangladesh', 'Belize', 'Benin', 'Bhutan', 'Bolivia', 'Bosnia and Herzegovina', 'Burkina Faso', 'Cambodia', 'Cameroon', 'Cape Verde', 'Central African Republic', 'China', 'Comoros', 'Cote d'Ivoire', 'Ecuador', 'Egypt', 'El Salvador', 'Eritrea', 'Fiji', 'Gambia', 'Georgia', 'Ghana', 'Guatemala', 'Guinea-Bissau', 'India', 'Jamaica', 'Jordan', 'Kenya', 'Kiribati', 'Lao', 'Madagascar', 'Malawi', 'Mali', 'Morocco', 'Mozambique', 'Namibia', 'Nepal', 'Niger', 'Pakistan', 'Paraguay', 'Peru', 'Philippines', 'Rwanda', 'Samoa', 'Senegal', 'Solomon Islands', 'Sri Lanka', 'Sudan', 'Tajikistan', 'Tanzania', 'Togo', 'Tonga', 'Tunisia', 'Uganda', 'Ukraine', 'Uzbekistan', 'Vanuatu', 'Yemen', 'Zambia'.

- All of them have low export to import ratios, GDPP and also high Inflation which indicates that these are economically backward.
- As these countries has very less income they are unable to spend on health due to which they are ending up losing their new born children and due to the same other aged people also resulting in least Life expectancy thus socially backward.