



# Clustering & PCA Assignment

SRI HARSHA RAVI

# Business Need

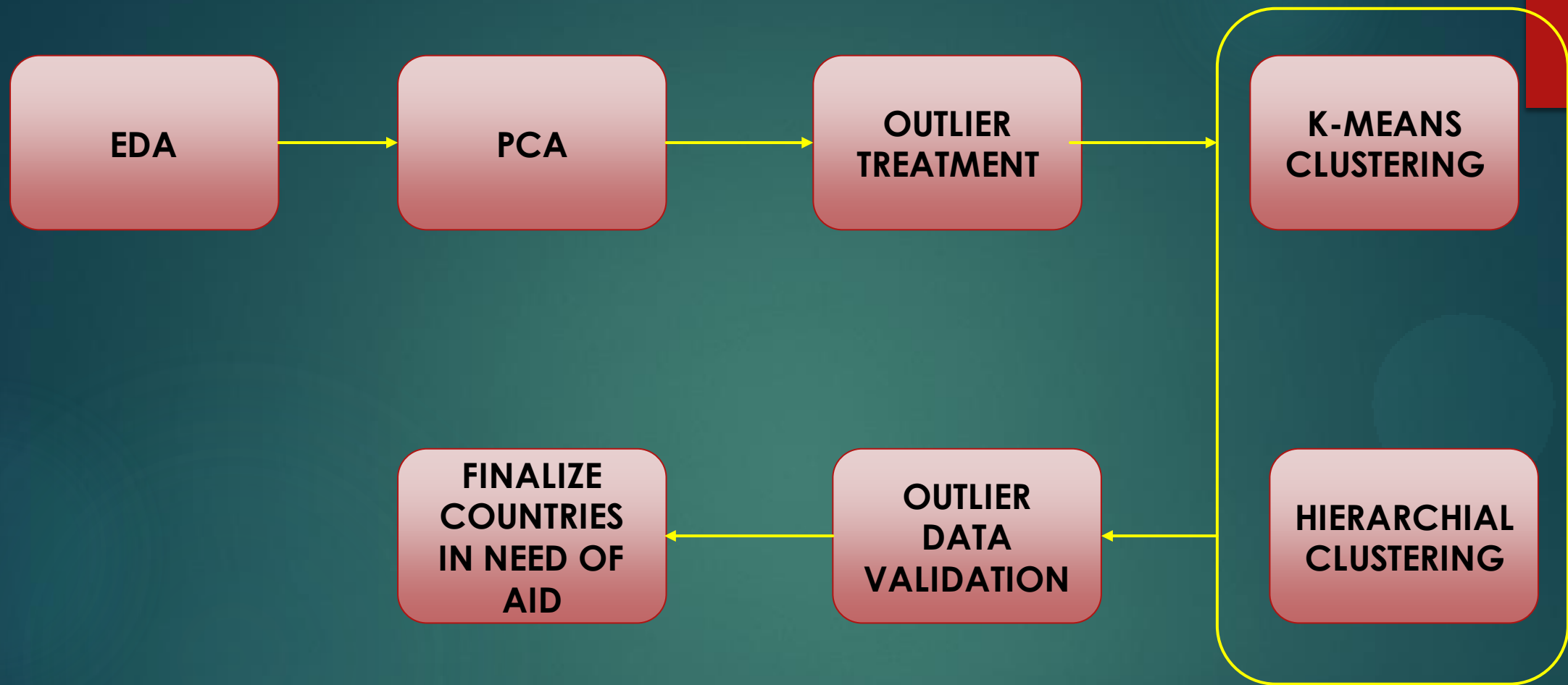
- ▶ Help International NGO has raised 10 Million \$ through awareness drives and funding programs to fight poverty
- ▶ Need to help NGO company to identify the countries which are in at most need to assist them with the raised amount
- ▶ Socio-Economic and health factors of 167 Countries have been collected and need to strategically segregate the countries with the direst need of aid

# Technical Approach

- ▶ Performed EDA by inspecting the data provided, checked for duplicates, null values and the data type validation
- ▶ Checked for the correlation among the variables
- ▶ Performed PCA in order to reduce the dimensionality which in turn takes reduces the correlation
- ▶ Performed the Outlier treatment to perform valid analysis
- ▶ Checked for the Hopkins measure to determine the K-Means compatibility of the data set
- ▶ K-Means clustering with different k-values (determined by Silhouette and Elbow Analysis)

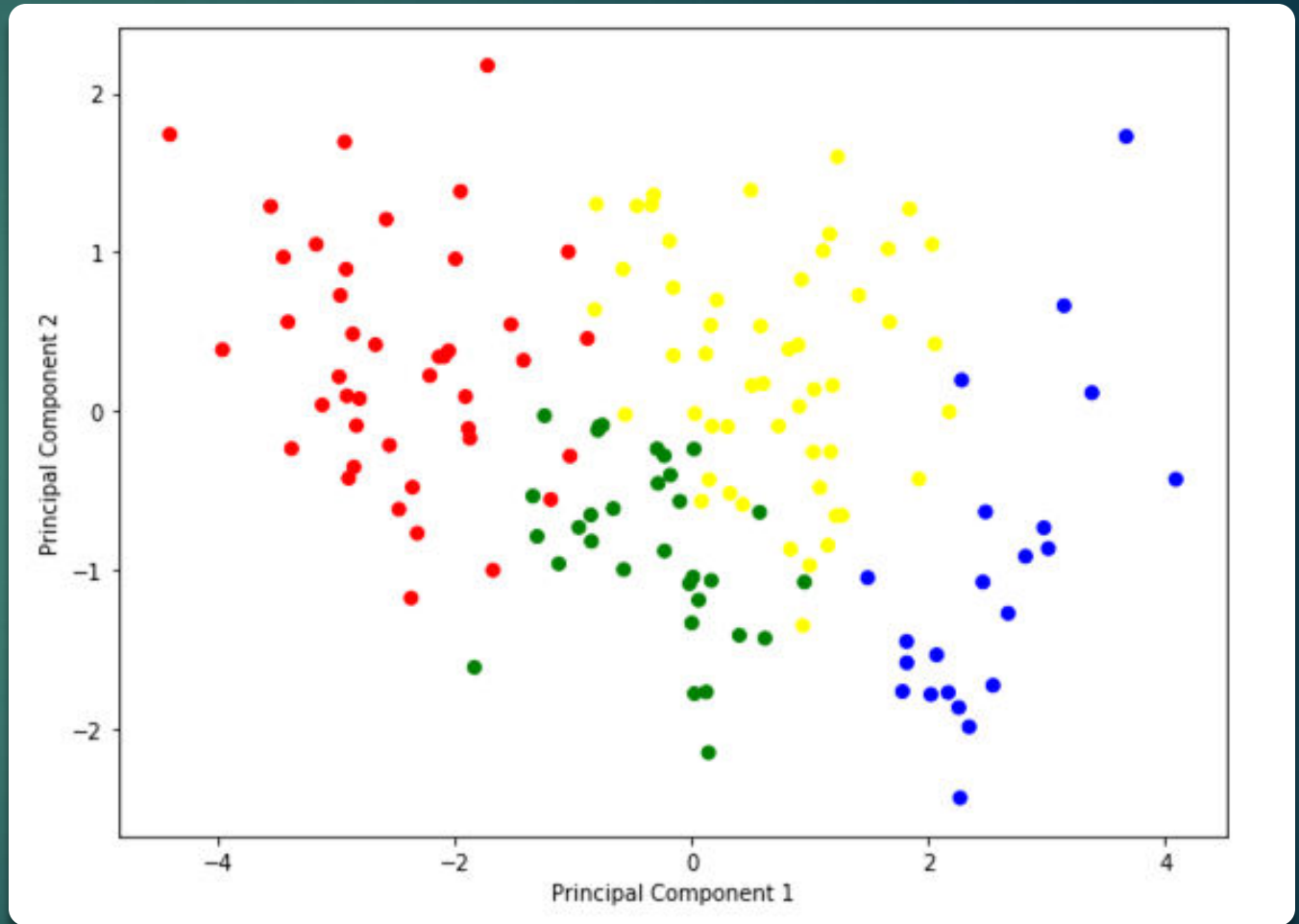
# Technical Approach Contd..

- ▶ Hierarchical Clustering
- ▶ Outlier Data Analysis
- ▶ Finalizing the Countries in need of Aid



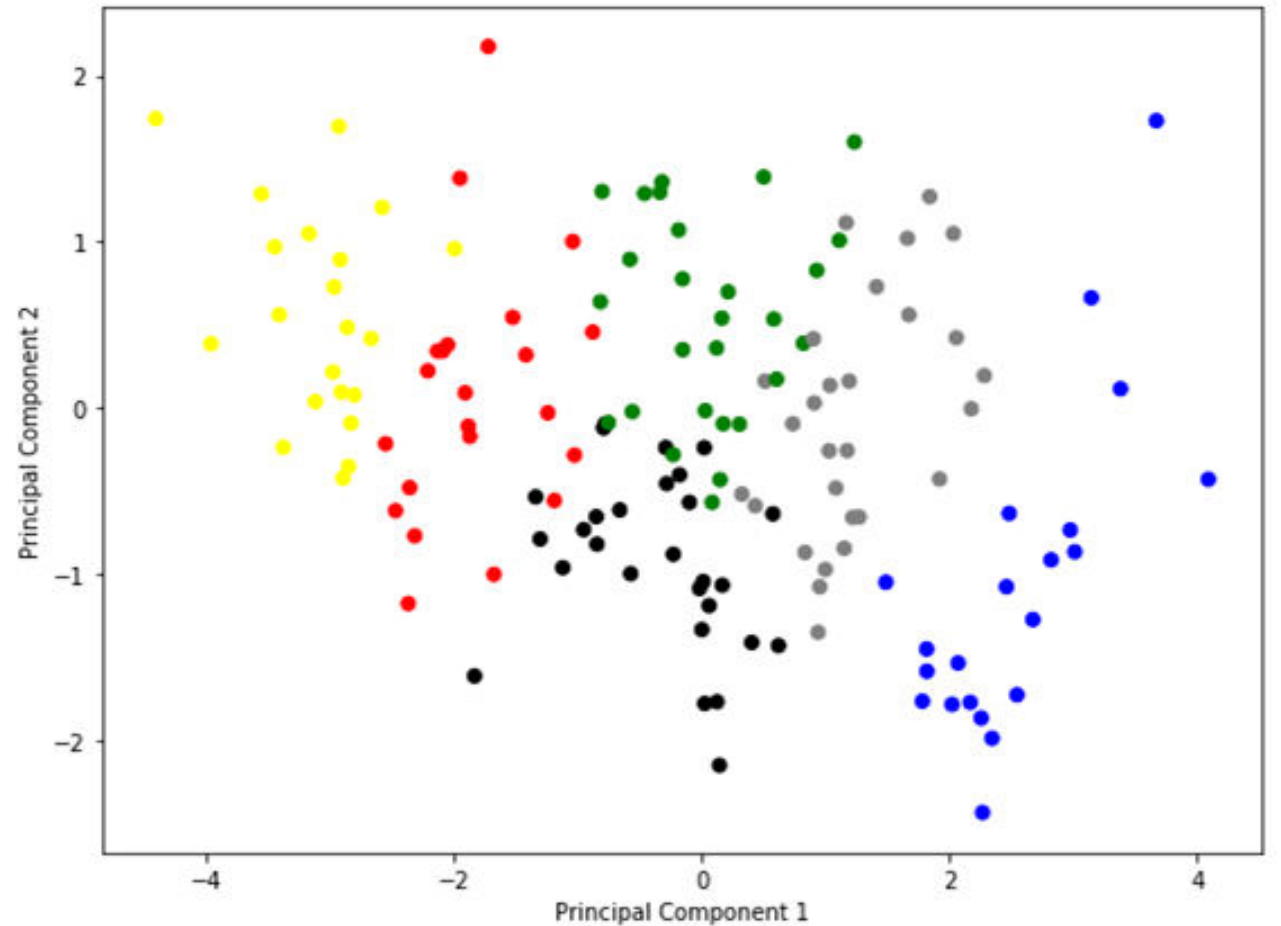
# Scatter Plot b/w PC1 and PC2

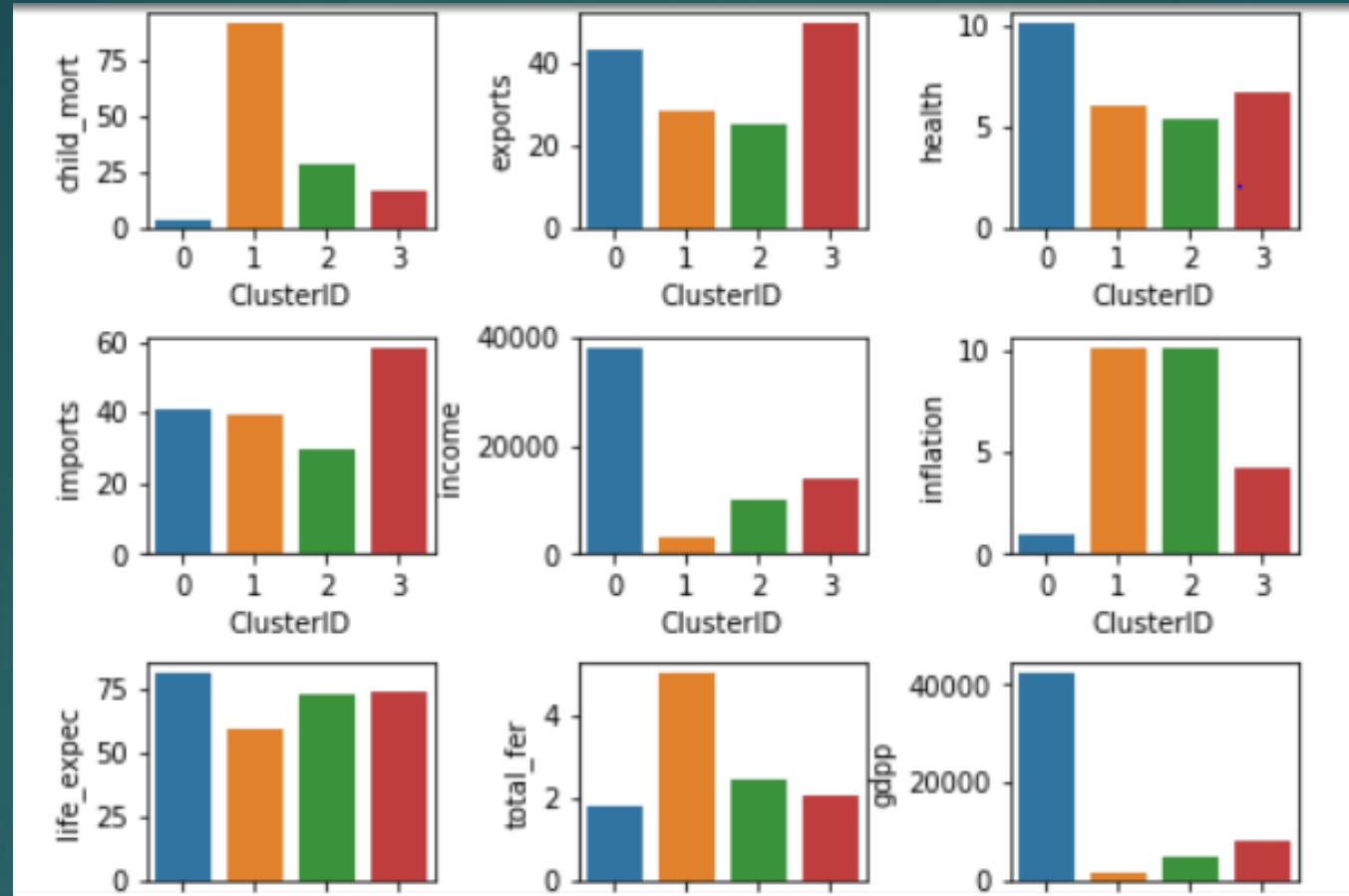
- Each color represents a cluster (k=4)
- Data points towards extreme left would be in need of the Aid



# Scatter Plot b/w PC1 and PC2

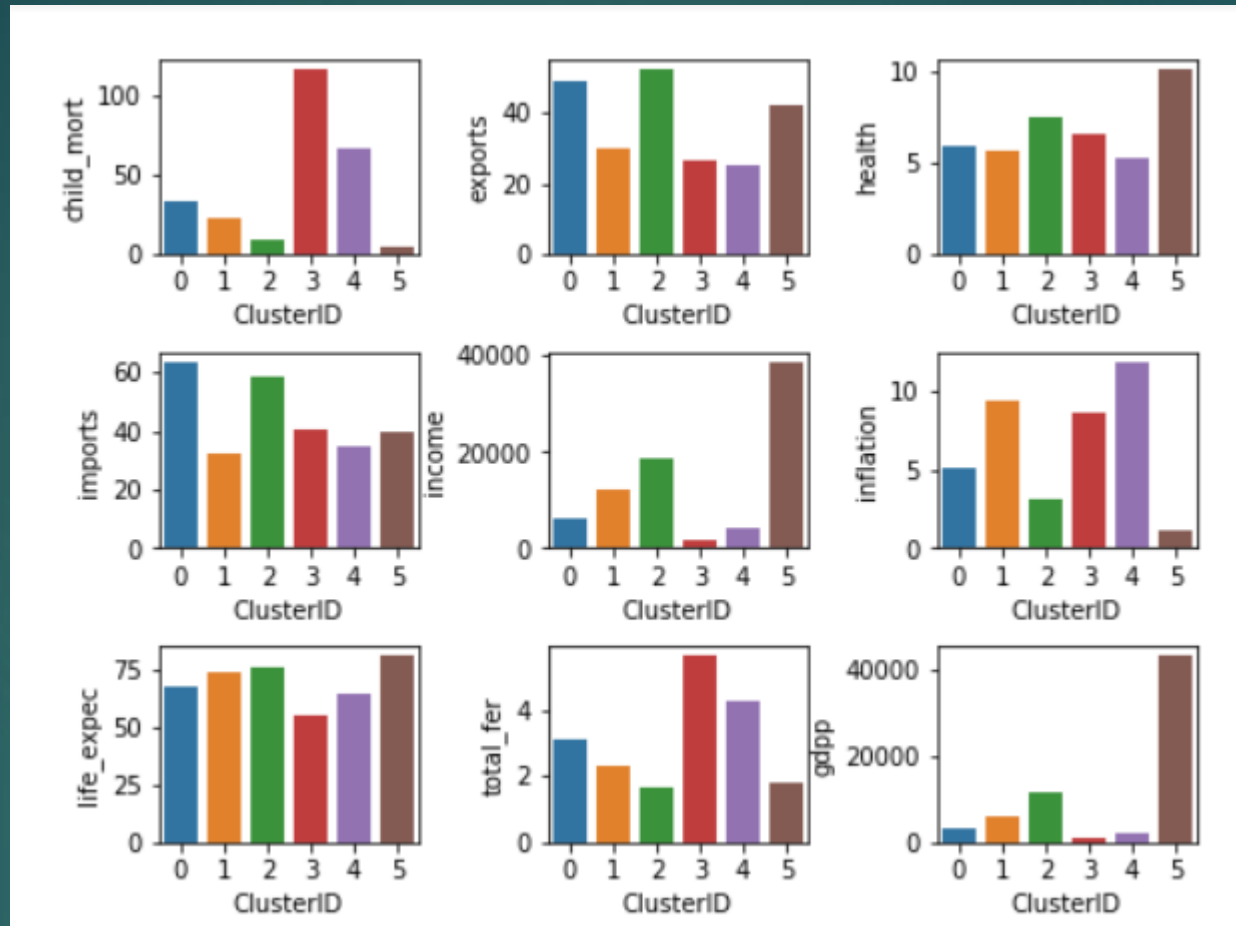
- Each color represents a cluster (k=6)
- Data points towards extreme left would be in need of the Aid



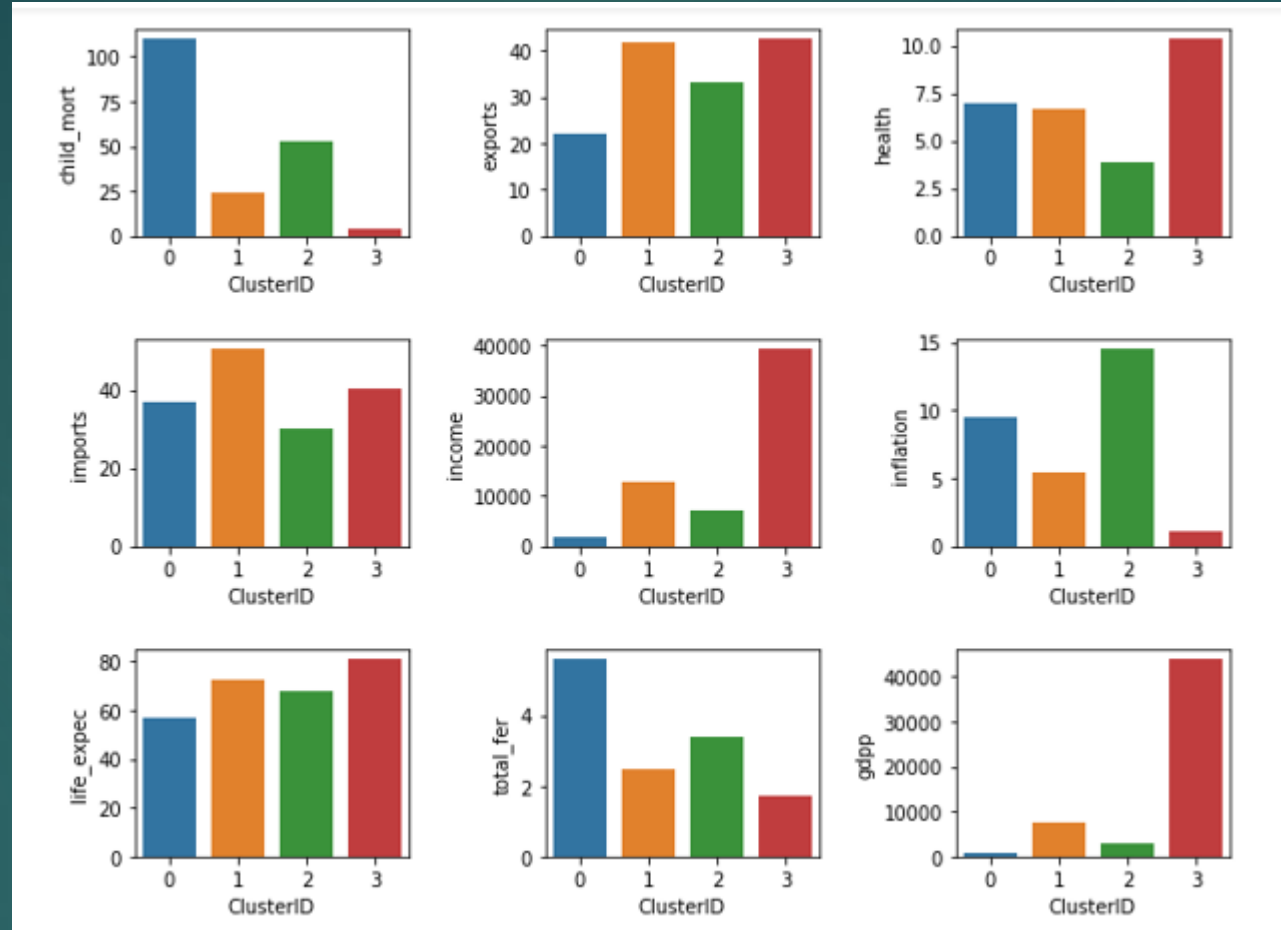


Above graphs are the output of the  $k = 4$  under K-Means  
Clearly, Countries in ClusterID = 1 are to be funded

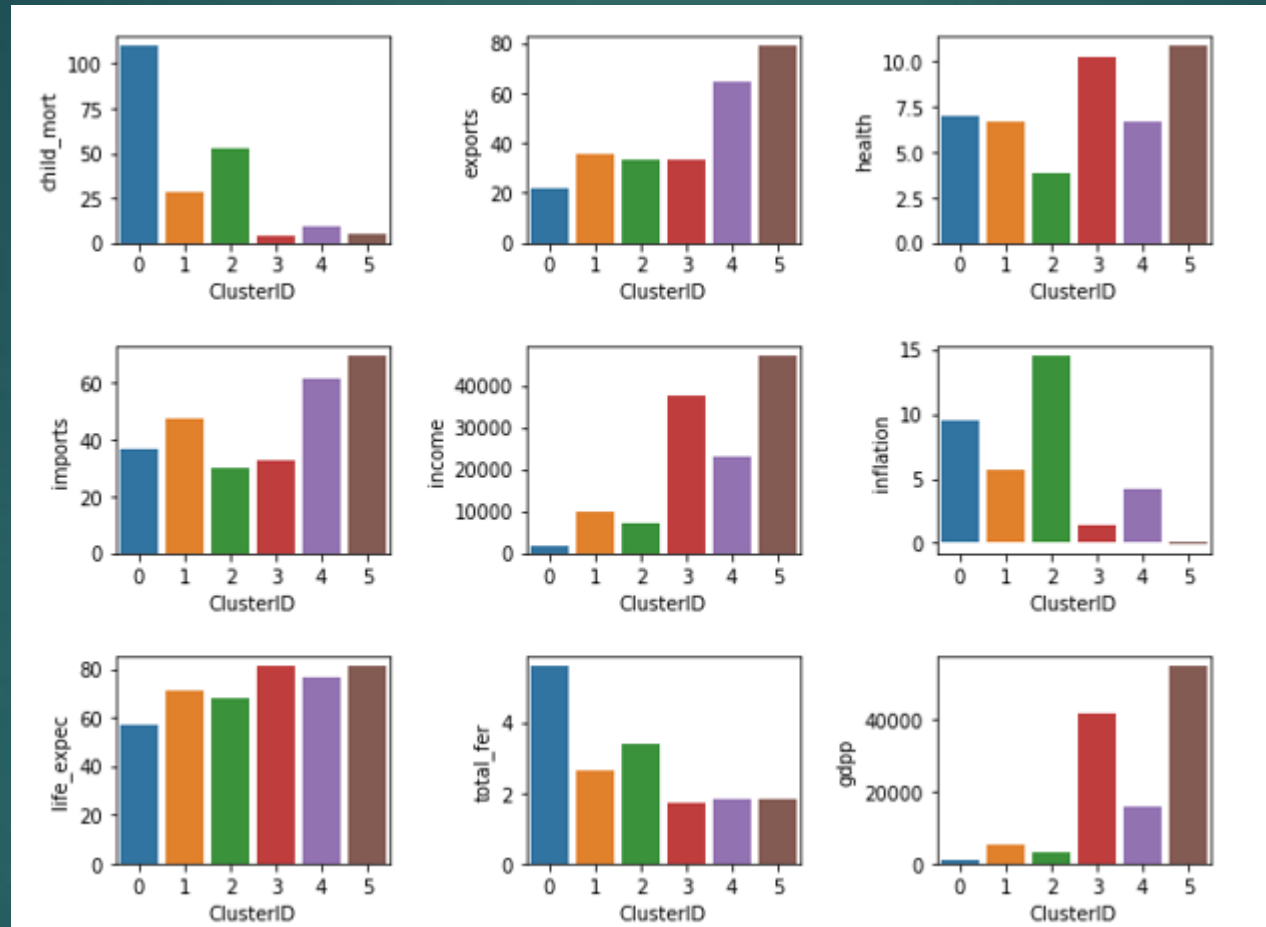






Above graphs are the output of the  $k = 4$  under K-Means  
Clearly, Countries in ClusterID = 3 are to be funded



Above graphs are the output of the cut =4 under Hierarchical Clustering  
Clearly, Countries in ClusterID = 0 are to be funded



Above graphs are the output of the cut=6 under Hierarchical clustering  
Clearly, Countries in ClusterID = 0 are to be funded

- 
- 
- After analyzing the Clusters with different k values of K-Means and Cut values of Hierarchical Clustering and also the Outlier data we had identified a set of Countries
  - 28 Countries would be needing first preference in gaining the assistance from the NGO
  - Few more Countries also would need assistance but may not be as needy as the identified Countries

# Countries listed below would definitely need 1<sup>st</sup> preference in receiving the Aid from 'Help International' NGO

## Aid needed Countries

Afghanistan  
Benin  
Burkina Faso  
Burundi  
Cameroon  
Central African Republic  
Chad  
Congo, Dem. Rep.  
Cote d'Ivoire  
Guinea

## Aid needed Countries

Guinea-Bissau  
Haiti  
Malawi  
Mali  
Mozambique  
Niger  
Rwanda  
Sierra Leone  
Sudan  
Tanzania

## Aid Needed Countries

Timor-Leste  
Togo  
Uganda  
Zambia  
Kiribati  
Lesotho  
Liberia  
Nigeria