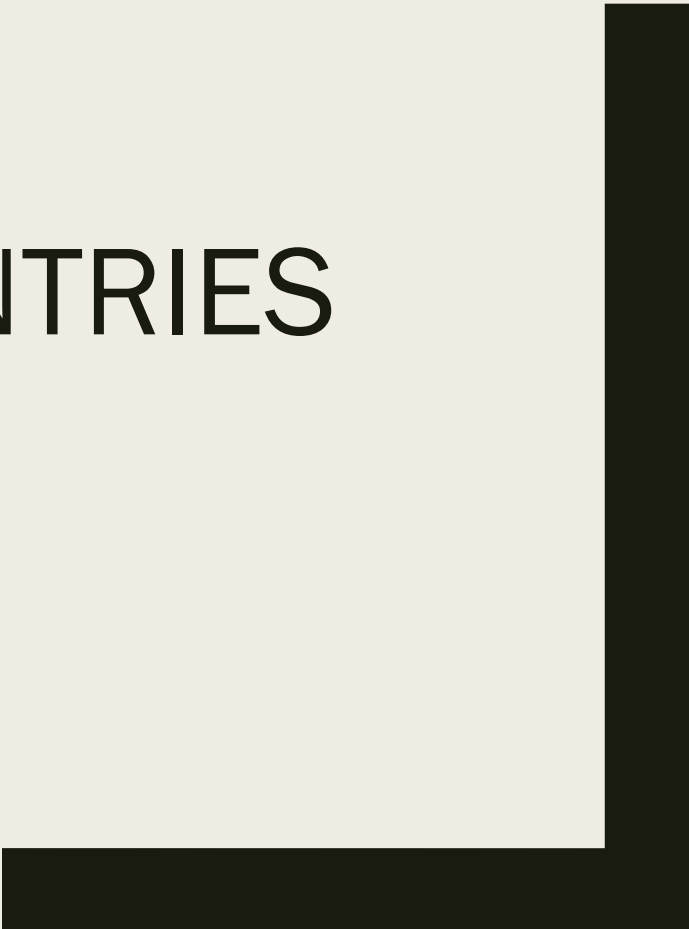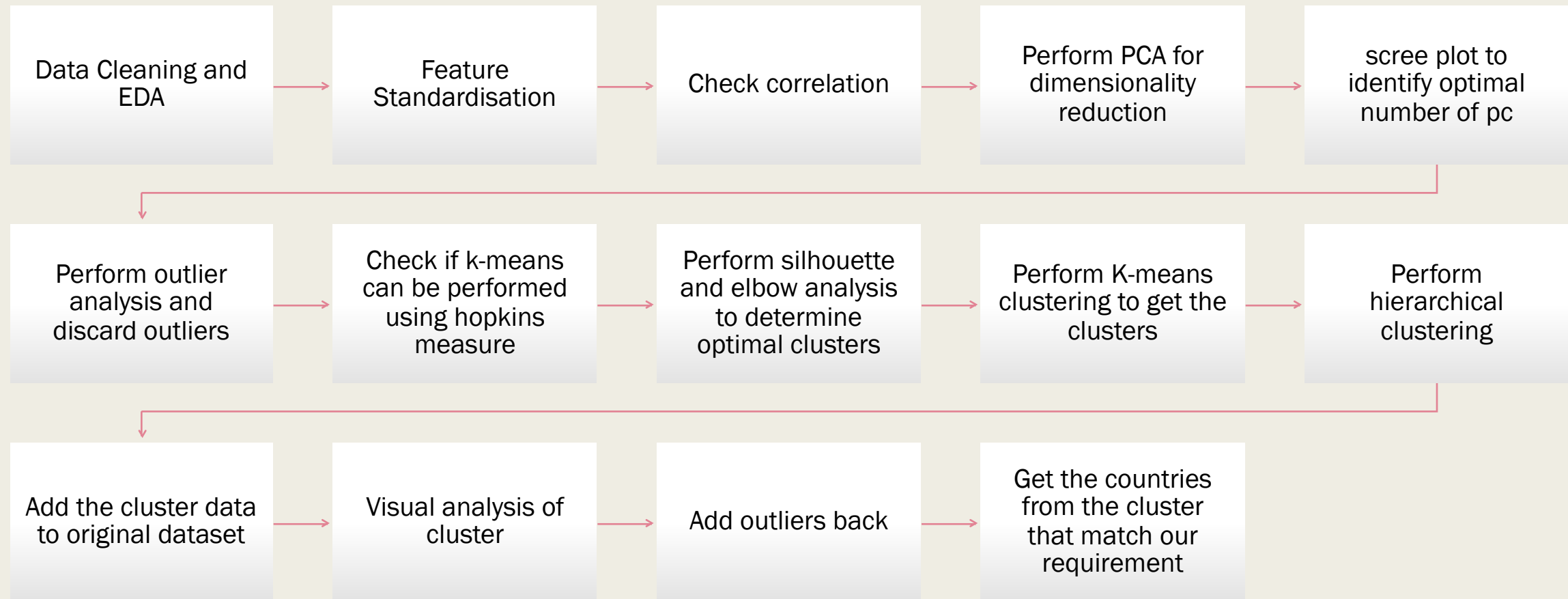# CLUSTERING OF COUNTRIES
HELP-INTERNATIONAL

Presented By,

Subarna Saha

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 project that included a lot of awareness drives and 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.

METHODOLOGY

Data Cleaning and EDA → Feature Standardisation → Check correlation → Perform PCA for dimensionality reduction → scree plot to identify optimal number of pc

Perform outlier analysis and discard outliers → Check if k-means can be performed using hopkins measure → Perform silhouette and elbow analysis to determine optimal clusters → Perform K-means clustering to get the clusters → Perform hierarchical clustering

Add the cluster data to original dataset → Visual analysis of cluster → Add outliers back → Get the countries from the cluster that match our requirement

RESULTS (PCA)

- We have high collinearity in the data. Therefore we perform PCA to reduce dimensions and get non collinear columns without losing useful data.

- From scree plot we get optimal number of PC as 5

- Obtain PCA dataset

- Check collinearity, PCA has helped to reduce dimension and collinearity in data.

- Income-gdp-life_expectancy has higher weightage in PC1

- Imports-exports has higher weightage in PC2
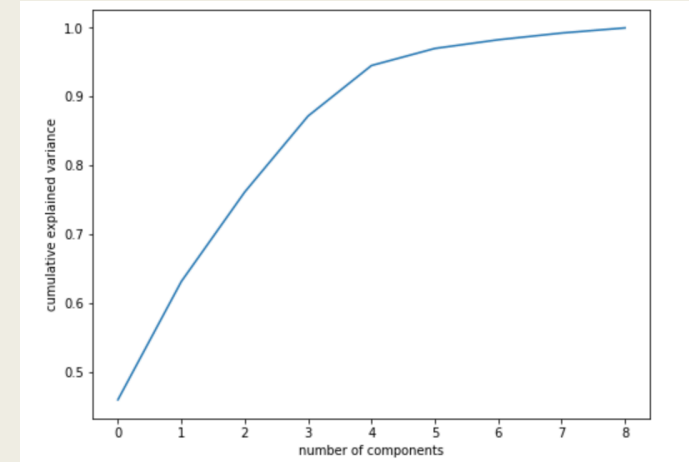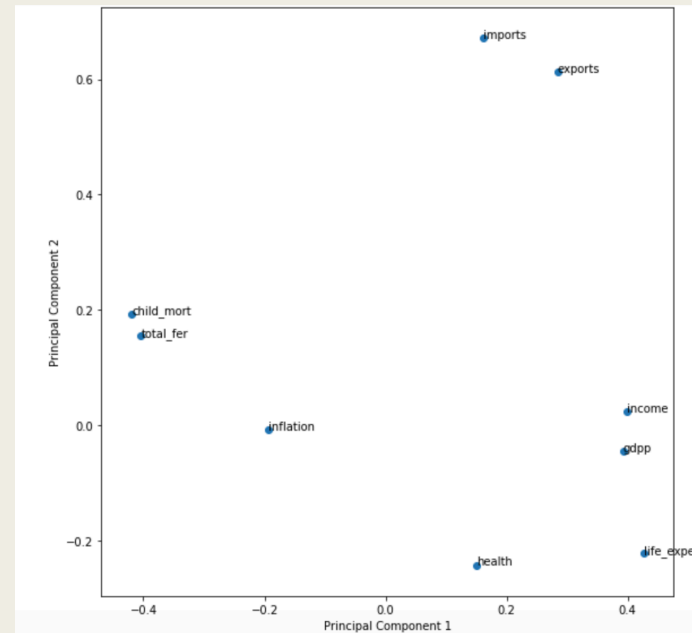
|	Feature	PC1	PC2	PC3	PC4	PC5
0	child_mort	-0.419519	0.192884	-0.029544	0.370653	-0.168970
1	exports	0.283897	0.613163	0.144761	0.003091	0.057616
2	health	0.150838	-0.243087	-0.596632	0.461897	0.518000
3	imports	0.161482	0.671821	-0.299927	-0.071907	0.255376
4	income	0.398441	0.022536	0.301548	0.392159	-0.247150
5	inflation	-0.193173	-0.008404	0.642520	0.150442	0.714869
6	life_expec	0.425839	-0.222707	0.113919	-0.203797	0.108220
7	total_fer	-0.403729	0.155233	0.019549	0.378304	-0.135262
8	gdpp	0.392645	-0.046022	0.122977	0.531995	-0.180167
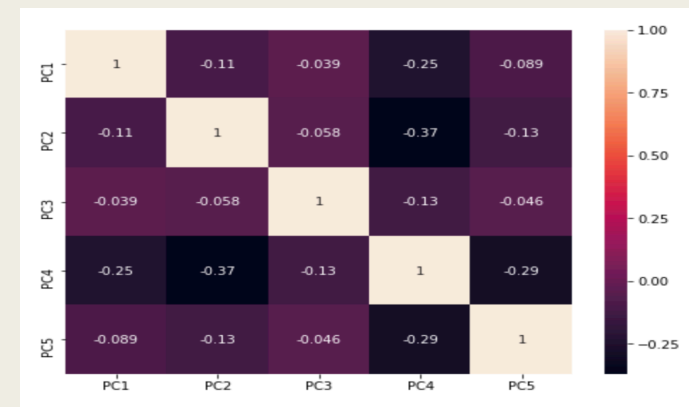
RESULTS (Clustering)

- Perform silhouette and elbow analysis to determine optimal clusters

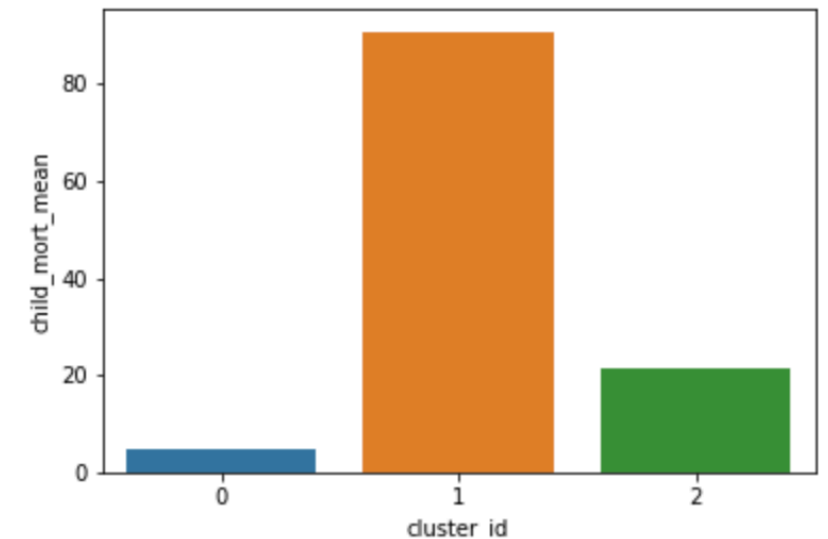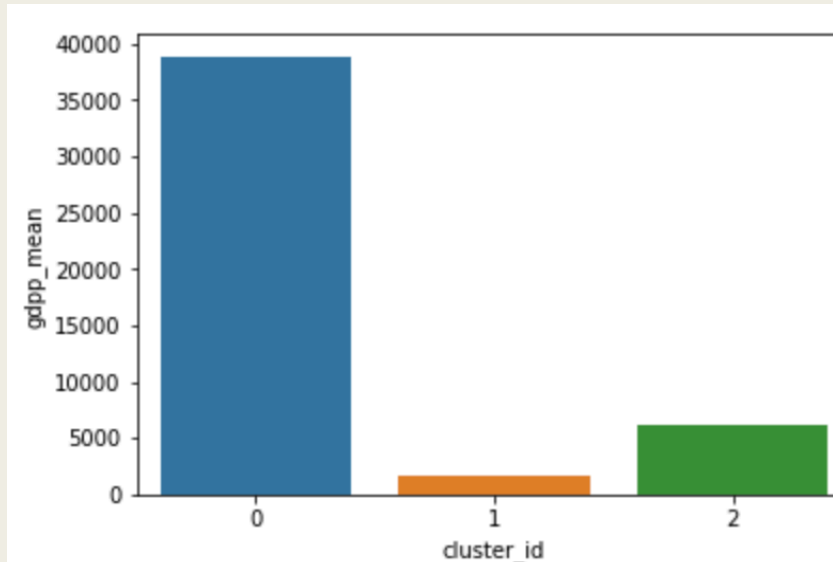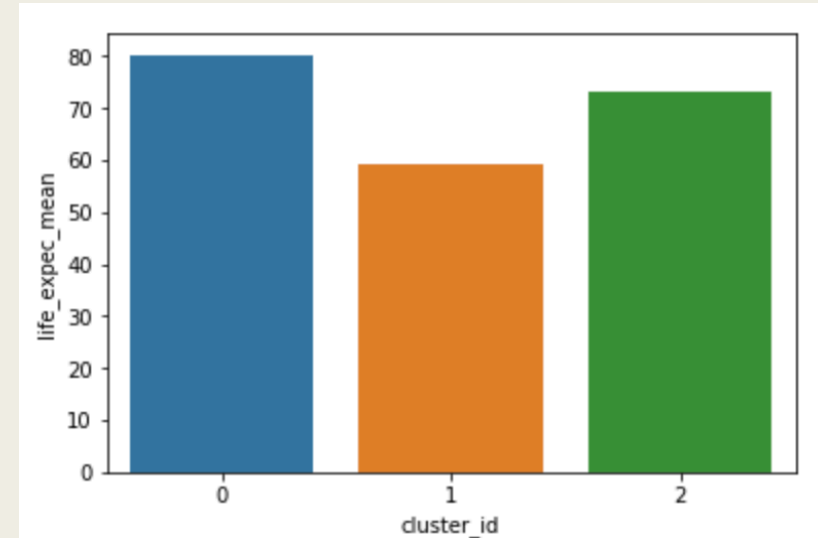- Optimal number of clusters are from 3 to 5
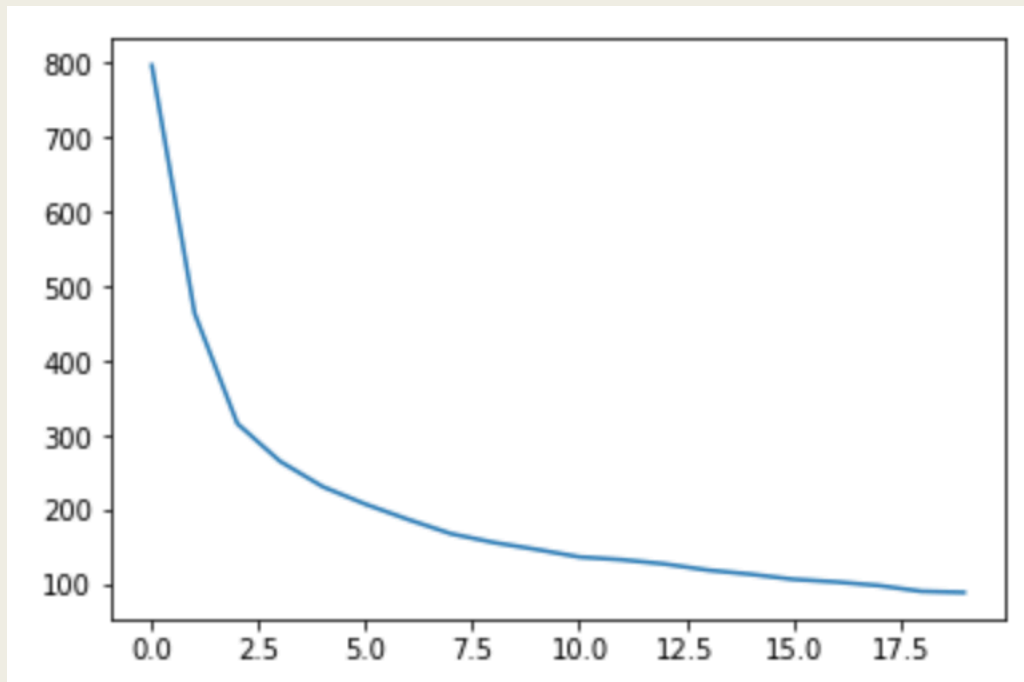
Elbow Method
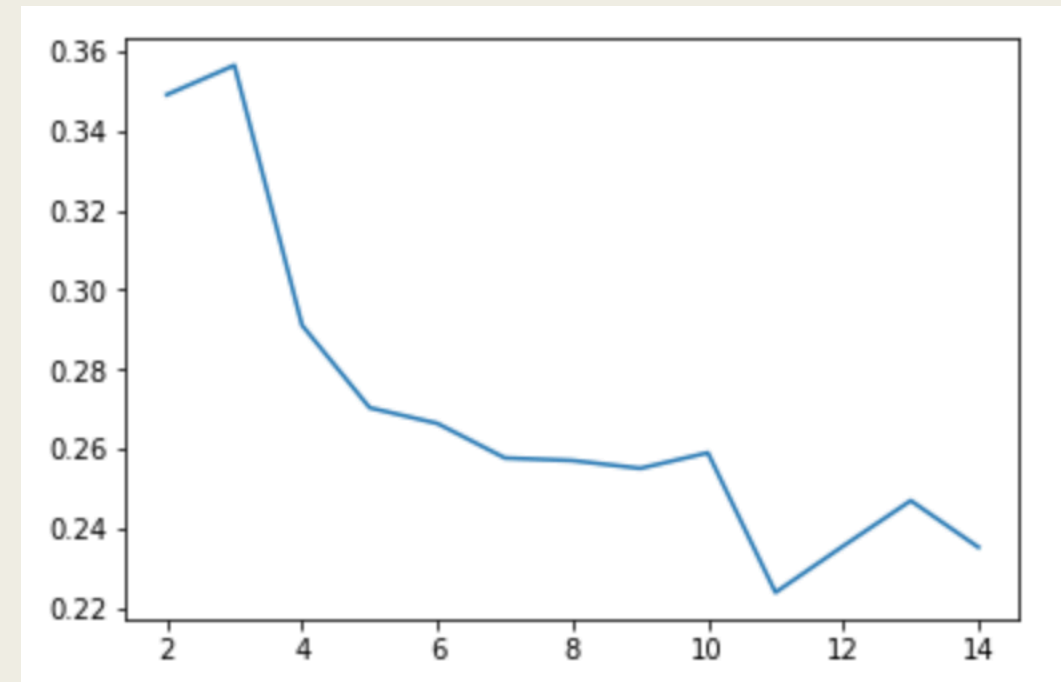
Silhouette Method

RESULTS(CLUSTERING) K-MEANS

- ■ Visualize the clusters on few factors

- ■ From the plots we see that cluster 1 fulfills the criterion for funding.

- ■ It has low gdpp

- ■ Low life expectancy

- ■ And very high child mortatlity

RESULTS(CLUSTERING)
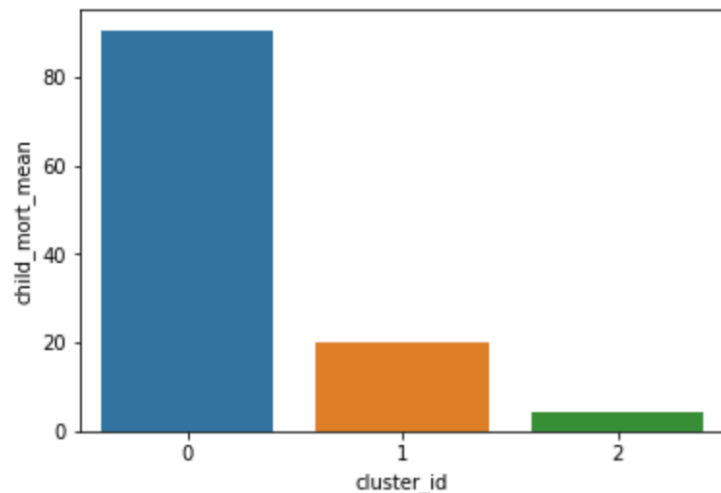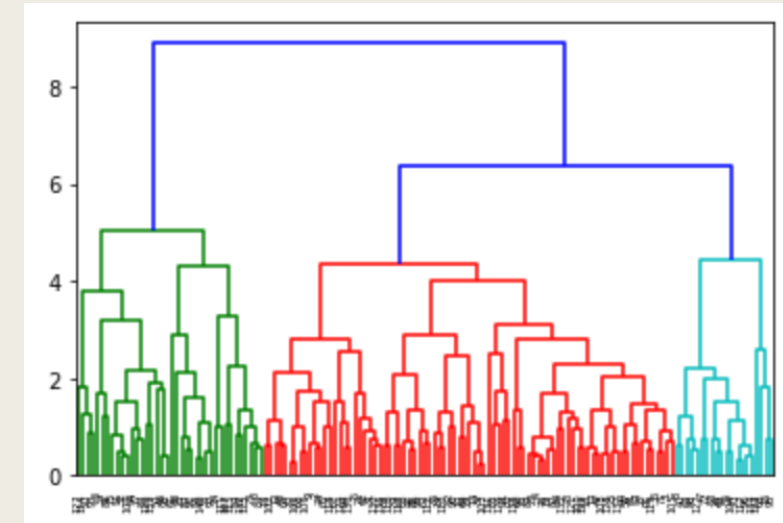# Hierarchical Algorithm

- We cut our dendogram at 3

- Visualize the clusters on few factors

- From the plots we see that cluster 0 fulfills the criterion for funding.

- It has low gdpp

- Low life expectancy

- And very high child mortatlity
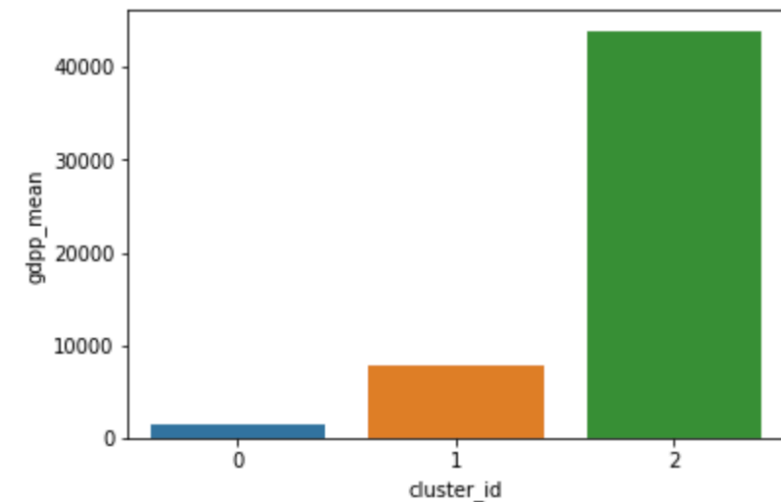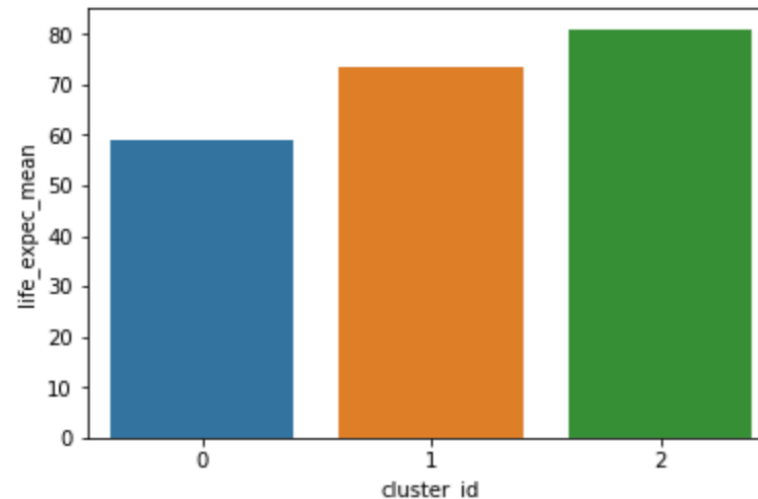
RESULTS
(Selection of countries)

- Based on the clusters formed we come to a list of countries that need funding for development.

- Afghanistan, Angola, Botswana, Burkina Faso, Burundi, Chad, Comoros, Congo-Dem. Rep, Congo-Rep, Cote d'Ivoire, 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, Togo, Uganda, Yemen, Zambia

- Also we can consider the below countries: Benin, Cameroon, Central African Republic, Lesotho, Liberia (were removed as outlier)

CONCLUSION

- After doing our analysis on data, performing pca and then clustering the countries based on the different socio economical factors we have suggested a list of countries that require funding the most.