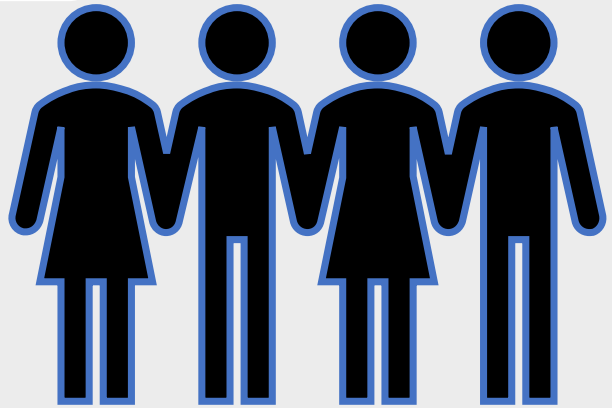
HELP

working towards a better world



Approach to find Countries that are in dire need of aid

EDA : Find out important variables to find out countries that needs HELP. Outliers are not treated since they form a different cluster altogether

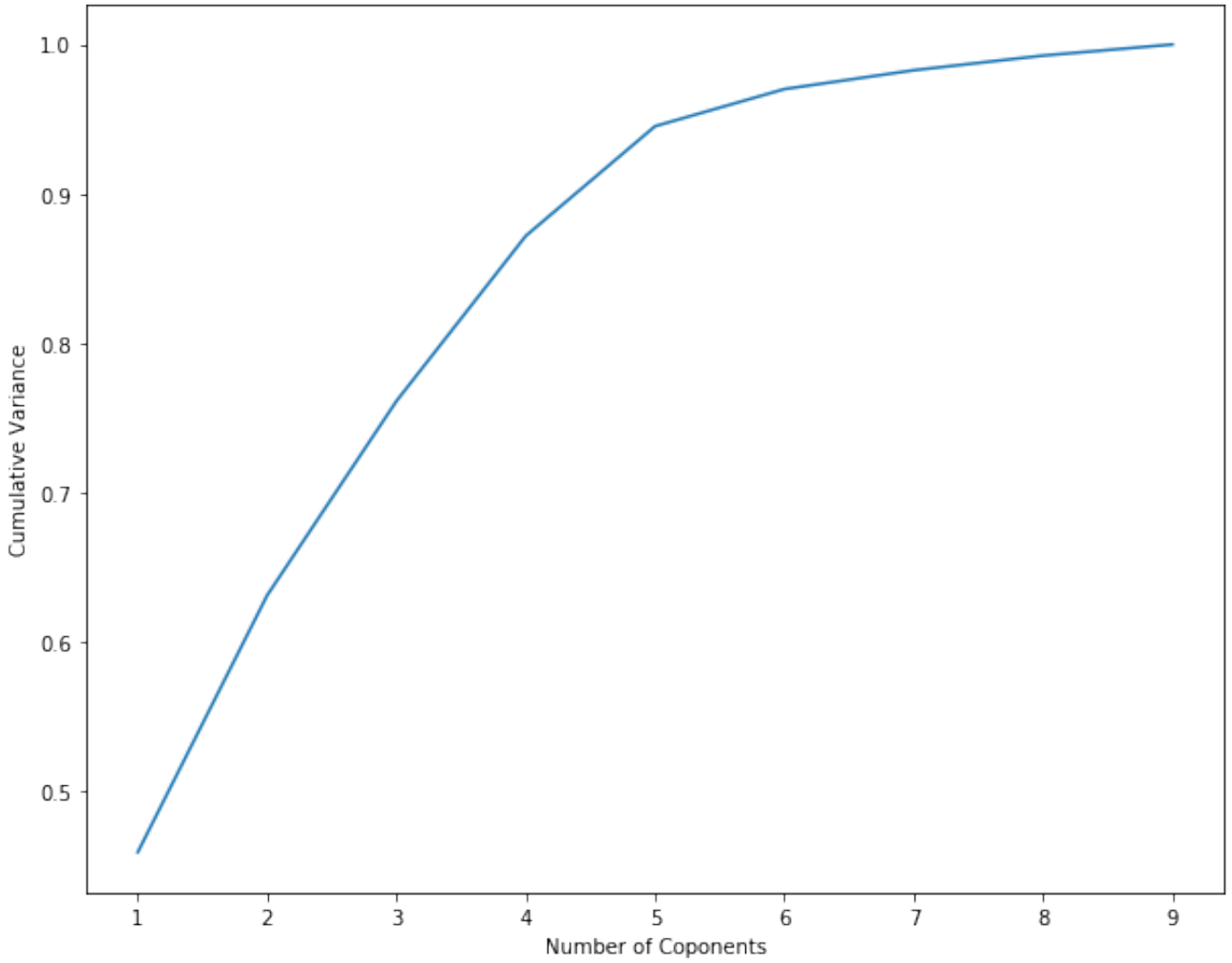
PCA : Helps in dimensionality reduction

Clustering : Classify similar data points based on principal components we got from PCA. Apply these on 3 important variables we got from EDA

gdpp
income
child_mortality

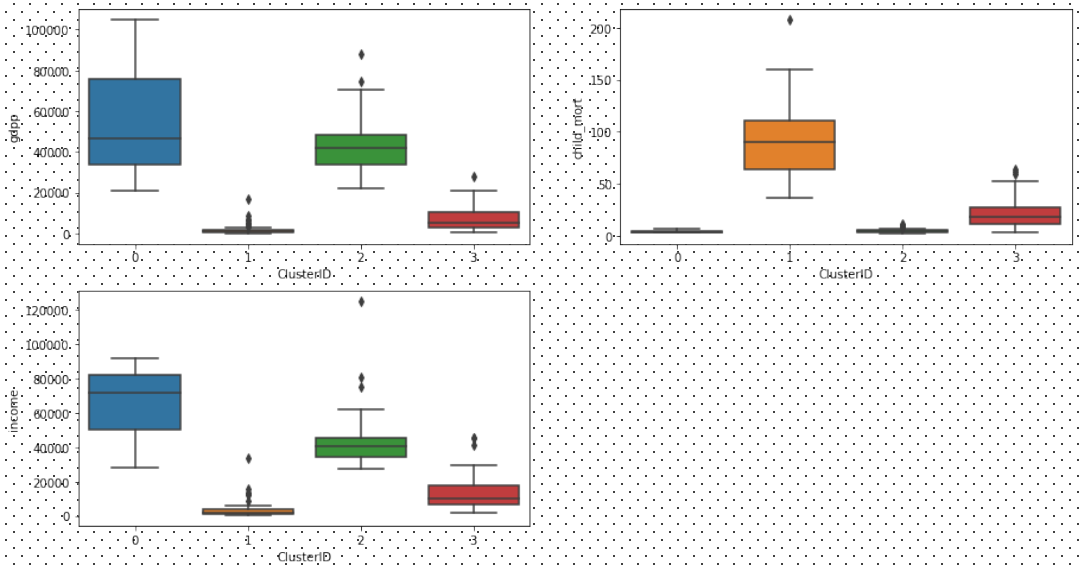
PCA

- We had various factors for each country
- Thus creating multiple dimensions of data
- After PCA number of factors explaining the variability got reduced
- This plot shows the variability explained by each components
- By looking at the variability it’s good to have 5 components



Clustering

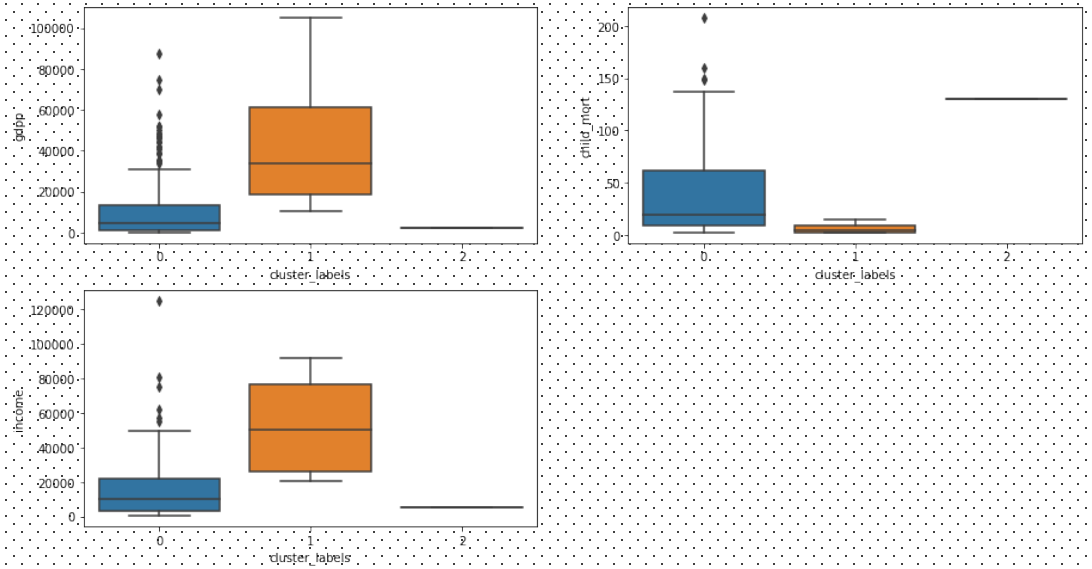
K -Means



We have 4 clusters explaining various combination of 3 variables

Cluster 1: This focuses on population which has low income, low gdpp and high child_mortality

Heirarchial



We have 3 clusters explaining various combination of 3 variables

Cluster 0: This focuses on population which has low income, low gdpp and high child_mortality

Conclusion

By looking at 2 methods of clustering we can find Countries which need help based on low income, low gdpp and high child mortality.

- Congo, Dem. Rep.
- Liberia
- Burundi
- Niger
- Central African Republic
- Sierra Leone
- Mozambique
- Malawi
- Guinea