HELP Country Data Research

- 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 funding programs, they have been able to raise around $ 10 million. 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.

Problem Statement

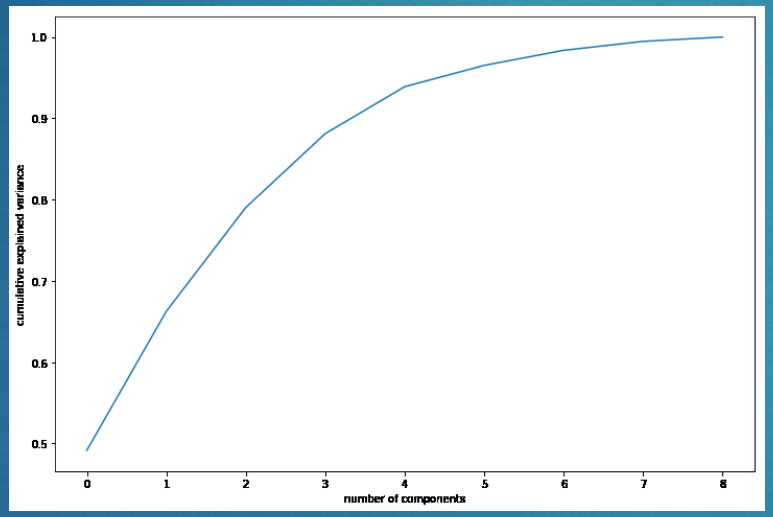
- Problem statement is to help the CEO of HELP International to find the countries which are the most underdeveloped and which need the direst of help among the countries based on certain factors and decide how to use the money strategically and effectively.
- The objective of this analysis is to categorize the countries using some socio-economic and health factors that determine the overall development of the country.
- Using the provided dataset with countries and respective socio-economic factors and health factors ,it is required to find the countries that need the direst help.

Approach taken for Analysis

- The approach chosen to achieve the final list of countries that need the most help is Unsupervised learning(**Clustering**)
- The countries are clustered based on the socio economic factors into different clusters and the clusters are later analysed to find the cluster which represents the countries that are underdeveloped.
- To do so we take the help of Principal Component Analysis to reduce the number of predictor variables.
- Two types of methods are used which are K means and Hierarchical Clustering.
- To find number of principal components and clusters Scree plots and Elbow method are used respectively

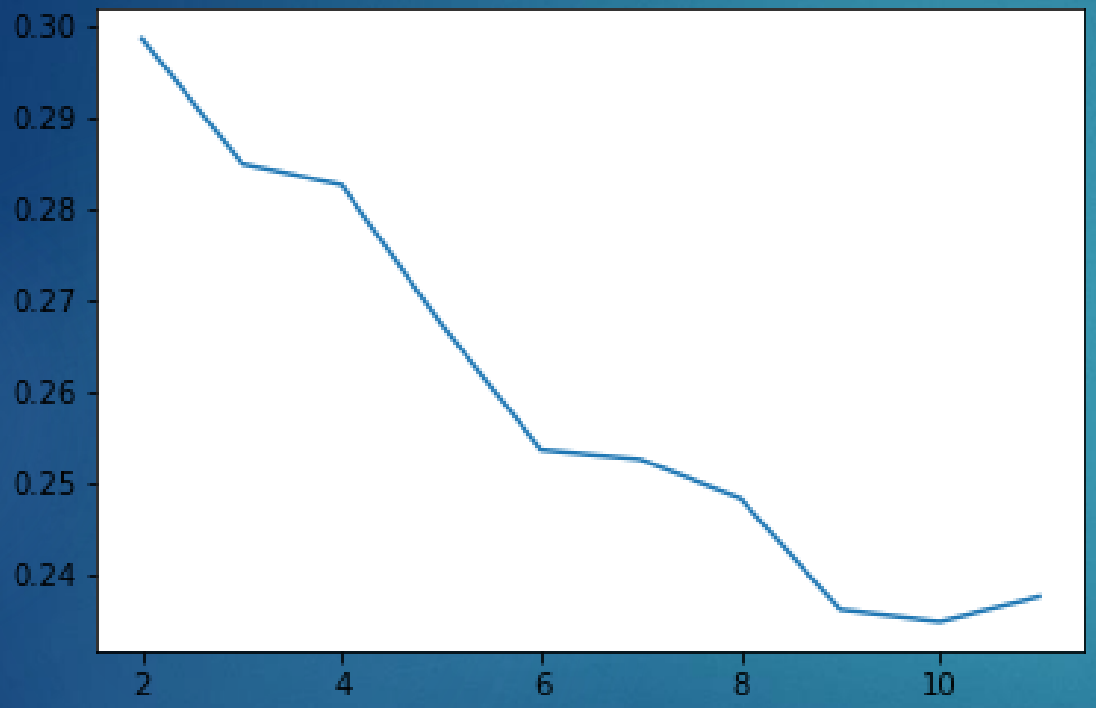
Results of Analysis

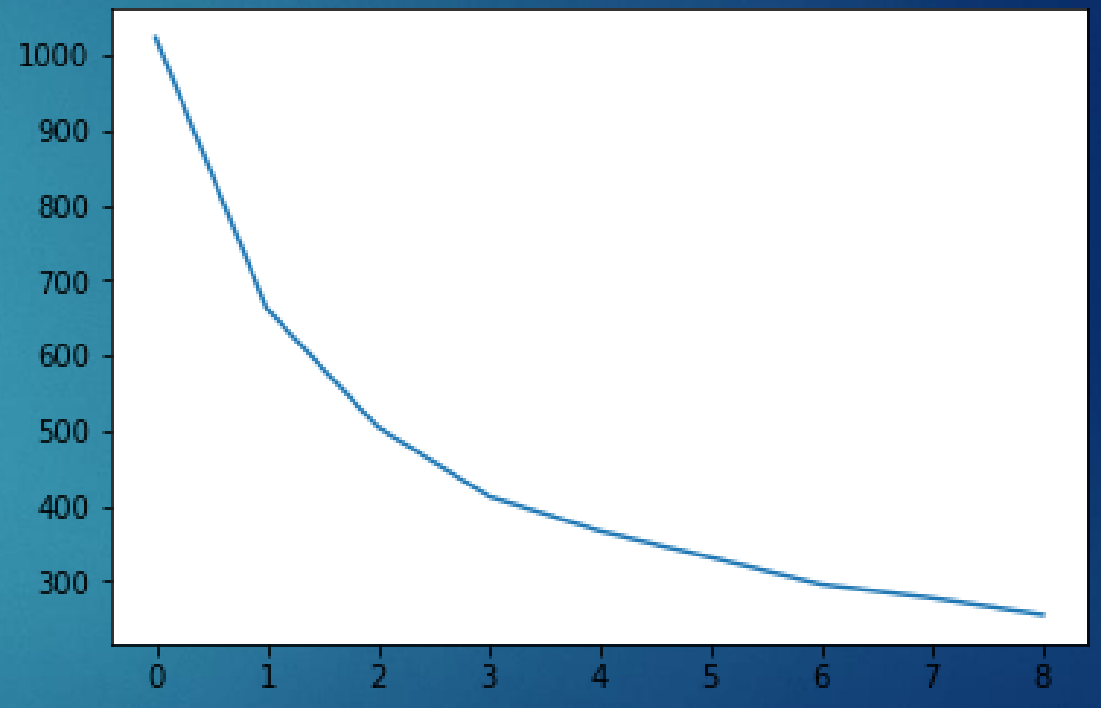
- To decide the number of principal components ,Scree plots were used .



- As evident from the Scree plot 5 components are able to explain 93.8% of the variance.

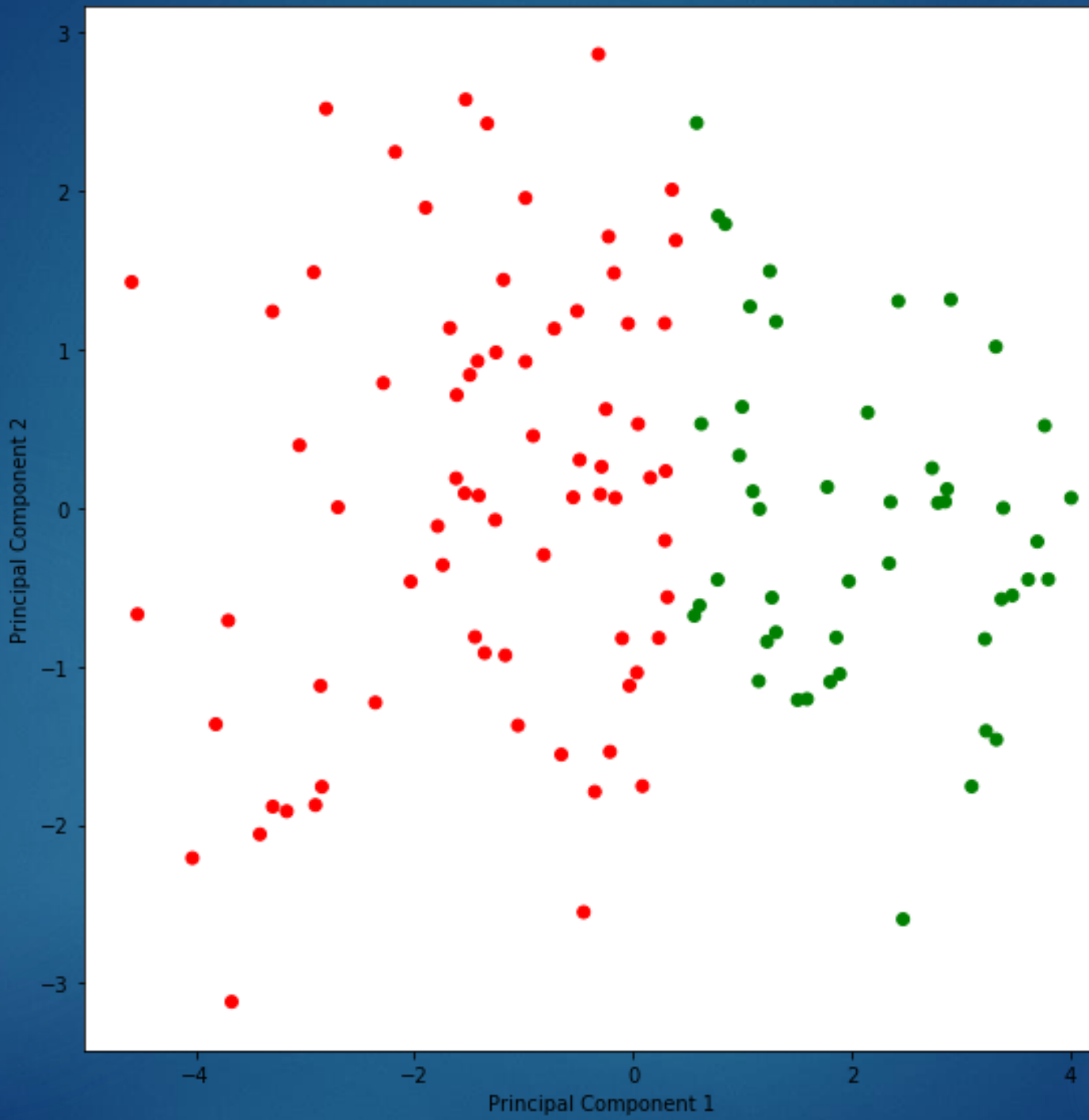
- Thus we take 5 principal components and run PCA again to find the final dataset with only 5 principal components.
- We then run Hopkin's statistic to find the ability of the dataset to form clusters and find that it gives a percentage of 72%.
- Next to find the number of clusters to be formed from the dataset we use the Silhouette method and elbow method as shown below.



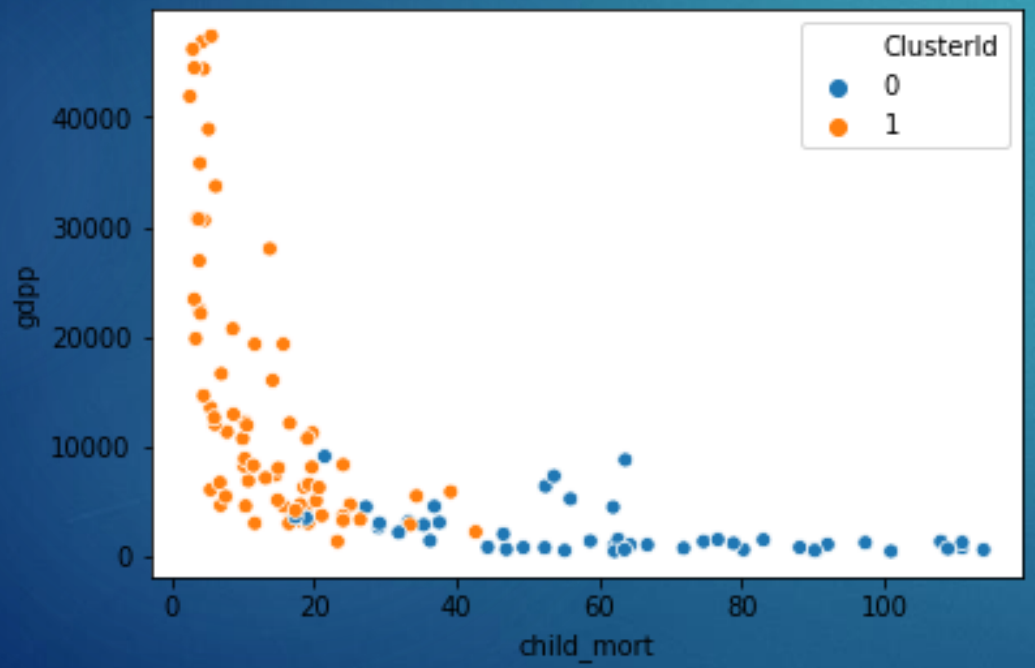


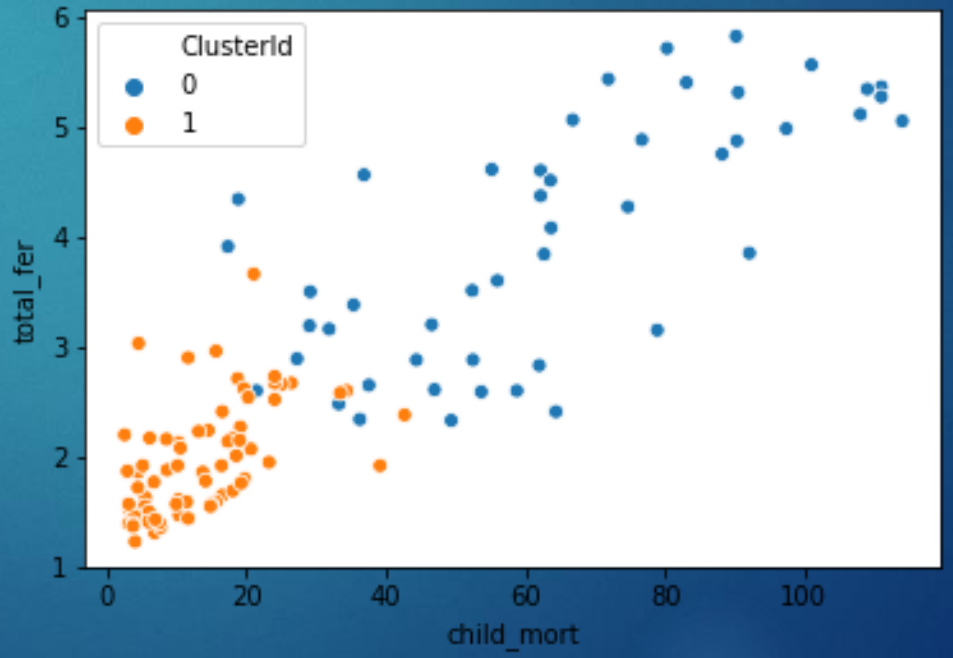
- It is evident that the number clusters to be used are 2 from both the above plots

- Using K means and Hierarchical Clustering methods with 2 clusters we find the final dataset divided into **2 Clusters.**
- Now we need to find the cluster which represents the underdeveloped countries, to do so we plot the original variables by representing each cluster with a different colour.
- First we plot the PCA components against each other to check if we have formed clusters by using different colours for each cluster
- The following plot is obtained when two principal components are plotted.


Principal Component 2
Principal Component 1
3
2
1
0
−1
−2
−3
−4
−2
0
2
4

- We can see 2 good clusters being formed , now we plot two of the original variables against each other with respective clusters , to obtain the following





- By Comparing the two clusters we can see a clear pattern. One cluster has low GDP per capota ,high child mortality rate and high fertility rate.
- Thus we can come to a conclusion that Cluster 0 represents countries that are underdeveloped.

Conclusion

- Thus we conclude cluster 0 has underdeveloped countries and they are the one's that have gdpp lesser than 1000 :
- **Madagascar,Mozambique,Malawi,Eritrea,Togo,Guinea-Bissau,Afghanistan,Gambia,Rwanda,Nepal,Guinea,Tanzania,Tajikistan,Benin,Bangladesh,Comoros,Cambodia,Kenya,Myanmar,Senegal**
- The countries that need the maximum help are :
- **Madagascar**
- **Mozambique**
- **Malawi**
- **Eritrea**
- **Togo**
- it should also be noted that the countries that need the maximum help belong to Africa thus that's where maximum amount of help needs to be provided