

# Clustering Assignment

[K means and Hierarchical Clustering]

Clustering Countries based on Socioeconomic factors for financial aid by NGO

Author: Sourabh Somvanshi

# Problem Statement

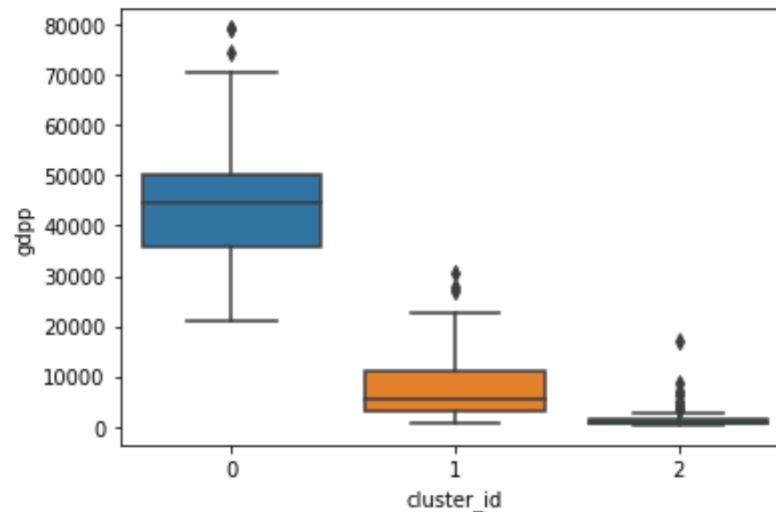
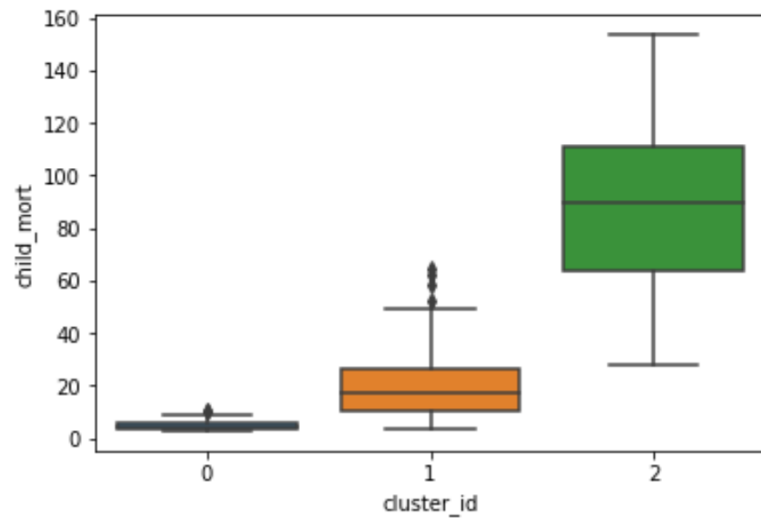
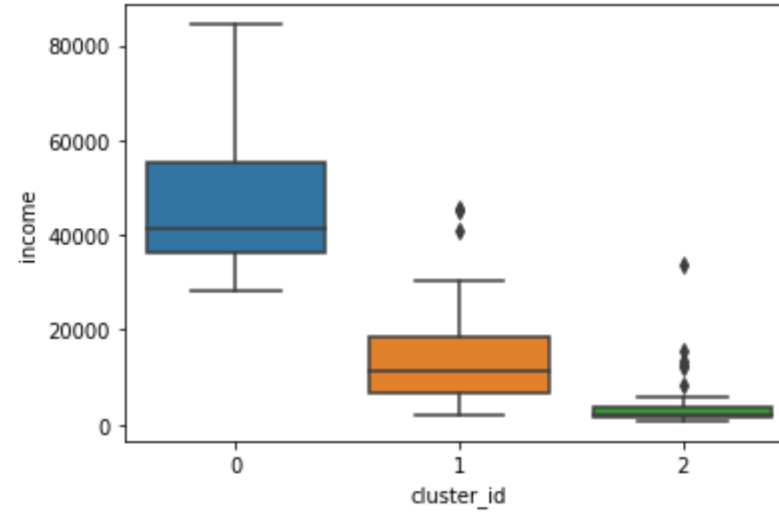
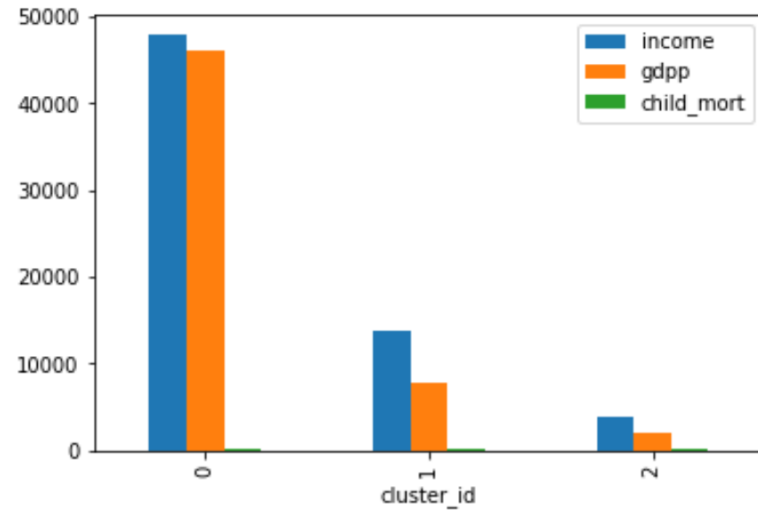
- ▶ HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.
- ▶ After the recent funding programs, 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.
- ▶ The problem statement is to identify the countries in need by using clustering to analyse the socioeconomic and health factors.

# Analysis Approach

- ▶ Read and validate the input data
- ▶ Perform EDA to understand data trends
- ▶ Outlier Analysis and Outlier Treatment
- ▶ Hopkins Statistics to check is clustering is possible
- ▶ Using Hierarchical clustering to identify the optimal cluster value
- ▶ Use Silhouette and Elbow method to validate the optimal cluster values.
- ▶ Use K-means cluster method to build the final cluster model
- ▶ Identify appropriate cluster for financial aid using cluster mean method
- ▶ Analyse the cluster statistics with other clusters.
- ▶ Choose top 10 countries from the final cluster based on higher mortality rate, lower income and lower gdpp
- ▶ Final Report

# Cluster Summary

We identified that cluster 2 has countries with most child mortality and least income and gdp.



# Cluster Statistics

- ▶ Below is the cluster statistics for the 3 clusters created using K-means method.
- ▶ As we can see that cluster 2 has the most child mortality and least income and gdpp.
- ▶ We have picked the top 10 countries from this cluster.

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
cluster_id									
0	4.989655	25405.359310	4253.879655	21316.695862	47784.413793	2.906731	80.479310	1.754138	46068.137931
1	20.547778	3477.250726	528.894338	3589.291996	13804.333333	7.131624	73.393333	2.242444	7808.577778
2	90.335417	879.097657	114.821765	827.327888	3901.010000	10.608604	59.239583	4.992083	1911.400833

# Final List of Countries

- We concluded on the top 10 list of countries from the final cluster(2) based on the cluster mean values of child\_mort, income and gdpp
- We have filtered countries with:
  - highest child mortality
  - lowest income
  - lowest gdpp

Top 10 Countries
Sierra Leone
Haiti
Chad
Central African Republic
Mali
Nigeria
Niger
Angola
Congo, Dem. Rep.
Burkina Faso