



# Clustering Assignment

---

BY :- AJAY BENDALE

# Objectives & Problem statement

---

## Objective:-

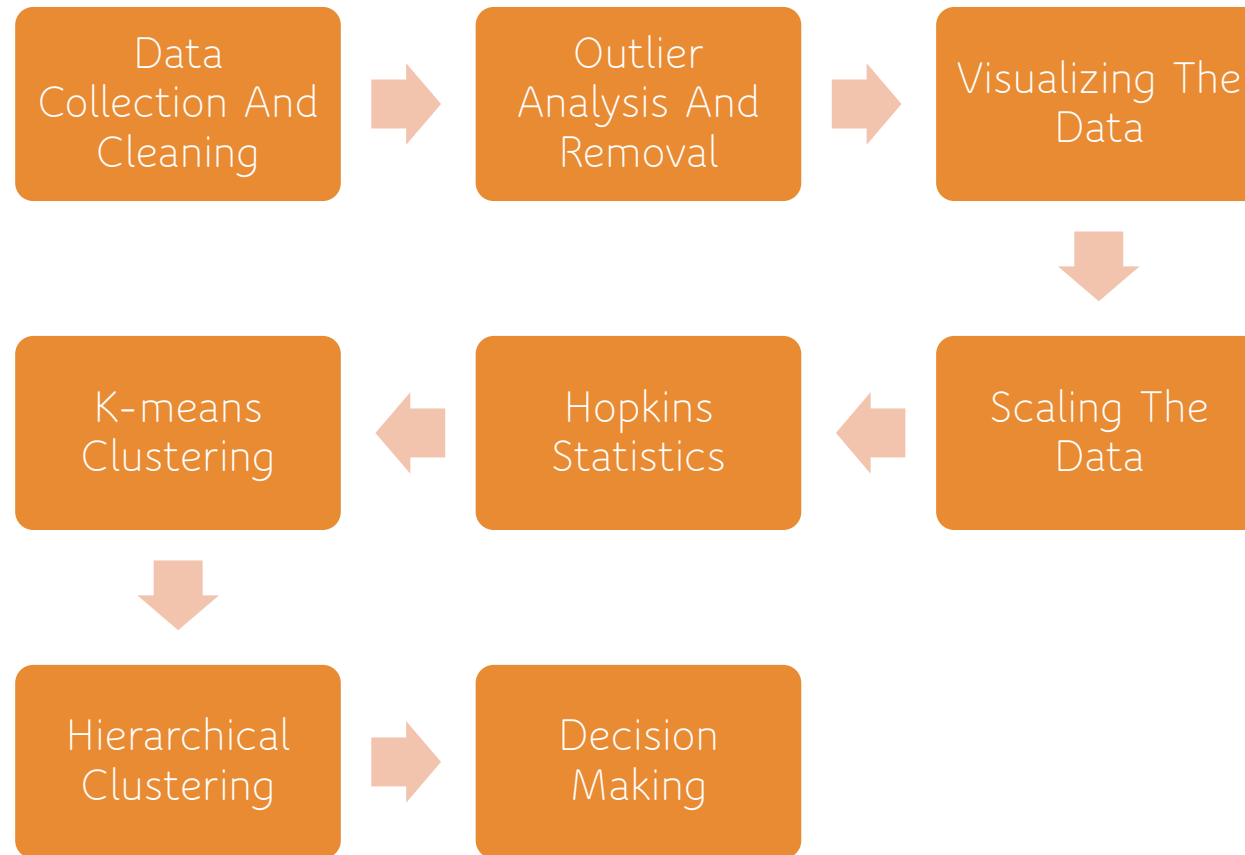
- We, HELP International humanitarian NGO, committed to fight poverty and provide the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. We run a lot of operational projects from time to time, along with advocacy, drives to raise awareness as well as for funding purposes.

## Problem statement:-

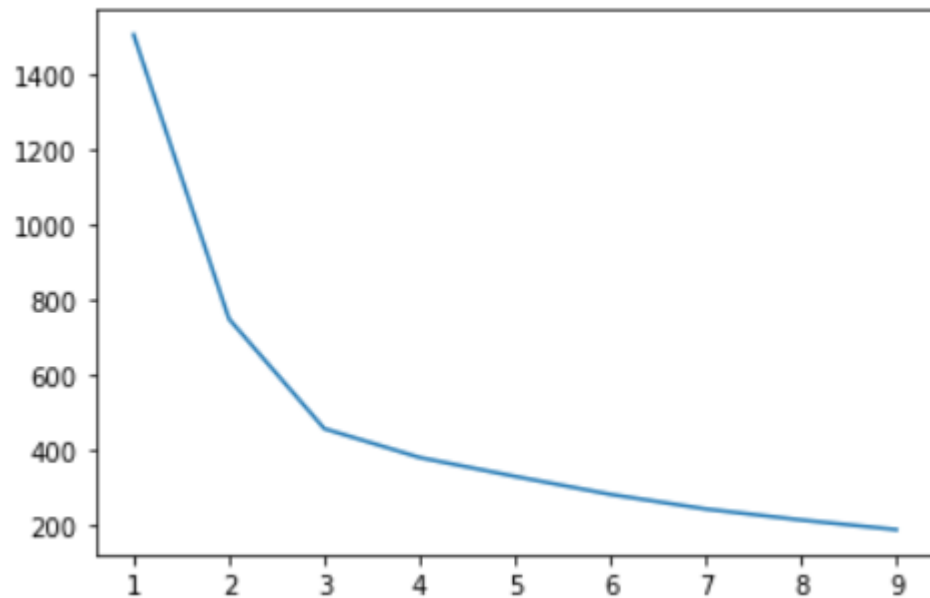
- During the recent funding programs, we have been able to raise around \$ 10 million. As an analyst, we have to come up with the countries list that are in the direst need of aid.

# Analysis methodology

---



# K-means Clustering



silhouette score

For n\_clusters=2, the silhouette score is 0.48447902987180524

For n\_clusters=3, the silhouette score is 0.4203919640787207

For n\_clusters=4, the silhouette score is 0.3832442500525417

For n\_clusters=5, the silhouette score is 0.2994963024128736

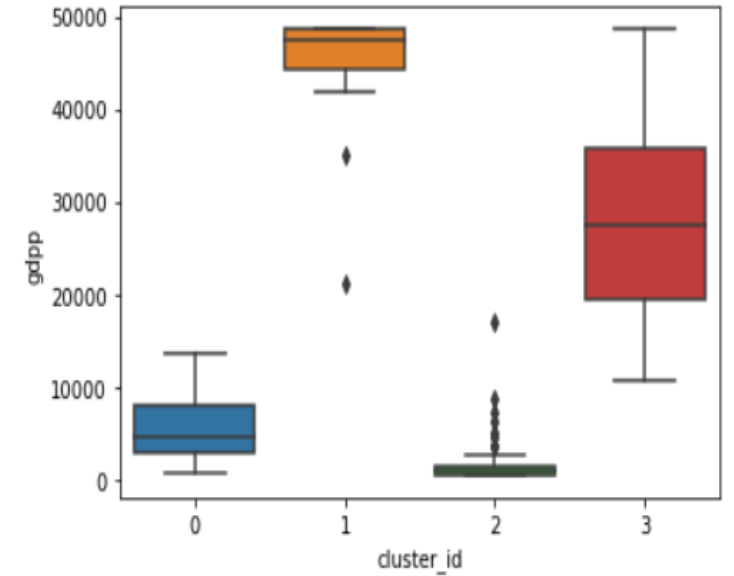
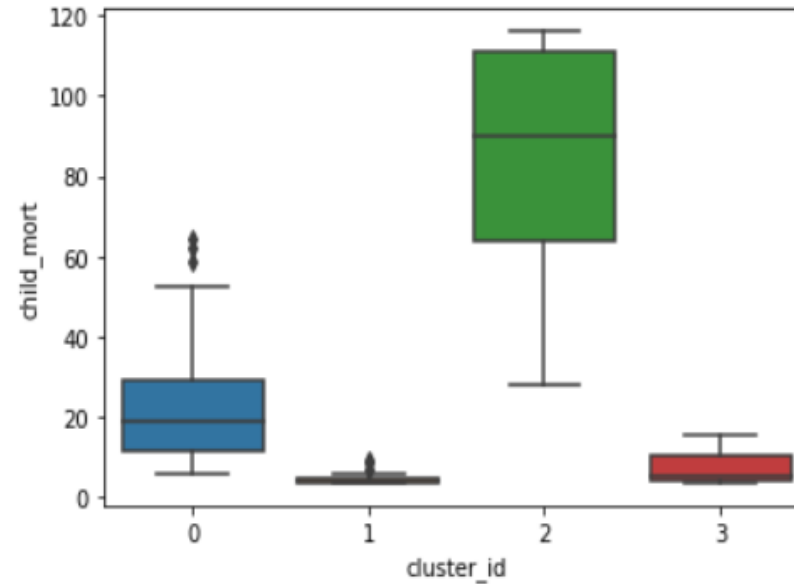
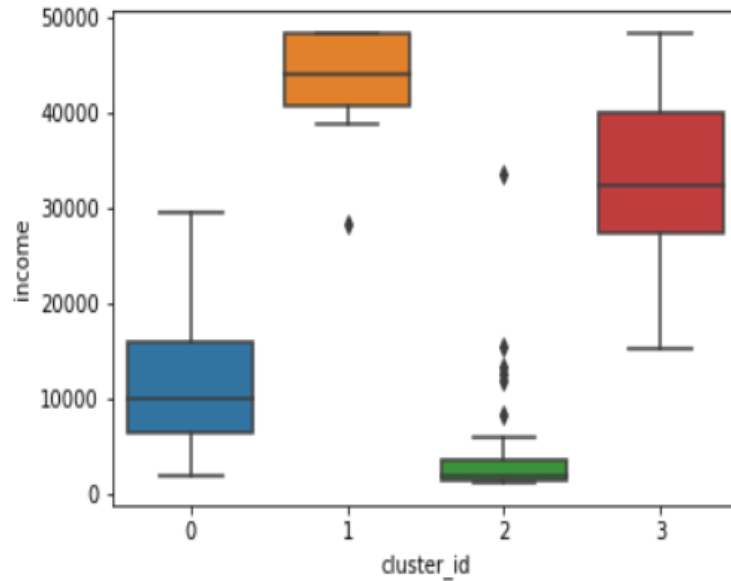
For n\_clusters=6, the silhouette score is 0.30307244388683313

For n\_clusters=7, the silhouette score is 0.3301983269977132

For n\_clusters=8, the silhouette score is 0.3432577893986418

- By looking silhouette score we can see  $k = 4$  and in sum of squared distances graph, we see that the elbow is in the range of 2 to 4, so we are going ahead with  $k$  as 4.

# K-means Clustering Boxplot

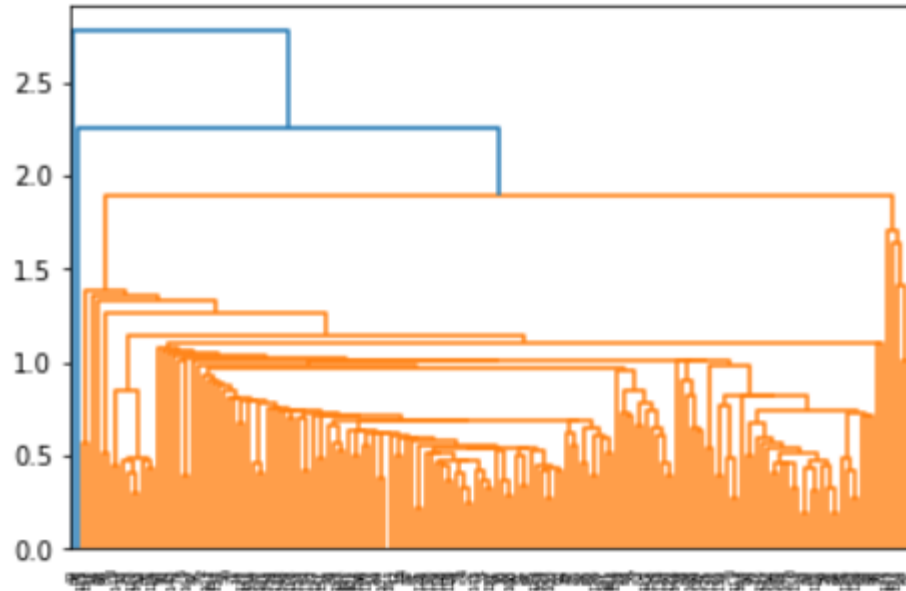


- Here **cluster\_id 2** has the lowest income and highest child mortality rate with a very a very low gdp so, this will be our concerned cluster

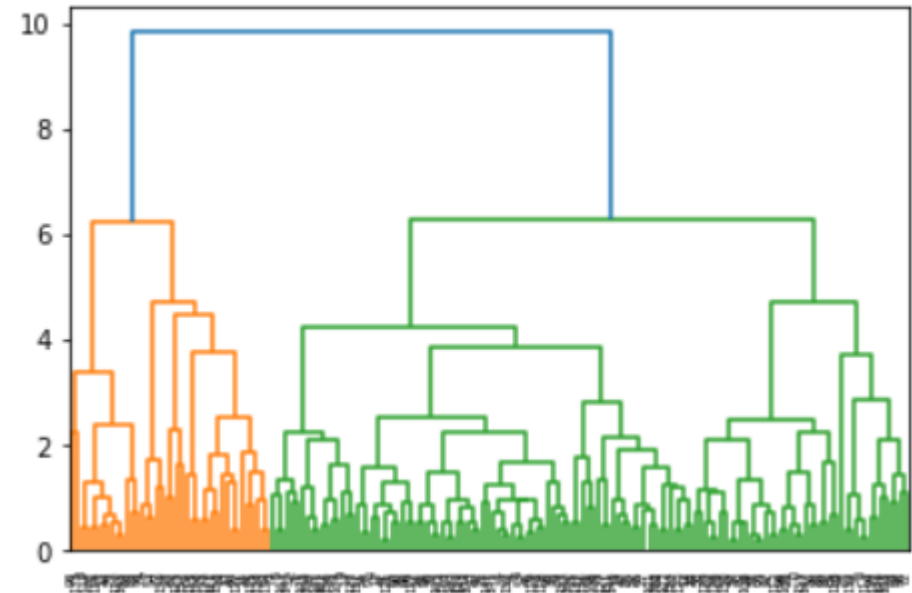
## Top 10 counties under cluster 2 using kMeans clustering

country	
0	Central African Republic
1	Congo, Dem. Rep.
2	Niger
3	Sierra Leone
4	Burkina Faso
5	Haiti
6	Mali
7	Chad
8	Nigeria
9	Angola

# Hierarchical Clustering



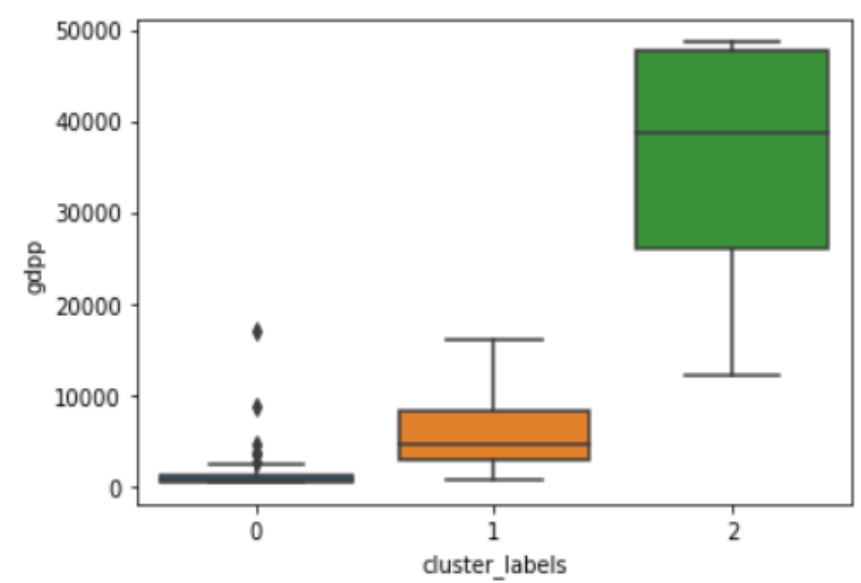
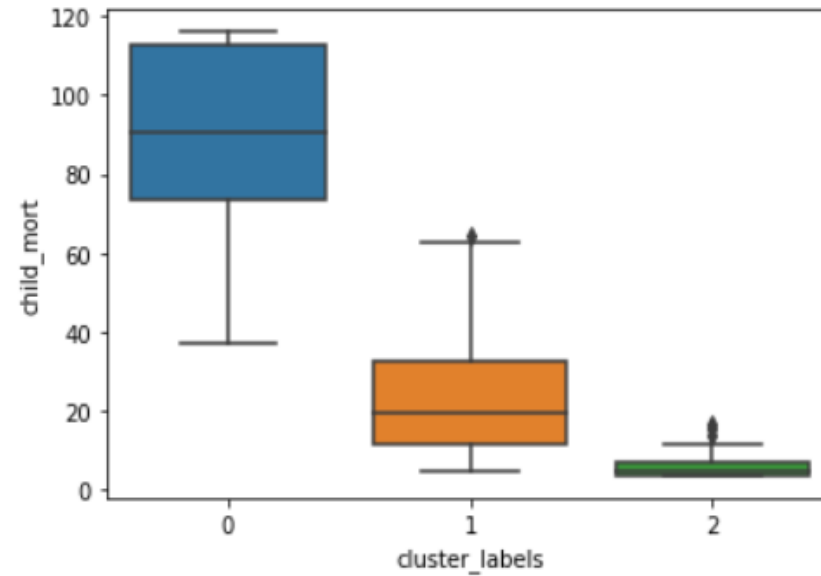
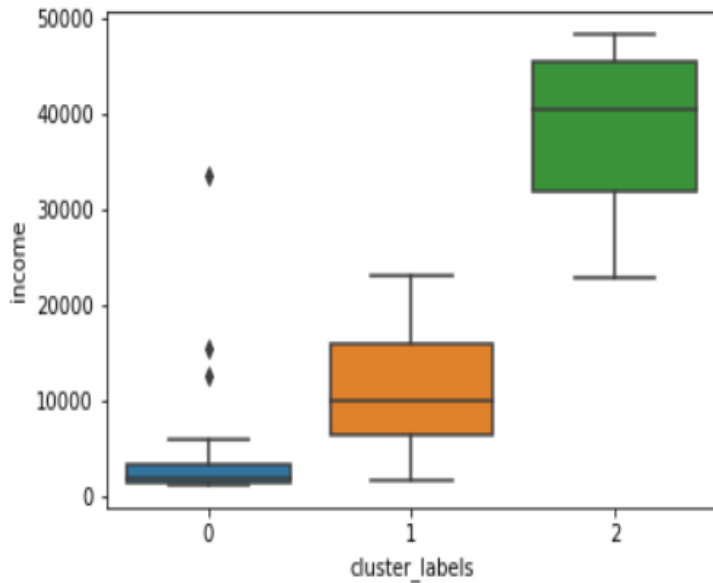
Single Linkage Method



Complete Linkage Method

- We are going for Complete method hierarchical clustering as single method clustering is not clear. By looking at this dendrogram taking  $n$ -clusters as 3.

# Hierarchical Clustering Boxplot



- Here **cluster\_id 0** has the lowest income and highest child mortality rate with a very a very low gdp so, this will be our concerned cluster



# Top 10 counties under cluster 0 using Hierarchical Clustering

country	
0	Central African Republic
1	Congo, Dem. Rep.
2	Niger
3	Sierra Leone
4	Burkina Faso
5	Haiti
6	Mali
7	Chad
8	Nigeria
9	Angola

# Summary

---

- As by both K means and Hierarchical clustering method - we have got same countries which requires aid.
- The following are the countries which are in direst need of aid:-

country	
0	Central African Republic
1	Congo, Dem. Rep.
2	Niger
3	Sierra Leone
4	Burkina Faso
5	Haiti
6	Mali
7	Chad
8	Nigeria
9	Angola



THANK  
You!