Clustering Assignment

Clustering Countries based on Socioeconomic Factors for Financial Aid by HELP International NGO

Author : Akash Agarwalla

Background

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.

Objective

After the recent 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.

Technical Approach

1. Using Hierarchical Clustering to identify optimal cluster value.
2. Using Silhouette and Elbow method to validate optimal cluster value.
3. Use K-Means cluster method to build final Cluster Model.
4. Identify appropriate cluster for financial aid using cluster mean method.
5. Analyze the final cluster statistics against other clusters and decision making on the final list based on the descriptive statistics of final cluster.
6. Choose the top 10 countries from the final cluster based on higher child mortality, lower gdpp and lower income.

Cluster Summary

-The final cluster generated **3** clusters.

-Based on our descriptive statistics, we can identify them as :

- Underdeveloped Countries
- Developing Countries
- Developed Countries

Cluster Underdeveloped Countries has the Highest average **Child Mortality rate of ~92** when compared to other 3 clusters, and Lowest average **GDPP & Income of ~ 1909 & 3897** respectively.

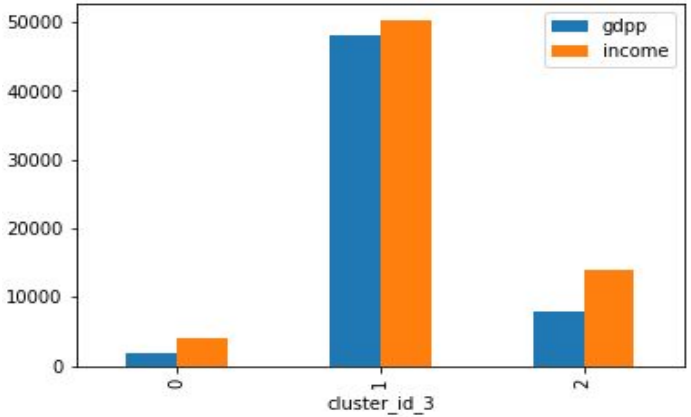
-All these figures clearly makes this cluster the best candidate for the financial aid from NGO.

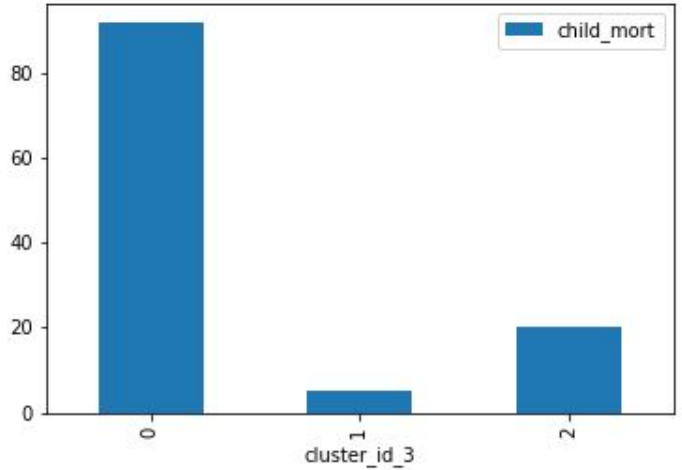
-We could also see that Cluster Underdeveloped countries comprises of **~29% of overall data**, and has ~48 observations in comparison to 167 total observations.

Analysis

cluster_id_3	Cluster
0	Underdeveloped Countries
2	Developing Countries
1	Developed Countries

cluster_id_3	child_mort	income	gdpp	Observations	Proportion
0	91.610417	3897.354167	1909.208333	48	0.29
2	20.357143	13968.021978	7979.912088	91	0.54
1	5.046429	50178.571429	48114.285714	28	0.17





Final List

We concluded top 10 countries from the final cluster(Underdeveloped Countries) based on the cluster median values of gdpp, income and child mortality and filtered the same with below order.

- **Lowest GDPP**

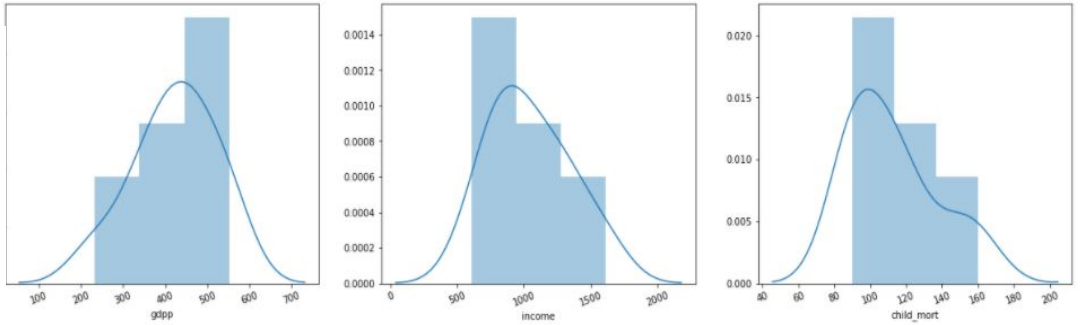
-**Lowest Income**

-**Highest Child Mortality**

Countries
Burundi
Congo,Dem. Rep.
Niger
Sierra Leone
Mozambique
Central African Republic
Malawi
Togo
Guinea-Bissau
Afghanistan

Statistics

country	gdpp	income	child_mort
Burundi	231	764	93.6
Congo, Dem. Rep.	334	609	116.0
Niger	348	814	123.0
Sierra Leone	399	1220	160.0
Mozambique	419	918	101.0
Central African Republic	446	888	149.0
Malawi	459	1030	90.5
Togo	488	1210	90.3
Guinea-Bissau	547	1390	114.0
Afghanistan	553	1610	90.2



	Min	Max	Median
GDPP	231	553	432
Income	609	1610	974
Child Mortality	90	160	107

Thank you.