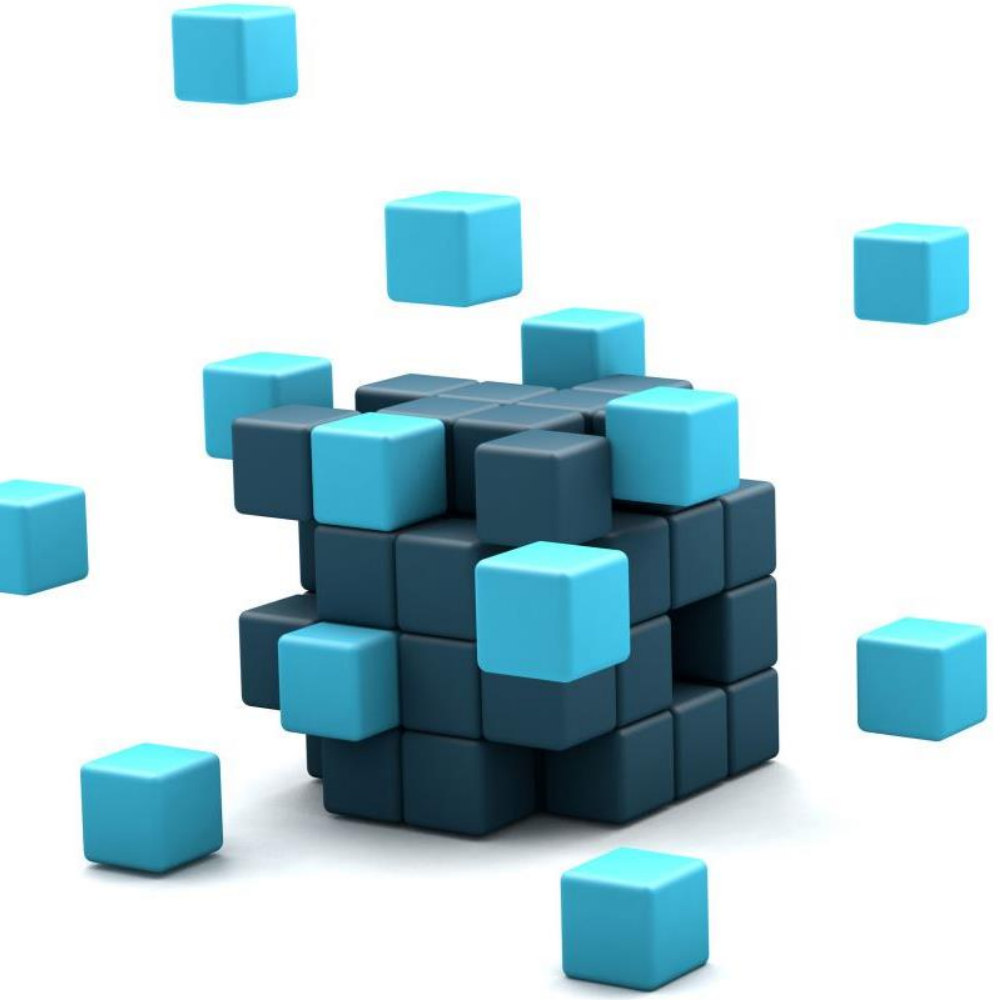


PRESENTATION
ON
CLUSTERING COUNTRIES BASED ON SOCIOECONOMIC FACTORS TO PROVIDE DIRE AID

Author: SHRUTI VYAS

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.

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:

- Performed Hierarchical Clustering – Chose Complete Linkage to proceed with the analysis.
- Hierarchical Clustering didn’t provide good results as 89% of data points were segmented into single cluster hence performed K-means clustering.
- Used Elbow Curve method and Silhouette Curve method to obtained optimal value of no. clusters. (Chose 3, 4 and 5 as value of K to proceed further)
- Used K-mean clustering method to build final cluster.
- Analysed the final cluster statistics against other clusters.
- Finally, Selected top 10 final countries which require aid from the NGO.

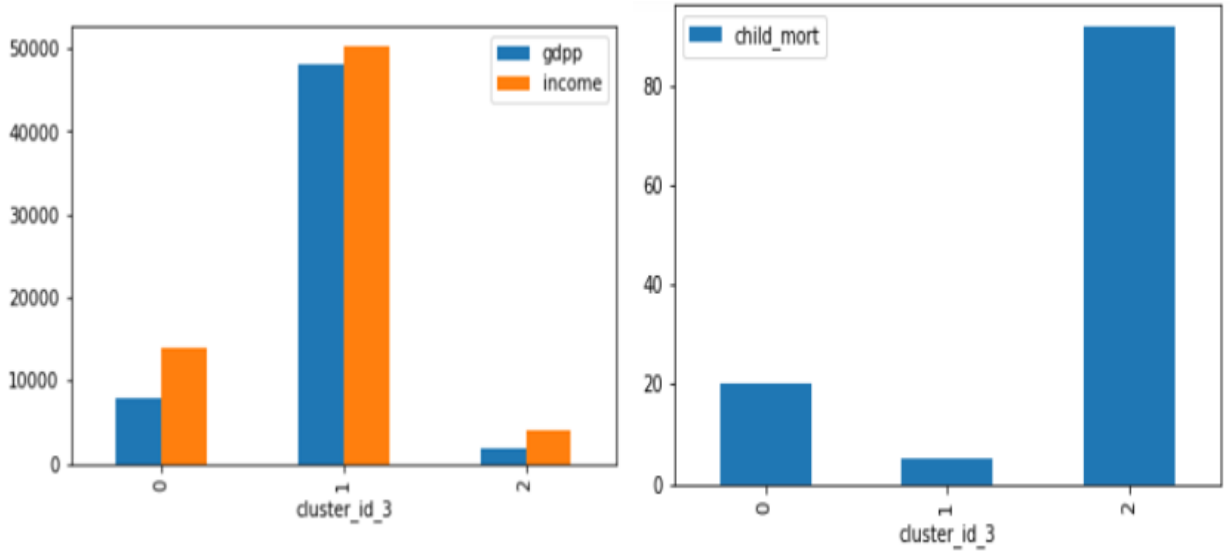
SUMMARY

- The final model generated 3 clusters. These clusters were identified after comparing various socio economic factors.
- The 3 clusters are categorised as:

Under Developed Countries, Developed Countries, Developing Countries

- Cluster 'Under Developed Countries' has **the Highest average Child Mortality rate of 92.61** when compared to other clusters, and **Lowest average GDPP & Income of 1909.20 & 3897.35** respectively.
- All these figures clearly make this cluster the best cluster for the aid. We could also see that this cluster comprises of **29% of overall data**, and has **48 count of observations compared to 167 total count.**

Cluster Name	Child Mortality Rate	Income	GDPP	Count	Percentage
Under Developed Countries	91.610417	3897.3542	1909.208	48	0.29
Developing Countries	20.357143	13968.022	7979.912	91	0.54
Developed Countries	5.046429	50178.571	48114.29	28	0.17



FINAL LIST OF COUNTRIES UNDER THE CLUSTER ‘UNDER DEVELOPED COUNTRIES’

- Top 10 countries are obtained based on the analysis done on the dataset which are in dire need of aid.
- These countries were shortlisted based on the Child mortality rate, Income and GDPP.
- The listed countries have following features in common:

a. Lowest GDPP

b. Lowest Income

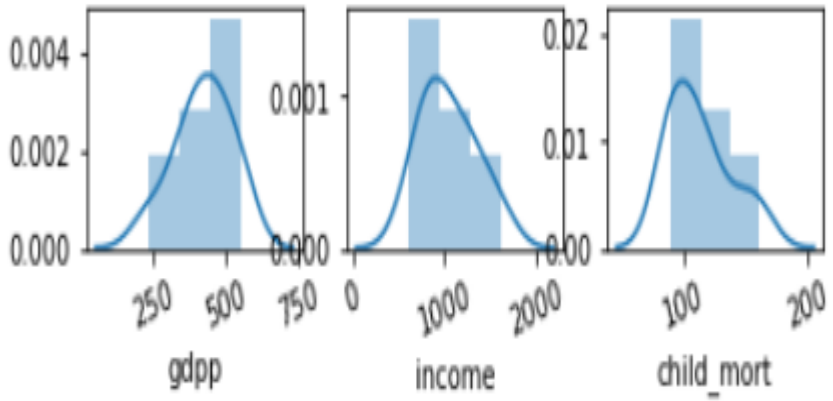
c. Highest Child Mortality Rate

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

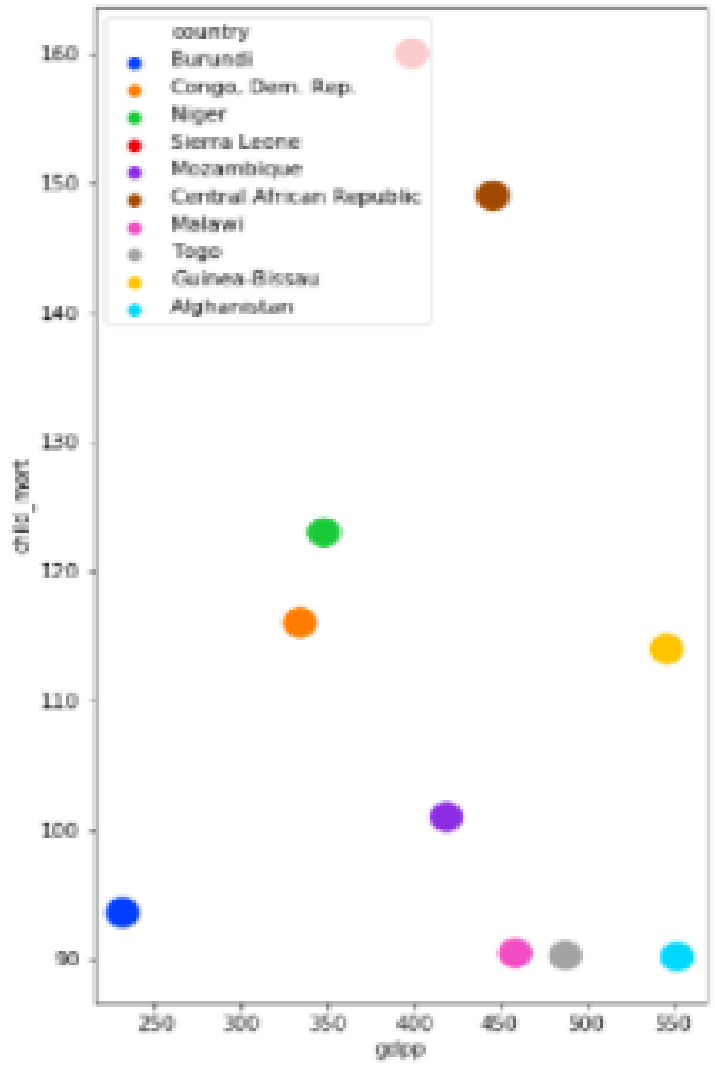
STATISTICS (OF FINAL LIST OF COUNTRIES)

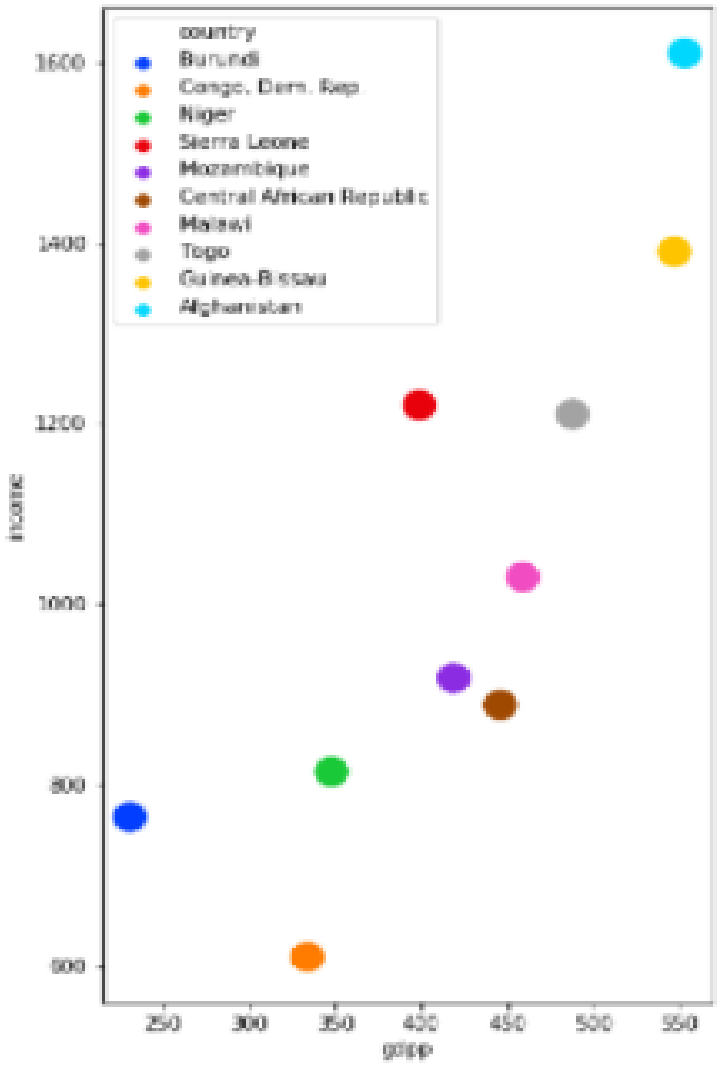
- Statistics as obtained from the analysis:

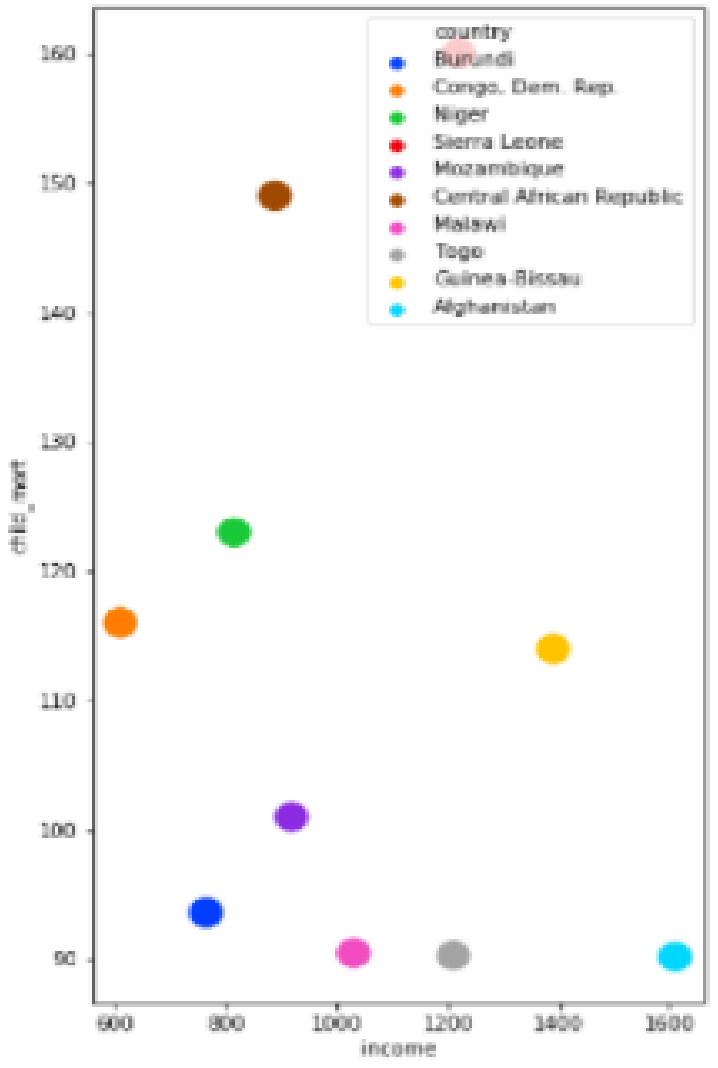
	Min	Max	Median
Income	609	1610	974
GDPP	231	553	432.5
Child Mortality Rate	90.2	160	107



	child_mort	income	gdpp
count	48.000000	48.000000	48.000000
mean	91.610417	3897.354167	1909.208333
std	34.319855	5590.168621	2925.911009
min	28.100000	609.000000	231.000000
25%	63.675000	1390.000000	551.500000
50%	89.750000	1860.000000	932.000000
75%	111.000000	3522.500000	1465.000000
max	208.000000	33700.000000	17100.000000


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


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


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

THANK YOU