CLUSTERING ASSIGNMENT

by G P Shiva Prasad

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.
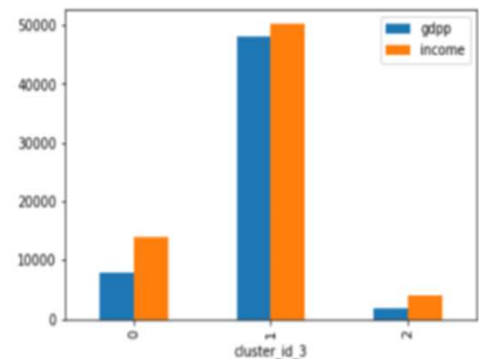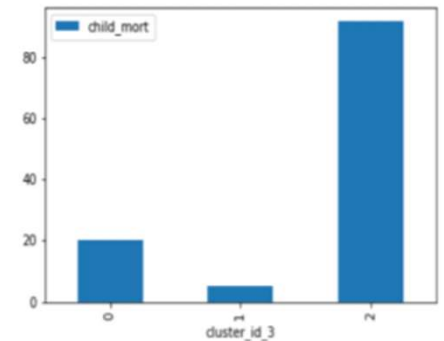
OBJECTIVES

- Categorising countries based on socio-economic and health factors that determine the overall development of the country.

- To suggest those countries which are in dire need of aid by analysing given data

Approach

- Using Hierarchical clustering to identify the optimal cluster value.

- Using Silhouette and Elbow method to validate the optimal cluster value.

- Building K-Means cluster model as the final model.

- Identify the cluster of countries which are in dire need of aid using cluster mean method.

- Analyse the final cluster statistics against other clusters.

- Making a decision on the final list based on the final cluster descriptive statistics.

- Choose top 10 countries which need help the most from the final cluster.

- Present a final report on this to CEO of HELP NGO

Cluster Summary

- Final Model Generated three clusters. We can identify them basing on the descriptive statistics. They are

    1. Underdeveloped countries.

    2. Developing countries.

    3. Developed countries.

- Underdeveloped countries has the highest average child mortality rate of ~92 and lowest average gdpp and income of ~1909 & 3897 respectively.

- Basing on this, we can say that "Underdeveloped" cluster is the cluster that needs aid from NGO

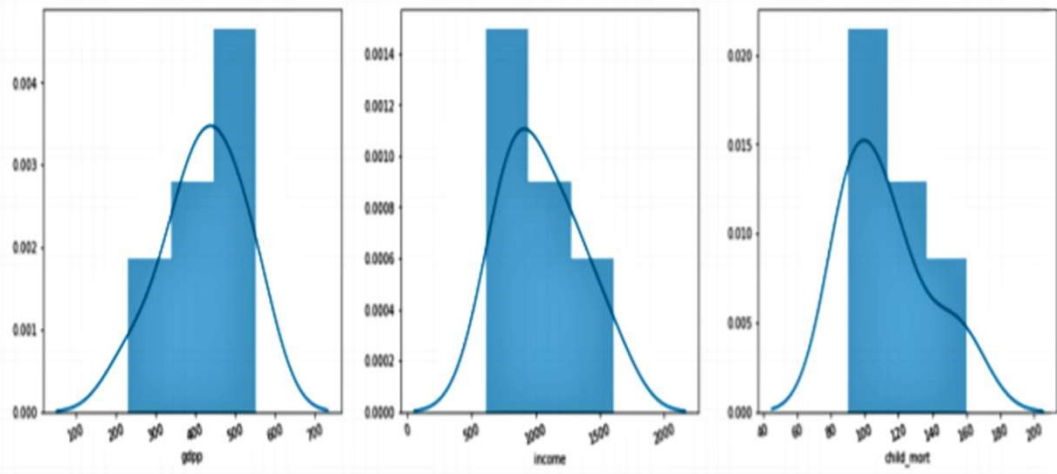| Cluster Name	child_mort	income	gdpp	obs	prop
Under_Developed Countries	91.61	389.35	1909.21	48	0.29
Developing Countries	20.36	13968.02	7979.91	91	0.54
Developed Countries	5.05	50178.57	48114.29	28	0.17

TOP 10 UNDERDEVELOPED COUNTRIES

- These are the top 10 underdeveloped countries based on the cluster median values of gdpp, income and child_mort

. Filtered countries based on

  1. Lowest gdpp
  2. Lowest income.
  3. Highest child_mort

TOP 10 COUNTRIES FOR FINANCIAL AID
Burundi
Congo, Dem. Rep.
Niger
Sierra Leone
Mozambique
Central African Republic
Malawi
Togo
Guinea-Bissau
Afghanistan

DESCRIPTIVE STATISTICS OF RECOMMENDED COUNTRIES

	Gdpp	Income	Child_mort
Min	231	609	90
Median	432.5	974	107
Max	553	1610	160

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

	gdpp	income	child_mort
count	10.00000	10.000000	10.00000
mean	422.40000	1045.300000	112.76000
std	99.62396	309.034356	25.11344
min	231.00000	609.000000	90.20000
25%	360.75000	832.500000	91.27500
50%	432.50000	974.000000	107.50000
75%	480.75000	1217.500000	121.25000
max	553.00000	1610.000000	160.00000

THANK YOU