CLUSTERING OF COUNTRIES

Author:

Rajeev Ranjan Sinha

PROBLEM STATEMENT

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 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.

BUSINESS GOALS:

- Our job is to categorize the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.

OVERALL APPROACH

1: Data understanding - > Asking Questions from Data

2: Data Munging/Wrangling

3: Data Visualization - EDA (Feature Engineering)

4: Data Preparation

5: PCA Application

6: Hopkins Statistics Test

7: Model Building

8: Final Analysis

CORRELATION IN THE DATA

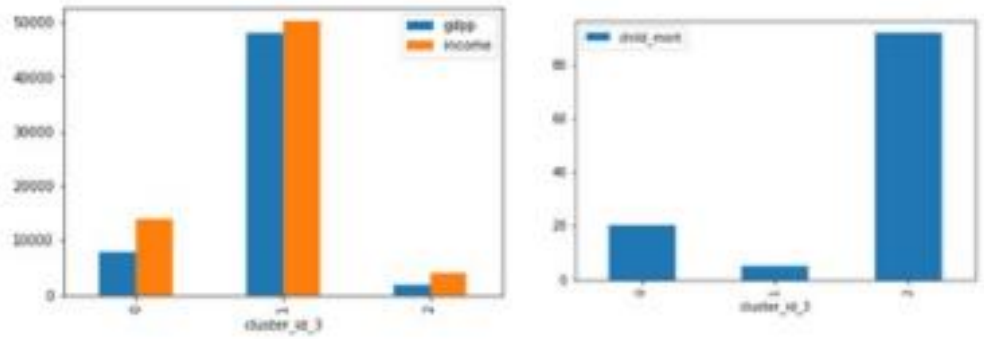
- After data cleaning, we removed outlier from gdpp column because the country with high gdpp would not require any aid as they are already doing good.

- We did standardized scaling to standardize all parameters on cleaned, outlier removed data.
- Looking at the heatmap, we see thatfew variables like (total fertility, child mortality), (income, gdpp) and (imports and exports) have high correlation.

CLUSTER SUMMARY

Cluster Name	child_mort	income	gdpp	Obs	Prop
Under_Developed Countries	91.61	3897.35	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

Cluster Mean Values along with no. of Observations & Proportion



Graphical representation of GDPP, Income & Child Mortality of 3 clusters

- The final model generated 3 clusters. Based on their descriptive statistics, we could identify them as:
  - Under developed countries
  - Developing countries
  - Developed countries
- Clusters 'under developed countries' has the Highest average Child Mortality rate of ~92, when compared to other 3 clusters. And lowest Avg. GDPP and 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 'Under developed countries' comprises of ~29% of overall data.

FINAL LIST OF COUNRIES FROM THE CLUSTER – **'UNDER DEVELOPED COUNTRIES'**

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

TOP 10 Countries recommended for financial Aid

We concluded on the top 10 list of countries from the final cluster('under developed countries') based on the cluster median values of gdpp, income and child mortality.

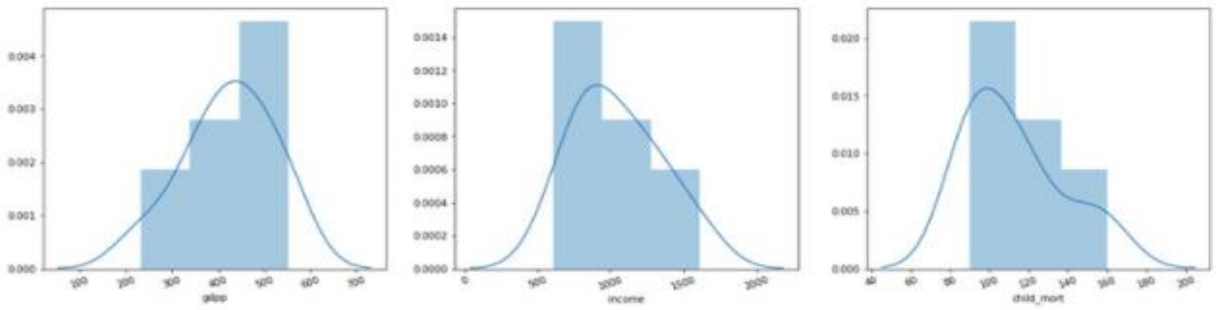
We filtered the countries with (in below order)-

- Lowest gdpp
- Lowest income
- Highest child_mortality

FINAL RECOMMENDATION

Statistics
(of our recommended countries)

Min GDPP = 231	Max GDPP = 553	Median GDPP= 432.5
Min INCOME = 609	Max INCOME = 1610	Median INCOME= 974
Min CHILD_MORT = 90	Max CHILD_MORT = 160	Median CHILD_MORT= 107



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_Mortality of top 10 recommended countries

	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

Descriptive Statistics of top 10 recommended countries

THANK YOU

Rajeev Sinha