Clustering Assignment

By
Sourav Acharya

Problem Statement

- HELP humanitarian NGO has 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.

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

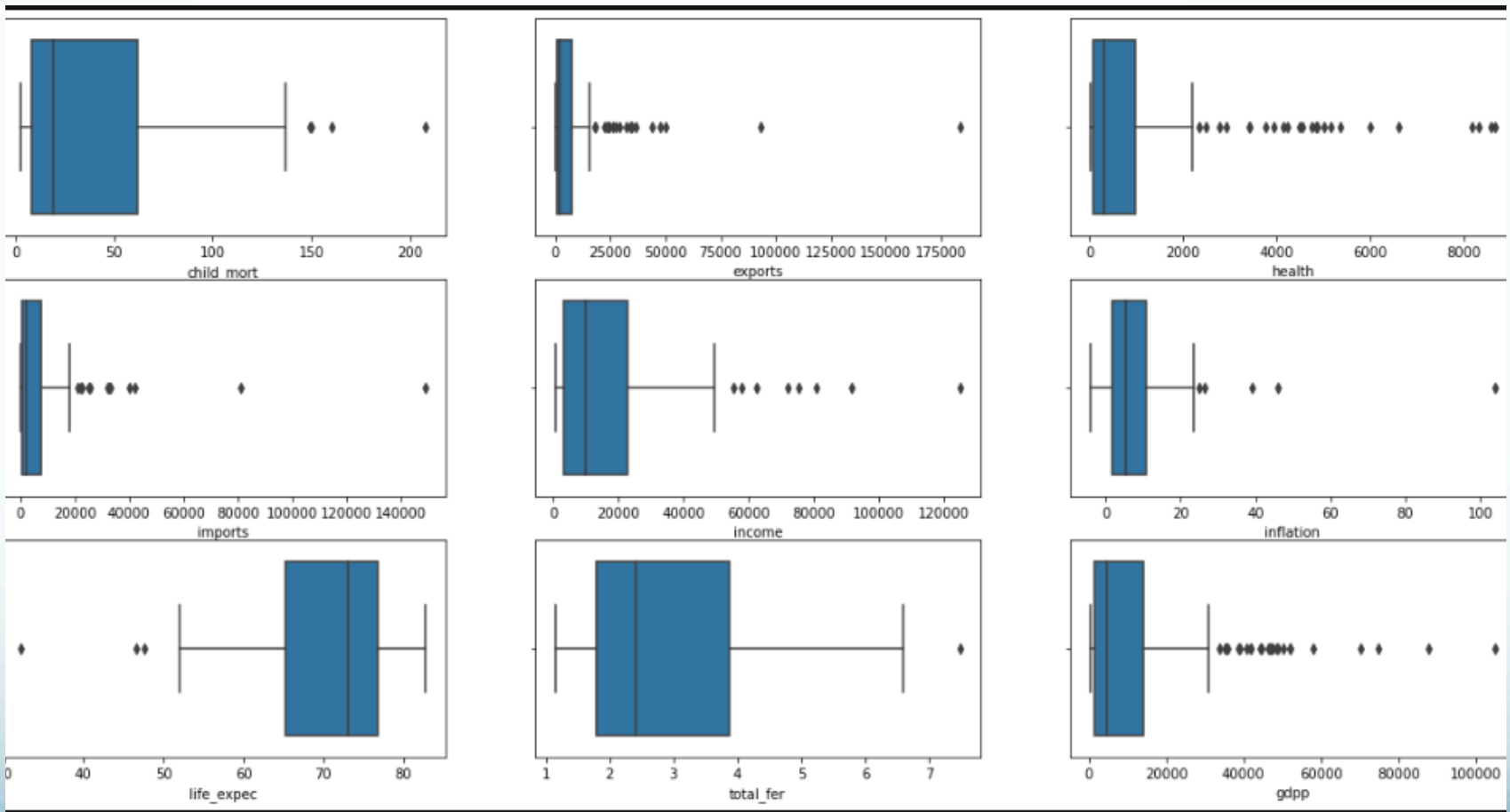# Analyzing Different Parameters for choosing the countries

Column Name	Description
country	Name of the country
child_mort	Death of children under 5 years of age per 1000 live births
exports	Exports of goods and services per capita. Given as %age of the GDP per capita
health	Total health spending per capita. Given as %age of GDP per capita
imports	Imports of goods and services per capita. Given as %age of the GDP per capita
Income	Net income per person
Inflation	The measurement of the annual growth rate of the Total GDP
life_expec	The average number of years a new born child would live if the current mortality patterns are to remain the same
total_fer	The number of children that would be born to each woman if the current age-fertility rates remain the same.
gdpp	The GDP per capita. Calculated as the Total GDP divided by the total population.

Analyzing Different Parameters for choosing the countries

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

Here is the sample data for 5 countries out of 167 countries data
We will doing out analysis on these 167 countries and try to find out
cluster countries based on their various socio-economic factors (gdpp,
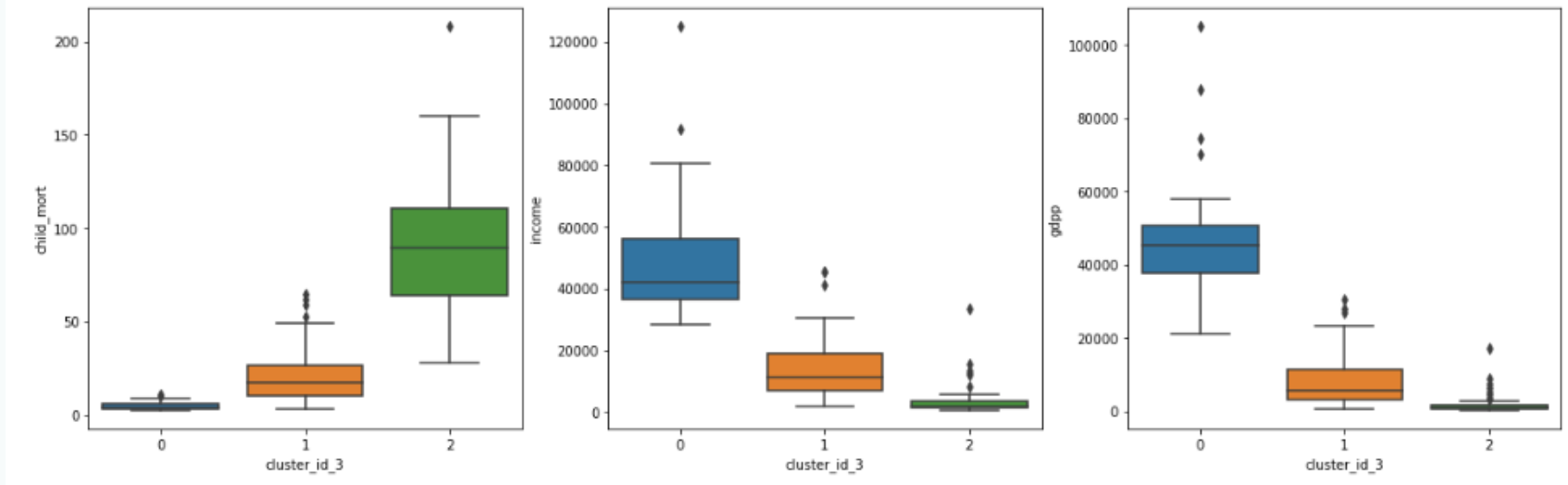income and child mortality)

Data Analysis

Data Analysis

- There are some data which can be called as outliers, we wont be removing those as it might lead to removal of countries which actually need funding

Data Analysis

- There are some data which can be called as outliers, we wont be removing those as it might lead to removal of countries which actually need funding

- Based on the analysis we have found 3 factors affect the decision making factor mostly they are –
  - Gdpp
  - Income
  - Child_mort

Data Analysis

Based on the boxplots above, we can see we have 3 main clusters, where –

- Cluster 0 – has low child mort but high gdpp and income
- Cluster 1 – has medium child mort with medium gdpp and income
- Cluster 3 – has high child mort and low gdpp and income

So Cluster 3 is best group for our further analysis and we can label it as Poor Countries.

List of Poor Countries

Haiti
Sierra Leone
Chad
Central African Republic
Mali
Nigeria
Niger
Angola
Congo, Dem. Rep.
Burkina Faso
Guinea-Bissau
Benin
Cote d'Ivoire
Equatorial Guinea
Guinea
Cameroon
Mozambique
Lesotho
Mauritania
Burundi
Pakistan

Malawi
Togo
Afghanistan
Liberia
Comoros
Zambia
Uganda
Gambia
Lao
Sudan
Ghana
Tanzania
Senegal
Congo, Rep.
Gabon
Rwanda
Kiribati
Timor-Leste
Madagascar
Kenya
Yemen
Namibia

Eritrea
South Africa
Botswana
Iraq
Solomon Islands

Analysis of Poor Countries

- During further analysis it was found that some countries have high child mortality but their gdpp and income was high too like Equatorial Guinea

Equatorial Guinea	111.0	33700	17100

- High income and gdpp means these countries are better off than those with low gdpp, income and high child mortality

- So countries with low gdpp, low income and high child mortality need to be filtered

Analysis of Poor Countries – Low GDPP, Low Income, High Child Mortality

- For this analysis we took median values of all the 3 parameters and filtered the countries

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
Burkina Faso	575	1430	116.0
Guinea	648	1190	109.0
Haiti	662	1500	208.0
Benin	758	1820	111.0

Top 5 Countries That Need Funding

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