

Segmenting Countries for NGO Aid using Clustering

Analysis & Result

Problem Statement:

HELP International is an international humanitarian NGO that runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.

The CEO of the NGO needs to decide how to use \$10 million strategically and effectively by choosing the countries that are in the direst need of aid.

Goal

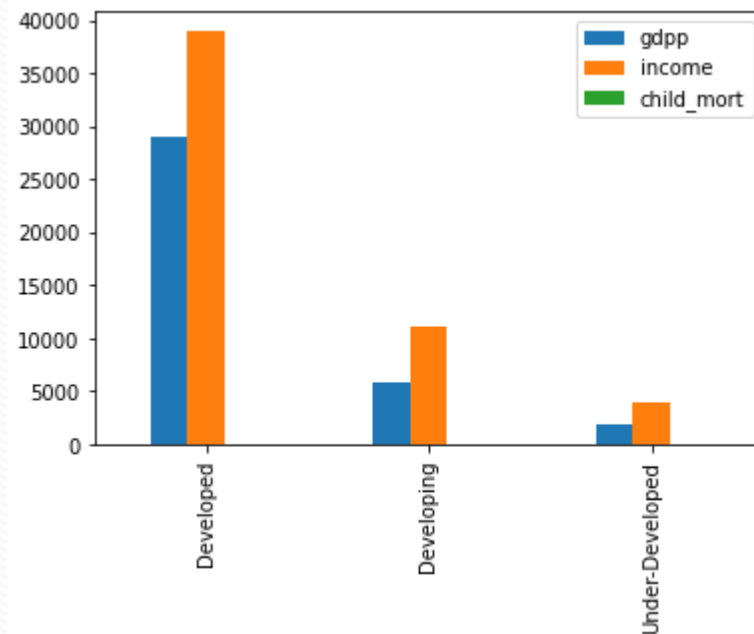
- We have to categorise the countries using some socio-economic and health factors that determine the overall development of the country.
- Then we need to suggest the countries which the CEO needs to focus on the most.

Approach to the Problem

- Sourcing and inspecting the data provided by the company
- Data Cleaning
- Exploratory Data Analysis.
- Data preparation
- Feature Scaling
- Cluster Formation
- Cluster Profiling
- Results Comparison and choosing the best algorithm
- Recommendation of the final list

Observations from K-Means Clustering

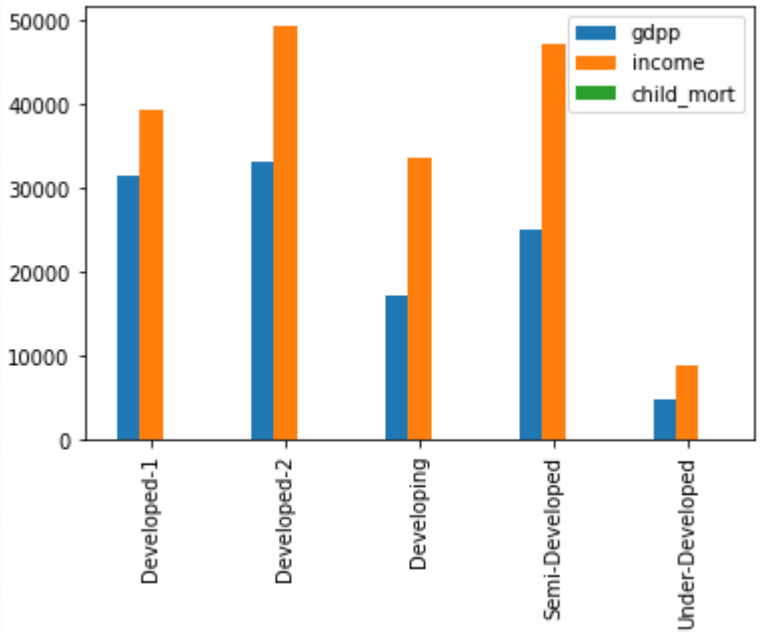
- There is a clear segregation of countries in respect to the gdpp, income and child mortality parameters



	gdpp	income	child_mort
Clustered_Countries			
Developed	28969.512195	38988.780488	5.651220
Developing	5759.897436	11196.025641	22.591026
Under-Developed	1909.208333	3897.354167	89.620833

Observations from Hierarchical Clustering (Single Linkage)

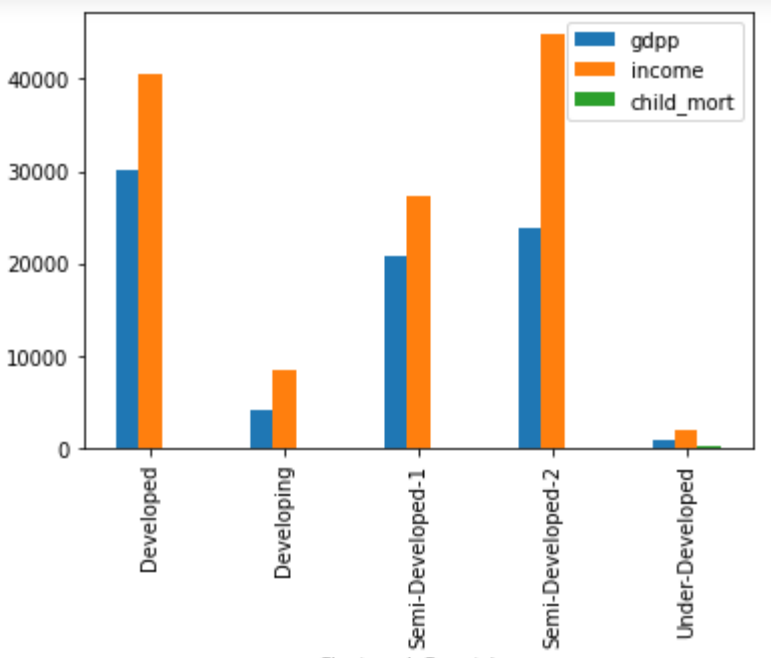
The total number of countries falling in the developed and semi developed countries are 6
While the list of under developed countries is 130
This figure is not good to make an informed analysis



	gdpp	income	child_mort
Clustered_Countries			
Developed-1	31428.666667	39483.500000	4.736667
Developed-2	33130.000000	49400.000000	7.300000
Developing	17100.000000	33700.000000	111.000000
Semi-Developed	25112.000000	47147.000000	11.460000
Under-Developed	4692.415385	8888.946154	45.983846

Observations from Hierarchical Clustering (Complete Linkage)

- There is a notable difference between the under developed countries and the developing ones
- The list of under developed countries is 25



	gdpp	income	child_mort
Clustered_Countries			
Developed	30081.739130	40504.565217	4.904348
Developing	4178.606742	8498.651685	34.133708
Semi-Developed-1	20869.166667	27200.000000	7.450000
Semi-Developed-2	23776.666667	44905.833333	28.050000
Under-Developed	854.720000	1947.320000	111.912000

Results Comparison and choosing the best algorithm

- We see that the top 10 countries by both K-means and Hierarchical (single Linkage) are the same
- The top 10 list of Hierarchical (complete linkage) is almost the same but varies by 2 countries
- The top 5 countries in all the list are the same

	K-Means Countries	Hierarchical-Single-Linkage Countries	Hierarchical-Complete-Linkage Countries
0	Burundi	Burundi	Burundi
1	Liberia	Liberia	Liberia
2	Congo, Dem. Rep.	Congo, Dem. Rep.	Congo, Dem. Rep.
3	Niger	Niger	Niger
4	Sierra Leone	Sierra Leone	Sierra Leone
5	Madagascar	Madagascar	Mozambique
6	Mozambique	Mozambique	Central African Republic
7	Central African Republic	Central African Republic	Malawi
8	Malawi	Malawi	Togo
9	Eritrea	Eritrea	Guinea-Bissau

Results Comparison and choosing the best algorithm

- The cluster size for the needy countries obtained from hierarchical Complete linkage is the smallest at 25
- This can help us in focussing resources better on the most needy countries of all
- So we can choose the Hierarchical Complete Linkage clustering algorithm in this case

Comparing number of countries obtained from all clustering methods

The total number of poor countries in need of aid obtained are:

K-Means Clustering : 48

Hierarchical-Single-Linkage Clustering : 130

Hierarchical-Complete-Linkage Clustering : 25

Recommendation of the final list

- The final list of top 5 countries in direst need of help are:
 - 1) Burundi
 - 2) Liberia
 - 3) Congo, Dem. Rep.
 - 4) Niger
 - 5) Sierra Leone