HELP INTERNATIONAL - AID ASSESSMENT

ARVIND R NAIR

PROBLEM STATEMENT

HELP International, an international humanitarian NGO, has raised around $ 10 million through funding programs and now needs to decide how to use this money strategically and effectively. The requirement is to categorise the countries using some socio-economic and health factors that determine the overall development of the country and then suggest the countries that are in the direst need of aid and which the CEO needs to focus on the most.

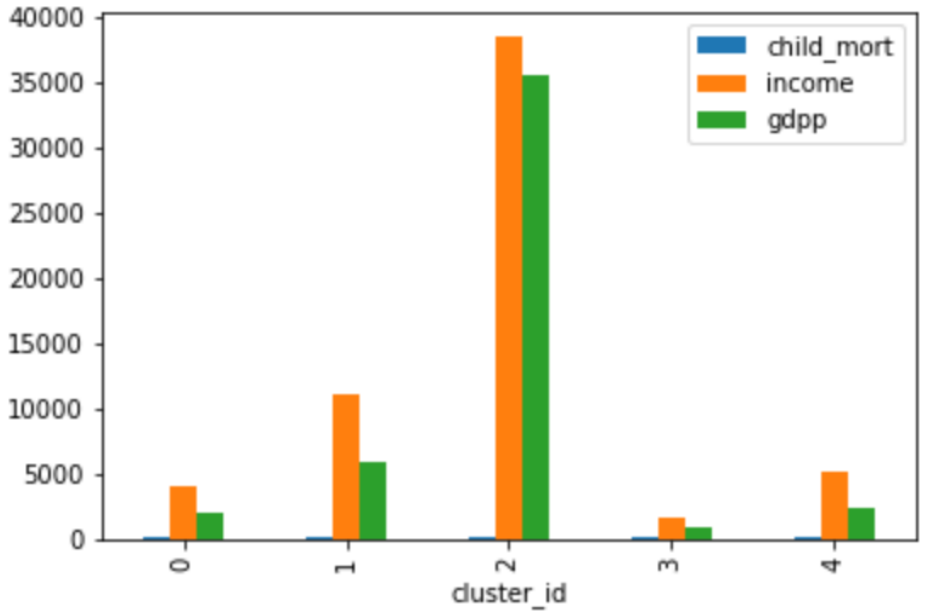
ANALYSIS APPROACH

- As this is a requirement to segment and identify countries in need of aid, we have used clustering – an unsupervised machine learning technique.
- The following steps were taken to ready the data in addition to clustering:-
  - Exploratory Data Analysis – This included data inspection and use of data visualiztions and correlation techniques to carry out univariate and bivariate analyses.
  - Data Preparation – This involved Outlier Analysis & Treatment as well as converting all numeric variables to a comparable scale using a scaling technique.
  - Clustering – Two clustering techniques were used, namely - K Means & Hiearchical Clustering and the results were compared using univariate and bivariate analysis techniques before arriving at a conclusion.

CLUSTERING MODEL RESULTS

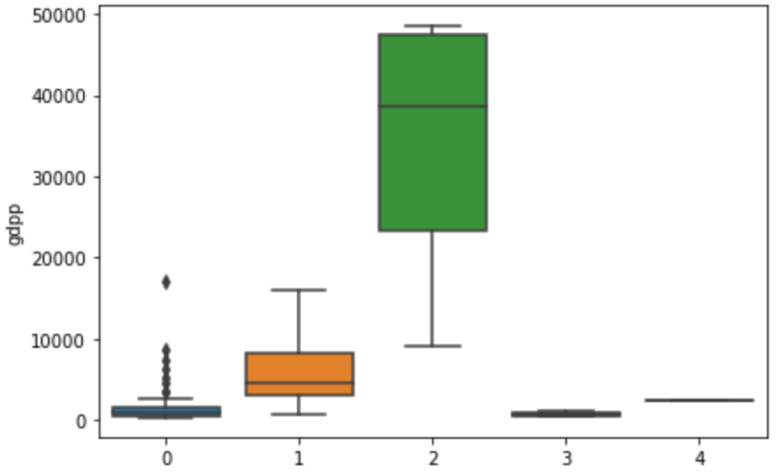
- The K-Means clustering model arrived at 3 clusters of countries.
- Our Hiearchical clustering model arrived at 4 main clusters with a 5th one that consisted of only a single country.
- We went with the results of the Hiearchical clustering model for the following reasons:-
  - External research *(link in assumptions slide),* in the absence of business inputs, showed that the world has moved away from a 3 cluster or developed vs developing vs underdeveloped countries to a state of '4 Levels'.
  - Hierarchical Clustering is known to be generally more accurate than K Means.

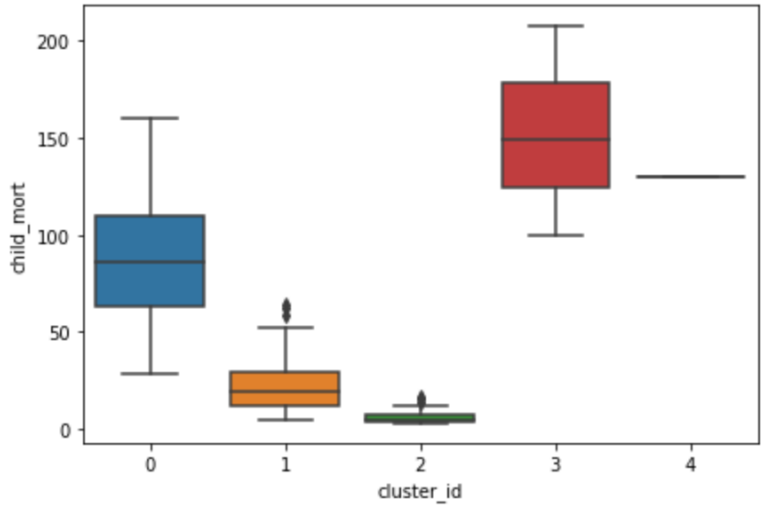
OBSERVATIONS - I

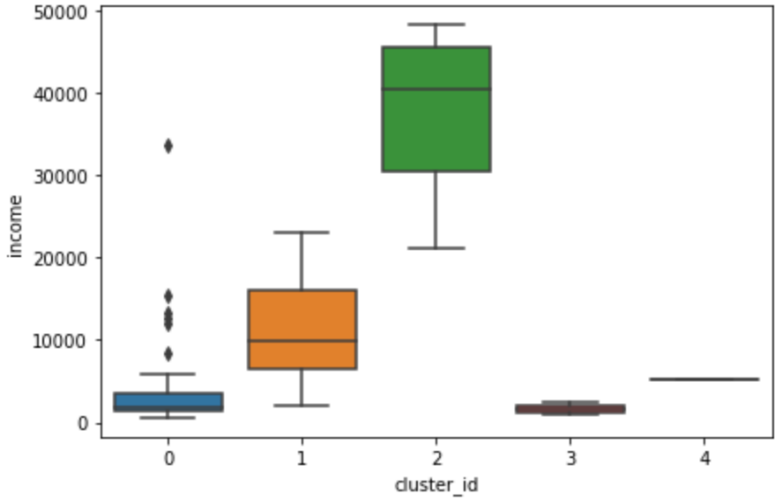


- There are 4 main clusters in terms of the 3 important variables - GDP, Child Mortality & Income. The 5th one - namely Cluster 4 - has only one country: Nigeria that does not seem to fit into other clusters.
- The differences in the 3 variables across all clusters are relatively smaller when compared to differences observed in K-Means clustering.
- Similar to observations made in K-Means clustering, there is one cluster that is poorly performing across all 3 variables: Cluster 3.

OBSERVATIONS - II







- As indicated in the boxplots, cluster 3 performed worst across all the important indicators followed by Cluster 0.
- While the median of child mortality hovers around 4 per 1000 children for Cluster 2, 18 per 1000 for Cluster 1 and 85 per 1000 for Cluster 0, the same parameter for Cluster 3 was 149 per 1000 which is significantly high just by itself.
- GDP & Income are very low for Cluster 3 countries as well.

CONCLUSION

In conclusion, the final list of countries were arrived at using two clusters of countries – Cluster 3 and Cluster 0. This is due to Cluster 3 only containing 3 countries and Cluster 0 following Cluster 3 in performing poorly across key indicators.

In the order of the highest need, sorted basis GDP, Child Mortality & Income, the final list of five countries that require aid are:-

- Central African Republic
- Haiti
- Lesotho
- Burundi
- Liberia

ASSUMPTIONS & REFERENCES

- In the absence of business inputs, external sources were referred to and research by Hans Rosling (Gapminder Foundation - https://www.gapminder.org/topics/four-income-levels/) was used to employ a logical process in arriving at the final number of clusters.T
- The research showed that the modern world has moved on from fitting into the developed vs developing vs underdeveloped countries narrative and that we now live in a world with '4 levels' defined basis several factors such as income, lifespan, GDP, child mortality etc.
- The results from Hiearchical Clustering was preferred over K Means Clustering based on above inputs.

THANK YOU!