WHERE TO SHELL OUT THE AID?

Deciding the countries where the monetary aid is most required.

PROBLEM STATEMENT

- HELP International, an international humanitarian NGO has acquired a fund of 10 million USD through recent fundings.
- The CEO wants to decide how should these funds be used in a strategic and effective manner by deciding which countries are in dire need of the aid.
- The countries needs to be clustered based on various socio-economic and health factors.

ANALYSIS OVERVIEW

- ➔ There are 167 distinct countries containing 13 different attributes ranging from net imports/exports per capita, health spending per capita, annual growth rate, average life expectancy, net income per person, child deaths per 1000.
- ➔ Based on these factors, we can cluster the countries.
- ➔ Initial inspection of the dataframe revealed few outliers in almost all the attributes. These outliers were of importance because they either represented developed countries or under-developed countries, therefore were not removed.

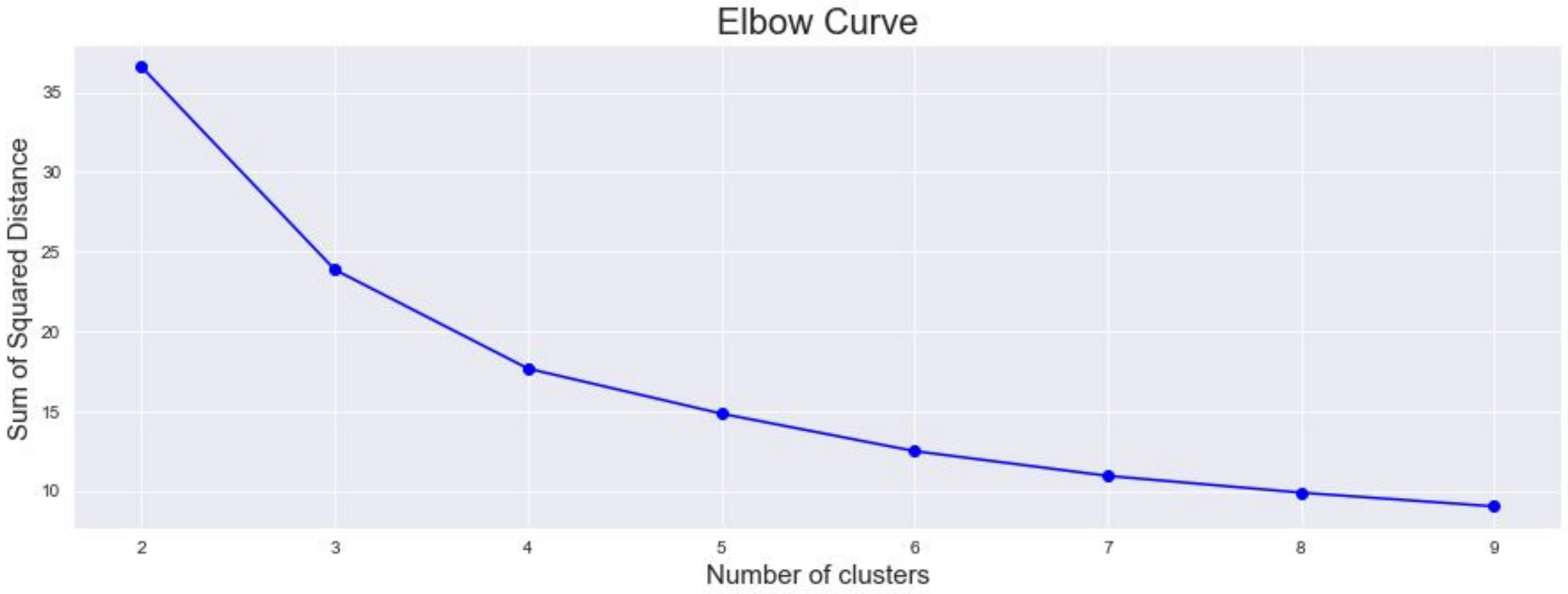
ANALYSIS OVERVIEW

- ➔ During Outlier treatment, the outliers in the features that represented the rich countries were capped to an appropriate value so that they would not impact the cluster centers and we would not have to remove them.
- ➔ For instance, outliers in the higher side of income, GDP, exports etc. were capped as they represented countries belonging to the developed category.
- ➔ The Hopkins Statistic, which denotes how much is the clustering tendency also gave a high value(0.89) which was desirable.

ANALYSIS OVERVIEW

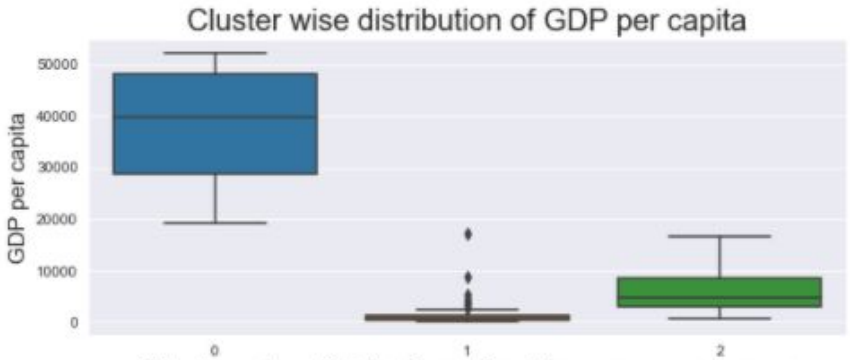
- Using the elbow curve and the Silhouette Score, the value of ‘k’ was chosen as 3. There was a clear distinction among the clusters when GDP, child mortality and net income per person were visualized.
- Then using the hierarchical clustering(complete linkage), we cut the dendrogram at k = 3.
- Based on the dendrogram, we finalized the value of ‘k’ as 4.
- We then rebuilt the KMeans model using the cluster centers returned by hierarchical clustering and retrieved the final list of countries which required the most aid.

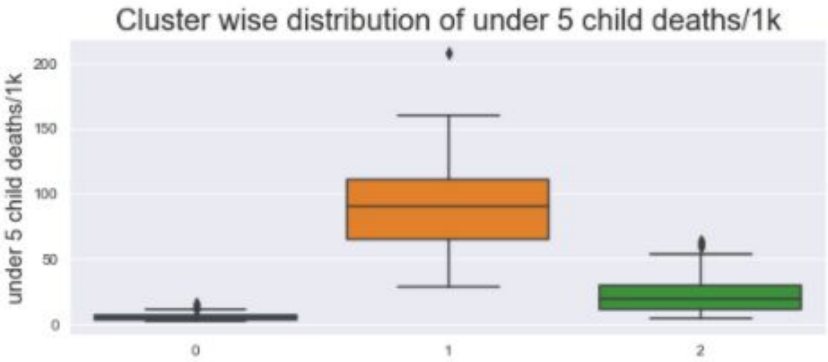
CLUSTERING-KMEANS

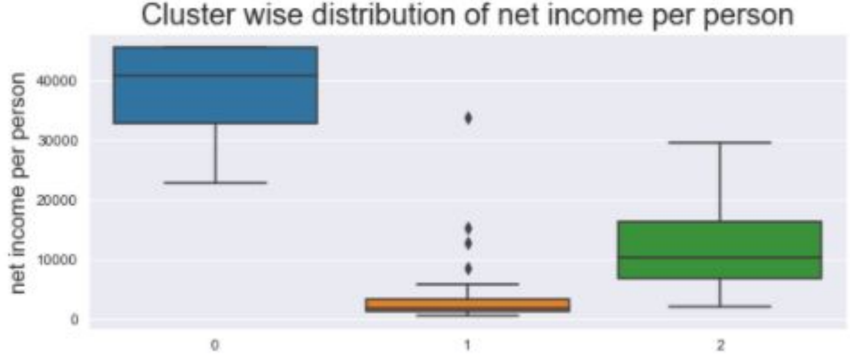


Based on the elbow curve and Silhouette score, the number of clusters were decided as 3.

CLUSTERING-KMEANS



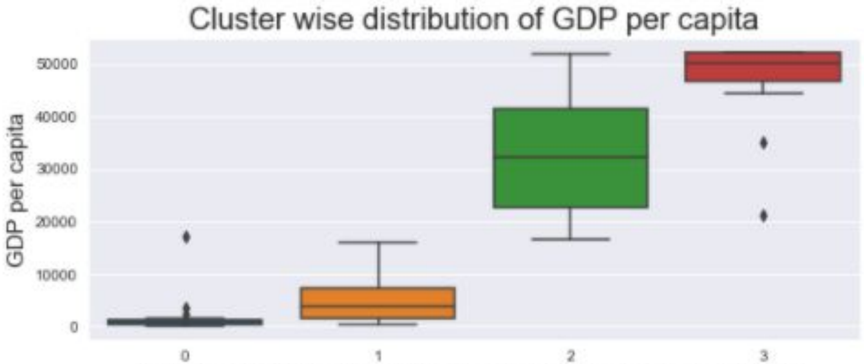


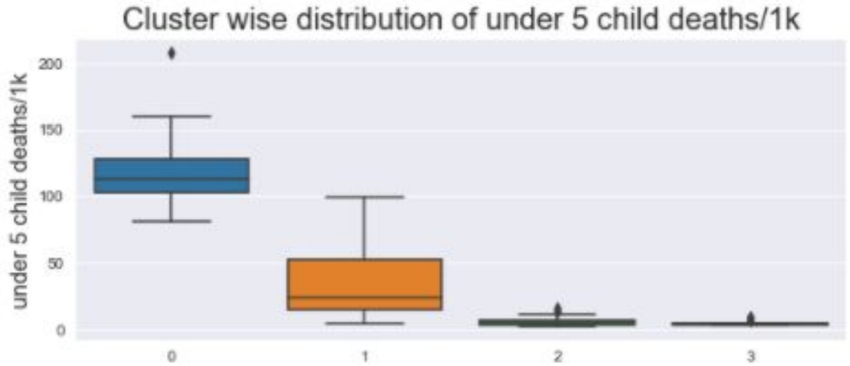


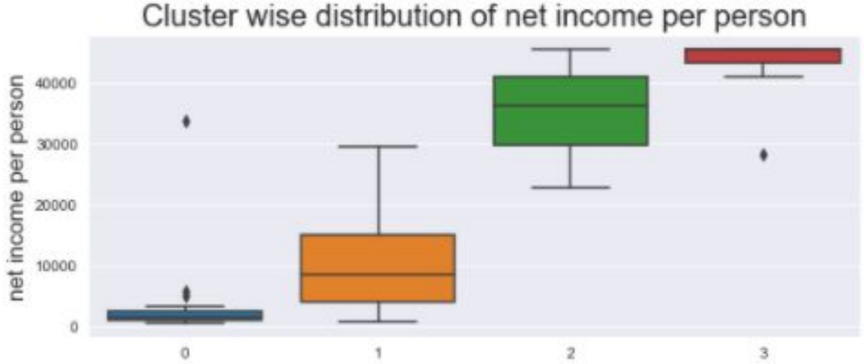
KMEANS CLUSTERING INFERENCES

- ➔ From the previous plots, we can cluster the countries into 3 distinct categories.
- ➔ Countries belonging to the Cluster 1 are considered to be weak countries. They have high under 5 child deaths/1k, low GDP per capita and low net income per person.
- ➔ Countries belonging to cluster 0 are strong countries. They have low child deaths/1k, high GDP per capita and high net income per person.

CLUSTERING-HIERARCHICAL



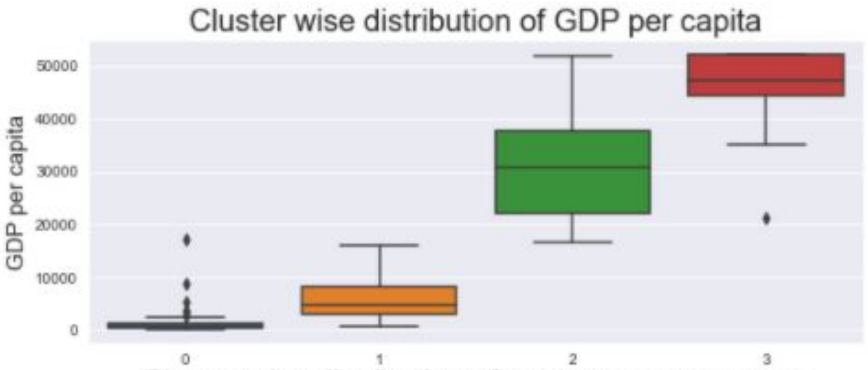


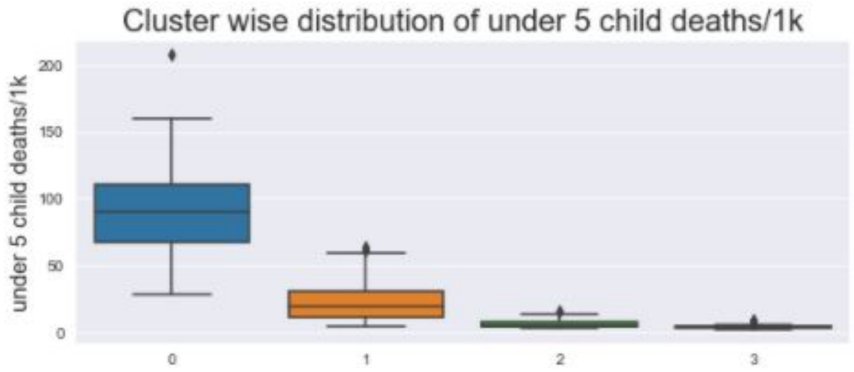


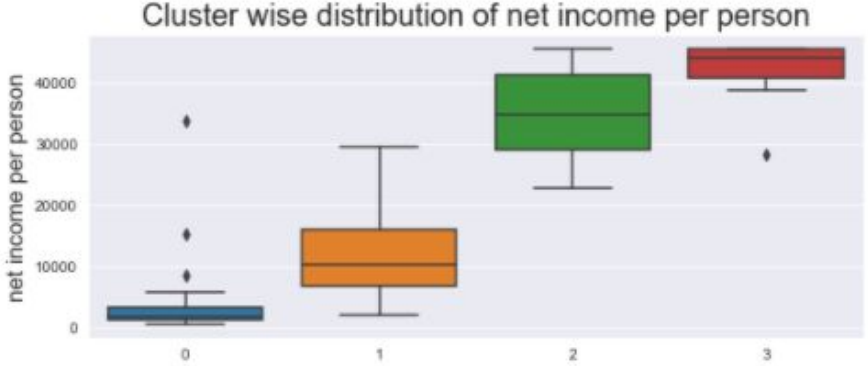
HIERARCHICAL CLUSTERING INFERENCES

- ➔ Using Hierarchical Clustering - Complete Linkage, 4 cluster centers were identified.
- ➔ Countries belonging to the Cluster 0 are considered to be extremely weak countries. They have high under 5 child deaths/1k, low GDP per capita and low net income per person.
- ➔ Countries belonging to Cluster 3 are strong countries. They have low child deaths/1k, high GDP per capita and high net income per person.
- ➔ The cluster centers calculated by the hierarchical clustering - complete linkage were finalized as there was a distinct identification among the clusters.

FINAL CLUSTERING USING KMEANS







UNDERDEVELOPED COUNTRIES

➔ The country names as per Hierarchical Clustering Complete Linkage are:

```
array(['Afghanistan', 'Angola', 'Benin', 'Burkina Faso', 'Burundi',
       'Cameroon', 'Central African Republic', 'Chad', 'Congo, Dem. Rep.',
       "Cote d'Ivoire", 'Equatorial Guinea', 'Guinea', 'Guinea-Bissau',
       'Haiti', 'Malawi', 'Mali', 'Mozambique', 'Niger', 'Nigeria',
       'Sierra Leone', 'Uganda', 'Zambia'], dtype=object)
```

➔ The country names as per KMeans are:

```
array(['Afghanistan', 'Angola', 'Benin', 'Burkina Faso', 'Burundi',
       'Cameroon', 'Central African Republic', 'Chad', 'Comoros',
       'Congo, Dem. Rep.', 'Congo, Rep.', "Cote d'Ivoire",
       'Equatorial Guinea', 'Eritrea', 'Gabon', 'Gambia', 'Ghana',
       'Guinea', 'Guinea-Bissau', 'Haiti', 'Kenya', 'Kiribati', 'Lao',
       'Lesotho', 'Liberia', 'Madagascar', 'Malawi', 'Mali', 'Mauritania',
       'Mozambique', 'Namibia', 'Niger', 'Nigeria', 'Pakistan', 'Rwanda',
       'Senegal', 'Sierra Leone', 'Solomon Islands', 'Sudan', 'Tanzania',
       'Timor-Leste', 'Togo', 'Uganda', 'Yemen', 'Zambia'], dtype=object)
```

CRITICAL COUNTRIES

Below is a list of the top 5 countries requiring immediate attention, decided based on sorting the dataframe in ascending order on gdpp(GDP per capita) and income(net income per person) and descending order on child_mort(child deaths under 5 years/1000).

```
cluster_0_df.sort_values(by = ['gdpp', 'income', 'child_mort'], ascending = [True, True, False]).head()
```

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	kmeans_labels
26	Burundi	93.6	20.6052	26.7960	90.552	764.0	12.30	57.7	6.26	231.0	0
88	Liberia	89.3	62.4570	38.5860	302.802	700.0	5.47	60.8	5.02	327.0	0
37	Congo, Dem. Rep.	116.0	137.2740	26.4194	165.664	609.0	20.80	57.5	6.54	334.0	0
112	Niger	123.0	77.2560	17.9568	170.868	814.0	2.55	58.8	7.49	348.0	0
132	Sierra Leone	160.0	67.0320	52.2690	137.655	1220.0	17.20	55.0	5.20	399.0	0

SUMMARY

- ➔ The number of clusters for KMeans Algorithm were decided as 3 based on the elbow curve.
- ➔ We could see some distinction between the clusters for features like net income per person, GDP per capita, and child mortality/1k.
- ➔ Using Hierarchical Clustering Single Linkage, cutting the dendrogram at number of clusters as 3, gave unwanted clusters.
- ➔ Then using the complete linkage hierarchical clustering, we could confirm that there were indeed 4 clusters in the data. Countries belonging to cluster 0 were the weakest and the ones belonging to cluster 3 the strongest.