Country Aid using Socio-economic and health factors

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

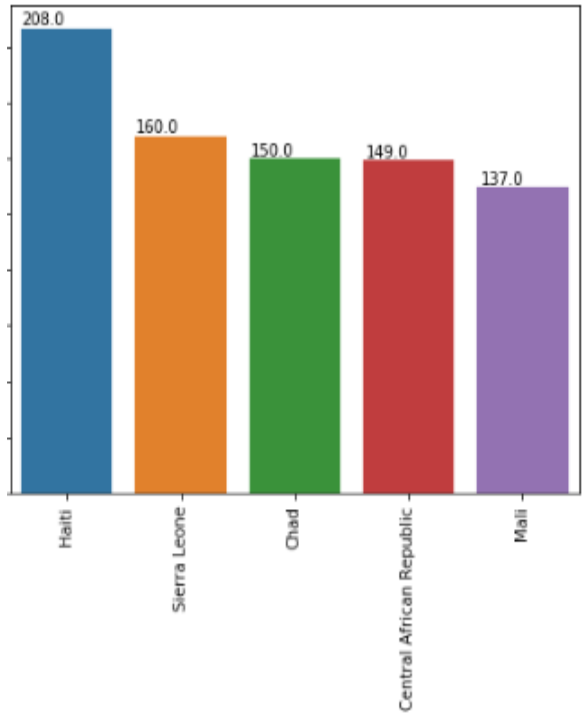
Business Goal

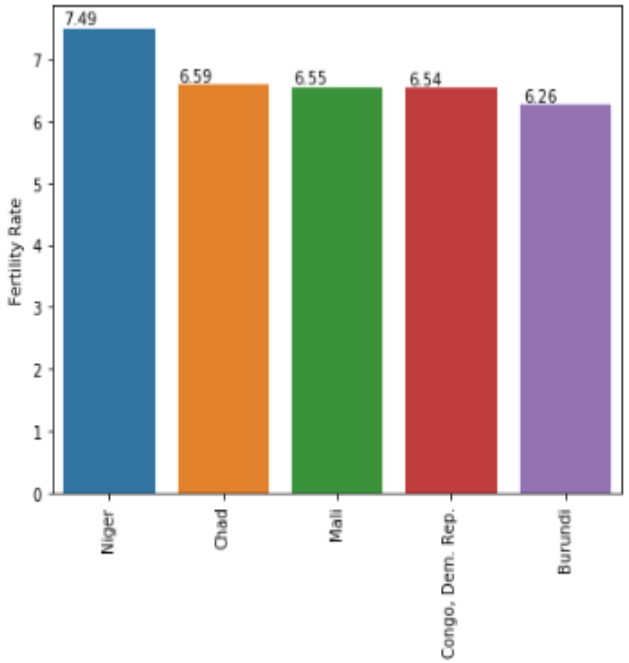
- Our job is to categorise the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.

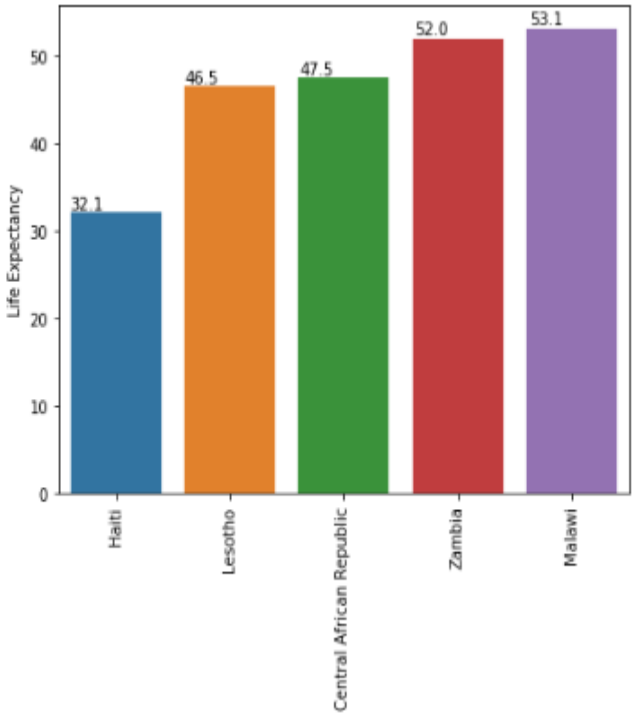
Steps for analysis followed:

1. Reading and Understanding the Data
2. Data Cleansing
3. Data Visualization
4. Data Preparation
5. Hopkins Statistics Test
6. Model Building
7. Step 8: Final Analysis

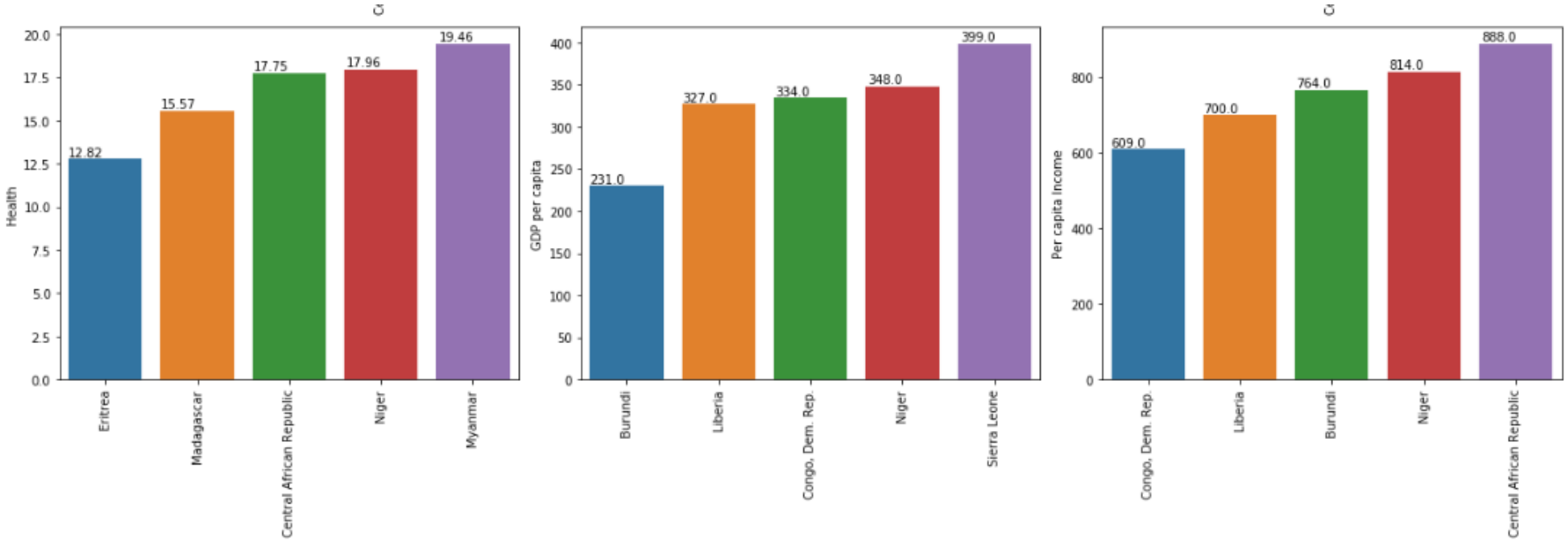
Lowest 5 countries for each factor:



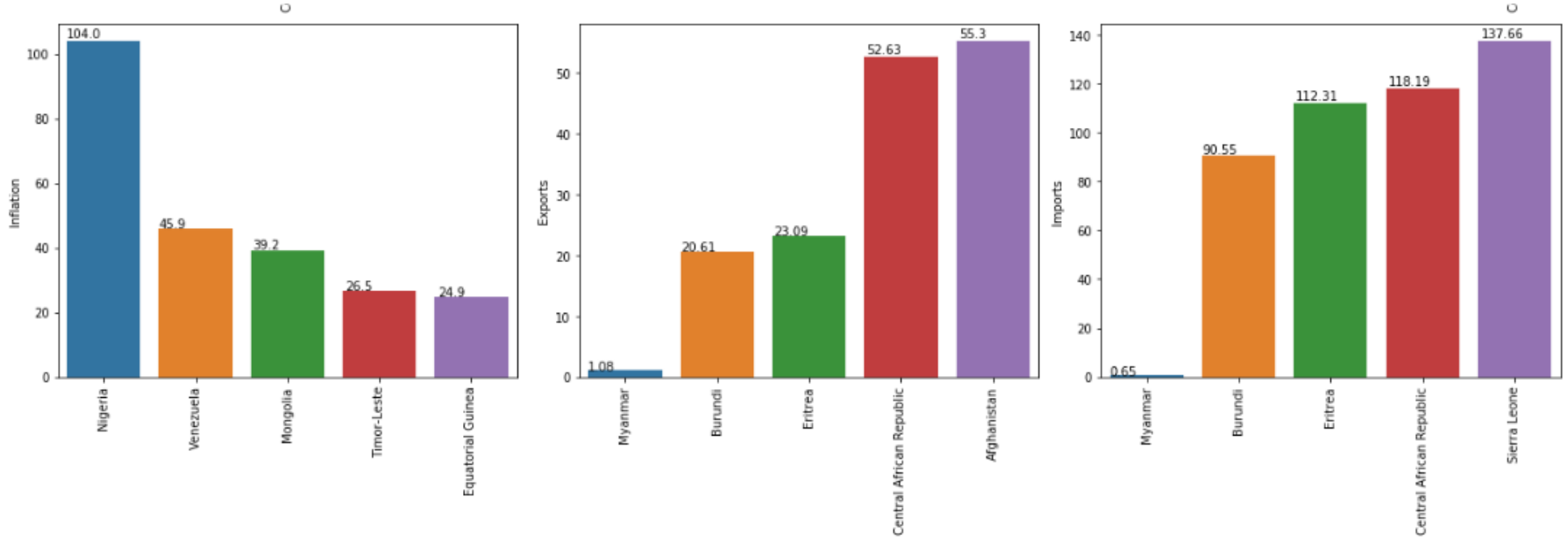




Lowest 5 countries for each factor:



Lowest 5 countries for each factor:



Correlation among variables:

- child_mortality and life_expentency are highly correlated with correlation of -0.89
- child_mortality and total_fertility are highly correlated with correlation of 0.85
- imports and exports are highly correlated with correlation of 0.99
- life_expentency and total_fertility are highly correlated with correlation of -0.76

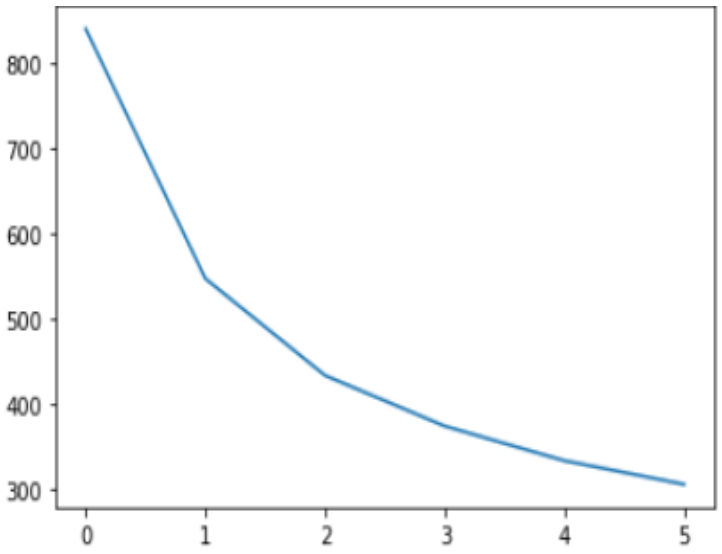
Hopkins Statistics Test

- The **Hopkins statistic** is a way of measuring the cluster tendency of a data set.
- We got a Hopkins score of 0.91 by performing Hopkins Statistics
- This means that this problem is ideal for clustering

Rescaling the Features(Standardization)

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	1.344012	-0.569638	-0.566983	-0.598844	-0.851772	0.263649	-1.693799	1.926928	-0.702259
1	-0.547543	-0.473873	-0.440417	-0.413679	-0.387025	-0.375251	0.663053	-0.865911	-0.498726
2	-0.272548	-0.424015	-0.486295	-0.476198	-0.221124	1.123260	0.686504	-0.035427	-0.477434
3	2.084186	-0.381264	-0.534113	-0.464070	-0.612136	1.936405	-1.236499	2.154642	-0.530950
4	-0.709457	-0.086754	-0.178431	0.139659	0.125202	-0.768917	0.721681	-0.544433	-0.032042

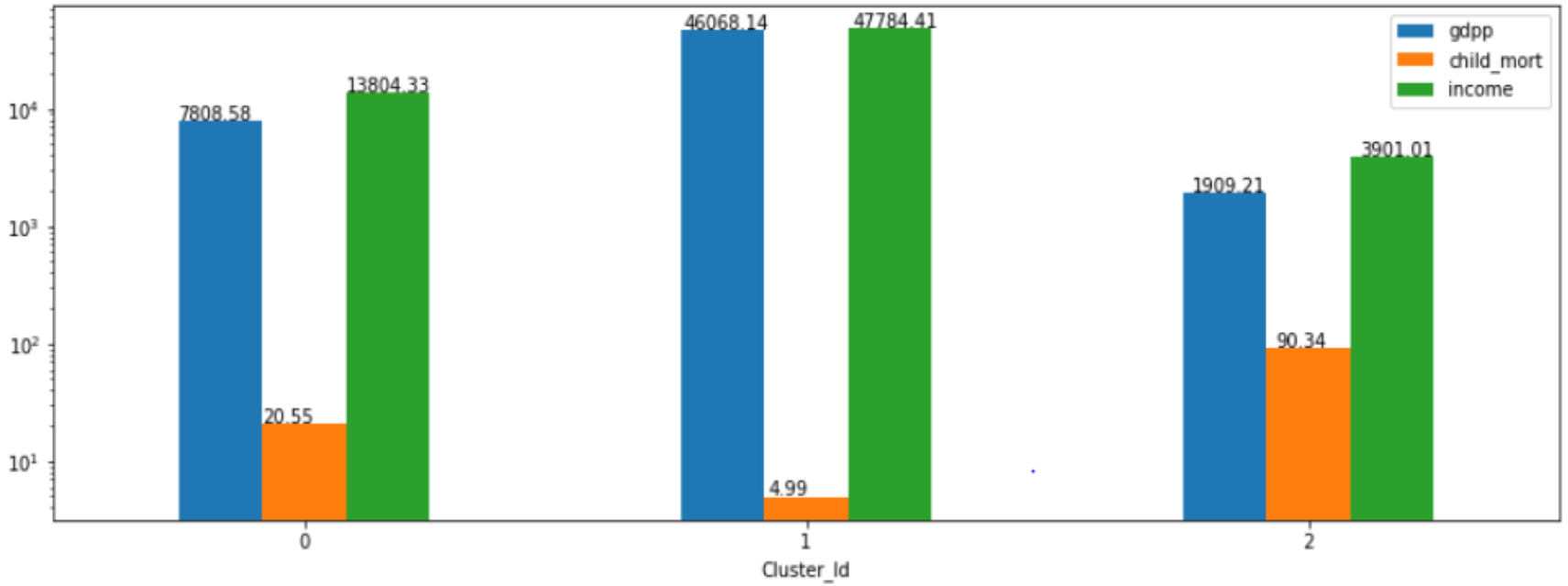
Model Building




```
For n_clusters=2, the silhouette score is 0.4693985523558032
For n_clusters=3, the silhouette score is 0.4070898884641155
For n_clusters=4, the silhouette score is 0.3944993340525435
For n_clusters=5, the silhouette score is 0.3864291132976758
For n_clusters=6, the silhouette score is 0.2993922830096998
For n_clusters=7, the silhouette score is 0.28829902269947366
For n_clusters=8, the silhouette score is 0.32145769851384637
```

From elbow curve and Silhouette score, we found out that 3 is the ideal number of clusters

Cluster Profiling(for K-means):

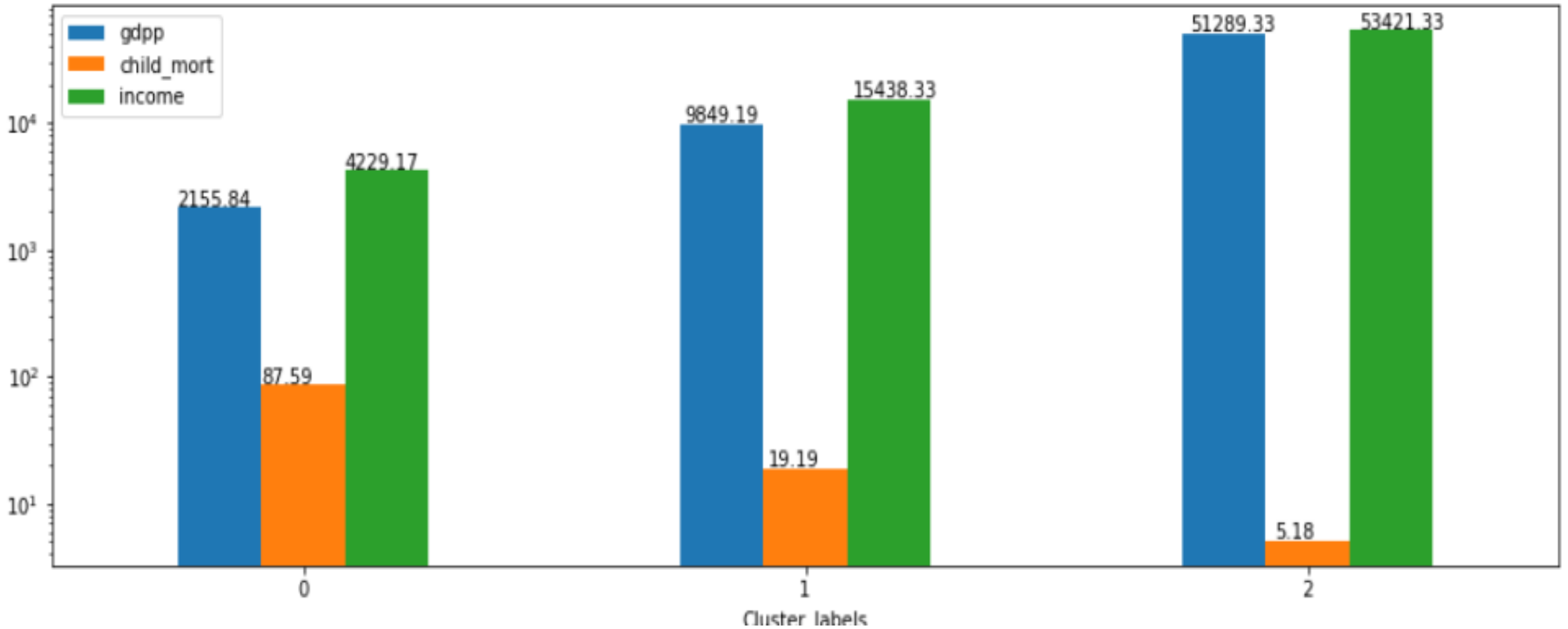


Cluster number 2 has the least developed countries and hence these are in dire need of aid

The top 5 countries in dire need of aid: (K-means)

- The 5 countries in dire need of aid:
- 1) Sierra Leone
- 2) Haiti
- 3) Chad
- 4) Central African Republic
- 5) Mali

Cluster Profiling(for Heirarichal method):



Cluster 0 has countries which are in dire need of aid

The top 5 countries in dire need of aid: (Heirarichal)

1) Sierra Leone
2) Haiti
3) Chad
4) Central African Republic
5) Mali

Final Analysis:

- We got the same country names for both K Means and Heirarichal algorithm.
- The final 5 countries which needs the most focus :

1) **Sierra Leone**
2) **Haiti**
3) **Chad**
4) **Central African Republic**
5) **Mali**