# CLUSTERING OF COUNTRIES

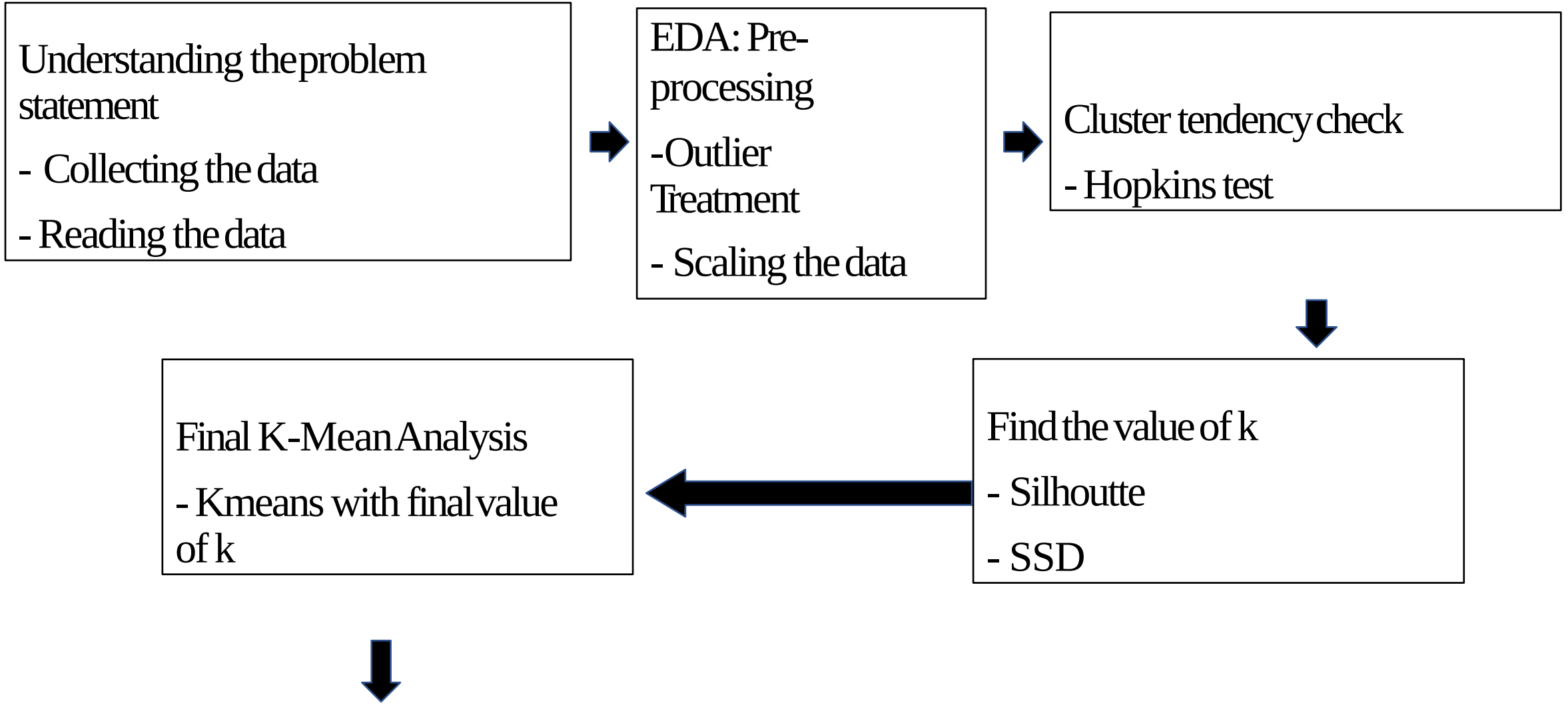
- Abstract :

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.

- Problem statement:

After the recent funding programs, 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. As a analyst we need to suggest the countries which the CEO needs to focus on the most

# ANALYSIS APPROACH



# ANALYSIS APPROACH CONT.

Cluster Profiling

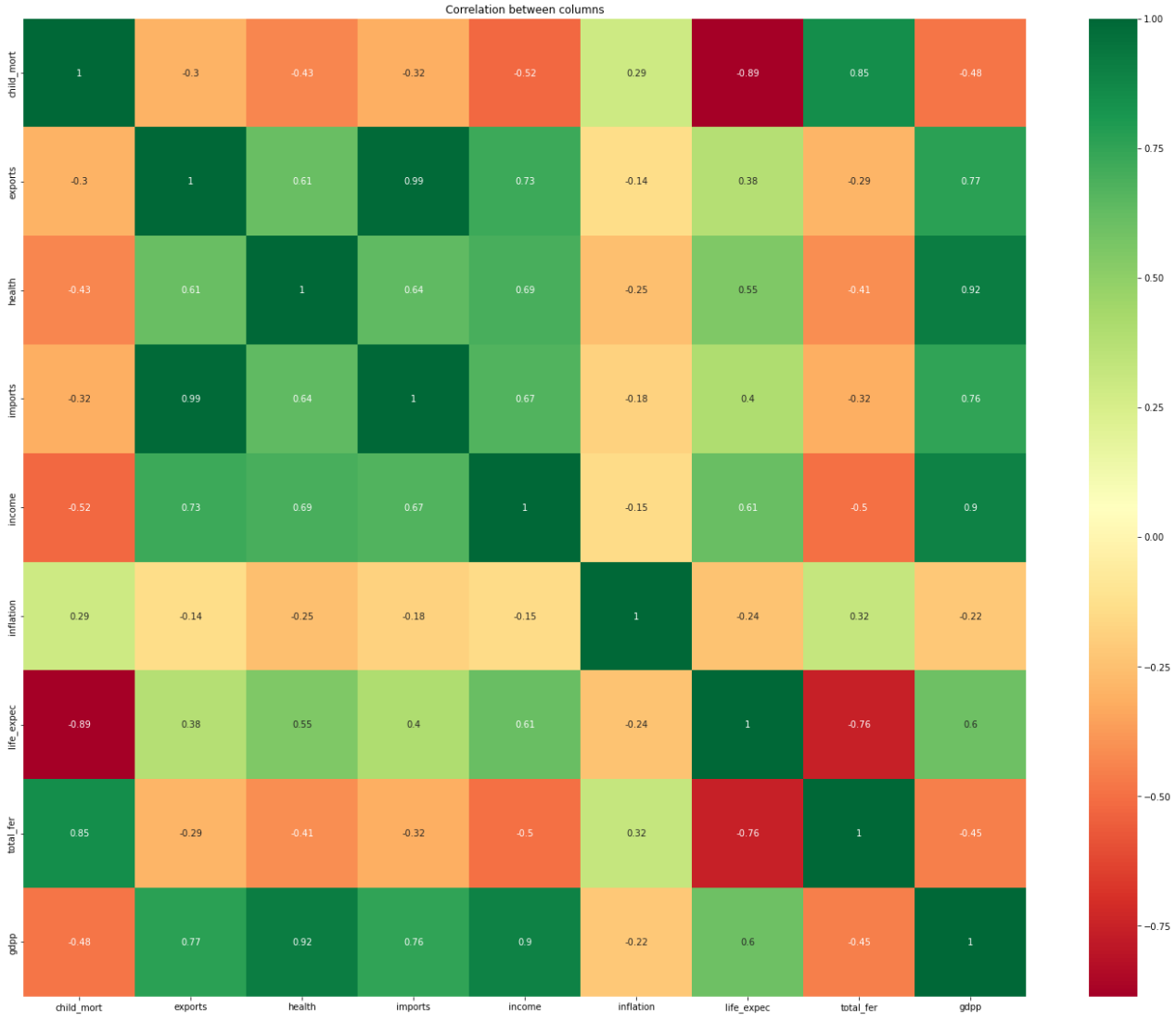
- Making sense out of various cluster

Hierarchical clustering

- Single linkage
- Complete linkage

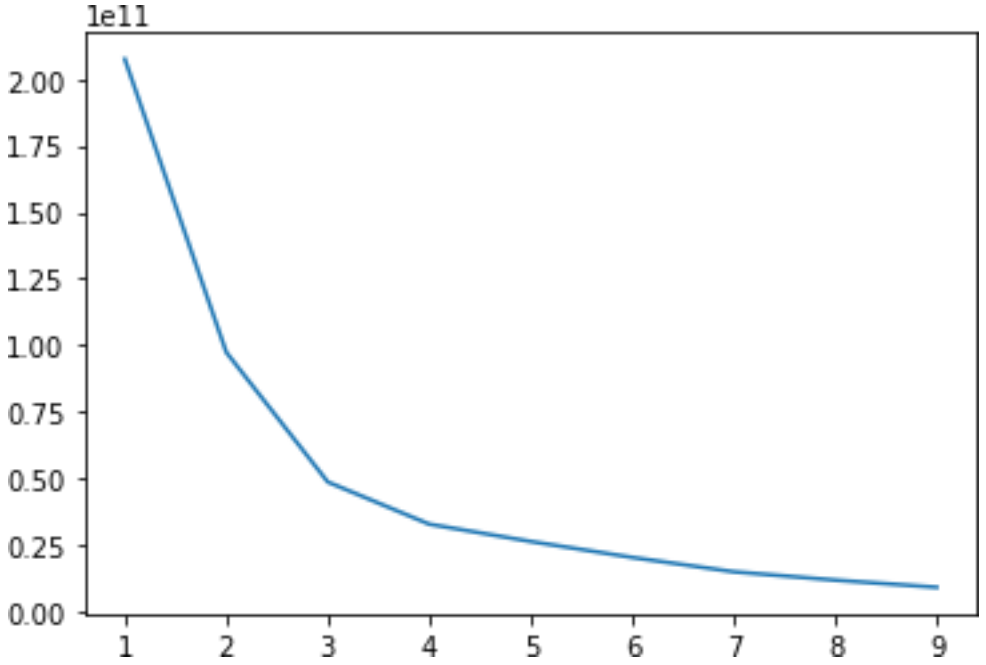
-profiling

# CO-RRELATION




- In data cleaning step, Here we first find out outliers in every columns in dataset, as we cannot remove the outliers as it can manipulate our result
- After that we did standard scalling, as export, health and imports were given in percentage, we convert it into back to normal
- Looking at the heatmap we cansee some variable like(total fertility and child morality) ,(income and gdpp) and (import and export) have high co-relation

# K-MEANS CLUSTERING



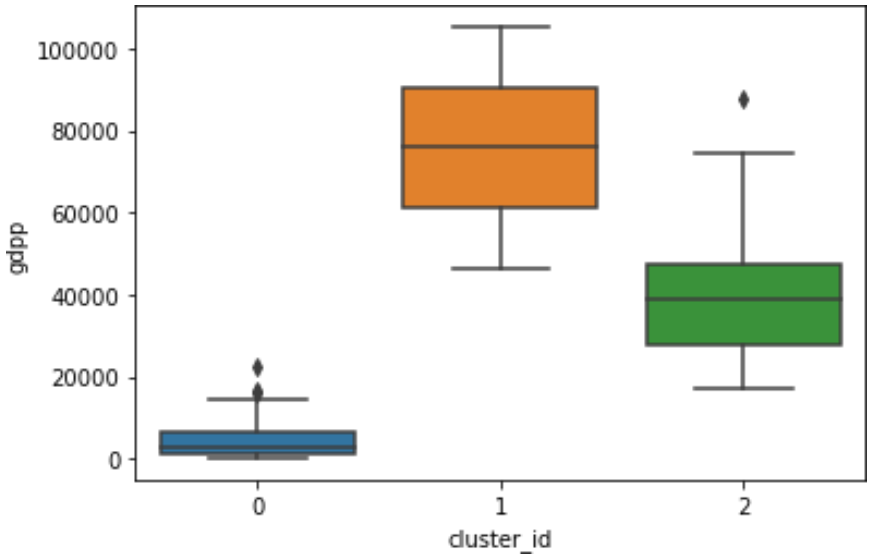

Sumof Squared Distance

```
For n_clusters=2, the silhouette score is 0.712586926709893
For n_clusters=3, the silhouette score is 0.7010999403856036
For n_clusters=4, the silhouette score is 0.6617474309946076
For n_clusters=5, the silhouette score is 0.5902589833204496
For n_clusters=6, the silhouette score is 0.5314887374931582
For n_clusters=7, the silhouette score is 0.5339607116039059
For n_clusters=8, the silhouette score is 0.5245652209238603
```

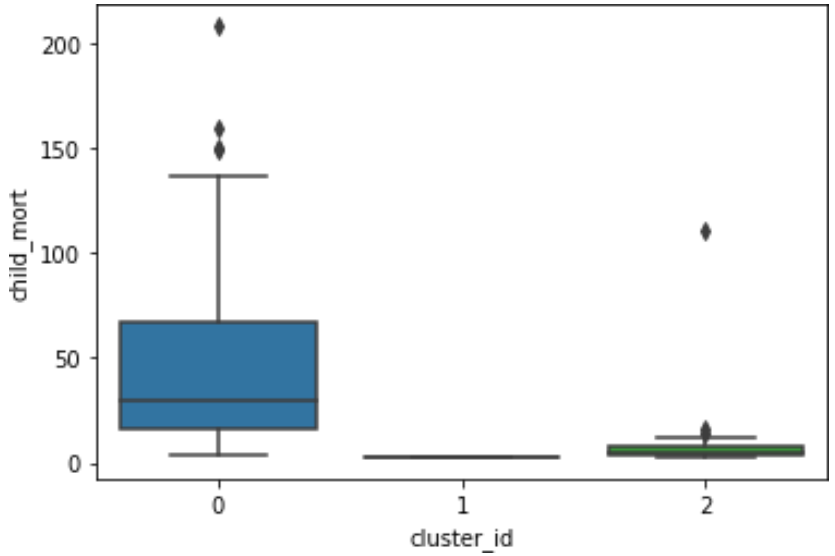
Silhoutte analysis

By looking at silhouette analysis we see that the silhouette score is 2nd maximum at k=3, and by looking at sum of squared method the elbow is at 2,3 and 4 , so we are going with k=3

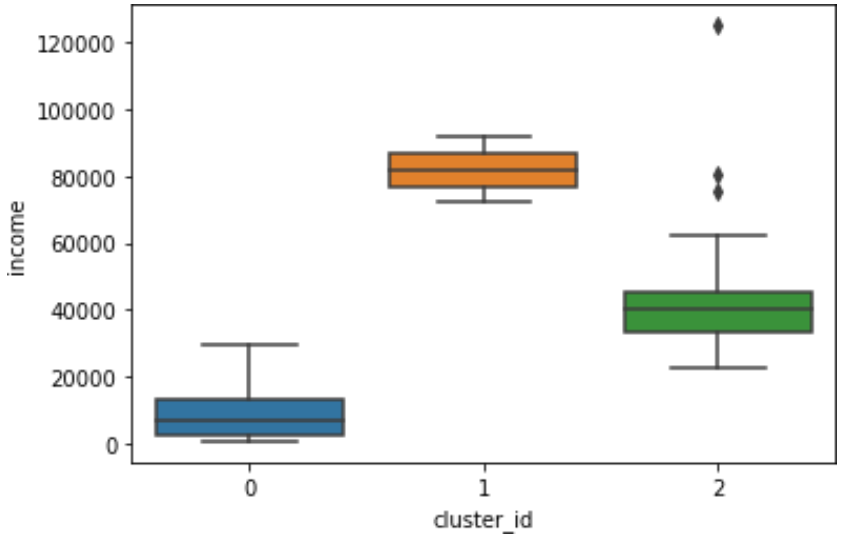
# K-MEANS CLUSTERING




Box plot for various cluster id to gdpp ,we see for cluster id 0 the gdpp is very low




Box plot for various cluster id to child morality ,we see that for cluster id 0 the child morality is high



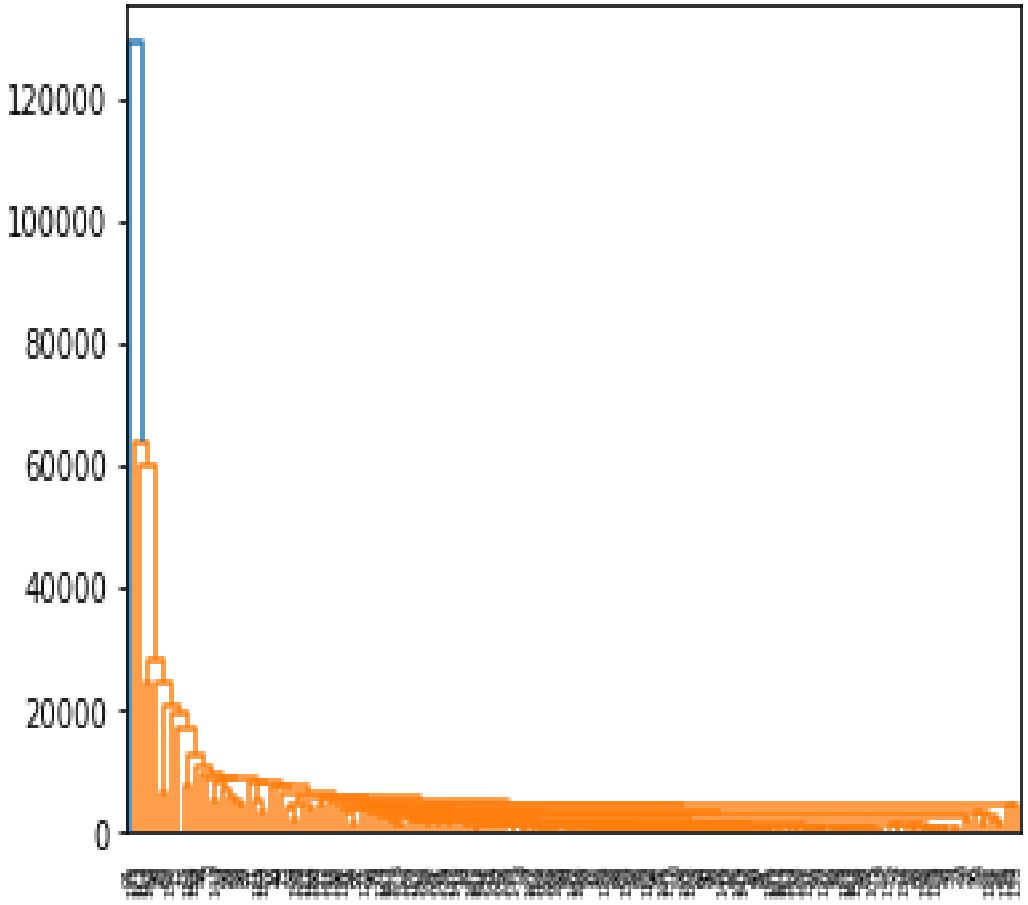

Box plot for various cluster id to income ,we see that for cluster id 0, the income is very low

# K-Means clustering

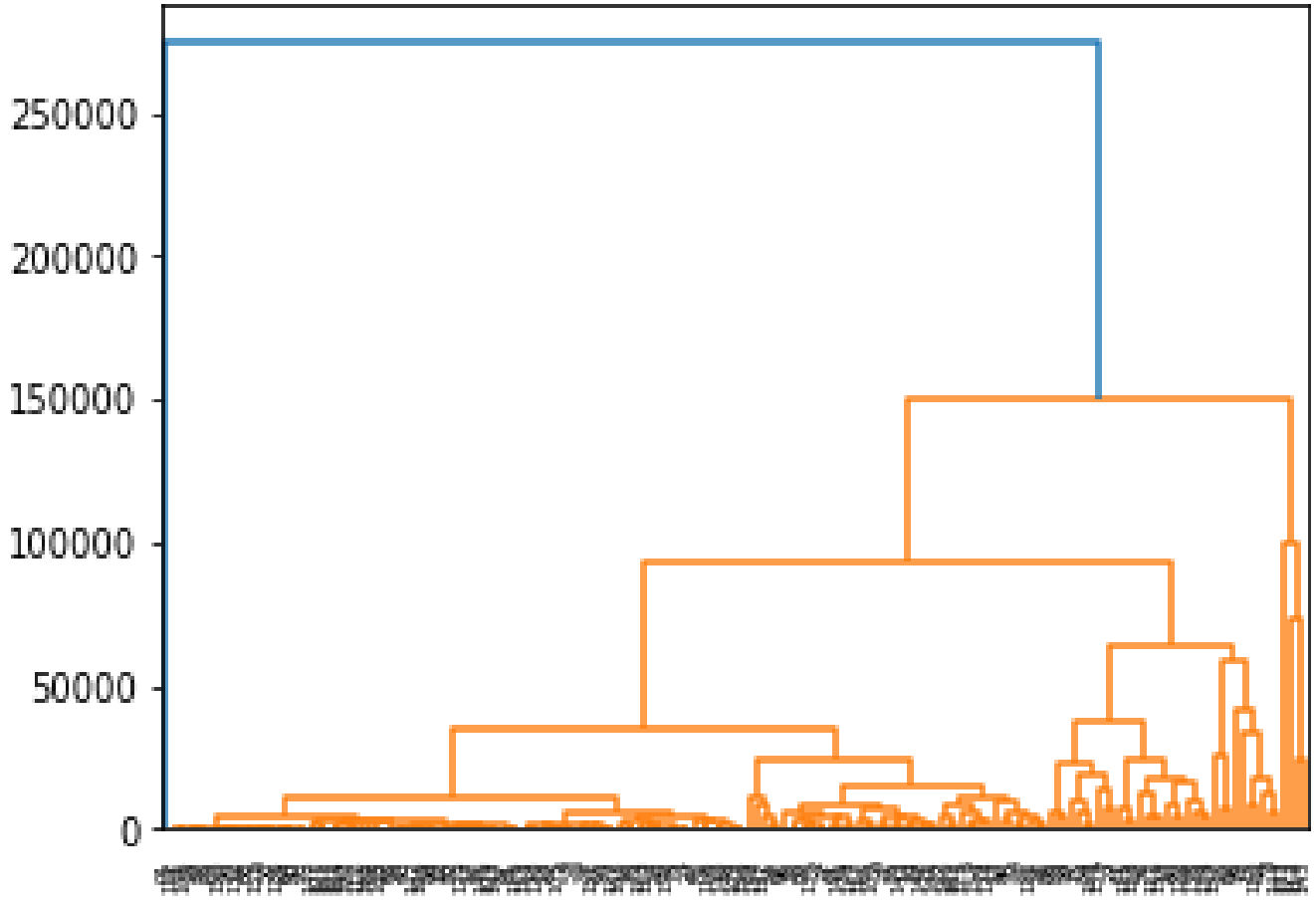
Top 10 countries under cluster id 0 are

| country | child_mort | exports | health | imports | income | inflation | life_expec | total_fer | gdpp | cluster_id |
|---|---|---|---|---|---|---|---|---|---|---|
| Haiti | 208.0 | 15.3 | 6.91 | 64.7 | 1500 | 5.45 | 32.1 | 3.33 | 662 | 0 |
| Sierra Leone | 160.0 | 16.8 | 13.10 | 34.5 | 1220 | 17.20 | 55.0 | 5.20 | 399 | 0 |
| Chad | 150.0 | 36.8 | 4.53 | 43.5 | 1930 | 6.39 | 56.5 | 6.59 | 897 | 0 |
| Central African Republic | 149.0 | 11.8 | 3.98 | 26.5 | 888 | 2.01 | 47.5 | 5.21 | 446 | 0 |
| Mali | 137.0 | 22.8 | 4.98 | 35.1 | 1870 | 4.37 | 59.5 | 6.55 | 708 | 0 |
| Nigeria | 130.0 | 25.3 | 5.07 | 17.4 | 5150 | 104.00 | 60.5 | 5.84 | 2330 | 0 |
| Niger | 123.0 | 22.2 | 5.16 | 49.1 | 814 | 2.55 | 58.8 | 7.49 | 348 | 0 |
| Angola | 119.0 | 62.3 | 2.85 | 42.9 | 5900 | 22.40 | 60.1 | 6.16 | 3530 | 0 |
| Congo, Dem. Rep. | 116.0 | 41.1 | 7.91 | 49.6 | 609 | 20.80 | 57.5 | 6.54 | 334 | 0 |
| Burkina Faso | 116.0 | 19.2 | 6.74 | 29.6 | 1430 | 6.81 | 57.9 | 5.87 | 575 | 0 |

# HIERARCHICAL CLUSTERING

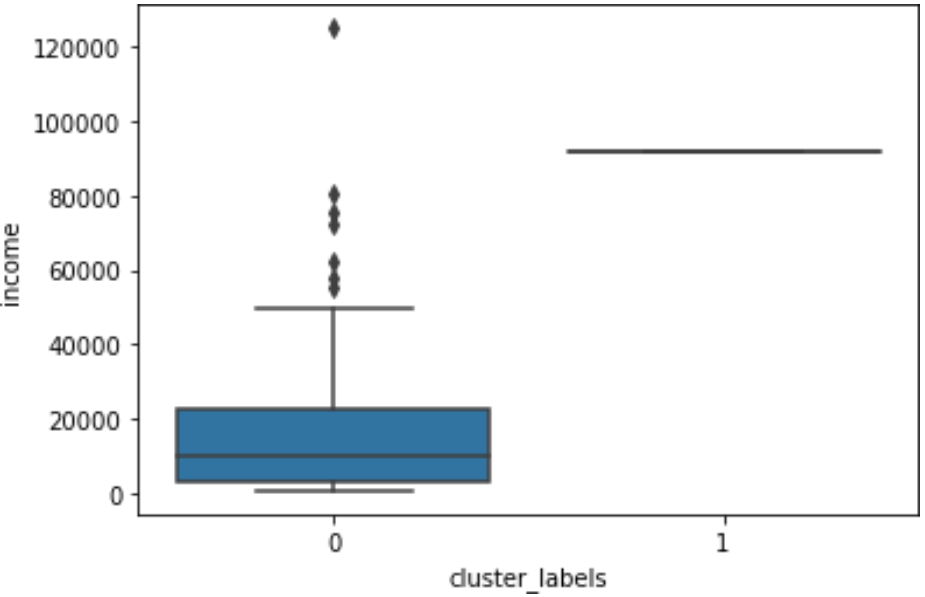



Single method hierarchicalclustering

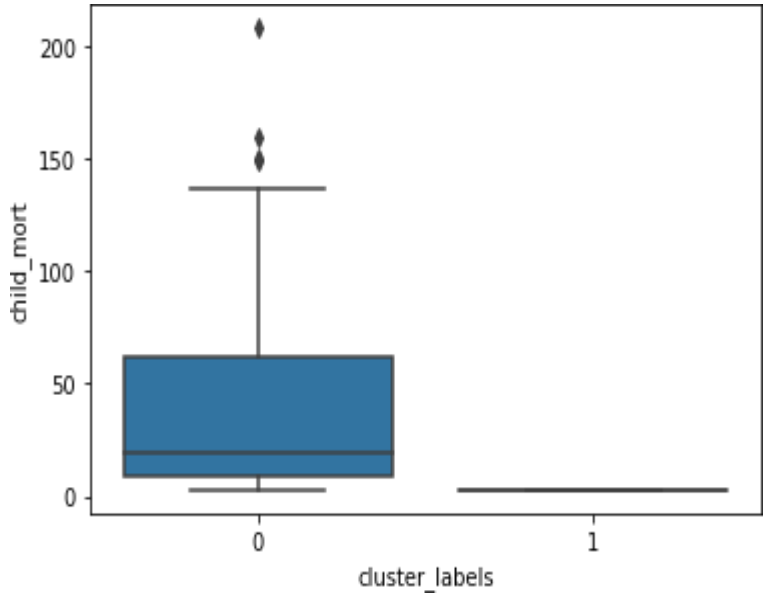



We are going for complete method hierarchical clustering as single method is not that clear, by looking at this dendogram, taking n-clusters as2

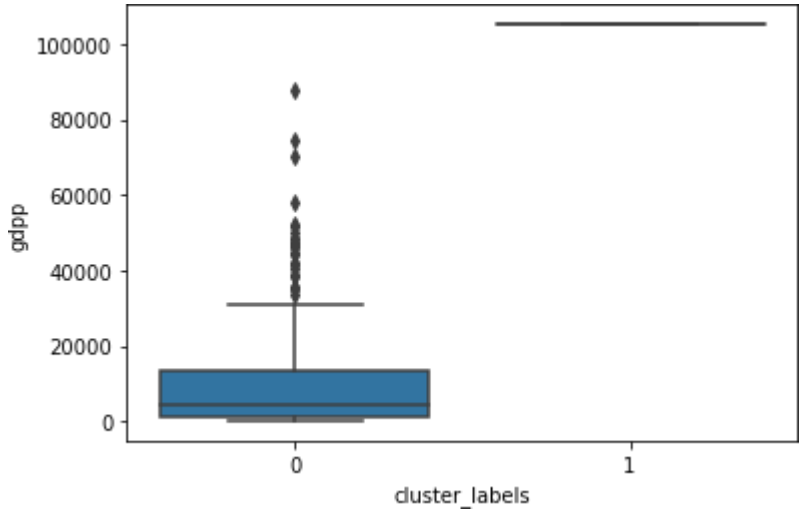
# HIERARCHICAL CLUSTERING




Box plot for various cluster labels to income ,we see for cluster label 0 the income is very low




Box plot for various cluster label to child morality ,we see for cluster label 0 the child morality is very low




Box plot for various cluster label to gdpp , we see for cluster label 0 the gdpp is very low

# Hierarchical Clustering

Top 10 countries under cluster labels 0 are

| country | child_mort | exports | health | imports | income | inflation | life_expec | total_fer | gdpp | cluster_labels |
|---|---|---|---|---|---|---|---|---|---|---|
| Haiti | 208.0 | 15.3 | 6.91 | 64.7 | 1500 | 5.45 | 32.1 | 3.33 | 662 | 0 |
| Sierra Leone | 160.0 | 16.8 | 13.10 | 34.5 | 1220 | 17.20 | 55.0 | 5.20 | 399 | 0 |
| Chad | 150.0 | 36.8 | 4.53 | 43.5 | 1930 | 6.39 | 56.5 | 6.59 | 897 | 0 |
| Central African Republic | 149.0 | 11.8 | 3.98 | 26.5 | 888 | 2.01 | 47.5 | 5.21 | 446 | 0 |
| Mali | 137.0 | 22.8 | 4.98 | 35.1 | 1870 | 4.37 | 59.5 | 6.55 | 708 | 0 |
| Nigeria | 130.0 | 25.3 | 5.07 | 17.4 | 5150 | 104.00 | 60.5 | 5.84 | 2330 | 0 |
| Niger | 123.0 | 22.2 | 5.16 | 49.1 | 814 | 2.55 | 58.8 | 7.49 | 348 | 0 |
| Angola | 119.0 | 62.3 | 2.85 | 42.9 | 5900 | 22.40 | 60.1 | 6.16 | 3530 | 0 |
| Congo, Dem. Rep. | 116.0 | 41.1 | 7.91 | 49.6 | 609 | 20.80 | 57.5 | 6.54 | 334 | 0 |
| Burkina Faso | 116.0 | 19.2 | 6.74 | 29.6 | 1430 | 6.81 | 57.9 | 5.87 | 575 | 0 |

# SUMMARY

As by both k-means and hierarchical clustering method- we have got same countries which required aid.
The following are the top 10 countries which are in direst need of help

| | country |
|---|---|
| 0 | Haiti |
| 1 | Sierra Leone |
| 2 | Chad |
| 3 | Central African Republic |
| 4 | Mali |
| 5 | Nigeria |
| 6 | Niger |
| 7 | Angola |
| 8 | Congo, Dem. Rep. |
| 9 | Burkina Faso |

# THANK YOU

-BHARGAV RAM