



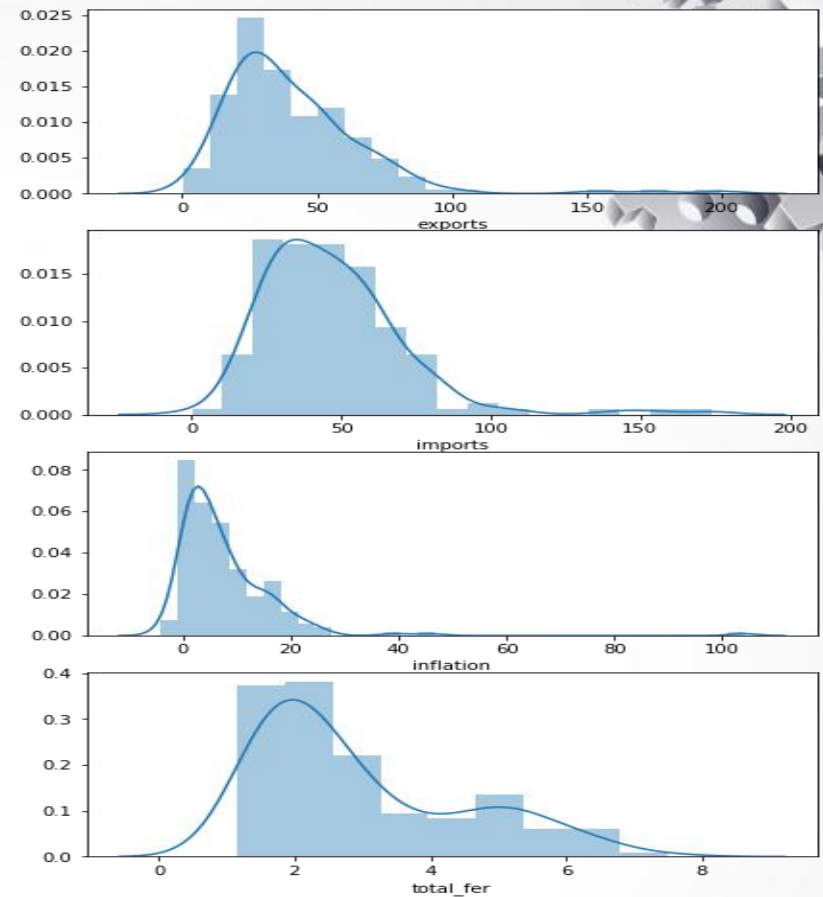
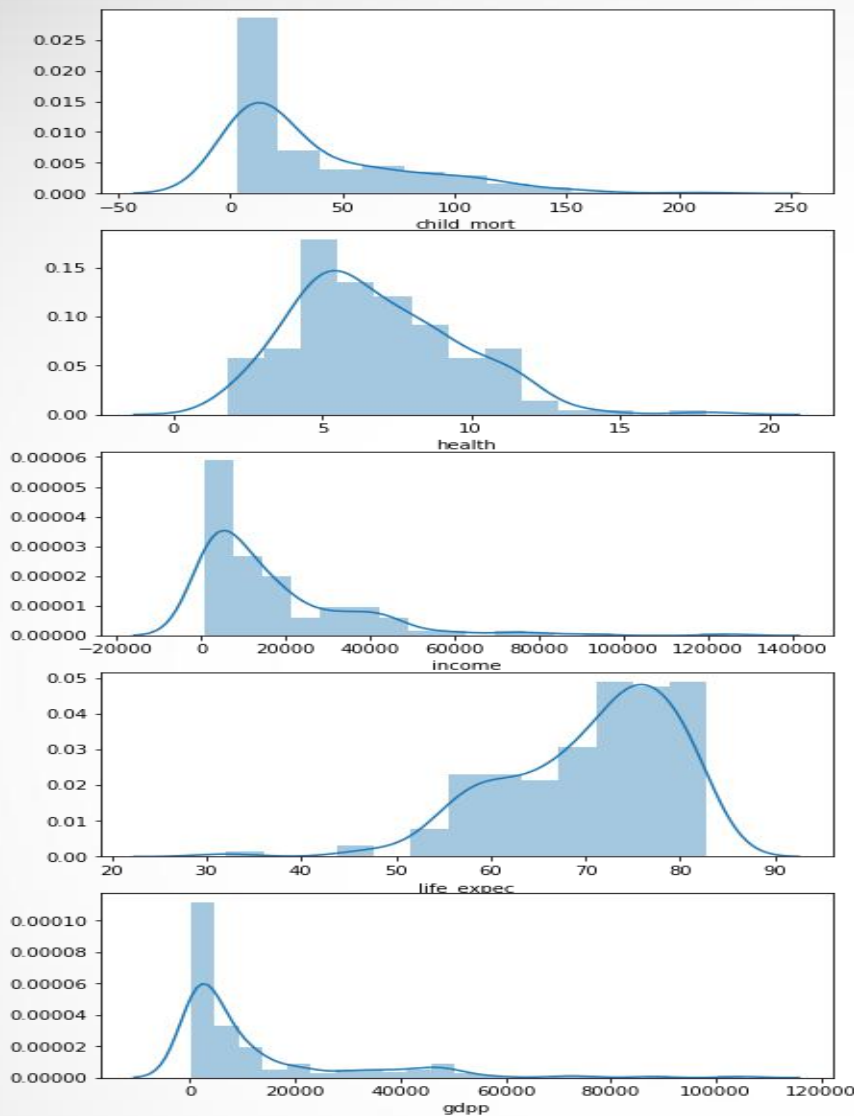
# "Clustering of Countries"

# • Problem Statement

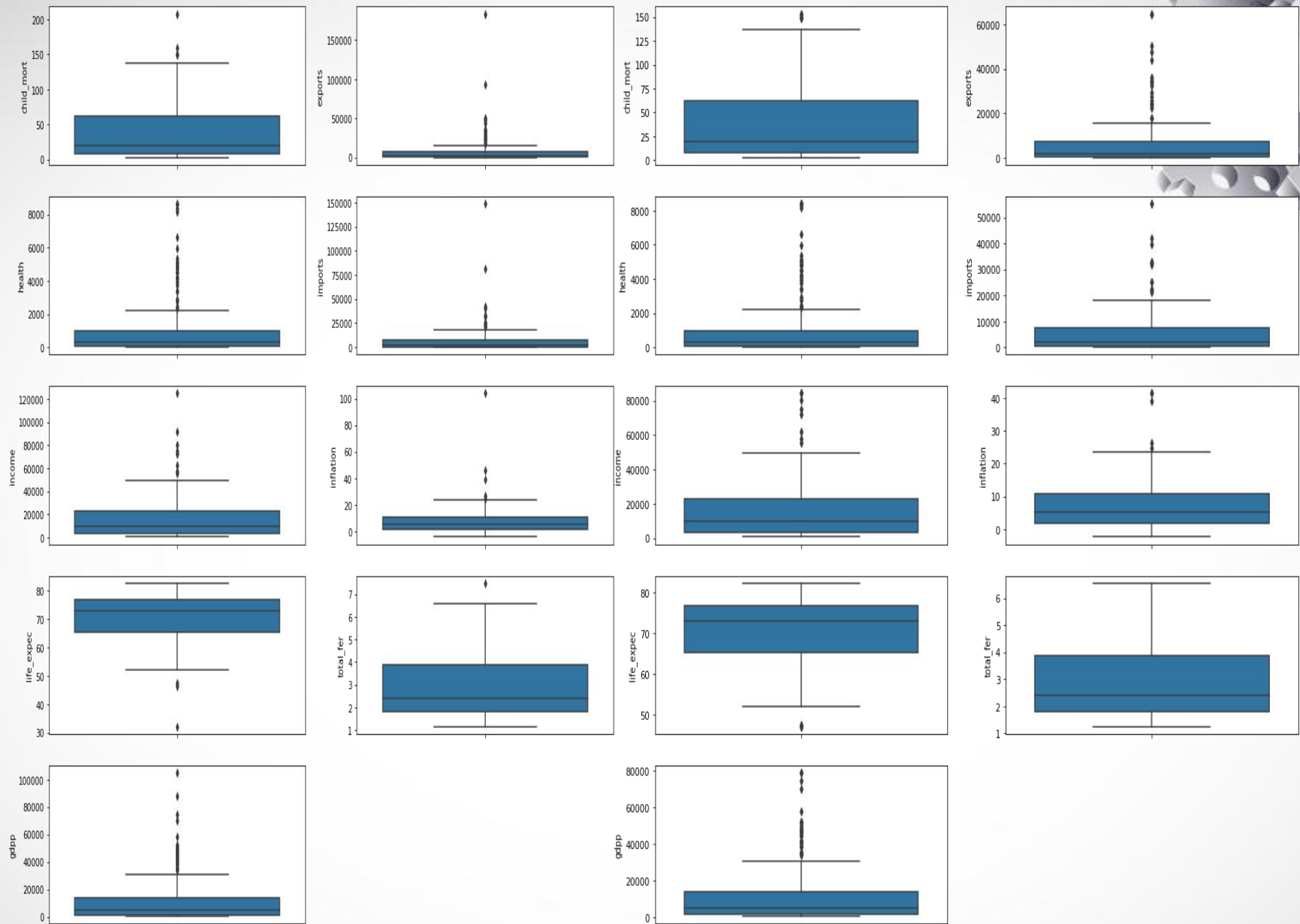
- HELP International is an international humanitarian NGO that is committed to fight against poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities.
- We need to cluster the countries & choosing the countries that are in the direst need of aid by these three variables - [gdpp, child\_mort and income] .

# • Solution Approach

- ❖ Analyse the Country Dataset
- ❖ Performed the Exploratory Data Analysis
- ❖ Visualise, Identified And Treated the outlier
- ❖ Choosing the optimum no of cluster for K-Mean Clustering Approach
- ❖ Visualise to find the significant cluster
- ❖ Visualise the cluster of the basis of [gdpp, child\_mort and income]
- ❖ Visualise the clustering using Hierarchical Clustering(single linkage & Complete linkage)
- ❖ Choosing the optimum no of cluster for Hierarchical Clustering Approach
- ❖ Visualise to find the significant cluster
- ❖ Visualise the cluster of the basis of [gdpp, child\_mort and income]
- ❖ Identify the the countries that are in the direst need of aid

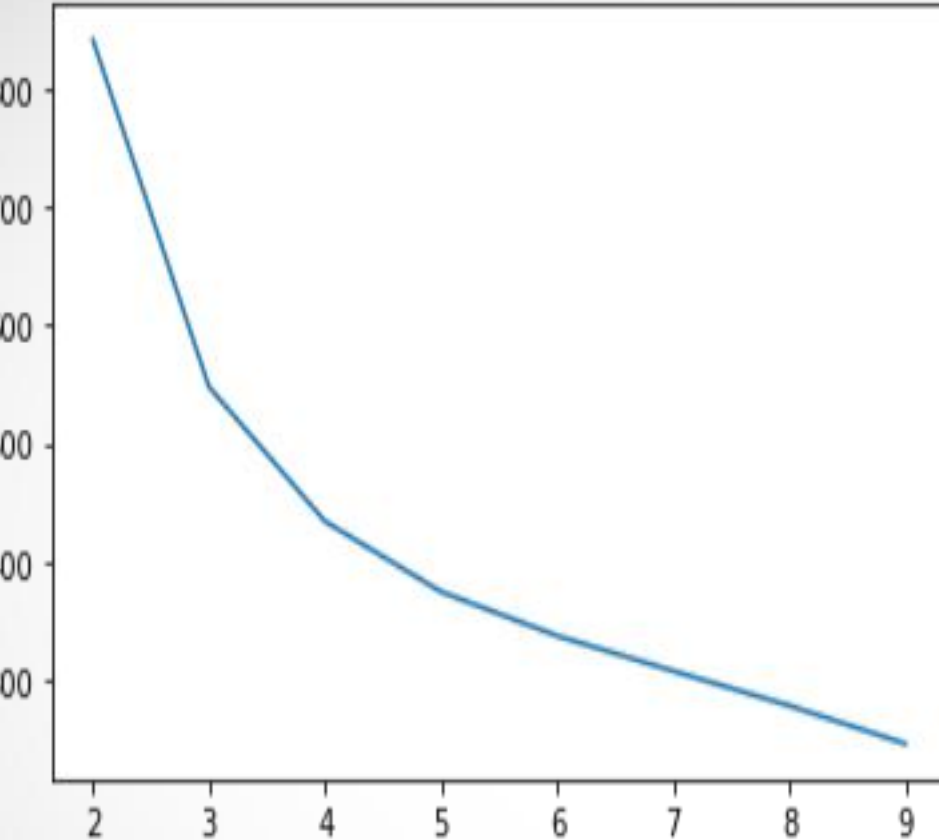


- ❖ From the above plot its found that 'child\_mort', 'exports', 'health', 'imports', 'inflation', 'gdpp', 'life\_expec' are almost normally distributed. But for 'income' there are much number of countries income around 400k and for 'total\_fer' there are much number of countries total fertility rate around 5

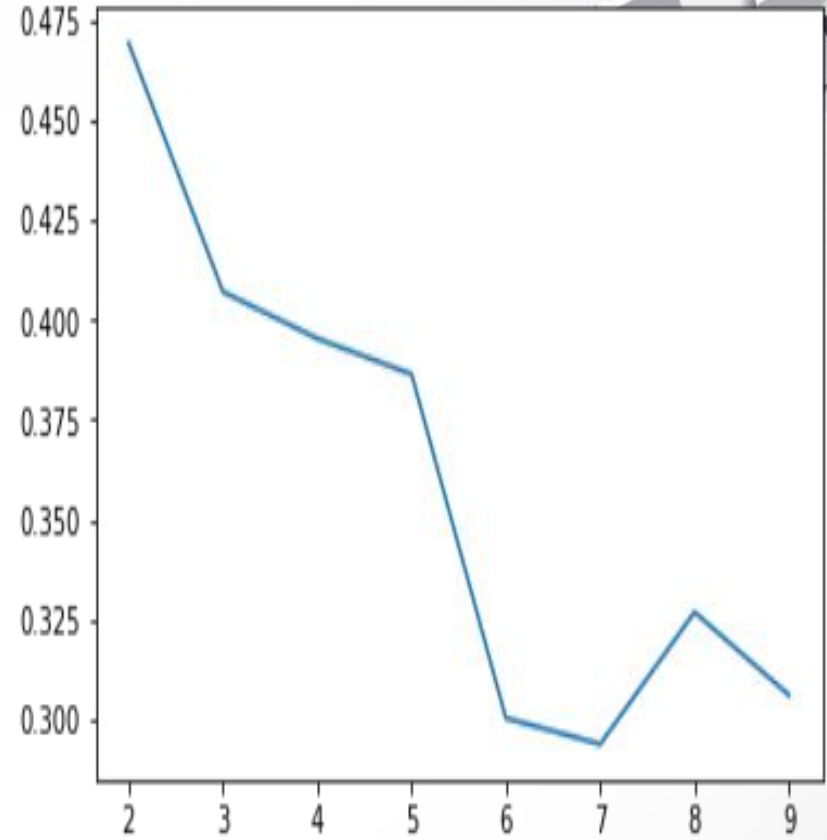


Though there are outliers we have not removed as we don't want to lose the data so we have capped at 1% to 99%.

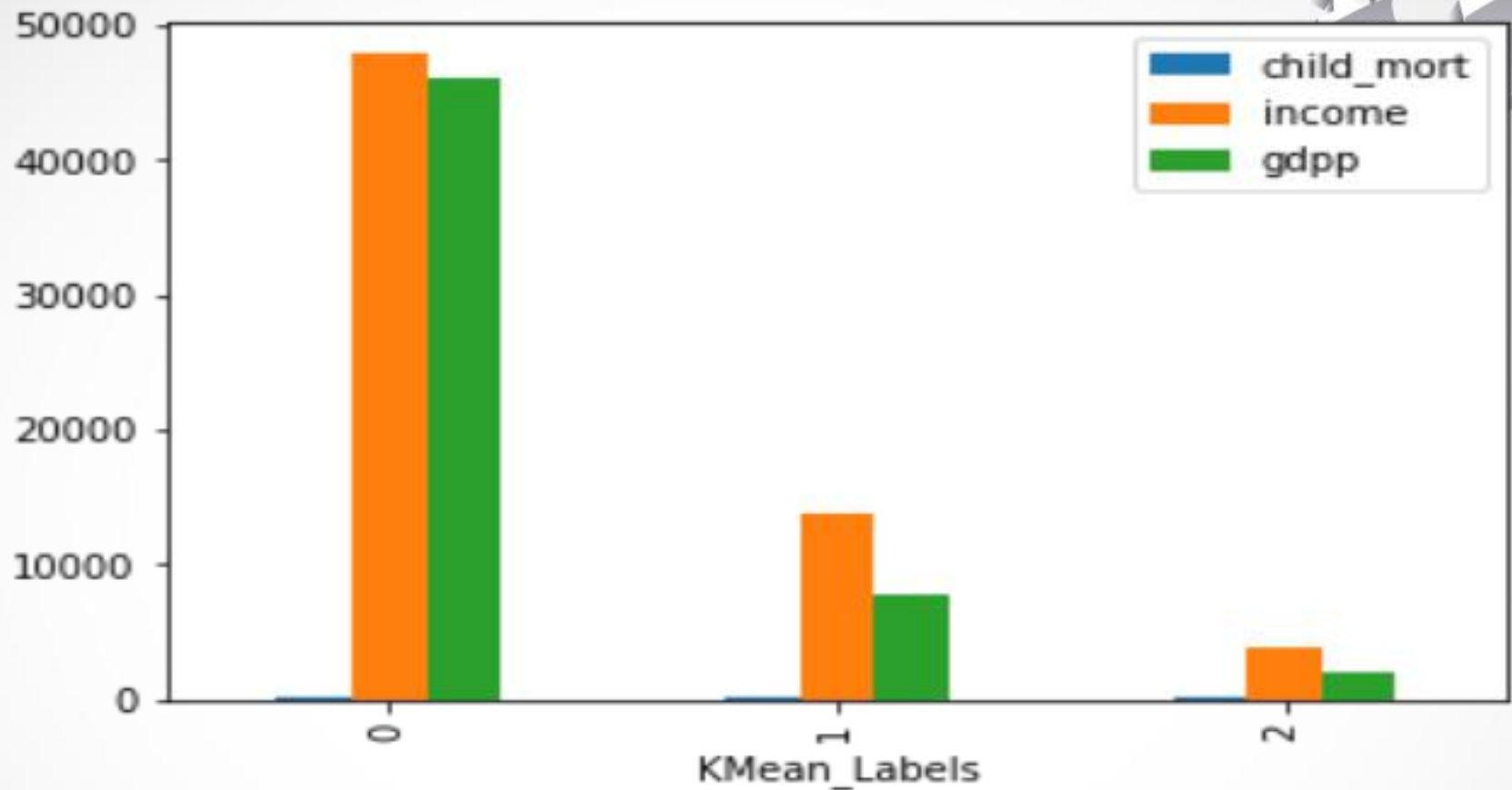
## Elbow Curve



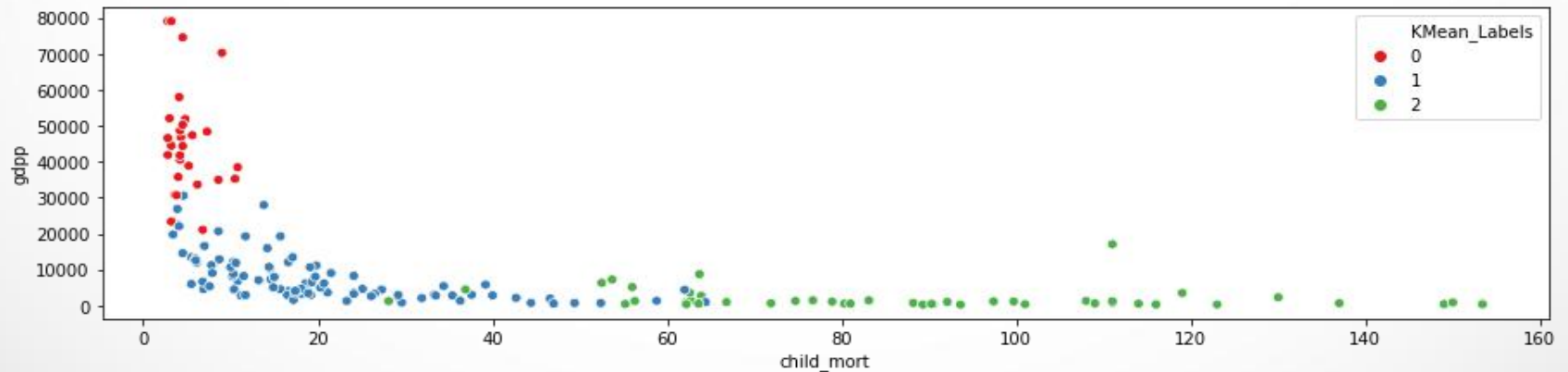
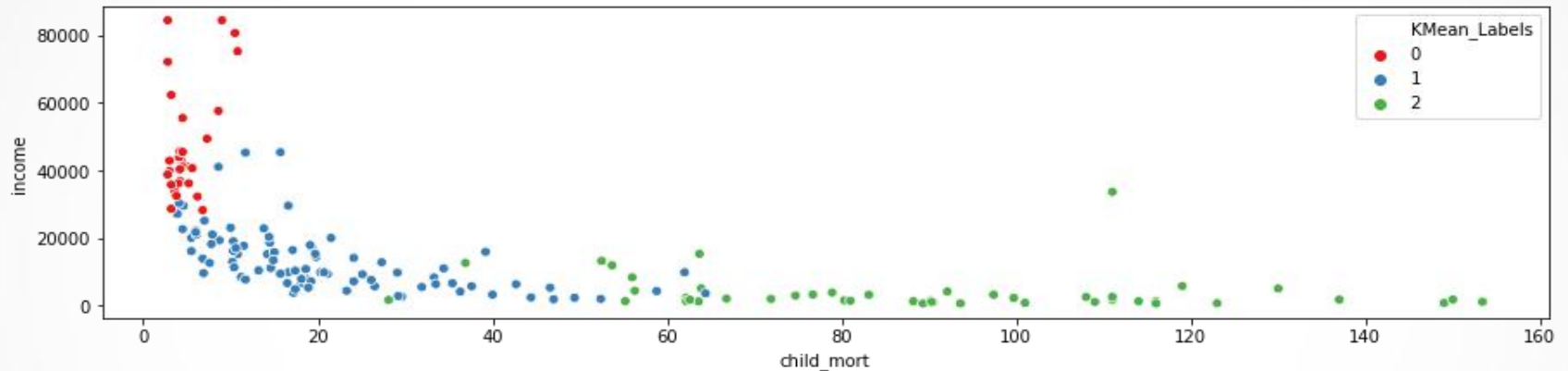
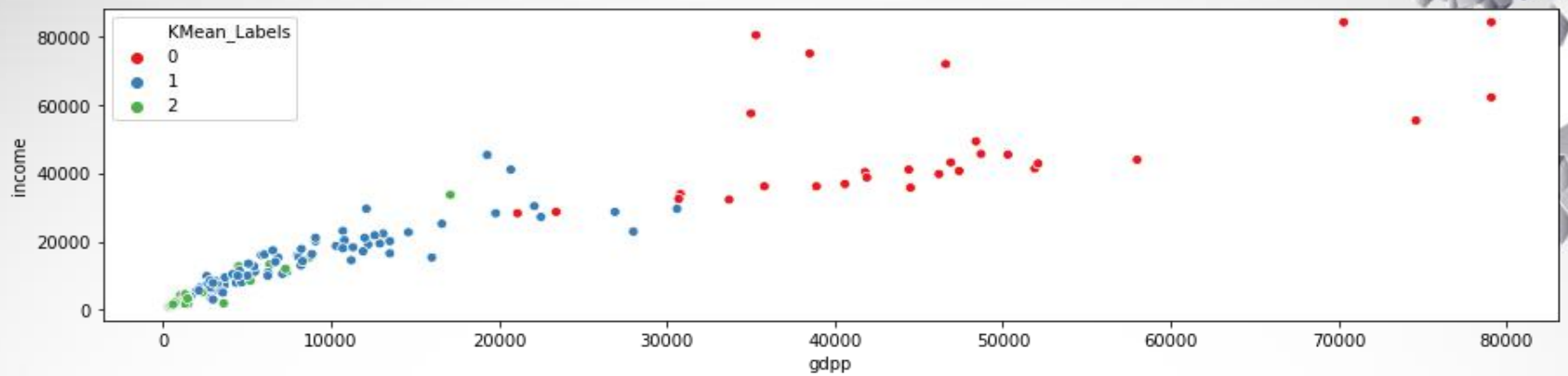
## Silhouette Curve



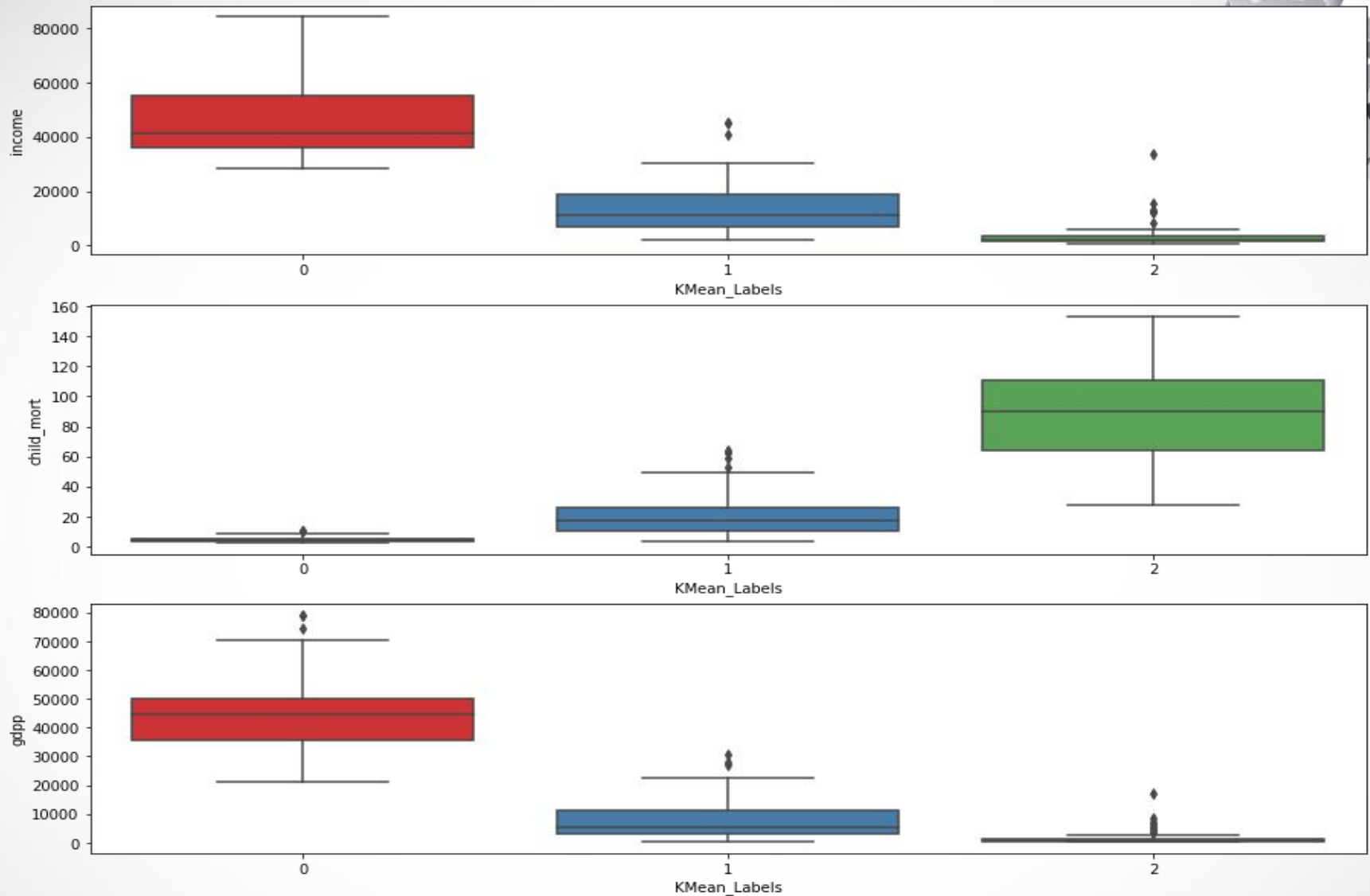
- ❖ As we can see at 2 we have maximum value of cost function but its not optimal to use as the 3 cluster looks significant in both Elbow curve and Silhouette curve.



- ❖ After plotting the 3 clusters on basis of Child mortality, income & GDP, it is found that the countries in Cluster no 2 are lowest among all as per K-mean clustering.



❖ Visualising the Clusters on the basis of **income, child mortality & gdp** by K-mean cluster labels



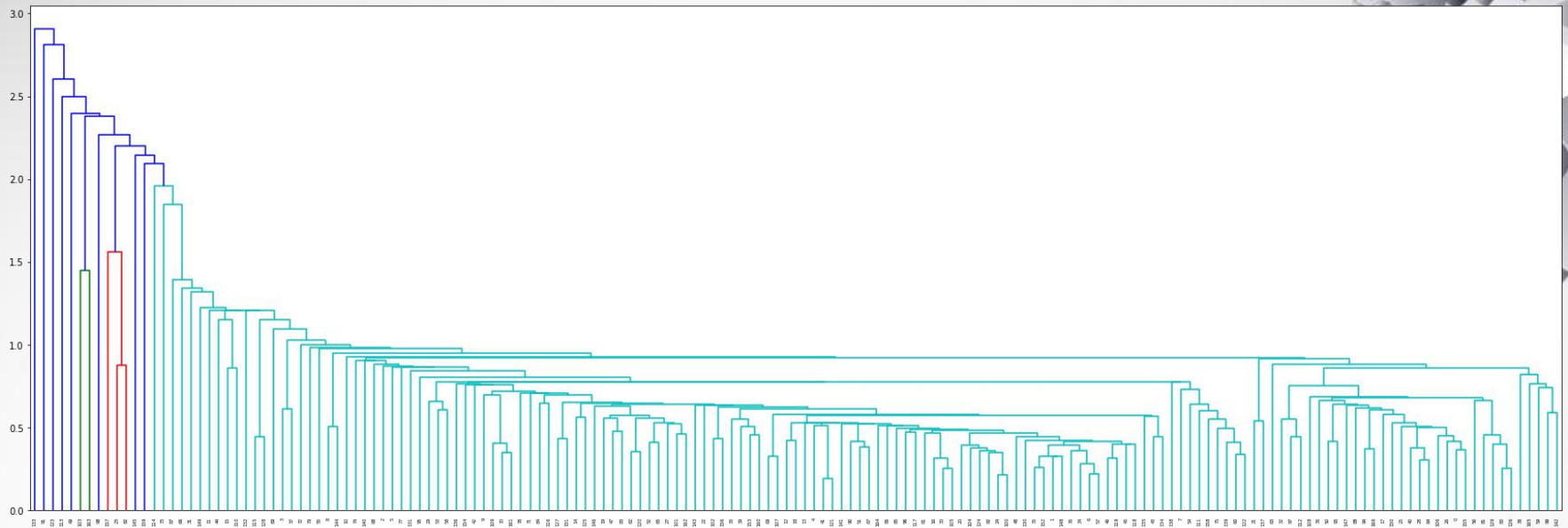
❖ As we can visualise, Cluster 2 countries has **Lowset income, highest child mortality & lowest gdp**



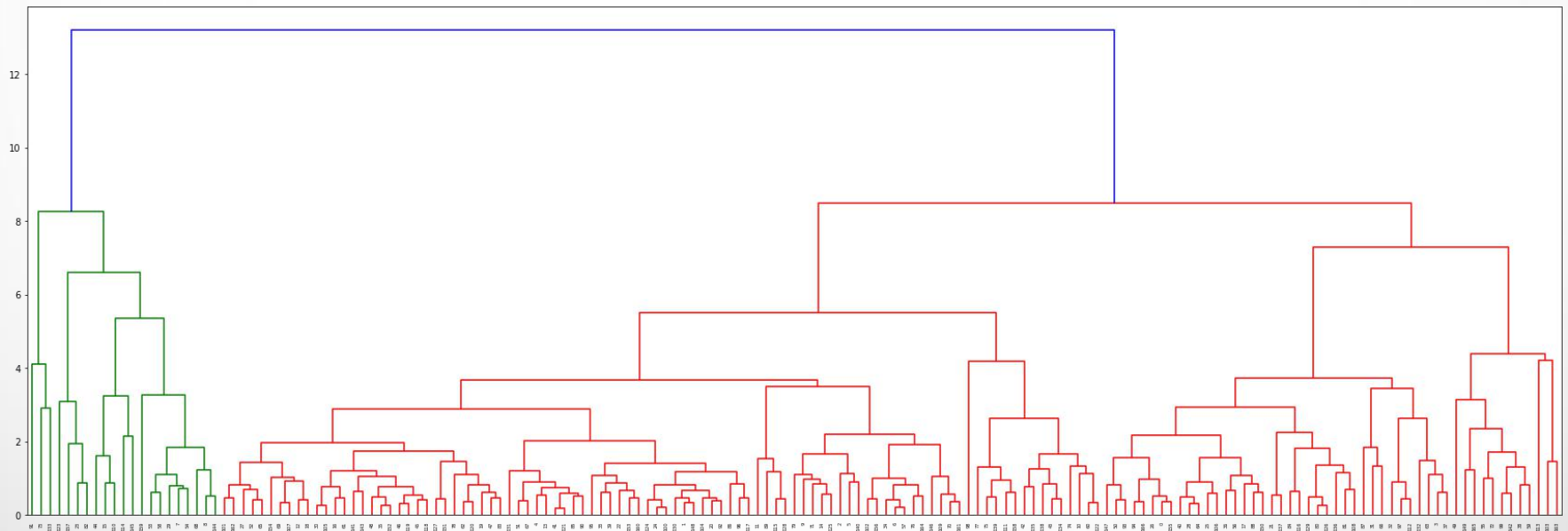
- ✓ **Suggested 10 Countries by K-Mean clustering that are in the direst need of aid, choosed on the basis of [gdpp, child\_mort and income]**

```
df_cluster[df_cluster['KMean_Labels']==2].sort_values(['gdpp','child_mort','income'],ascending=[True,False,True]).head(10)
```

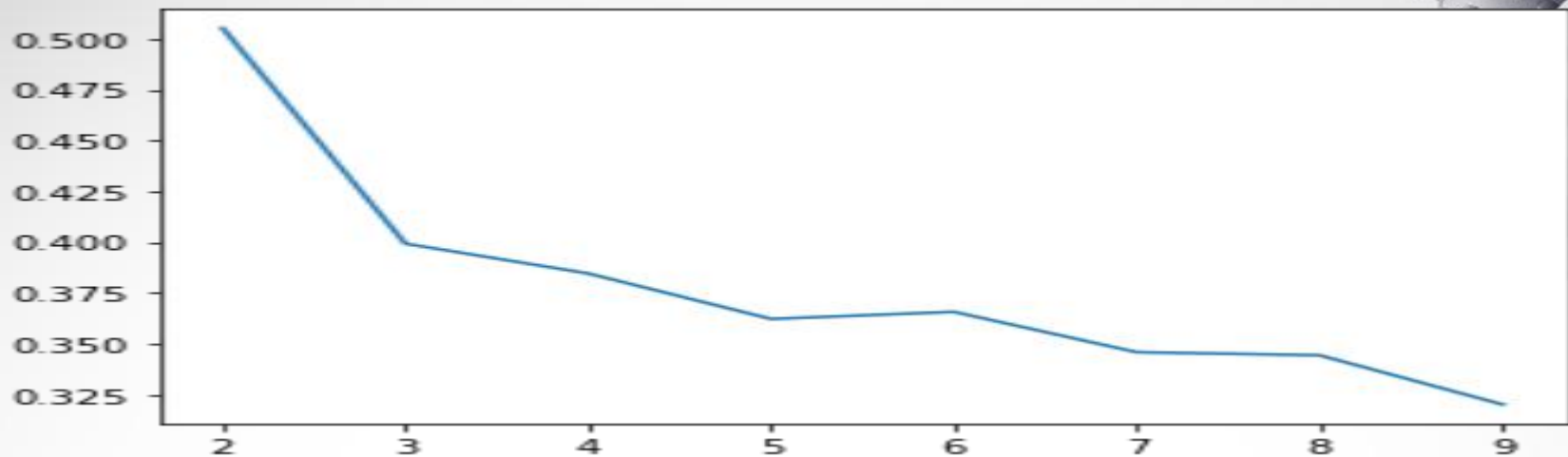
	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	KMean_Labels
26	Burundi	93.6	22.243716	26.796000	104.90964	764.00	12.30	57.7	6.2600	331.62	2
88	Liberia	89.3	62.457000	38.586000	302.80200	742.24	5.47	60.8	5.0200	331.62	2
37	Congo, Dem. Rep.	116.0	137.274000	26.419400	165.66400	742.24	20.80	57.5	6.5400	334.00	2
112	Niger	123.0	77.256000	17.956800	170.86800	814.00	2.55	58.8	6.5636	348.00	2
132	Sierra Leone	153.4	67.032000	52.269000	137.65500	1220.00	17.20	55.0	5.2000	399.00	2
93	Madagascar	62.2	103.250000	17.009362	177.59000	1390.00	8.79	60.8	4.6000	413.00	2
106	Mozambique	101.0	131.985000	21.829900	193.57800	918.00	7.64	54.5	5.5600	419.00	2
31	Central African Republic	149.0	52.628000	17.750800	118.19000	888.00	2.01	47.5	5.2100	446.00	2
94	Malawi	90.5	104.652000	30.248100	160.19100	1030.00	12.10	53.1	5.3100	459.00	2
50	Eritrea	55.2	23.087800	17.009362	112.30600	1420.00	11.60	61.7	4.6100	482.00	2



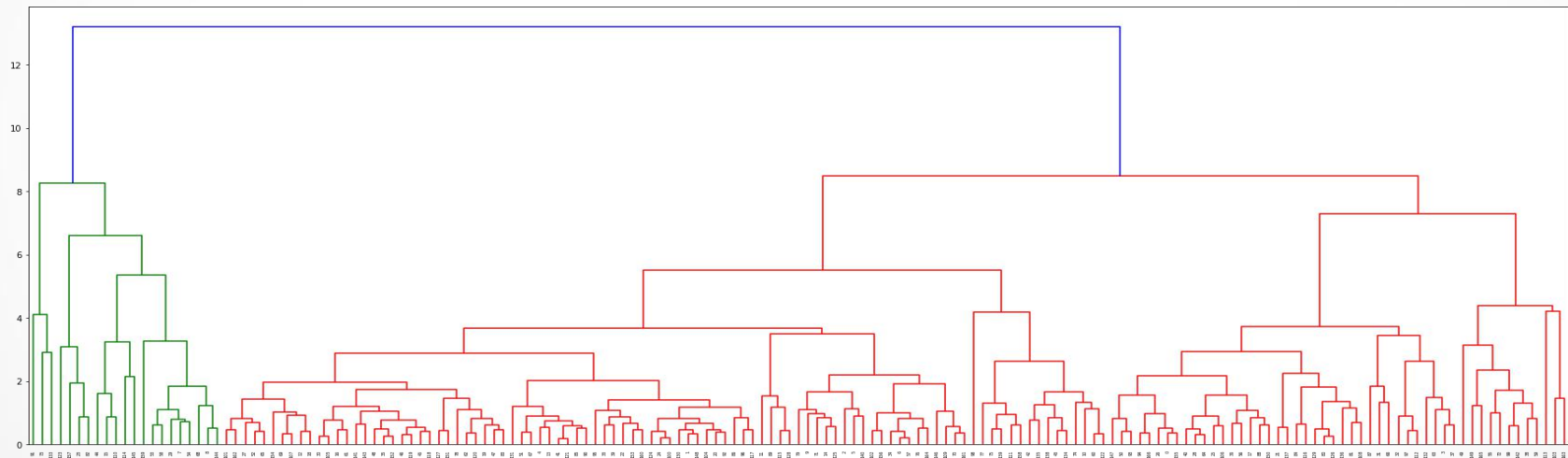
Hierarchical Clustering using single linkage



Hierarchical Clustering using complete linkage

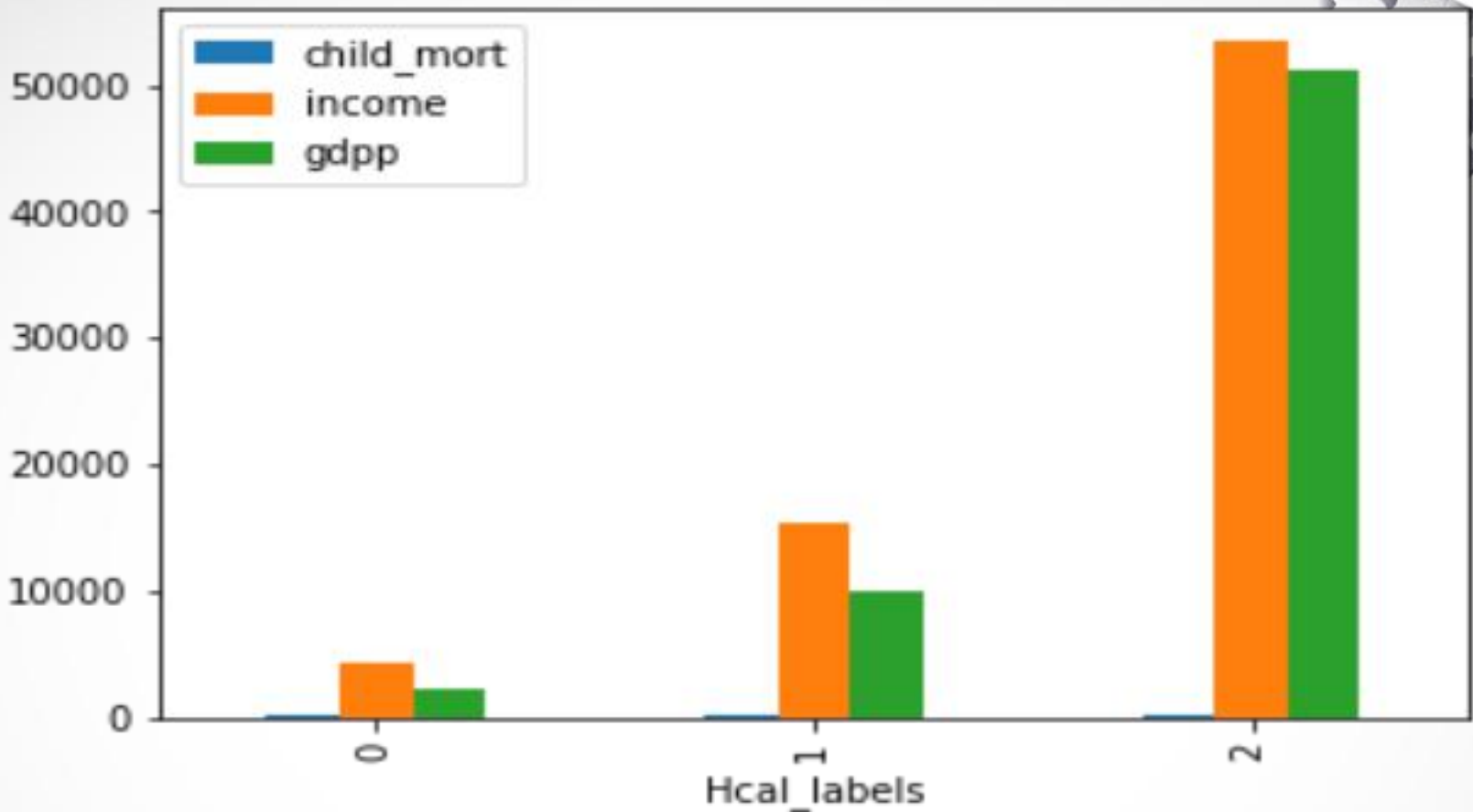


**Silhouette Curve**

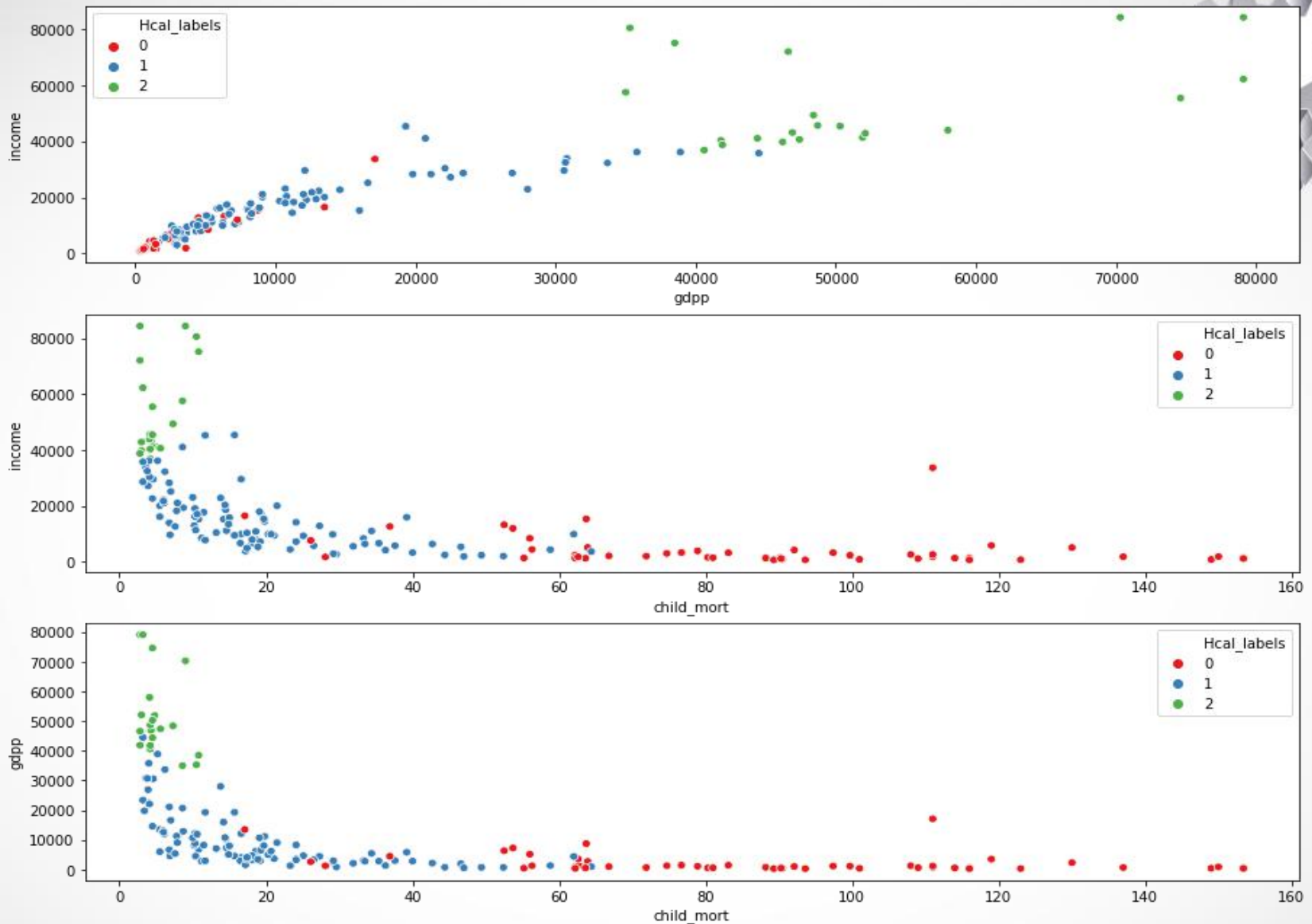


**Hierarchical Clustering using complete linkage**

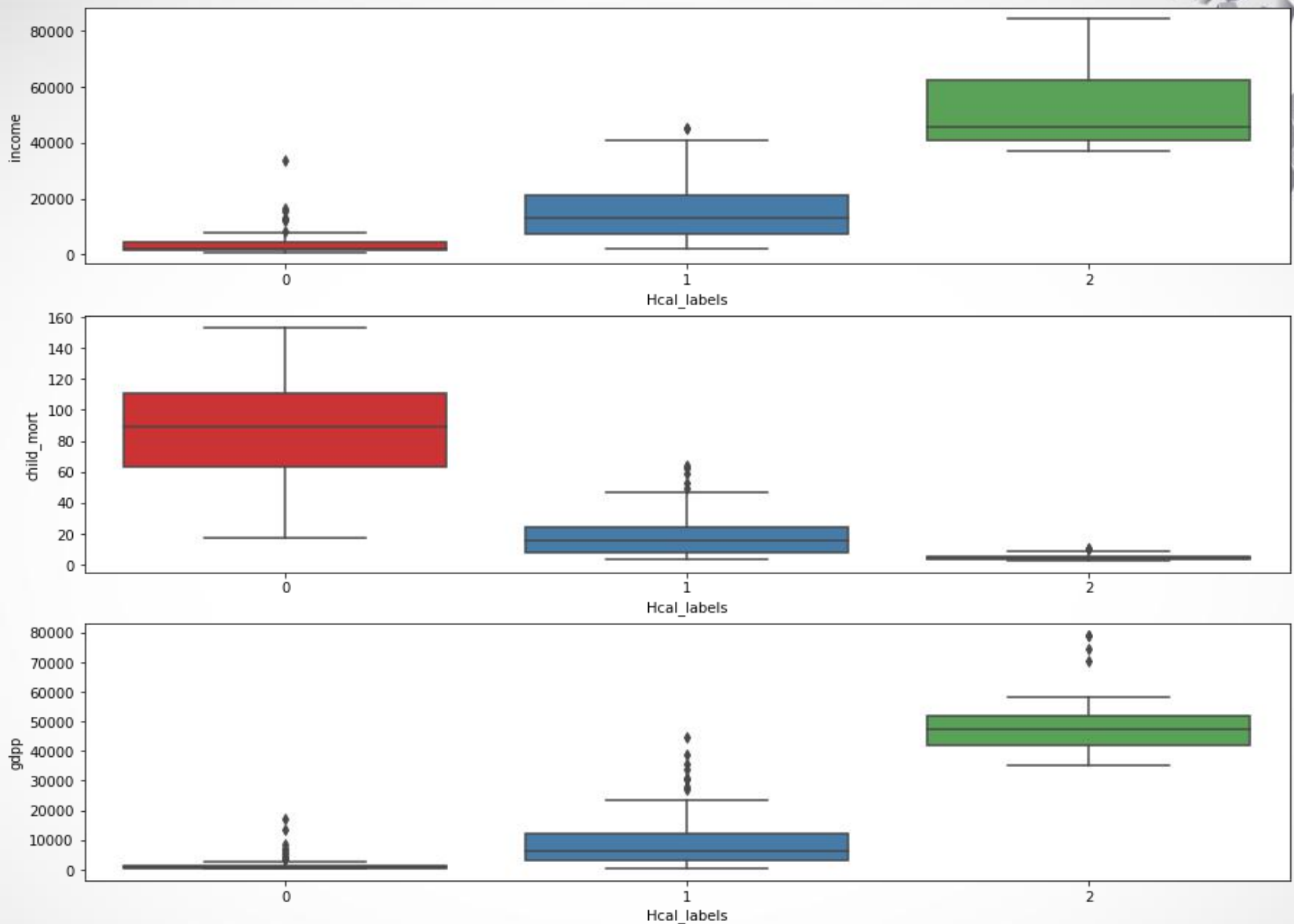
By the Silhouette Curve we found 3 clusters are optimum and by Hierarchical Clustering using complete linkage, x axis consists of countries and y axis consists of Euclidean distance between the clusters, to get the largest distance we count the number of lines on the diagram and determine optimal numbers of clusters, so we have choosed 3 clusters as optimum



- ❖ After plotting the 3 clusters on basis of Child mortality, income & GDP, it is found that the countries in Cluster no 0 are lowest among all as per Hierarchical clustering.



❖ Visualising the Clusters on the basis of **income, child mortality & gdp** by **Hierarchical cluster labels**



❖ As we can visualise, Cluster 0 countries has **Lowset income, highest child mortality & lowest gdp by Hierarchical clustering**



- ✓ **Suggested 10 Countries by Hierarchical clustering that are in the direst need of aid, choosed on the basis of [gdpp, child\_mort and income]**

```
df_cluster[df_cluster['Hcal_labels']==0].sort_values(['gdpp','child_mort','income'],ascending=[True,False,True]).head(10)
```

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	KMean_Labels	Hcal_labels
26	Burundi	93.6	22.243716	26.796000	104.90964	764.00	12.30	57.7	6.2600	331.62	2	0
88	Liberia	89.3	62.457000	38.586000	302.80200	742.24	5.47	60.8	5.0200	331.62	2	0
37	Congo, Dem. Rep.	116.0	137.274000	26.419400	165.66400	742.24	20.80	57.5	6.5400	334.00	2	0
112	Niger	123.0	77.256000	17.956800	170.86800	814.00	2.55	58.8	6.5636	348.00	2	0
132	Sierra Leone	153.4	67.032000	52.269000	137.65500	1220.00	17.20	55.0	5.2000	399.00	2	0
93	Madagascar	62.2	103.250000	17.009362	177.59000	1390.00	8.79	60.8	4.6000	413.00	2	0
106	Mozambique	101.0	131.985000	21.829900	193.57800	918.00	7.64	54.5	5.5600	419.00	2	0
31	Central African Republic	149.0	52.628000	17.750800	118.19000	888.00	2.01	47.5	5.2100	446.00	2	0
94	Malawi	90.5	104.652000	30.248100	160.19100	1030.00	12.10	53.1	5.3100	459.00	2	0
50	Eritrea	55.2	23.087800	17.009362	112.30600	1420.00	11.60	61.7	4.6100	482.00	2	0



country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	KMean_Labels	Hcal_labels
Burundi	93.6	22.243716	26.7960	104.90964	764.00	12.30	57.7	6.2600	331.62	2	0
Liberia	89.3	62.457000	38.5860	302.80200	742.24	5.47	60.8	5.0200	331.62	2	0
Congo, Dem. Rep.	116.0	137.274000	26.4194	165.66400	742.24	20.80	57.5	6.5400	334.00	2	0
Niger	123.0	77.256000	17.9568	170.86800	814.00	2.55	58.8	6.5636	348.00	2	0
Sierra Leone	153.4	67.032000	52.2690	137.65500	1220.00	17.20	55.0	5.2000	399.00	2	0

## **Suggestion:**

- ✓ The listed countries are in the direst need of aid, sorted by these three variables - [gdpp, child\_mort and income] as found by both the algorithm(K-mean & Hierarchical Clustering)

1. **Burundi**
2. **Liberia**
3. **Congo, Dem. Rep**
4. **Niger**
5. **Sierra Leone**



Thank You

