HELP NGO CASE STUDY

SUBMISSION

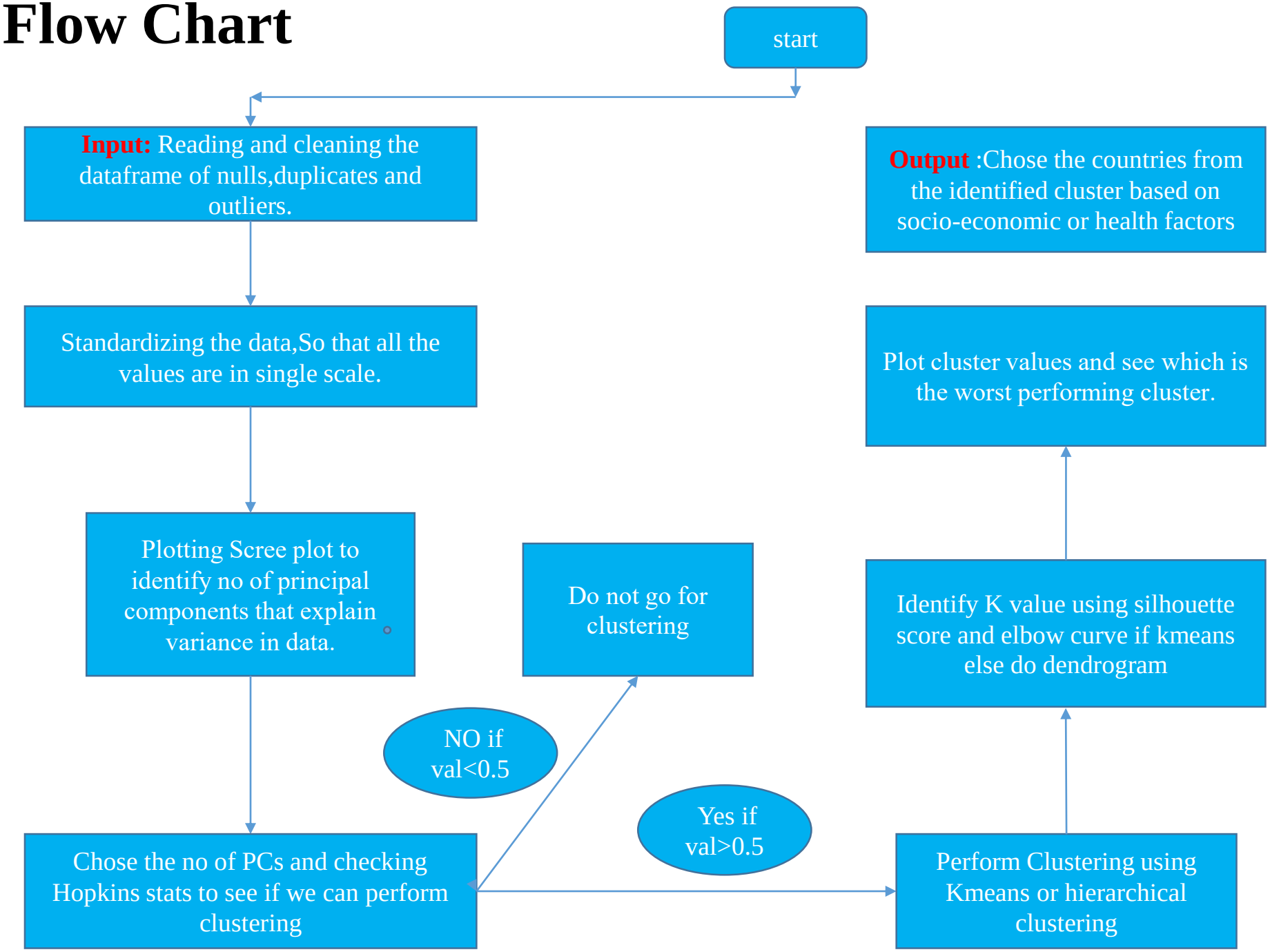
By:

Ravi Biradar

Abstract

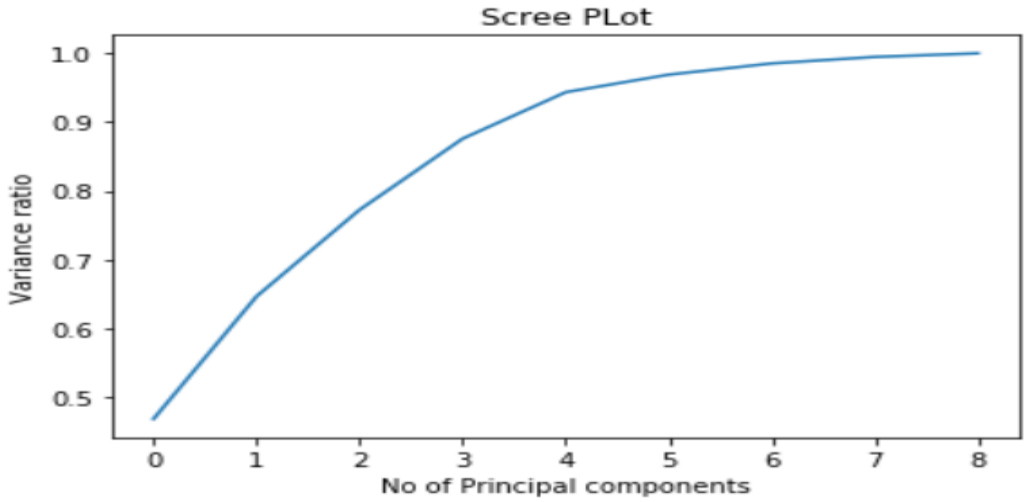
HELP have been able to raise around $ 10 million. Now the CEO of NGO wants to decide how to use this money strategically and effectively. Our job is to categorize the countries using some socio-economic and health factors that determine the overall development of the country and suggest the countries which the CEO needs to focus on the most.

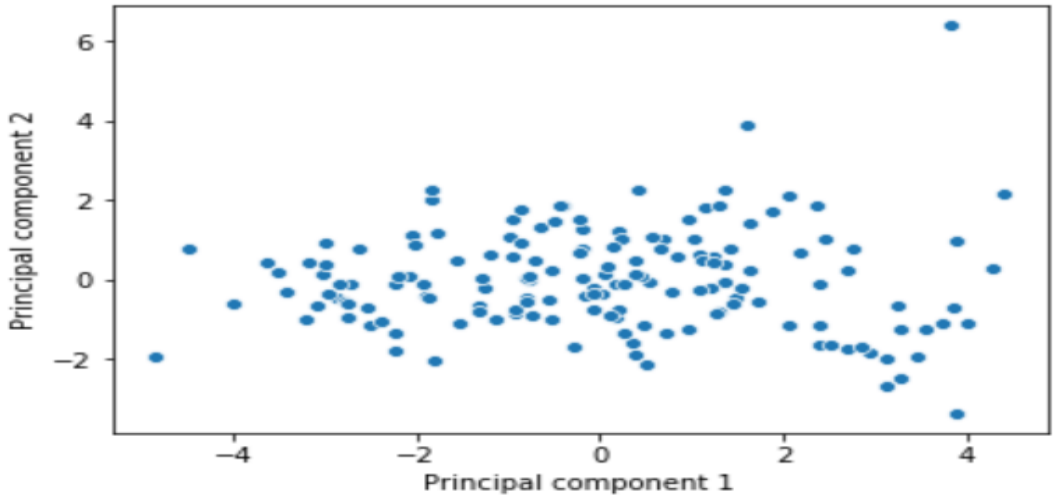
Flow Chart


start
Input: Reading and cleaning the dataframe of nulls,duplicates and outliers.
Standardizing the data,So that all the values are in single scale.
Plotting Scree plot to identify no of principal components that explain variance in data.
Chose the no of PCs and checking Hopkins stats to see if we can perform clustering
NO if val<0.5
Do not go for clustering
Yes if val>0.5
Perform Clustering using Kmeans or hierarchical clustering
Identify K value using silhouette score and elbow curve if kmeans else do dendrogram
Plot cluster values and see which is the worst performing cluster.
Output :Chose the countries from the identified cluster based on socio-economic or health factors

PCA(Principal Component Analysis)

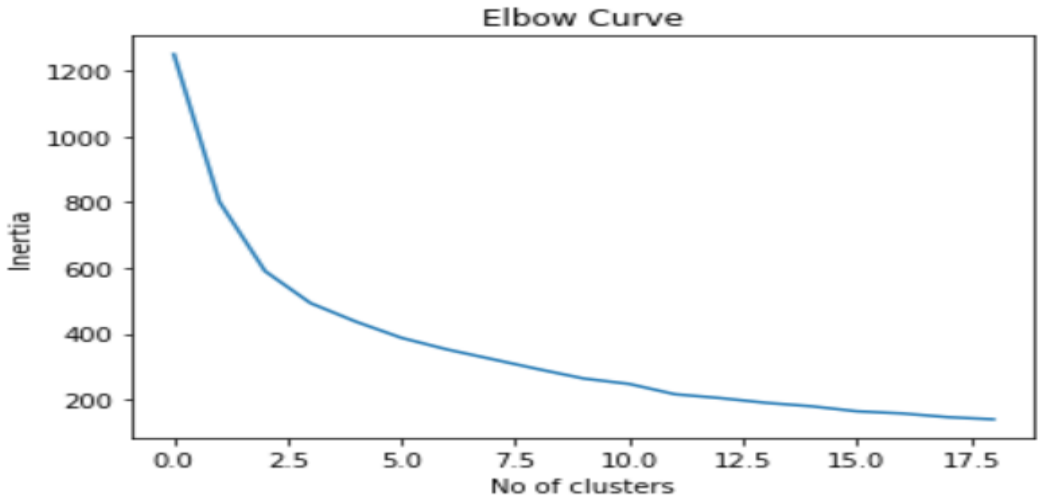
- After Standardizing the data we perform PCA on standardized dataset and determine no of PC we need to choose.
- From Scree plot we can see that just 4 PCs are able to explain more than 90% of the data.
- These features are reduced to just 4 PCs
- From the second plot we can see how the PC 1 and PC 2 have transformed the data.



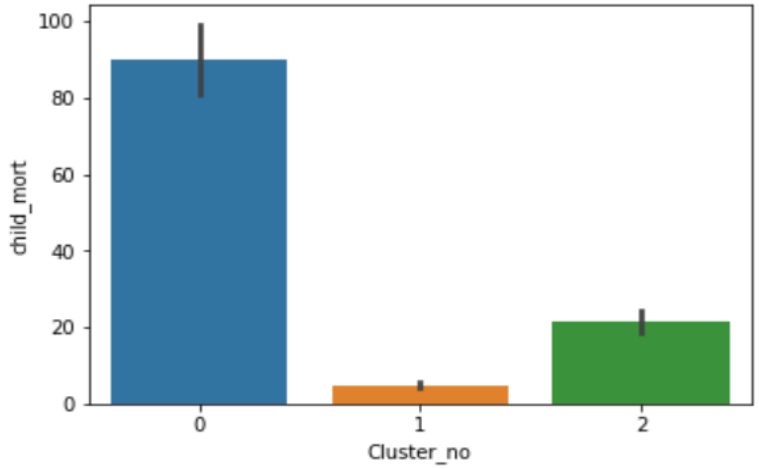


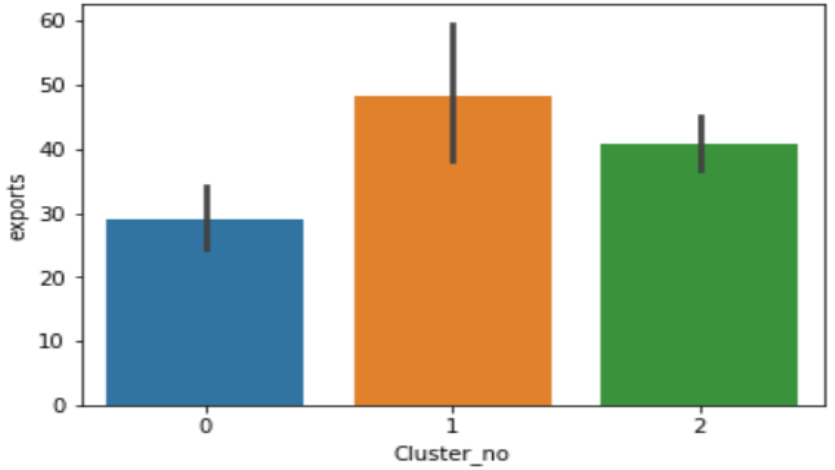
Kmeans Clustering (Choosing K)

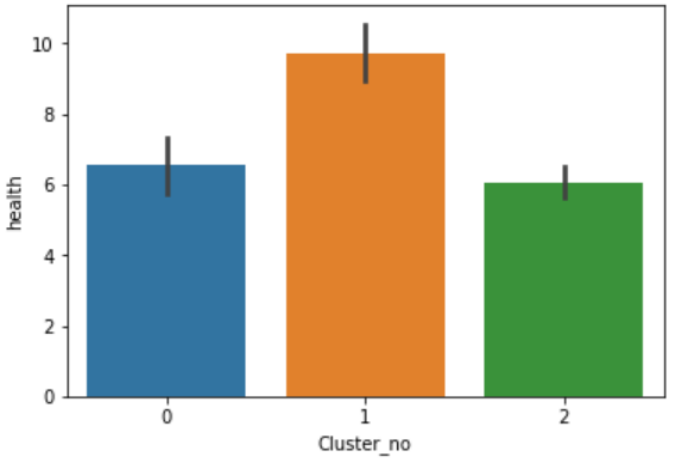
- In order to do determine the value of K for K means clustering, we plot elbow curve and silhouette plots.
- We can see from elbow curve that after the 3 clusters the value of inertia is almost constant.
- In the silhoutte plot we can see the max score is when the no of clusters is 3.
- This we choose k=3.

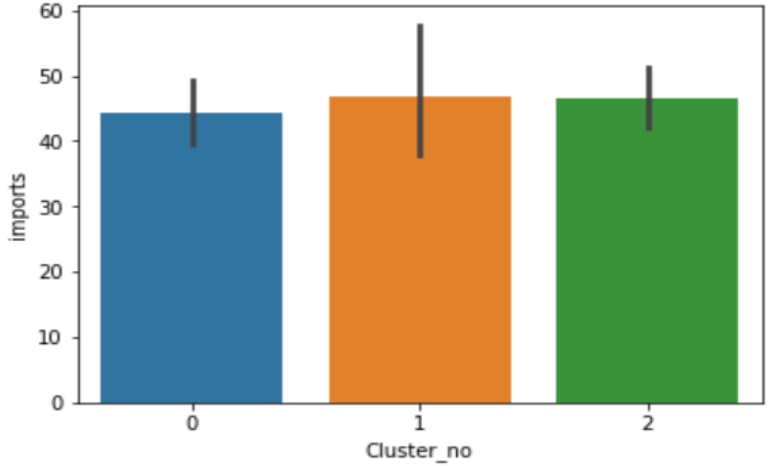


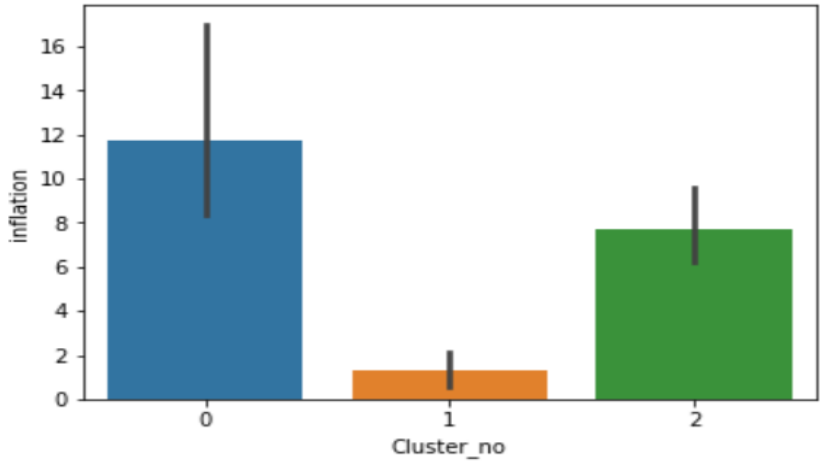
Kmeans Analysis

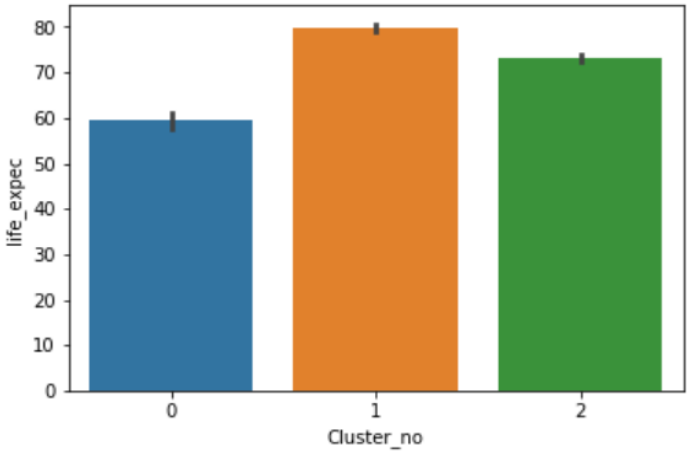










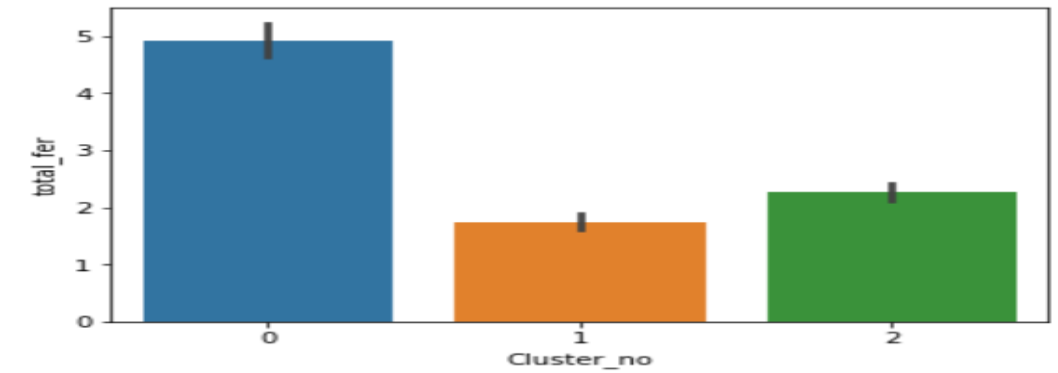


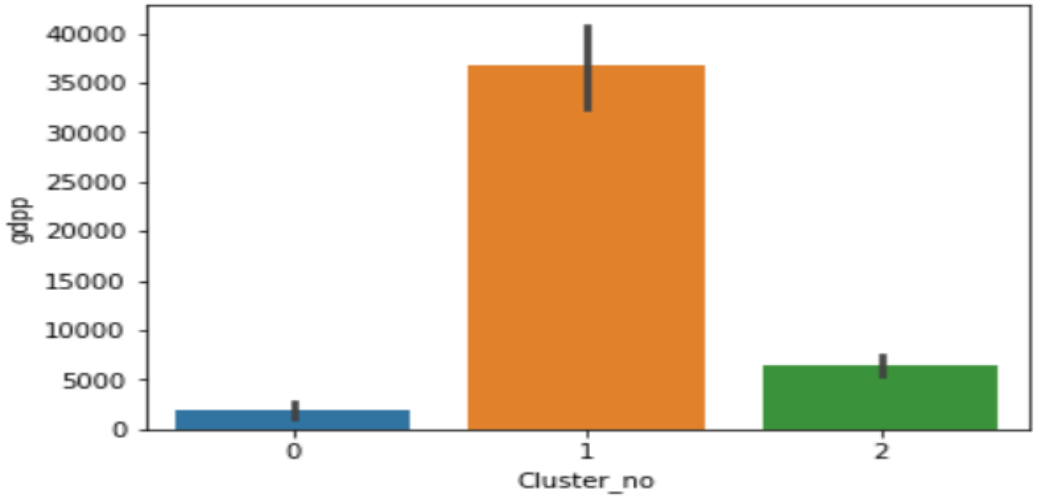
Kmeans Analysis

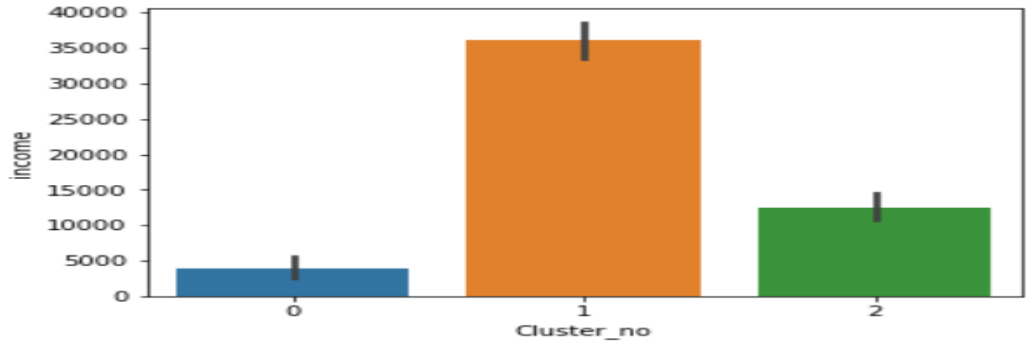
- From the previous slide and current we see that Cluster 0 has Higher child mort rate ,lower exports, lower imports, higher inflation and low health spending.

- Life expectancy is within 60,Total fertility rate is high and GDP is within 5K and even income per person is within 5k.

- This we can conclude that cluster 0 is worst performing and countries listed within it are in need of help.







Kmeans Analysis

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
Burkina Faso	116.0	19.2	6.74	29.6	1430	6.81	57.9	5.87	575
Central African Republic	149.0	11.8	3.98	26.5	888	2.01	47.5	5.21	446
Chad	150.0	36.8	4.53	43.5	1930	6.39	56.5	6.59	897
Congo, Dem. Rep.	116.0	41.1	7.91	49.6	609	20.80	57.5	6.54	334
Guinea-Bissau	114.0	14.9	8.50	35.2	1390	2.97	55.6	5.05	547
Haiti	208.0	15.3	6.91	64.7	1500	5.45	32.1	3.33	662
Mali	137.0	22.8	4.98	35.1	1870	4.37	59.5	6.55	708
Niger	123.0	22.2	5.16	49.1	814	2.55	58.8	7.49	348
Sierra Leone	160.0	16.8	13.10	34.5	1220	17.20	55.0	5.20	399

- We consider 2 factors to choose the countries with direst need of help within cluster 0.We look at the child mort and GDP of the countries in cluster 0.
- Cut of is if child mort is above 75% of cluster 0 and GDP is below 50% of the cluster GDP we chose countries as the once need of help.
- This above listed table gives the countries in direst need of help.

Conclusion

This we can conclude that we have found the worst performing cluster and the countries within it which are in need of help based on their socio-economic and health factors.