
Clustering and PCA
ASSIGNMENT

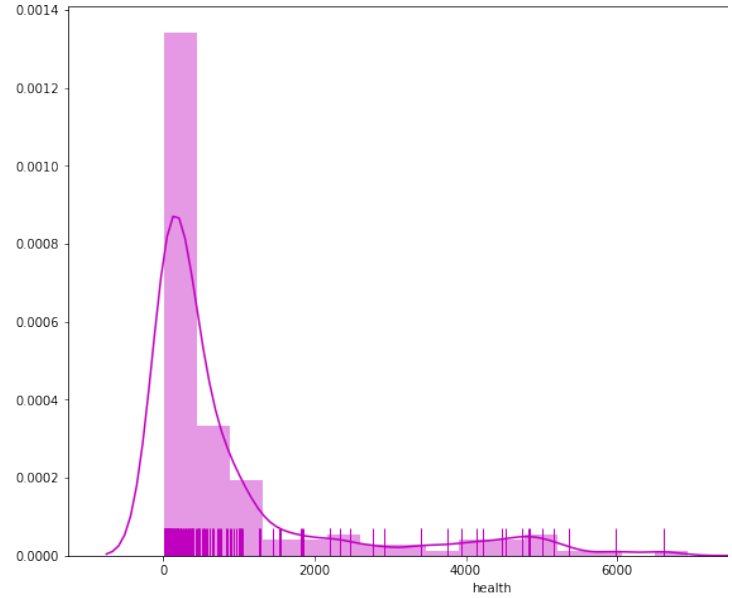
Problem Statement

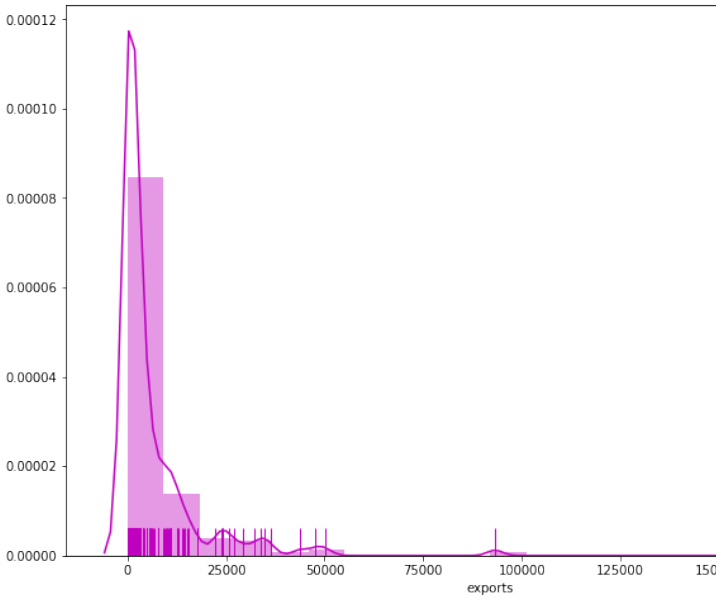
HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.

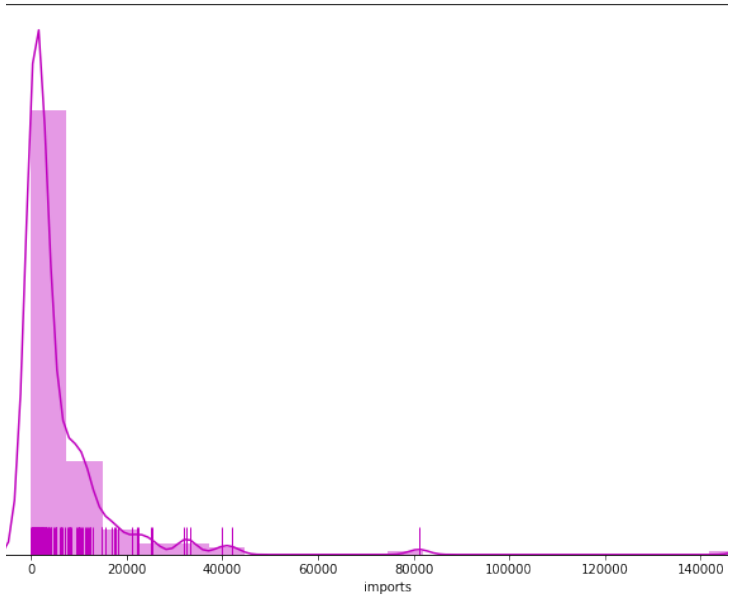
After the recent funding programmes, they have been able to raise around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. The significant issues that come while making this decision are mostly related to choosing the countries that are in the direst need of aid.

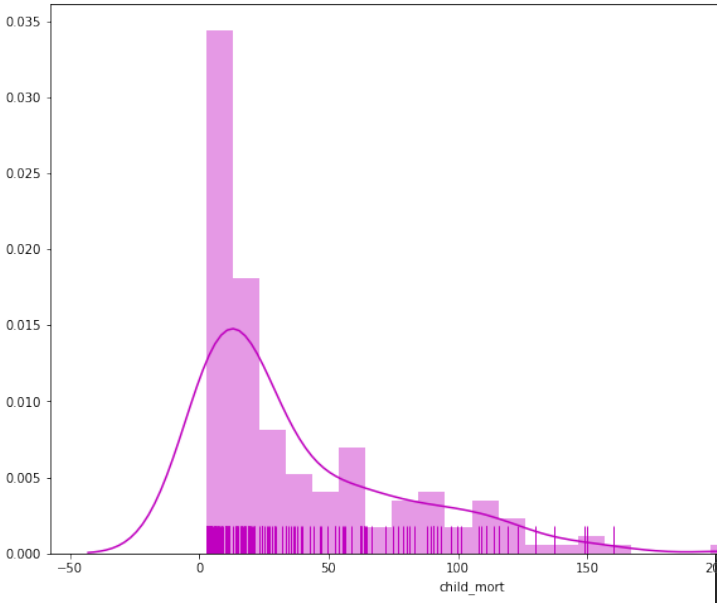

Analysis
CLUSTERING AND PCA

Univariate Analysis

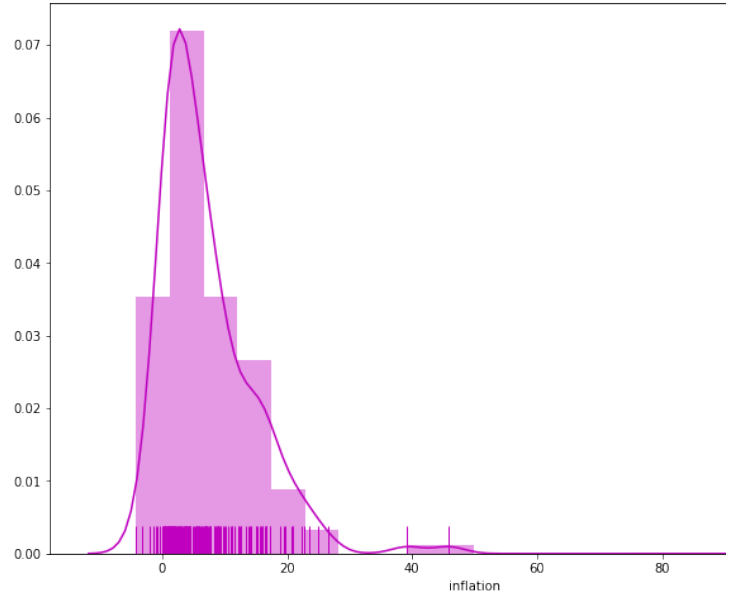


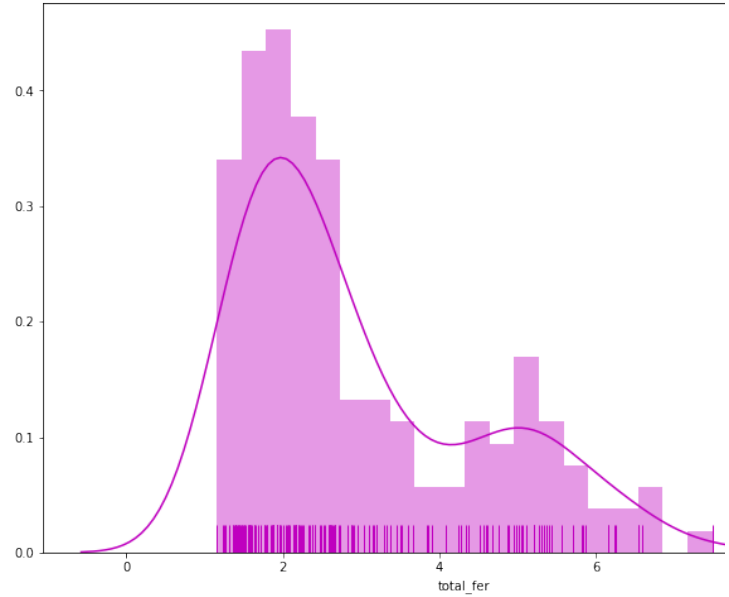


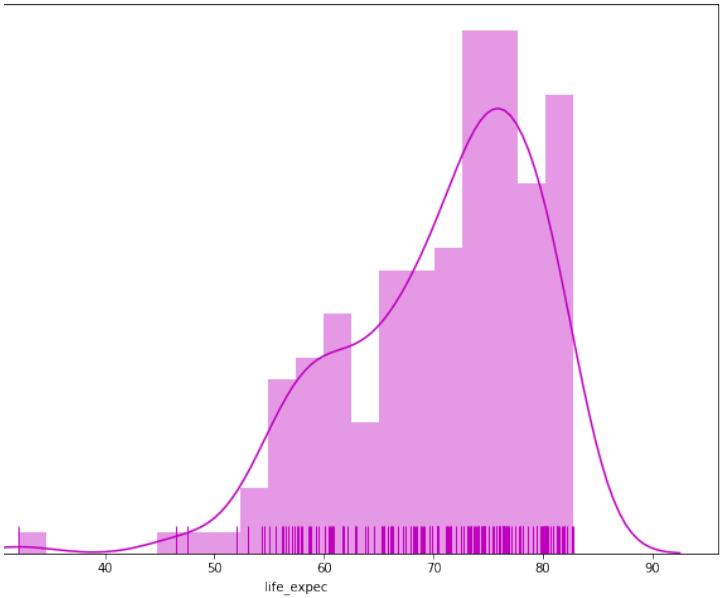


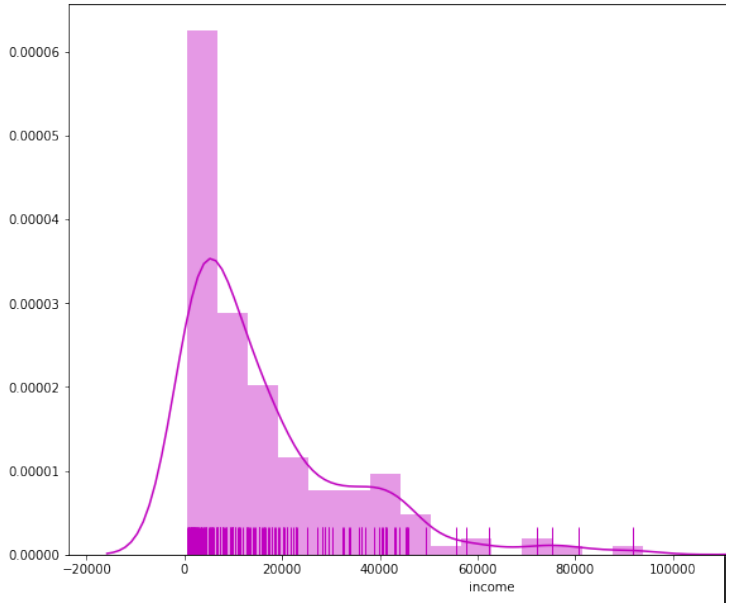


Univariate Analysis

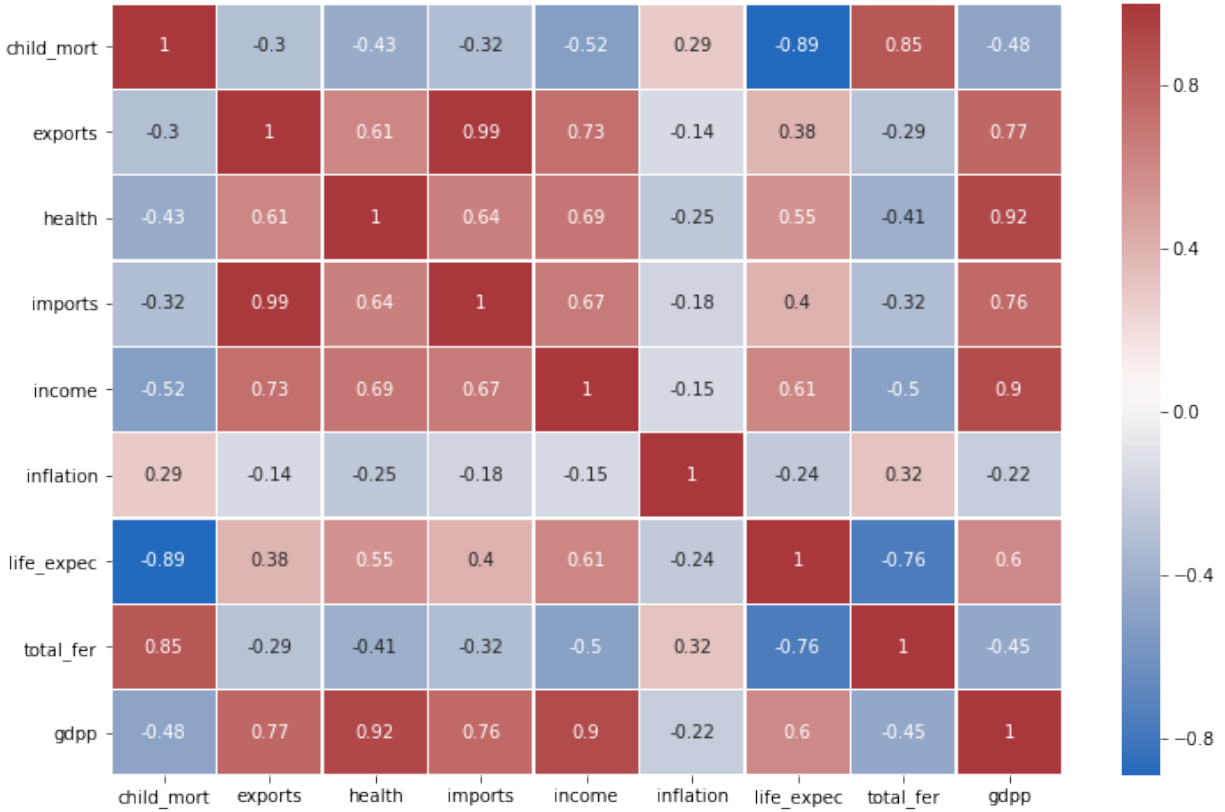


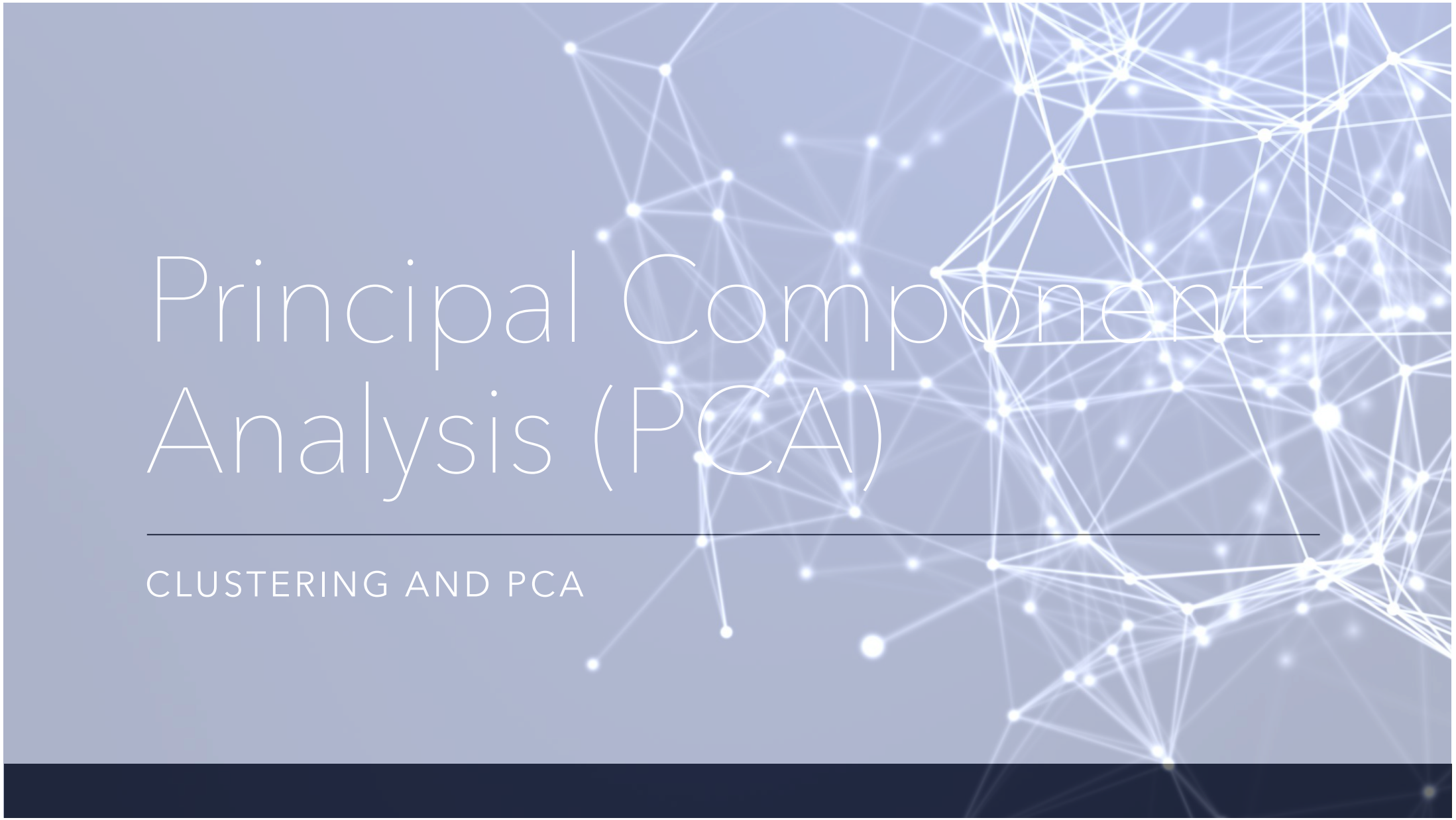






Bivariate Analysis



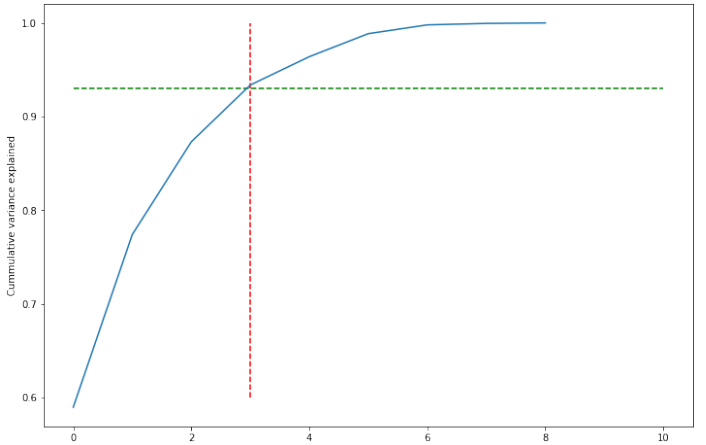


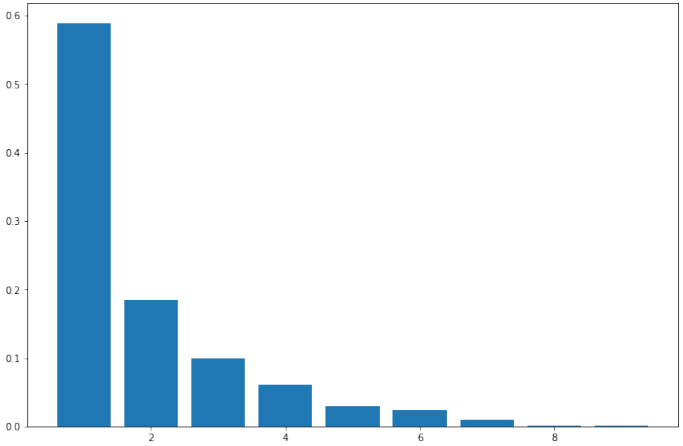
Principal Component Analysis (PCA)

CLUSTERING AND PCA

Principal Component Analysis

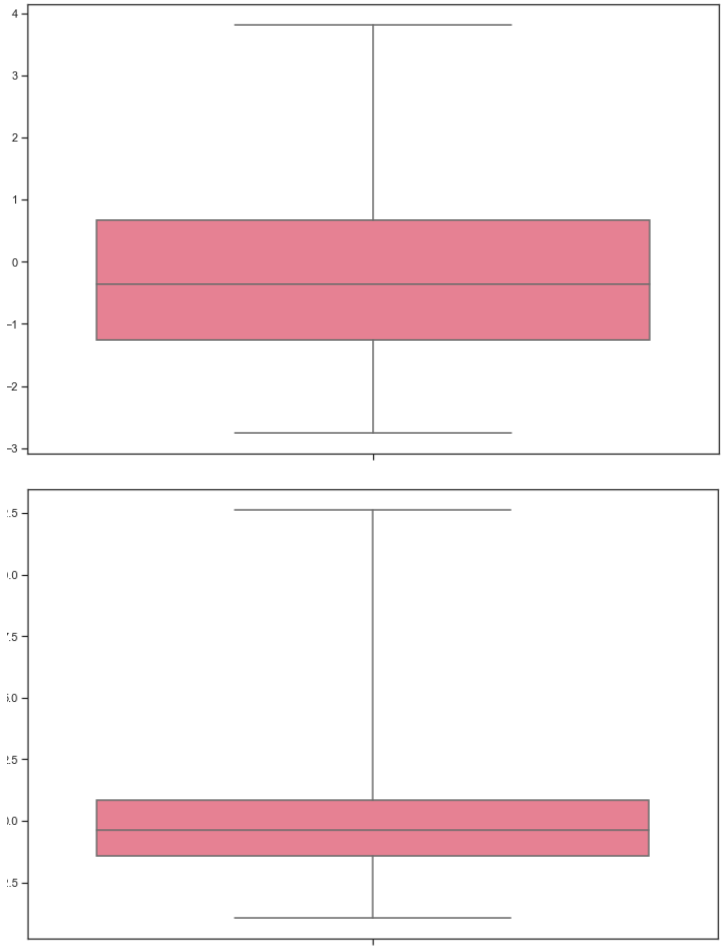
- Scaling : maximum Value is 9.84
- PCA Explained Variance
- Scree Plot
  - dimension reduction to 3
  - cumulative Value of 0.93





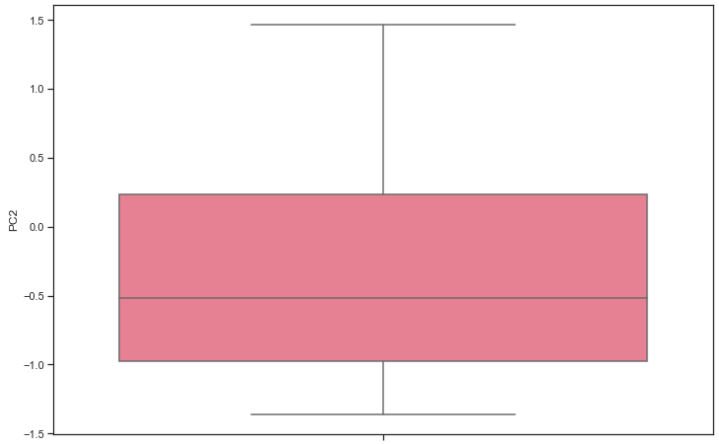
Outlier Treatment - PC1

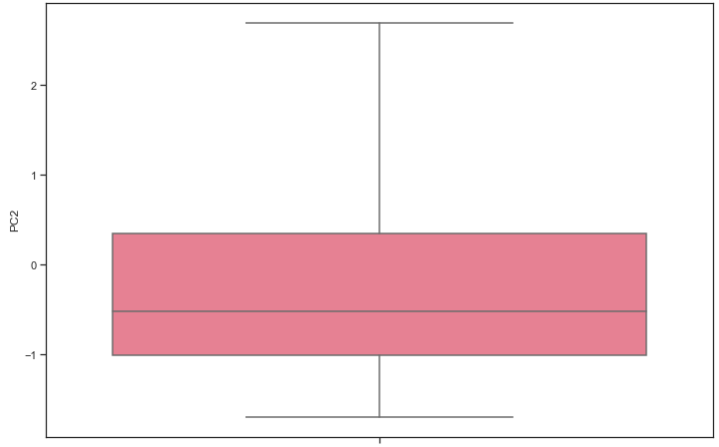
- 0.05 Percentile is : -2.761010104920181 6
- 0.90 Percentile is : 6.678328963972401
- 0.95 Percentile is : 3.917318859052219
- PC1 - Total Number of Records 167, after removing outliner 149, Percentage dropped 10.78%



Outlier Treatment – PC2

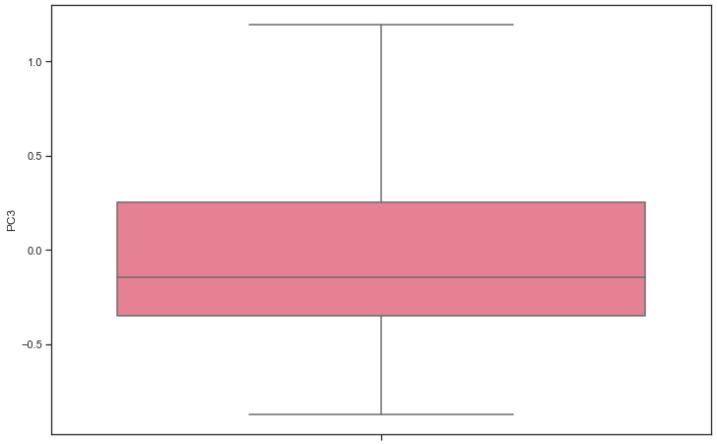
- 0.05 Percentile is : -1.3684306776672928
- 0.90 Percentile is : 2.8524088307670876
- 0.95 Percentile is : 1.483978153099795
- PC2 - Total Number of Records 149, after removing outliner 133, Percentage dropped 10.74%

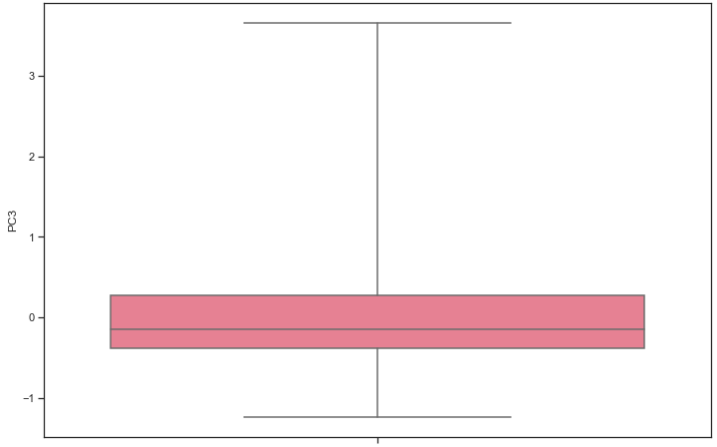




Outlier Treatment – PC3

- 0.05 Percentile is : -0.8914343685972016
- 0.90 Percentile is : 2.130859684882185
- 0.95 Percentile is : 1.2394253162849833
- PC3 - Total Number of Records 133, after removing outliner 119, Percentage dropped 10.53%





Hopkins

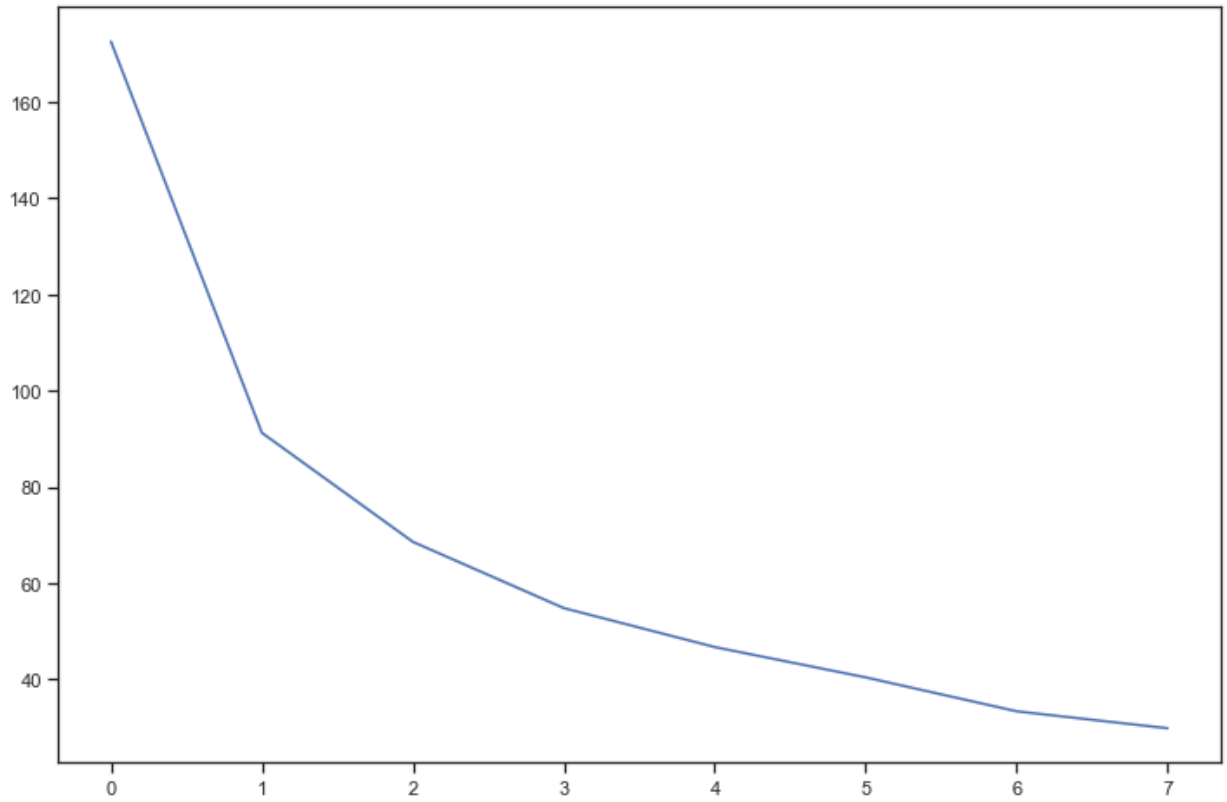
THE GIVEN DATA IS GOOD FOR PERFORMING CLUSTERING **0.78**

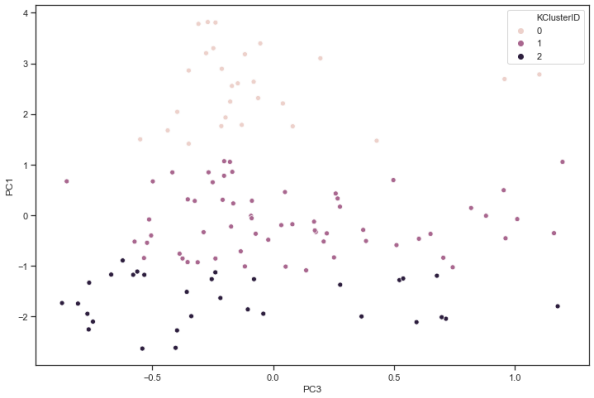
Silhouette Analysis

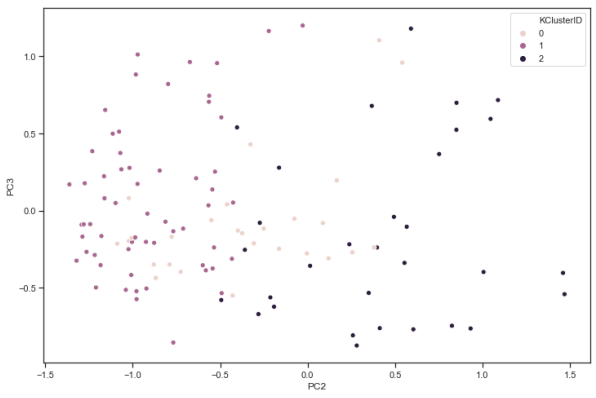
THE SILHOUETTE SCORE ANALYSIS SHOWS **0.49** FOR N_CLUSTERS=**2**

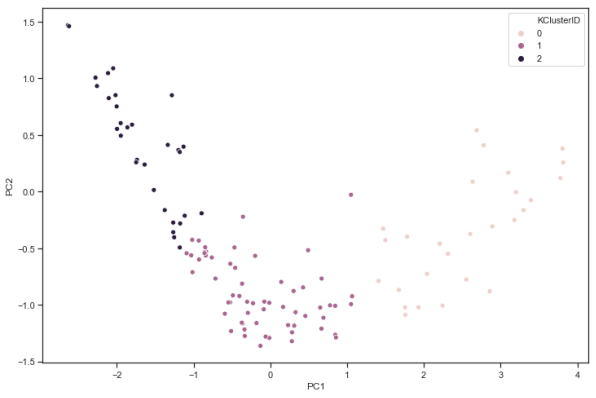
Elbow / SSD

SSD / Elbow curve shows optimal number of cluster is 3.



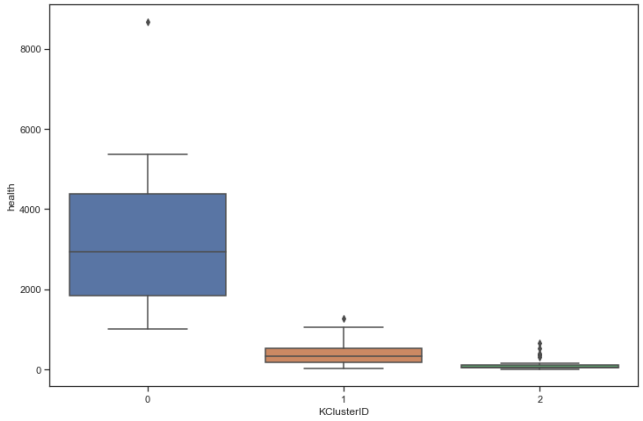


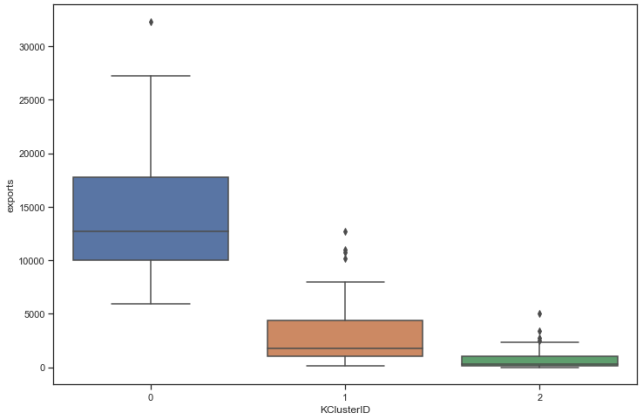




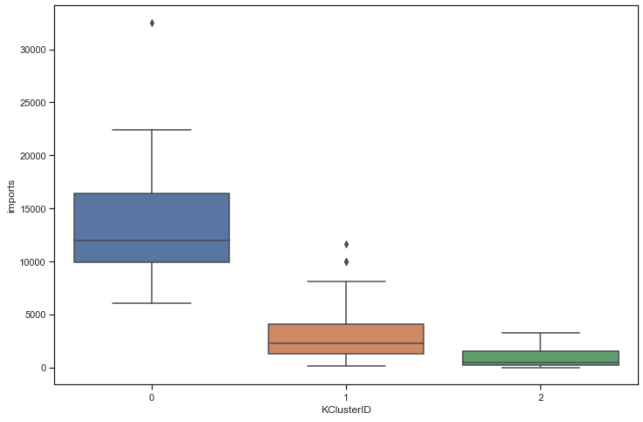
KMeans

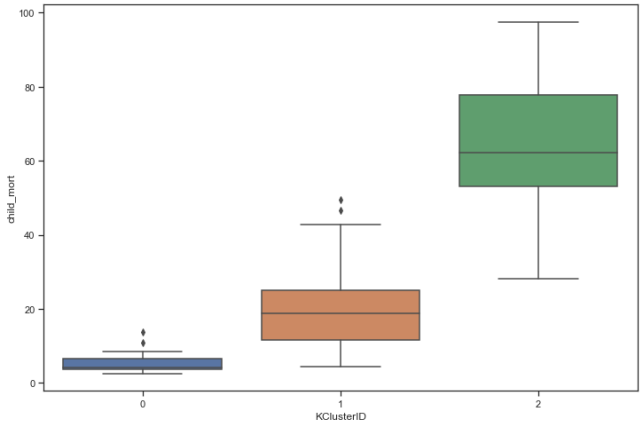
CLUSTER 3

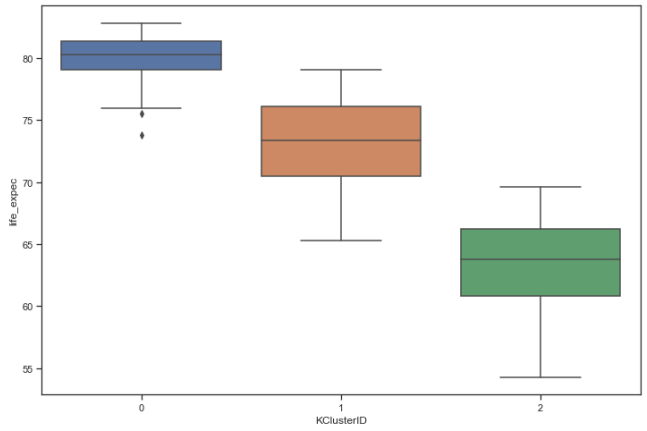


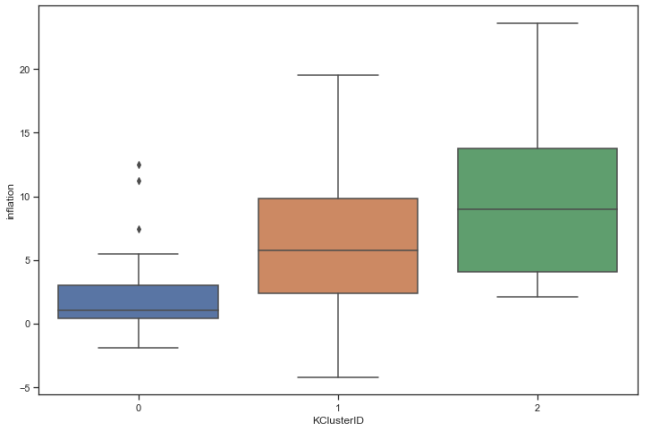


KMeans Boxplot

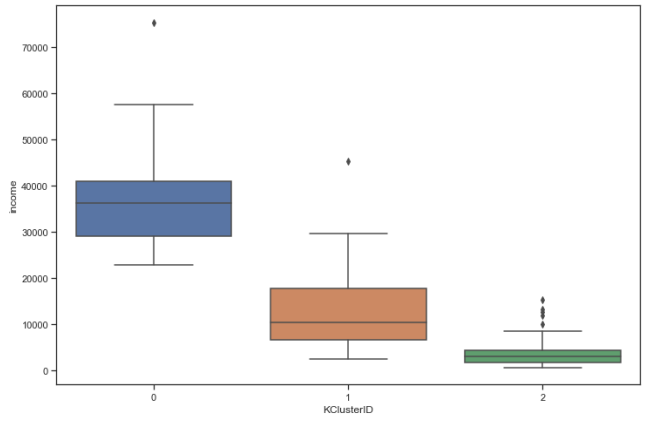


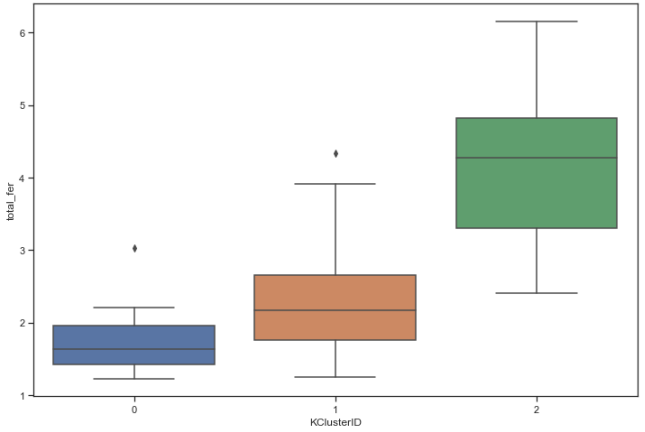


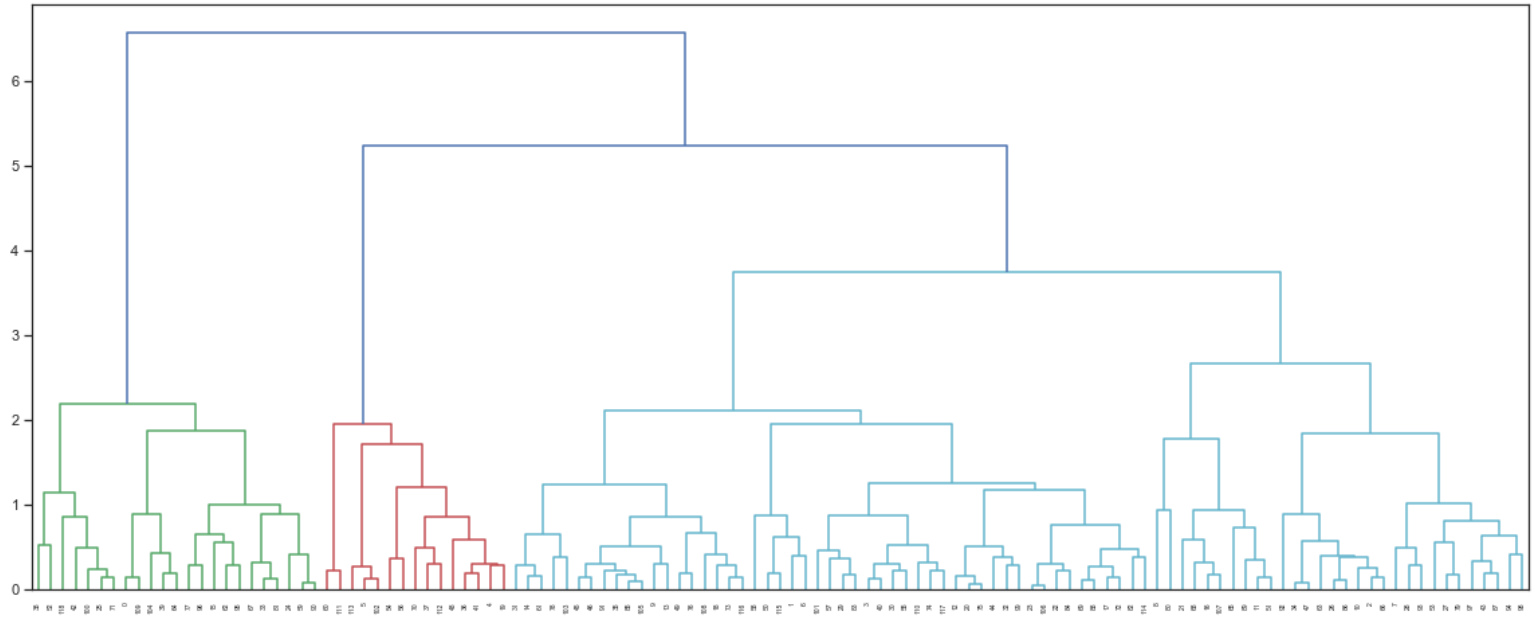




KMeans Boxplot

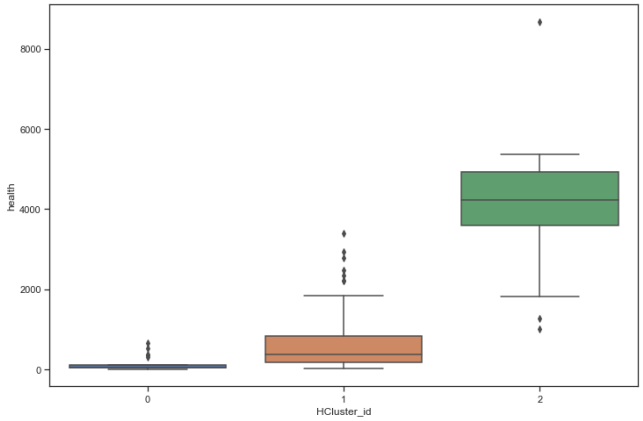


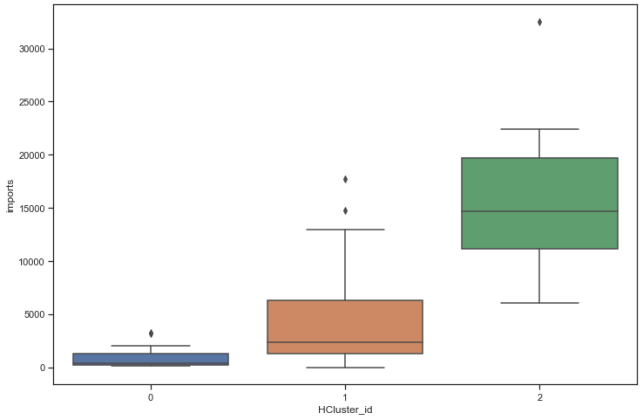


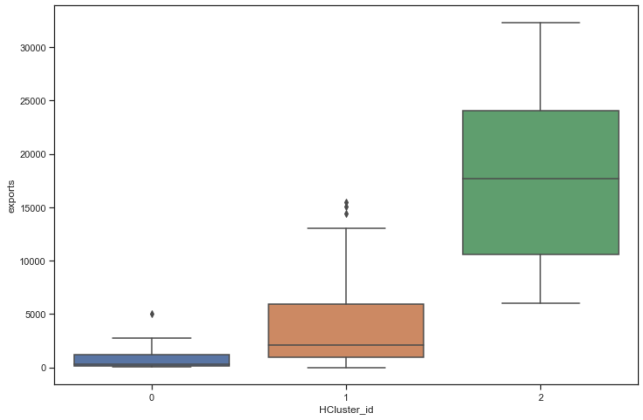


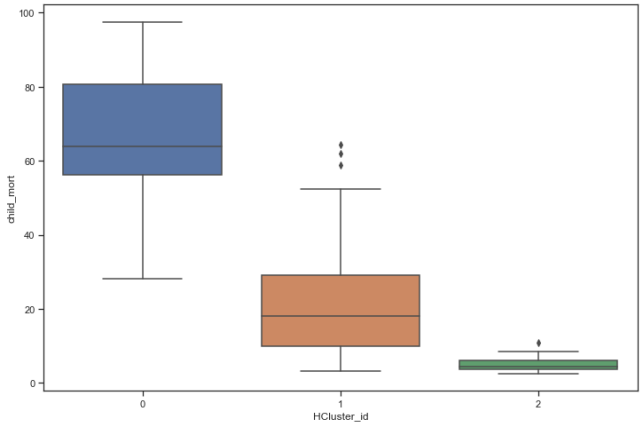
Hierarchical Clustering

Hierarchical Clustering - Box Plot

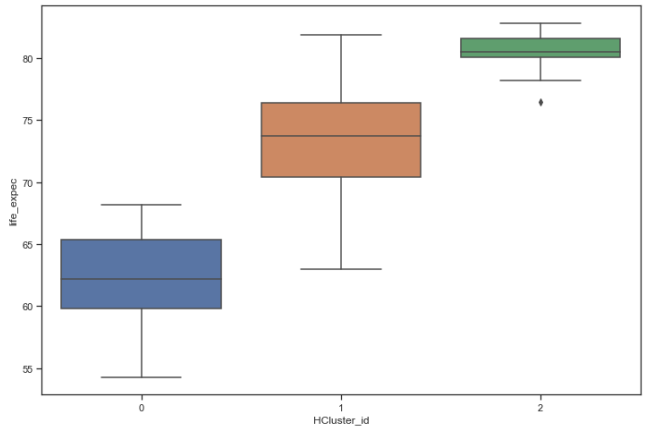


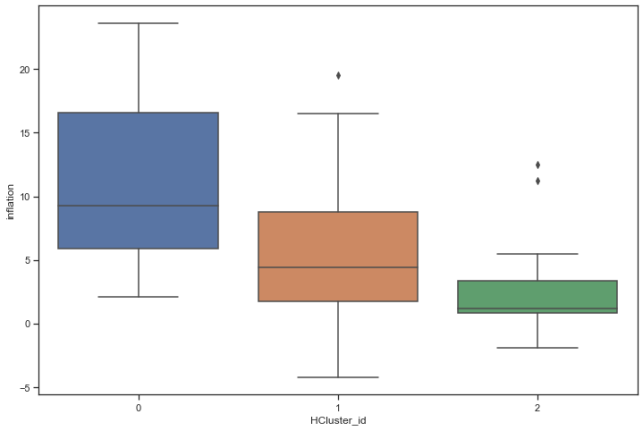


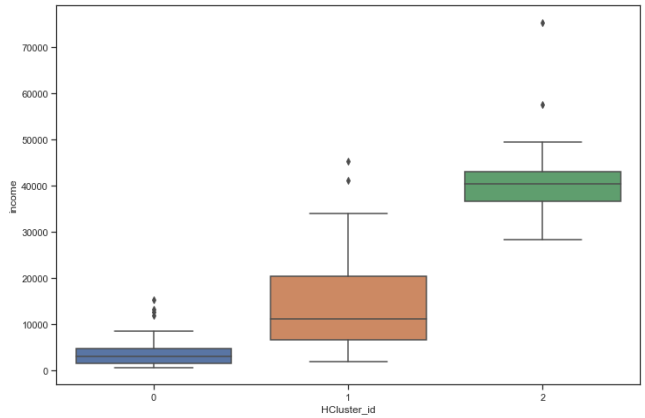


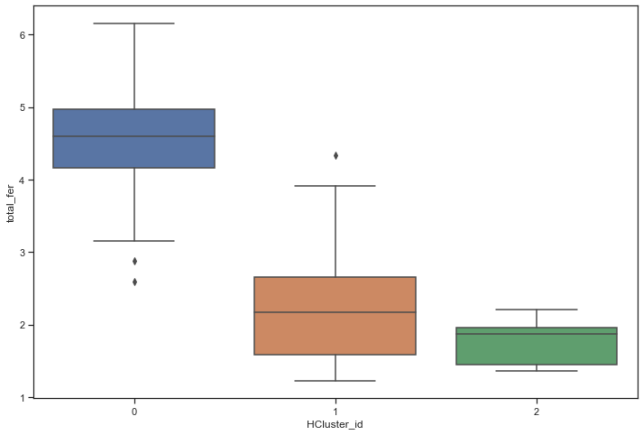


Hierarchical Clustering – Box Plot









Analysis of the Clustering

CLUSTERING AND PCA

KMeans and Hierarchical Clustering Analysis

KClusterID	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	5.322222	14504.02593	3249.133704	13650.45556	37170.37037	2.204556	79.777778	1.749259	34696.2963
1	20.221311	3071.619836	406.979256	3160.474656	12796.55738	6.627721	73.208197	2.272295	6636.360656
2	63.651613	861.023623	128.77869	858.977519	4450	9.736452	63.329032	4.111935	2072.16129

HCluster_id	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	68.478261	884.774513	135.873852	838.634174	4650.869565	10.711304	62.169565	4.507391	2152.73913
1	20.960494	4030.595414	635.004141	4139.667544	14473.7037	5.88916	73.403704	2.25963	8954.839506
2	5.173333	17256.08	4132.27	15558.52667	42853.33333	2.817467	80.486667	1.774	42066.66667

	country
0	Afghanistan
1	Botswana
2	Comoros
3	Congo, Rep.
4	Eritrea
5	Gabon
6	Gambia
7	Ghana
8	Iraq
9	Kenya
10	Lao
11	Liberia
12	Madagascar
13	Mauritania
14	Namibia
15	Pakistan
16	Rwanda
17	Solomon Islands
18	South Africa
19	Sudan
20	Tanzania
21	Uganda
22	Yemen

Data Analysis

Data Analysis perfomed using KMeans and Hierarchical Clustering shows that the countries that re in direst need of aid are listed below, by taking into the consideration of Variables:

- child_mort : Dath of childern under 5 is more
- exports : Less % of goods and service
- health : Less % of spedning
- imports: Less % of imports
- income : Less % of Income
- life_expec : Less average number of uears a new born would live
- gdpp : Less gdpp growth