HELP International

Identification of Countries Qualifying for Financial Aid

AGENDA

- Effective and strategic use of the fund of $10Mn as part of financial aid to countries that are in it's dire need.
- To identify top 5 from the existing list of 167 countries that qualify the eligibility metrics of HELP International.

Resources & Approach

- Data Source- Data of 167 countries updated with numeric values of certain socio-economic factors
- Approach- Applied Business logic and Analytical Methods of grouping data into best clusters identified
- Platform Used-Python

Certain pre-analytics hypothesis of the features of qualifying countries..

Socio-Economic Factors

- The countries can have low nominal GDP thus resulting in high inflation, high imports and low exports.
- Low net income per person could be a consequential factor of low nominal GDP.
- Low net income could result into low buying capacity of the population. This shall result in low inflation thus emerging recession situations in these countries.
- The qualifying countries have high mortality rate.
- High mortality rate could be a resulting factor of low health expenditure per capita.
- High mortality rate could also be due to high fertility rate.
- Inadequate health facilities can also be a reason for low life expectancy in these countries.

Analytical Approach-Clustering

Steps taken after importing data in Python

Step1-
Exploratory Data Analysis

- Understanding data and its features
- Missing value identification
- Appropriate data conversions *(eg.; relative values converted into absolute values)*
- Outlier detection and treatment
- Understanding data distribution
- Understanding correlation between variables

Step2-
Clustering

- Hopkins Statistics- To confirm data eligibility for a clustering approach
- Scaling Data- To standardize ML approach towards all variables in the data available
- Clustering approach techniques used:
  - K-Means Clustering
  - Hierarchical Clustering

Step3-
Cluster Profiling

- Identify common features in each cluster
- 3 clusters identified using K-Means clustering
- 4 clusters identified using Hierarchical clustering
- Feature comparison of the clusters obtained using the two methods w.r.t the factors prescribed for cluster profiling
- Finalizing list of top 5 qualified countries

Exploratory Data Analysis-

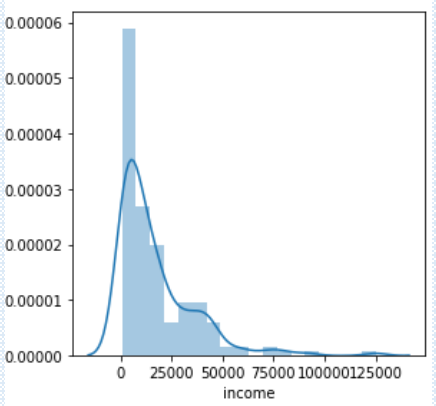
Data Variables and type available for clustering:

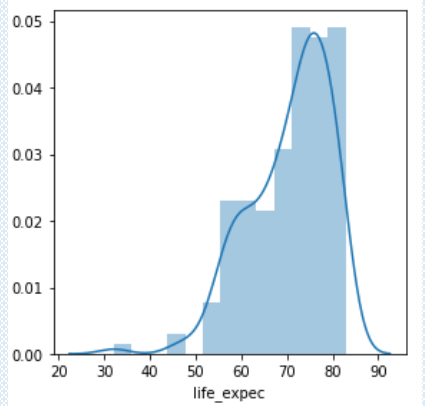
Variable Name	Description	Data Type	Data Treatment applied
country	Name of the country	Object	None
child_mort	Death of children under 5 years of age per 1000 live births	Numeric (float)	Converted to absolute values
exports	Exports of goods and services per capita. Given as %age of the GDP per capita		
health	Total health spending per capita. Given as %age of GDP per capita		
imports	Imports of goods and services per capita. Given as %age of the GDP per capita		
Income	Net income per person	Numeric (Integer)	None
Inflation	The measurement of the annual growth rate of the Total GDP	Numeric (float)	
life_expec	The average number of years a new born child would live if the current mortality patterns are to remain the same		
total_fer	The number of children that would be born to each woman if the current age-fertility rates remain the same		
gdpp	The GDP per capita. Calculated as the Total GDP divided by the total population	Numeric (Integer)	

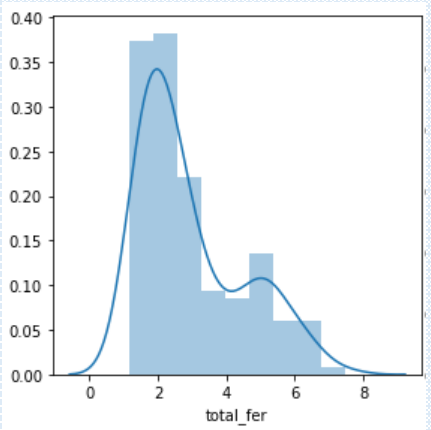
Data Distribution and Missing Values-

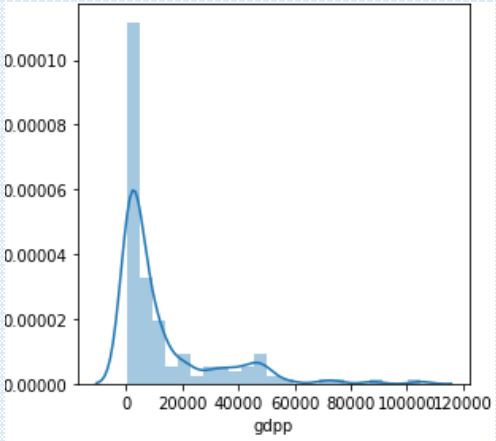
- Data columns that displayed distribution variation are:
  - ✓ ‘child_mort’(child mortality rate)
  - ✓ 'life_expec’ (life expectancy ratio)
  - ✓ ‘total_fer’ (fertility rate)
  - ✓ ‘gdpp’ (nominal GDP)
  - ✓ ‘income’
- No missing values identified in the data

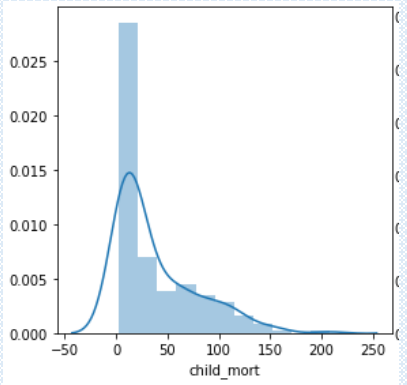
Variables with distribution anomaly-





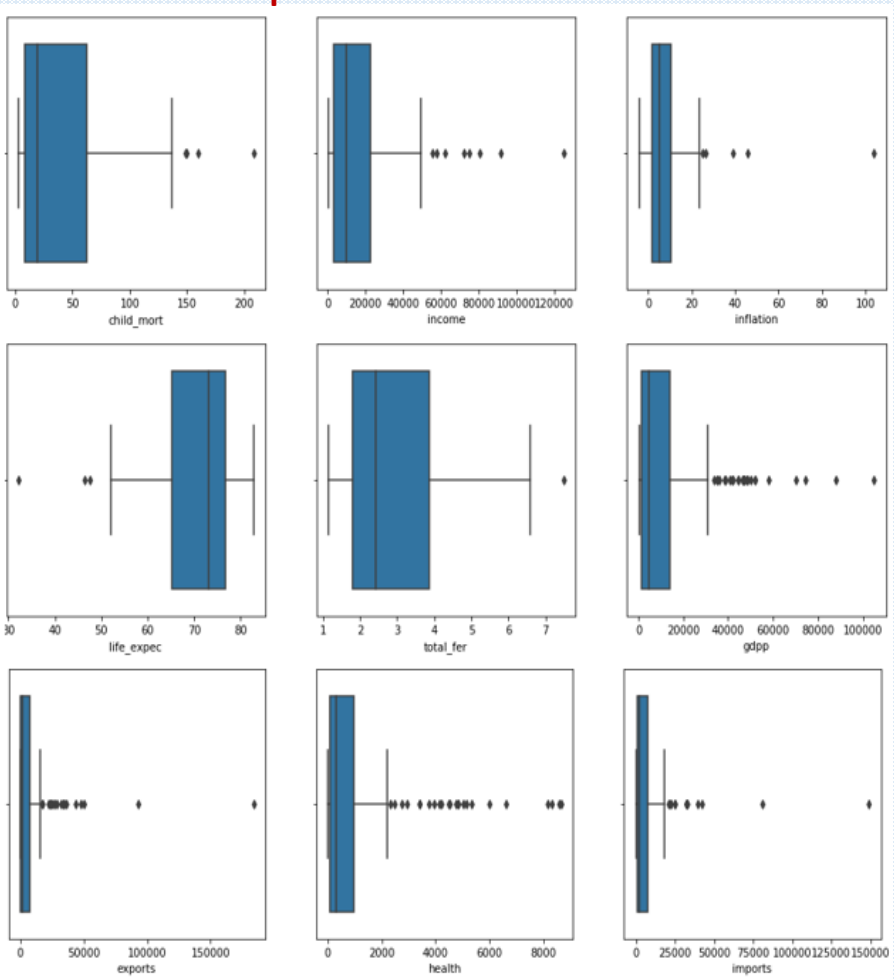






Outlier Detection and Treatment

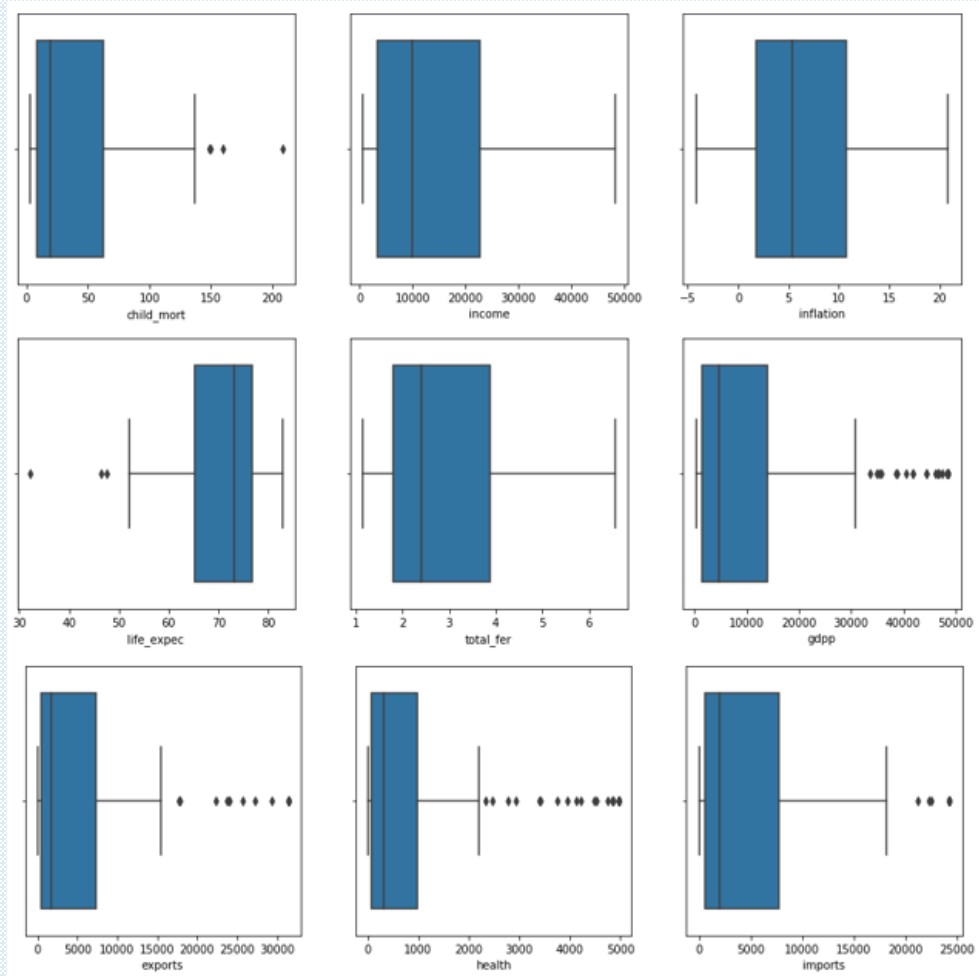
Outlier detection prior to treatment



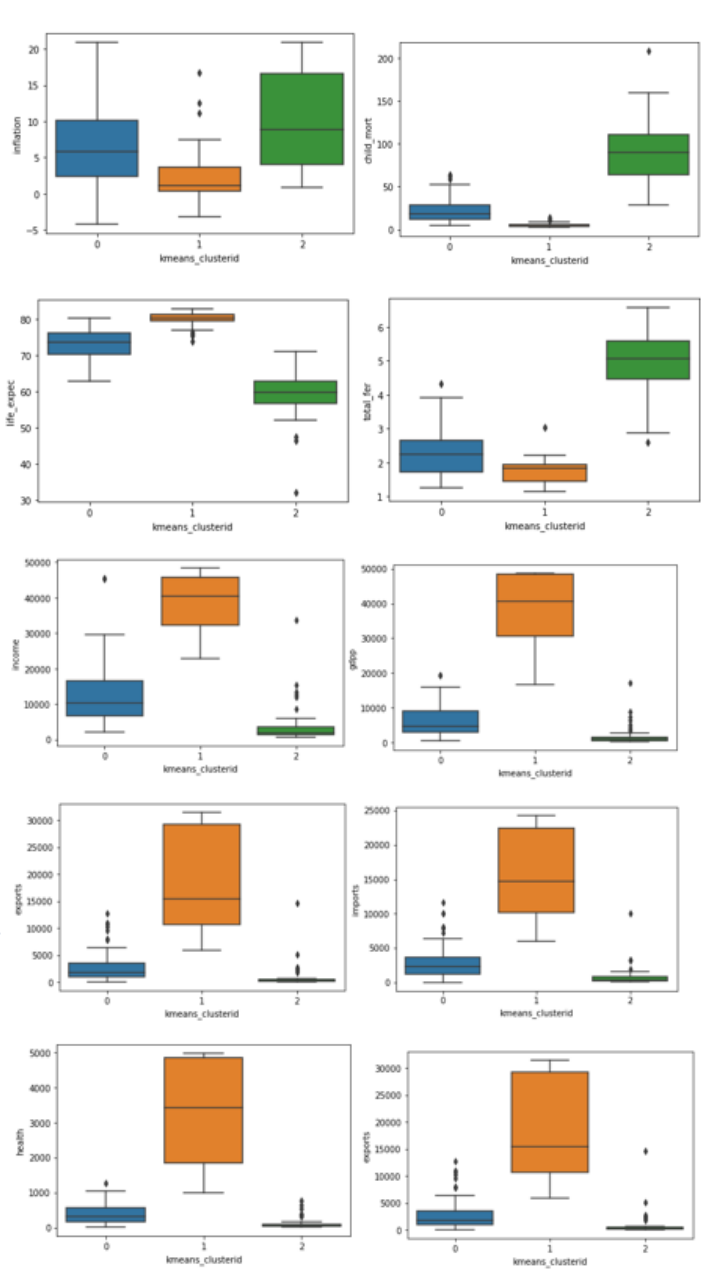
- All data columns have outliers

- Data columns for income, inflation, GDP, exports, health, & imports capped at the higher range of 95th percentile.
- Data column for total fertility capped at 95th percentile.
- Data column for child mortality *(child_mort)* has not been treated.

Post-treatment outlier detection



Summary Statistics- KMeans Clustering (k=3)

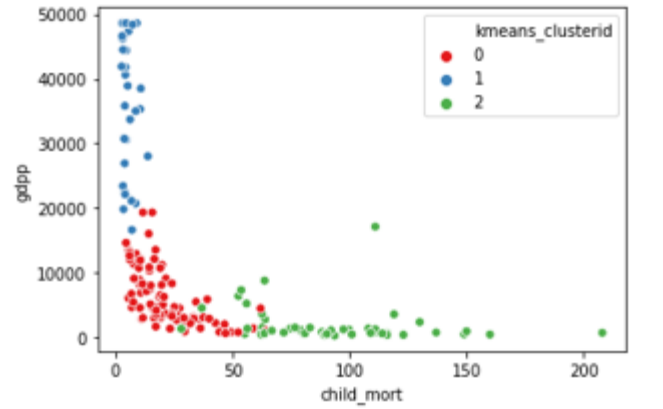


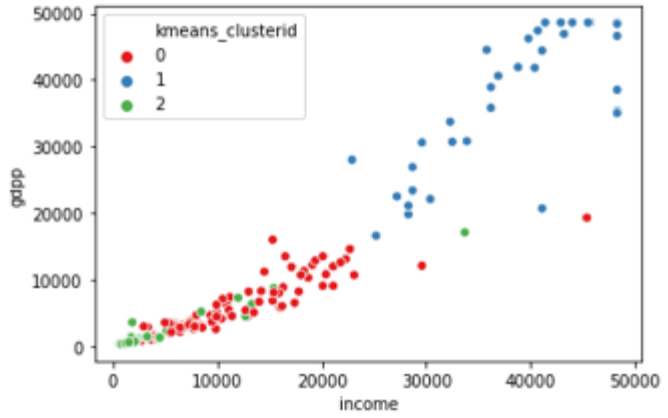
Cluster 2 as compared to other clusters has displayed following attribute behavior:

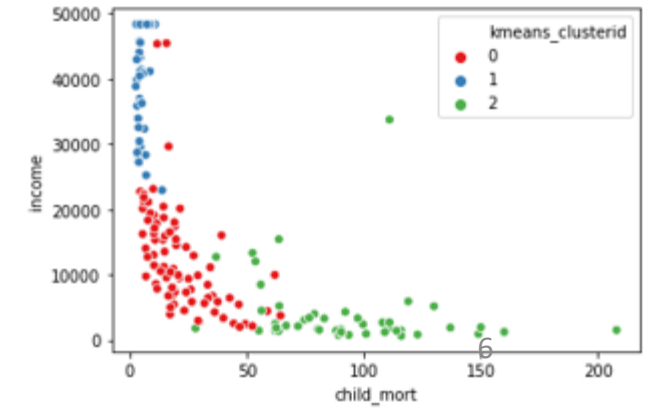
- ✓ High to moderate variation for child mortality, & inflation
- ✓ Low data distribution for Income, Nominal GDP, Exports & Imports, &Health

Relative behavior of each cluster with the prescribed factors for cluster profiling i.e.; GDP, Child Mortality, & Income per person:

- ✓ Cluster 2 displays a higher inversely proportional relationship between GDP and Child Mortality than other clusters.
- ✓ Cluster 1 has a high range of income per person and increases significantly with increase in GDP
- ✓ Cluster 2 claims that countries with low income have high child mortality rate.







Cluster Profiling & Final List-KMeans Clustering

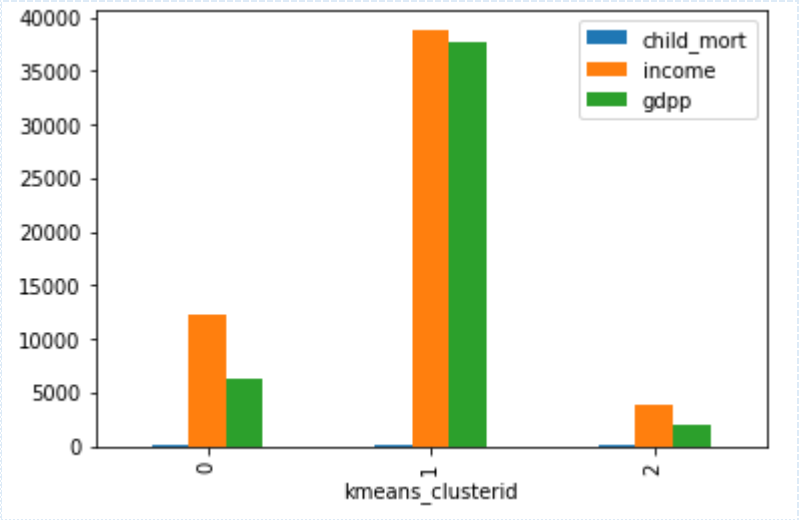
Based on the summary statistics and cluster profiling…

- Countries belonging to cluster 2 are in direst need of financial aid
- These countries have high child mortality rate, low income per person, & low nominal GDP

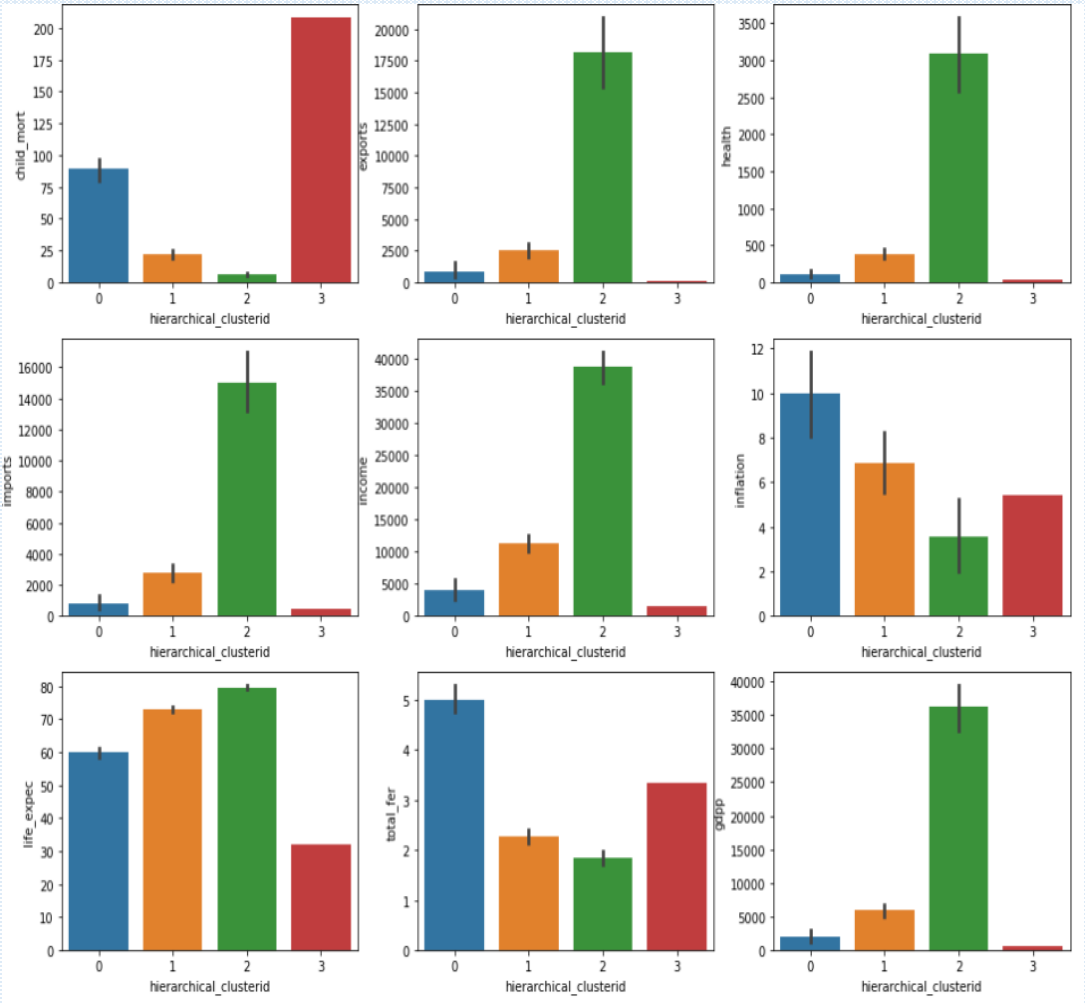
The top 5 countries that are qualify from cluster 2 are:

Country	Child Mortality	Net Income per person	Nominal GDP	kmeans_clusterid
Haiti	208	1500	662	2
Sierra Leone	160	1220	399	2
Chad	150	1930	897	2
Central African Republic	149	888	446	2
Mali	137	1870	708	2

Cluster	Data Count
0	82
1	48
2	37

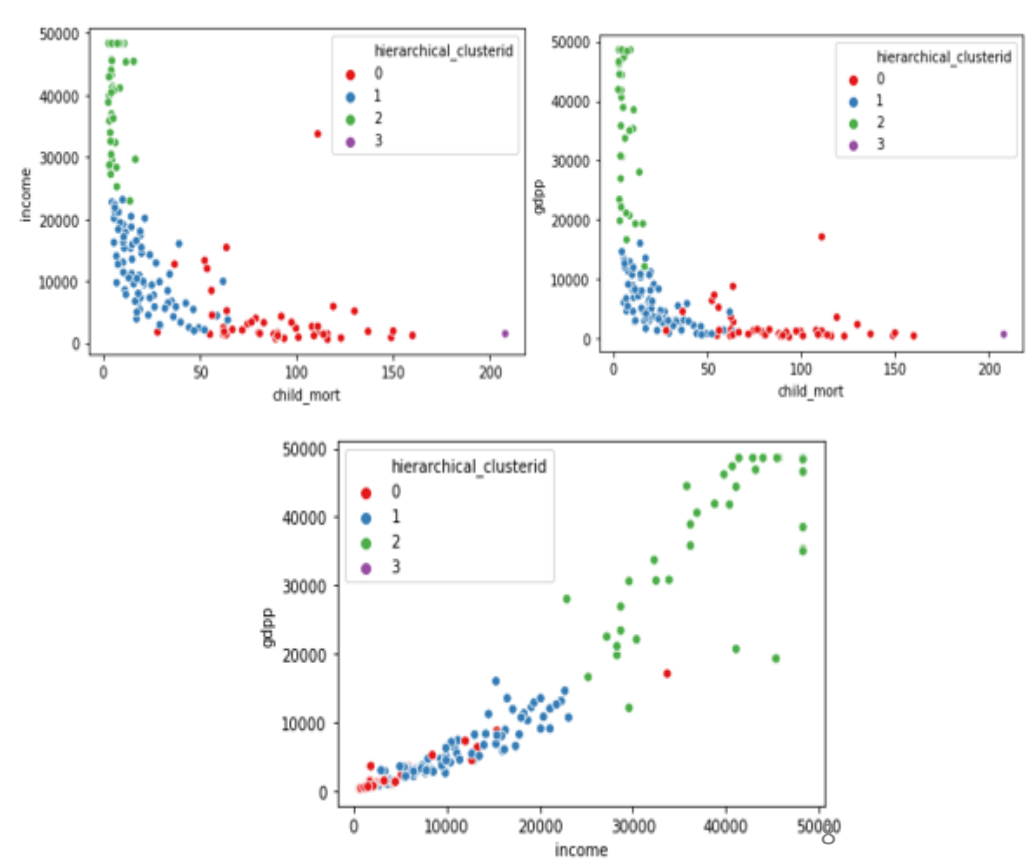


Summary Statistics- Hierarchical Clustering (k=4)



Relative behavior of each cluster with the prescribed factors for cluster profiling i.e.; GDP, Child Mortality, & Income per person:

- ✓ Cluster 0 displays a higher inversely proportional relationship between GDP and Child Mortality than other clusters.
- ✓ Cluster 2 has a high range of income per person and increases significantly with increase in GDP followed by cluster 1 and cluster 0 in descending order.
- ✓ Income Vs Child Mortality scatter plot claims that countries with high income have low child mortality rate and vice-versa.



Cluster 0 as compared to other clusters has displayed following attribute behavior:

- ✓ High to moderate data variation for child mortality, Inflation, Life Expectancy, Total Fertility
- ✓ Low data distribution for Export & Import, Health, Income, & GDP

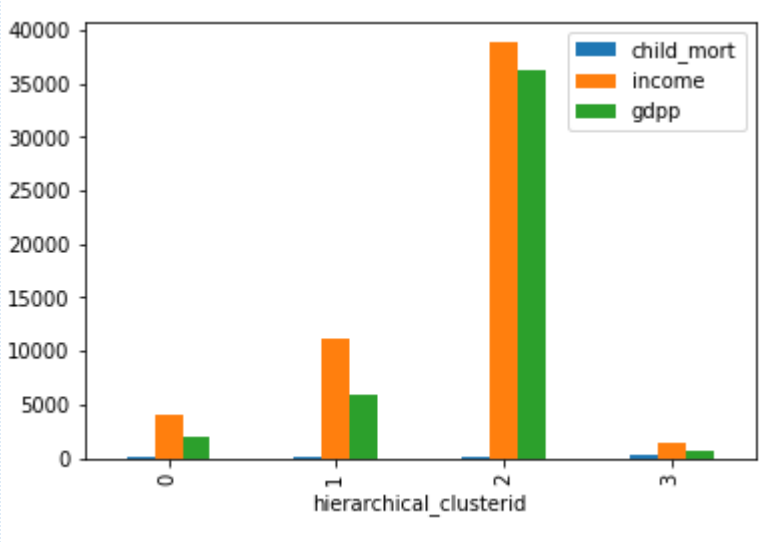
Cluster Profiling & Final List-Hierarchical Clustering

Based on the summary statistics and cluster profiling…

- Since cluster 3 comprises on only one datapoint, hence cluster 0 becomes the next best cluster for the business case.
- Countries belonging to cluster 0 are in direst need of financial aid
- These countries have high child mortality rate, low income per person, & low nominal GDP

Country	Child Mortality	Net Income per person	Nominal GDP	kmeans_clusterid
Sierra Leone	160	1220	399	0
Chad	150	1930	897	0
Central African Republic	149	888	446	0
Mali	137	1870	708	0
Nigeria	130	5150	2330	0

Cluster	Data Count
0	79
1	47
2	40
3	01



Conclusion

- Common countries identified as result of both the clustering techniques of K Means & Hierarchical:
  - Sierra Leone
  - Chad
  - Central African Republic
  - Mali
- Exception in final selection of top countries that require financial aid:
  - Haiti-Identified during K-Means Clustering
  - Nigeria- Identified during Hierarchical Clustering
- Comparing other socio-economic factors for both of these countries, the one that supersedes in qualification for financial aid is:
  - Haiti

Country	Child Mortality	Income	Inflation	Life Expectancy	Total fertility	Nominal GDP	Export	Health	Import
Haiti	↑ 208	↓ 1500	↓ 5.45	↓ 32.1	↓ 3.33	↓ 662	↓ 101	↓ 46	↑ 428
Nigeria	↓ 130	↑ 5150	↑ 20.87	↑ 60.5	↑ 5.84	↑ 2330	↑ 589	↑ 118	↓ 405

- Therefore, the final list of top 5 countries that qualify for financial aid from HELP International is based on **K-Means clustering**.

Country	Child Mortality	Income	Inflation	Life Expectancy	Total fertility	Nominal GDP	Export	Health	Import	K-Means ClusterID
Haiti	208	1500	5.5	32.1	3.3	662	101	46	428	2
Sierra Leone	160	1220	17.2	55.0	5.2	399	67	52	138	2
Chad	150	1930	6.4	56.5	6.6	897	330	41	390	2
Central African Republic	149	888	2.0	47.5	5.21	446	53	18	118	2
Mali	137	1870	4.4	59.5	6.6	708	161	35	249	2

Thank you