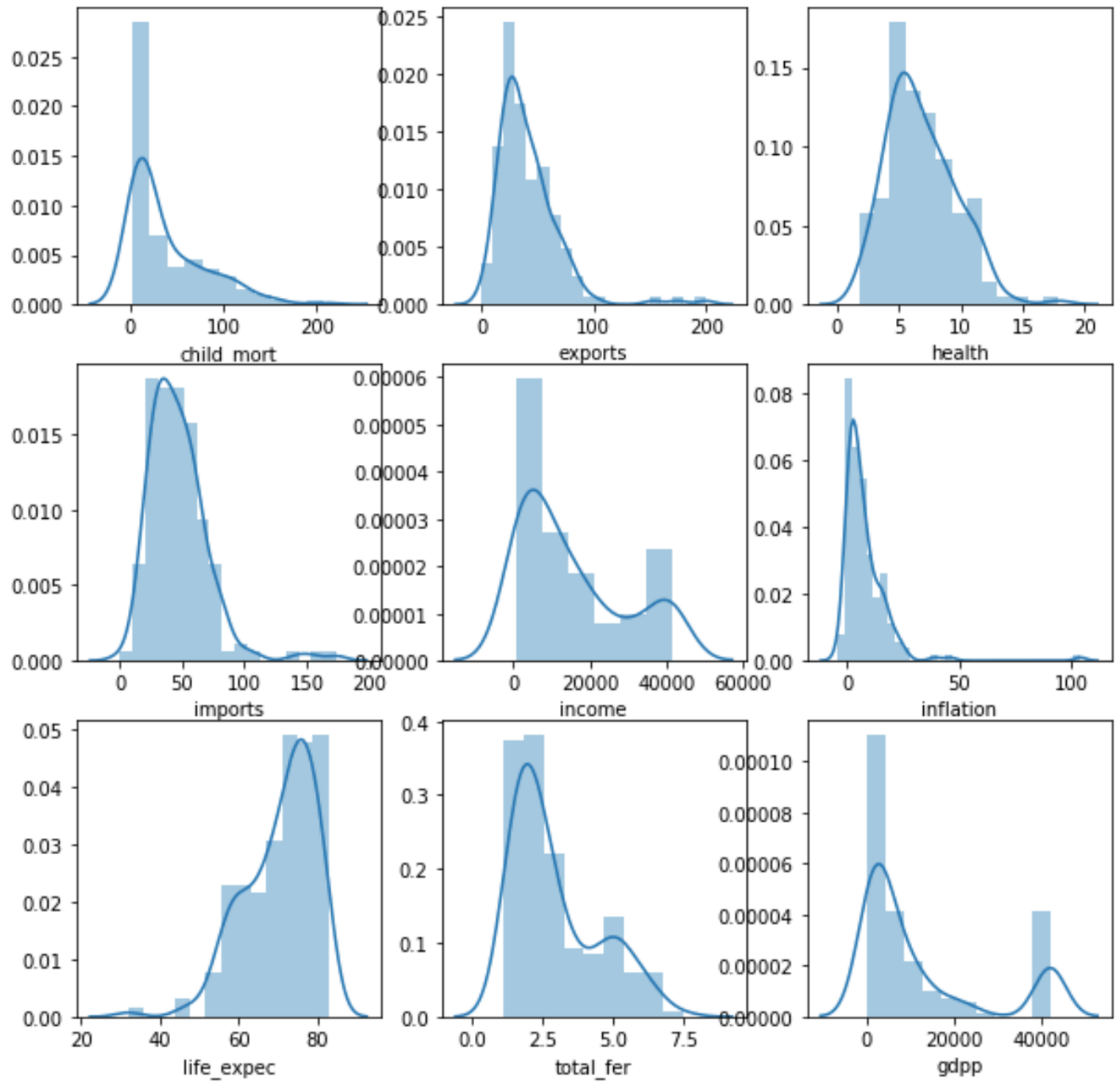
Country Clustering

Exploratory Data Analysis

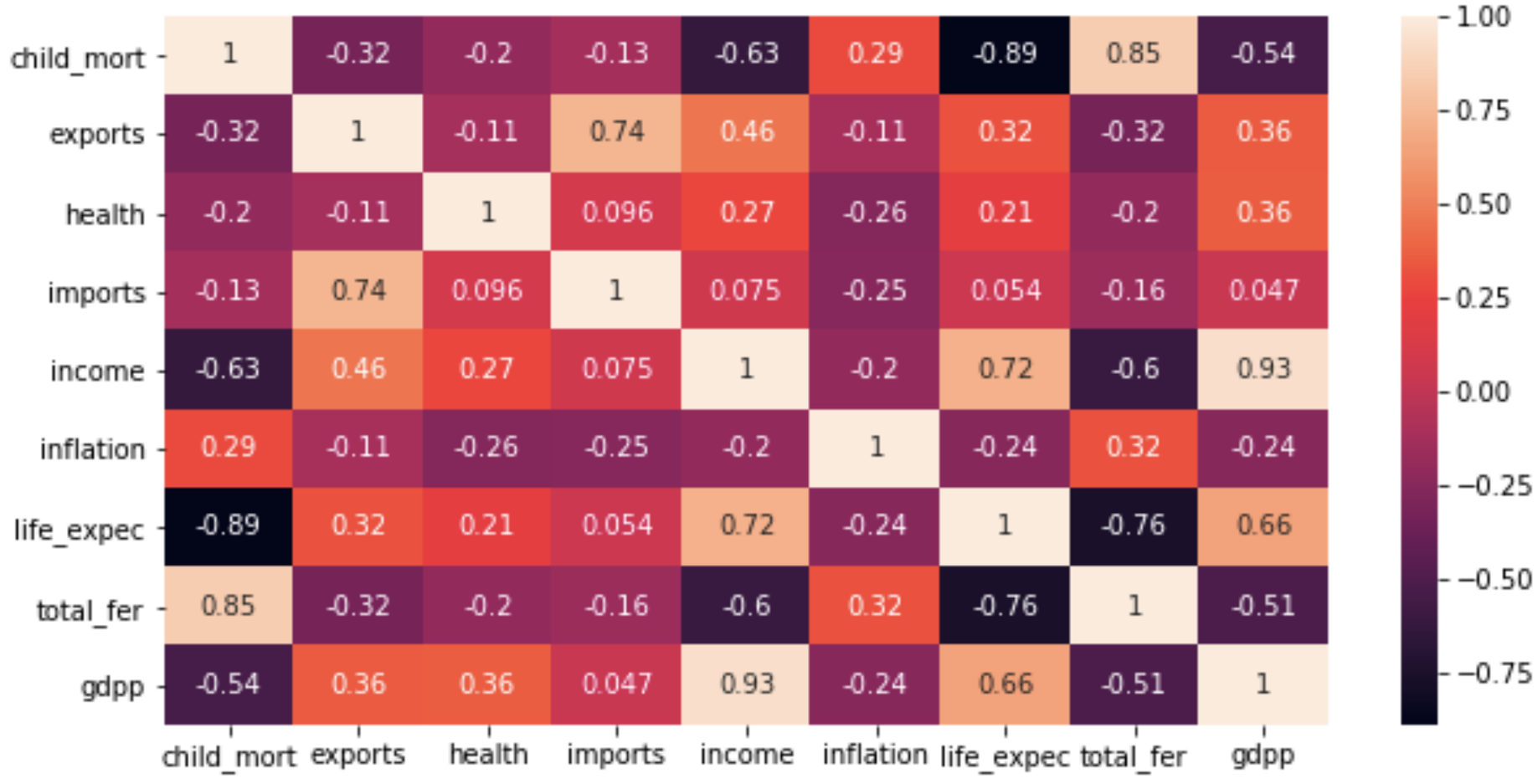
Univariate Analysis

The variables 'life_expec', 'total_fer', 'gdpp','income' and 'child_mort' seem to have more predictive power than the other variables based on the shape of the graph and the variance in data we can observe



Exploratory Data Analysis

Bivariate Analysis



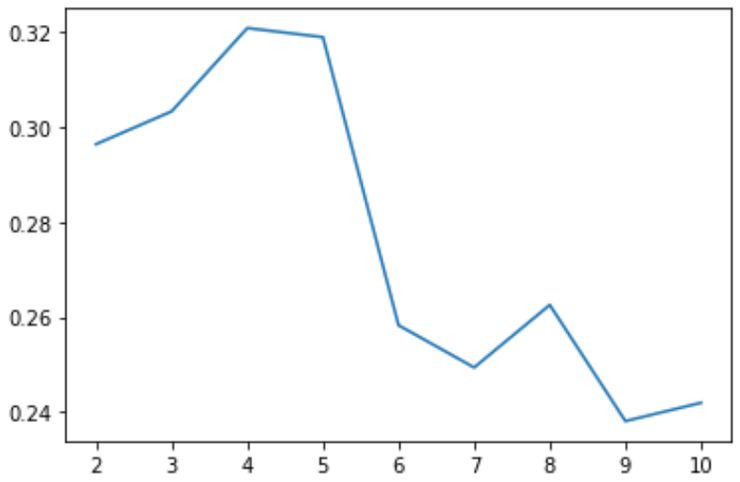
- **'child_mort' is highly correlated with 'life_expec'**
- **'total_fer' is also highly correlated with 'child_mort'**
- **We can just use 'child_mort' and not loose a significant predictive power of 'life_expec' and 'total_fer'**
- **The selection will make clustering easier**

Hopkins Score

0.962565830592441
0.959646719564542 7
0.952453222947465 3
0.967699269147393 3
0.896249777408795 7
0.936531790979283 3
0.950957017211189 2
0.903931700954016 1
0.951211771368799 4
0.897476113277423 7

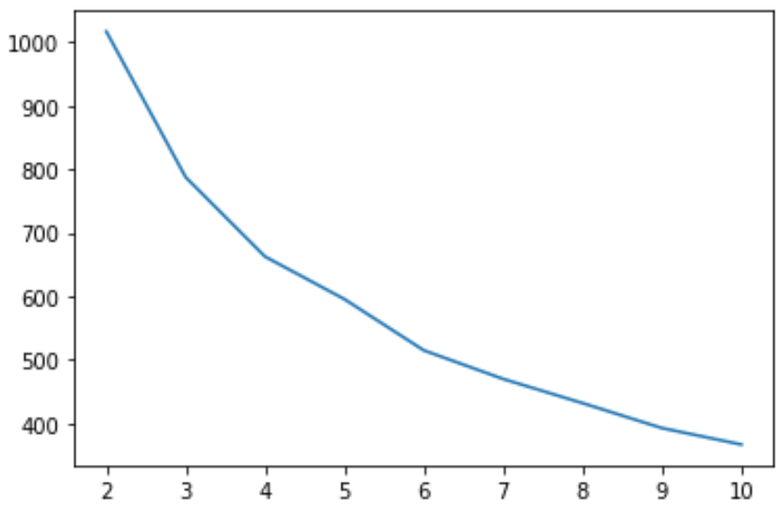
The Hopkins score is greater that 80 percentage even after runining it more than 10 times. So we can understand that the data has good clustering tendency

K Means Clustering



Selecting k based on Silhouette Score

- **1. Peak value of silhouette score is in between 3 and 4**

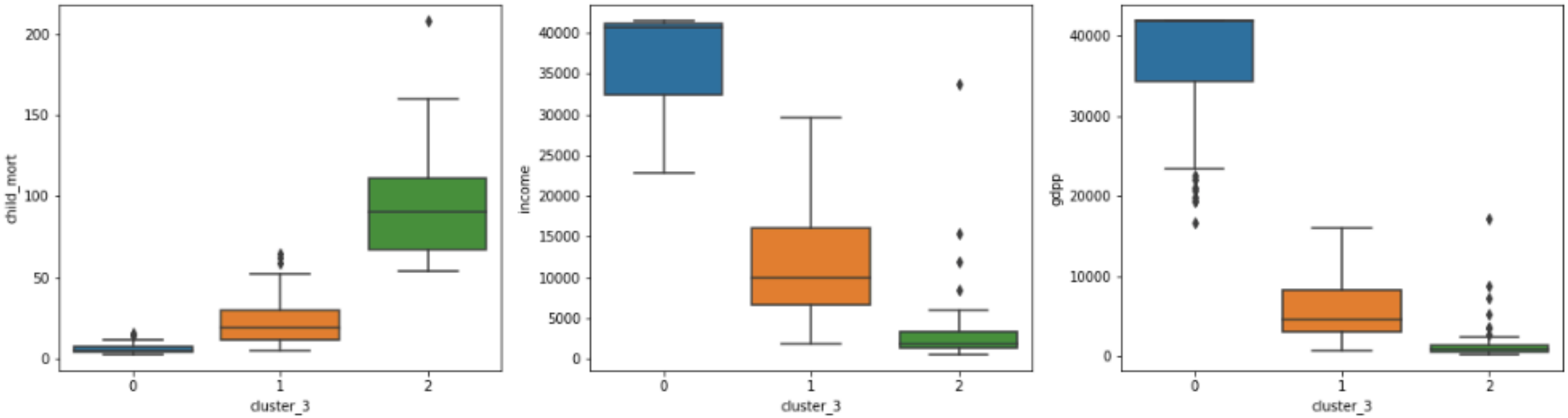


Selecting k based on Elbow Curve

- **1. In this curve we have a break point at 3,4 and 6**
- **2. In elbow curve we prefer choosing the lesser clusters in case of multiple bends**

▶ We chose, k = 3/4

K Means Clustering with k =3



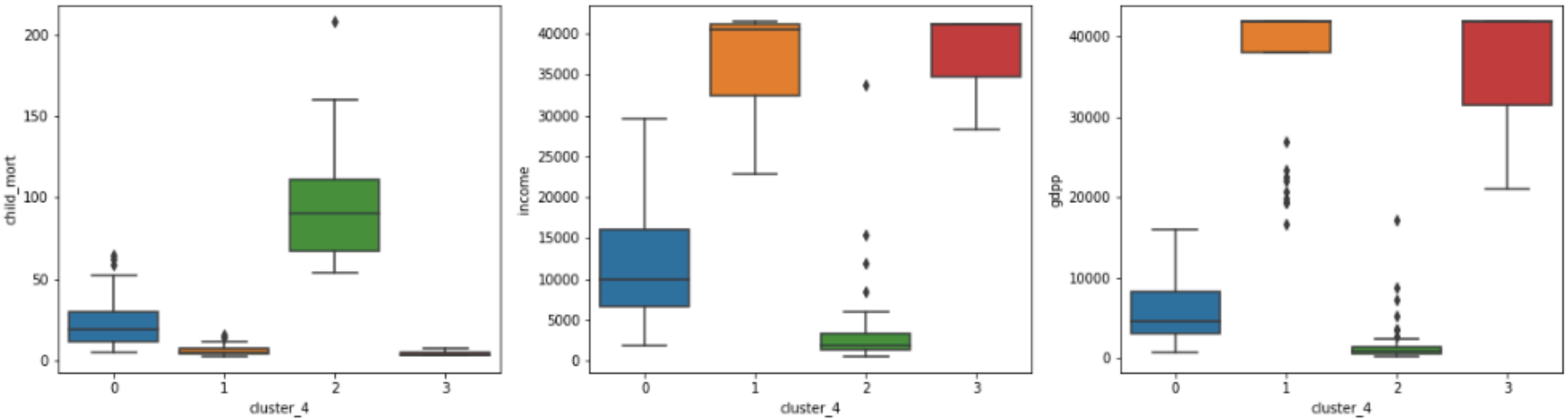
1. Cluster 2 has very high child mortality, low net income per person and very low GDP per capita
2. Cluster 1 has medium range child mortality , medium range net income per person and low GDP per capita
3. Cluster 0 has low child mortality , high net income per person and high GDP per capita

---

The countries we can recommend the CEO to focus on based on K Means clustering as k = 3 are :

1. Haiti
2. Sierra Leone
3. Chad
4. Central African Republic
5. Mali

K Means Clustering with k =4



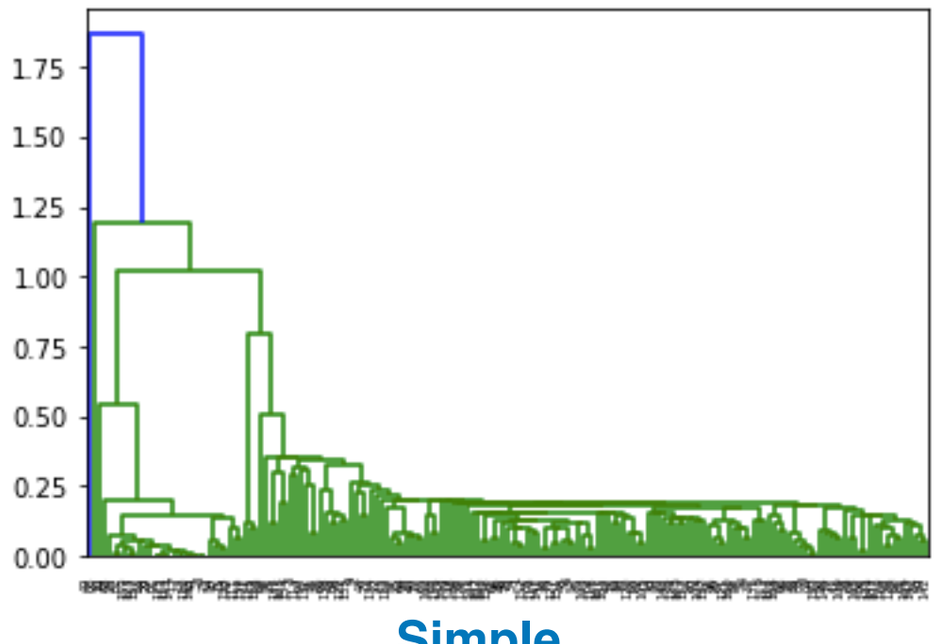
1. Cluster 0 has low child mortality, low net income per person and low GDP per capita
2. Cluster 1 has low range child mortality , high net income per person and high GDP per capita
3. Cluster 2 has high child mortality , very low net income per person and very low GDP per capita
4. Cluster 3 has low child mortality , high net income per person and high GDP per capita

---

The countries we can recommend the CEO to focus on based on K Means clustering as k = 4 are :

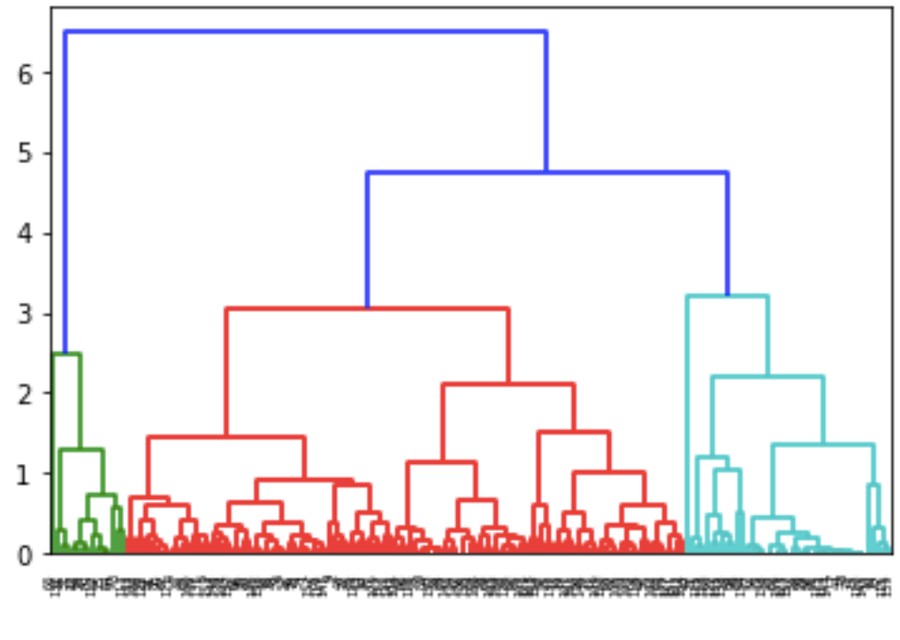
1. Haiti
2. Sierra Leone
3. Chad
4. Central African Republic
5. Mali

Hierarchical Clustering



Simple Linkage

1. It is very difficult to read the results based on single linkage
2. The formation of clusters do not seem to be appropriate



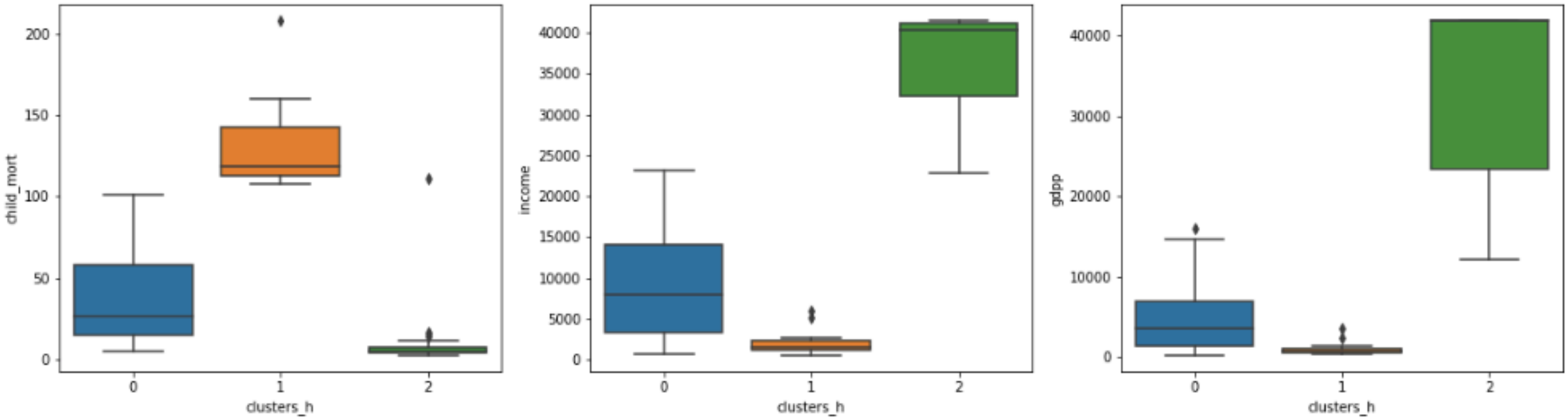
Complete Linkage

1. It is easier to read the result based on complete linkage
2. The formation of clusters seem appropriate

▶ We chose, clusters = 3

Based on the complete linkage dendrogram

Hierarchical Clustering with clusters =3



1. Cluster 0 has medium range child mortality , medium range net income per person and medium GDP per capita
2. Cluster 1 has very high child mortality, very low net income per person and very low GDP per capita
3. Cluster 2 has low child mortality , high net income per person and high GDP per capita

---

The countries we can recommend the CEO to focus on based on Hierarchical clustering are :
1. Haiti
2. Sierra Leone
3. Chad
4. Central African Republic
5. Mali

Final Recommendation

We found the same focus countries based on both K Means clustering and Hierarchical Clustering.

The focus countries are:

1. **Haiti**
2. **Sierra Leone**
3. **Chad**
4. **Central African Republic**
5. **Mali**