Clustering Assignment

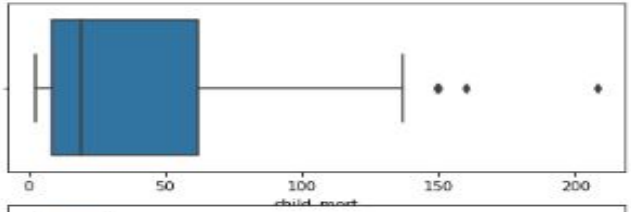
Problem Statement

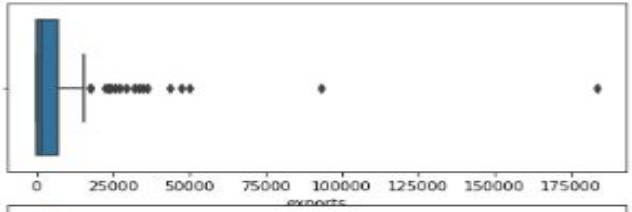
HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy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

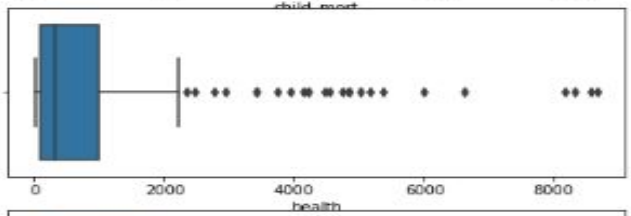
After the recent funding programmes, they have been able to raise around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. The significant issues that come while making this decision are mostly related to choosing the countries that are in the direst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

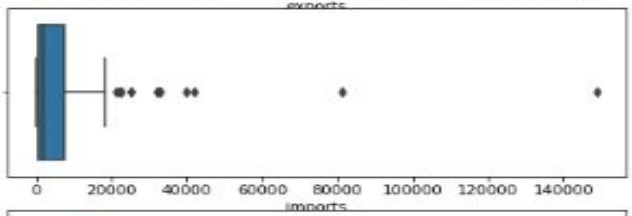
Step 1: Outlier treatment

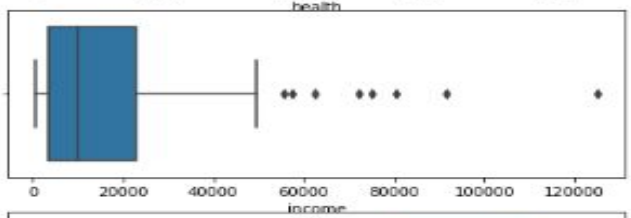
```
sns.boxplot(x=i[1],data=countries_data)
```

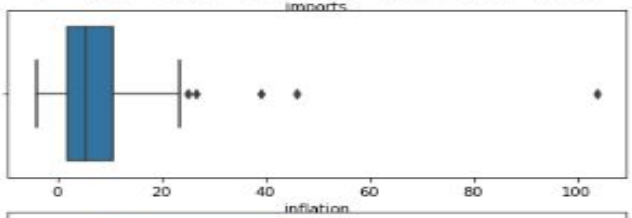


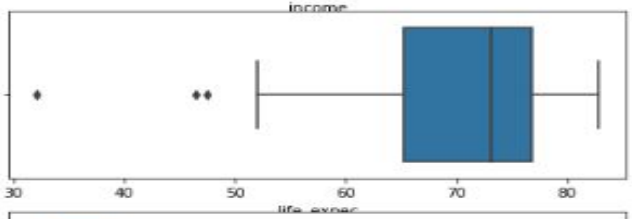


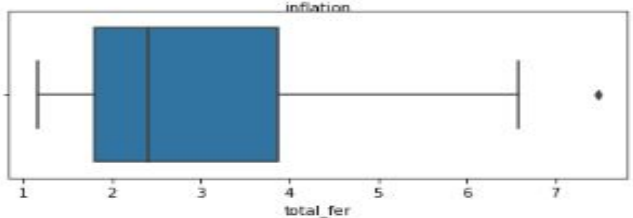


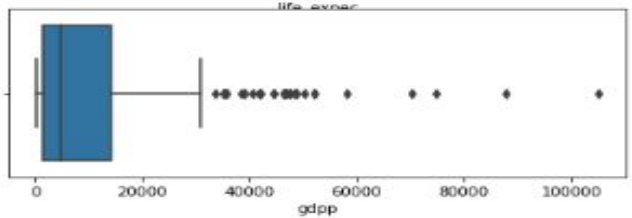






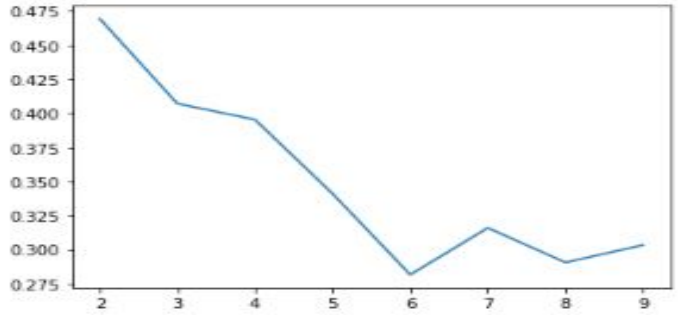






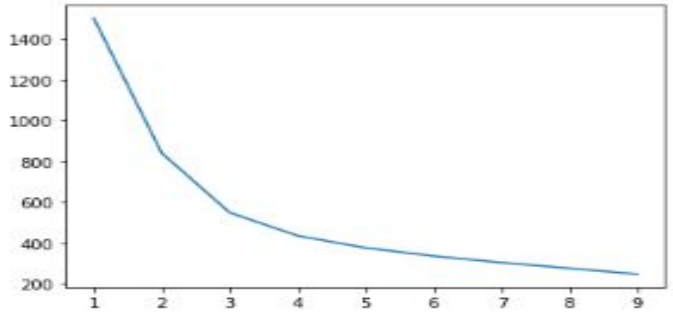
Elbow curve and silhoutte score analysis

```
plt.plot(pd.DataFrame(ssd)[0], pd.DataFrame(ssd)[1]);
```



Looking at the elbow curve

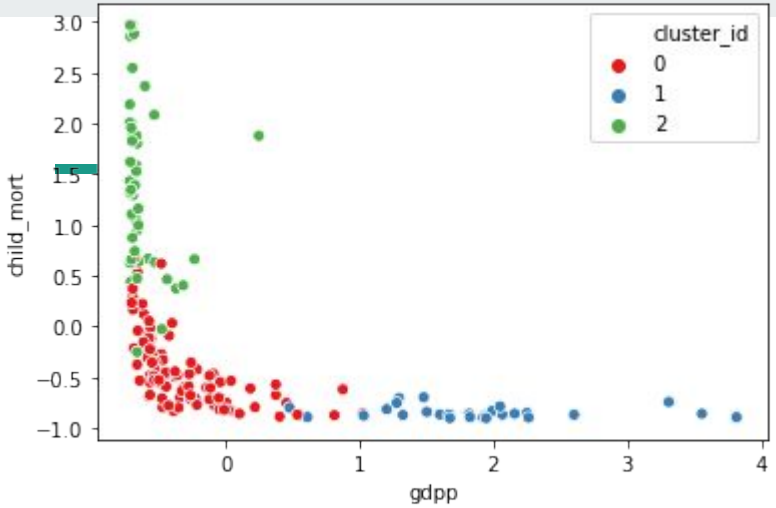
In [340]:
```
#Now let's proceed to the elbow curve method
ssd = []
for k in list(range(1,10)):
    model = KMeans(n_clusters = k, max_iter = 50).fit(countries_data)
    ssd.append([k, model.inertia_])

plt.plot(pd.DataFrame(ssd)[0], pd.DataFrame(ssd)[1]);
```

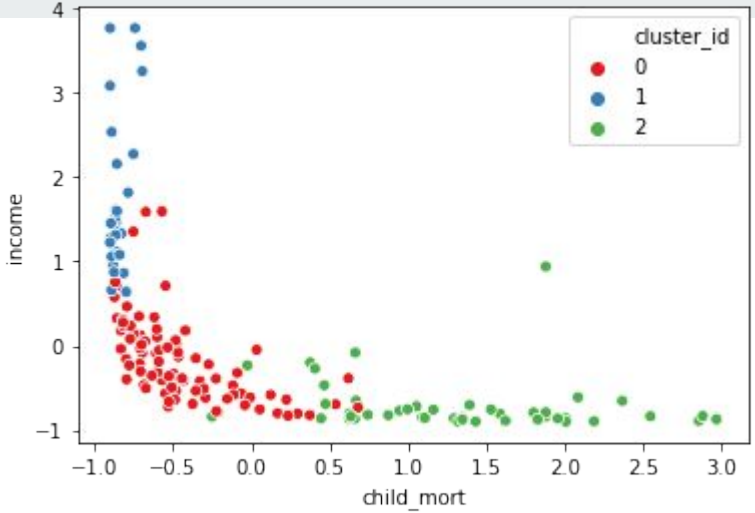


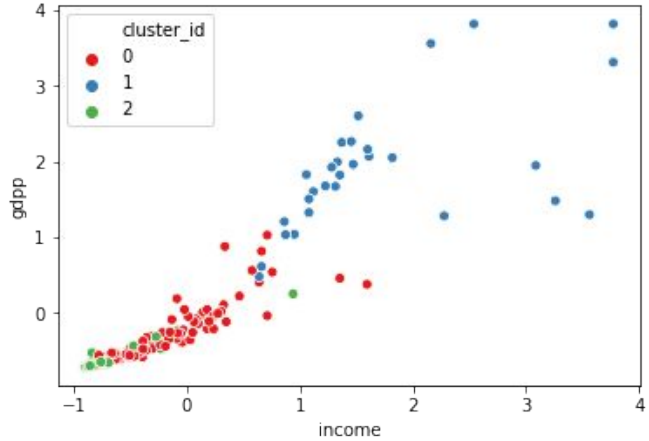
K Means with k=3

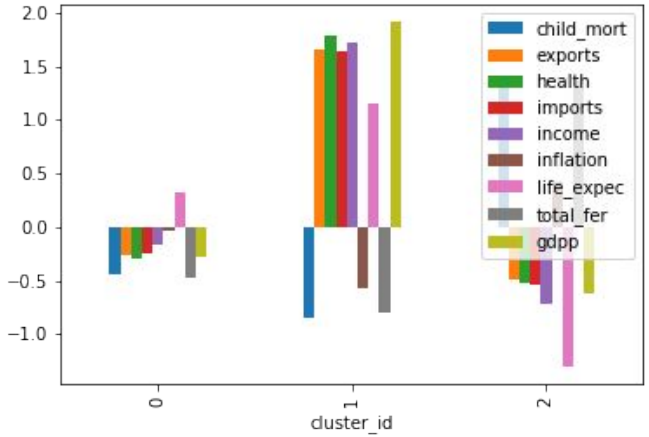
Analysis

After calculating the silhouette score for various number of cluster starting from 2 to 10., there I was able to draw meaningful insights on the number of clusters that we should go ahead with. The optimal number of clusters chosen was 3. Then we did plot the elbow curve, which represents the distance between the point and its assigned center, a lower value means the clustering is good but as the values decreases the number of clusters increases so we need to decide an optimal number.


cluster_id
0
1
2
child_mort
gdpp


cluster_id
0
1
2
income
child_mort


cluster_id
0
1
2
gdpp
income

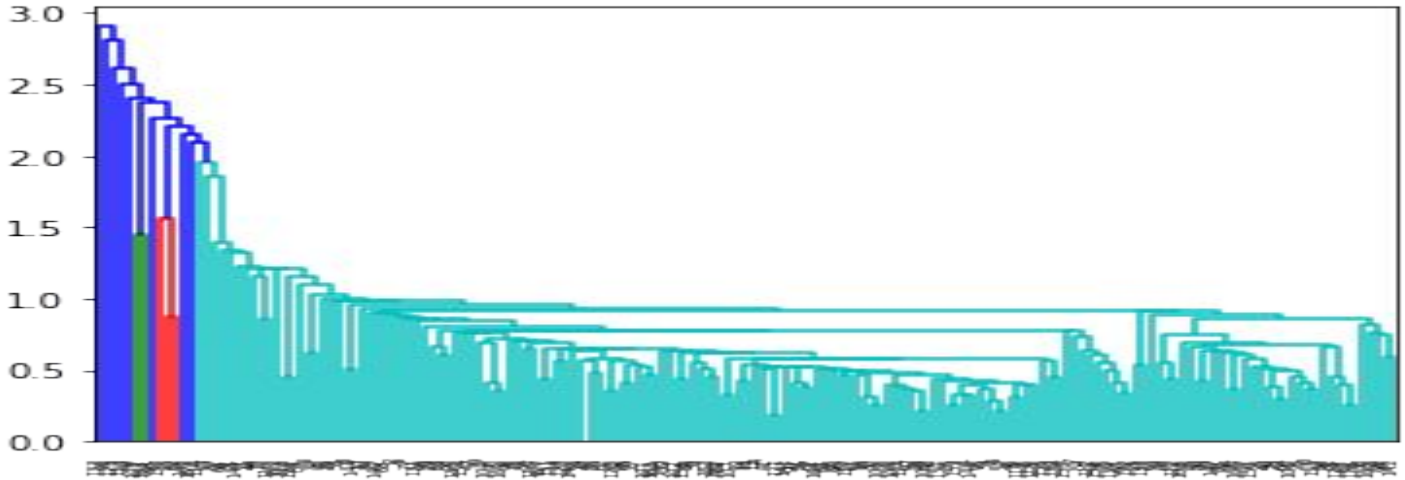

child_mort
exports
health
imports
income
inflation
life_expec
total_fer
gdpp
cluster_id

Findings using K means

Counteries from cluster 2 is in dire need of the aid, these counties are are in cluster 2 arranges in increasing order of GDPP.

Burundi
Liberia
Congo, Dem. Rep.
Niger
Sierra Leone
Madagascar
Mozambique
Central African Republic
Malawi
Eritrea

Analysis using Hierarchical Clustering- Single Linkage



Complete Linkage