

# **NGO DATA-DRIVEN MODEL TO USE RESOURCES STRATEGICALLY AND EFFECTIVELY.**

- **HARSHIT PAUNIKAR**  
( IIIT Bangalore )

# 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.
- Our job is to categorize the countries using some socio-economic and health factors that determine the overall development of the country. We need to suggest the countries which the CEO needs to focus on the most.





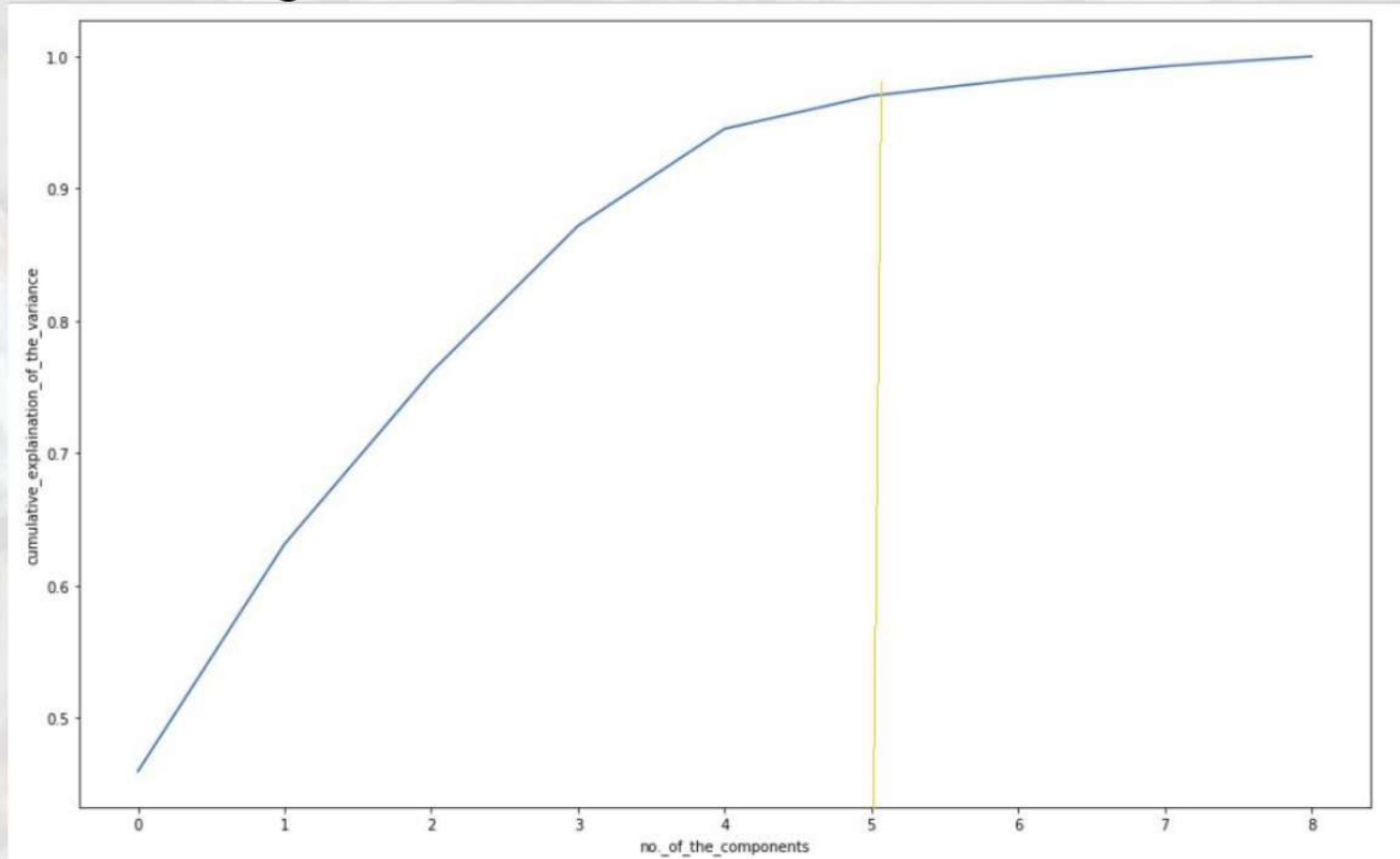
# APPROACH OF ANALYSIS :

- We first check the details about the data present using the data exploration process.
- Then we scaled the data to bring all the components on the same scale so that they are comparable for further analysis.
- After then we performed PCA operation and found Principal components for the available variables.
- By using skree plot we selected the components that could explain most of the variance in the data
- After that we did outlier analysis using boxplot and removed the data points that were beyond the 95 % values as we are interested in the lower performing countries only.
- We checked Hopkins score to make sure data is cluster able (here we get the Hopkins value more than 0.50)
- By using silhouette score and elbow curve we performed K-Means clustering and selected  $k = 3$ .
- Performed Hierarchical clustering (complete linkage) and cut the tree to make 3 clusters as the most of the height was attained there.
- Then we merged the original dataset with the PC data set to check different values of variables for each country in a clusters.
- Then we choose the best model based on the results attained for countries.
- As the result we reported the countries that are in dire need of aid based on the values for the column of GDP, income along with child mortality rate.



# RESULTS OF PCA

- We plot a skree plot after performing PCA to select the Principal Components that help to explain the most of the variance in the data which is about 95% or more. So as the result we took five principal components. There are nine variables that we had reduced to five variables for further analysis without losing much information of the dataset.



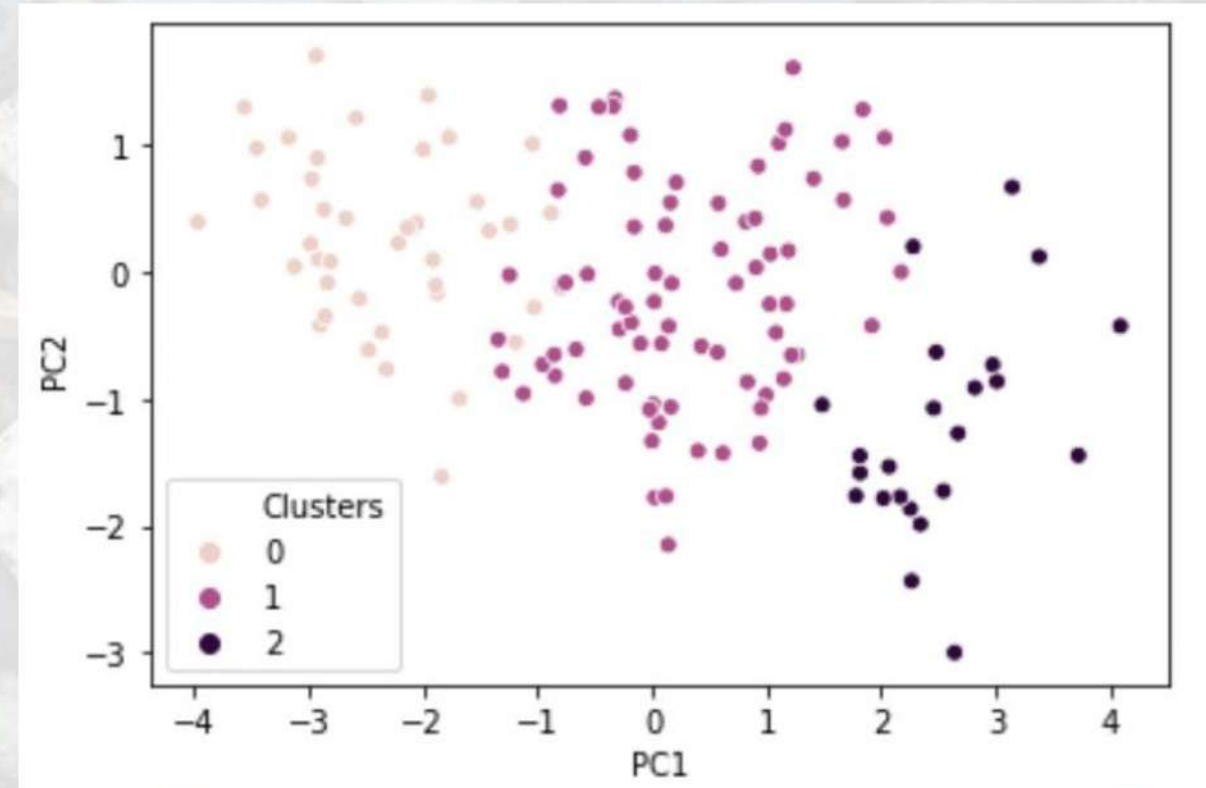


# RESULTS OF CLUSTERING

AFTER WE PERFORMED KMEANS CLUSTERING AND HIEARCHICAL CLUSTERING ON THE DATA WE ANALYZE AND REPORT THE BEST RESULTS AS FOLLOWS:

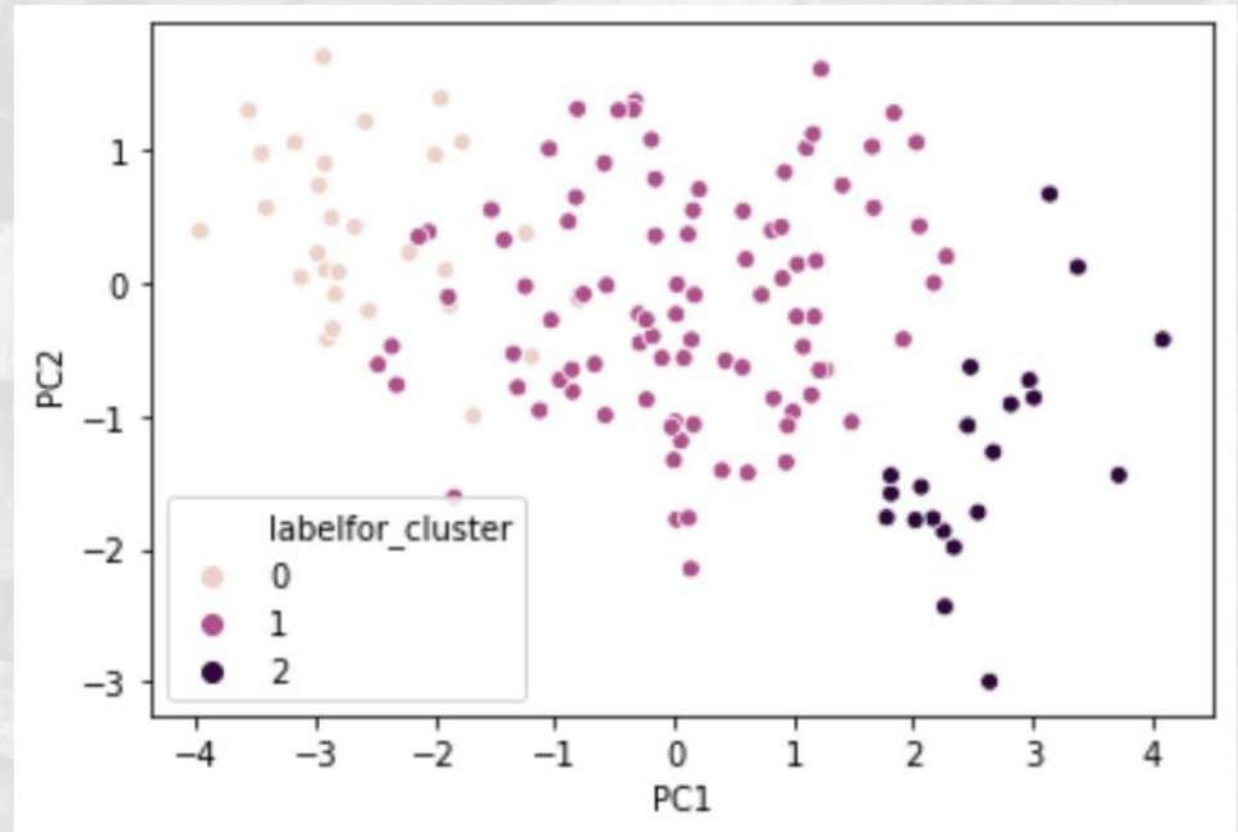
**After performing KMeans clustering we get:**

- a) 41 no. of Countries present in cluster-0
- b) 78 no. of Countries present in cluster-1
- c) 23 no. of Countries present in cluster-2

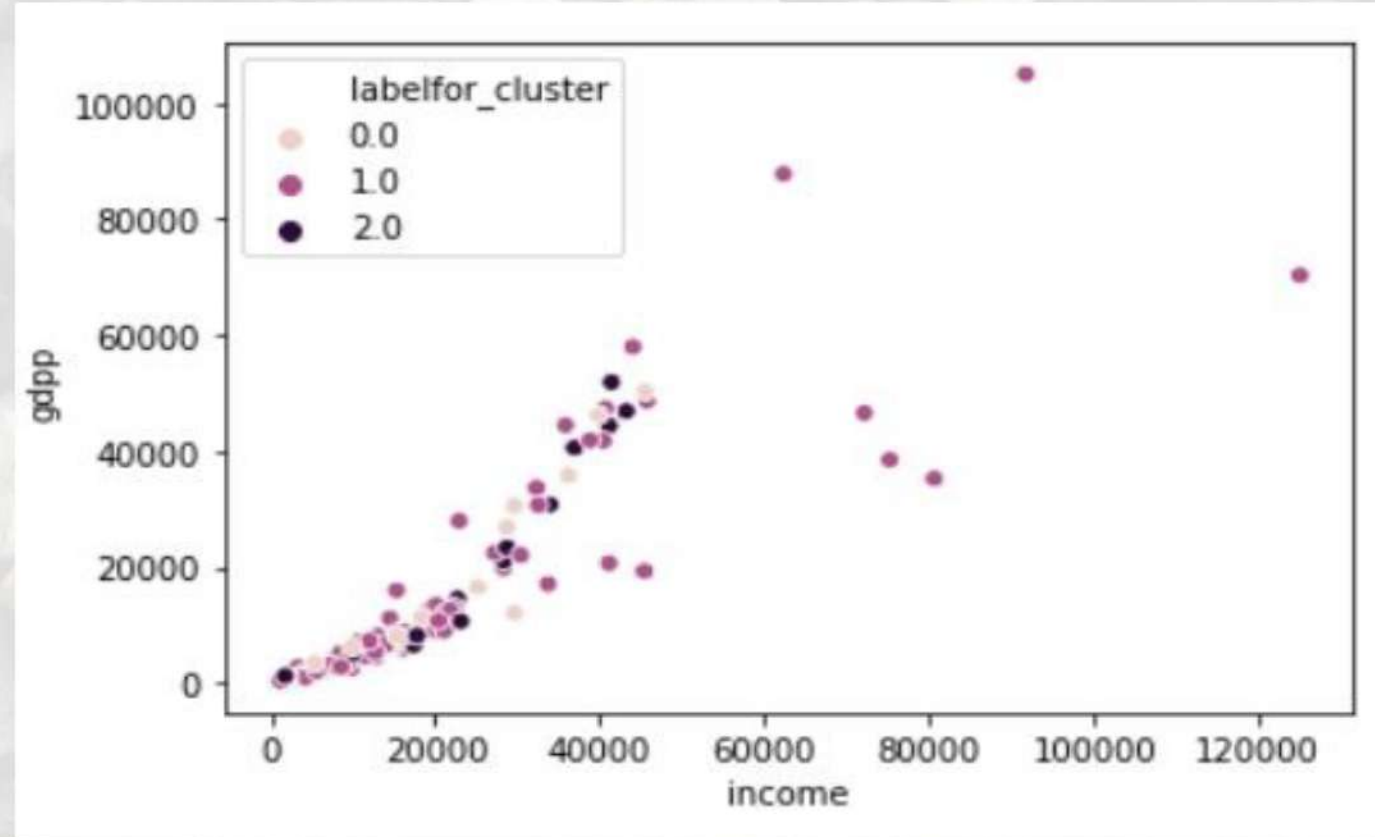


**After performing Hierarchical clustering we get :**

- a) 29 no. of Countries present in cluster-0
- b) 92 no. of Countries present in cluster-1
- c) 21 no. of Countries present in cluster-2

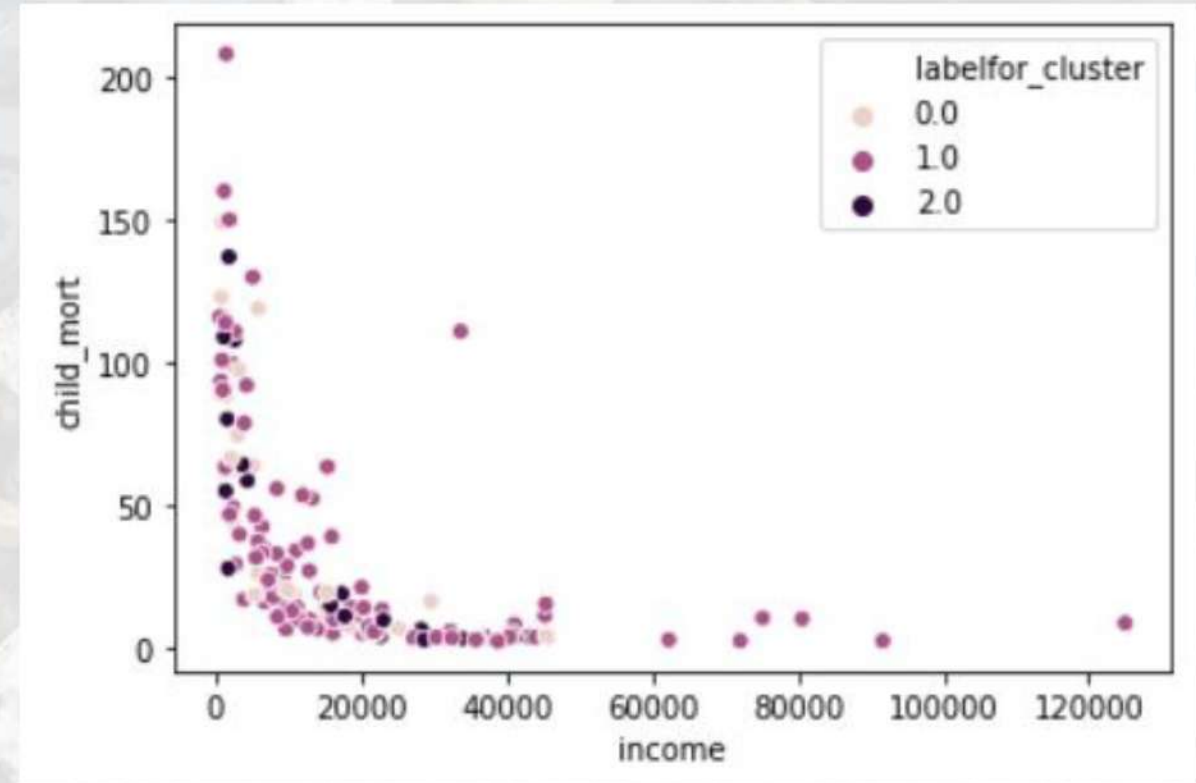


- By performing hierarchical clustering we analyze that it give us better results as most of the countries are having low GDP, income and high child mortality rate so we select this cluster to do our further analysis.
- There is the positive correlation relation between the income and GDP of countries and found that the countries who have high GDP tend to have income of people as high income and viceversa. And most of the countries in cluster-0 are near to origin in the graph.



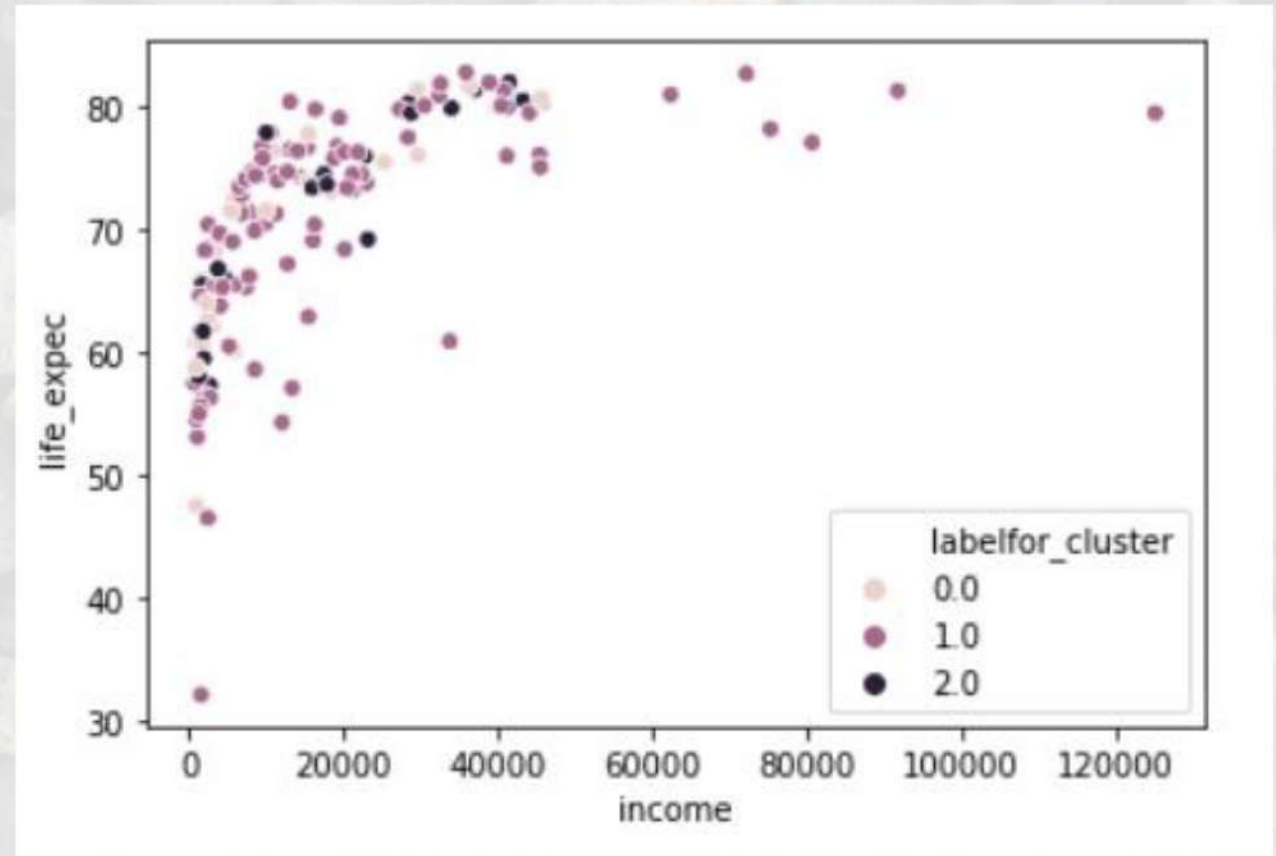


- We found that there is no positive correlation among the income of people and the child mortality rates in these countries.
- By observing the childmortality vs income graph we observe that the income & childmortality are not positively coorelated variables.
- In this the childmortality is low in a country then income of people of country will high and vice versa
- We observe the same pattern in this scatter plot between the income and child mortality rate. As the income is decreasing the child mortality is increasing viceversa.

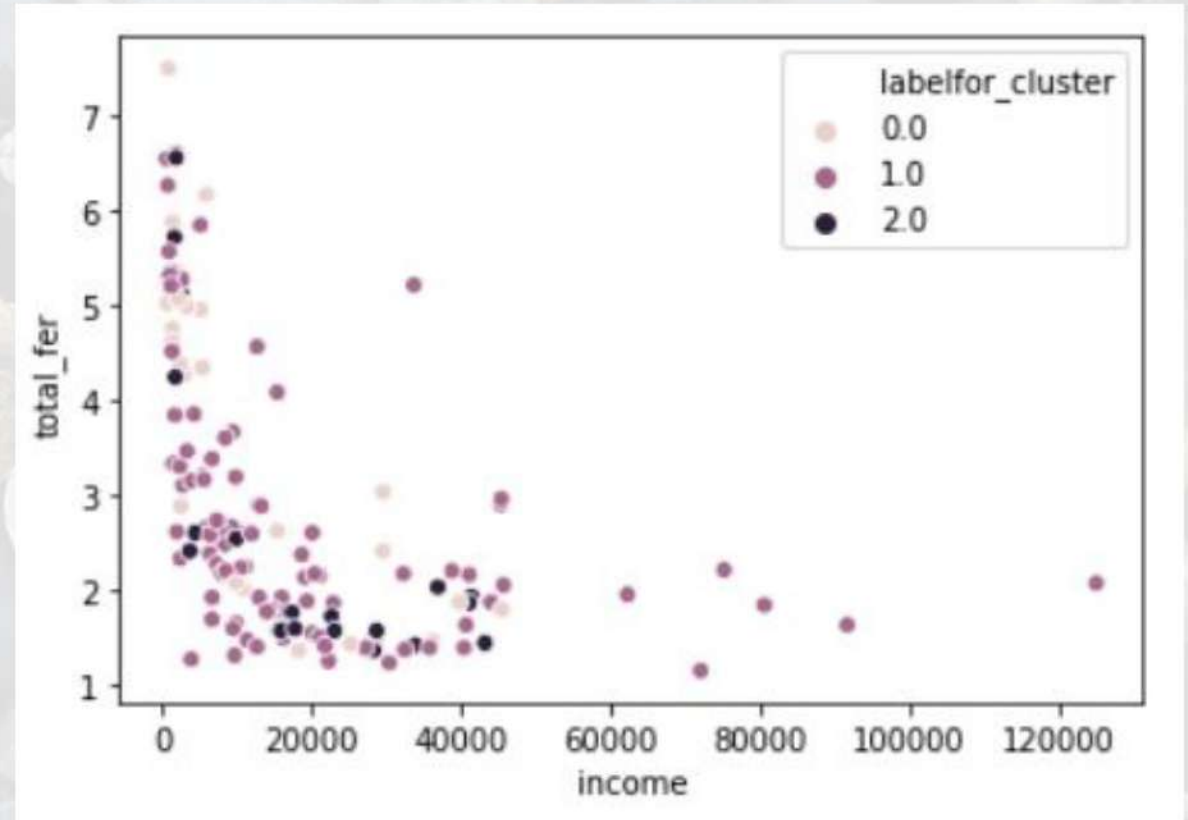




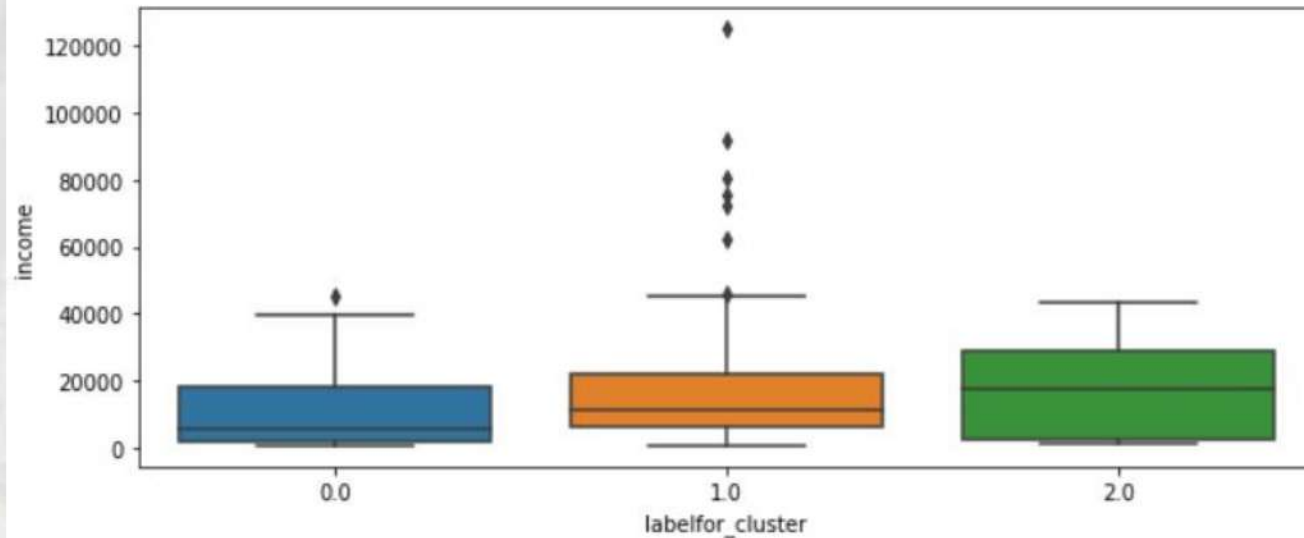
- By observing the graph of life expectancy and income we observe the non linear relation among the income & life expectancy variables.
- As the income is low then the life expectancy is relatively low but when we increase the income from about 10000 then life expectancy will increase significantly and become constant as the income increases.



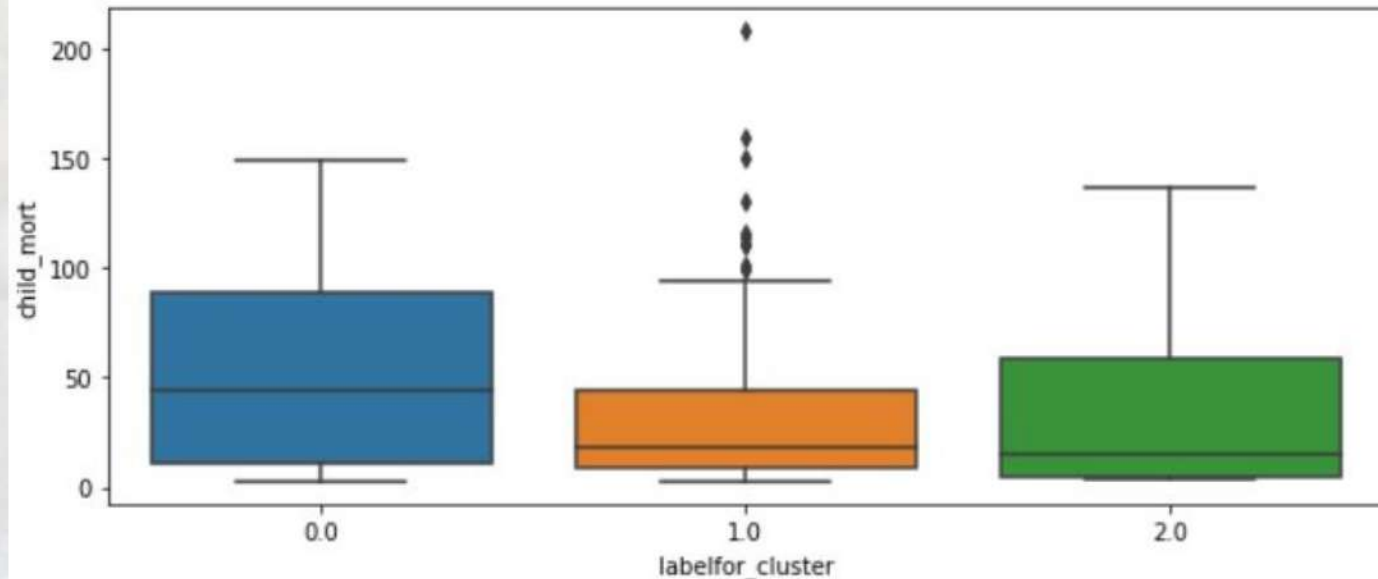
- By observing the graph of total fertility and income we observe that the income & total fertility are not positively coorelated variables.
- In this the total fertility is low in a country then income of people of country will high and vice versa.





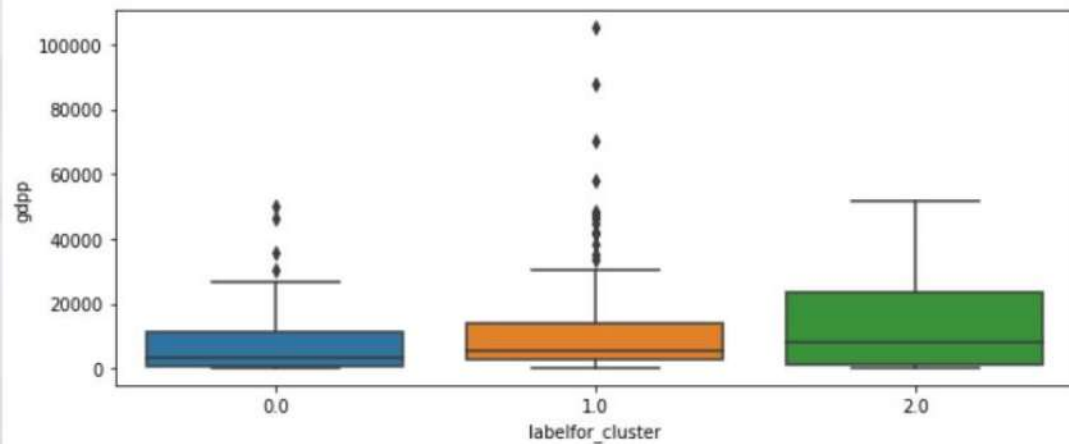


according to the box plot of income vs labelof\_cluster we observe that the avg income for label cluster-2 is comparatively more than other clusters.

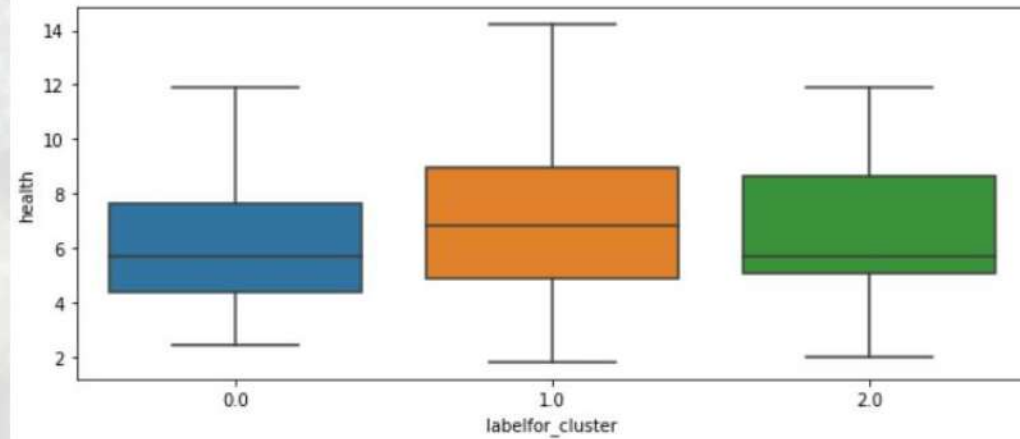


according to the box plot of child mortality vs labelof\_cluster we observe that the avg child mortality for label cluster-0 is comparatively more than other clusters.

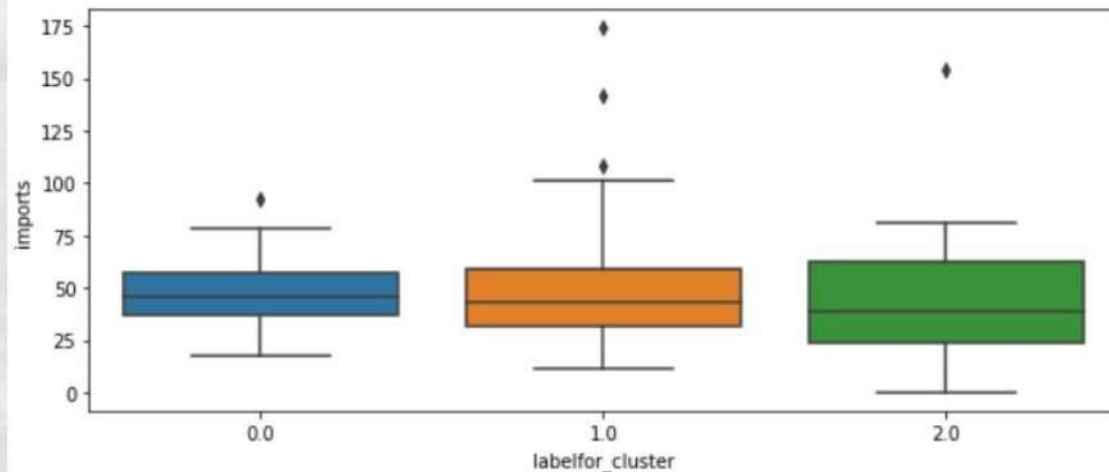




according to the box plot of GDP vs labelof\_cluster we observe that the avg gdp for label cluster-2 is comparatively more than other clusters.



according to the box of health vs labelof\_cluster we observe that plot the avg health for label cluster-2 is comparatively more than other clusters.



according to the box plot of imports vs labelof\_cluster we observe that the avg imports for all label clusters is comparatively same but for label cluster-1 the over all imports are more due to present of outliers.





# LIST OF TOP COUNTRIES THAT NEED AID ARE:

- - Chad (having 150 child mortality)
- - Sierra Leone (having 160 child mortality)
- - Central African Republic (having 149 child mortality)
- - Mali (having 137 child mortality)
- - Nigeria (having 130 child mortality)
- - Niger (having 123 child mortality)
- - Angola (having 119 child mortality)
- - Burkina Faso (having 116 child mortality)
- - Congo demo rep (having 116 child mortality)
- - Benin (having 111 child mortality)
- - Cote d'Ivoire (having 111 child mortality)
- - Guinea (having 109 child mortality)
- - Cameroon (having 108 child mortality)
- - Gambia (having 80 child mortality)
- - Myanmar (having 64 child mortality)

