

By: Kartik Mehra  
[kartikmhr97@gmail.com](mailto:kartikmhr97@gmail.com)  
+91-9878440421

# MOTIVE

PROVIDE LIST OF COUNTRIES THAT ARE IN DIREST NEED OF FINANCIAL SUPPORT SO THAT THEY CAN BE PRIORITIZED BEFORE OTHER COUNTRIES.



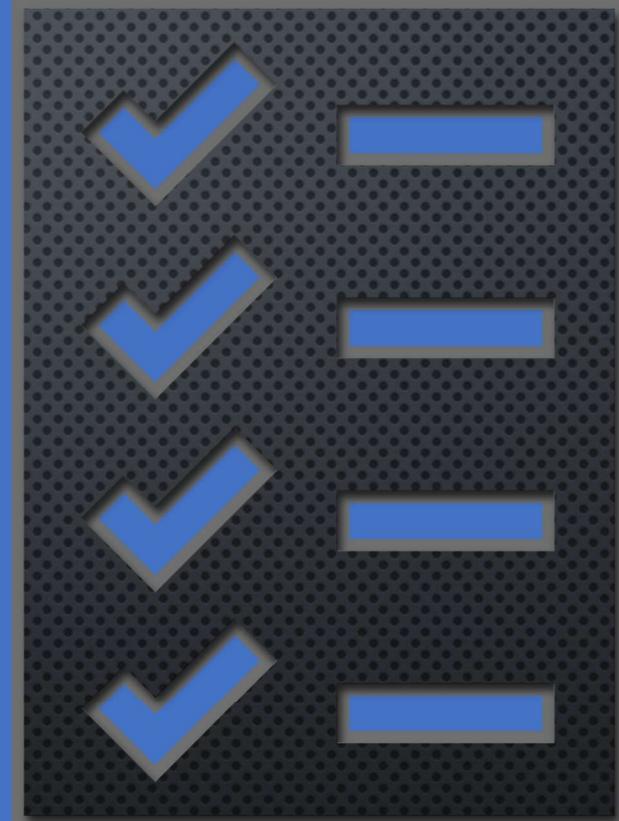
# APPROACH

LET'S DIVE IN



# DATASET SANITY CHECK

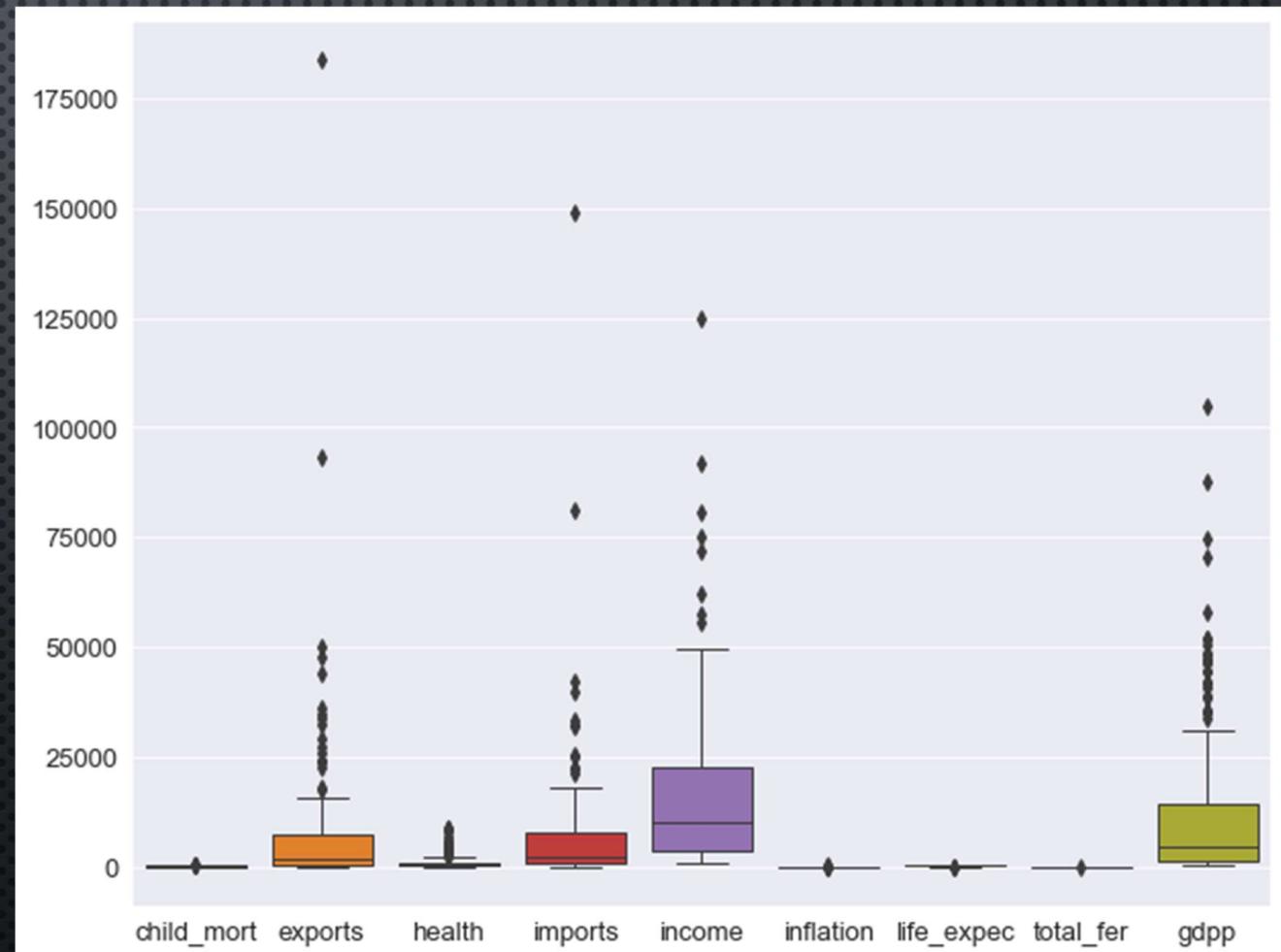
- NULL VALUES VERIFICATION
- DATATYPES VERIFICATION
- CONVERSION OF % VALUES TO  
ABSOLUTE VALUES( EXPORTS,  
INCOME, HEALTH )



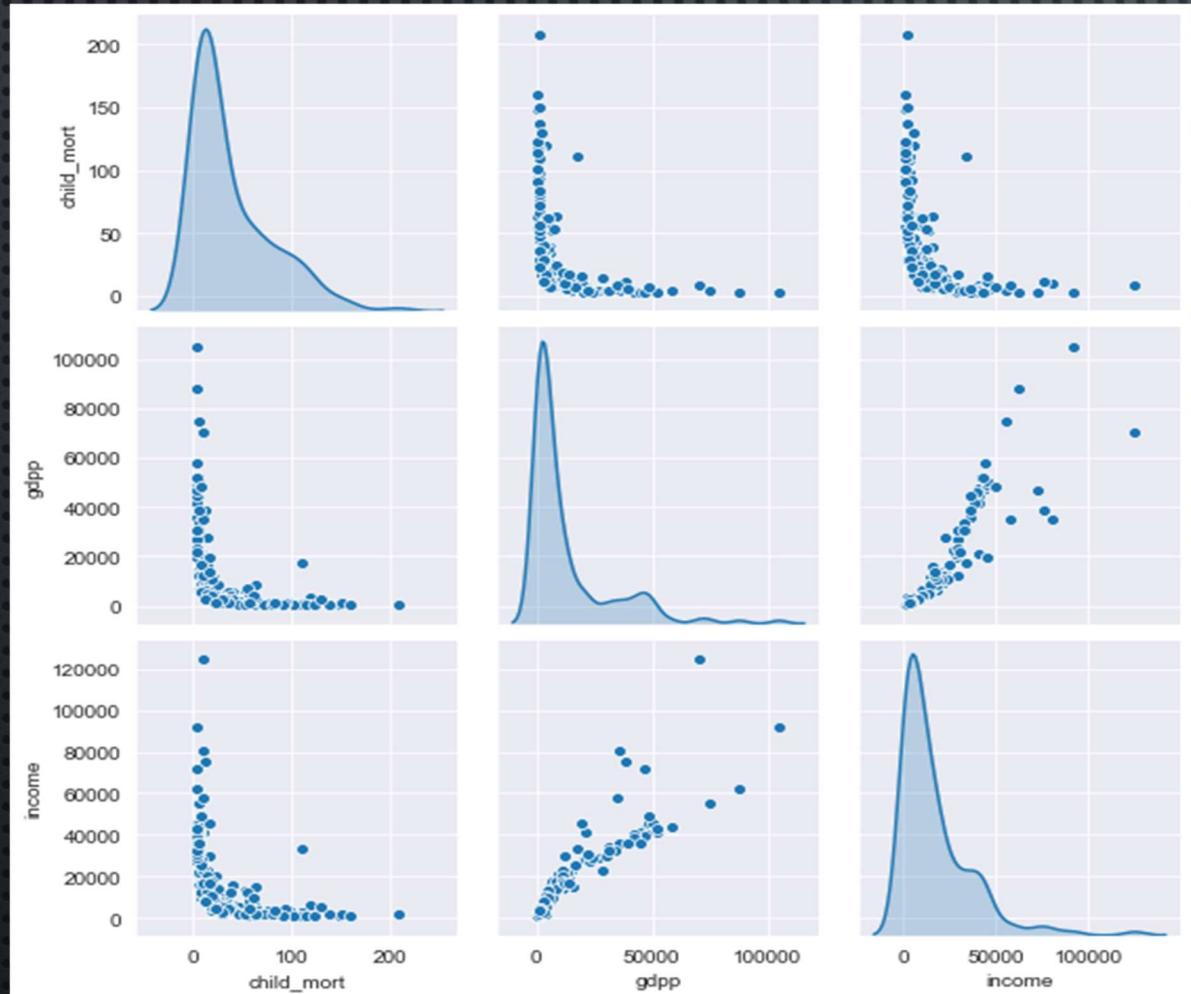
# OUTLIERS( EXTREME HIGH/LOW VALUES)

Here we can see that there are some outliers present in the data provided for exports, gdpp, income & imports.

But after analyzing those values, they seemed to be reasonable as highly developed country can have extremely greater value as compared to the under-developed countries.



# SCALING THE DATA

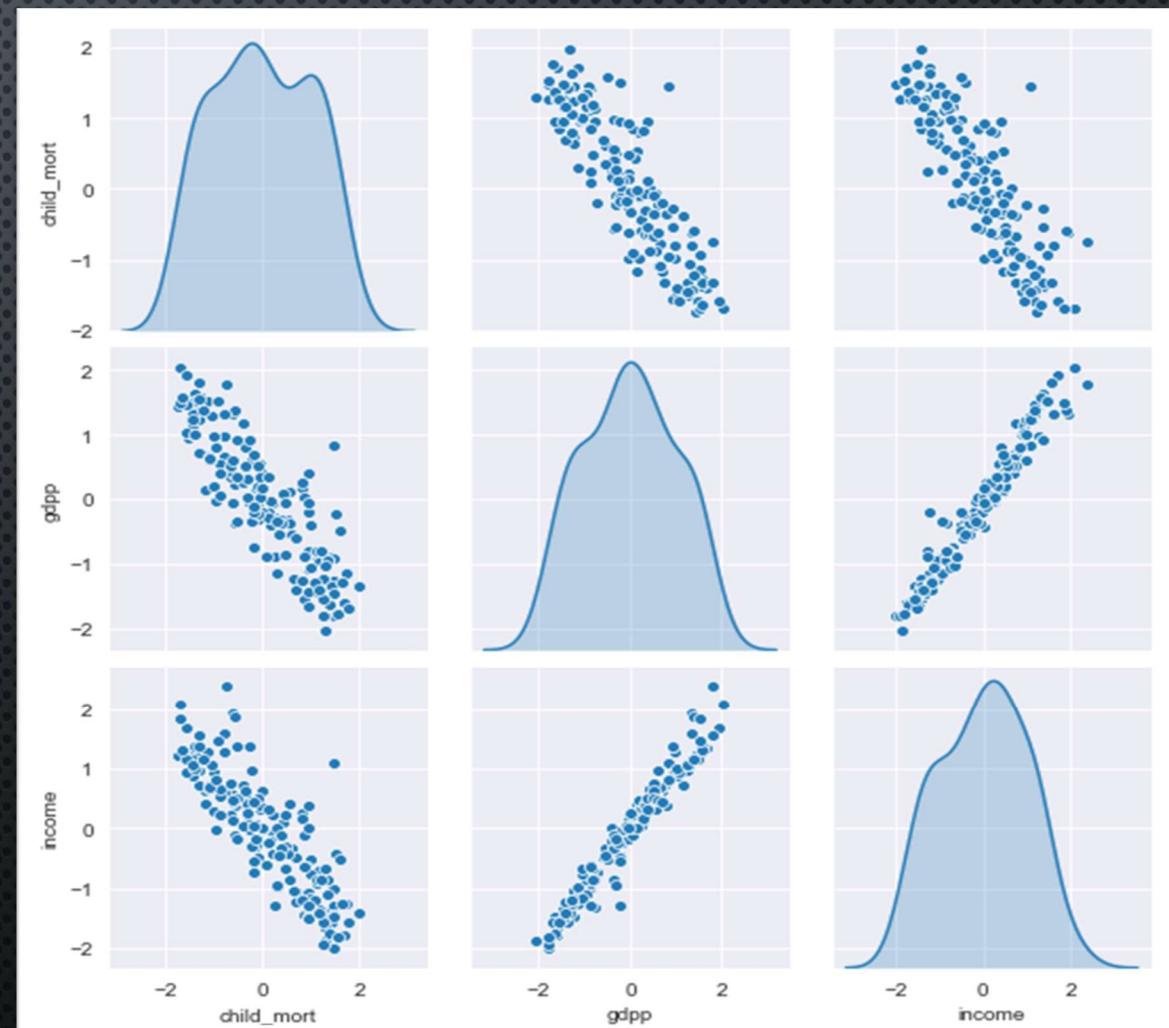
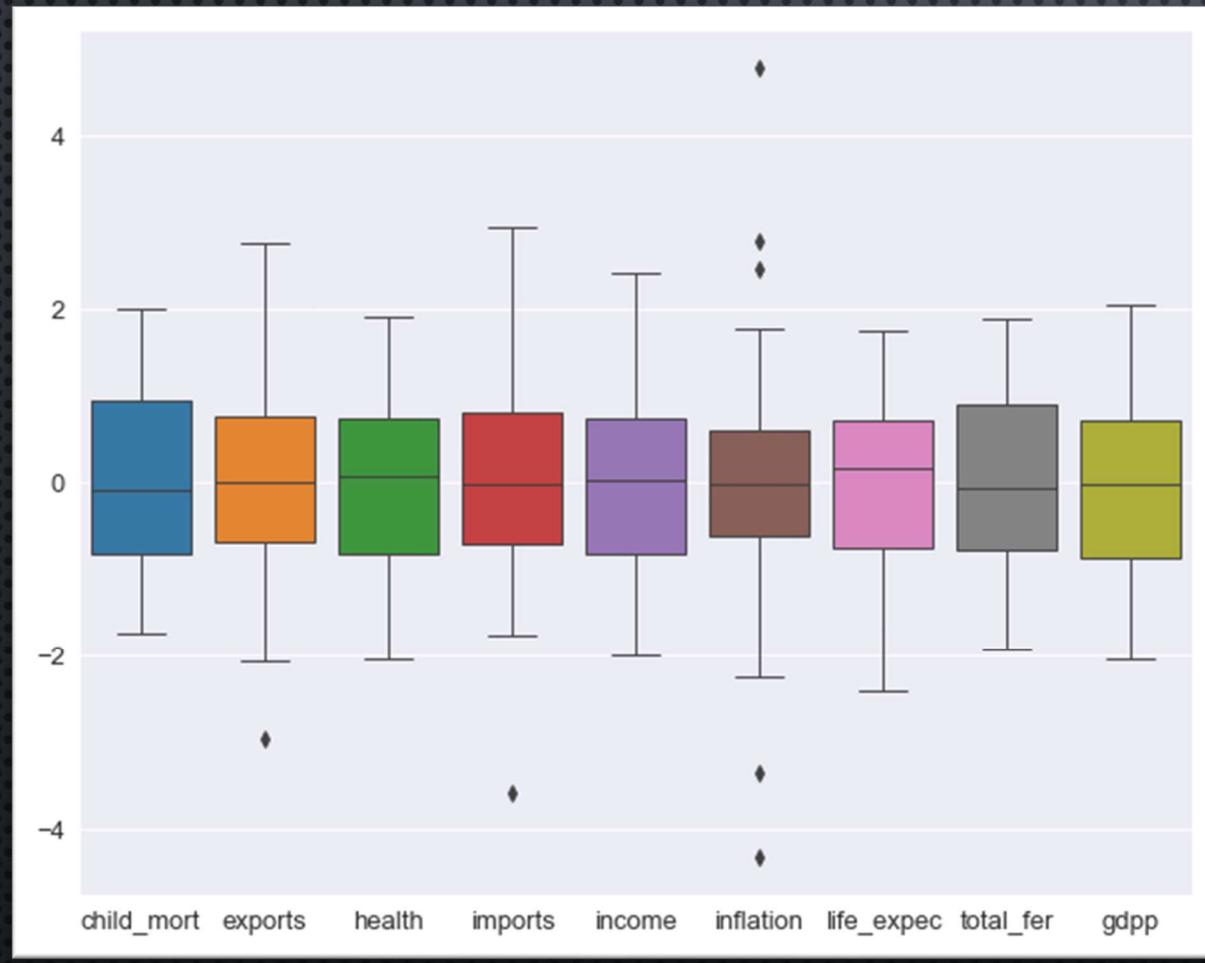


There was a lot of variance in our data and the data was skewed so to make it more normally distributed power transformation was used.

After performing the operation the correlation between variables became clear.

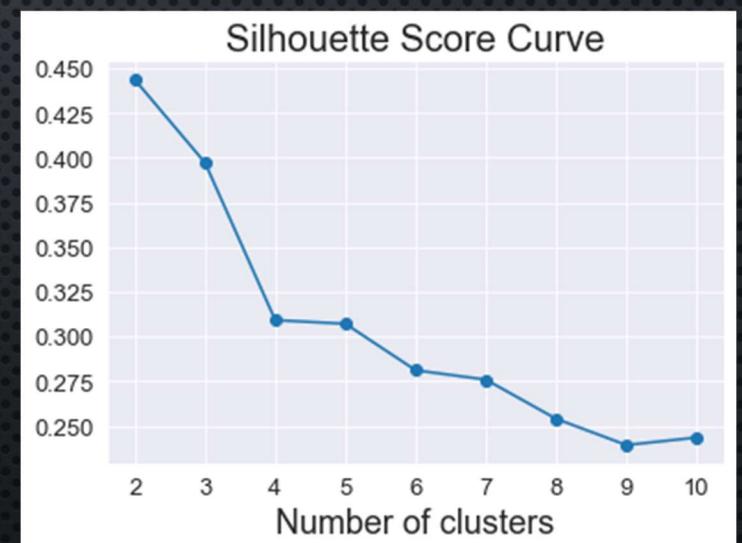
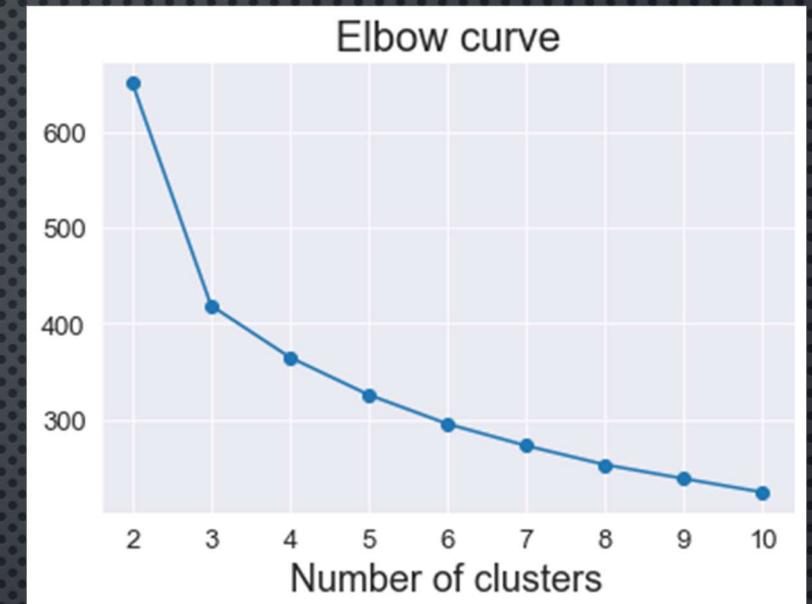
# AFTER TRANSFORMATION

Correlation became clear and outliers reduced

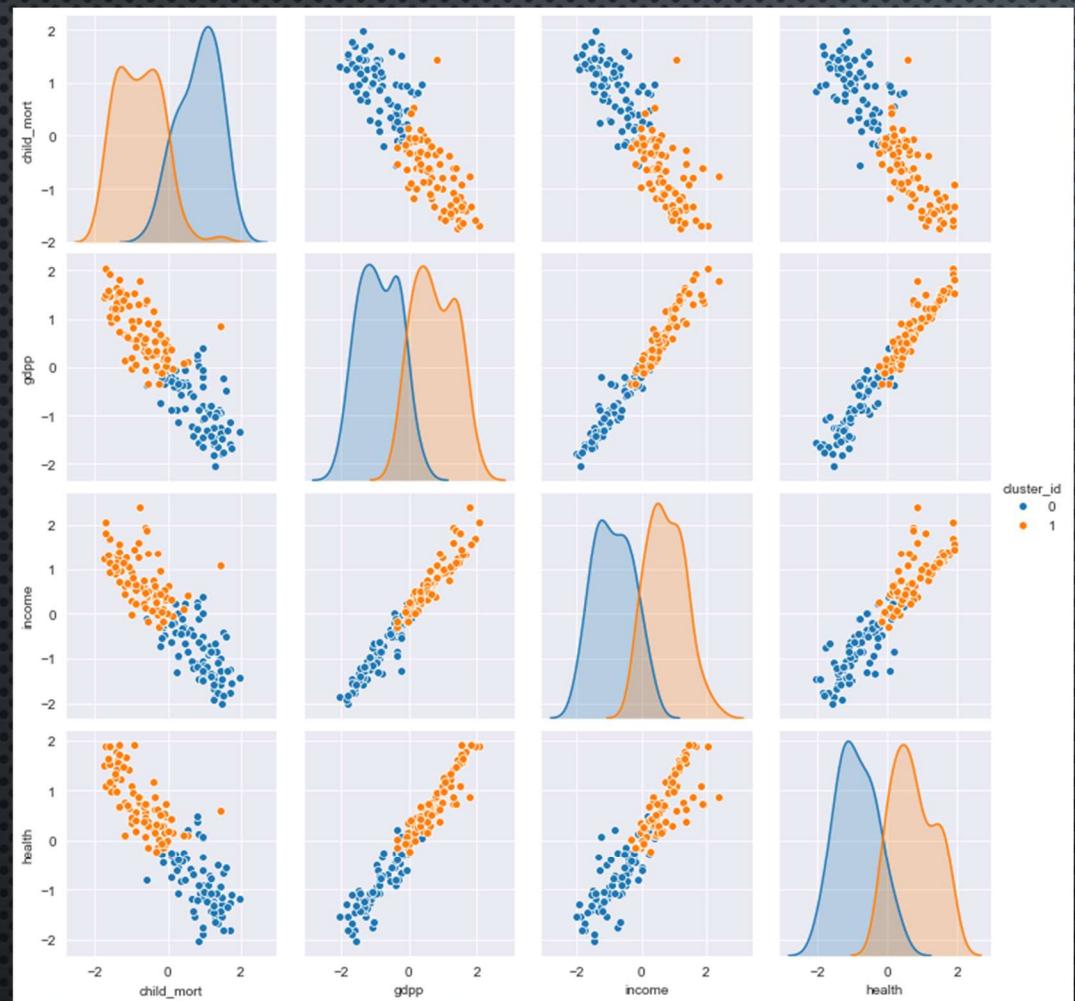


# K-MEANS CLUSTERING

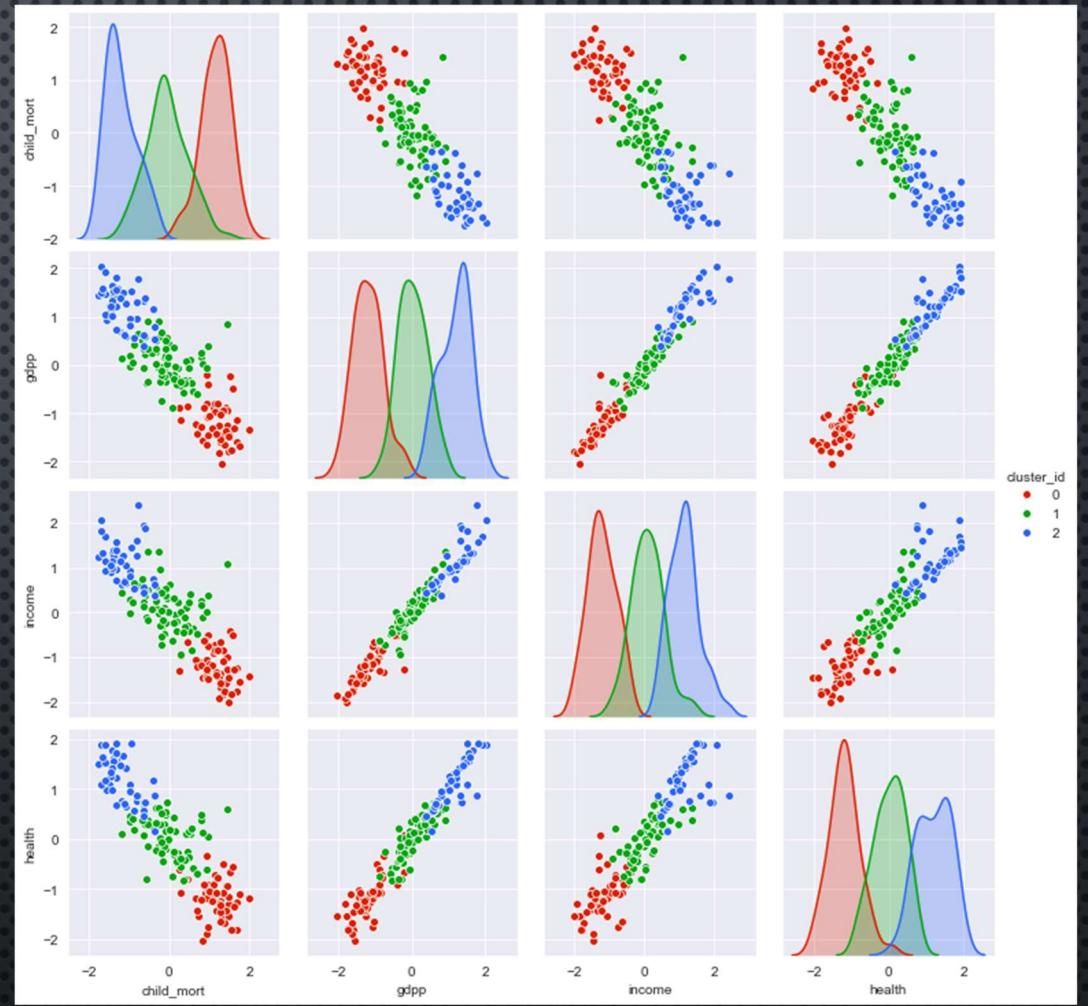
AFTER ANALYZING ELBOW CURVE AND AVERAGE SILHOUETTE CURVE THEIR OPTIMAL NUMBER OF CLUSTERS CAN BE 2 OR 3.



# RESULTS

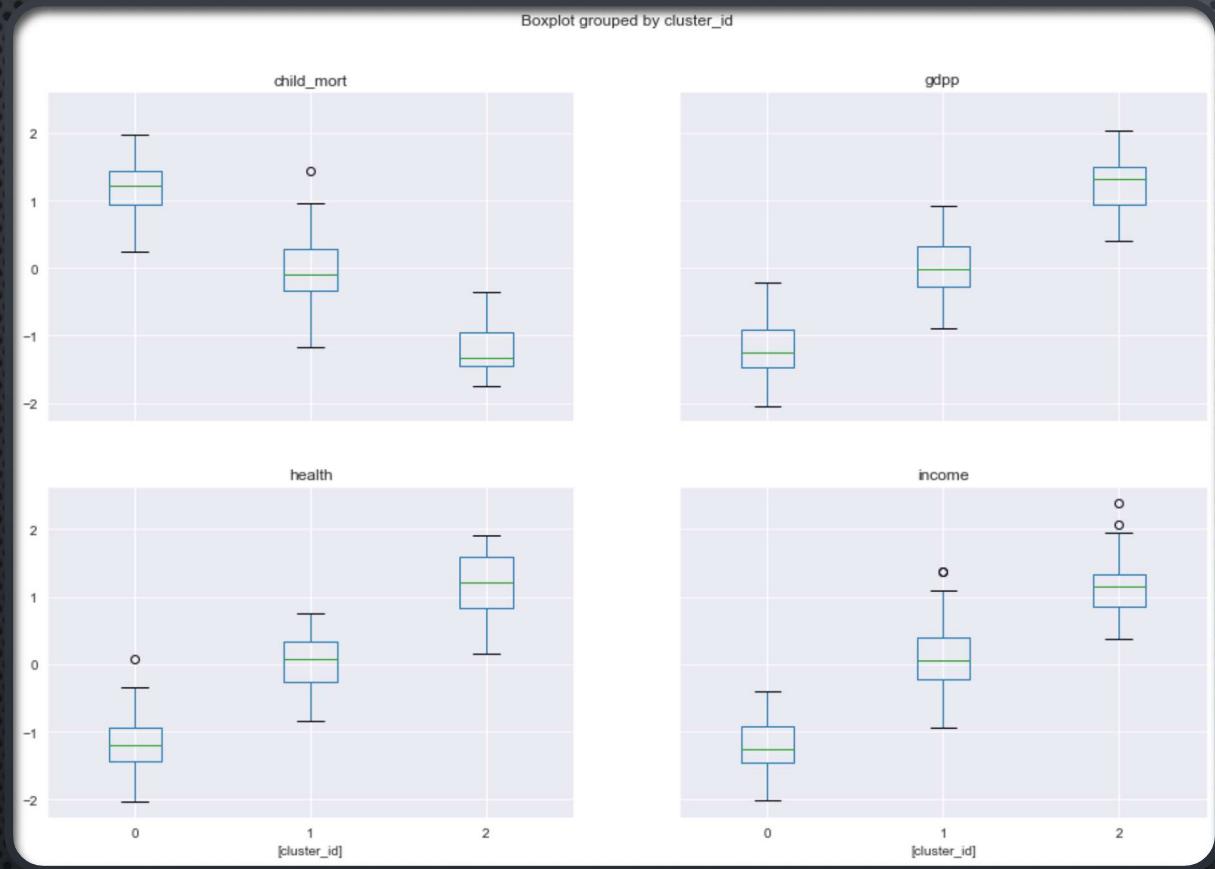


2 groups created



3 groups created

# BOX PLOT ( K-MEANS )



Note: cluster\_id represents group number

We used 3 cluster model as it also gave well grouped data like 2 cluster model but with more clusters and according to the business, countries are labeled as under-developed, developing or developed.

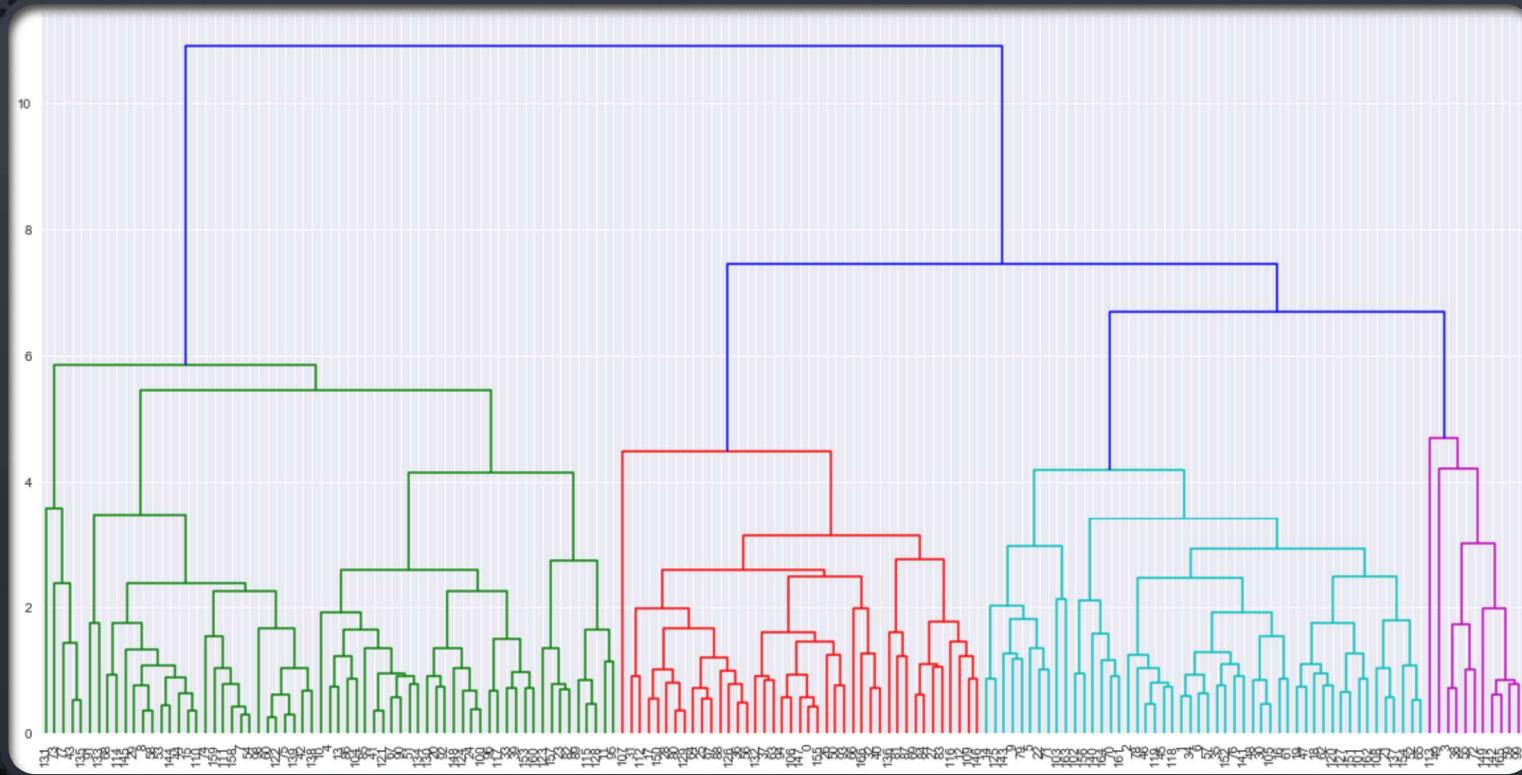
From the attached plot we can see that group 0 has the most child death rate as compared to other groups and has least gdp, income & health conditions.

- 0 as **Under-Developed**
- 1 as **Developing**
- 2 as **Developed**

Below are the 50 countries categorized as under-developed :

- Afghanistan
- Angola
- Bangladesh
- Benin
- Burkina Faso
- Burundi
- Cambodia
- Cameroon
- Central African Republic
- Chad
- Comoros
- Congo, Dem. Rep.
- Congo, Rep.
- Cote d'Ivoire
- Eritrea
- Gambia
- Ghana
- Guinea
- Guinea-Bissau
- Haiti
- India
- Kenya
- Kiribati
- Kyrgyz Republic
- Lao
- Lesotho
- Liberia
- Madagascar
- Malawi
- Mali
- Mauritania
- Mozambique
- Myanmar
- Nepal
- Niger
- Nigeria
- Pakistan
- Rwanda
- Senegal
- Sierra Leone
- Solomon Islands
- Sudan
- Tajikistan
- Tanzania
- Timor-Leste
- Togo
- Uganda
- Uzbekistan
- Yemen
- Zambia

# HIERARCHICAL CLUSTERING

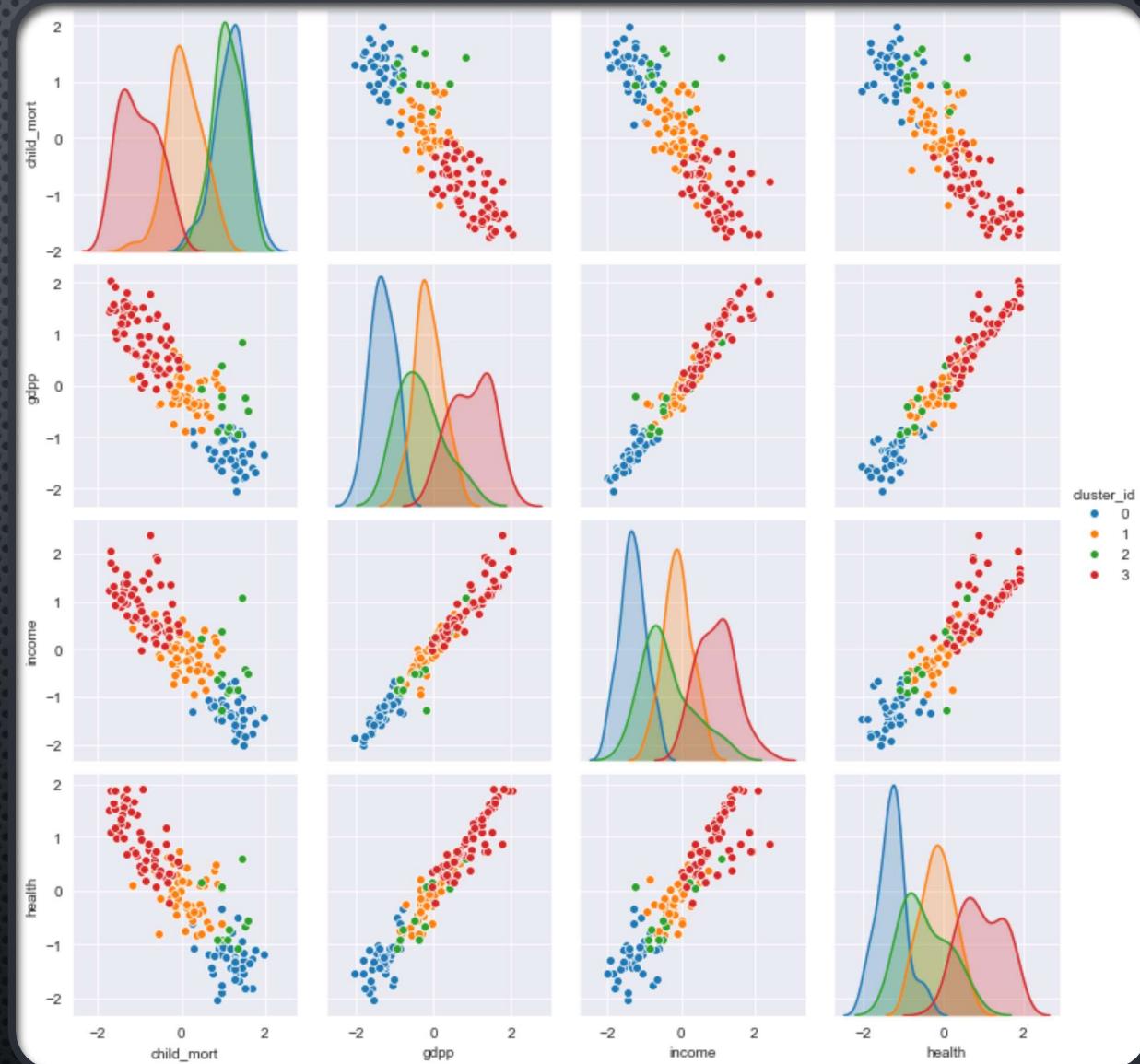


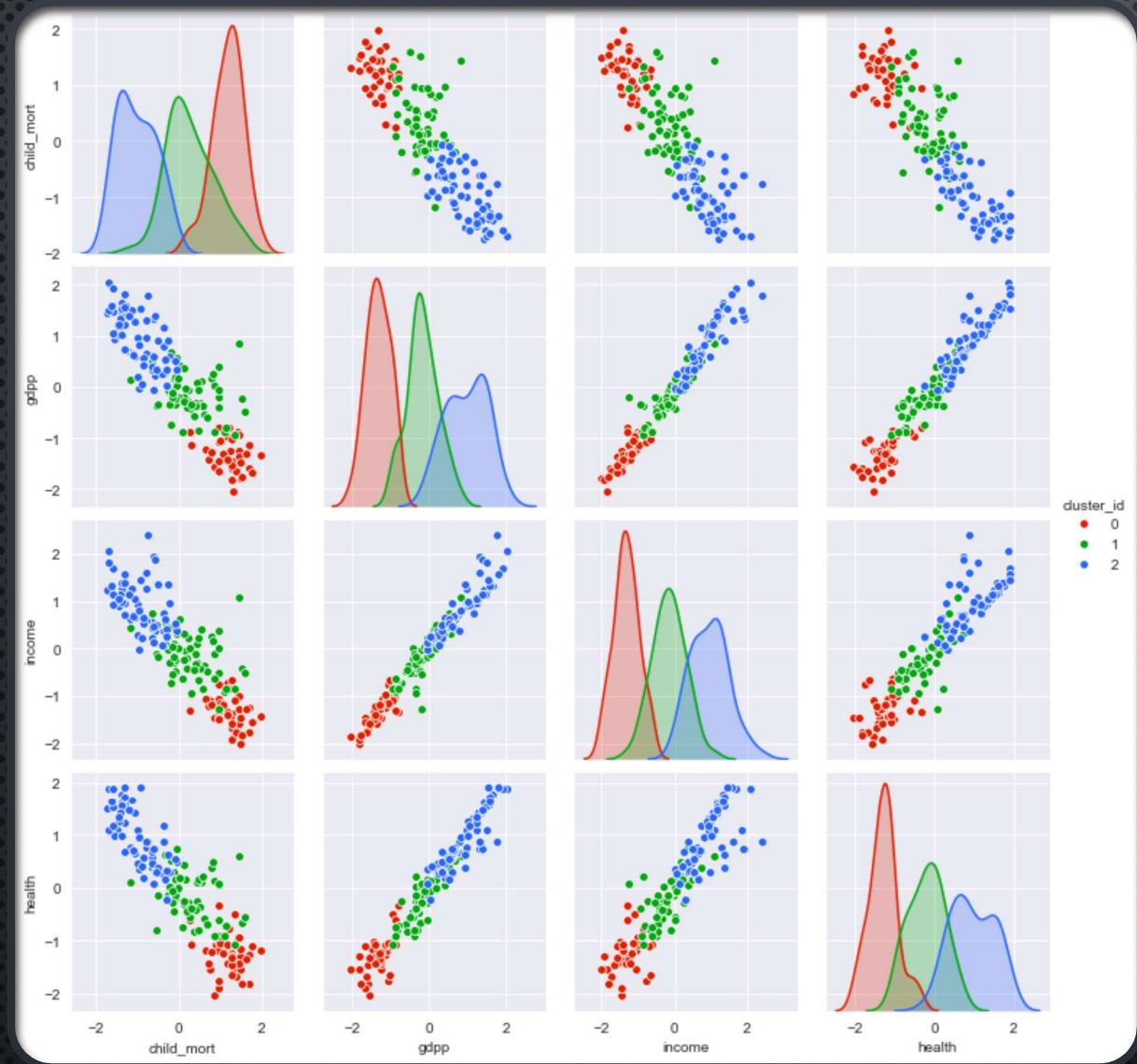
We used complete linkage so get clusters which uses maximum distance between 2 points.

We can see 4 clusters being formed marked in different colors so we will start with 4 and reduce the number until we get a better grouping.

# 4 CLUSTERS

Group 2 is merging into other groups and there are total 11 countries that all under this group next we will use 3 as total number of groups.





## 3 CLUSTERS

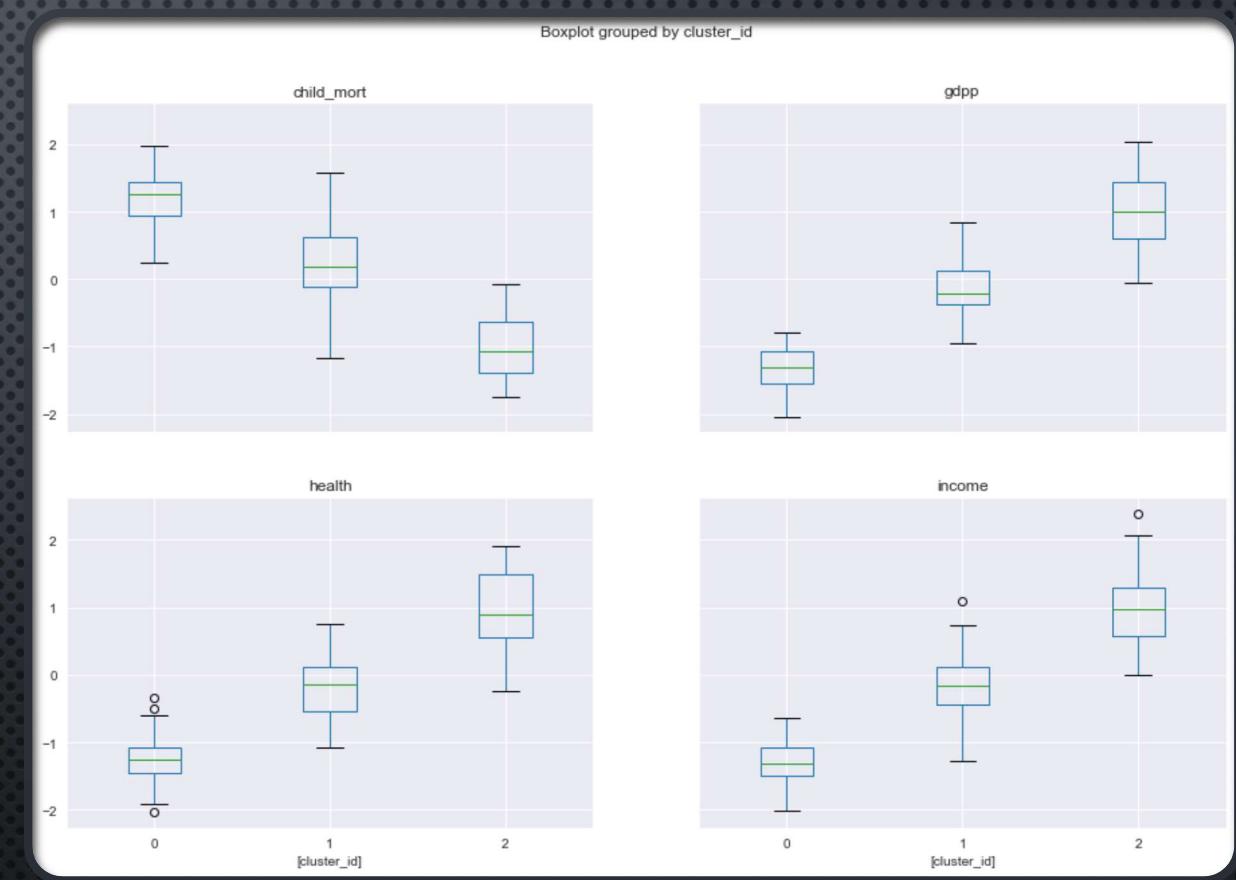
3 clusters looked good as compared to 4 clusters because on removing one group the clusters become distinct and most of the overlapping of datapoints got resolved.

# BOX PLOT ( HIERARCHICAL)

Like K-means result, all the groups follow the same pattern.

So according to the business group numbers are mapped in the following manner:

- 0 as Under-Developed
- 1 as Developing
- 2 as Developed



Note: cluster\_id represents group number

Hierarchical clusters were used to determine groups because variables are well organized like under-developed group have low gdpp, low income & high child\_mort as compared to the same group in k-means.

Below are the 41 countries categorized as under-developed :

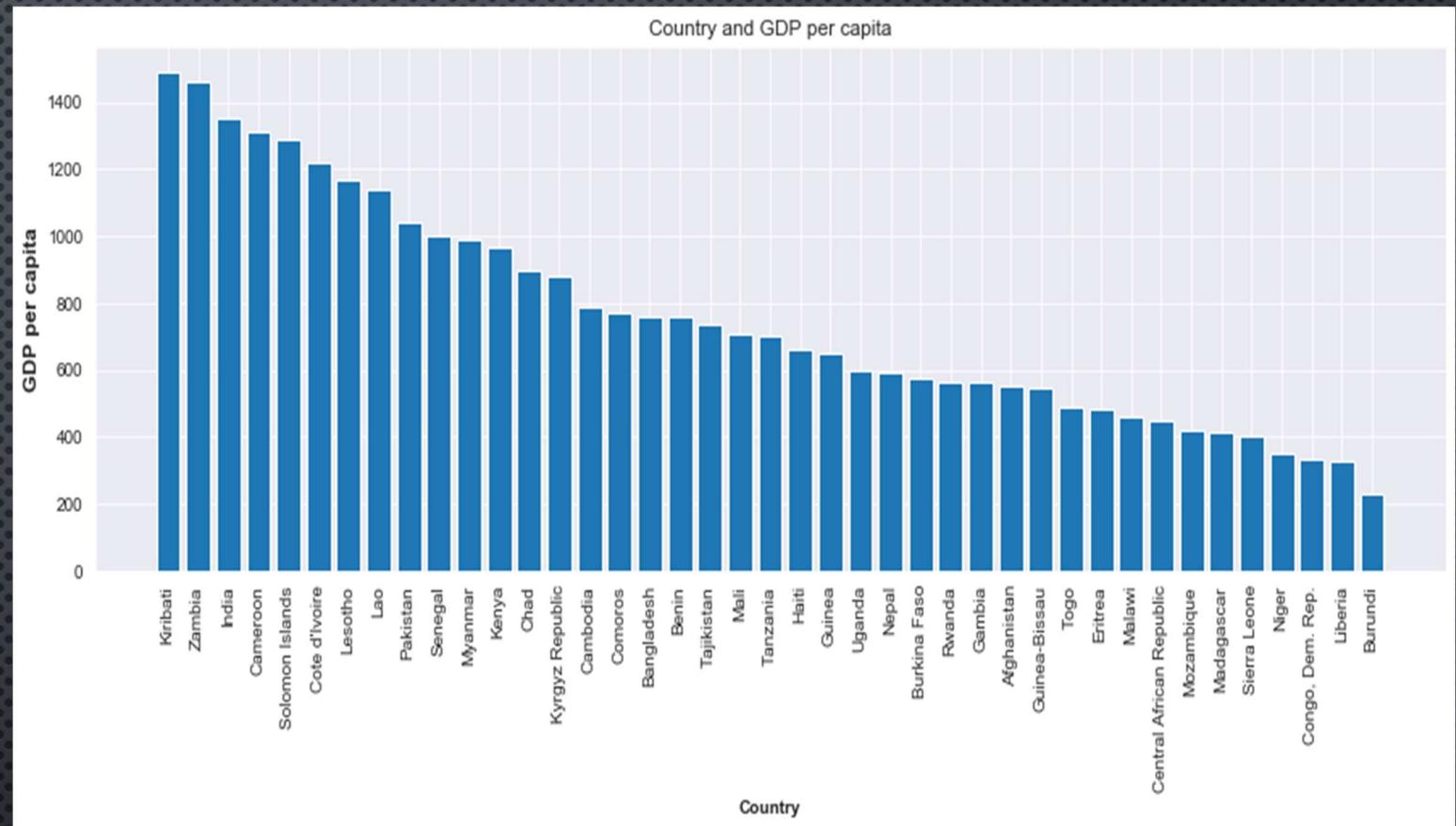
- Afghanistan
- Bangladesh
- Benin
- Burkina Faso
- Burundi
- Cambodia
- Cameroon
- Central African Republic
- Chad
- Comoros
- Congo, Dem. Rep.
- Cote d'Ivoire
- Eritrea
- Gambia
- Guinea
- Guinea-Bissau
- Haiti
- India
- Kenya
- Kiribati
- Kyrgyz Republic
- Lao
- Lesotho
- Liberia
- Madagascar
- Malawi
- Mali
- Mozambique
- Myanmar
- Nepal
- Niger
- Pakistan
- Rwanda
- Senegal
- Sierra Leone
- Solomon Islands
- Tajikistan
- Tanzania
- Togo
- Uganda
- Zambia

# CONCLUSION

Providing help to the countries with low GDP is important because it plays an important role in the development and the well-being of a country.

Here is the list of top 10 countries with least GDP from under-developed category.

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



KARTIK MEHRA  
KARTIKMHR97@GMAIL.COM



# THANK YOU