



CLUSTERING ASSIGNMENT: PART I

By: Simranjeet Singh
E-mail: sjsingh1809@gmail.com

PROBLEM STATEMENT

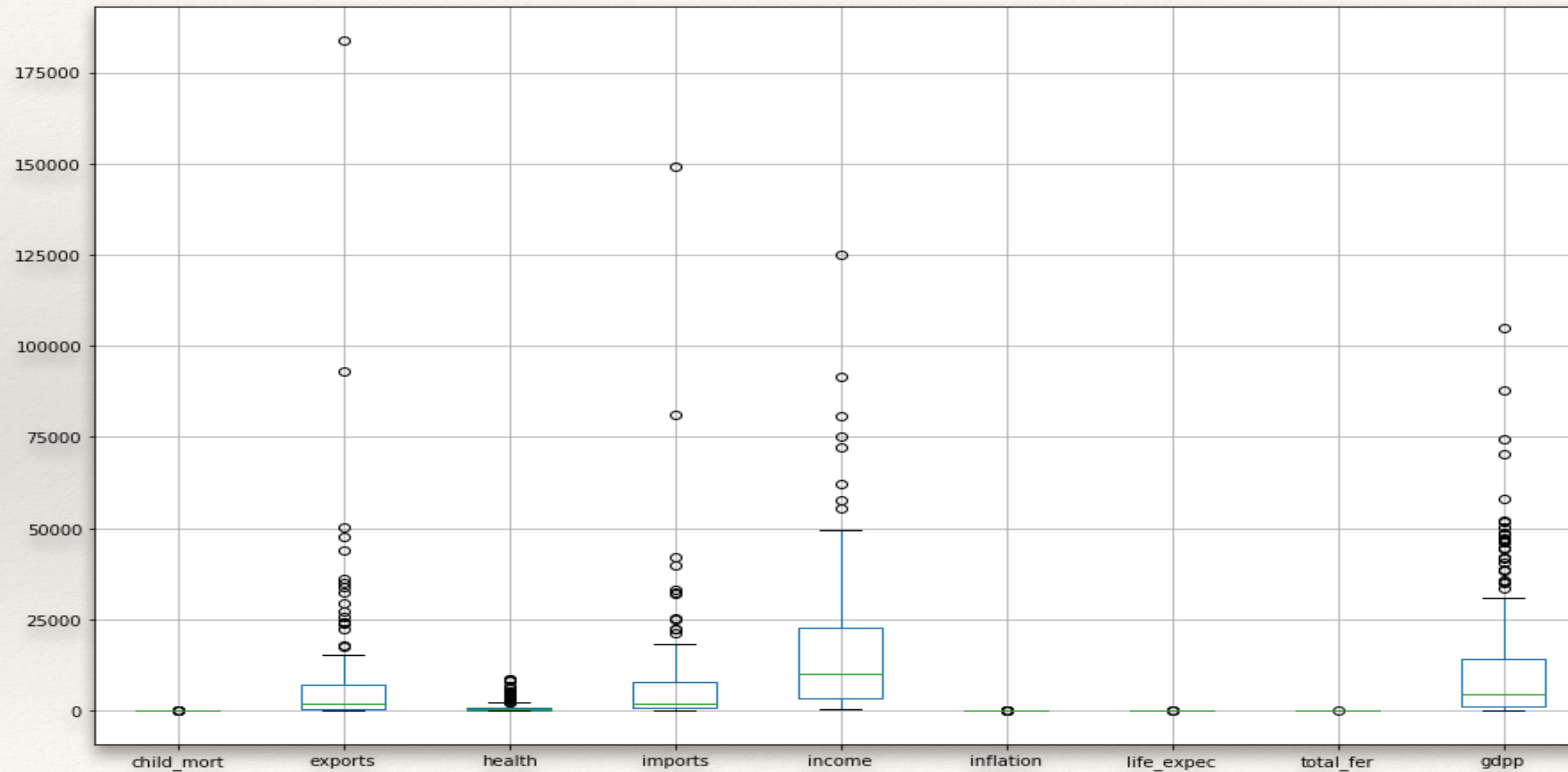
- ❖ HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy 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.

ANALYSIS APPROACH

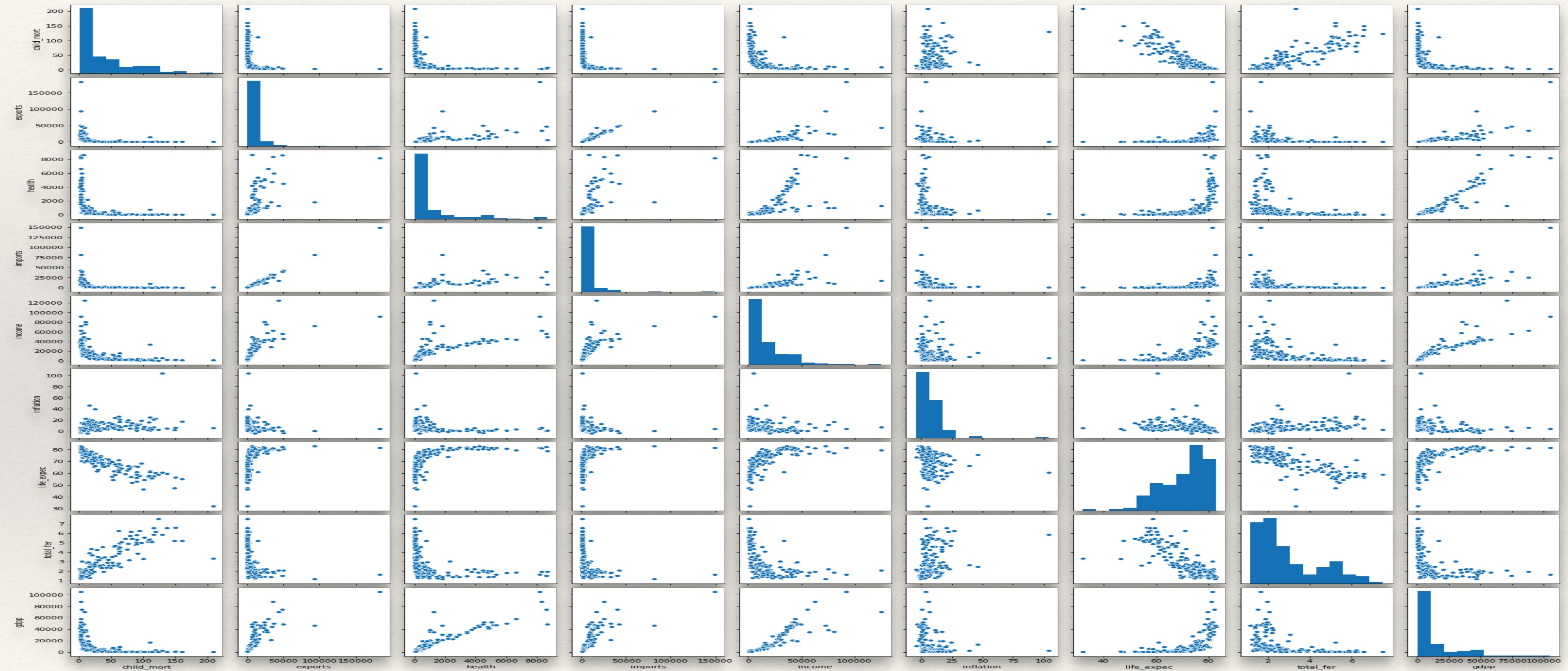
PERFORMING THE BASIC CHECK

- ❖ Data set contains 167 rows and 10 columns.
- ❖ There is no NULL values present in the data set.
- ❖ Converting the % columns values into absolute values. Ex: exports, imports and health.

CHECKING THE OUTLIER's



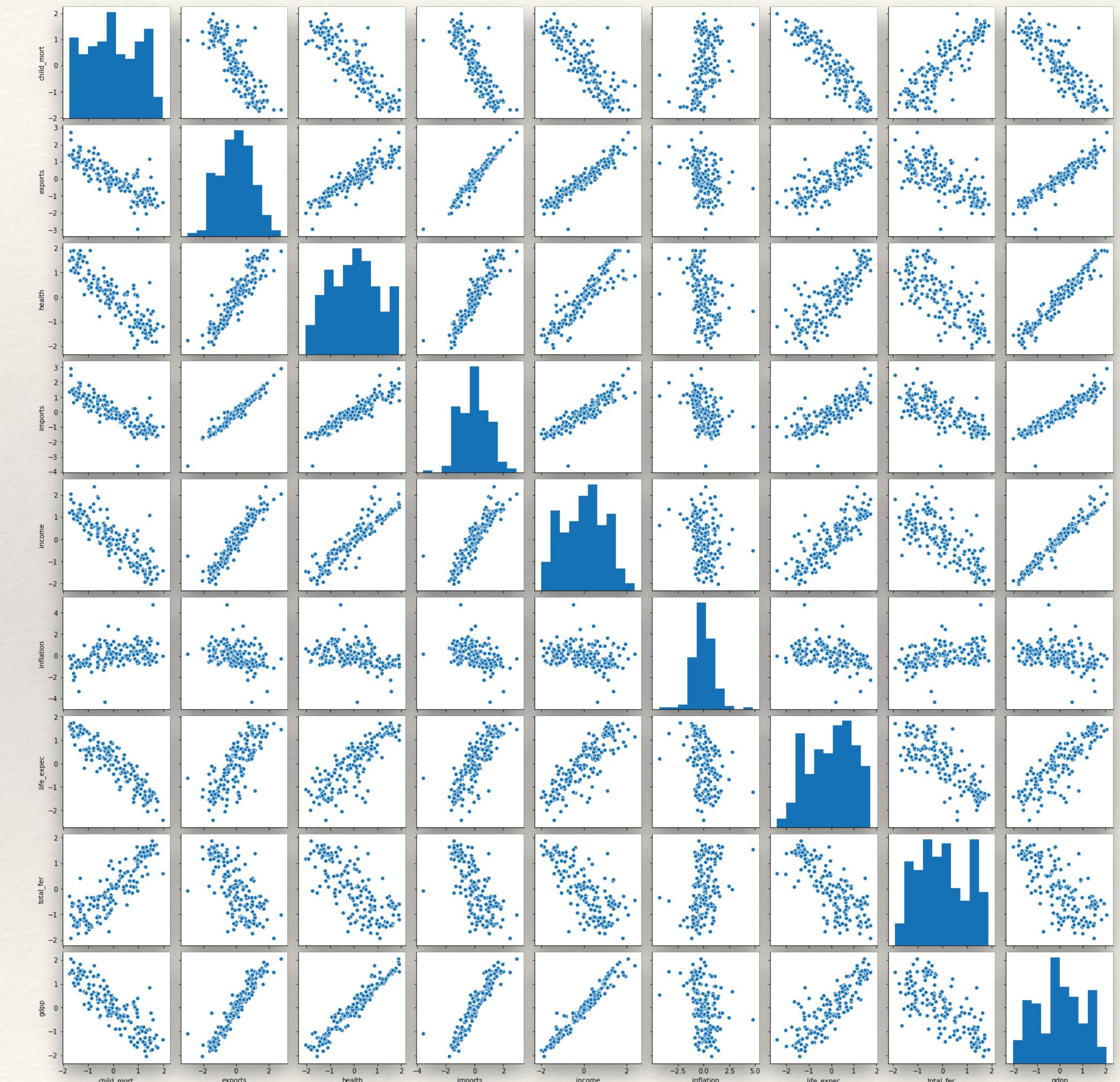
CHECKING THE PAIRPLOT



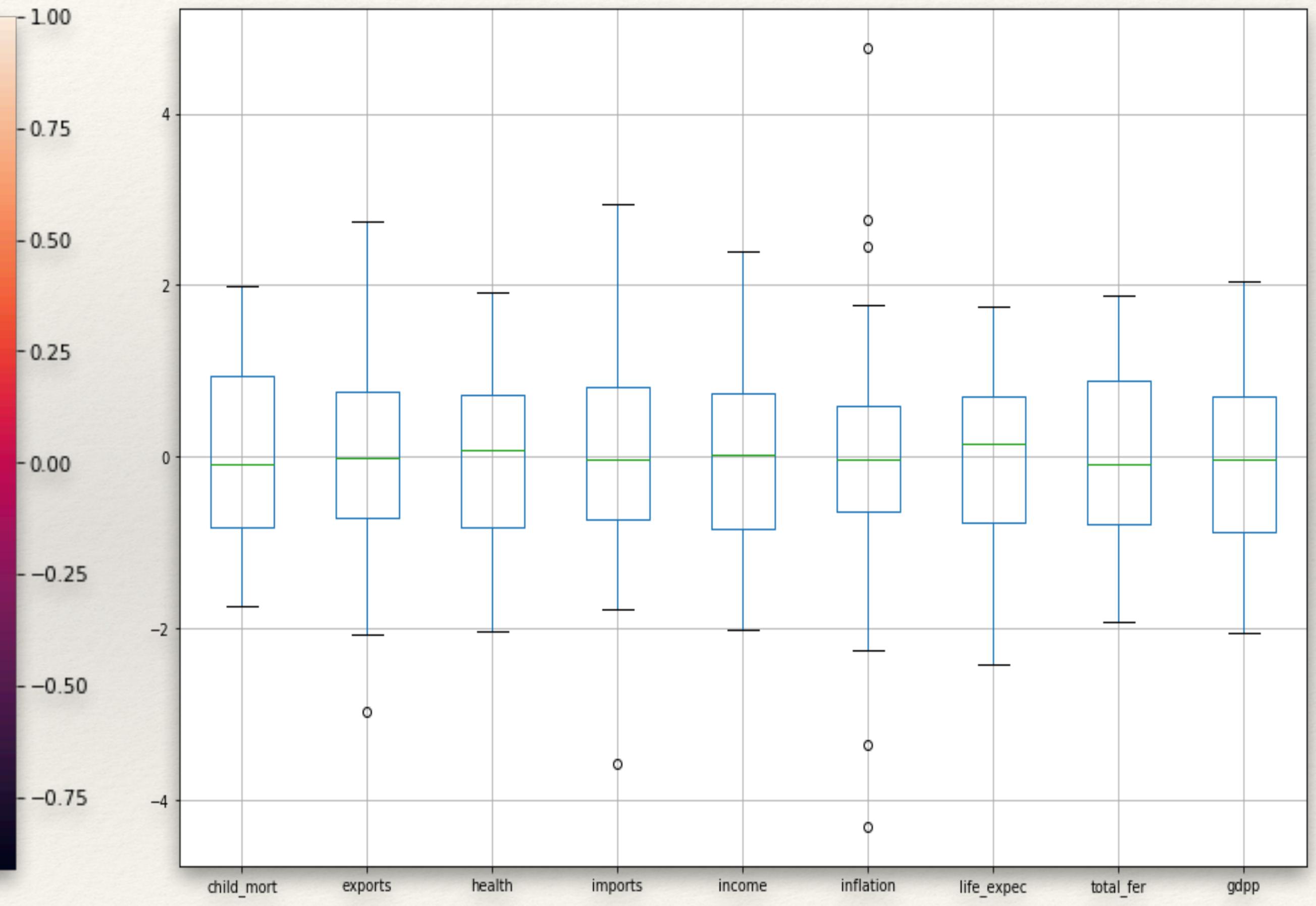
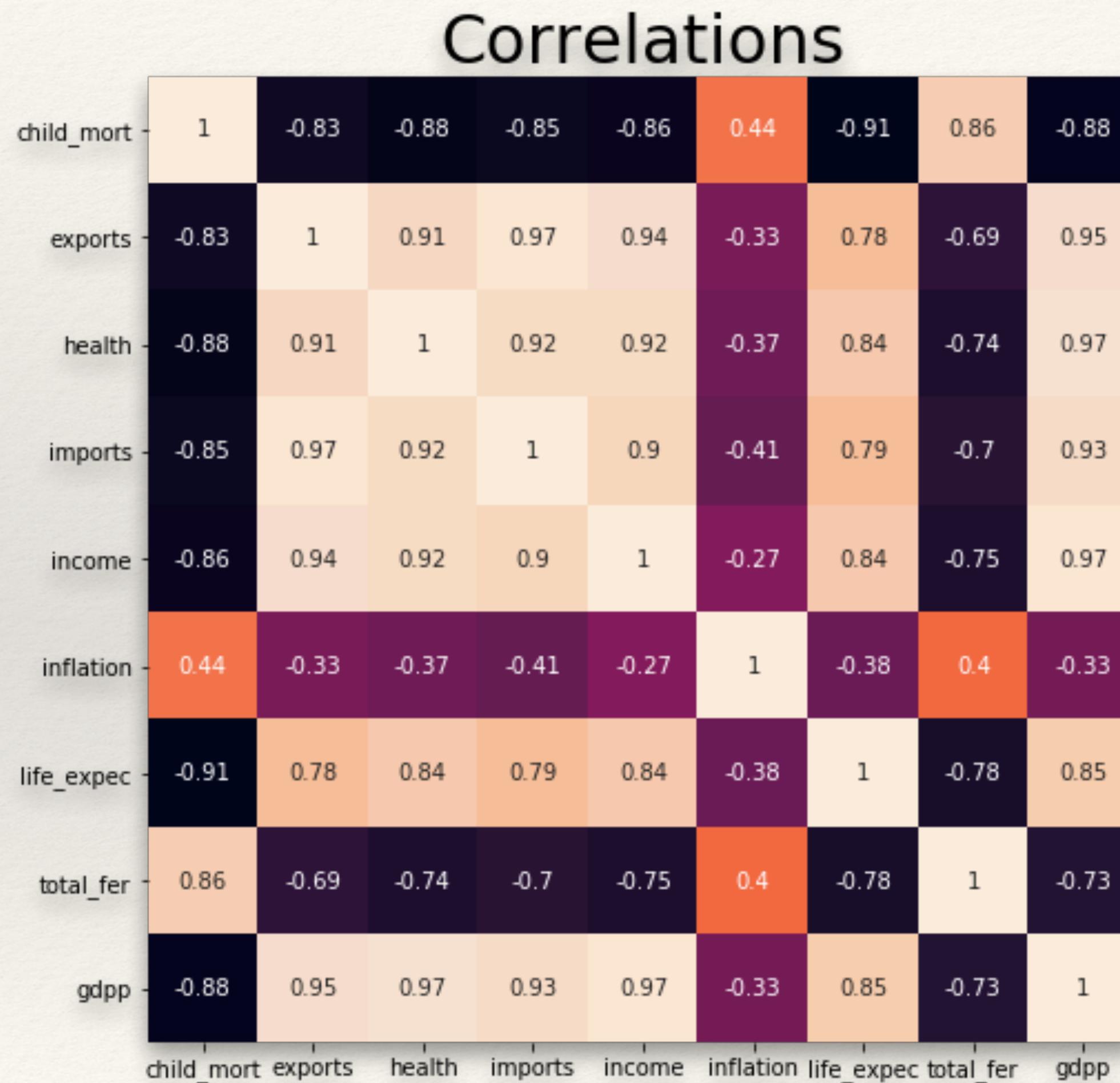
Data is either left-skewed or right-skewed. So we have done power transformation

POWER TRANSFORMATION

- ❖ In the previous ppt we saw that data is left-skewed or right-skewed.
- ❖ So, we have done the power transformation to normalise the data set.
- ❖ Now the data is more clear.



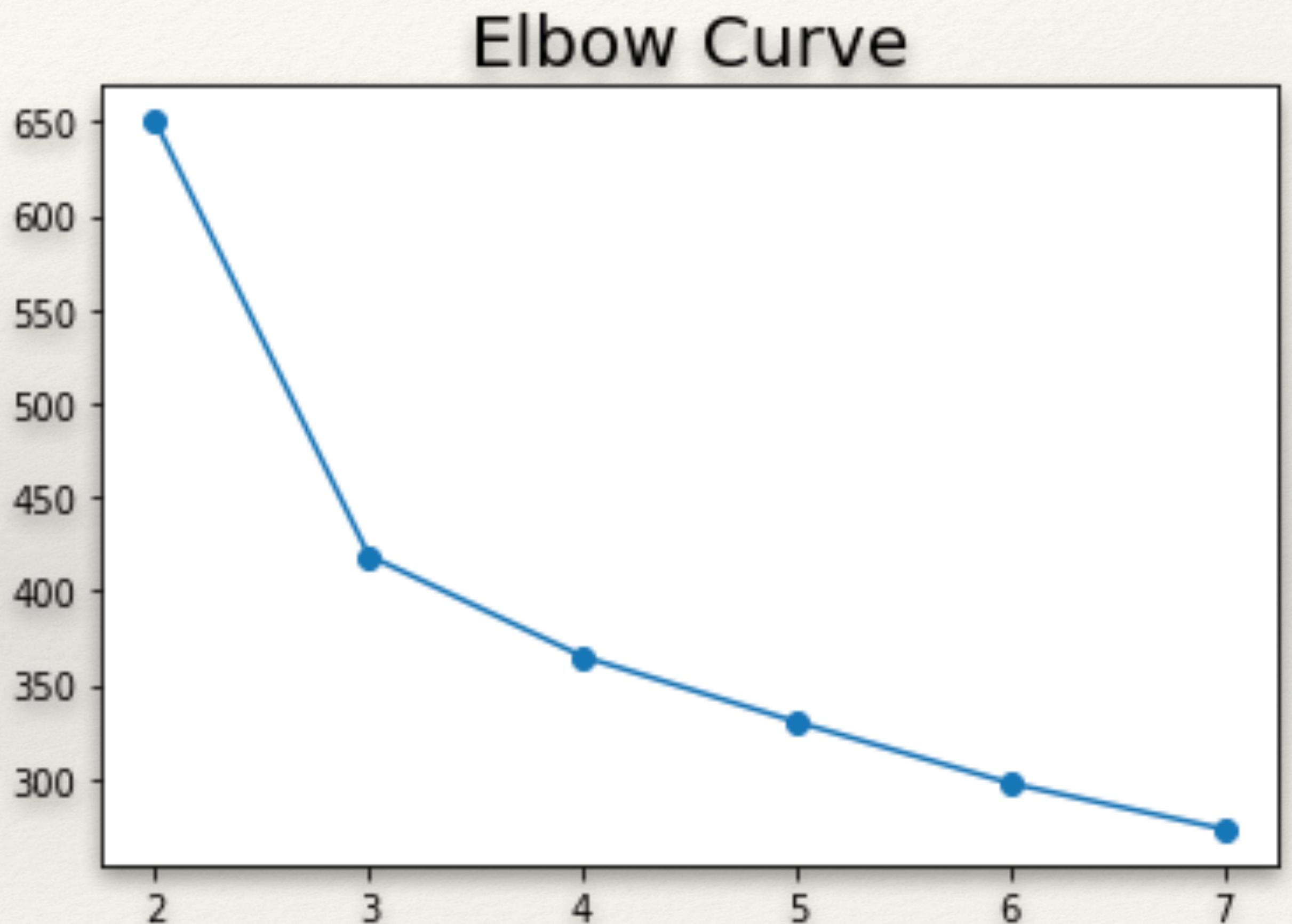
CHECKING THE CORRELATIONS AND BOXPLOTS



K-MEANS CLUSTERING

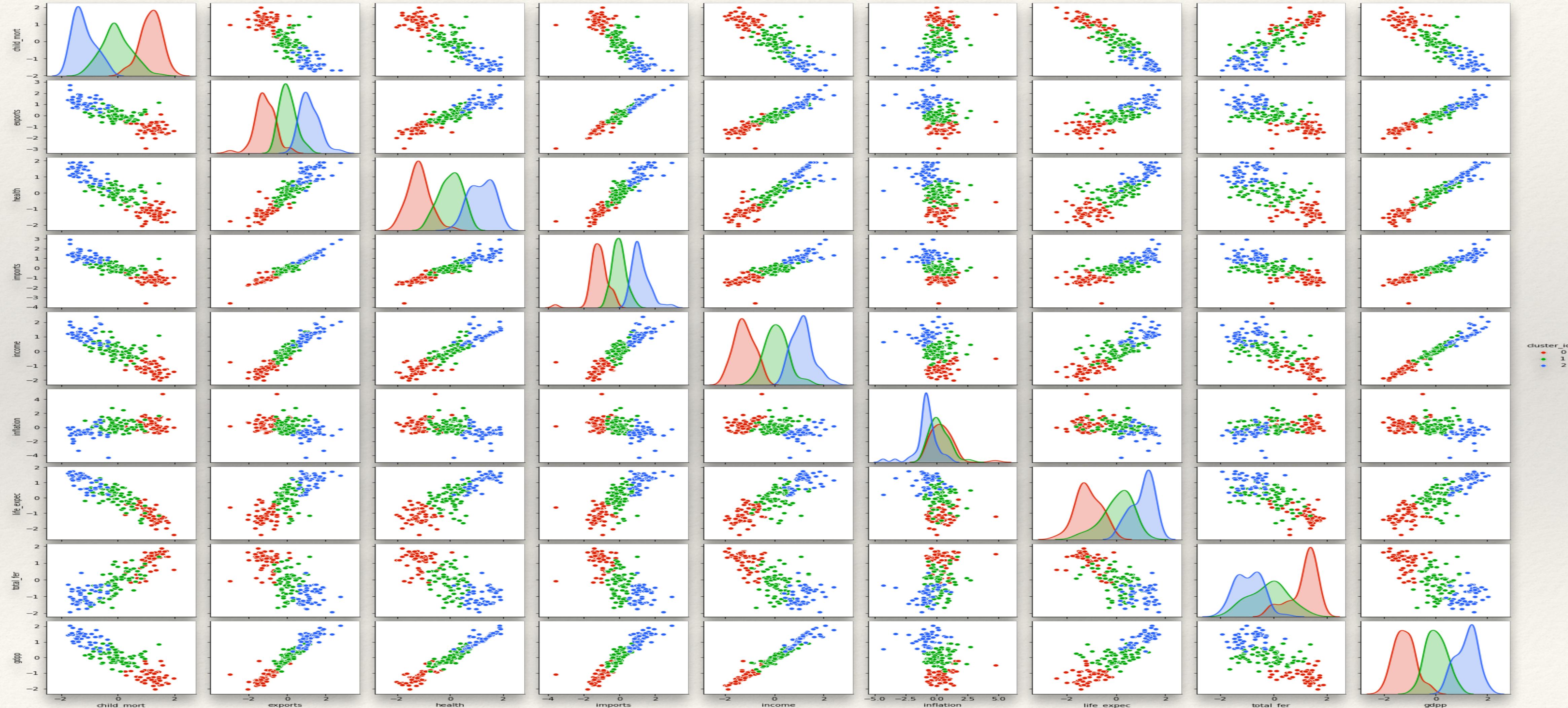
CHECKING THE OPTIMAL VALUE OF ‘K’

- ❖ Silhouette Analysis:
 - ❖ For n_clusters=2, the silhouette score is 0.44351069200040283
 - ❖ For n_clusters=3, the silhouette score is 0.3970904804043125
 - ❖ For n_clusters=4, the silhouette score is 0.31278212110754555
 - ❖ For n_clusters=5, the silhouette score is 0.30514793370029347
 - ❖ For n_clusters=6, the silhouette score is 0.2754916013454332
 - ❖ For n_clusters=7, the silhouette score is 0.27838811963816



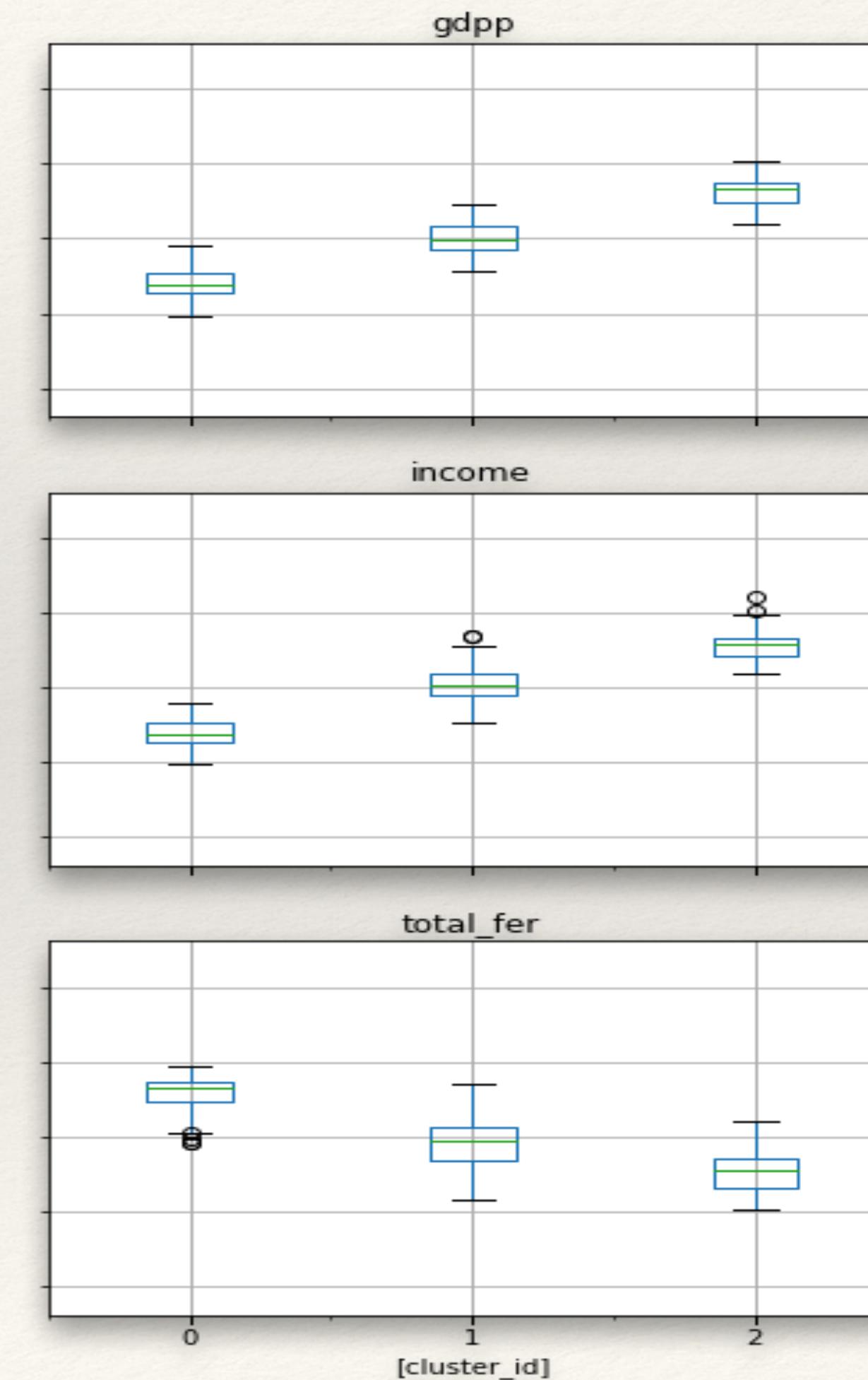
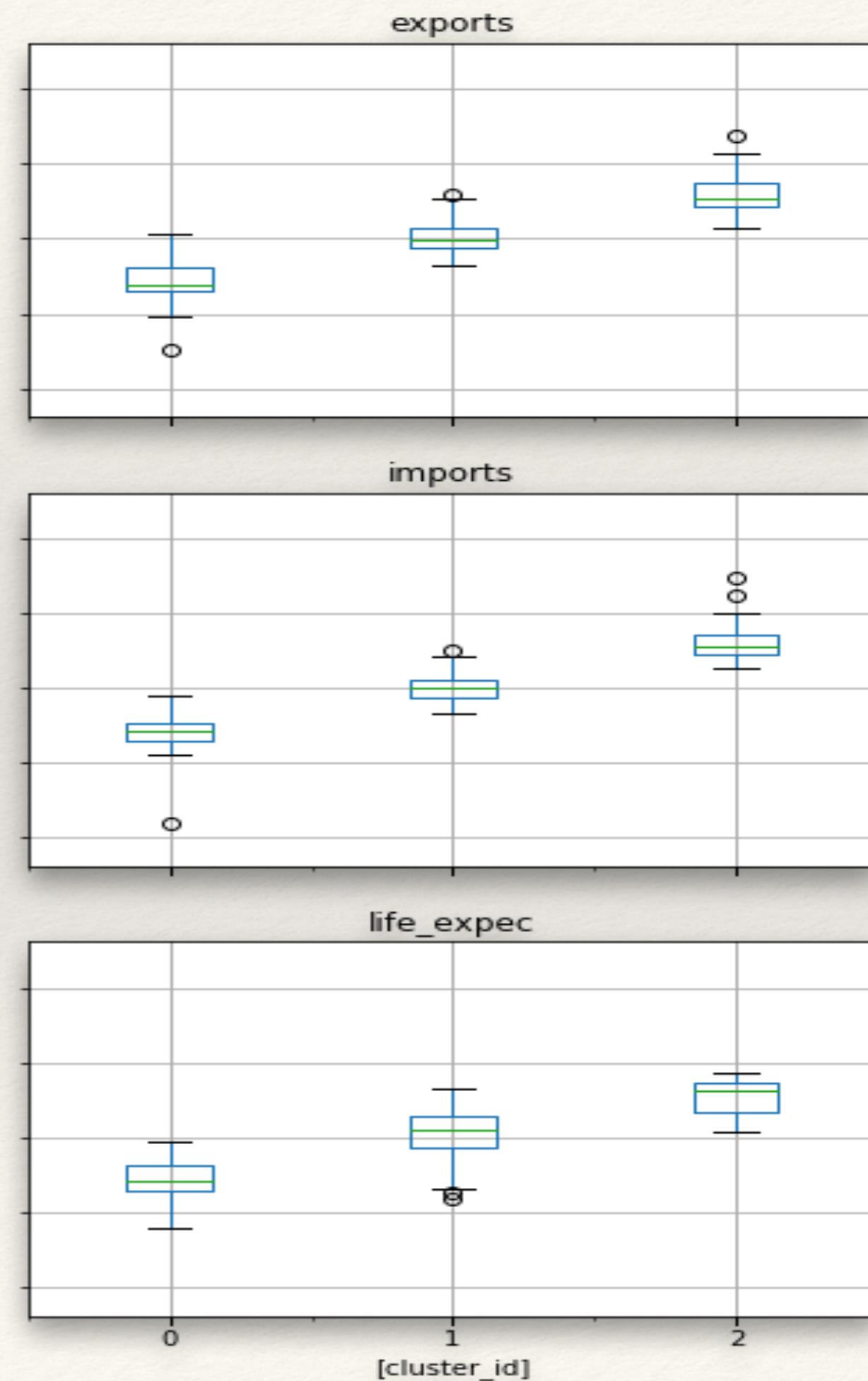
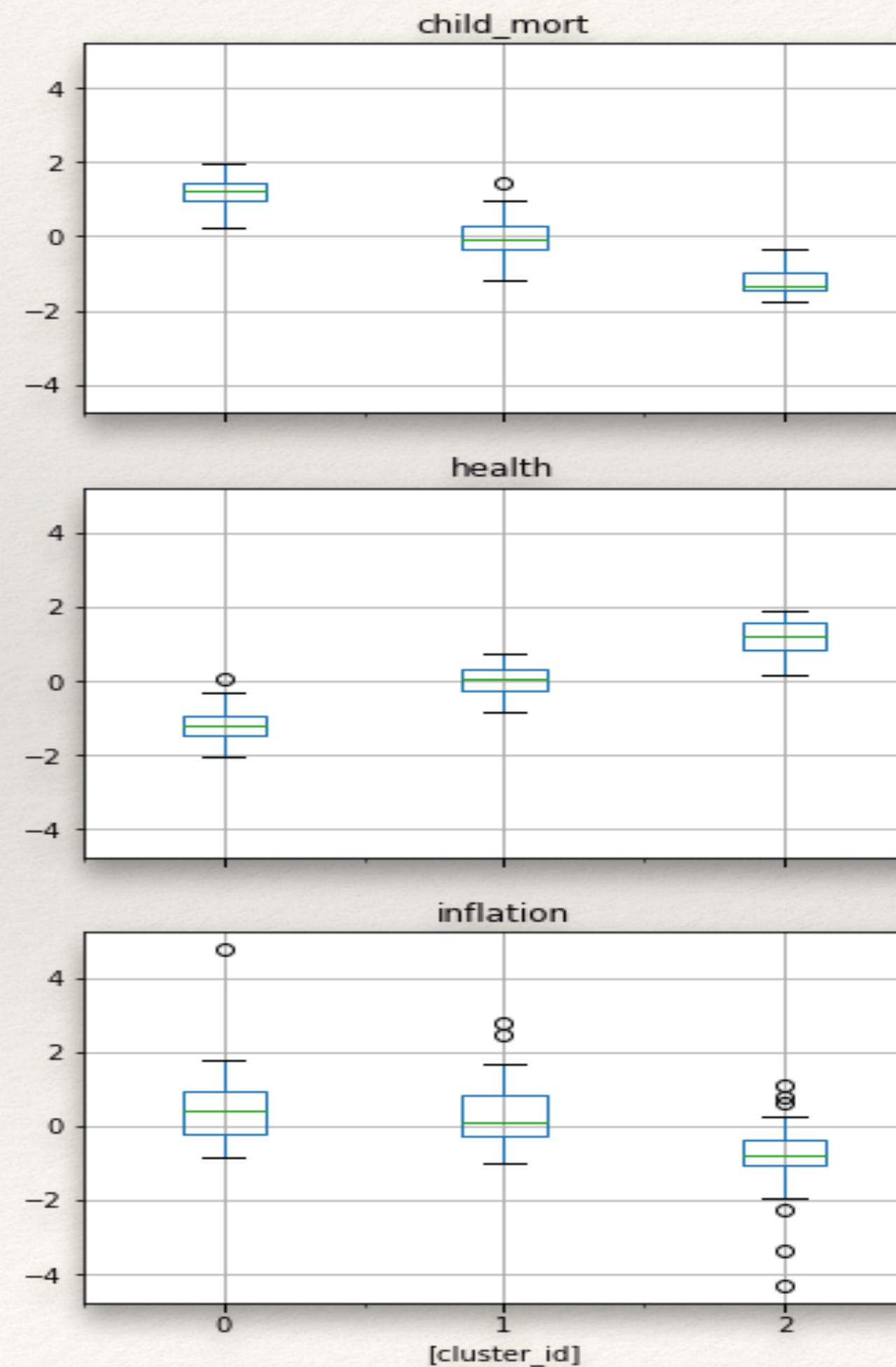
So the optimal value of k is 3.

K-MEANS RESULTS



K-MEANS RESULT

Boxplot grouped by cluster_id

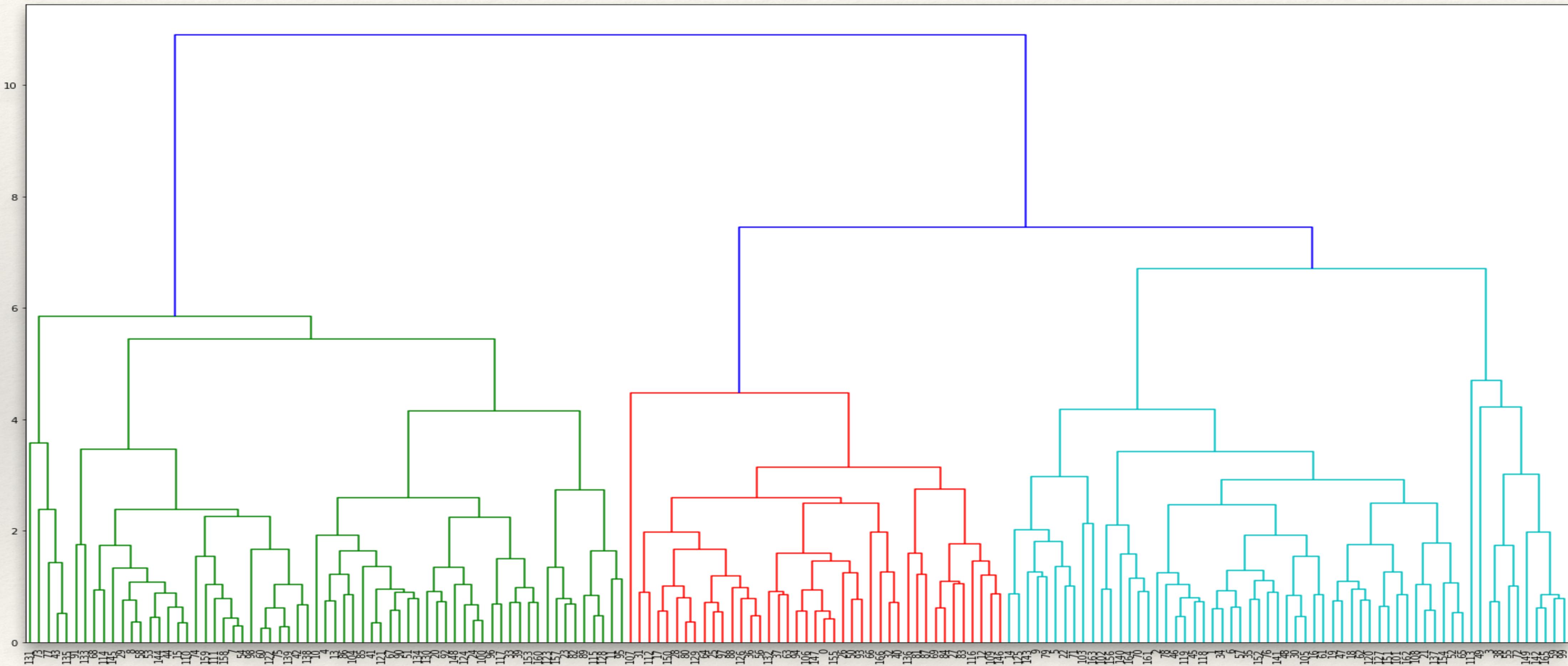


CONCLUSION OF K-MEANS

- ❖ Child death is inversely proportional to exports, imports, income and gdpp.
- ❖ Higher the GDP, higher the income, higher the import, higher the export.

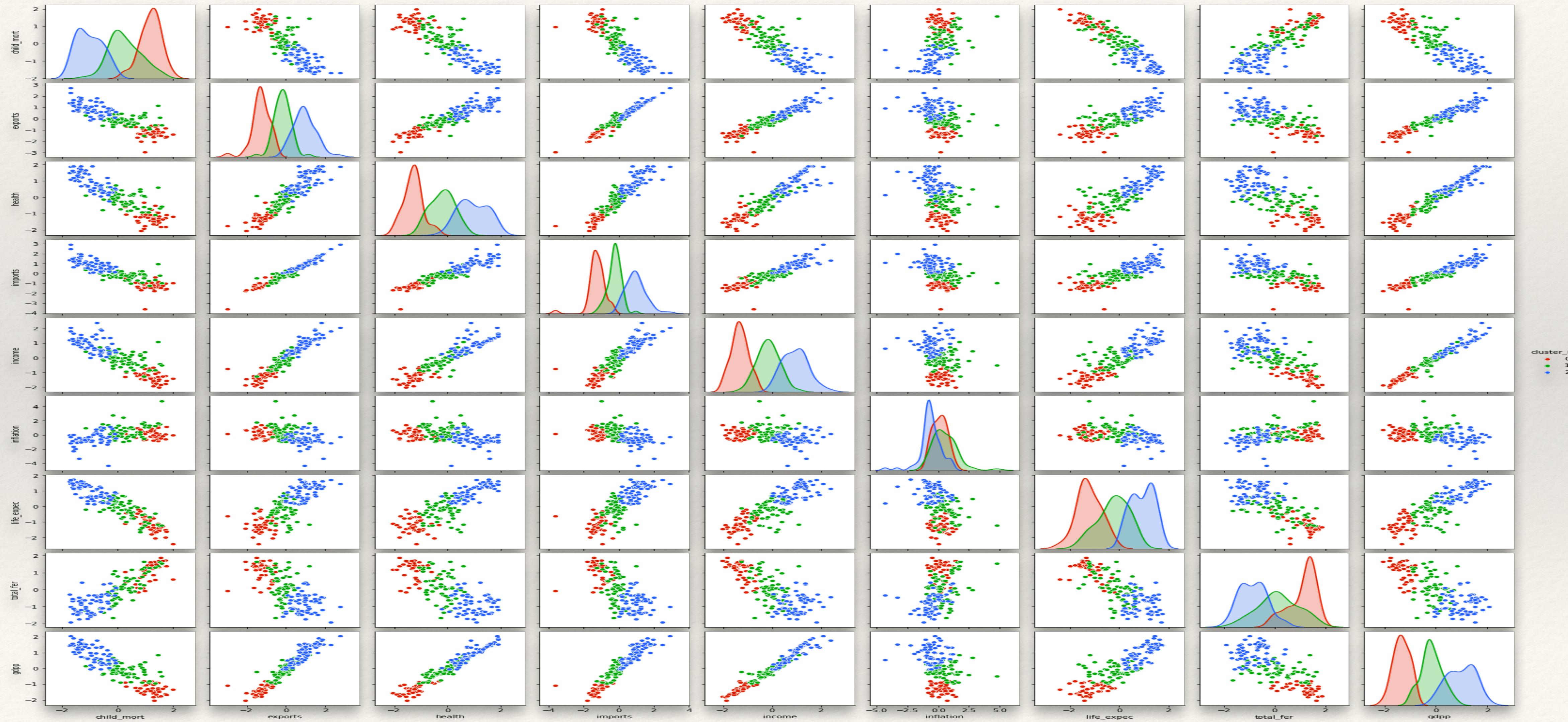
HIERARCHICAL CLUSTERING

DENDROGRAM



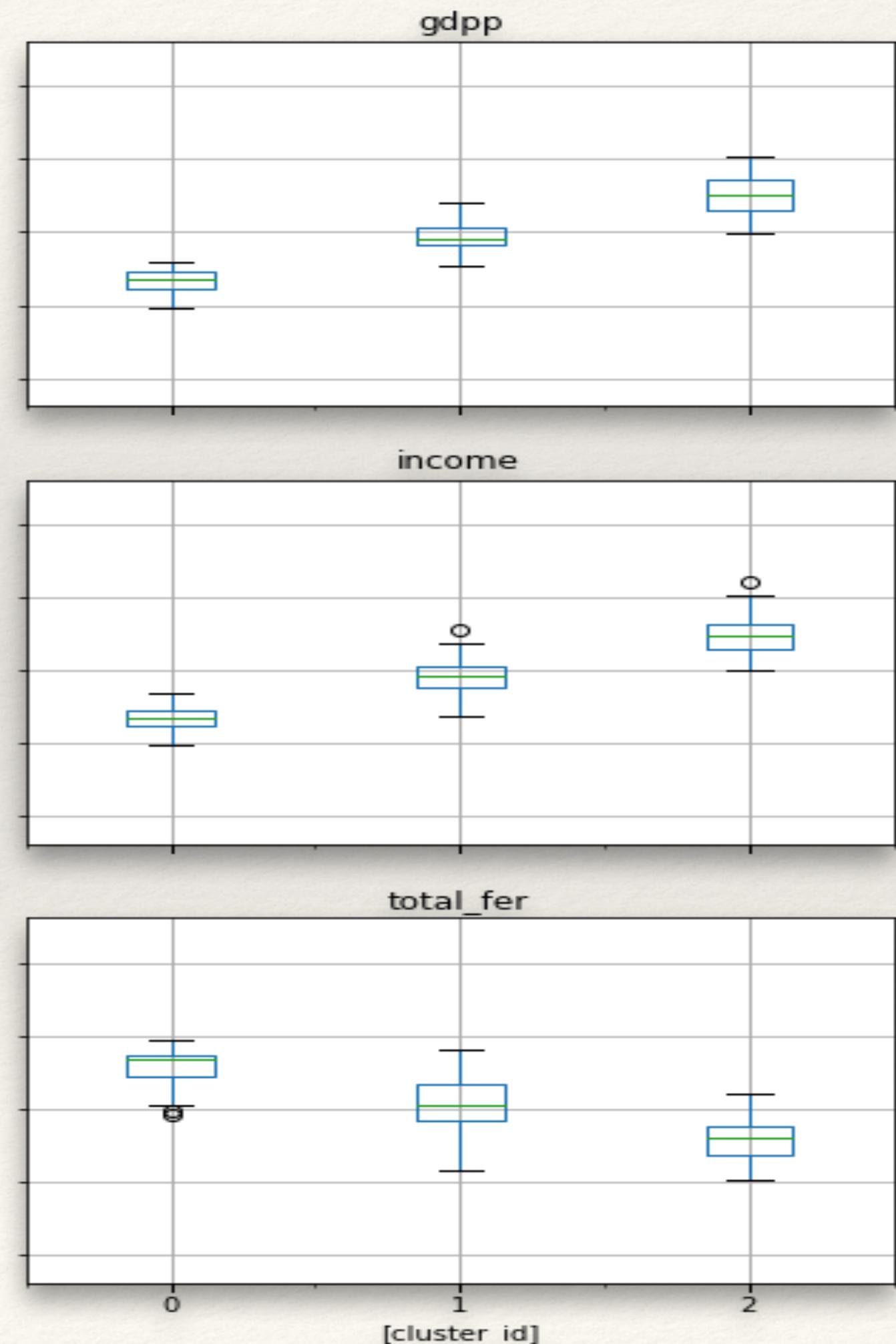
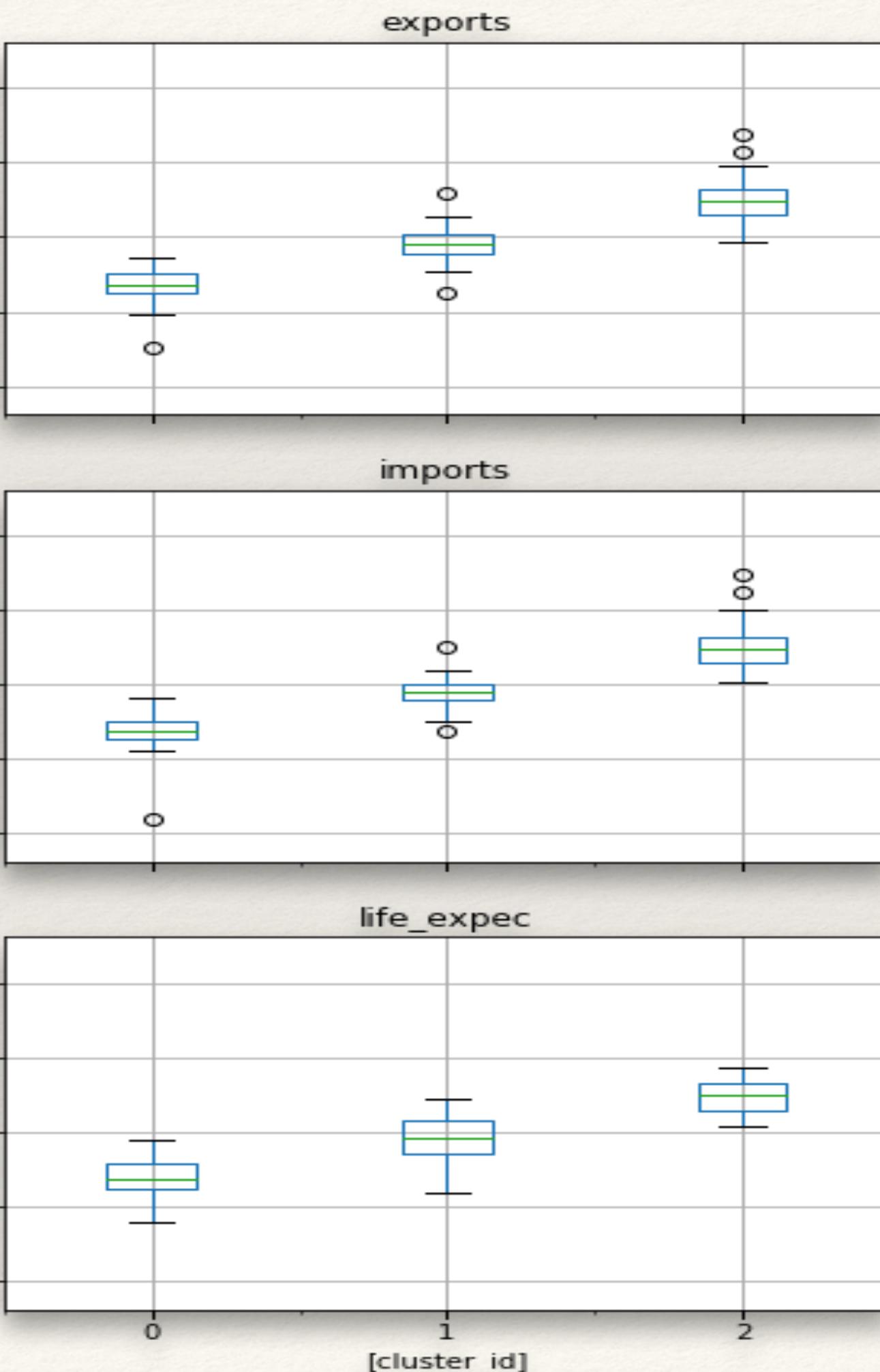
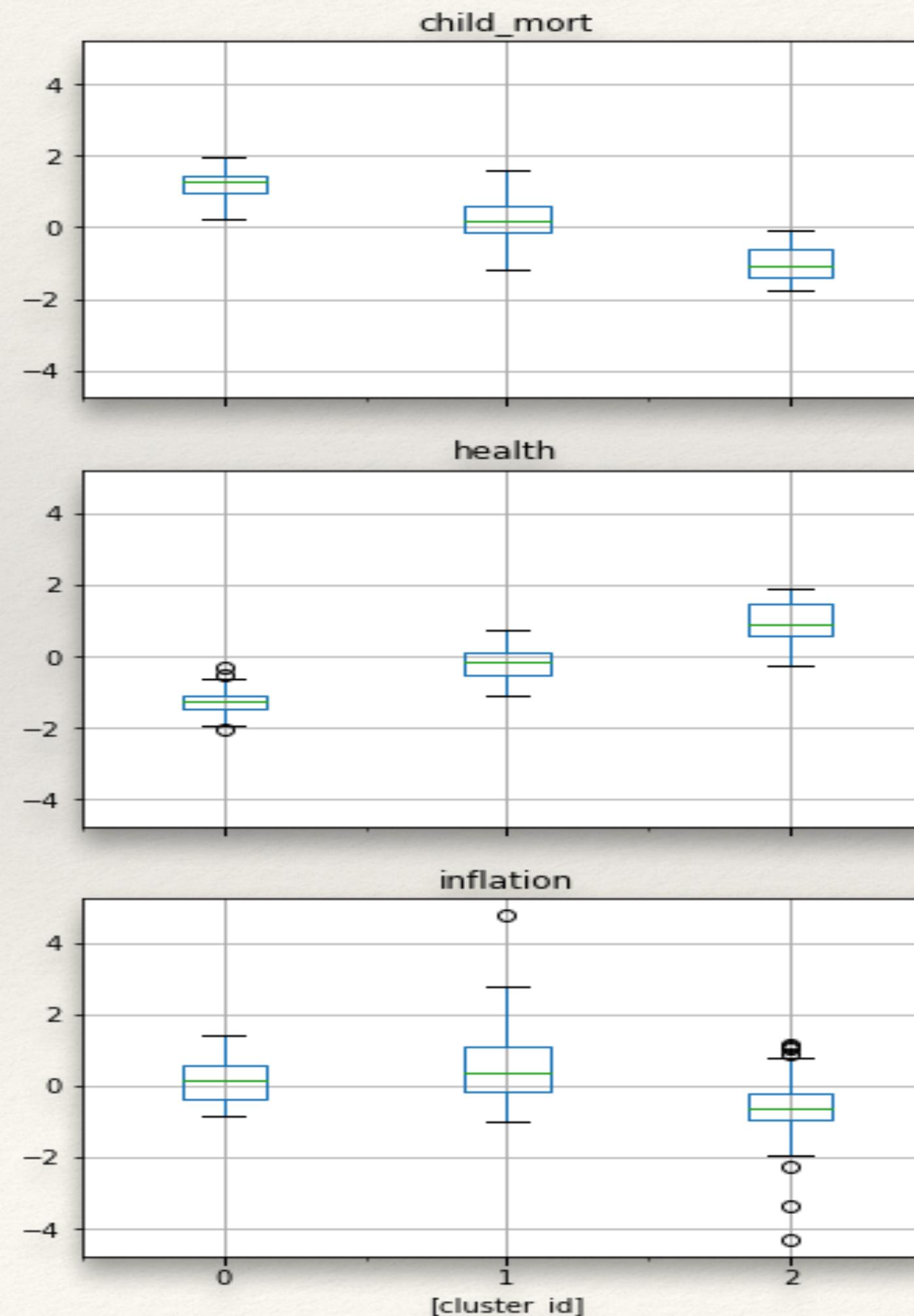
Here we see that we can cut the dendrogram from 7 and divide the data set into 3 clusters.

RESULT HIERARCHICAL CLUSTERING



RESULT HIERARCHICAL CLUSTERING

Boxplot grouped by cluster_id

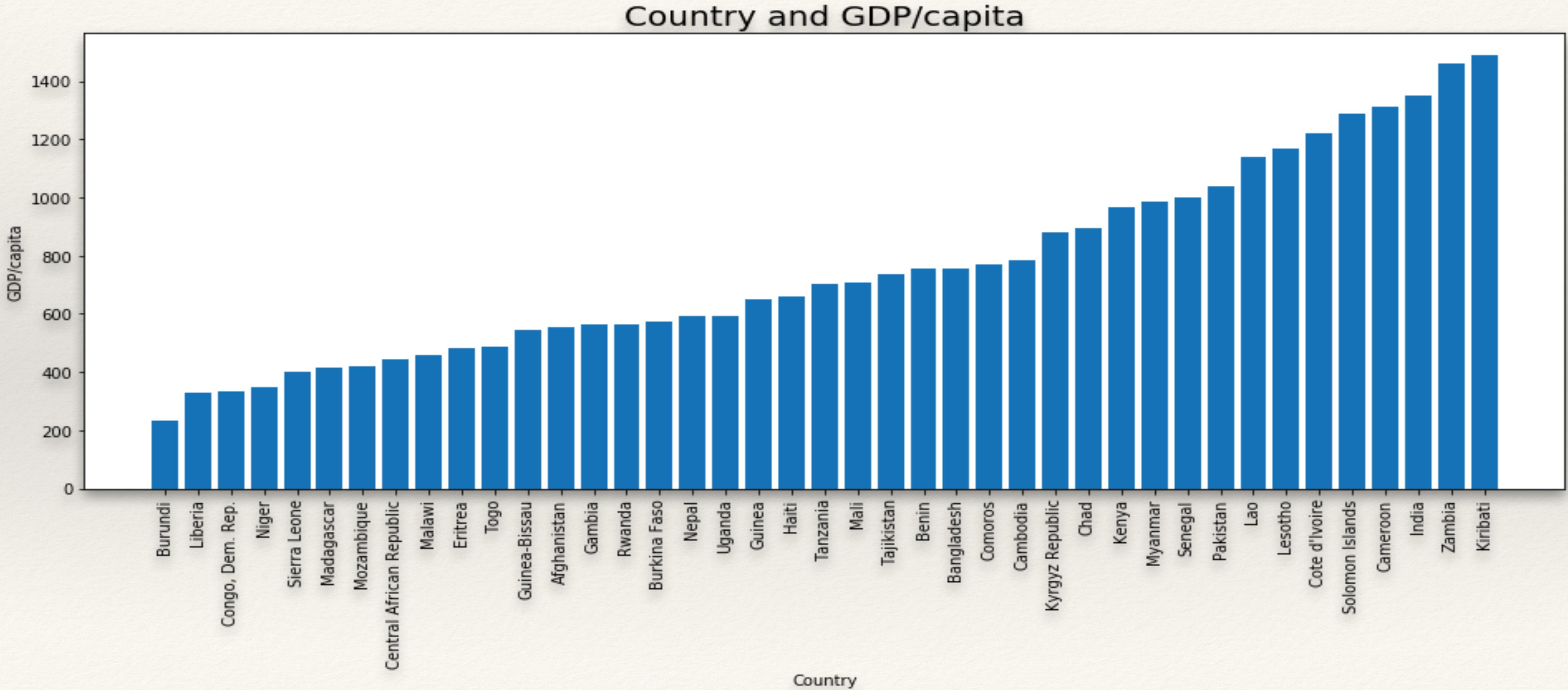


CONCLUSION OF HIERARCHICAL CLUSTERING

- ❖ Child death is inversely proportional to exports, imports, income and GDP.
- ❖ Higher the GDP, higher the income, higher the import, higher the export.
- ❖ Total_fer is inversely proportional to life_expenc

SOLUTION: COUNTRIES THAT NEEDS
ATTENTION

LIST OF COUNTRIES FROM LOWEST GDPP TO HIGHEST GDPP



FINAL LIST OF COUNTRIES WHO NEEDS DIREST OF AID

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	cluster_id
Burundi	93.6	20.6052	26.7960	90.552	764	12.30	57.7	6.26	231	0
Liberia	89.3	62.4570	38.5860	302.802	700	5.47	60.8	5.02	327	0
Congo, Dem. Rep.	116.0	137.2740	26.4194	165.664	609	20.80	57.5	6.54	334	0
Niger	123.0	77.2560	17.9568	170.868	814	2.55	58.8	7.49	348	0
Sierra Leone	160.0	67.0320	52.2690	137.655	1220	17.20	55.0	5.20	399	0
Madagascar	62.2	103.2500	15.5701	177.590	1390	8.79	60.8	4.60	413	0
Mozambique	101.0	131.9850	21.8299	193.578	918	7.64	54.5	5.56	419	0
Central African Republic	149.0	52.6280	17.7508	118.190	888	2.01	47.5	5.21	446	0
Malawi	90.5	104.6520	30.2481	160.191	1030	12.10	53.1	5.31	459	0
Eritrea	55.2	23.0878	12.8212	112.306	1420	11.60	61.7	4.61	482	0

So these countries need most attention because their GDP is very low as compare to other countries. If GDP increases then it will help the country to be self dependent.

THANK YOU