



PCA & Clustering Assignment

DONE BY
DAYANA ANN

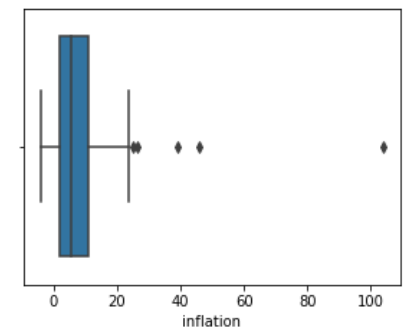
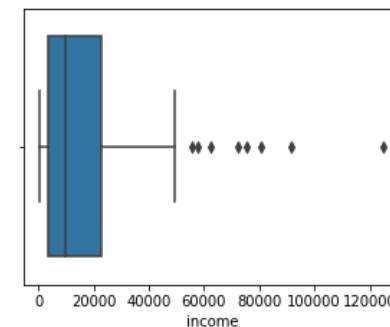
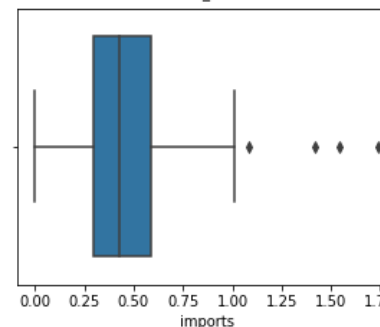
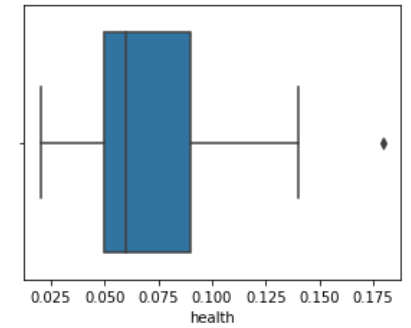
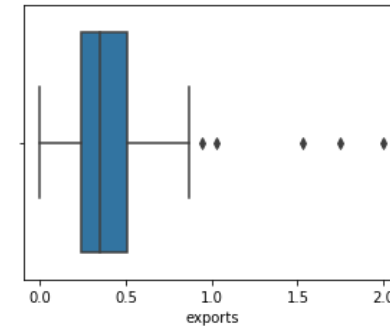
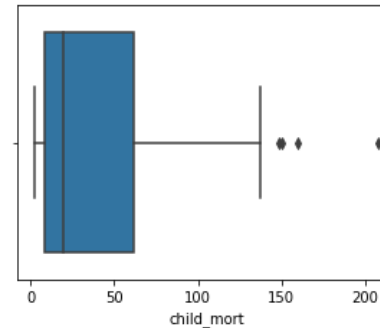
Peek into the data

Data contains countries and corresponding income, gdp, child mortality values etc.
EDA and visualization has been done to understand the dataset.

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

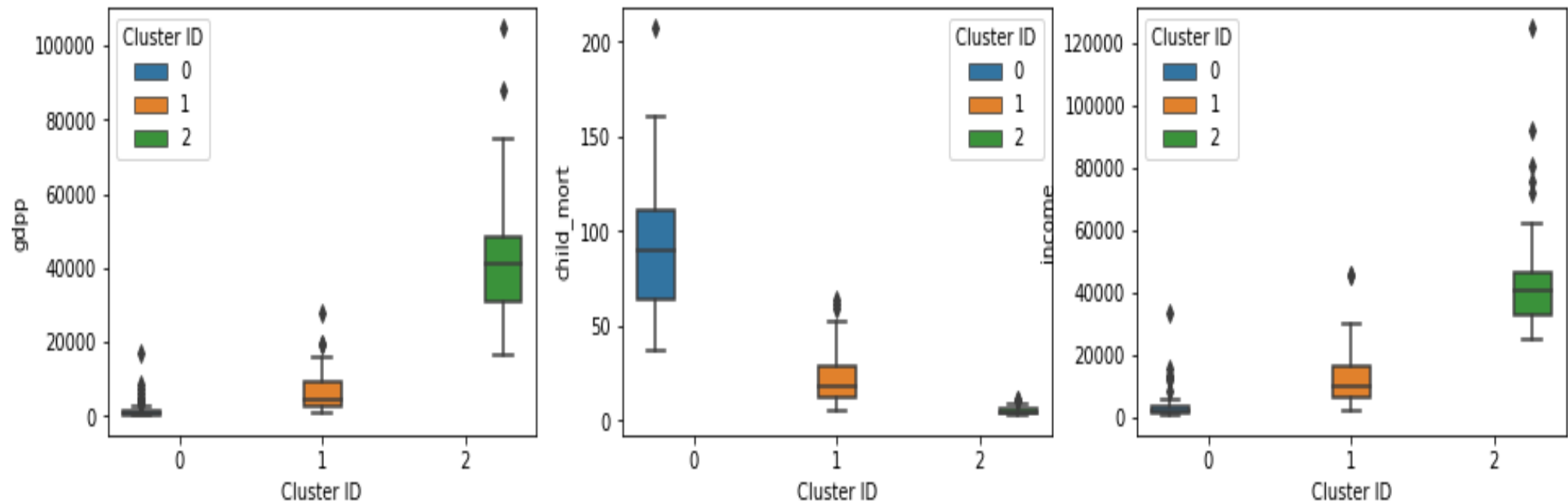
Data Reading and Visualisation

- By referring to the Boxplot we can see there are outliers in the data for most of the variables like *child_mort*, *exports*, *imports*, *income*, *gdpp* etc. Hence here we are not removing the outliers now.¶
- Few variables have few outliers hence it will not have much impact on the Analysis and hence keeping as it is.

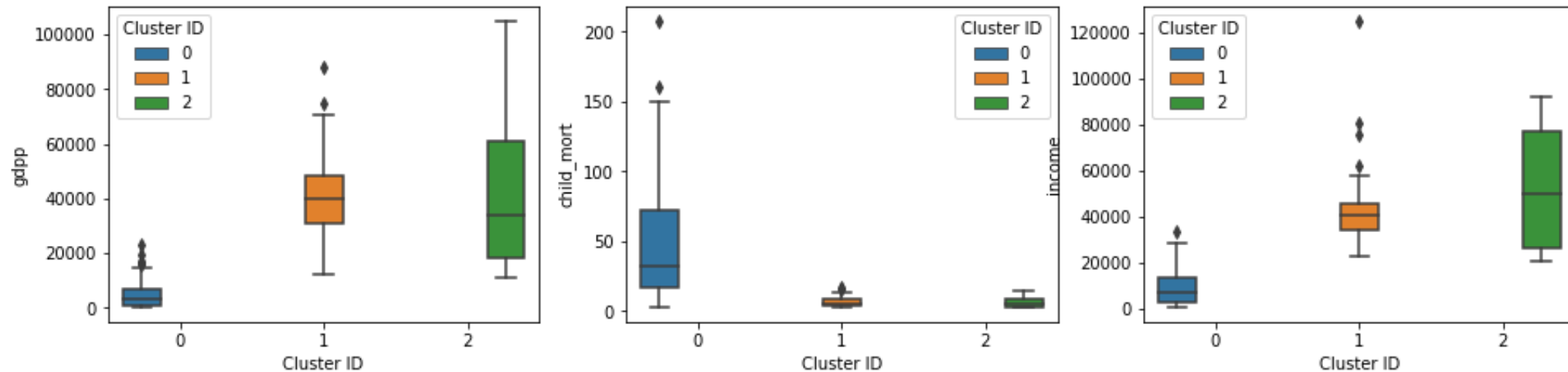


K means result

KMeans Clustering conclusion : From the analysis it is found that the Countries that needs focus on are: Congo, Dem. Rep, Niger, Central African Republic, Sierra Leone & Guinea-Bissau



Hierarchal Clustering result



Hierarchal Clustering conclusion : From the analysis it is found that the Countries that needs focus on are:

Congo, Dem. Rep,

Niger,

Central African Republic,

Sierra Leone &

Guinea-Bissau

Countries that needs focus are

- ▶ Congo, Dem. Rep,
- ▶ Niger,
- ▶ Central African Republic,
- ▶ Sierra Leone &
- ▶ Guinea-Bissau