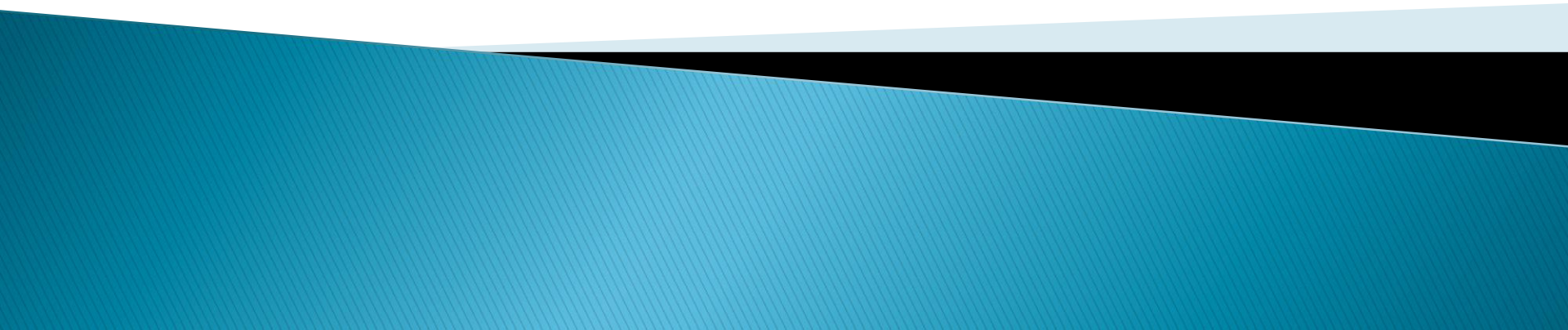
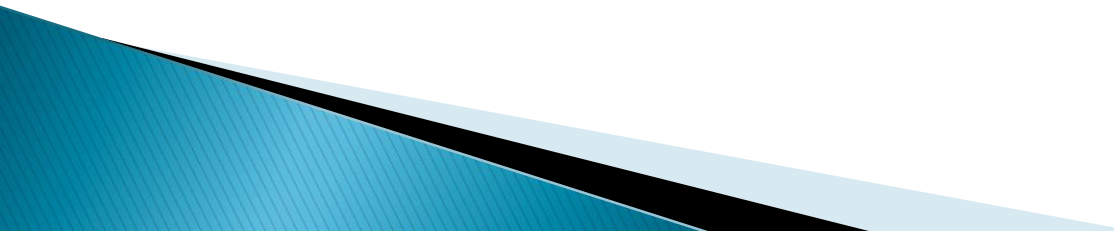


# Clustering Assignment

Garima Sharma



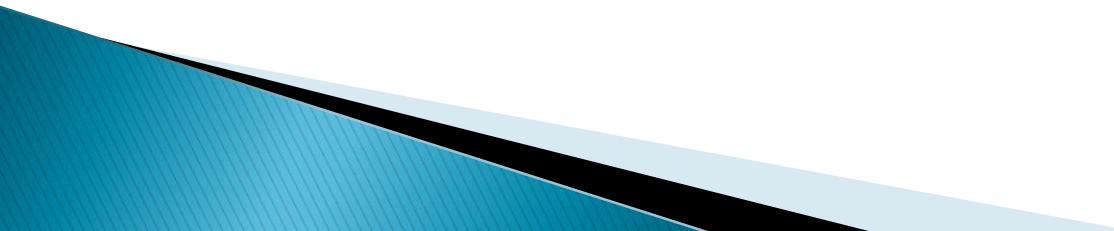
# Brief about the assignment

- ❖ 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 lots of operational projects from time along with advocacy drives to raise awareness as well as funding proposes.
  - ❖ This assignment has preformed on the socio-economic and health factors that determine the overall development of the country.
- 

# Continue....

- ▶ To categories 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
- ▶ Based on the few steps taken to determined the top 10 countries which needed to suggest the countries which the CEO need to focus on the most.

# Step Performed:–

- ▶ Read and visualise data
  - ▶ Clean the data
  - ▶ Visualisation of data
  - ▶ Prepare the data for modelling
  - ▶ Modelling
  - ▶ Final Analysis
- 

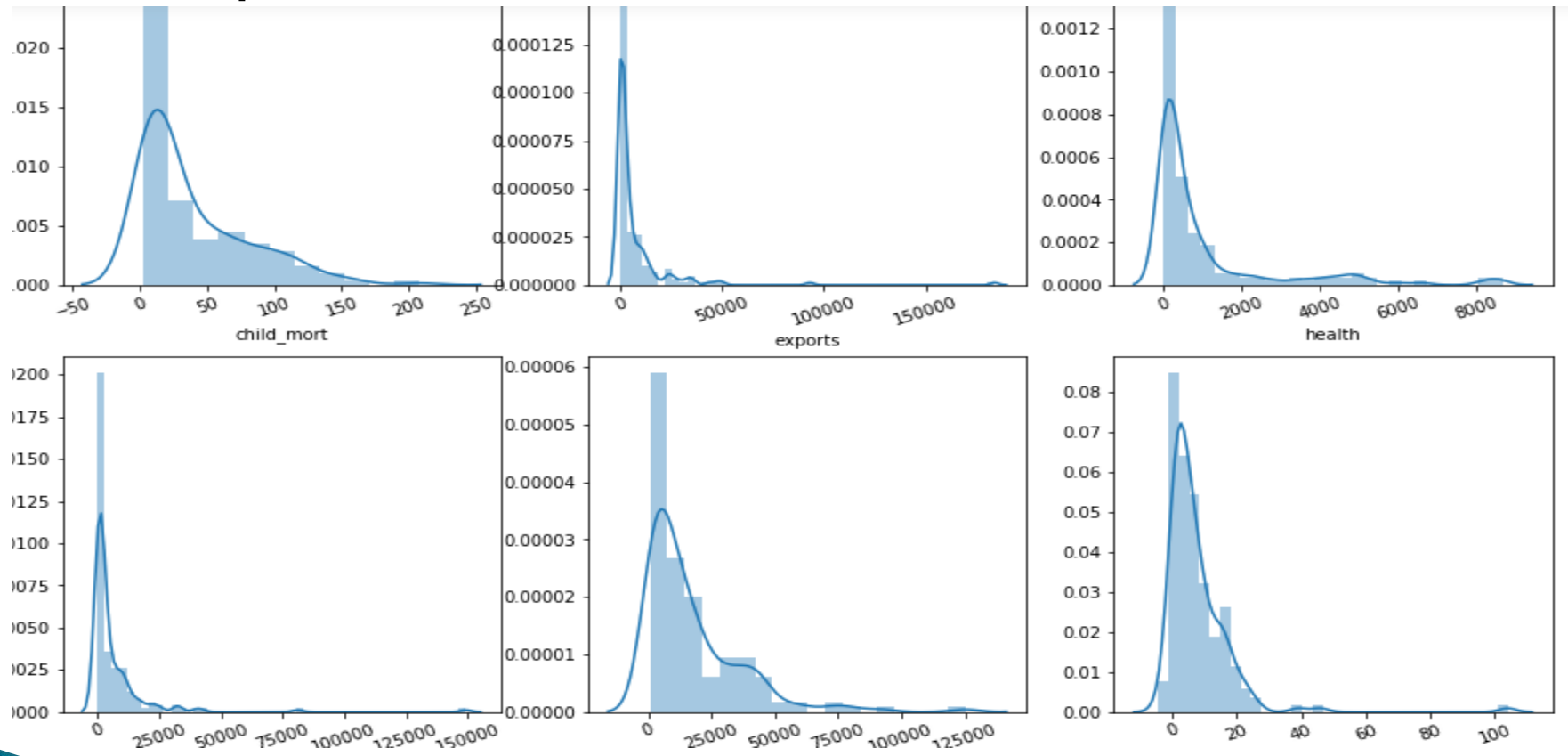
# Read the dataset:-

- ❖ In the dataset they have given countries wise child\_mort, exports, health, import ,income and etc..
- ❖ Based on the income and gdpp we suppose to suggest the CEO which countries are need

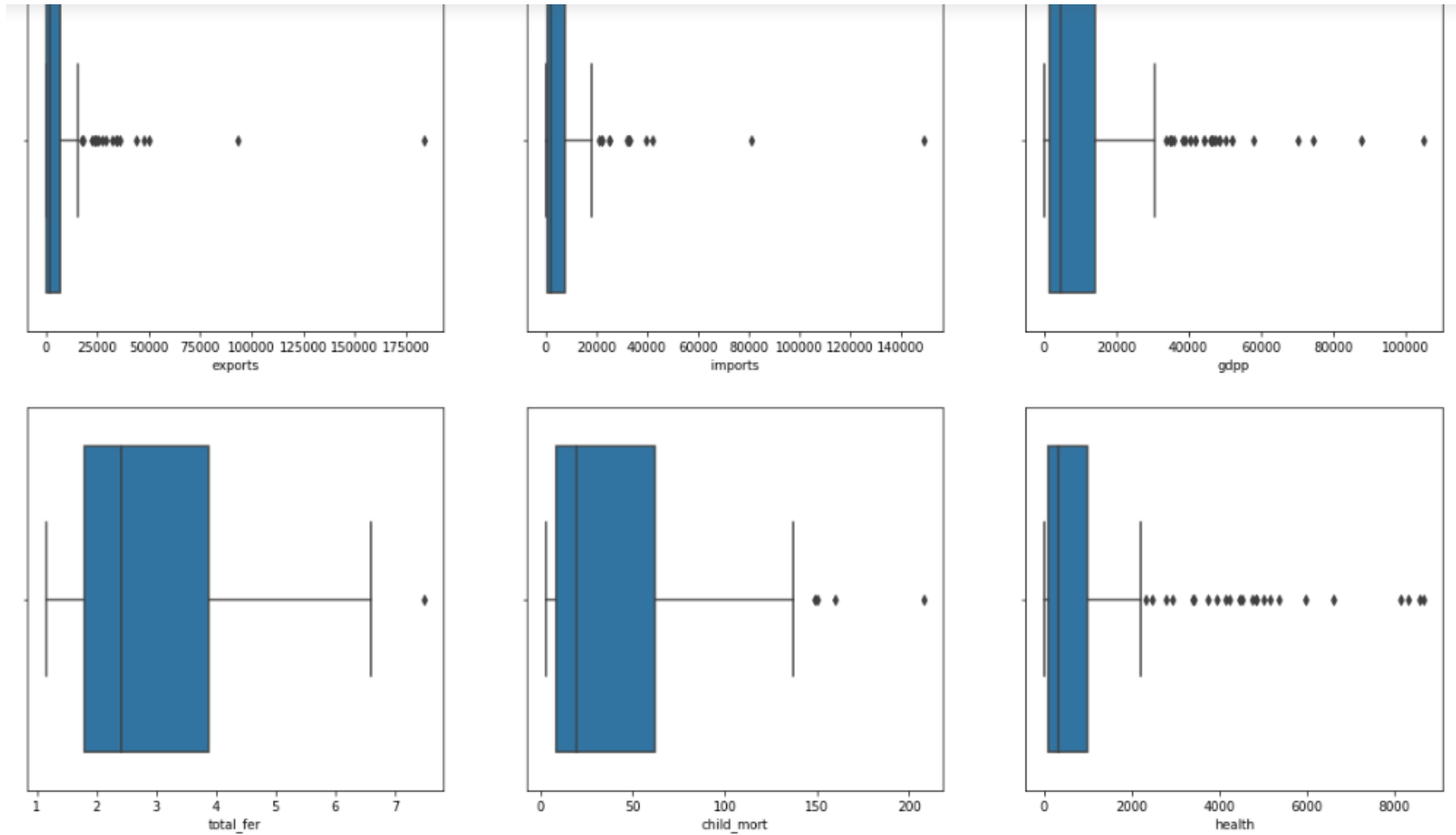
	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1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

# Visualise the data:-

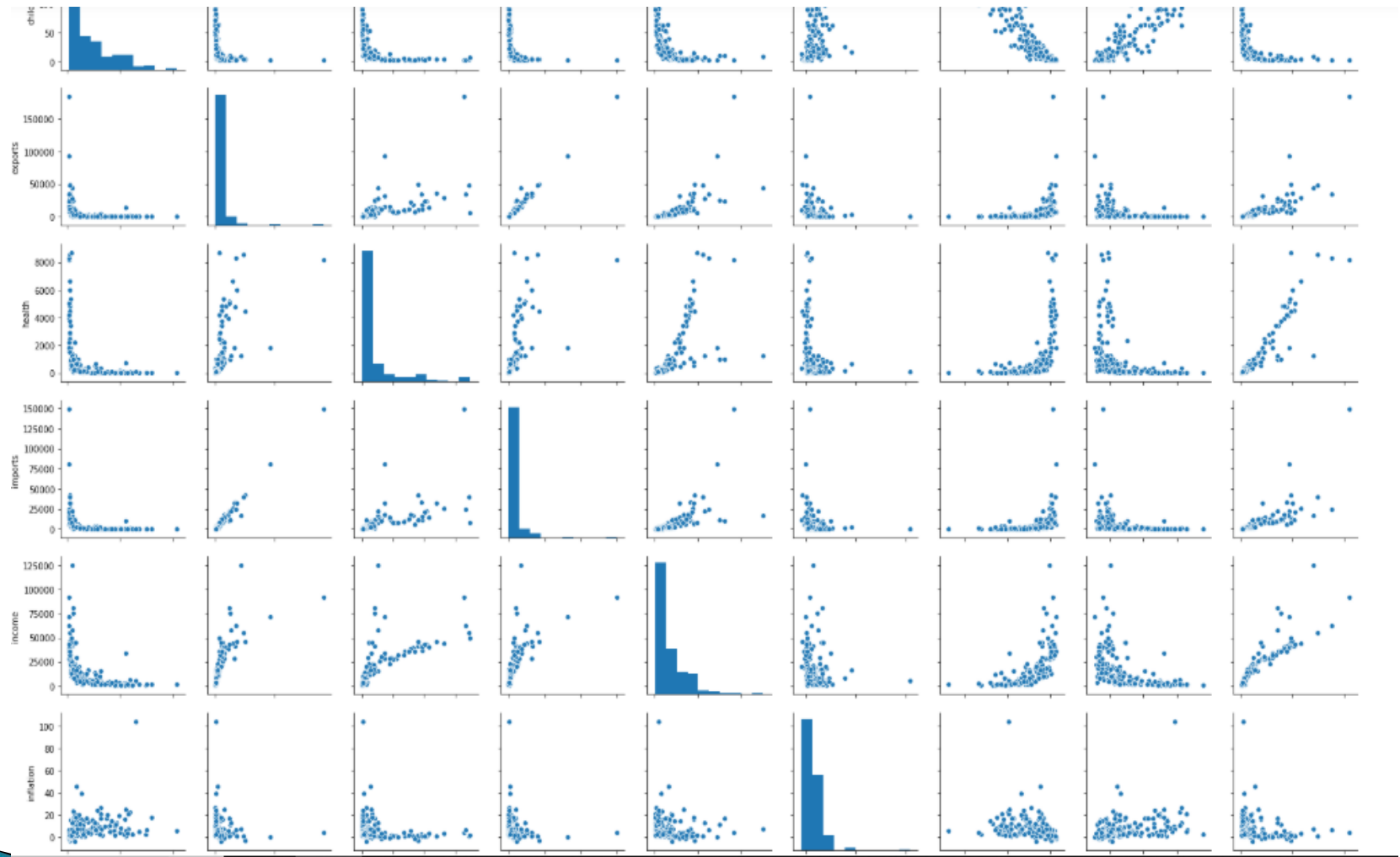
## ► Distplot:-



# BoxPlot:-

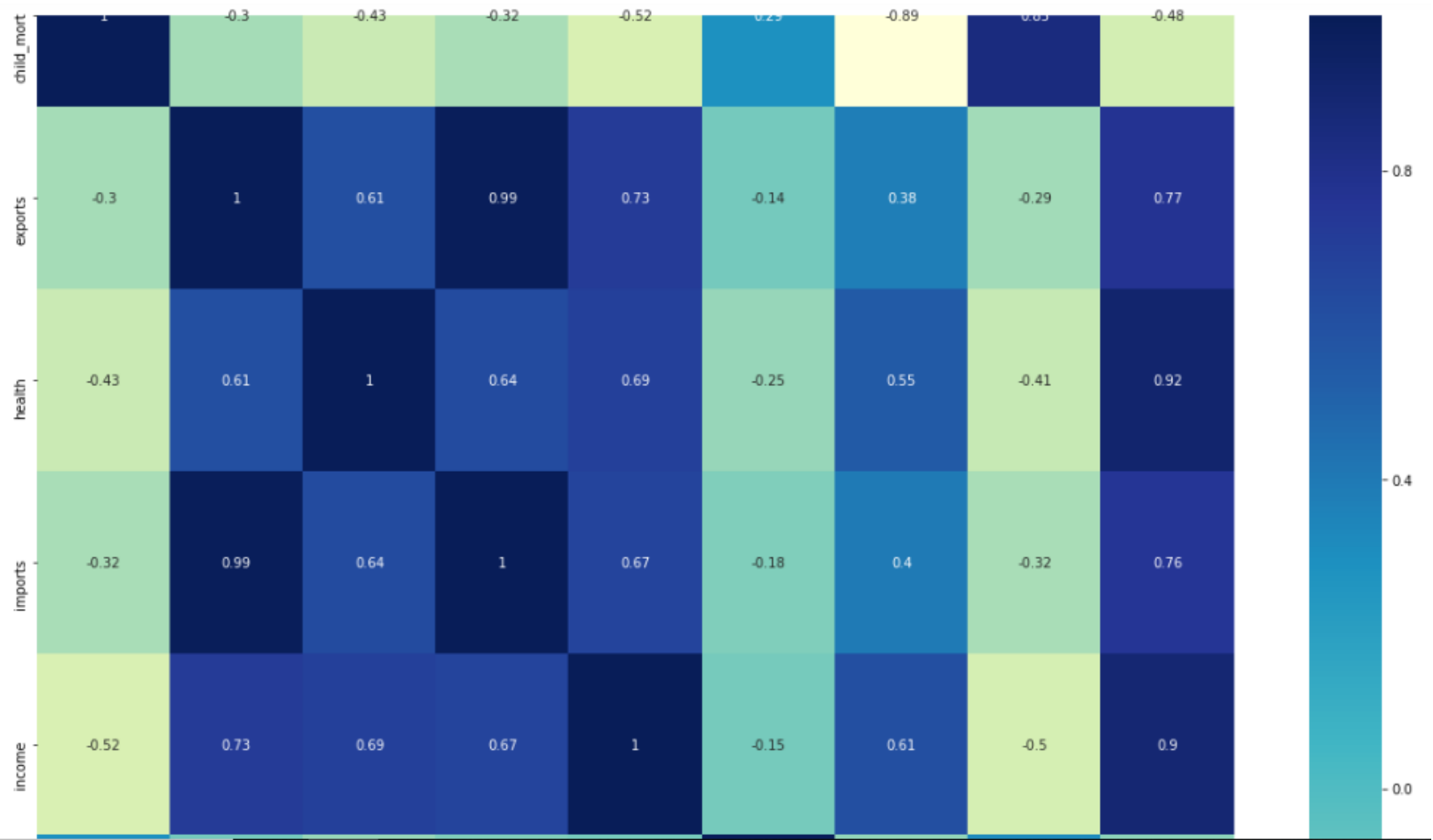


# Pairplot:-

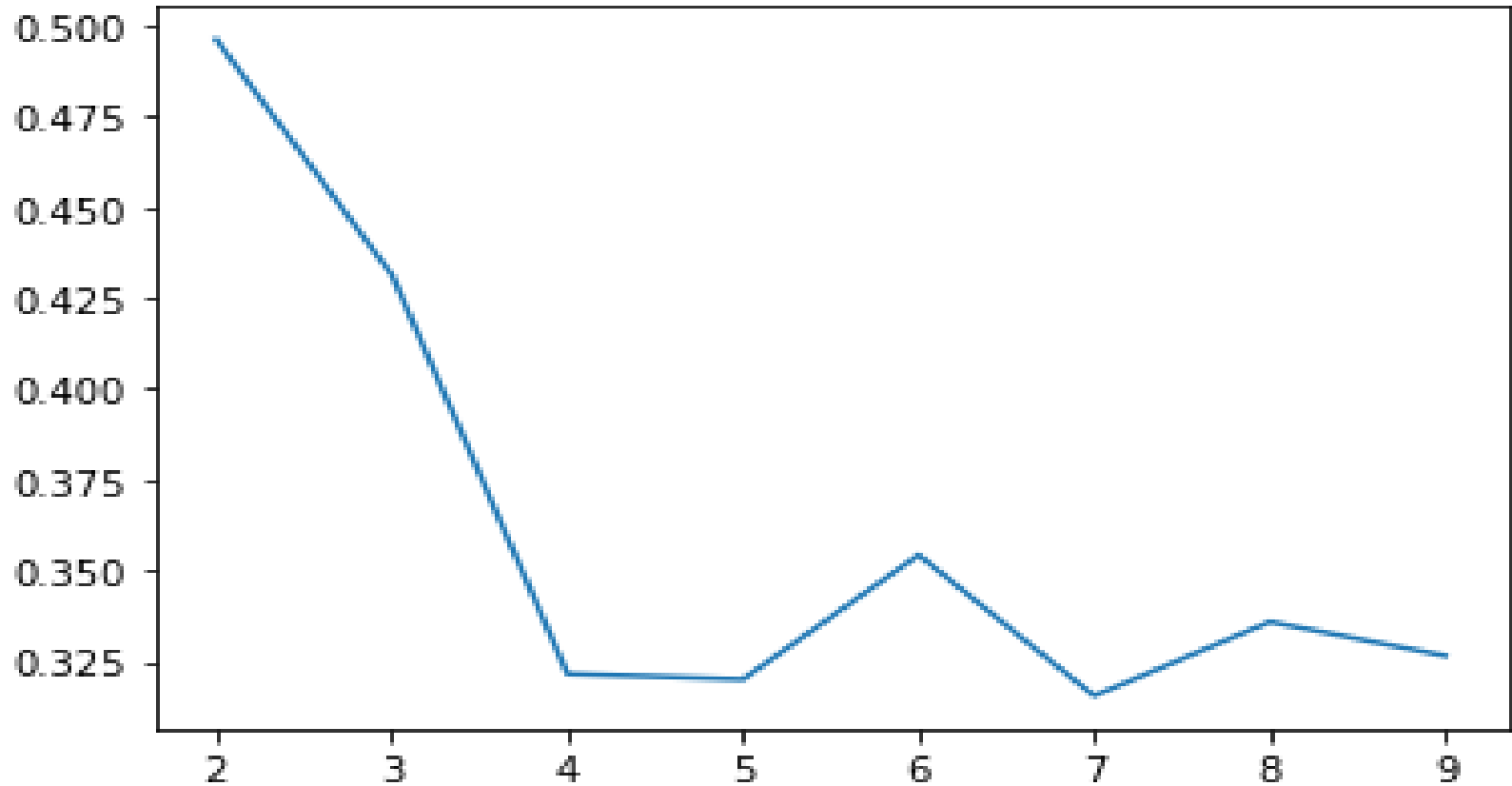




# Heatmap:-

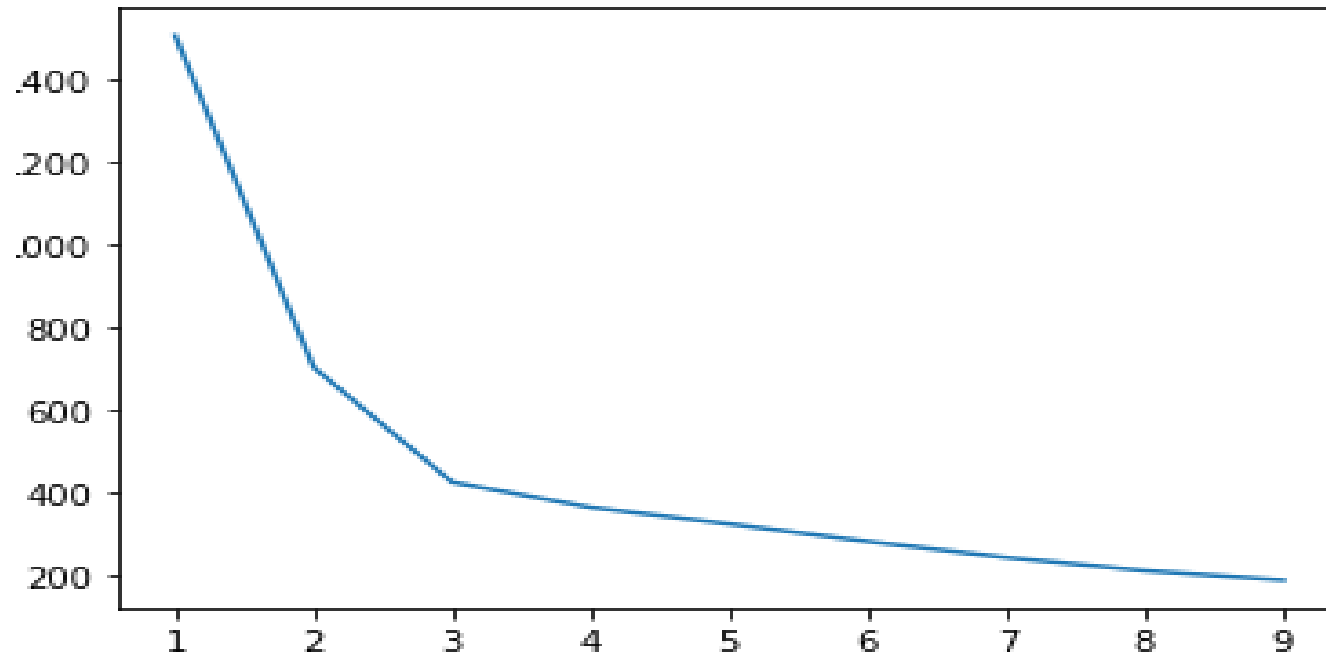


# Finding the optimal number of cluster



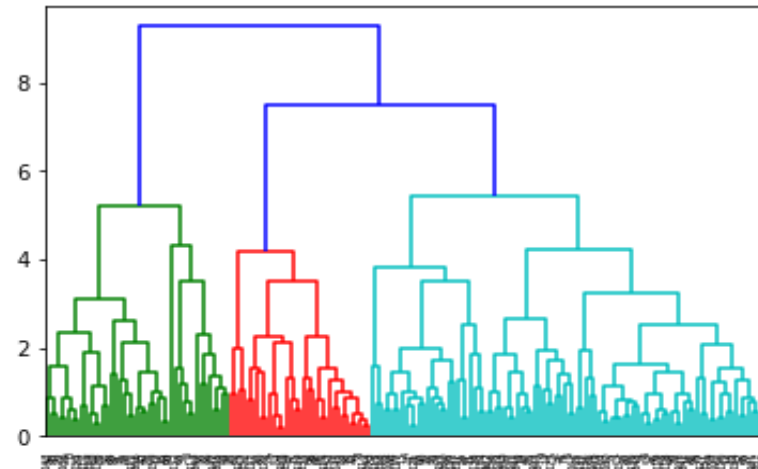
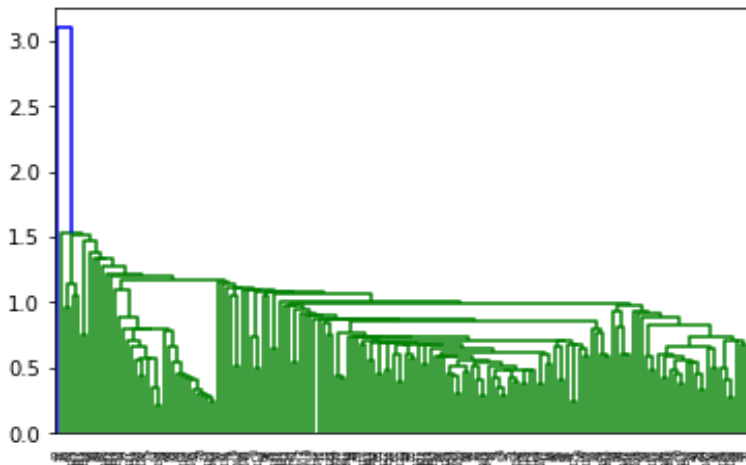
# Elbow curve plot

- ▶ Elbow curve help us to find the cluster
- ▶ In this 3 ,5 are the cluster

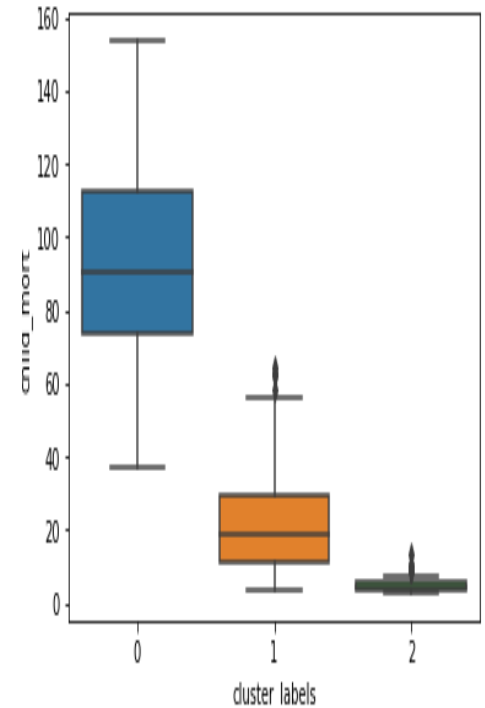
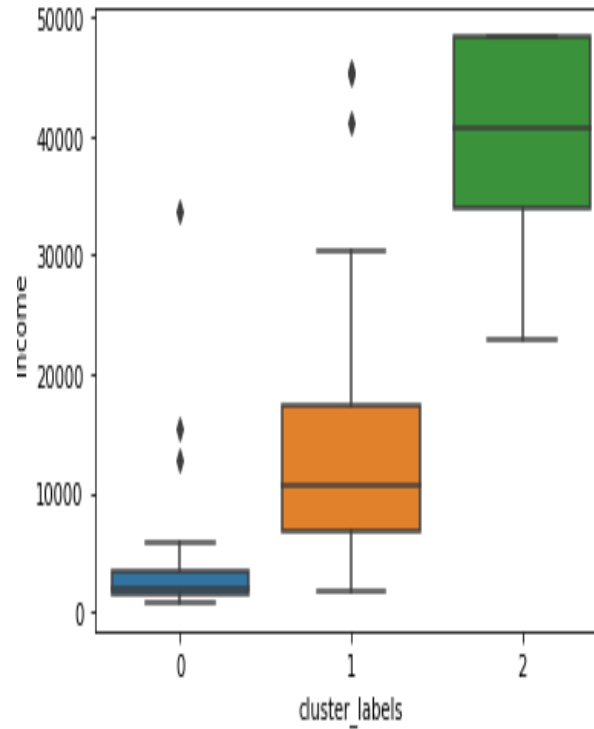
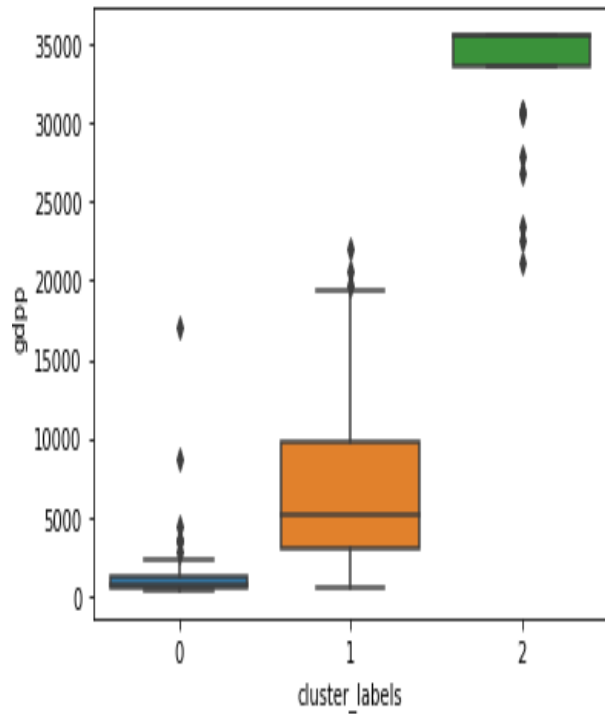


# Single linkage & complete linkage

- ▶ Single linkage is not showing clear picture.
- ▶ Complete linkage we can see 3 cluster.

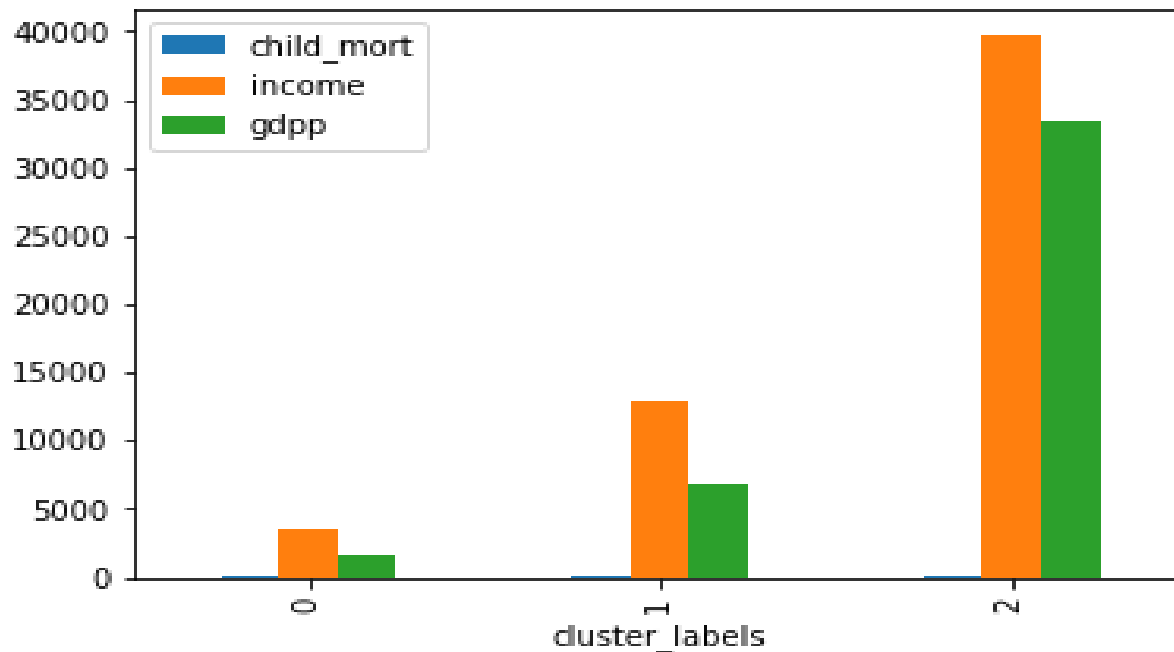


# Visualise with box plot



# Visualise with bar plot

- ▶ Countries with low Gdpp, low income per person have high child\_mort.



# Final Analysis:–

- ▶ These are the top countries which need fund by the NGO.

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	cluster_id	cluster_labels
88	Liberia	89.3	62.457000	38.586000	302.80200	742.24	5.470	60.800	5.0200	331.62	0	0
26	Burundi	93.6	22.243716	26.796000	104.90964	764.00	12.300	57.700	6.2600	331.62	0	0
37	Congo, Dem. Rep.	116.0	137.274000	26.419400	165.66400	742.24	20.800	57.500	6.5400	334.00	0	0
112	Niger	123.0	77.256000	17.956800	170.86800	814.00	2.550	58.800	6.5636	348.00	0	0
132	Sierra Leone	153.4	67.032000	52.269000	137.65500	1220.00	17.200	55.000	5.2000	399.00	0	0
93	Madagascar	62.2	103.250000	17.009362	177.59000	1390.00	8.790	60.800	4.6000	413.00	0	0
106	Mozambique	101.0	131.985000	21.829900	193.57800	918.00	7.640	54.500	5.5600	419.00	0	0
31	Central African Republic	149.0	52.628000	17.750800	118.19000	888.00	2.010	54.276	5.2100	446.00	0	0
94	Malawi	90.5	104.652000	30.248100	160.19100	1030.00	12.100	54.276	5.3100	459.00	0	0
50	Eritrea	55.2	23.087800	17.009362	112.30600	1420.00	11.600	61.700	4.6100	482.00	0	0