

PCA and Clustering

PRESENTED BY-

SHIVAM GUPTA

Problem Statement:-

1. Task is to cluster the countries by the factors like socio-economic and health factors that determine the overall development of the country.
2. We need to suggest the countries which the CEO needs to focus on the most so that the fund is distributed to the countries that need it the most.

Analysis Approach:-

The analysis approach that we have used here is

1. We have first checked the data.
2. Then we have removed the attributes which had some outliers.
3. Then we have used to form clusters i.e. tried to group the data which will help us to group the countries which are un need of the help.

Results of Principal Component Analysis and Clustering :

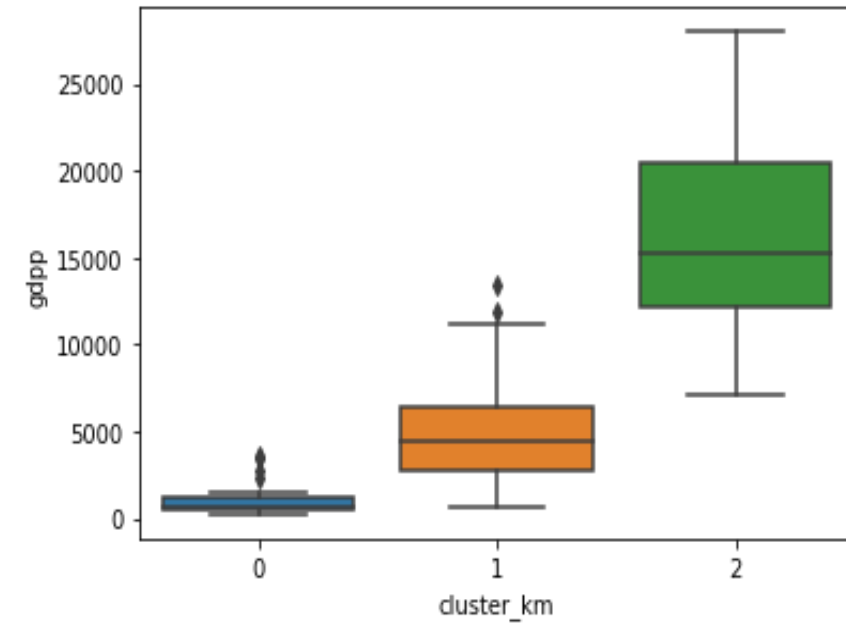
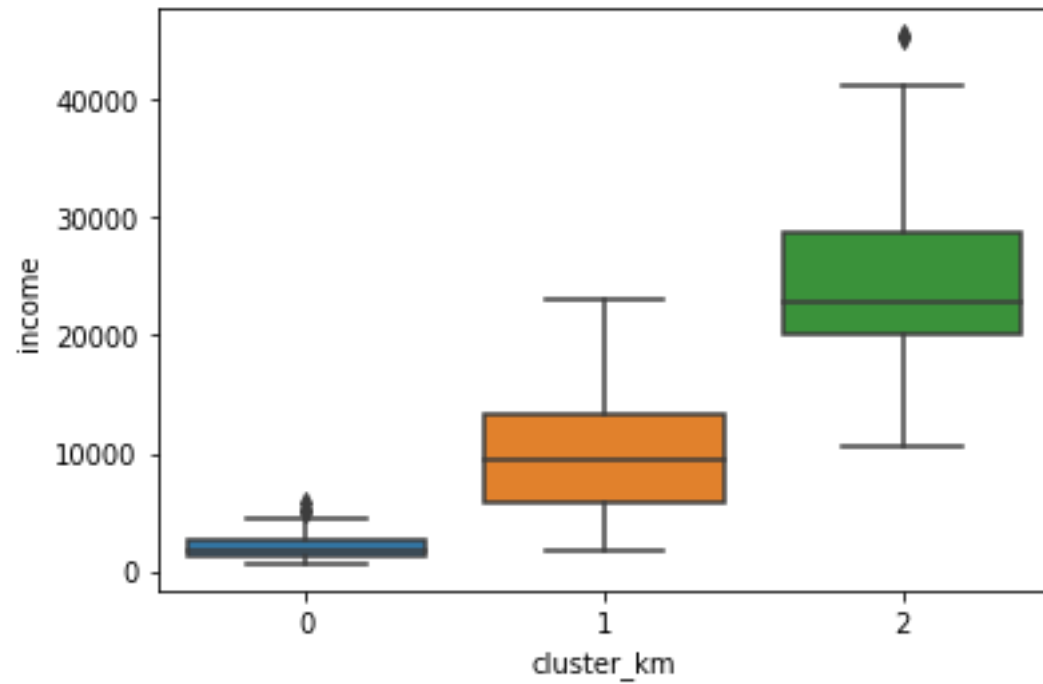
On applying PCA after cleaning the data and removing the outliers we have found that the optimal no of clusters are 3.

When we did clustering we found out that 3 clusters are the best option using silhouette score

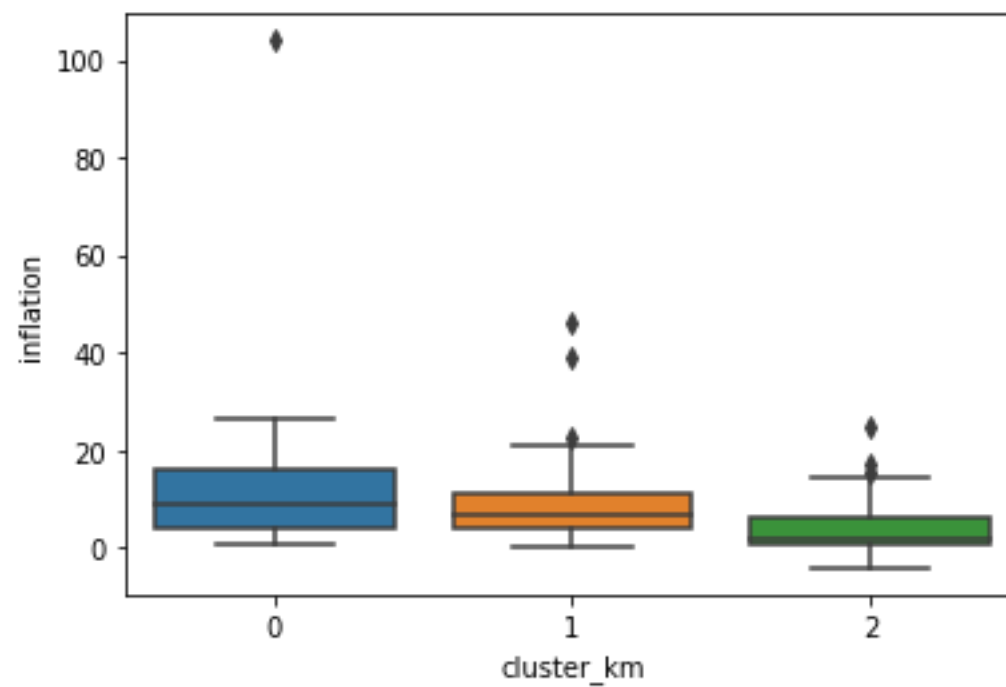
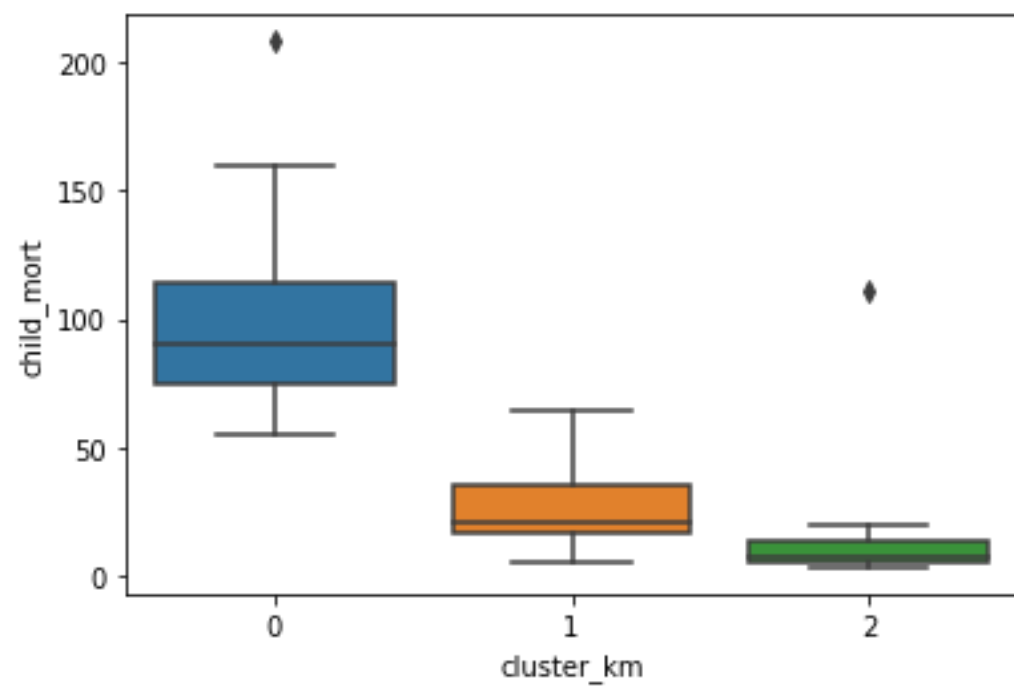
We also concluded that cluster 0 is found to be the best that suits for low gdpp and income and also for high child_mort.

Moreover applying k-means and hierarchical on the same data set we found out that k-means proves out to be better as we have smaller data.

Visualisations :-



b.



Summary of above visualization:-

From above we can conclude that cluster 0 is best suited for the NGO to get the donation to develop

Top five countries

Out[235]:

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	cluster_km	cluster_h
90	Myanmar	64.4	0.109	1.97	0.0659	3720	7.04	66.8	2.41	988	1	2
46	Gabon	63.7	57.700	3.50	18.9000	15400	16.60	62.9	4.08	8750	1	2
129	Turkmenistan	62.0	76.300	2.50	44.5000	9940	2.31	67.9	2.83	4440	1	2
58	India	58.8	22.600	4.05	27.1000	4410	8.98	66.2	2.60	1350	1	1
91	Namibia	56.0	47.800	6.78	60.7000	8460	3.56	58.6	3.60	5190	1	0

From here we can conclude that Myanmar is the most needed country if we study in terms of low gdpp low income and high child_mort.

Thankyou