



PCA and Clustering Assignment:

Submitted By-

Simran Jain




Problem Statement: -

1. 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.
2. Task is to cluster the countries by the factors like socio-economic and health factors that determine the overall development of the country.
3. We need to filter out the 5 countries which are in need of the help.



Analysis Approach:-

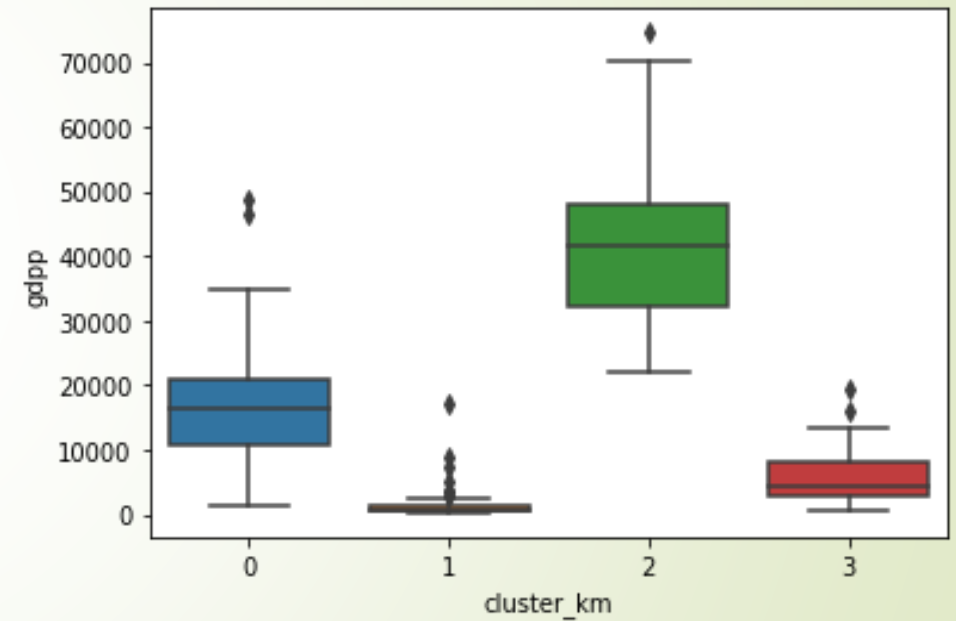
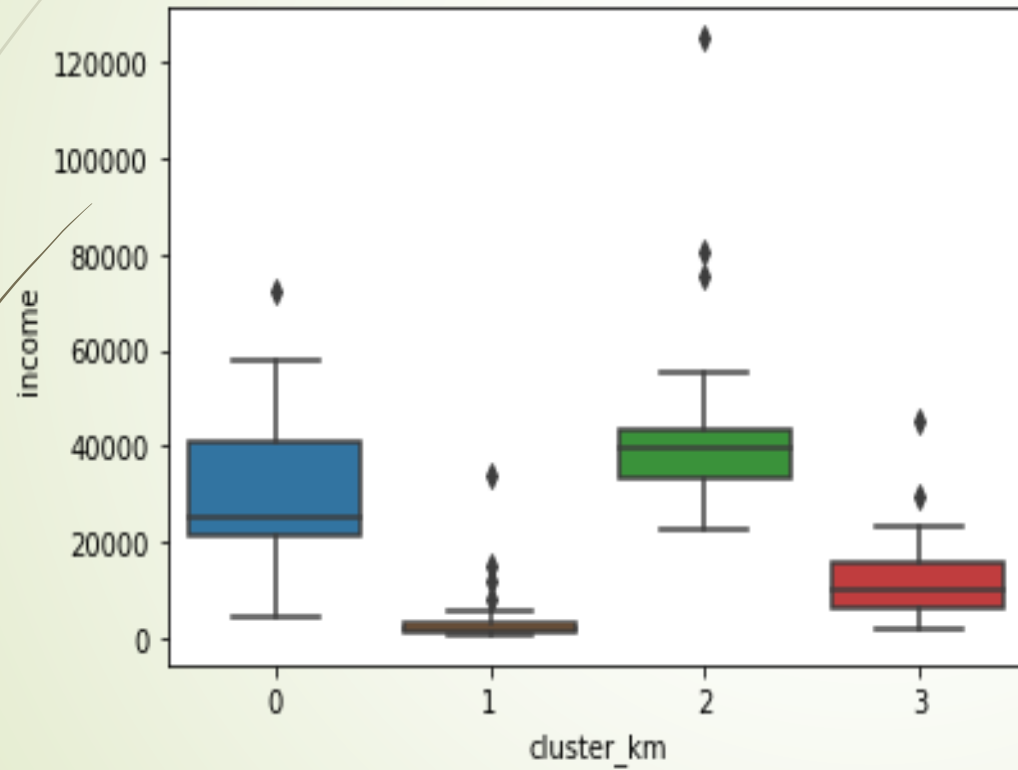
- ▶ We have analyzed the data for both k means and hierarchical clustering.
- ▶ Then we have used to form clusters i.e. tried to group the data which will help us to group the countries which are un need of the help.
- ▶ On conclusion we have found out that k means proves out to be better .



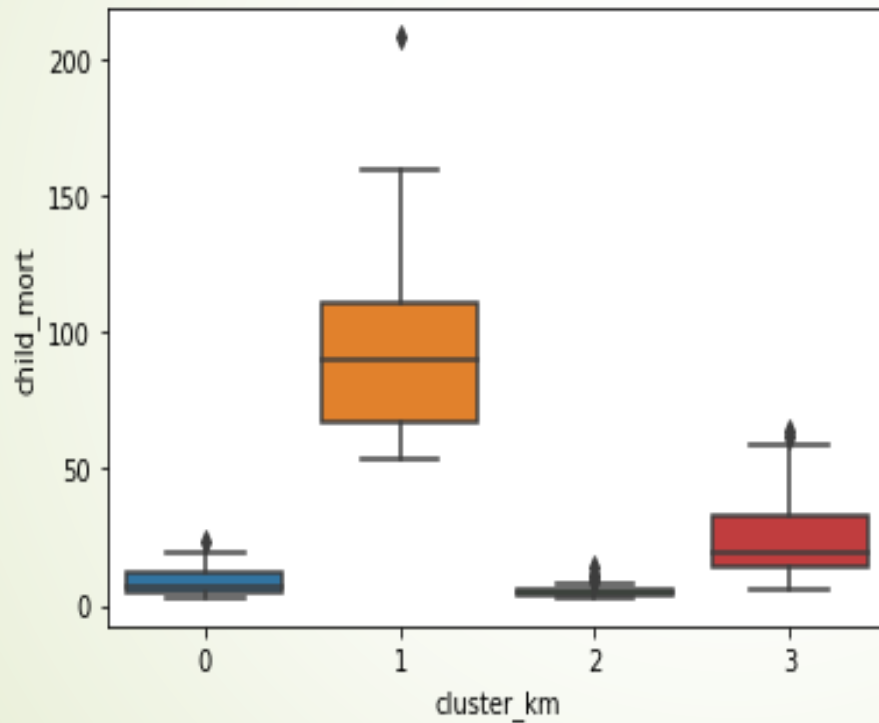
Results of Principal Component Analysis and Clustering :

- On applying PCA after cleaning the data and removing the outliers we have found that the optimal no of clusters are 4.
- When we did clustering we found out that 4 clusters are the best option using silhouette score
- We also concluded that cluster 1 is found to be the best that suits for low gdpp and income and also for high child_mort.

Visualizations:- (Part a)

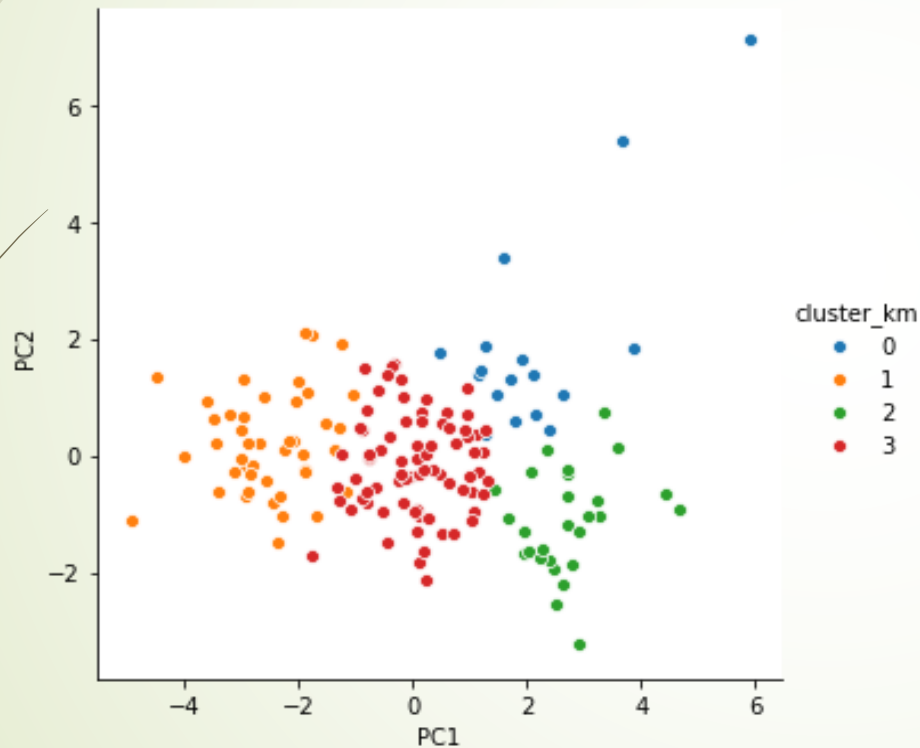


Visualizations:- (Part b)



- On analyzing all the three major aspects of income gdp and child_mort we conclude that cluster 1 is the best cluster that can be used to judge the countries.

Analysis of the pair plot



- On analyzing PC1 and PC2 we can see that we have clear cluster through k means.
- While we have analyzed hierarchical we could not cluster



Top 5 Countries based on the analysis:

- 1. Haiti
- 2. Sierra Leone
- 3. Chad
- 4. Central African Republic
- 5. Mali

Based on analyzing we observed that these 5 countries are the best.



Thankyou