

# Clustering Assignment

## Problem statement

Cluster the countries, and then compare them with these three variables - [gdpp, child\_mort and income]. Report back at least 5 countries which are in direst need of aid.

## Analysis Approach

- 1) Read and understand the data
- 2) Clean the data, and make necessary modifications
- 3) Prepare the data for modelling (Perform PCA)
- 4) Modelling (KMeans, and Hierarchical clustering method)
- 5) Final analysis (Choose one method, and analyze the results)

## **Results of Principal Component Analysis (PCA)**

It was used as a dimensionality reduction technique in the dataset. Even though columns were removed, PCA dataset was able to explain around 95% variance of the dataset. PCA also removed the multicollinearity from the dataset.

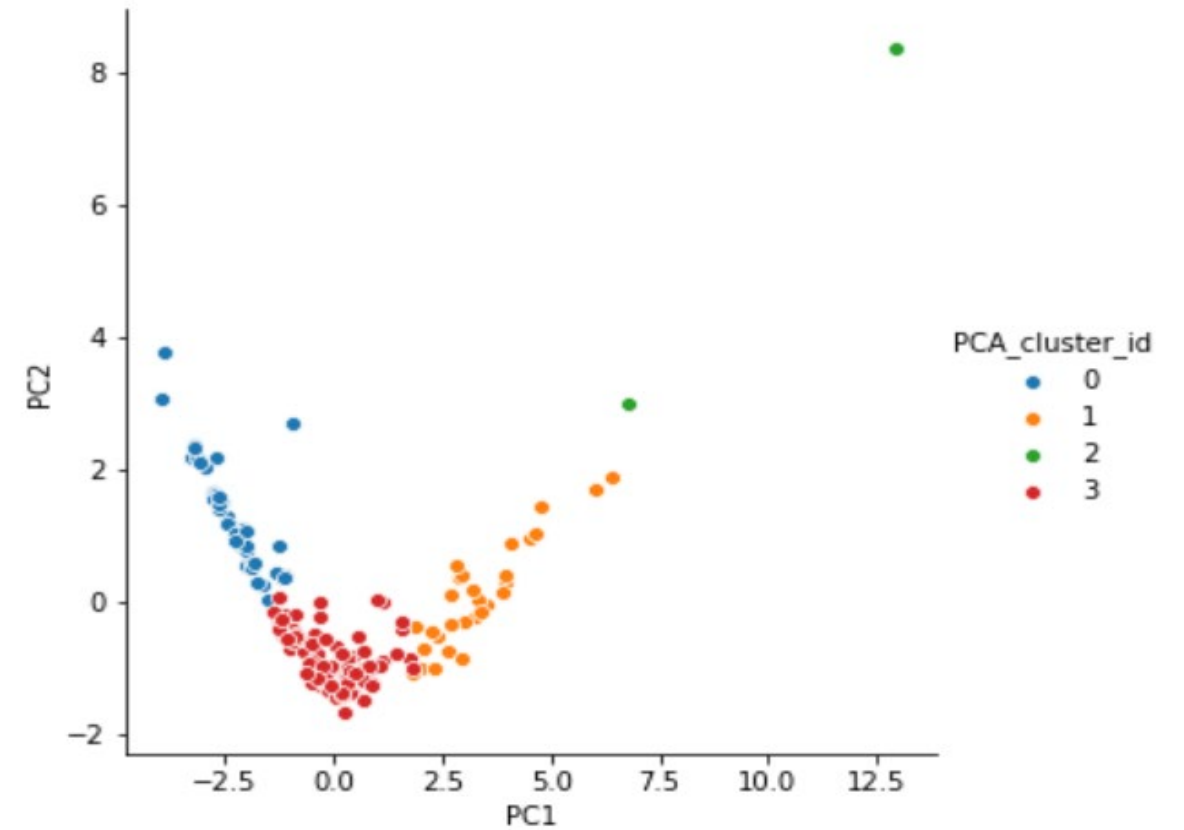
## **Results of Clustering**

Clustering techniques were able to segregate developed, underdeveloped, and developing countries. It clearly identified the number of countries which are in dire need of Aid. Hopkins statistics also showed that data has high tendency to cluster. KMeans clustering technique was chosen over Hierarchical technique, as KMeans clusters were more correctly segregated.

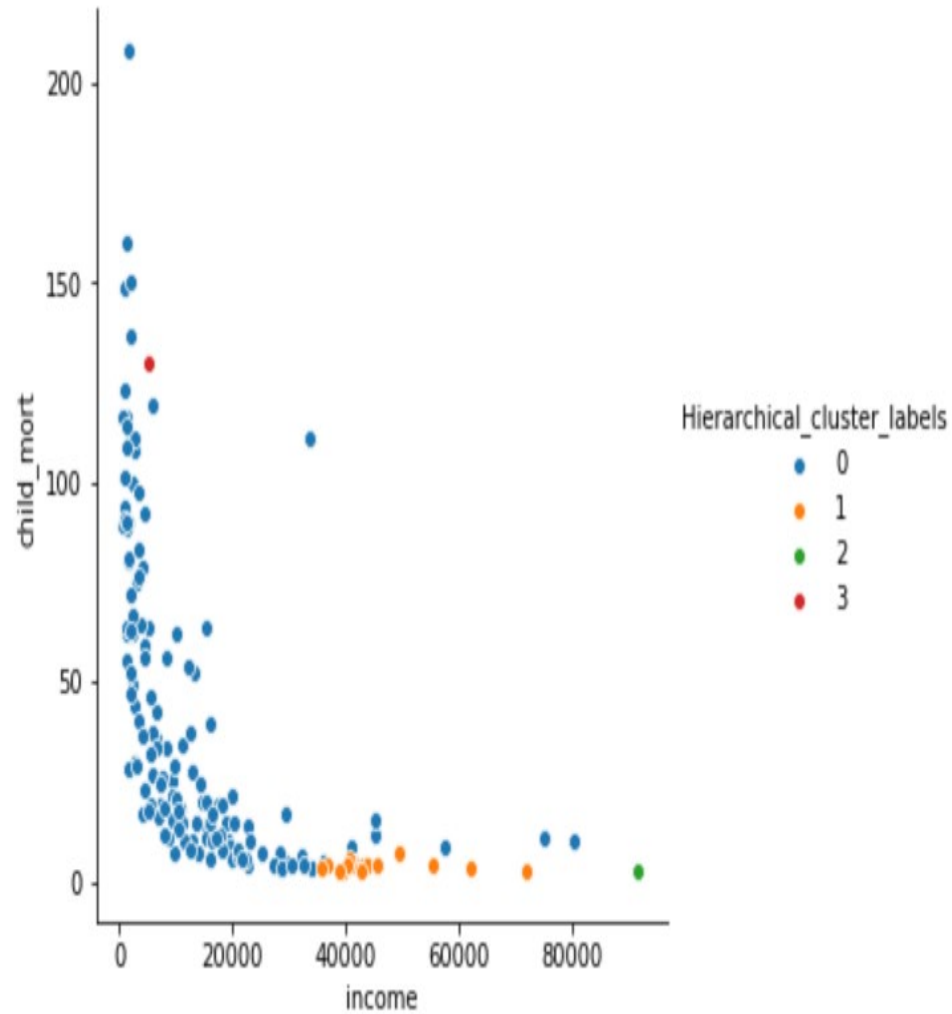
## Original visualization of PCA dataset



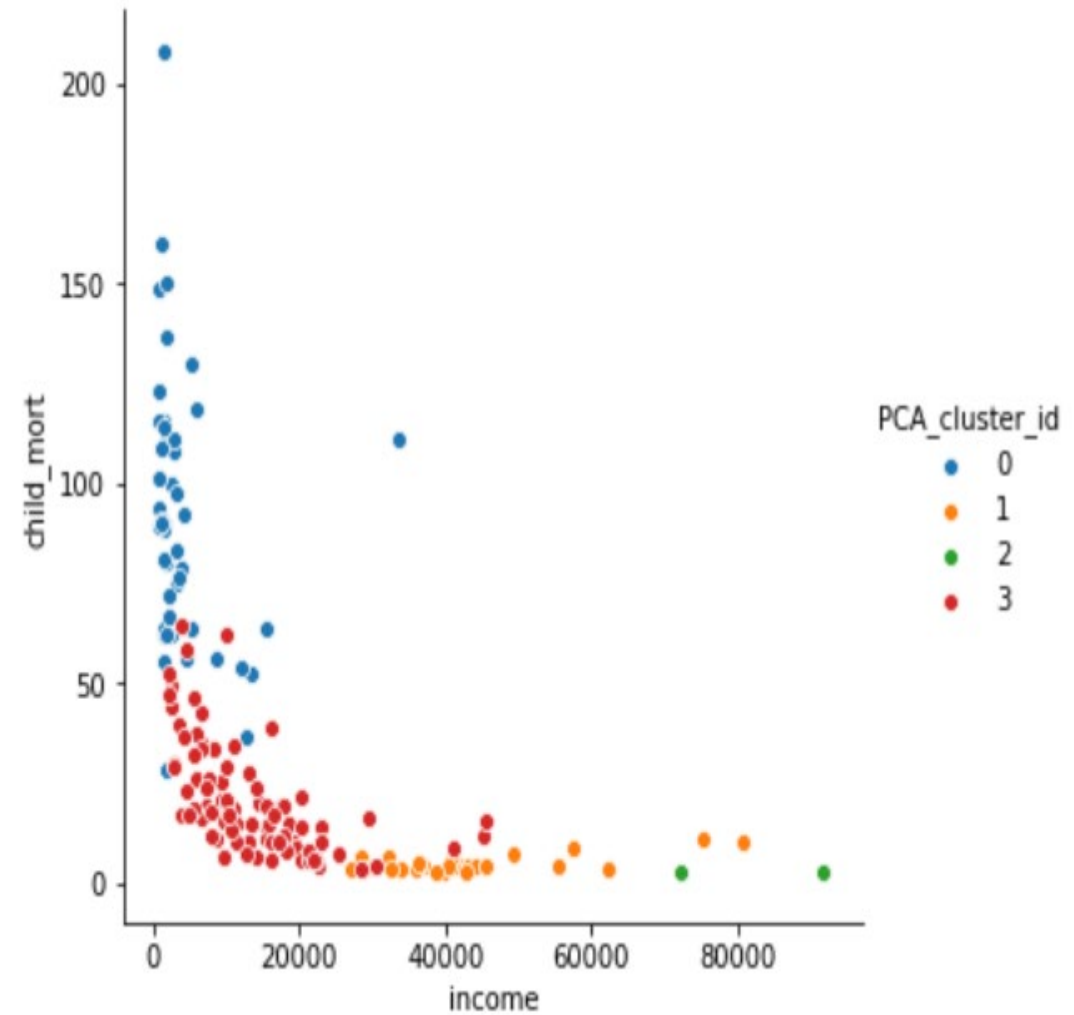
## Visualization after clustering



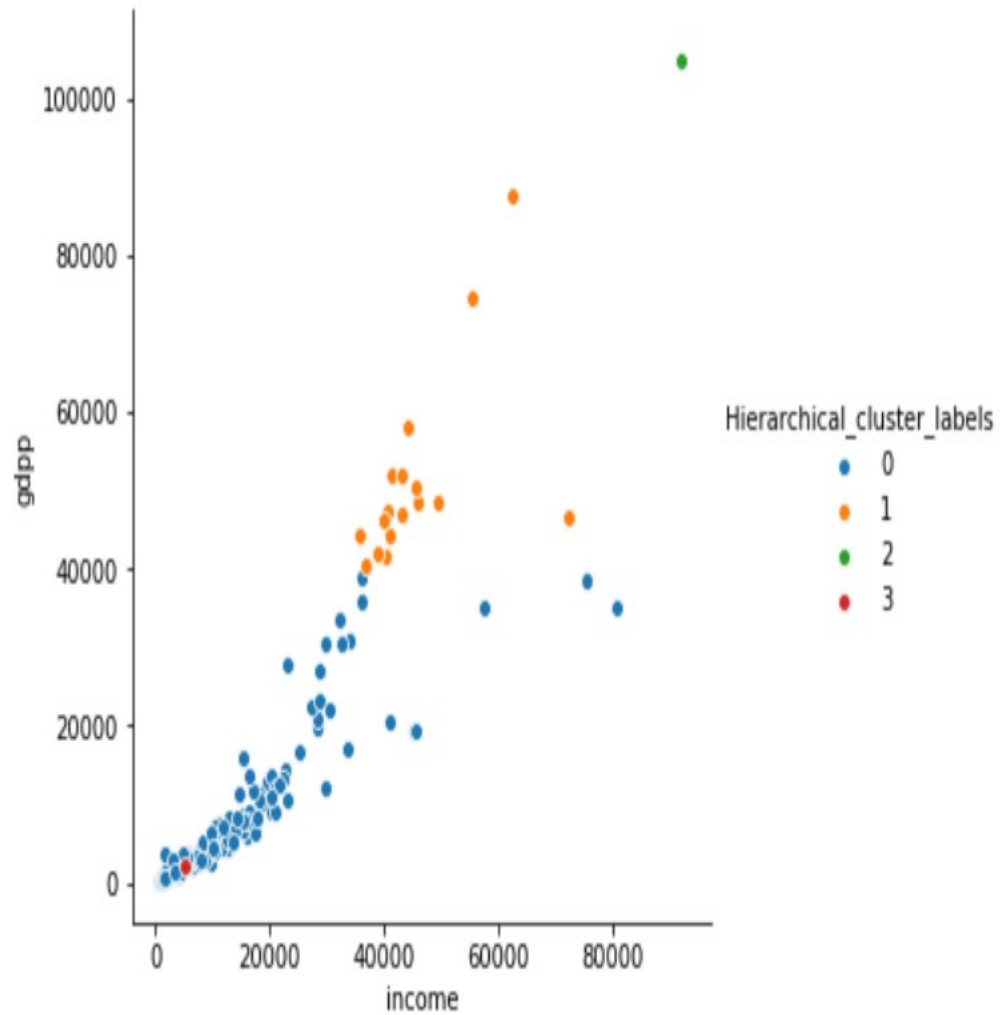
## Hierarchical clustering visualization



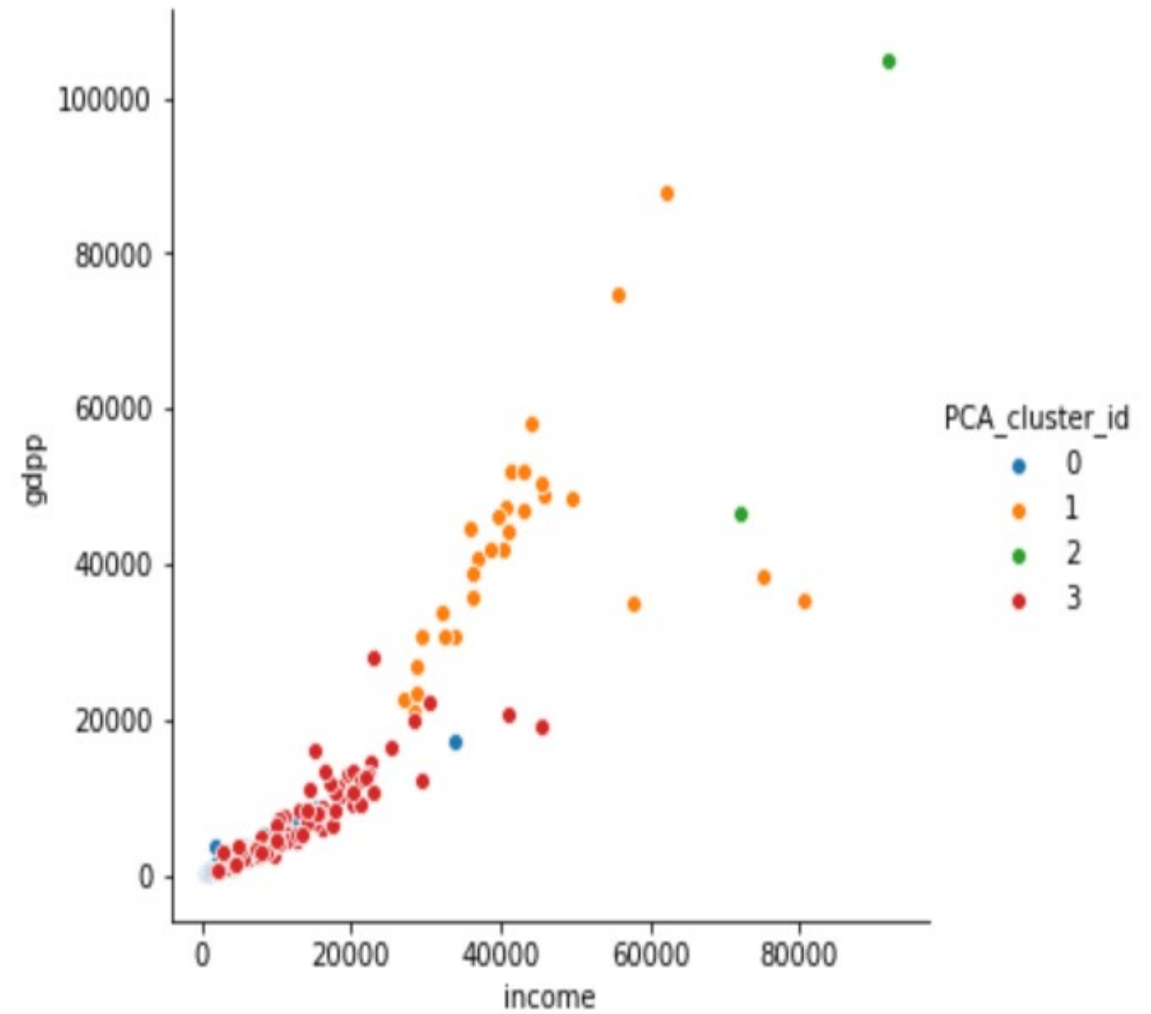
## PCA clustering visualization



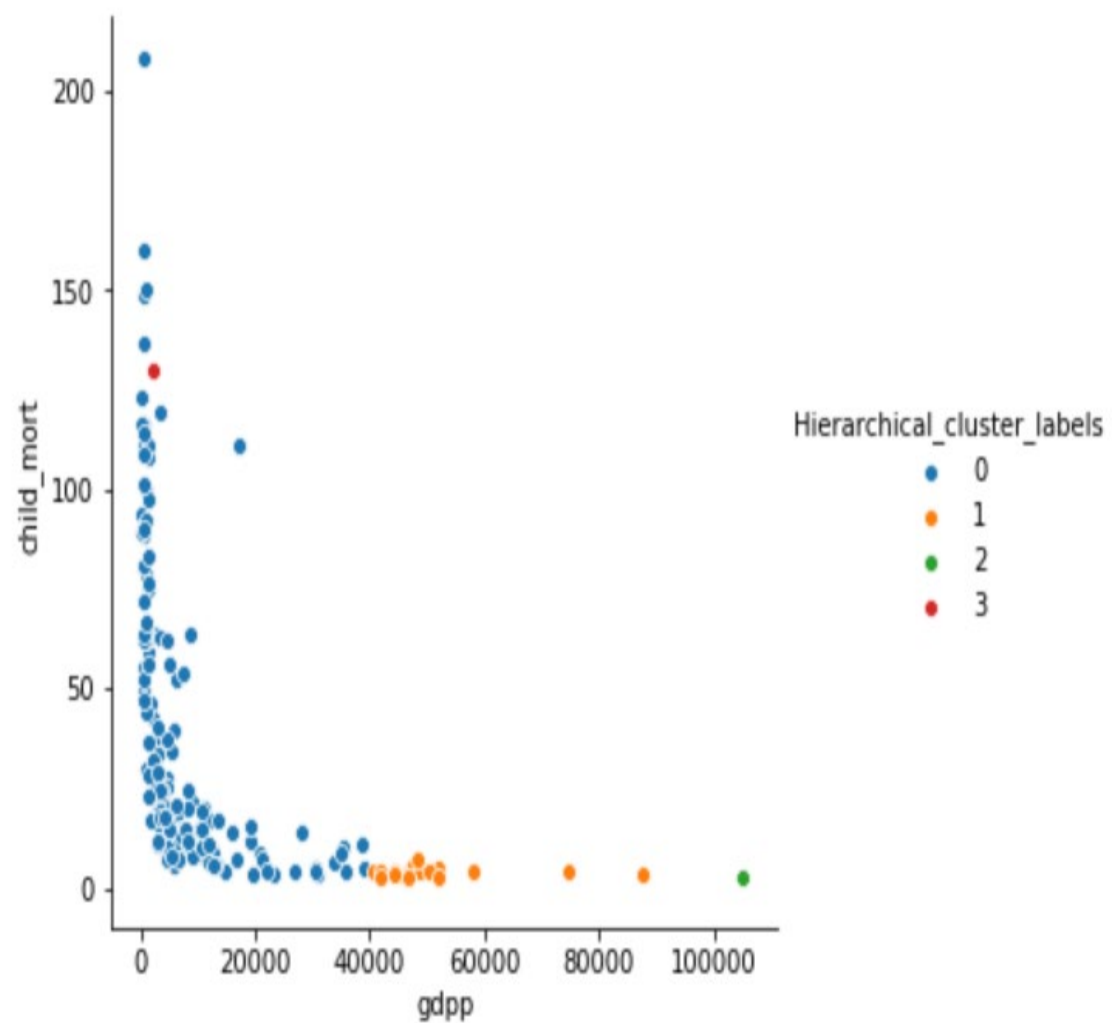
## Hierarchical clustering visualization



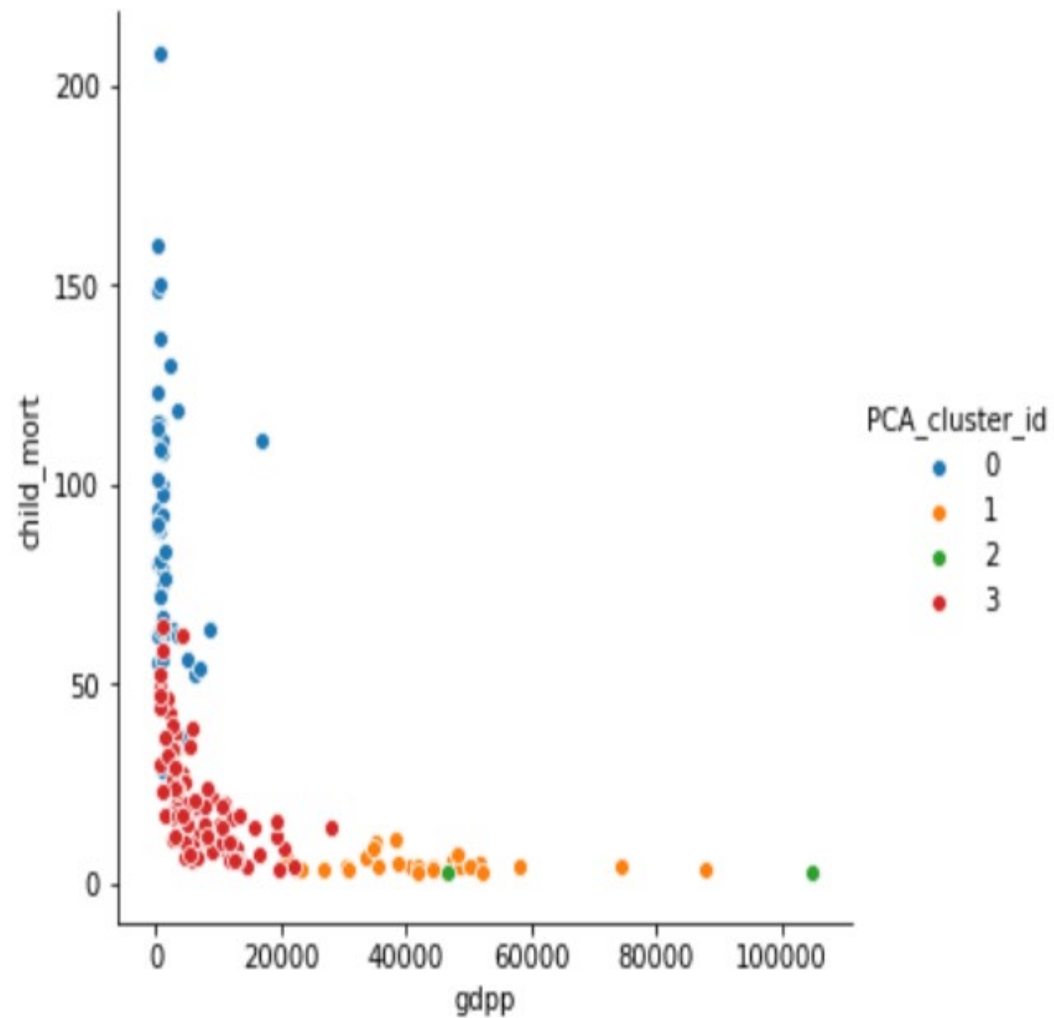
## PCA clustering visualization



## Hierarchical clustering visualization



## PCA clustering visualization



## Important results

Clustering techniques along with PCA were able to create 4 clusters of countries from the given dataset. PCA dataset was able to explain around 95% variance of the dataset, even with less number of columns. KMeans technique was chosen over Hierarchical over the type of clusters formed.

Following observations were noted after analyzing clusters formed by KMeans:

- 1) PCA cluster 0 contains poor countries which are in need of aid
- 2) PCA cluster 1 majorly contains developed countries having high living standards
- 3) PCA cluster 2 contains countries with high imports, and exports
- 4) PCA cluster 3 contains developing countries

## List of countries in need of Aid (In Decreasing order of severity)

1) Haiti	7) Niger	13) Cote d'Ivoire	19) Mauritania	25) Liberia	31) Sudan	37) Rwanda	43) Namibia
2) Sierra Leone	8) Angola	14) Equatorial Guinea	20) Burundi	26) Comoros	32) Ghana	38) Kiribati	44) Eritrea
3) Chad	9) Congo, Dem. Rep.	15) Guinea	21) Pakistan	27) Zambia	33) Tanzania	39) Timor-Leste	45) South Africa
4) Central African Republic	10) Burkina Faso	16) Cameroon	22) Malawi	28) Uganda	34) Senegal	40) Madagascar	46) Botswana
5) Mali	11) Guinea-Bissau	17) Mozambique	23) Togo	29) Gambia	35) Congo, Rep.	41) Kenya	47) Iraq
6) Nigeria	12) Benin	18) Lesotho	24) Afghanistan	30) Lao	36) Gabon	42) Yemen	48) Solomon Islands