

HELP International - NGO

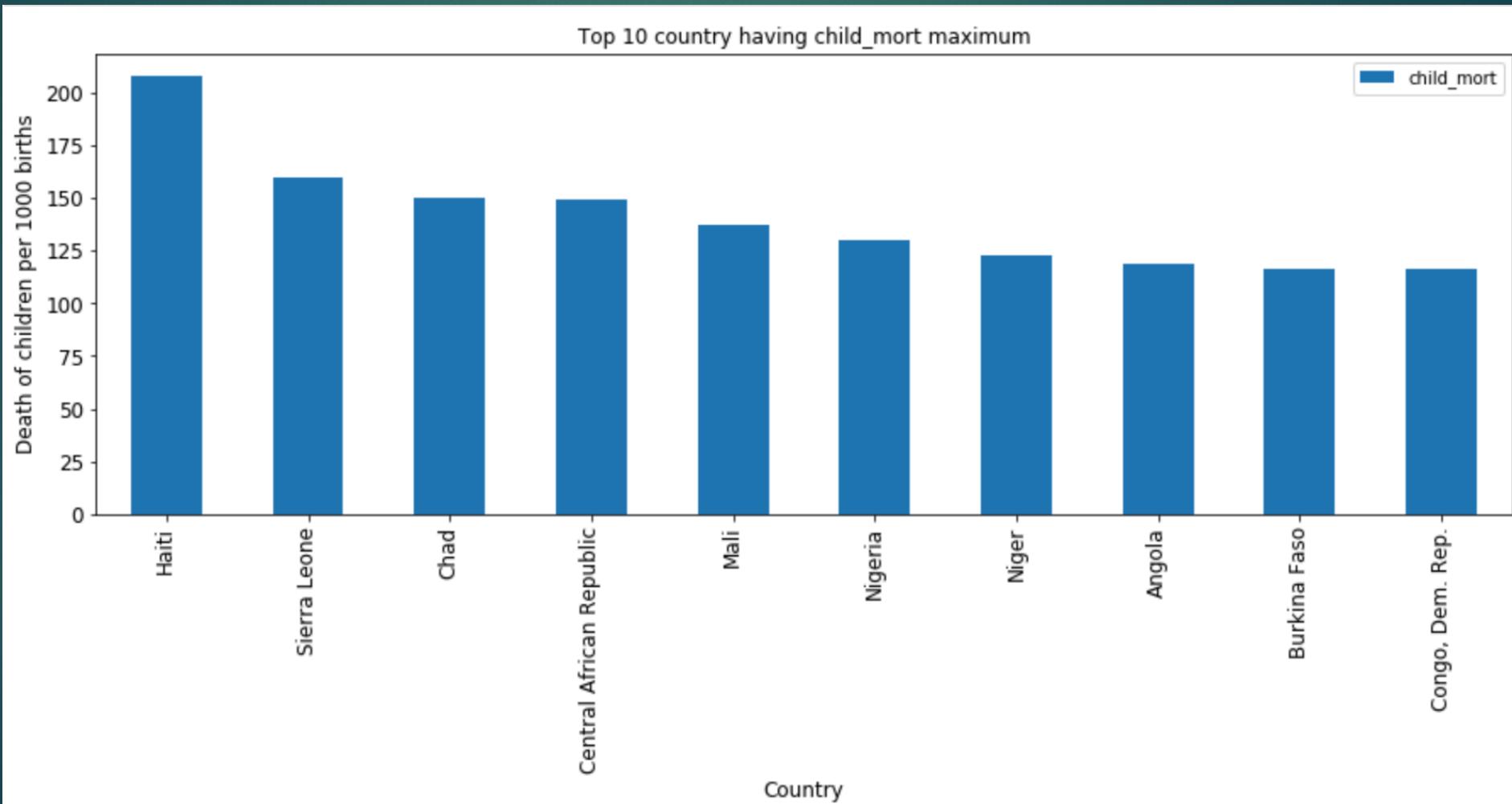
PRADEEP SAIN

- ▶ We have the dataset for all the countries that has below features for each country, Depending upon these factors we need to decide that which all countries are in direst need of aid.
 1. Child_mort - Death of children under 5 years of age per 1000 live births
 2. Exports - Exports of goods and services. Given as %age of the Total GDP
 3. Health - Total health spending as %age of Total GDP
 4. Imports - Imports of goods and services. Given as %age of the Total GDP
 5. Income - Net income per person
 6. Inflation - The measurement of the annual growth rate of the Total GDP
 7. Life_expec - The average number of years a new born child would live if the current mortality patterns are to remain the same.
 8. total_fer - The number of children that would be born to each woman if the current age-fertility rates remain the same.
 9. Gdpp - The GDP per capita. Calculated as the Total GDP divided by the total population.

- ▶ Depending upon given features and the domain knowledge we can see that below features are the major criteria to decide upon the aid granted to which country. We will be analysing the dataset on all the features, but at the end after segmenting the countries we will decide upon below major factors.
1. Child_mort - Death of children under 5 years of age per 1000 live births
 2. Income - Net income per person
 3. Inflation - The measurement of the annual growth rate of the Total GDP
 4. Gdpp - The GDP per capita. Calculated as the Total GDP divided by the total population.

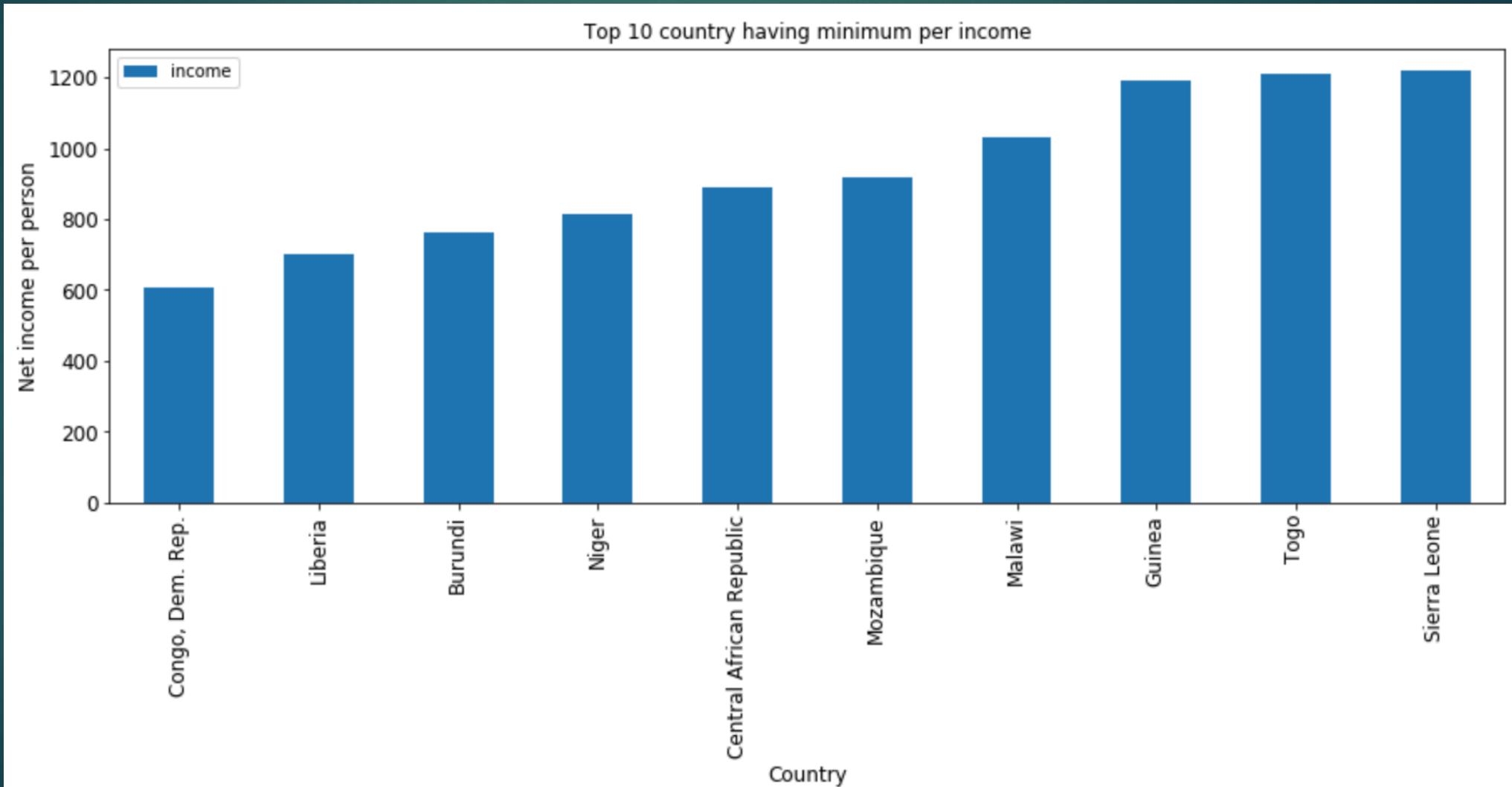
HELP International - Analysing given dataset.

- ▶ **Child Mort** - Below are the top 10 countries, which has child_mort (Death of children under 5 years of age per 1000 live births) maximum number, that means country which has maximum child death would be requiring more funds to sustain.



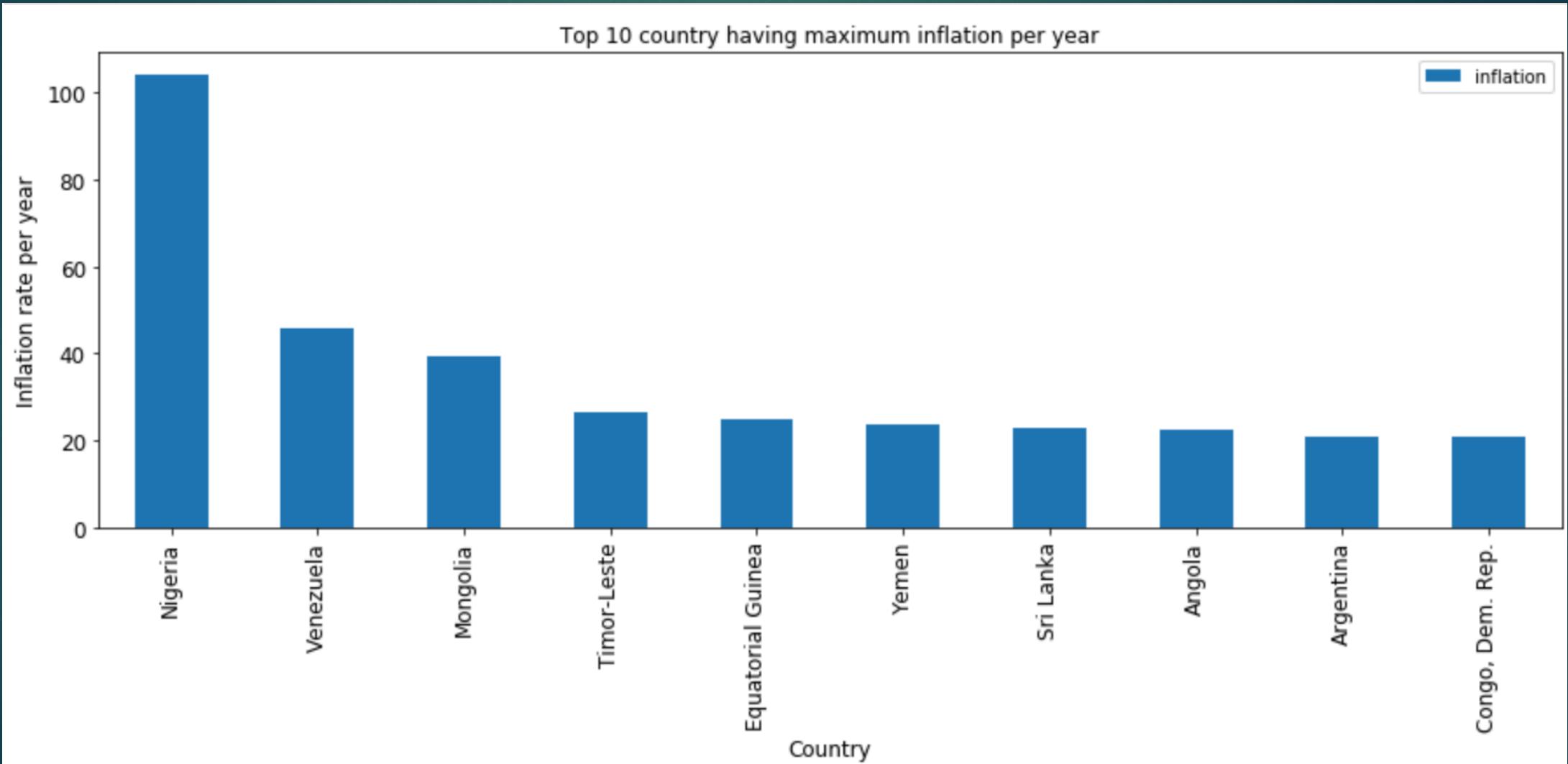
HELP International - Analysing given dataset

- ▶ **Income** - Below are the top 10 countries, which has net income per person very minimum, that means country which has minimum income would be requiring more funds to sustain.



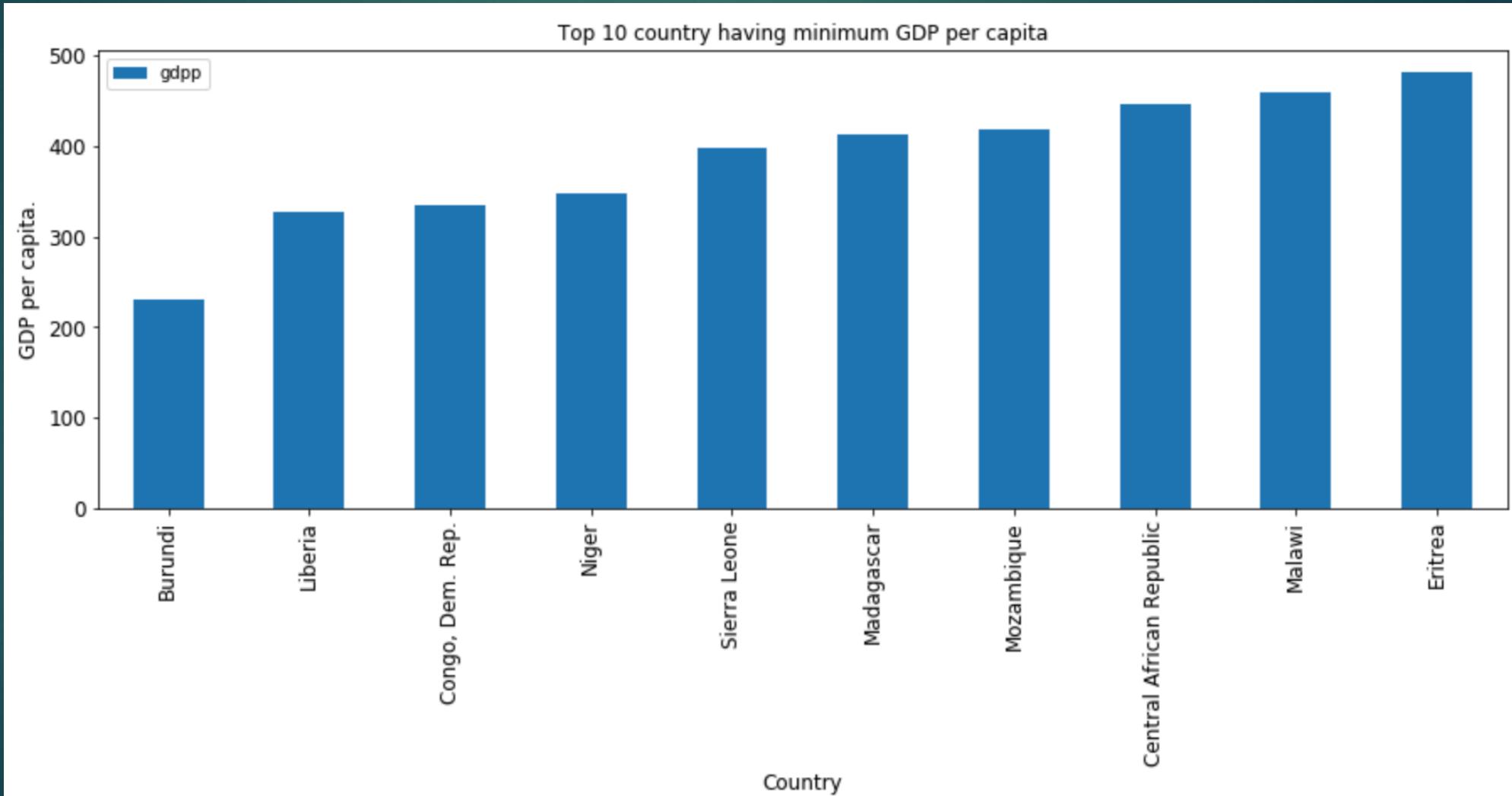
HELP International - Analysing given dataset

- ▶ **Inflation** - Below are the top 10 countries, which has inflation rate per year very high, that means country which has maximum inflation would be requiring more funds to sustain.



HELP International - Analysing given dataset

- ▶ **GDP per capita** - Below are the top 10 countries, which has GDP per capita per year very low, that means country which has minimum GDP per capita would be requiring more funds to sustain.



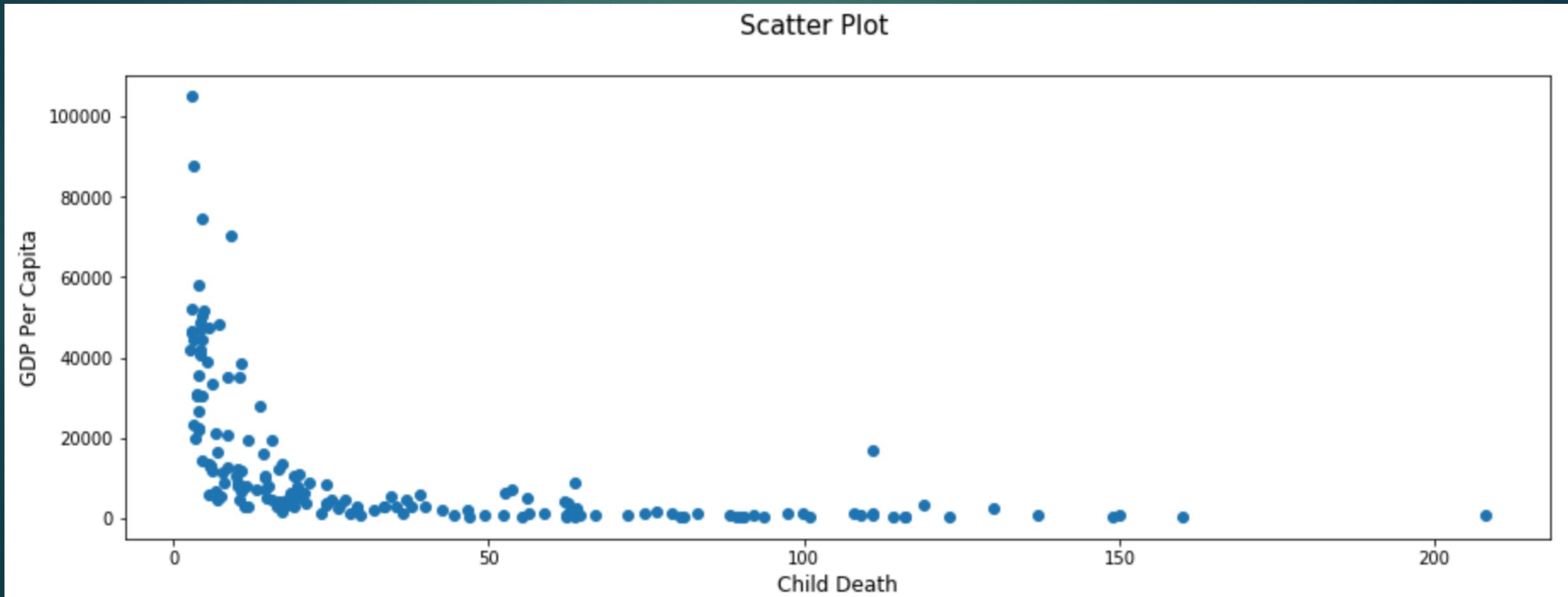
HELP International - Analysing given dataset



- ▶ Depending upon above analysis we can find common countries which has maximum child death, minimum income, maximum inflation and minimum GDP per capita. But that is not the right way to conduct analysis. And more importantly we need to consider all the features hence we have to segment these countries into different cluster depending upon their features values and as we need to include all the features to conduct clustering we will be first doing PCA then clustering on the result of the PCA.
- ▶ So we get the different clusters and on these resulted cluster we need to decide that which country requires the funds depending upon our analysis.

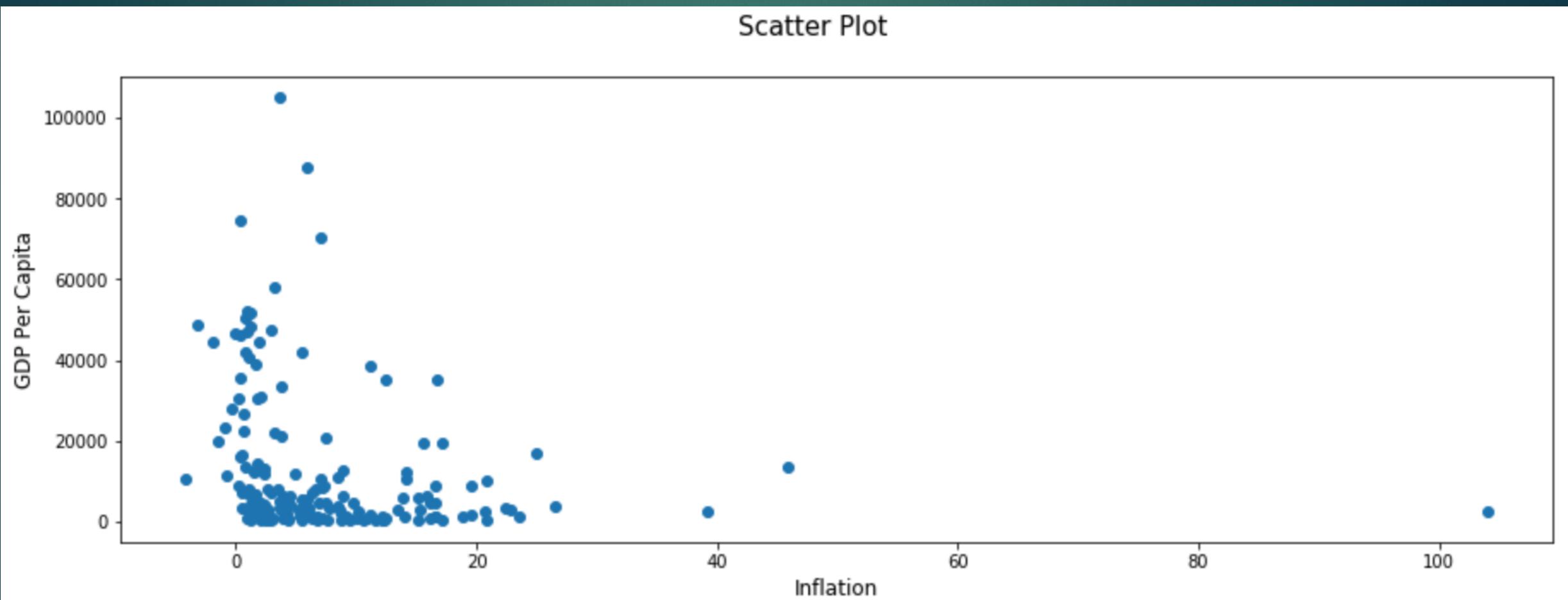
BEFORE PCA - Relationship B/W 2 datapoint

- ▶ **GDP vs Child Death** - Below is the relationship for GDP to Child death for all the countries, before conducting PC Analysis.



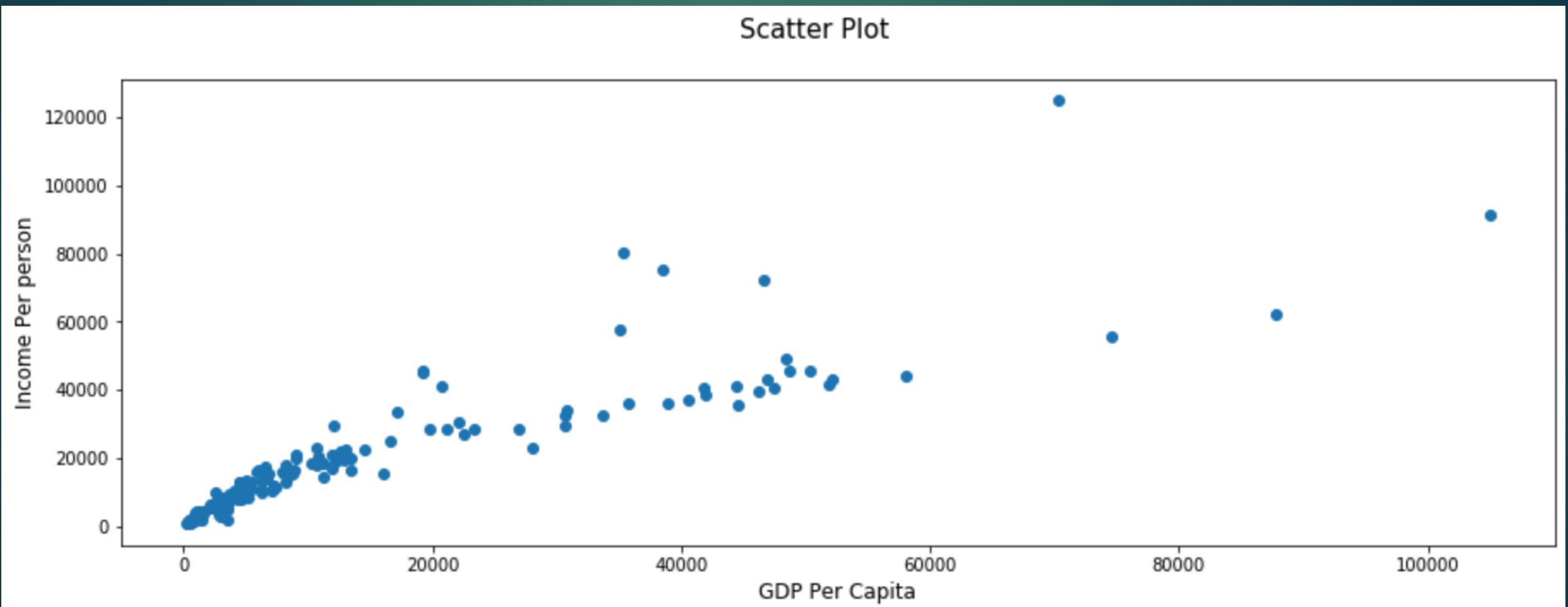
BEFORE PCA - Relationship B/W 2 datapoint

- ▶ **GDP vs Inflation** - Below is the relationship for GDP to Inflation for all the countries, before conducting PC Analysis.



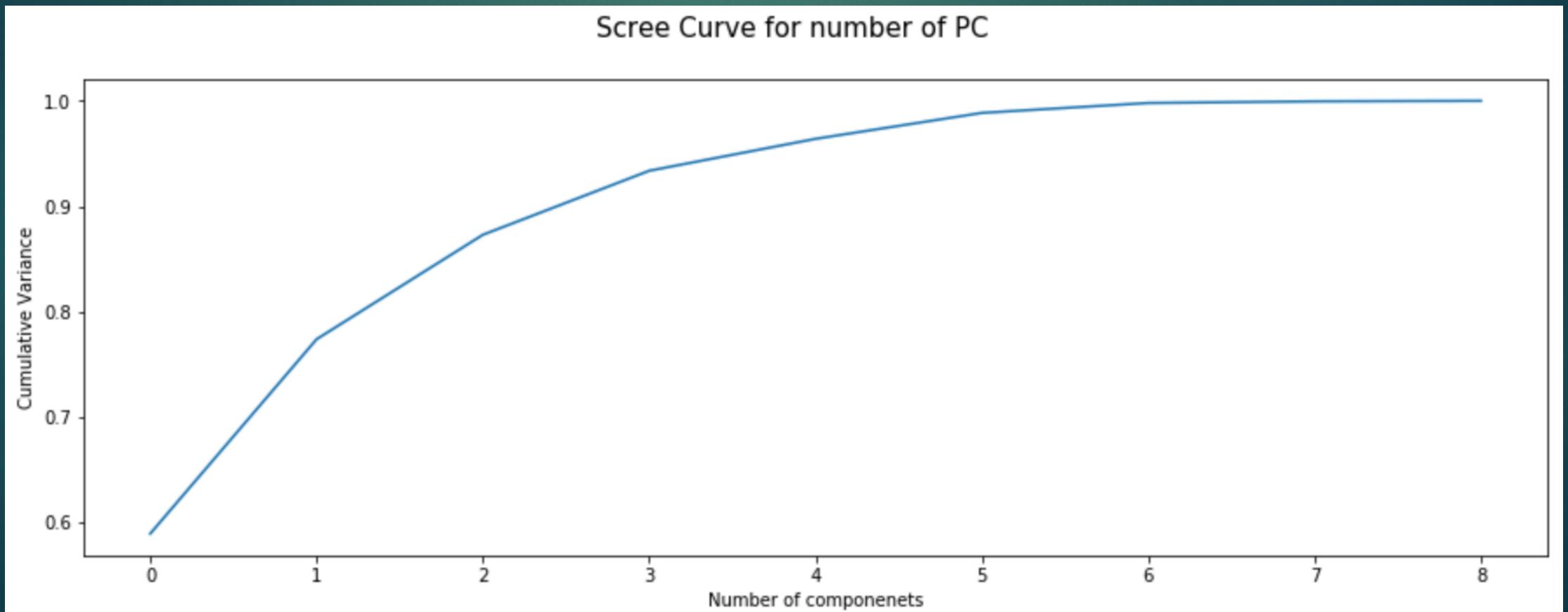
BEFORE PCA - Relationship B/W 2 datapoint

- ▶ **GDP vs Income** - Below is the relationship for GDP to Income for all the countries, before conducting PC Analysis.



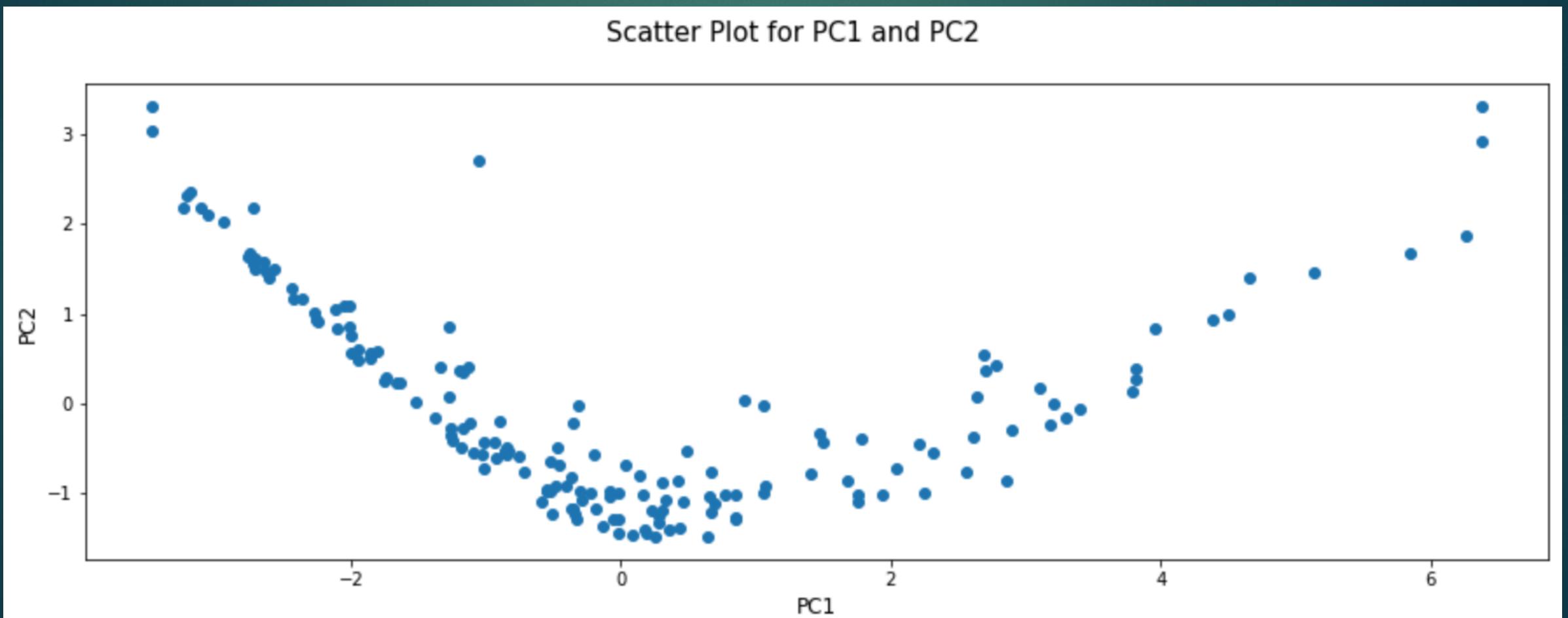
PERFORM PCA

- Now we apply PCA algorithm to reduce the number of features, having more than 90% covariance. After doing PCA we will get below scree plot, By looking at plot we can decide upon the number of minimum feature that we need to consider for our further analysis which cover minimum of 90% variance. So by below curve we can say that we can go with 4 principle component.



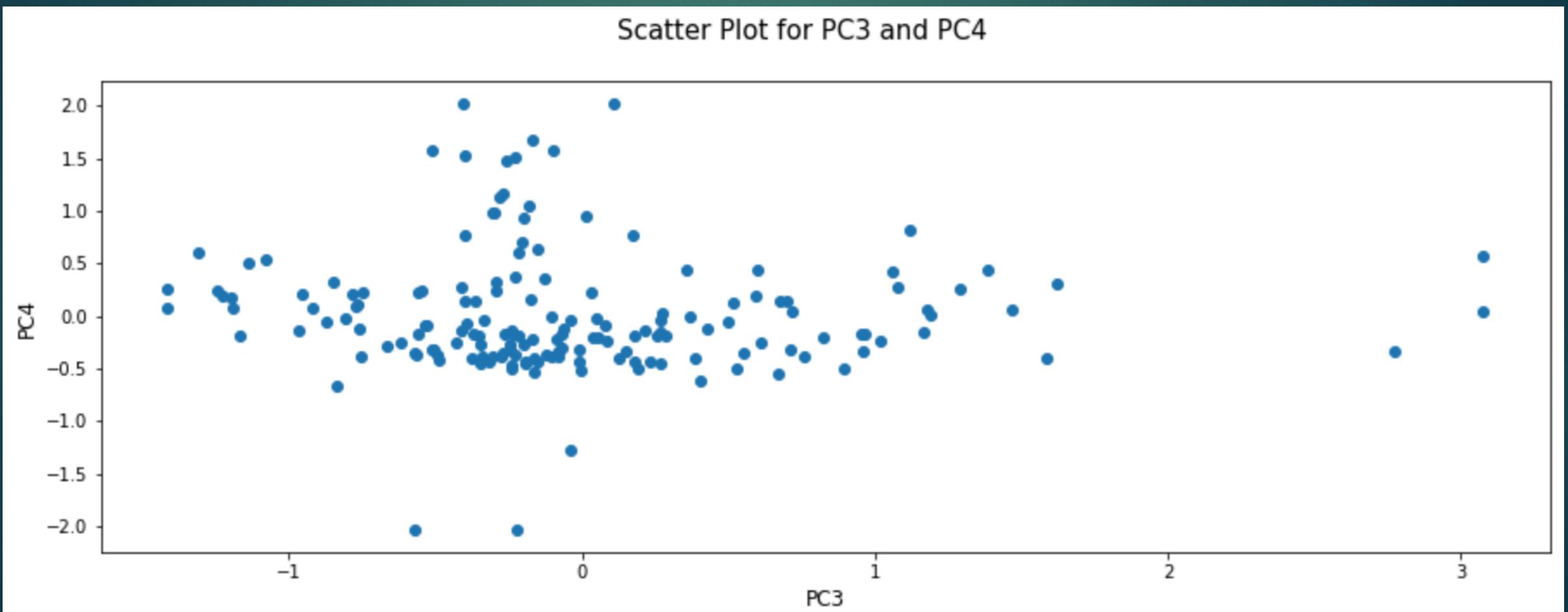
AFTER PCA -

- ▶ So after conducting PCA, We get 4 beautiful and minimum feature which define 95% of variance for the complete dataset. Where $PC1 > PC2 > PC3 > PC4$.
- ▶ Relationship between PC1 and PC2 can be seen below.



AFTER PCA -

- ▶ So after conducting PCA, We get 4 beautiful and minimum feature which define 95% of variance for the complete dataset. Where $PC1 > PC2 > PC3 > PC4$.
- ▶ Relationship between $PC3$ and $PC4$ can be seen below.

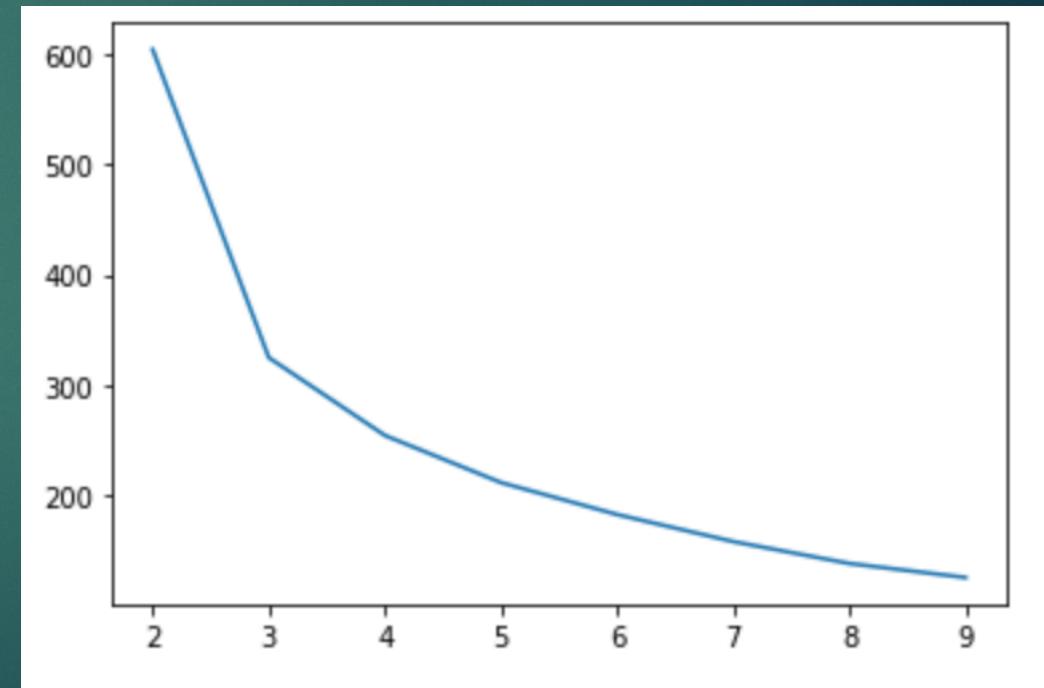


- ▶ So after analysing above two scatter plot we can not actually decide that how many clusters can be formed by these.
- ▶ So we will apply clustering algorithm to decide that how many cluster can be formed. Below are well know clustering algorithm that will tell us how many optimal cluster can be formed.
 - K-Means
 - Hierarchical
 - Single Linkage
 - Complete Linkage

Clustering -

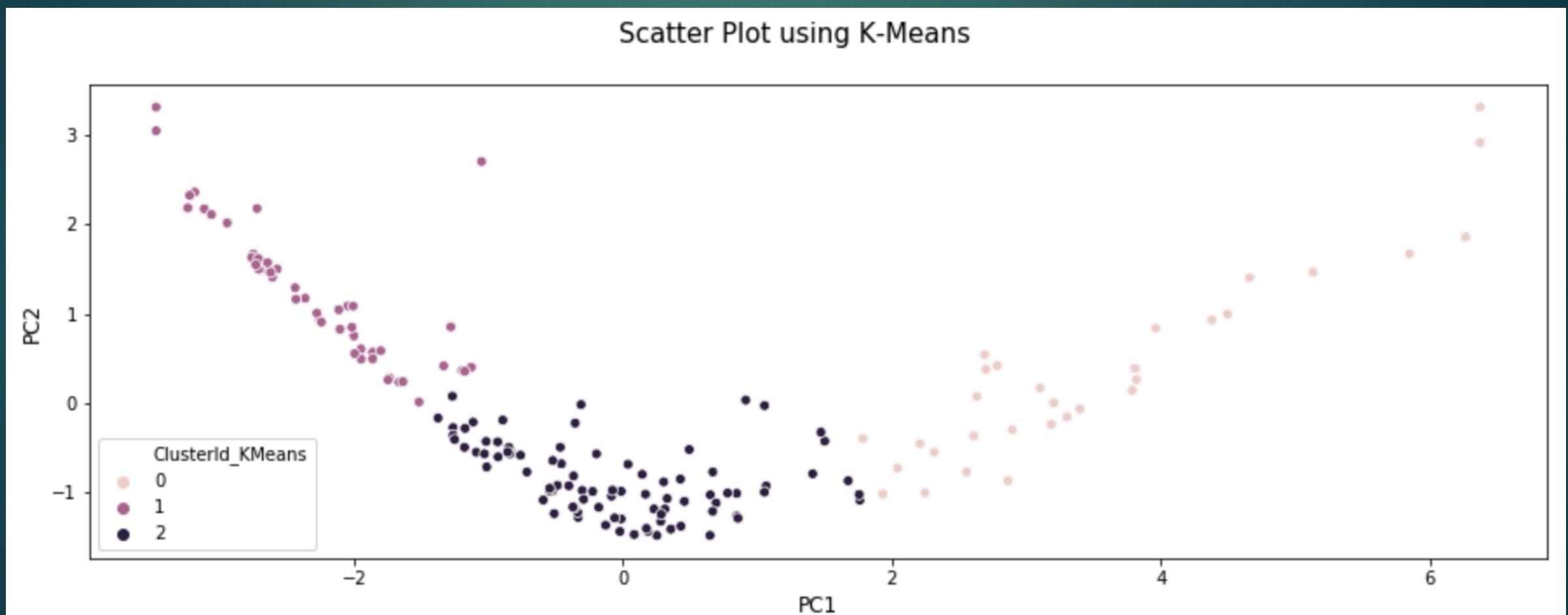
- ▶ So on the basis of silhouette score and elbow curve we will decide that how many cluster we need to take to get the optimal and good clusters.
- ▶ Below are the silhouette score and the elbow curve.
- ▶ So based on the below analysis we can say that we can take number of cluster as 3, as it has maximum silhouette score and elbow at 3.

```
For n_clusters=2, the silhouette score is 0.4880872677284147
For n_clusters=3, the silhouette score is 0.48381847399132566
For n_clusters=4, the silhouette score is 0.4626187378318367
For n_clusters=5, the silhouette score is 0.3813712886396705
For n_clusters=6, the silhouette score is 0.37785382183091437
For n_clusters=7, the silhouette score is 0.386640499924541
For n_clusters=8, the silhouette score is 0.38228625569937935
```



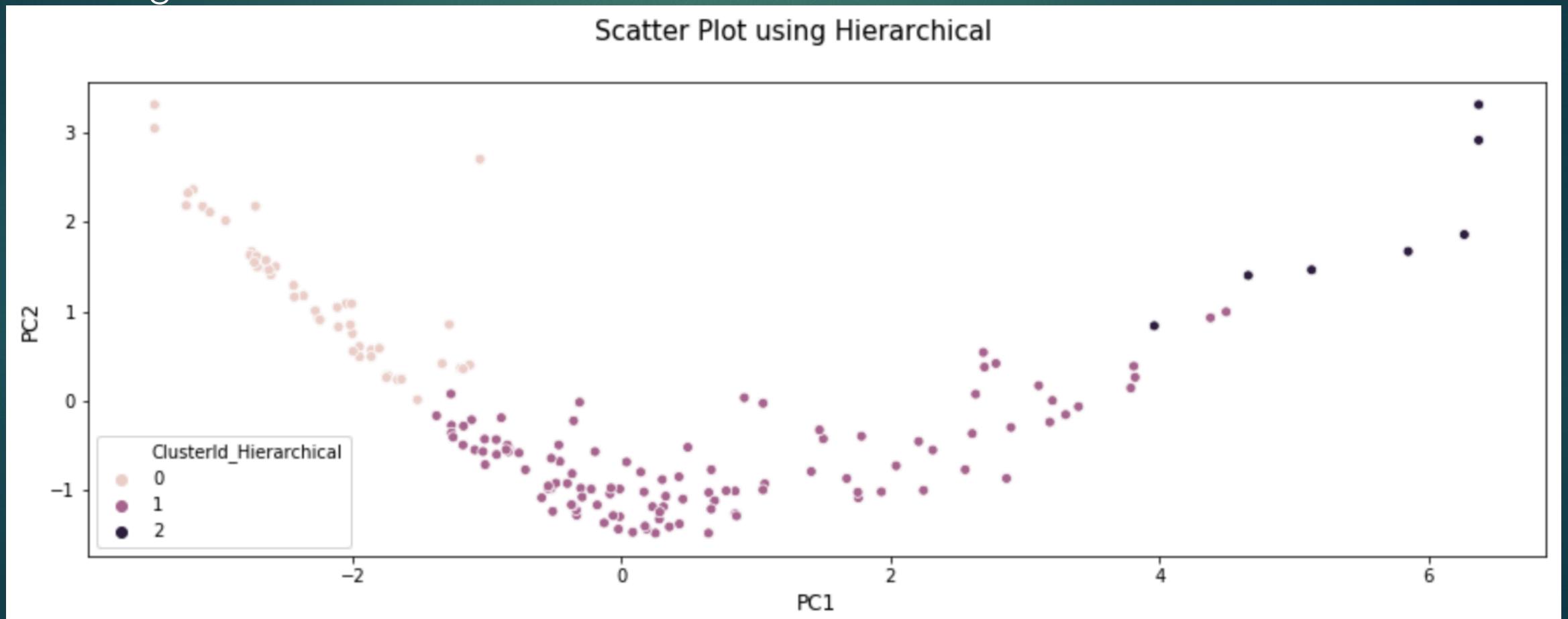
K-Means Clustering -

- When conducting K-Means with k=3, All the clusters looks good and can be seen below and seems not randomly distributed.



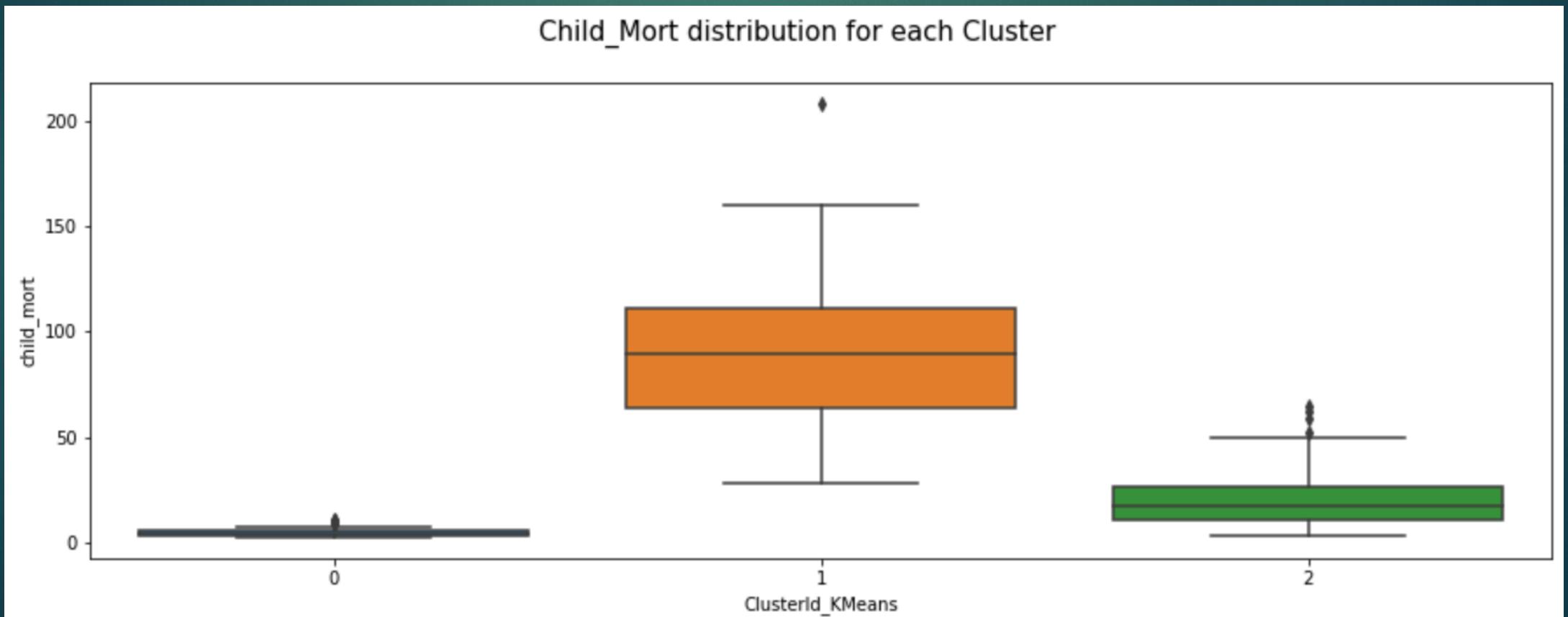
Hierarchical Clustering -

- When conducting Hierarchical clustering with k=3, Clusters can be found more randomly distributed. And it can be seen that cluster 1 has almost 112 countries and cluster 2 has only 7 countries, so we can see that clusters are not equally distributed. Hence we will go with K-Means algorithm.



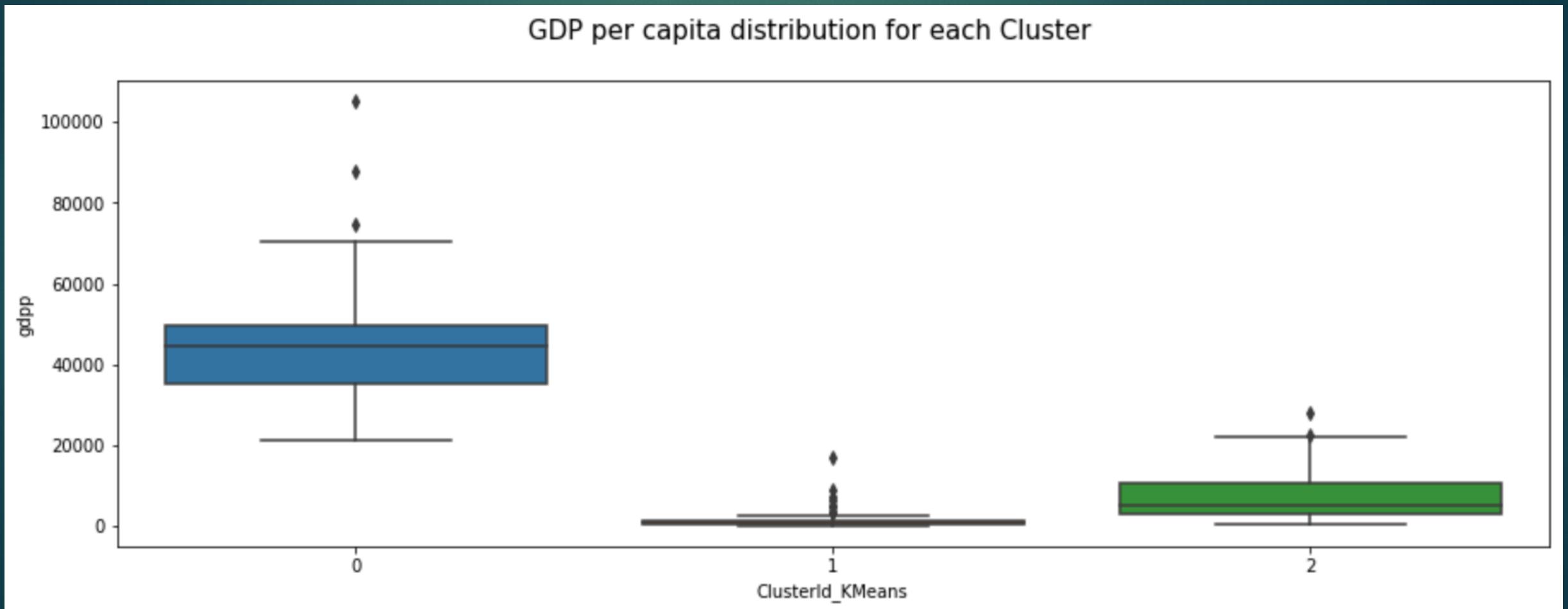
Analysing Clusters -

- ▶ Using k-means we have created 3 clusters and we will analyse them on their features, So for that we will create box plot and will see which cluster has the min or max distribution of the feature.
- ▶ From below we can clearly see that child death ratios are way high for cluster 1, hence cluster 1 may require more funds to sustain.



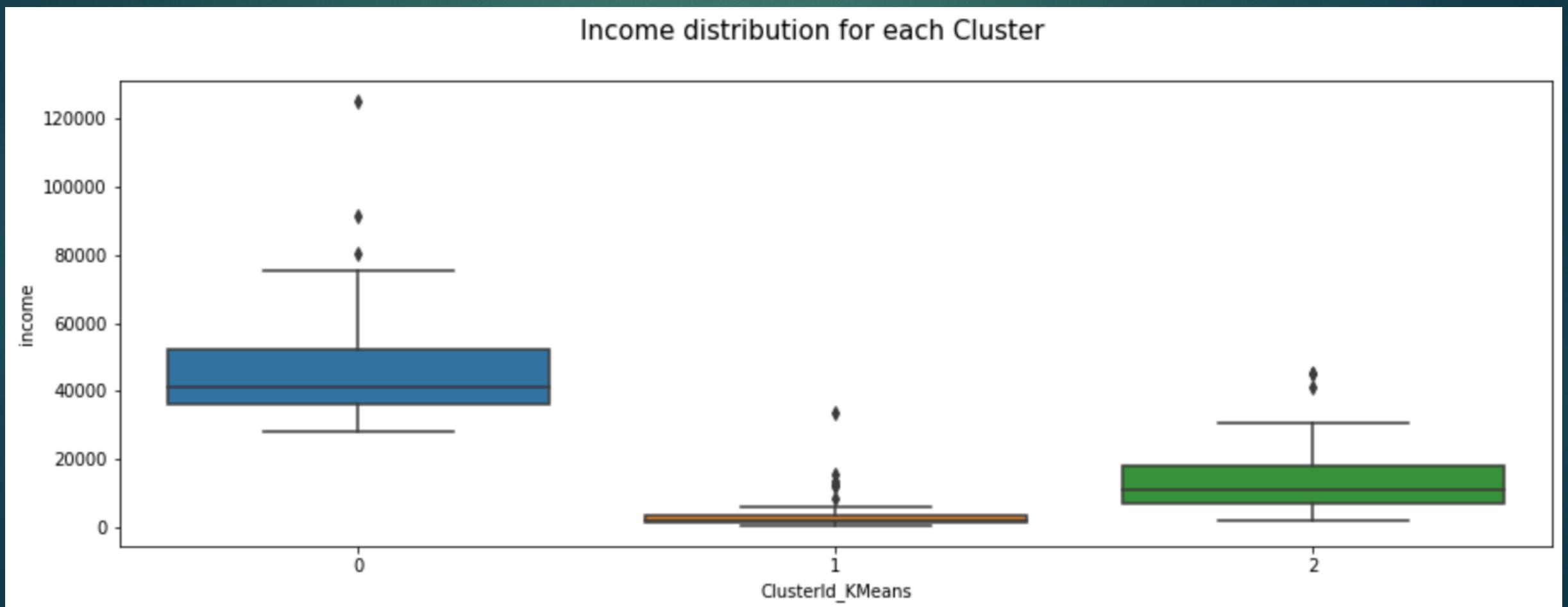
Analysing Clusters -

- ▶ From below plot we can clearly see that GDP per capita are very low for cluster 1, hence cluster 1 may require more funds to sustain.



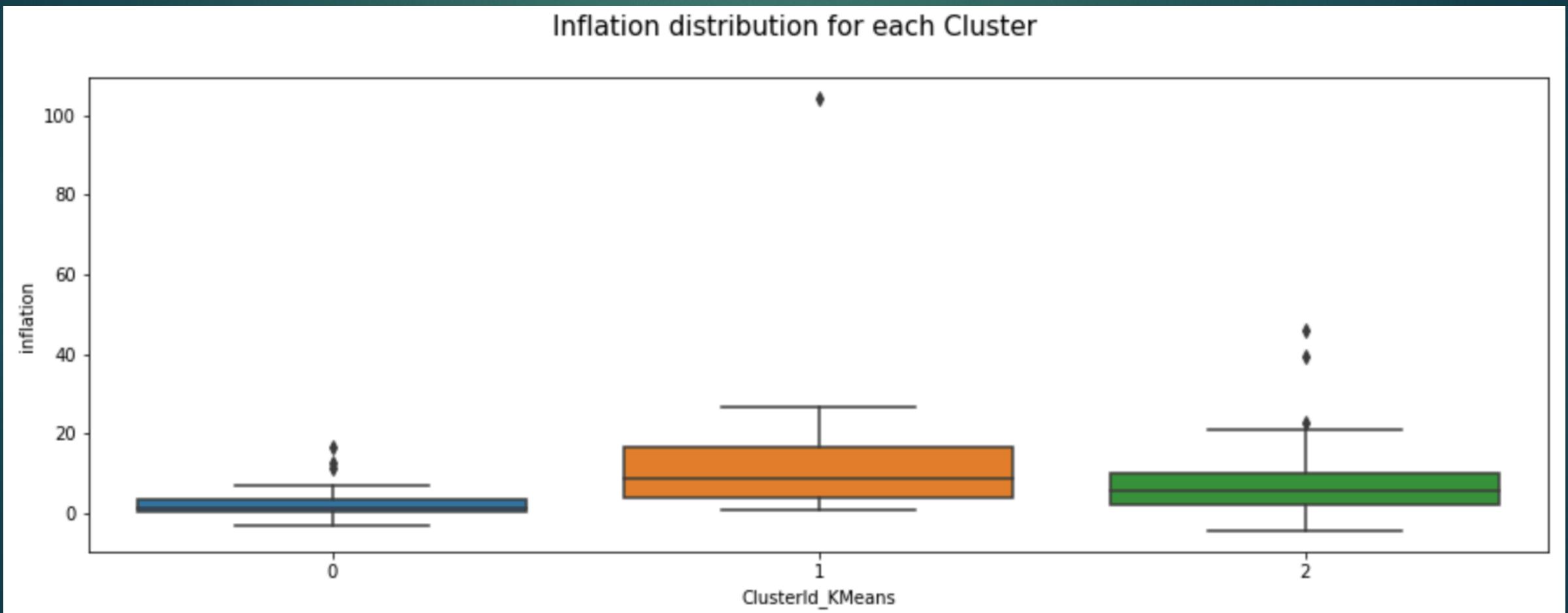
Analysing Clusters -

- ▶ From below plot we can clearly see that Income per person is very low for cluster 1, hence cluster 1 may require more funds to sustain.



Analysing Clusters -

- ▶ From below plot we can clearly see that Inflation is very high for cluster 1, hence cluster 1 may require more funds to sustain.



Analysing Clusters -

- ▶ From below plot we can clearly see that Life Exception is very Low for cluster 1, hence cluster 1 may require more funds to sustain.



Analysing Clusters -

- ▶ From all above analysis we can see that cluster 2 has very low income rate, very high inflation, very low GDP per capita, high child death and very low life expectation.
- ▶ By this analysis we can conclude that countries which falls under cluster require more funds and is in direst need of aid.
- ▶ So by our analysis we get total 48 countries which falls under cluster 1.
- ▶ So by sorting these countries on the basis of low income rate, high inflation, low GDP and high child_mort, we get the countries which need more funds and are more suitable for our organisation to funds them.
- ▶ In the next slide you will see the countries who are in direst need of aid in sorted.
- ▶ First country need more aid then second and so-on.
- ▶ So depending upon their GDP and income our organisation can decide on the amount they will gonna invest, Like how much they can invest on particular country.

Final Result - Country list who are direst need of aid.



- So from our analysis we can say that below 10 countries are in direst need of need, So we decide upon how much budget we can spend into these country.

1. Haiti
2. Sierra Leone
3. Chad
4. Central African Republic
5. Mali
6. Nigeria
7. Niger
8. Angola
9. Congo, Dem. Rep.
10. Burkina Faso

Haiti
Sierra Leone
Chad
Central African Republic
Mali
Nigeria
Niger
Angola
Congo, Dem. Rep.
Burkina Faso
Guinea-Bissau
Benin
Cote d'Ivoire
Equatorial Guinea
Guinea
Cameroon
Mozambique
Lesotho
Mauritania
Burundi
Pakistan

THANK - YOU