

# Clustering & PCA Assignment

Recommend countries to HELP International based on given data

# Business Objective

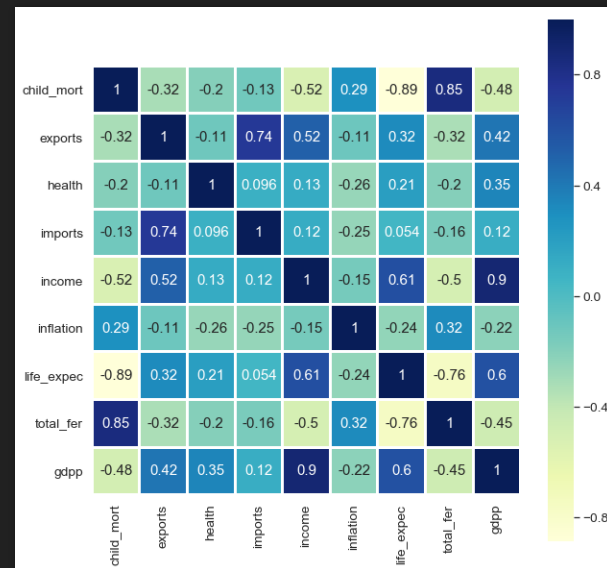
- We need to help the CEO of HELP International, an international humanitarian NGO, in the process of selection of countries they should focus on. The selection will be based on some socio-economic and health factors that determine the overall development of the country.
- In other words, it is required to cluster the countries by the factors mentioned above and then present our solution and recommendations to the CEO. We are also supposed to use dimensionality reduction using PCA to get the visualizations of the clusters in a 2-D form.

# Steps of Analysis

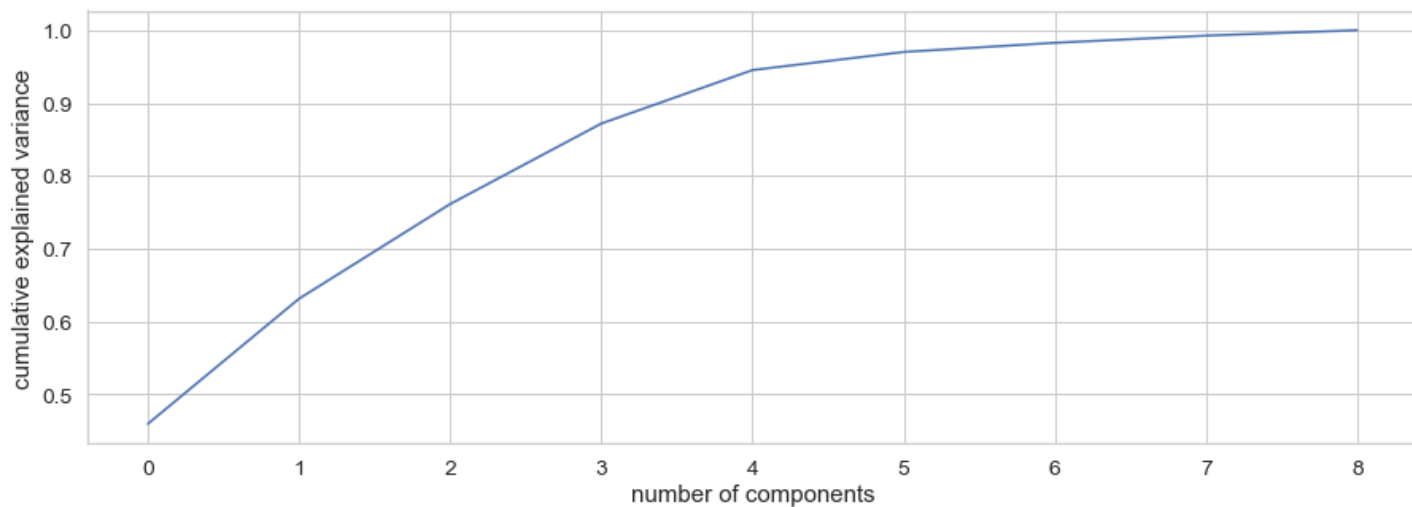
- Loading Dataset
- Cleaning Dataset
- Exploratory Data Analysis
- Data Preparation for PCA and Clustering
  - Feature Standardization or Normalization
- Selection of Principal Components using PCA
- Outlier Analysis and Treatment
- Visualizing Principal components
- Pre K-Means Clustering Analysis
  - Hopkins Analysis
  - Silhouette and Elbow Analysis
- K Means with K as 5
- K Means with K as 6
- K Means with K as 7
- Hierarchical Clustering Single Method
- Hierarchical Clustering Complete Method
- Combined list of countries from K-Means and Hierarchical Clustering
- Analysis of Outlier that were dropped
- Combining, Visualizing and Summarizing Results

# Co-relation of Original Columns

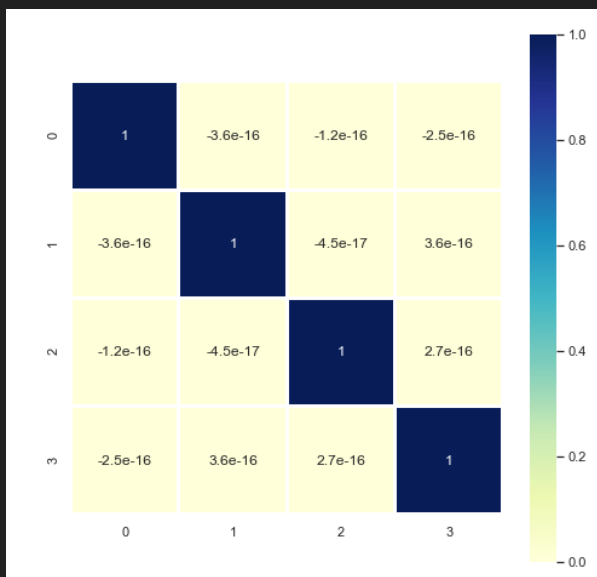
- We can say that a lot of original columns are actually co-related



# Selection of Principal Components using PCA

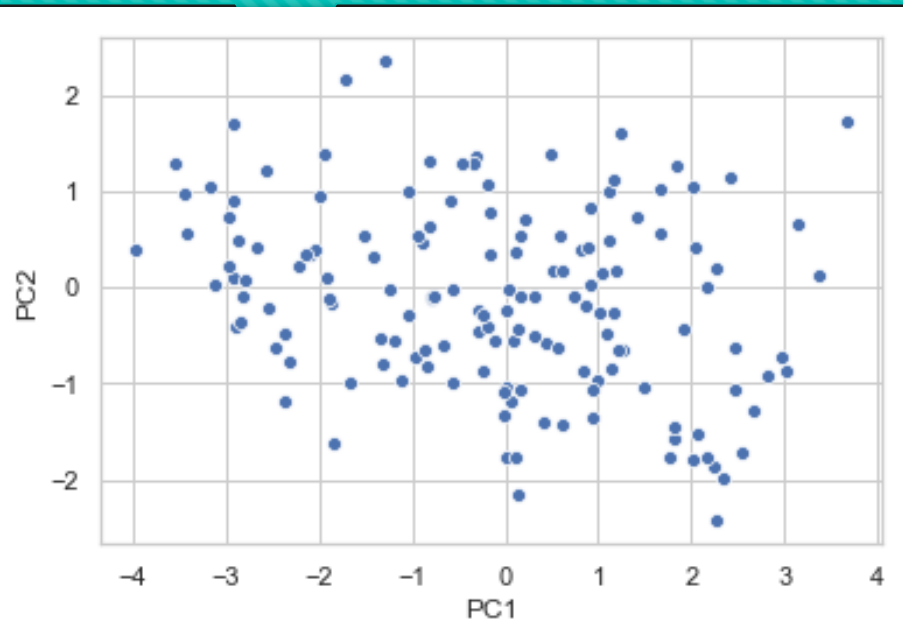


- Selecting 4 as number of Principal components since it explains ~ 95% variance.



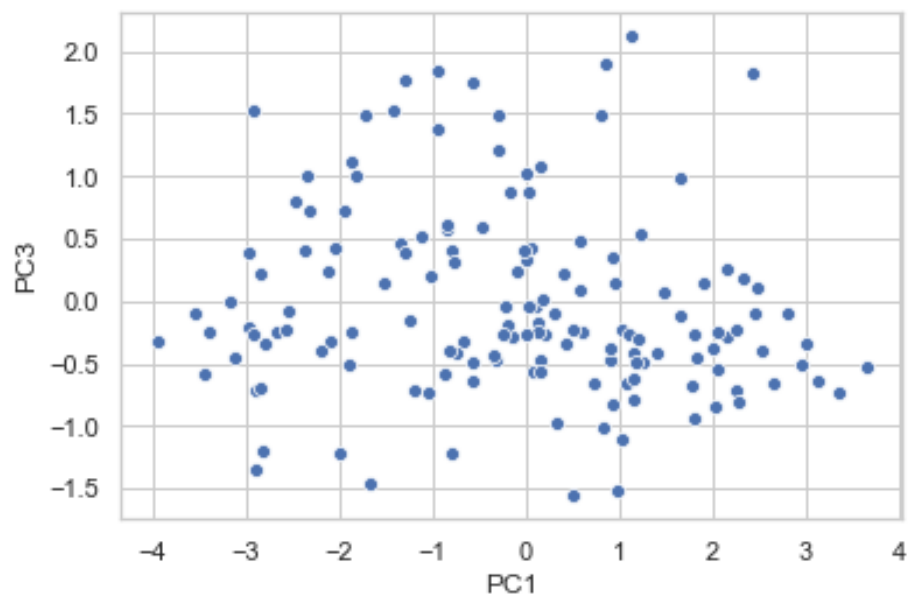
- We see there is no correlation between the 4 Principal Components

# Scatterplot (Principal Components)

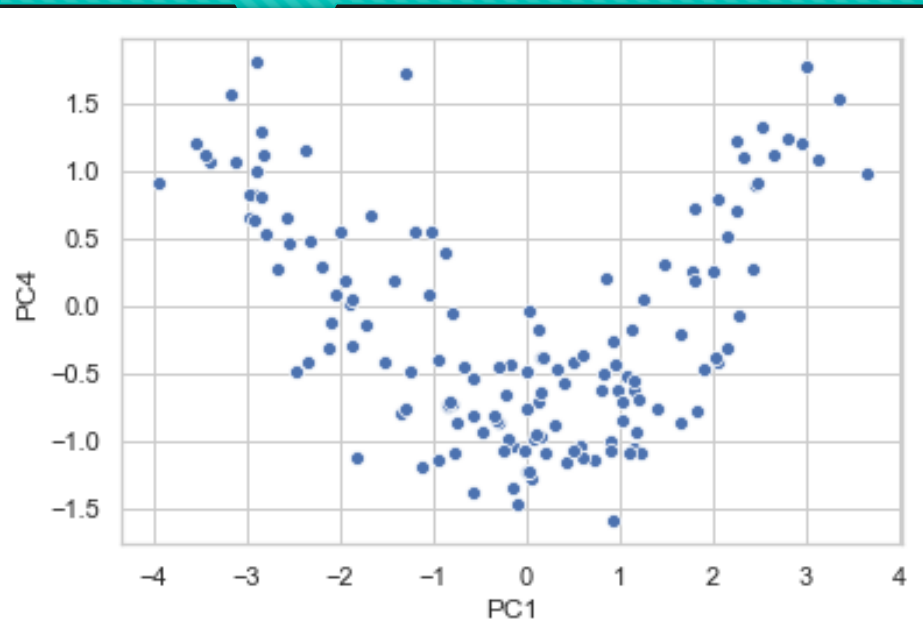


○ Principal Component 1  
vs. Principal Component  
2

○ Principal Component 1  
vs. Principal  
Component 3

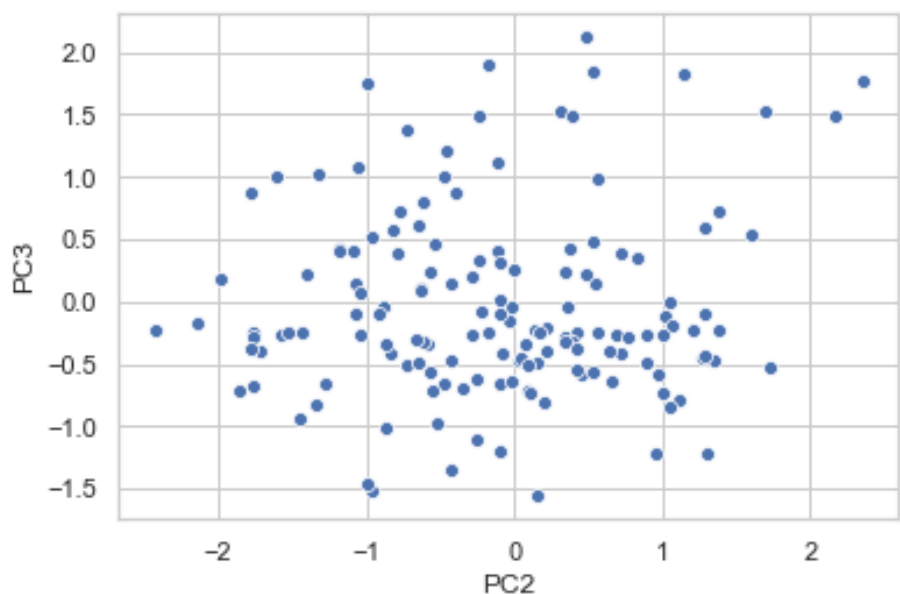


# Scatterplot (Principal Components)

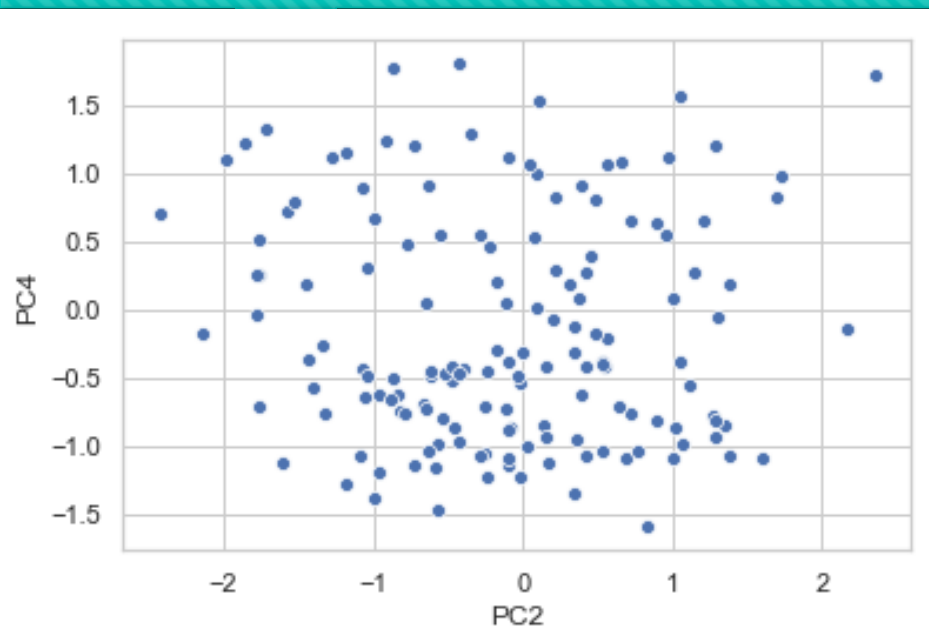


○ Principal Component 1  
vs. Principal  
Component 4

○ Principal Component 2  
vs. Principal  
Component 3

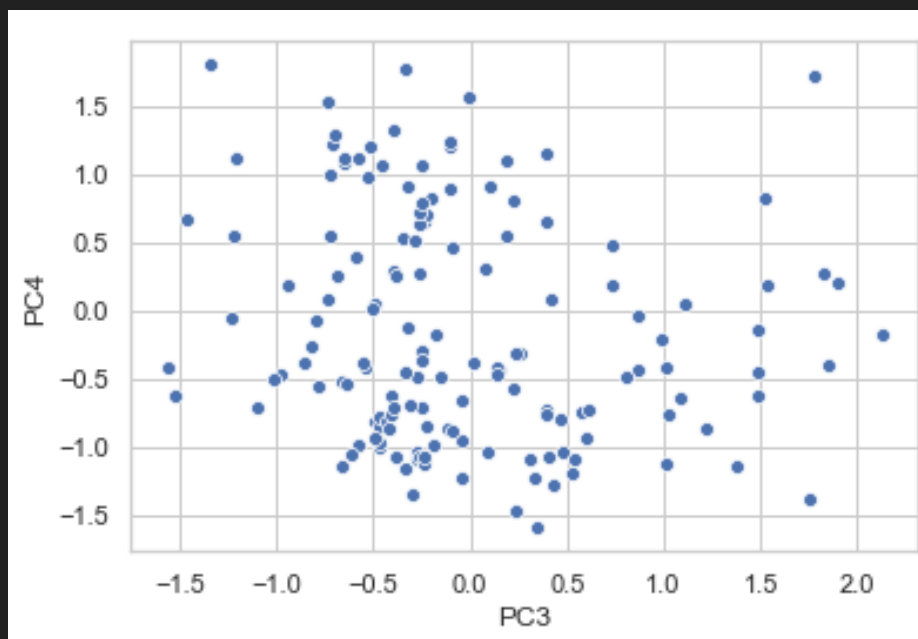


# Scatterplot (Principal Components)



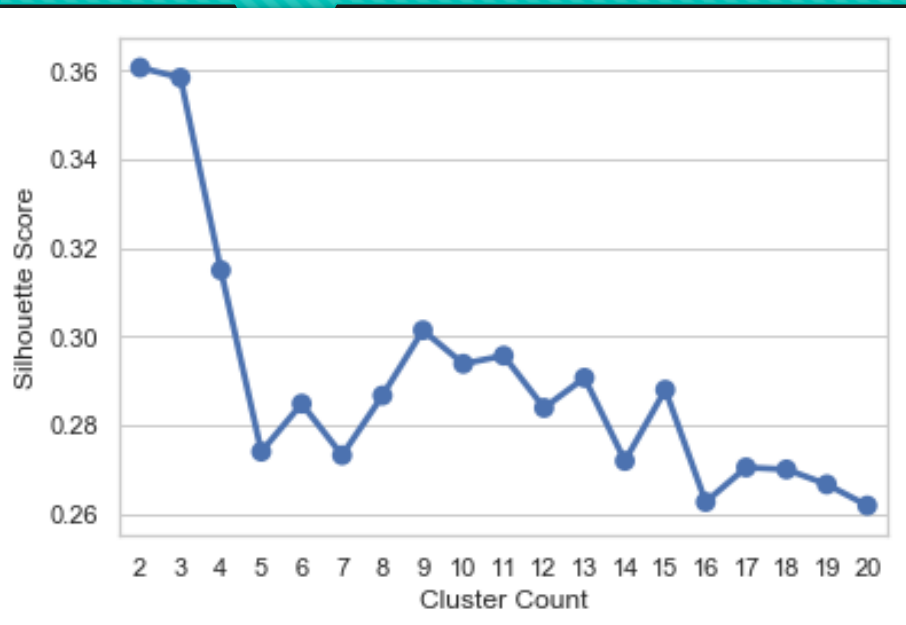
○ Principal Component 2  
vs. Principal  
Component 4

○ Principal Component 3  
vs. Principal  
Component 4



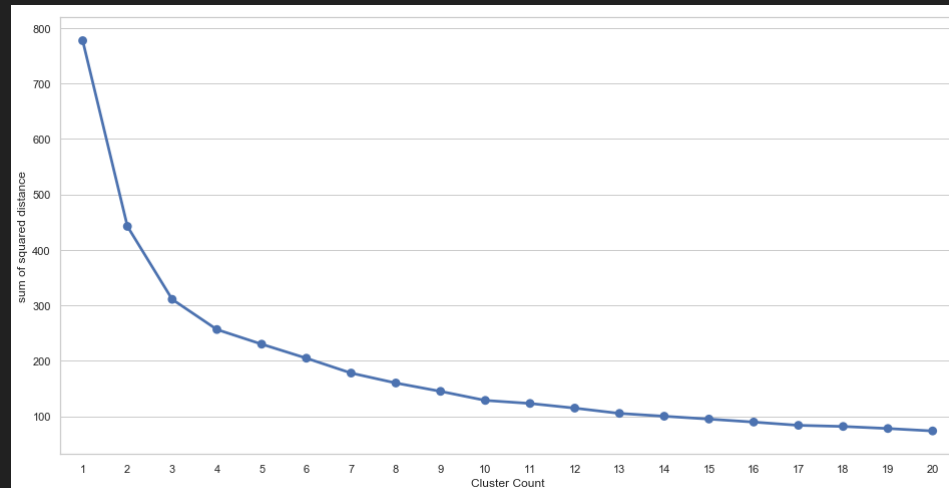


# Pre K-Means Clustering Analysis

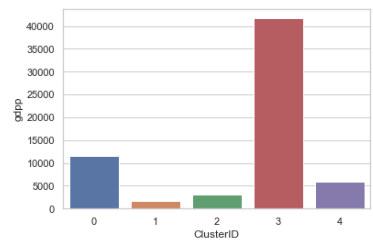
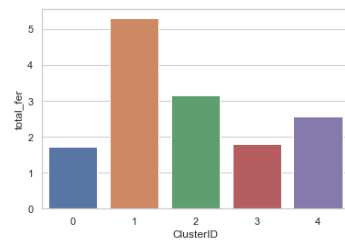
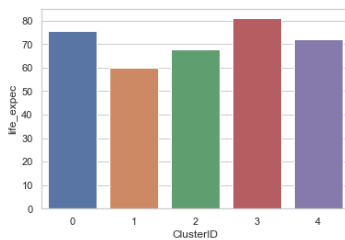
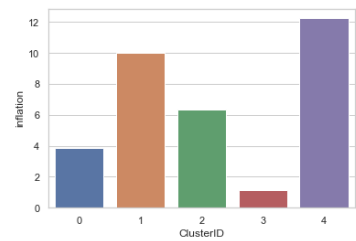
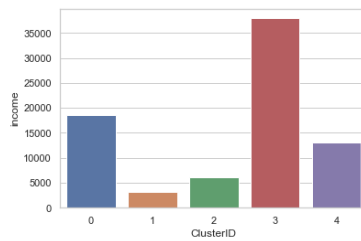
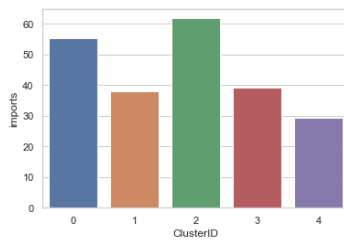
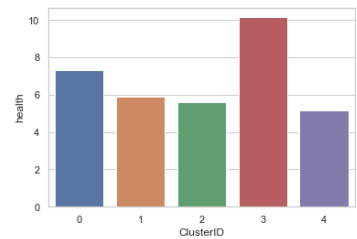
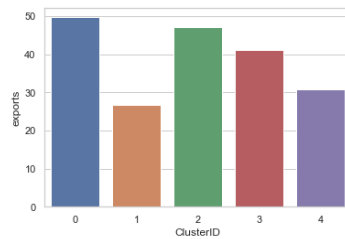
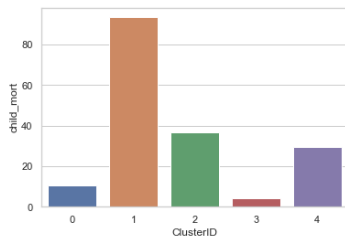


○ Silhouette Analysis

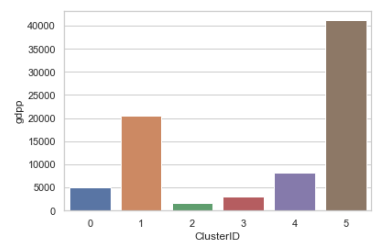
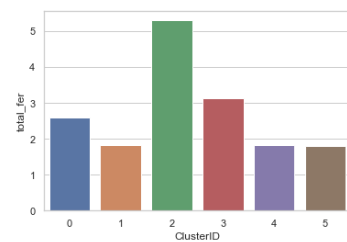
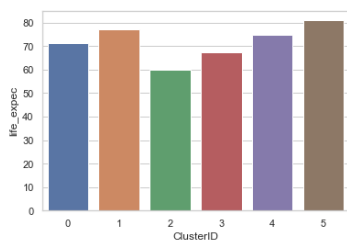
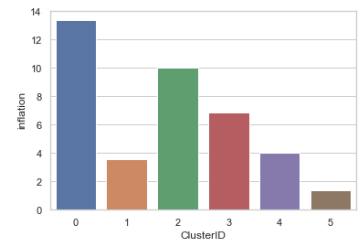
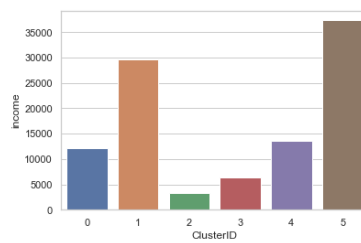
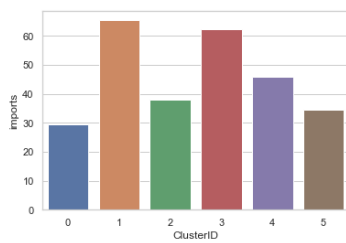
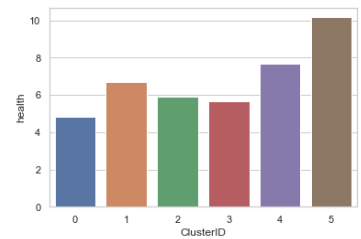
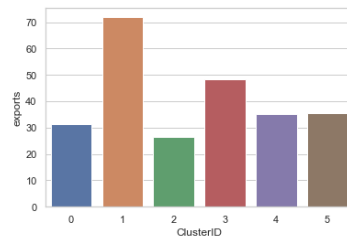
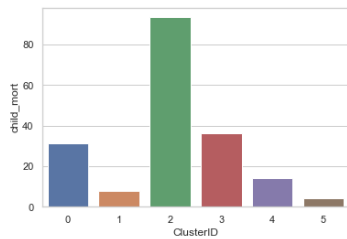
○ Sum of Squared Distances (Elbow Curve Analysis)



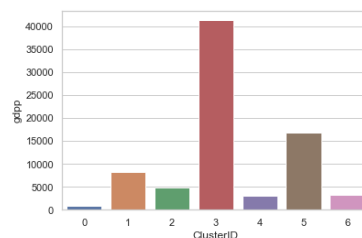
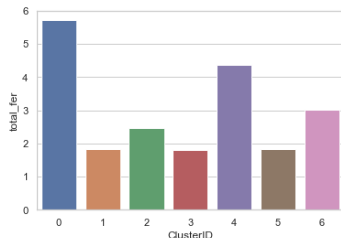
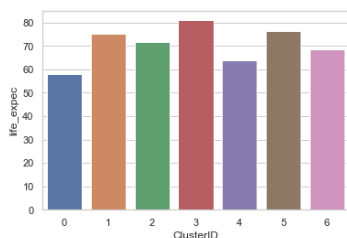
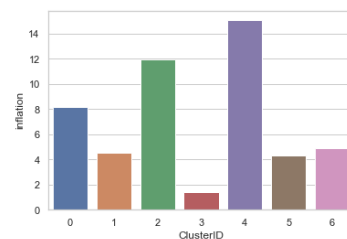
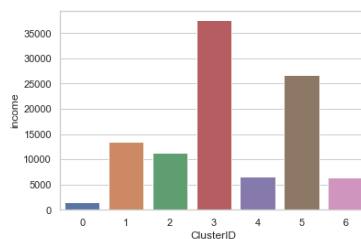
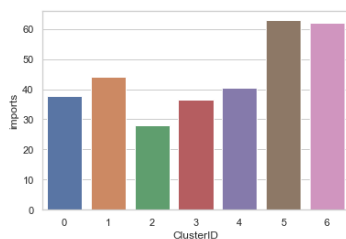
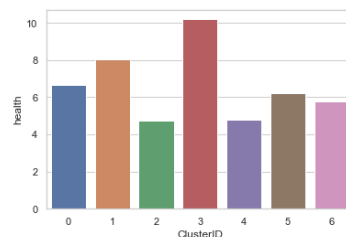
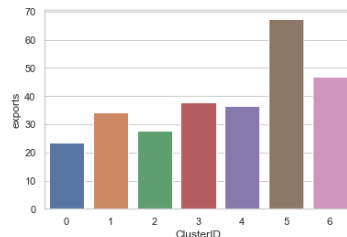
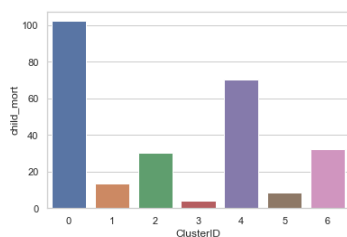
# K Means with K as 5



# K Means with K as 6



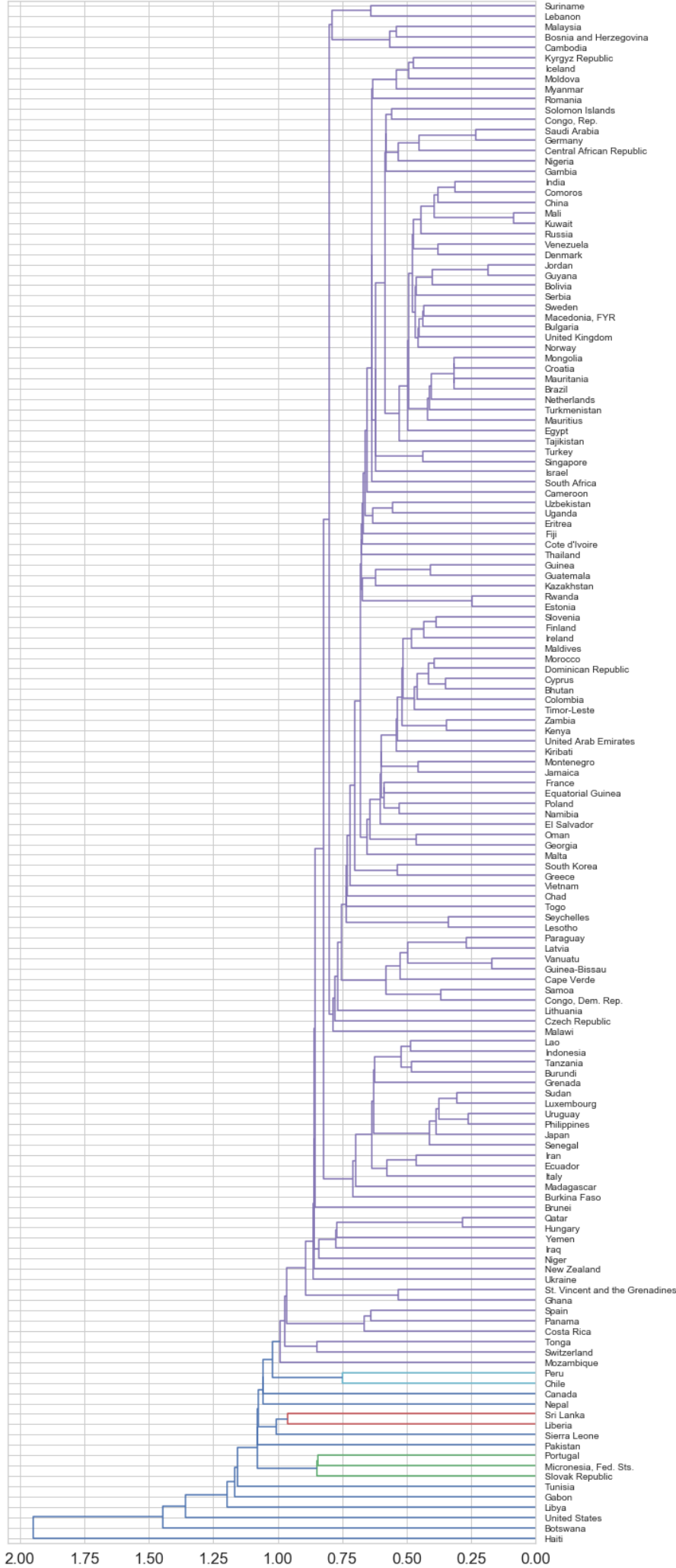
# K Means with K as 7



# Summary of K Means Clustering

- The set of countries which needs funding, as per the analysis with respect to business logic:
- 'Comoros', 'Angola', 'Kenya', 'Benin', 'Congo, Dem. Rep.', 'Pakistan', 'Afghanistan', 'Niger', 'Uganda', 'Togo', 'Mali', 'Cote d'Ivoire', 'Ghana', 'Guinea-Bissau', 'Malawi', 'Chad', 'Eritrea', 'Rwanda', 'Madagascar', 'Equatorial Guinea', 'Mauritania', 'Central African Republic', 'Guinea', 'Sudan', 'Mozambique', 'Timor-Leste', 'Cameroon', 'Tanzania', 'Burkina Faso', 'South Africa', 'Gambia', 'Senegal', 'Yemen', 'Zambia', 'Burundi'

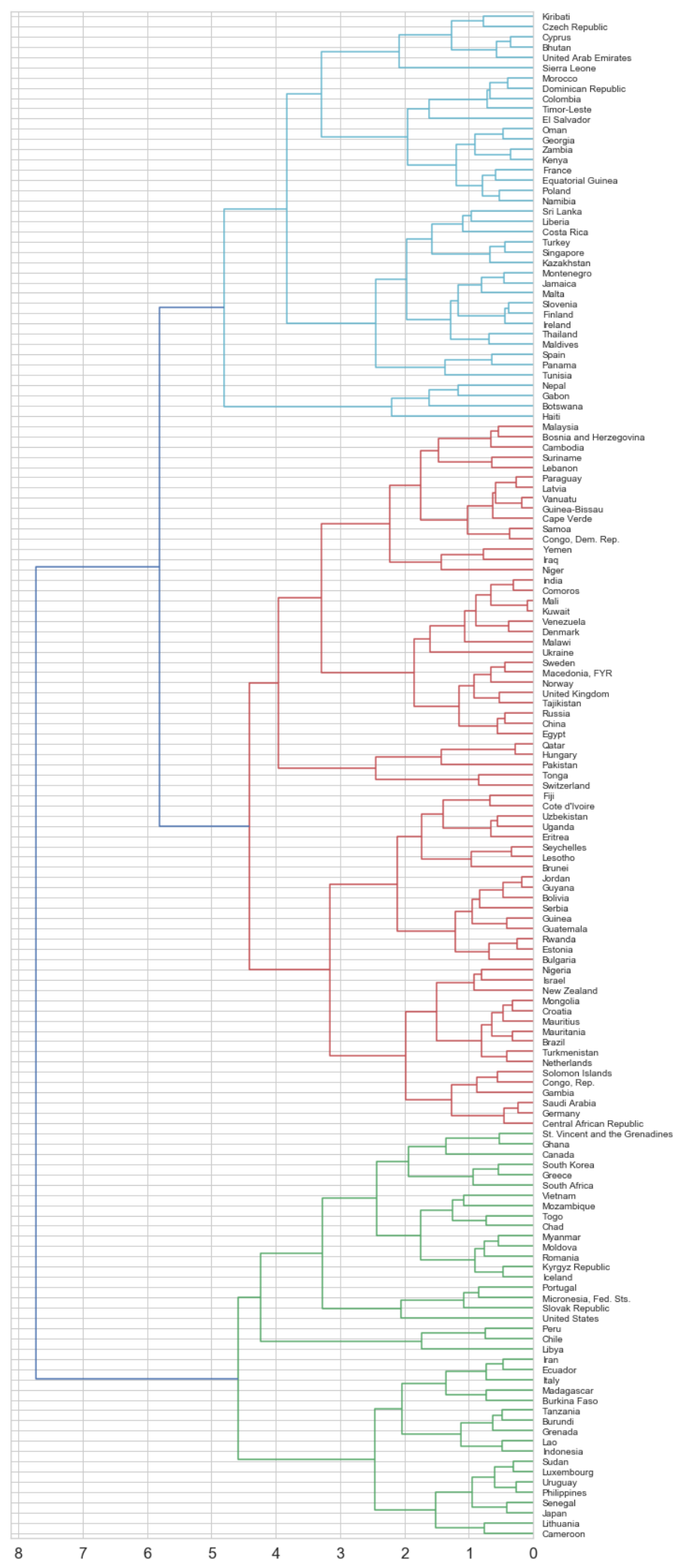
Hierarchical  
Clustering  
With  
Method  
As Single



# Summary of Hierarchical Clustering with Method Single

- The cluster formation is not good in this case
- We got the following individual country upon clustering and looking for cluster according to business requirement:
- {'Angola'}
- Hence, we move forward to Method Complete Hierarchical Clustering.

Hierarchical  
Clustering  
With  
Method  
As Complete





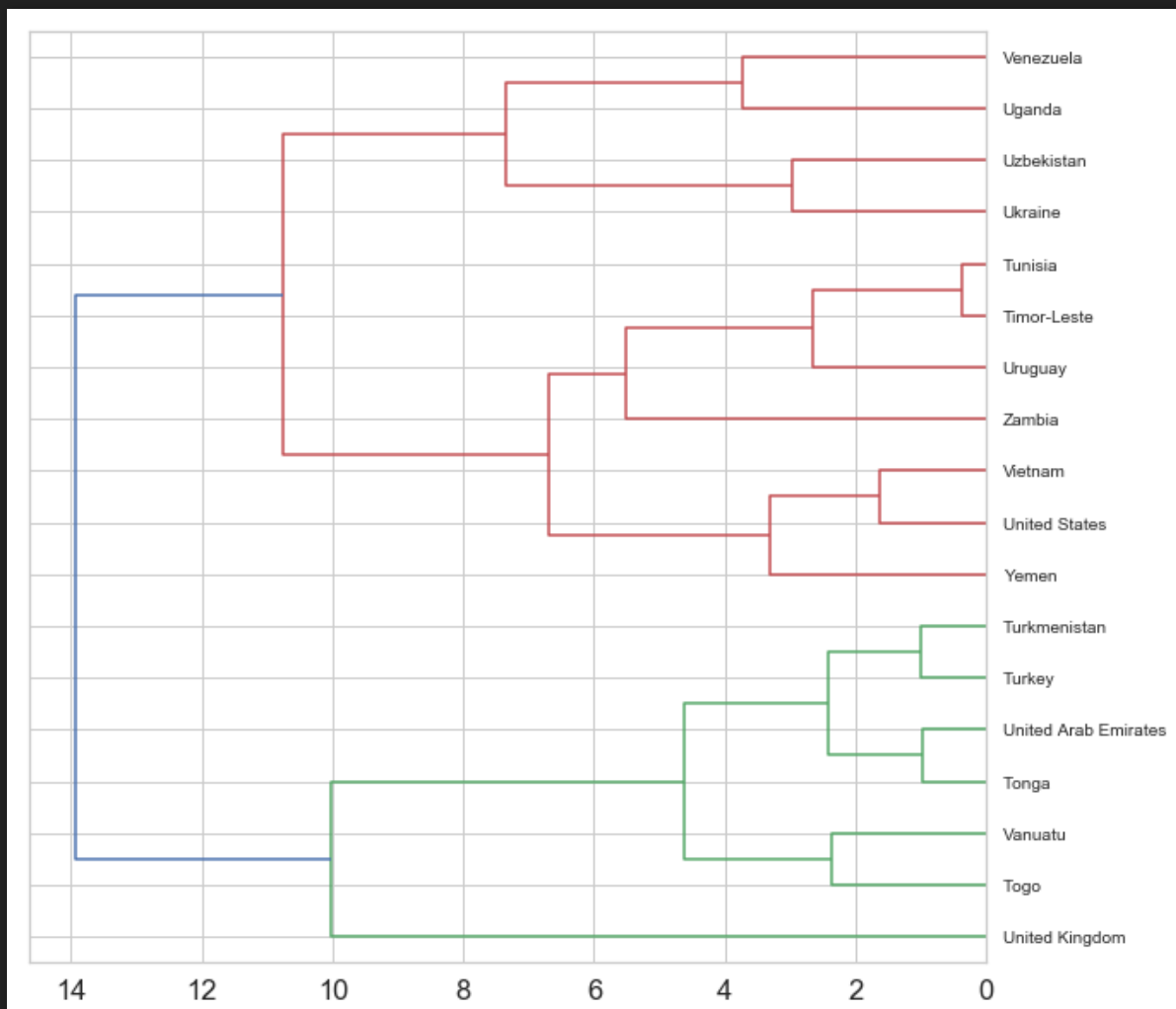
# Summary of Hierarchical Clustering with Method Complete

- The set of countries which needs funding, as per the analysis with respect to business logic:
- 'Comoros', 'Lao', 'Angola', 'Kenya', 'Benin', 'Tonga', 'Congo, Rep.', 'Namibia', 'Congo, Dem. Rep.', 'Guatemala', 'Tajikistan', 'Afghanistan', 'Niger', 'Uganda', 'Togo', 'Mali', 'Cote d'Ivoire', 'Ghana', 'Guinea-Bissau', 'Malawi', 'Chad', 'Rwanda', 'Madagascar', 'Solomon Islands', 'Equatorial Guinea', 'Mauritania', 'Botswana', 'Samoa', 'Central African Republic', 'Guinea', 'Mozambique', 'Cameroon', 'Tanzania', 'Iraq', 'Burkina Faso', 'South Africa', 'Gambia', 'Senegal', 'Zambia', 'Burundi'

# Combined Result: K Mean and Hierarchical Clustering

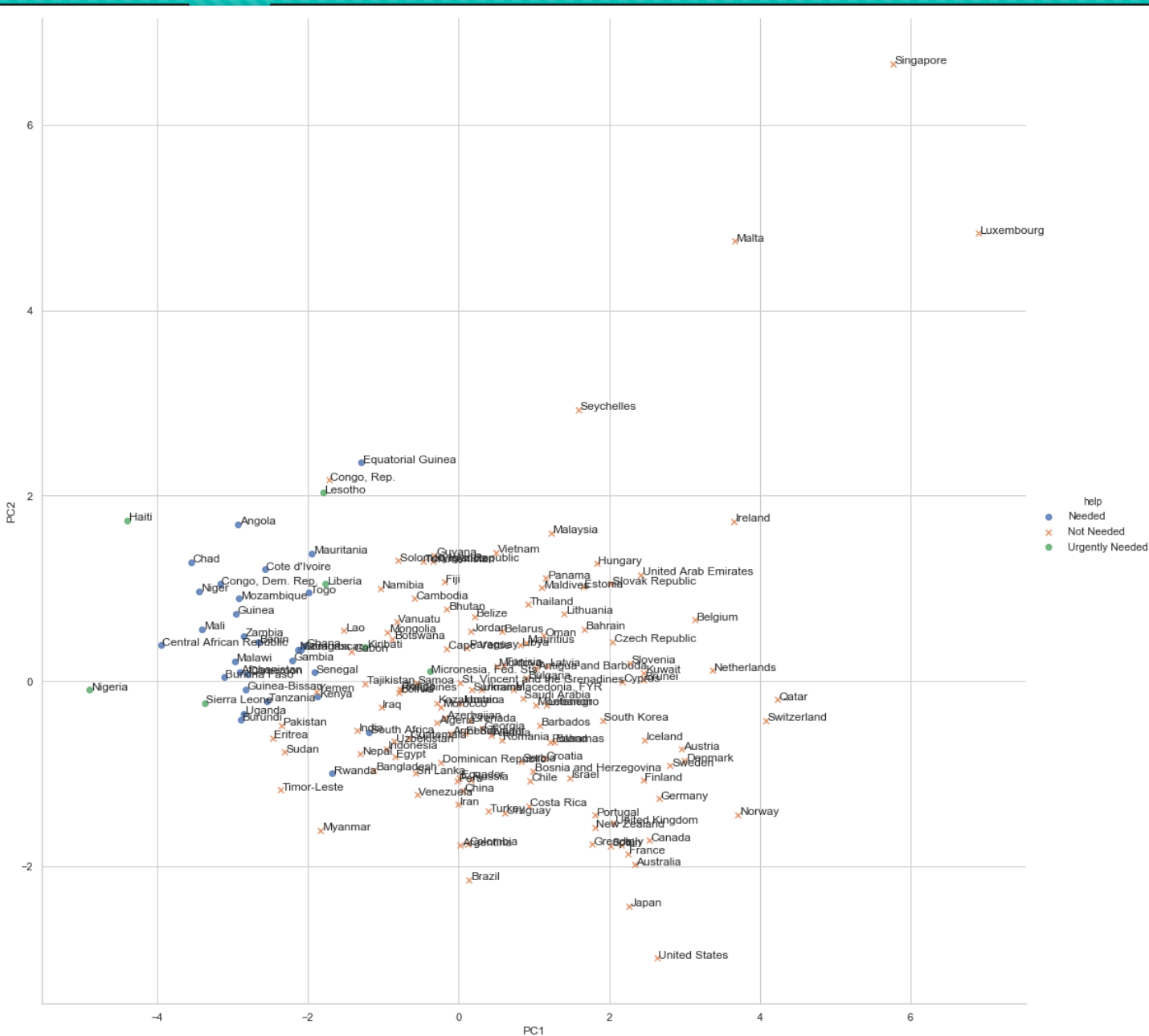
- The set of countries which needs funding, as per the analysis with respect to business logic:
- 'Comoros', 'Angola', 'Kenya', 'Benin', 'Congo, Dem. Rep.', 'Afghanistan', 'Niger', 'Uganda', 'Togo', 'Mali', 'Cote d'Ivoire', 'Ghana', 'Guinea-Bissau', 'Malawi', 'Chad', 'Rwanda', 'Madagascar', 'Equatorial Guinea', 'Mauritania', 'Central African Republic', 'Guinea', 'Mozambique', 'Cameroon', 'Tanzania', 'Burkina Faso', 'South Africa', 'Gambia', 'Senegal', 'Zambia', 'Burundi'

# Outlier Analysis

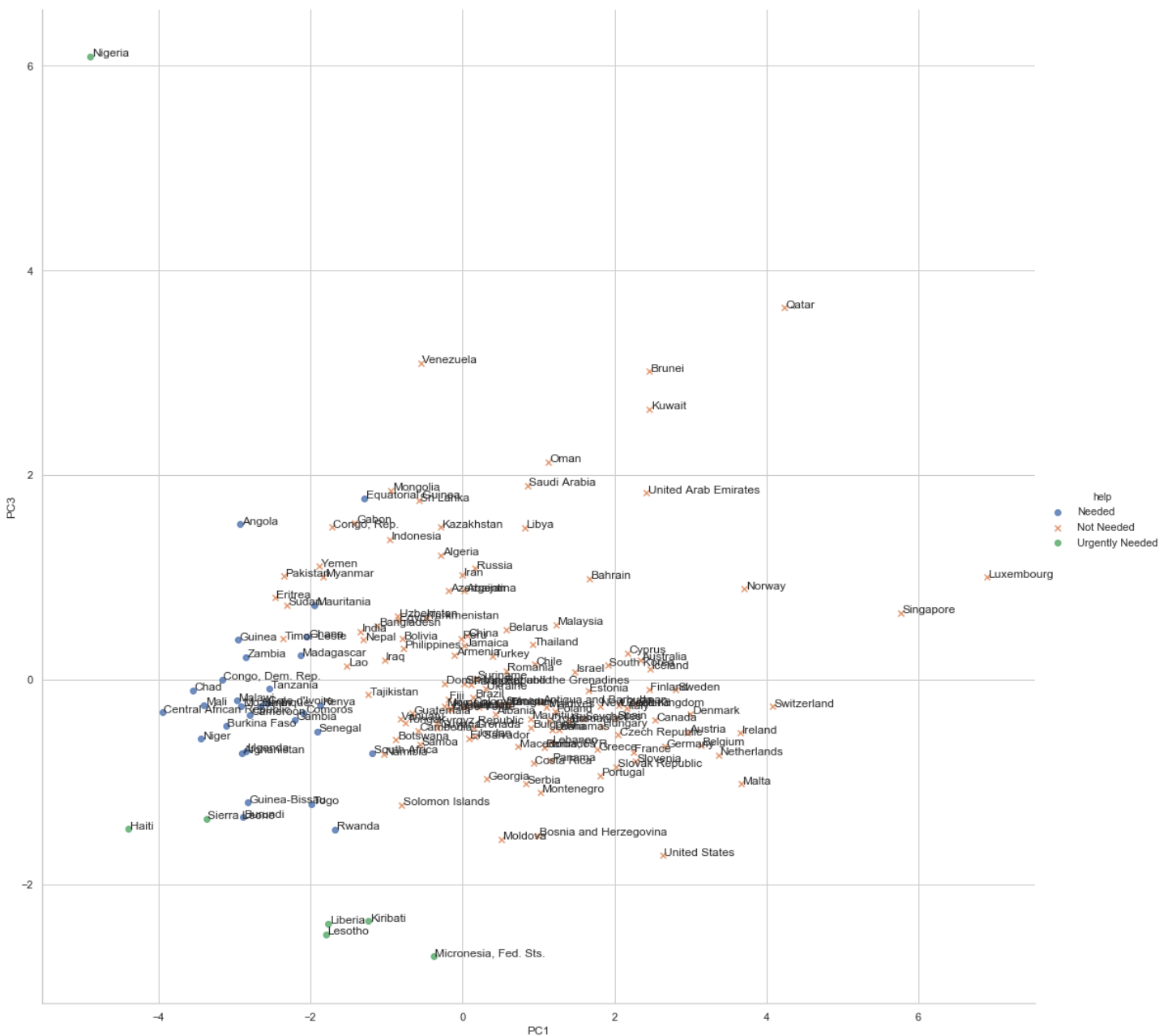


- Upon creating clusters and business logic analysis, we decide to add the outliers in our results, where socio-economic parameters are critical and need help on priority

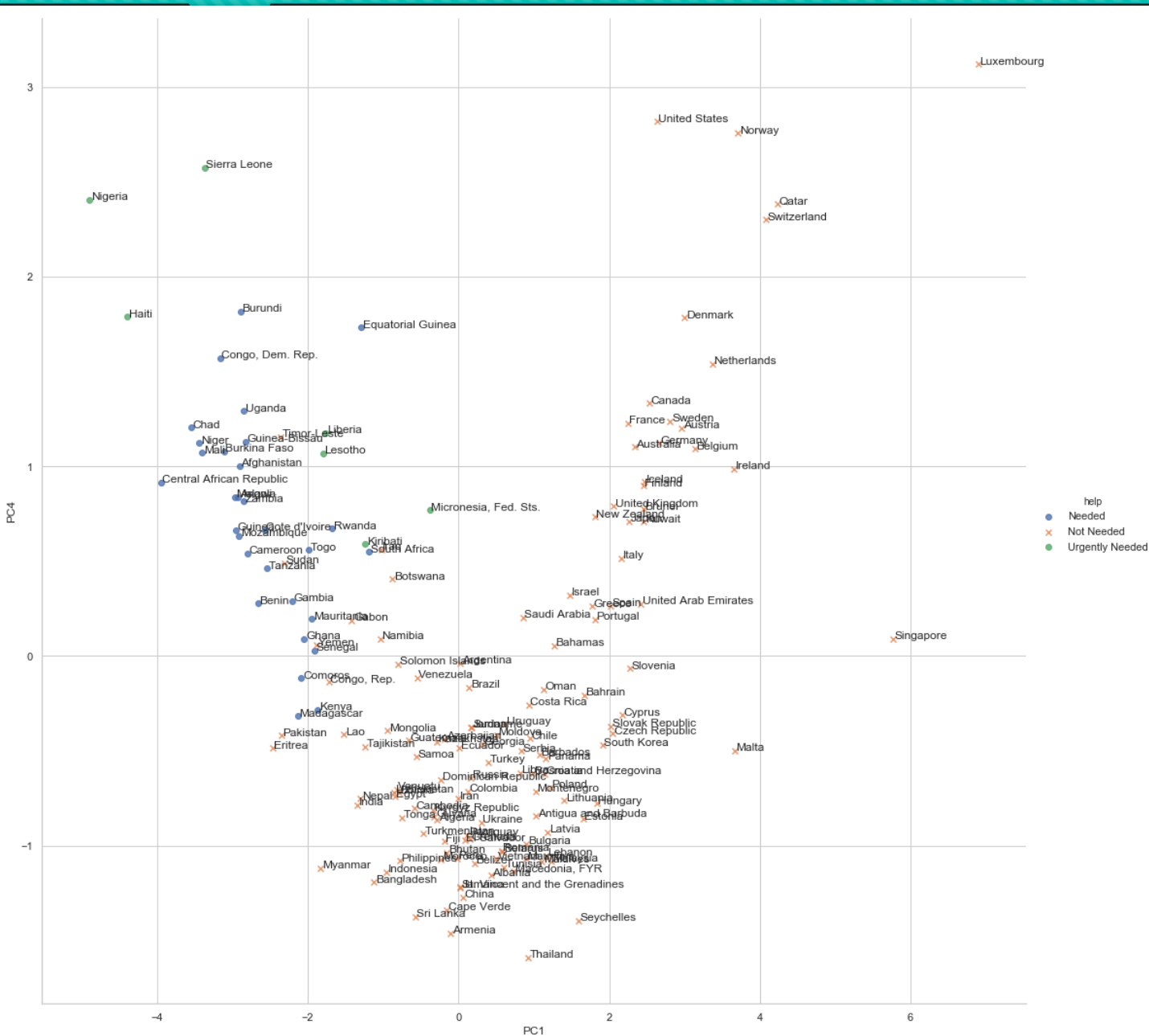
## PC1 vs PC2 hue by country (need help or not)



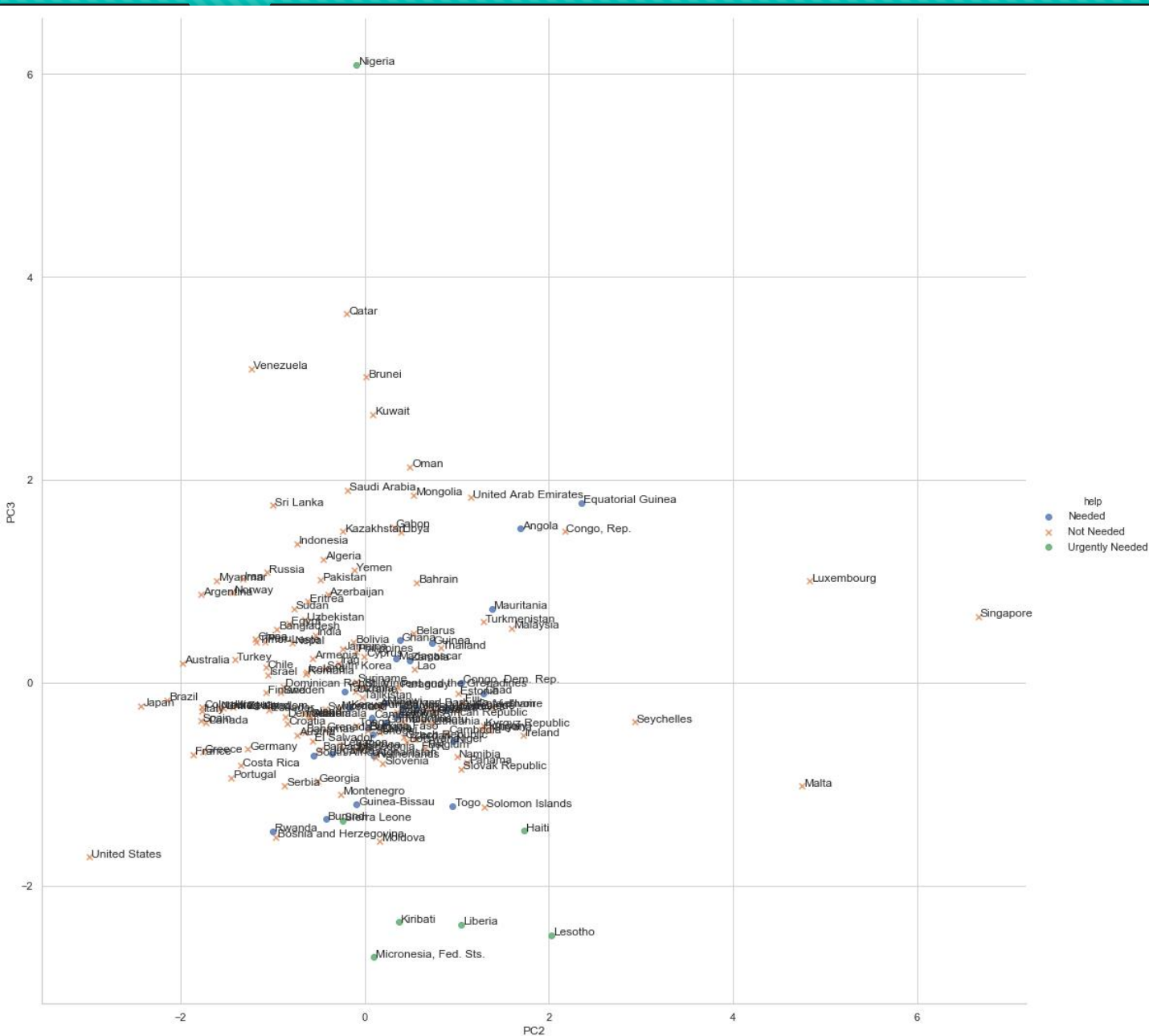
## PC1 vs PC3 hue by country (need help or not)



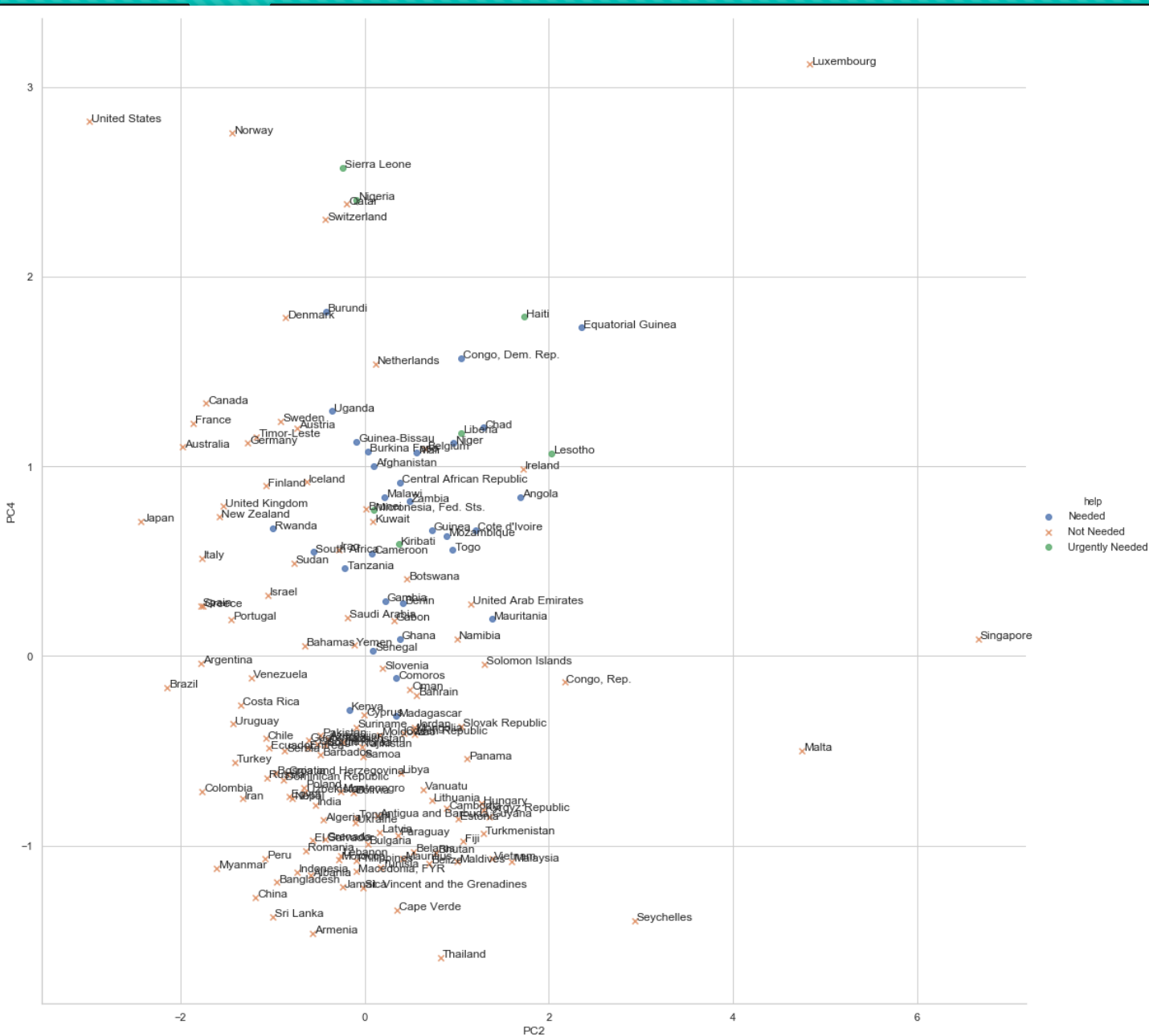
## PC1 vs PC4 hue by country (need help or not)



## PC2 vs PC3 hue by country (need help or not)



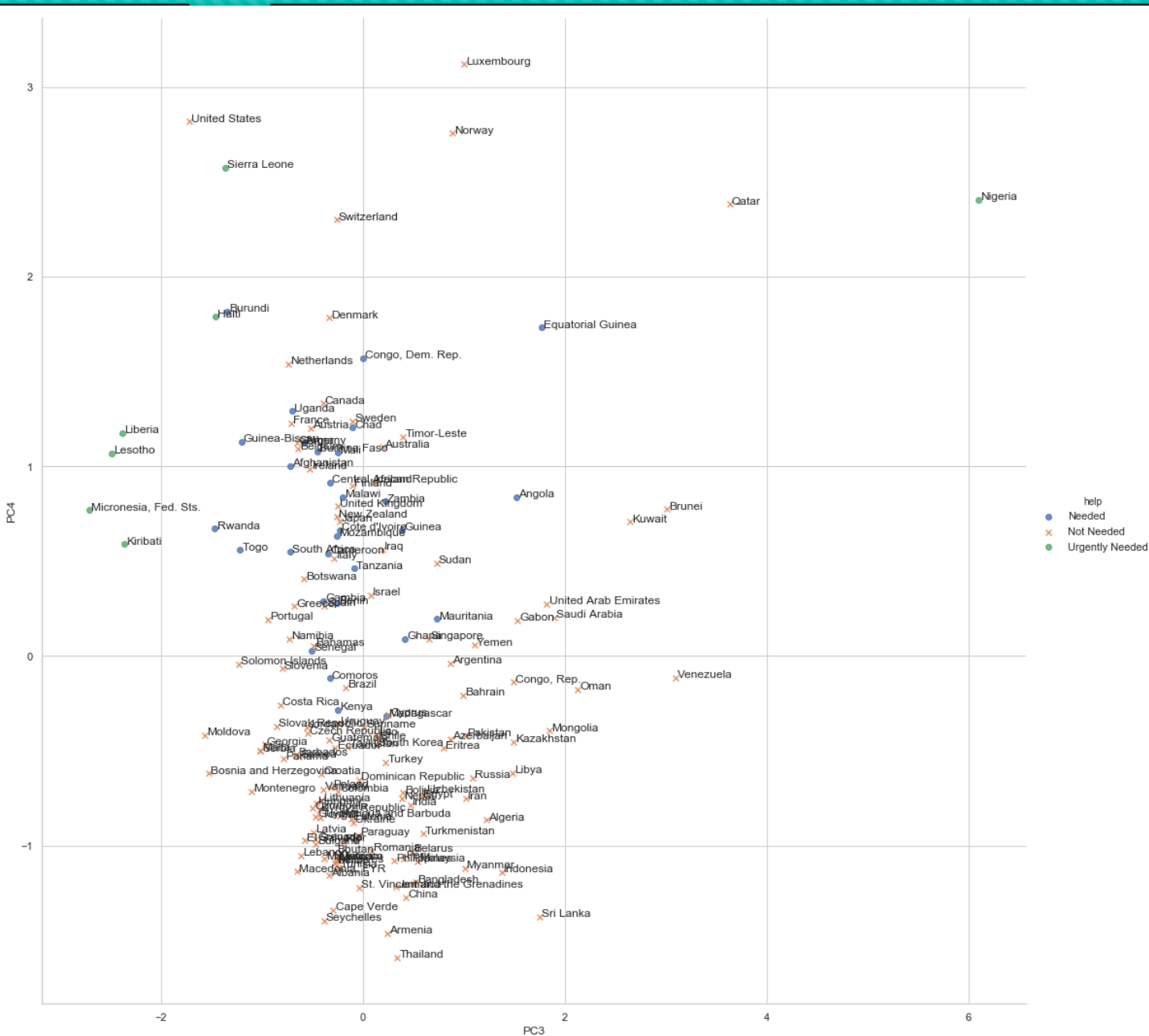
## PC2 vs PC4 hue by country (need help or not)





# Visually verifying

## PC3 vs PC4 hue by country (need help or not)



# Final Consolidated Result: 1

## FUNDING NEEDED AT HIGHER PRIORITY

- 1. Micronesia, Fed. Sts.
- 2. Liberia
- 3. Nigeria
- 4. Haiti
- 5. Sierra Leone
- 6. Kiribati
- 7. Lesotho

# Final Consolidated Result: 2

## FUNDING NEEDED AT MODERATE PRIORITY

- 1. Comoros
- 2. Angola
- 3. Kenya
- 4. Benin
- 5. Congo, Dem. Rep.
- 6. Afghanistan
- 7. Niger
- 8. Uganda
- 9. Togo
- 10. Mali
- 11. Cote d'Ivoire
- 12. Ghana
- 13. Guinea-Bissau
- 14. Malawi
- 15. Chad
- 16. Rwanda
- 17. Madagascar
- 18. Equatorial Guinea
- 19. Mauritania
- 20. Central African Republic
- 21. Guinea
- 22. Mozambique
- 23. Cameroon
- 24. Tanzania
- 25. Burkina Faso
- 26. South Africa
- 27. Gambia
- 28. Senegal
- 29. Zambia
- 30. Burundi