PCA and Clustering Assignment

T Amarnath

amarnath.t22@gmail.com

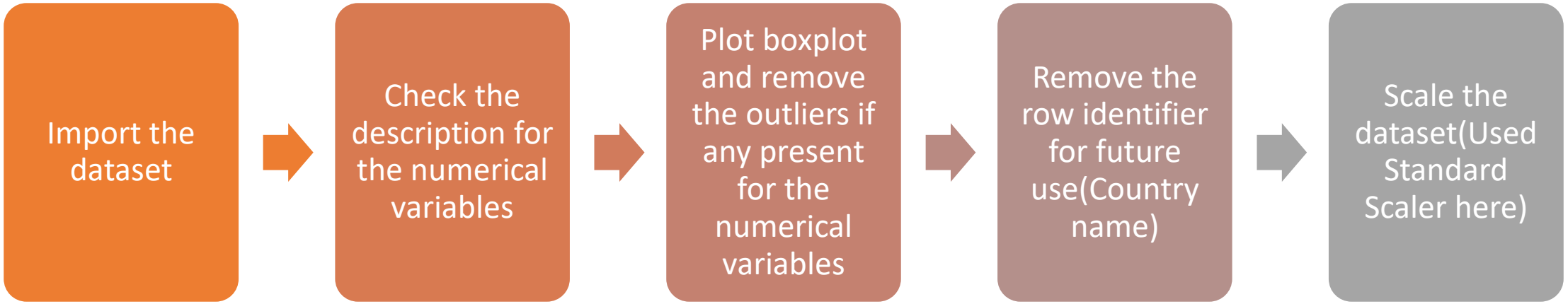
Problem Statement

- HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.
- After the recent funding programmes they have been able to raise around $ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. The significant issues that come while making this decision are mostly related to choosing the countries that are in the direst need of aid.

Our Role

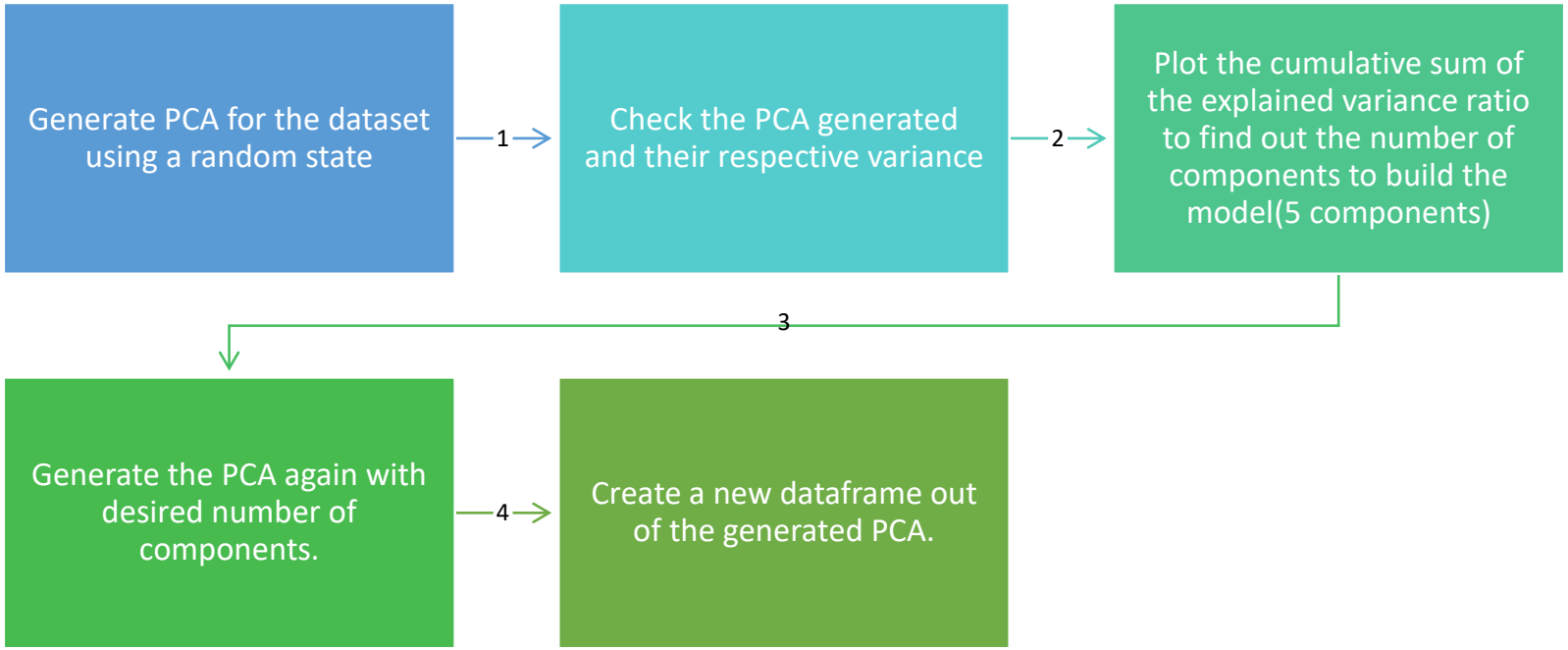
- This is where we can come in as a data analyst. Out job is to categorise the countries using some socio-economic and health factors that determine the overall development of the country. Then we need to suggest the countries which the CEO needs to focus on the most. The datasets containing those socio-economic factors and the corresponding data dictionary are provided.

Steps for the analysis

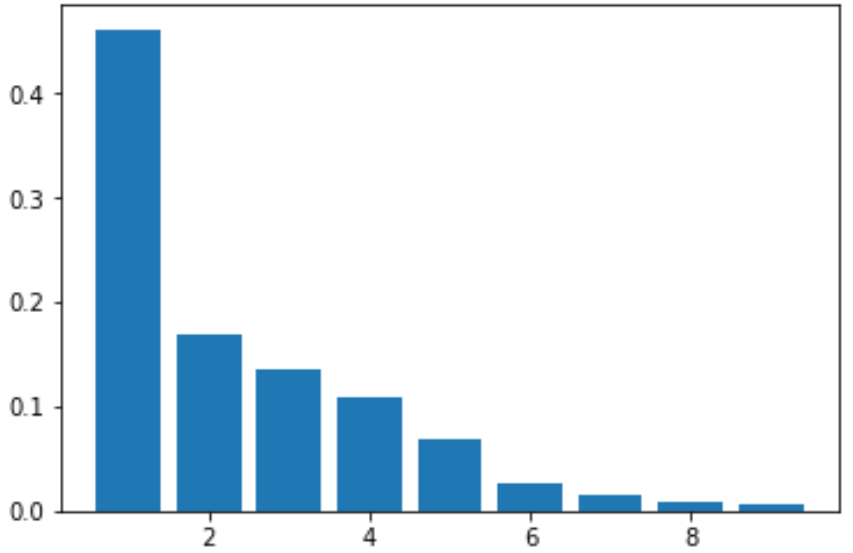


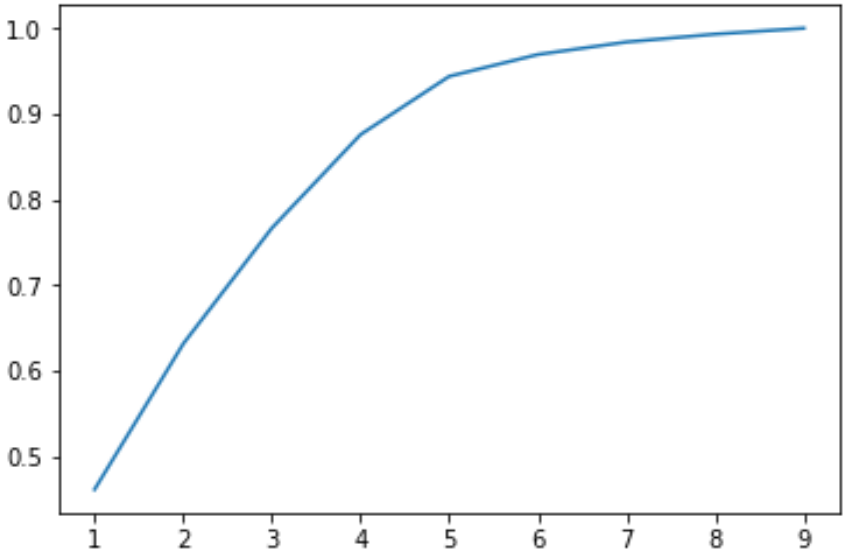
Preparing the data

Generating the PCA

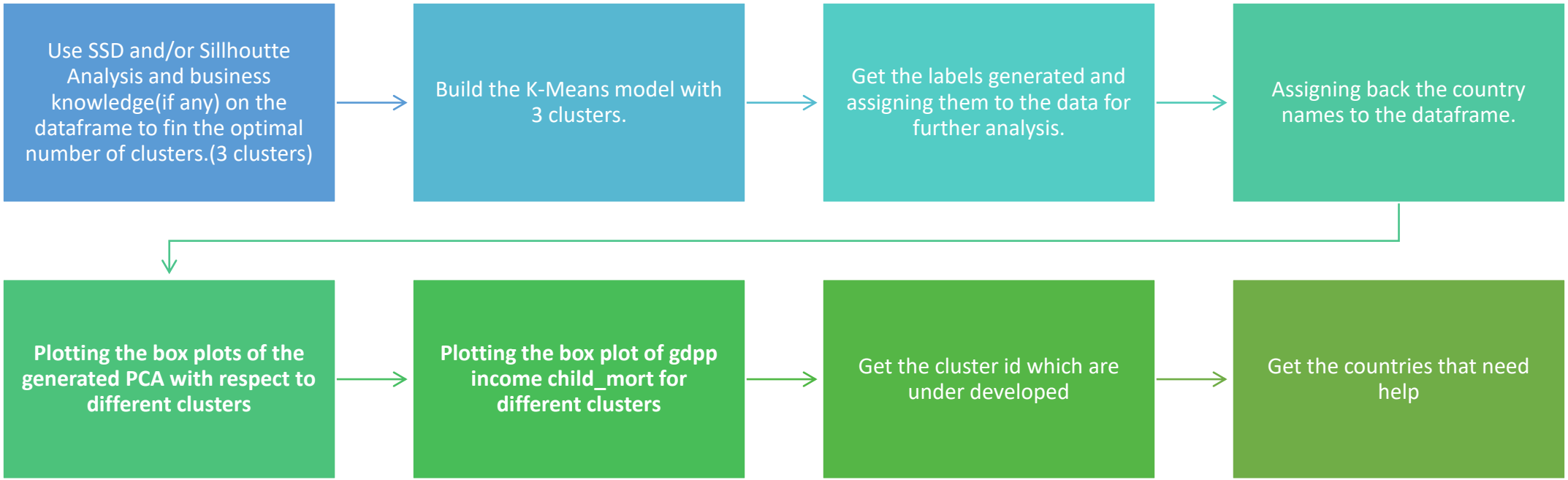


Variance Ratio and Cumulative variance ratio

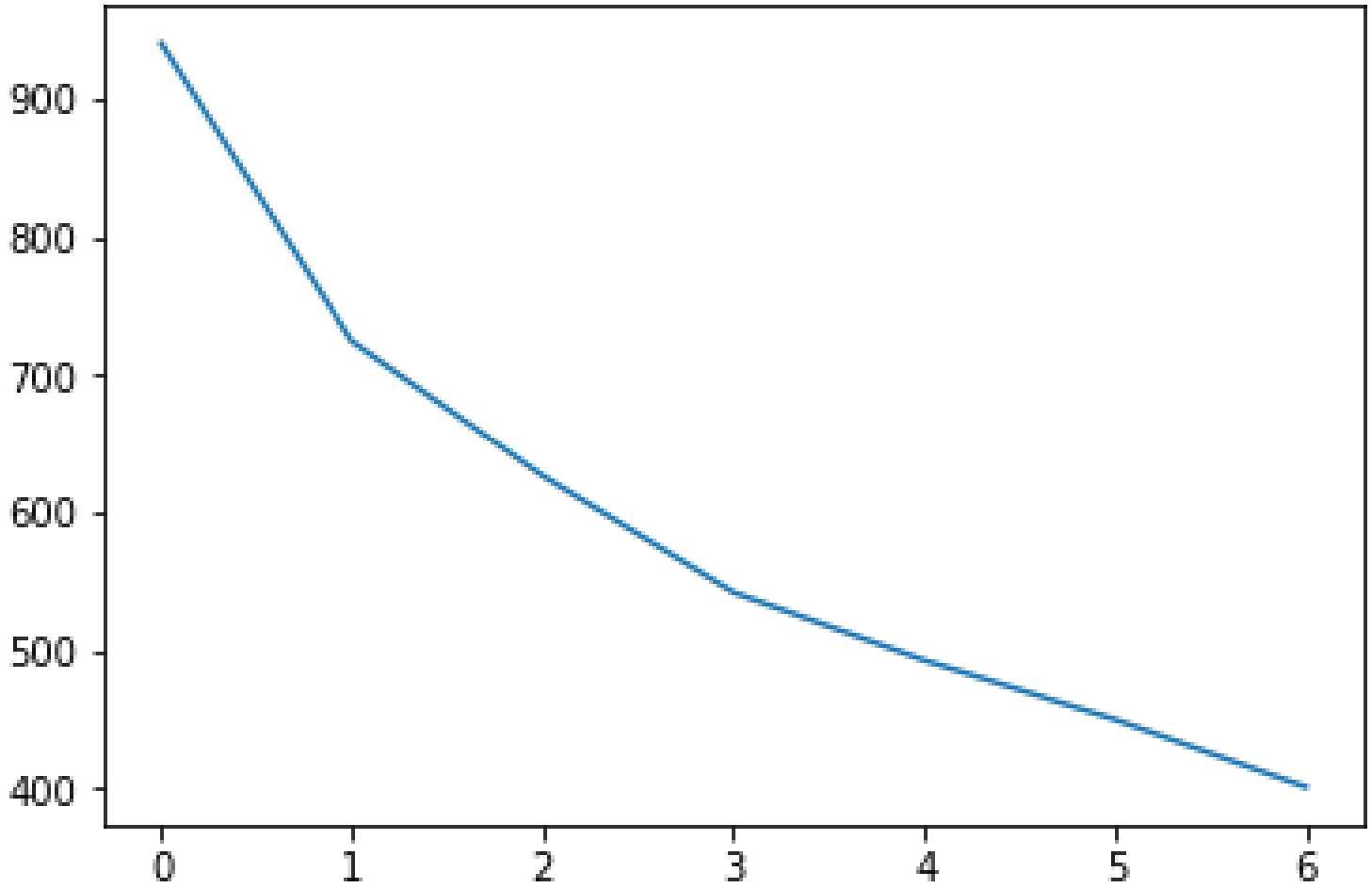




K-Means Clustering

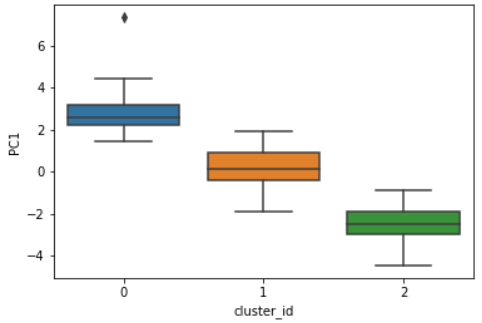


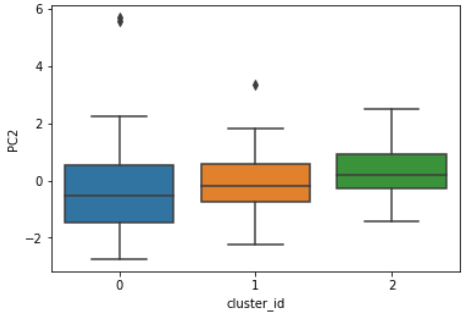
SSD Curve

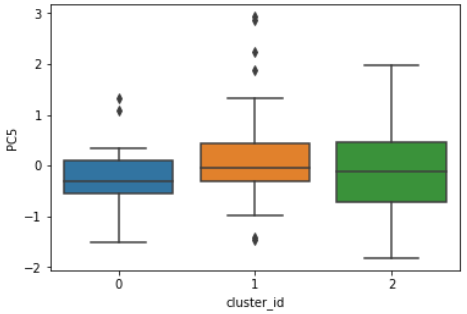


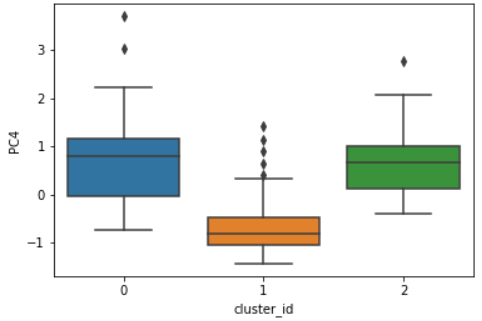
Silhouette Values for different values of k

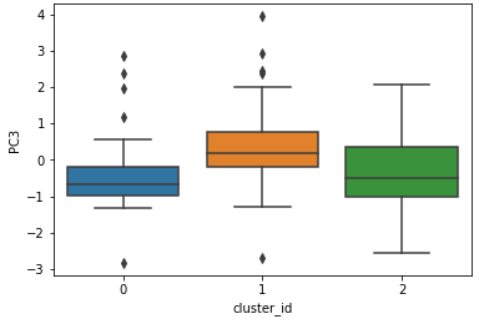
2-> 0.29476062148370336
3-> 0.2953780315630965
4-> 0.2596423194957667
5-> 0.26952687731571406
6-> 0.29590871881976166
7-> 0.2459678280890902
8-> 0.25064291767244673



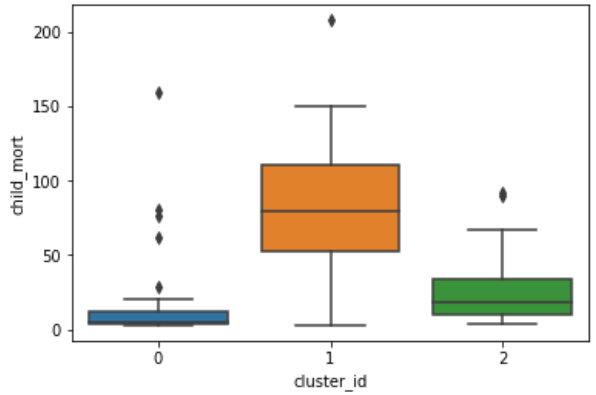


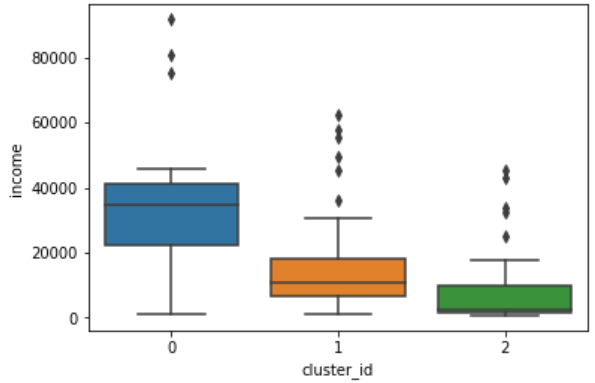


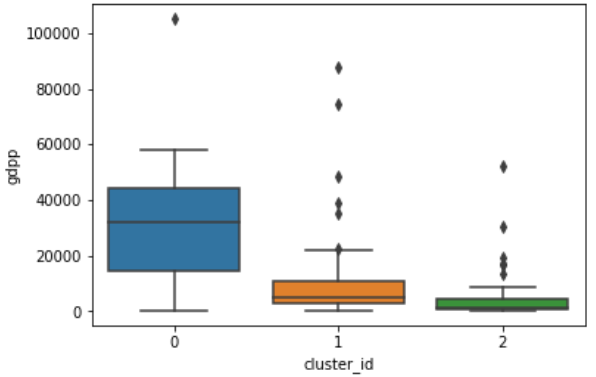




Comparison of PC with different clusters

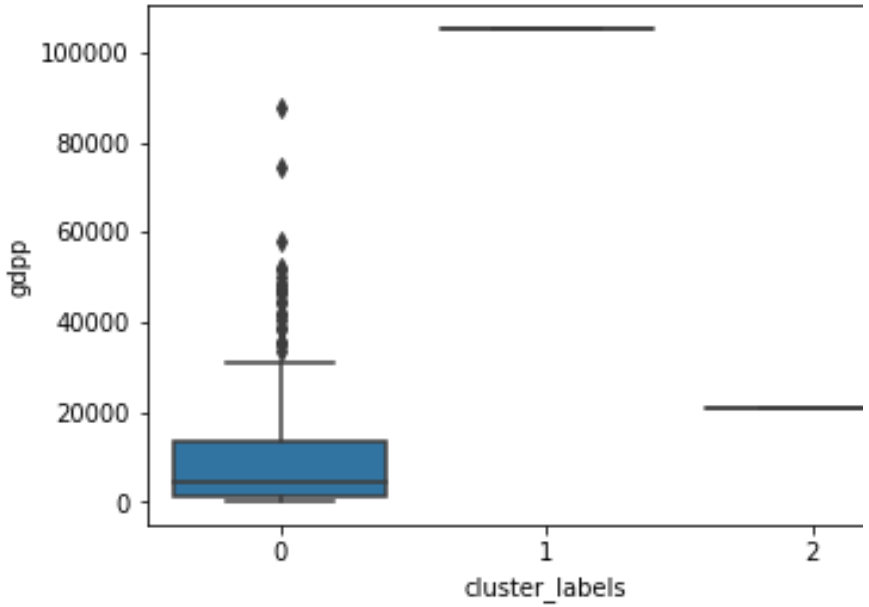

child_mort
200
150
100
50
0
0
1
2
cluster_id

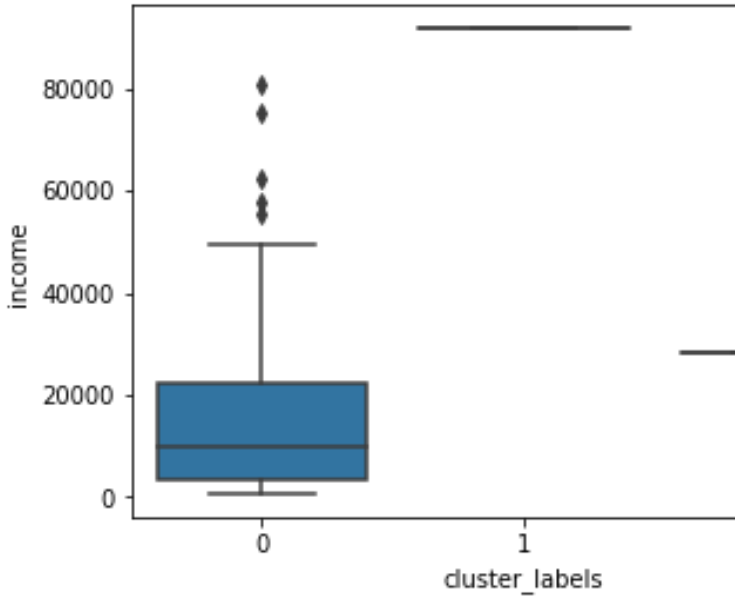

income
80000
60000
40000
20000
0
0
1
2
cluster_id

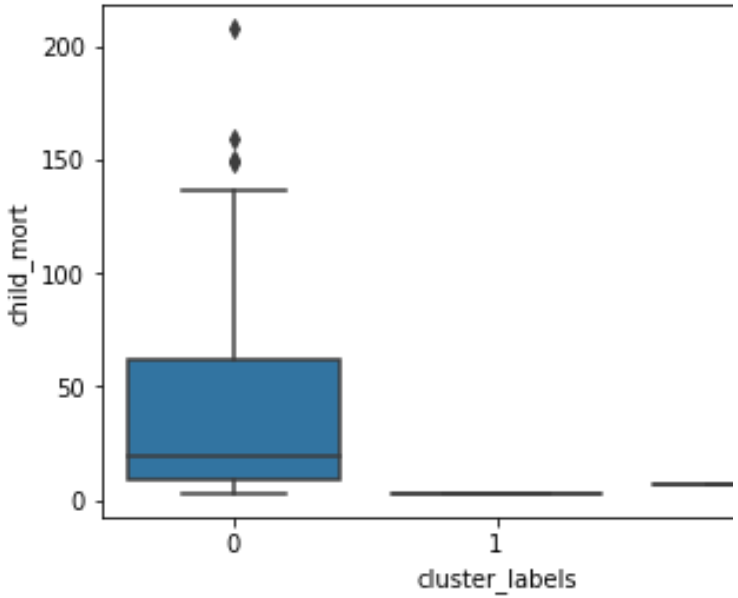

gdpp
100000
80000
60000
40000
20000
0
0
1
2
cluster_id

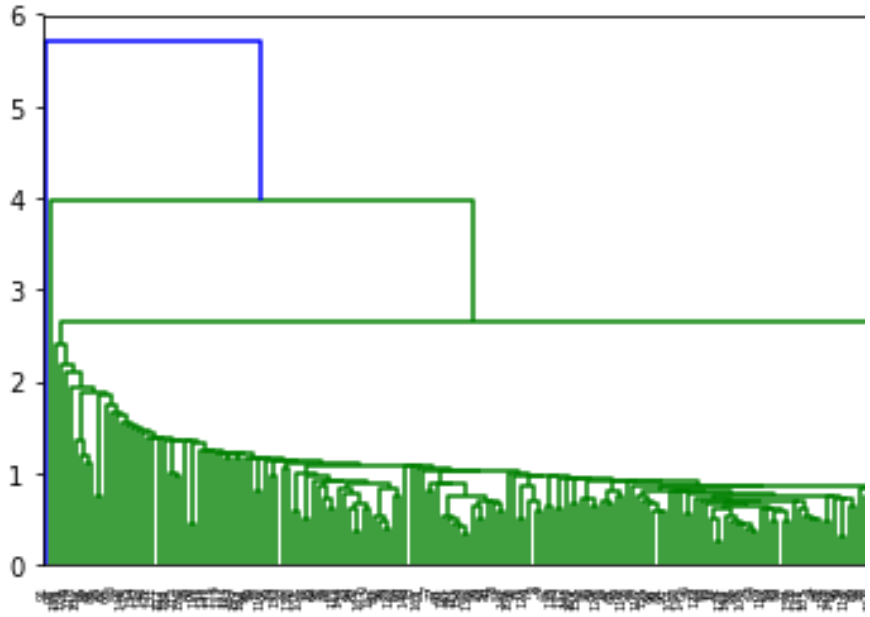
Hierarchical Clustering

1. Create single and complete linkage models.
2. Draw the dendograms.
3. Use the same number of clusters(3) used in k means to cut the dendograms using the cut_tree method
4. Add the columns back to the results to compare the clusters
5. Plot boxplots to visualize the clusters for gdpp income and child_mort.
6. Get the countries that are under-developed.

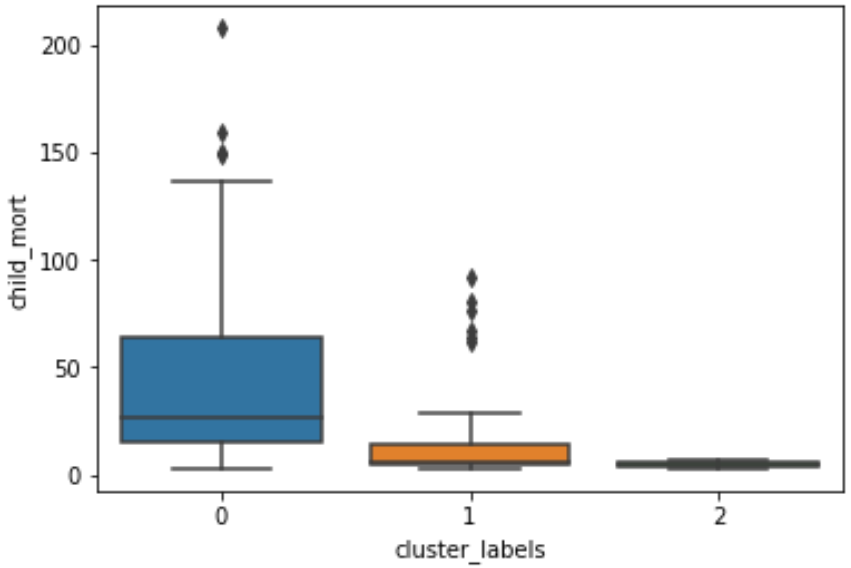


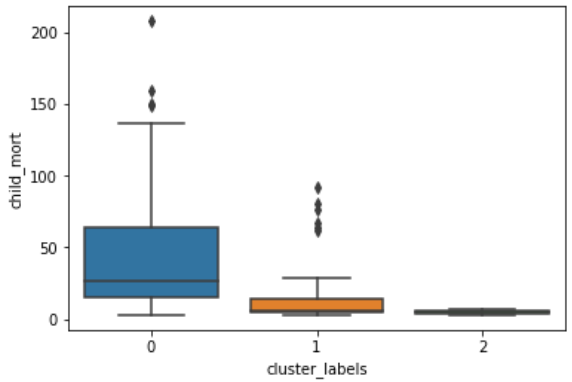


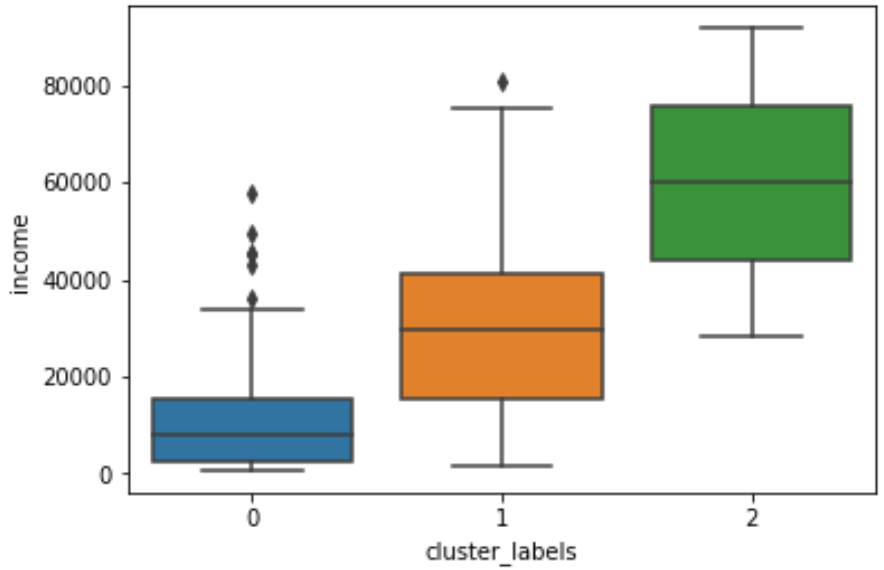


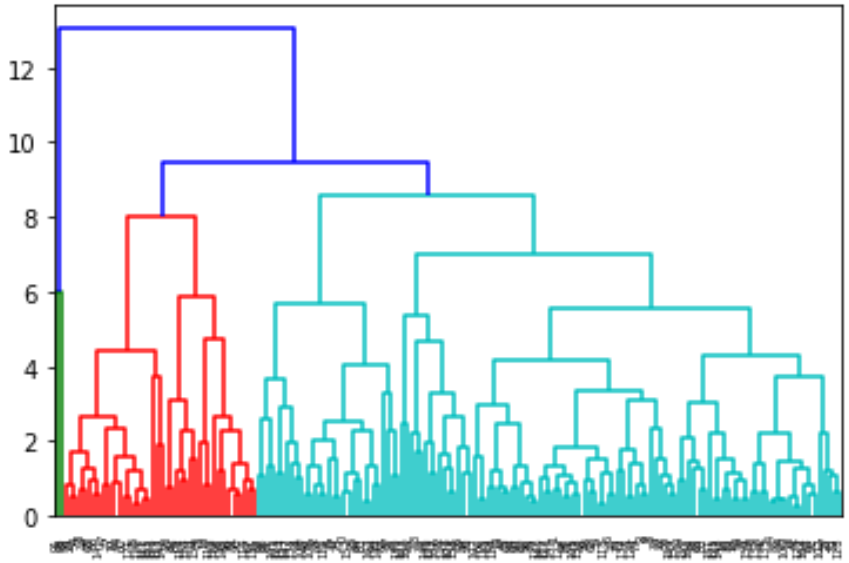


Hierarchical Clustering- Single Linkage









Hierarchical Clustering- Complete Linkage

Countries that need help

1. Afghanistan
2. Angola
3. Benin
4. Botswana
5. Burkina Faso
6. Burundi
7. Cameroon
8. Central African Republic
9. Chad
10. Comoros
11. Congo Dem. Rep.
12. Congo Rep.
13. Cote dIvoire
14. Equatorial Guinea
15. Eritrea
16. Gabon
17. Gambia
18. Ghana
19. Guinea
20. Guinea-Bissau
21. Haiti
22. Iraq
23. Kenya
24. Kiribati
25. Lao
26. Lesotho
27. Liberia
28. Madagascar
29. Malawi
30. Mali
31. Mauritania
32. Mozambique
33. Namibia
34. Niger
35. Oman
36. Romania
37. Samoa
38. Serbia
39. Slovak Republic
40. Spain
41. Sweden
42. Tajikistan
43. Tanzania
44. Tunisia
45. Vanuatu
46. Venezuela

CONCLUSION

CEO of HELP should focus on the above said 46 countries as they are under developed and high need of aid.