HELP INTERNATIONAL

PCA & Clustering Assignment

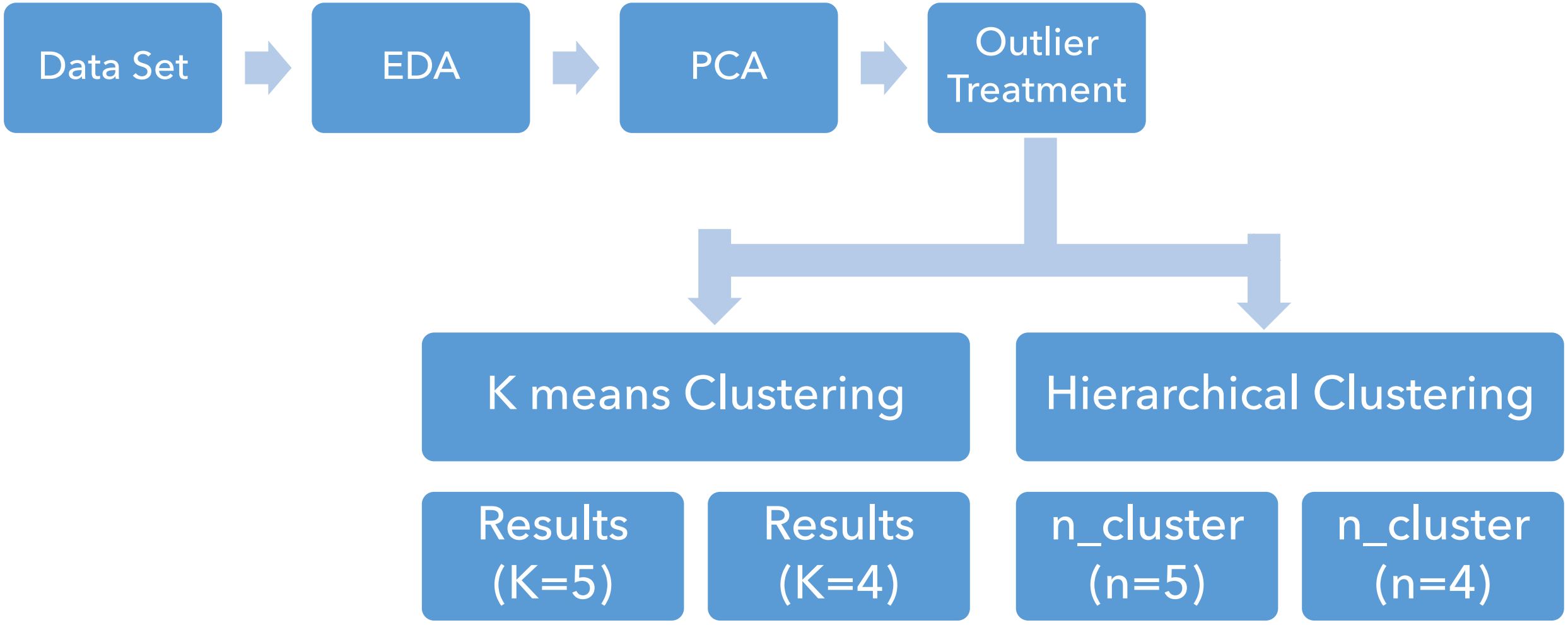
Submission:

Sachita Chauhan

Business Objectives

The aim of analysis is to enable HELP International to make an investment decision, in the most strategic and effective manner. We will categorize countries on basis of certain socio-economic and health factors, that are indicators of the overall development of the country. The categorization will suggest the clusters of countries which require HELP International.

Analysis Approach

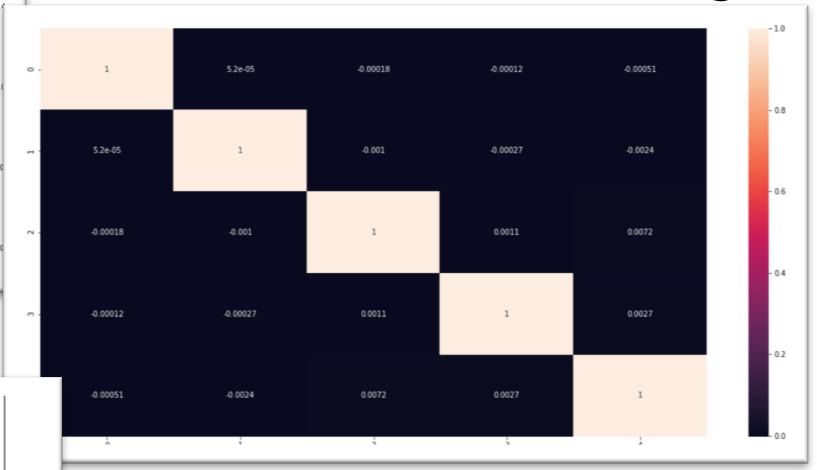


Principal Component Analysis - PCA

- High correlation in the available data set (Fig 1.)
- PCA Model is typically used in such scenarios
  - It will identify the principal components covering the maximum variance for accurate analysis of data set
  - The principal components identified through PCA are less co-related (Fig 2.)
- Scree plot suggests that 5 components are sufficient to cover 90% variance (Fig 3.)

Fig 1

Fig 2



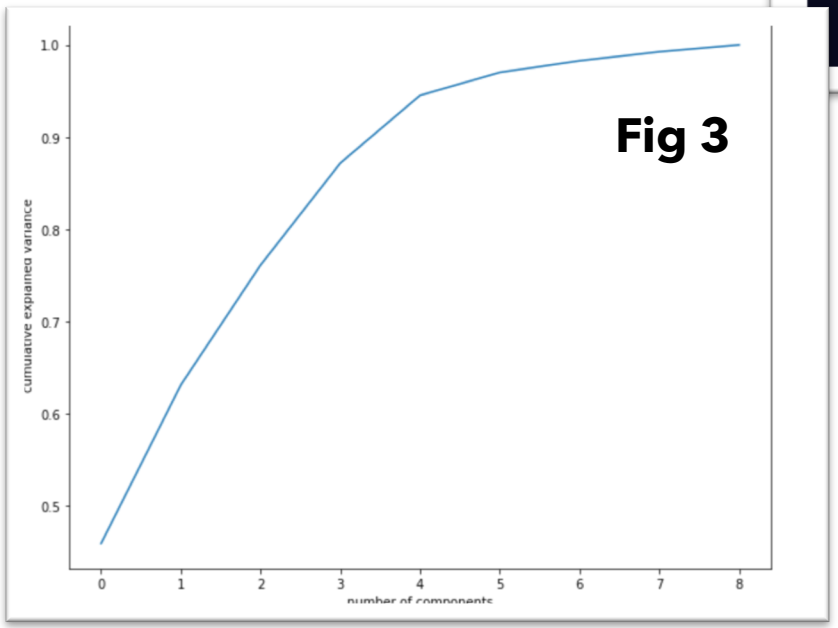


Fig 3

Clustering

- Hopkins Statistics indicate how well the data can be clustered. Our data set indicates that it has a good clustering tendency.
- Silhouette Analysis (Fig 4.) indicates that 4 and 5 is a good value for clustering
- Sum of Squared Distance analysis (Fig 5.) also indicate that 4 and 5 is a good value for clustering

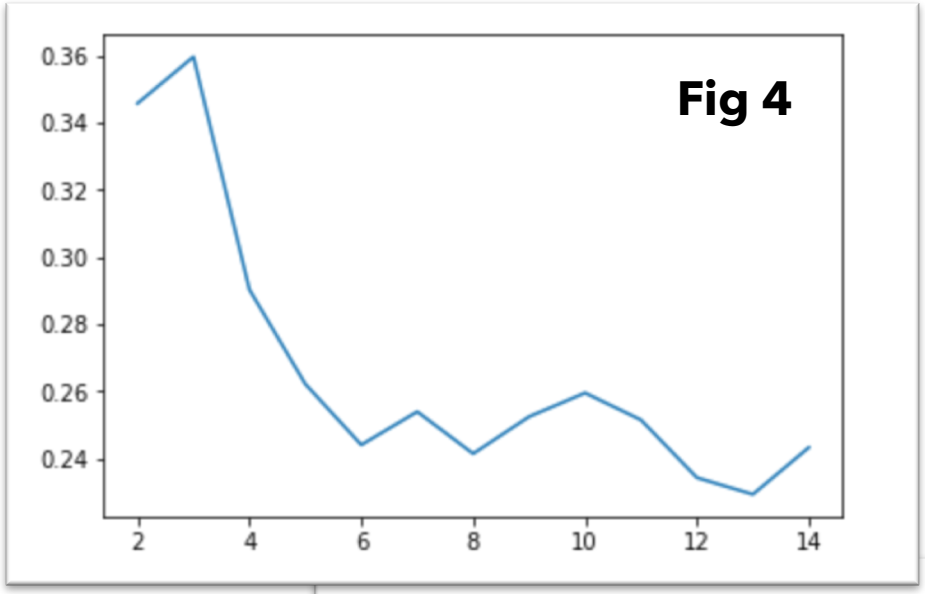


Fig 4

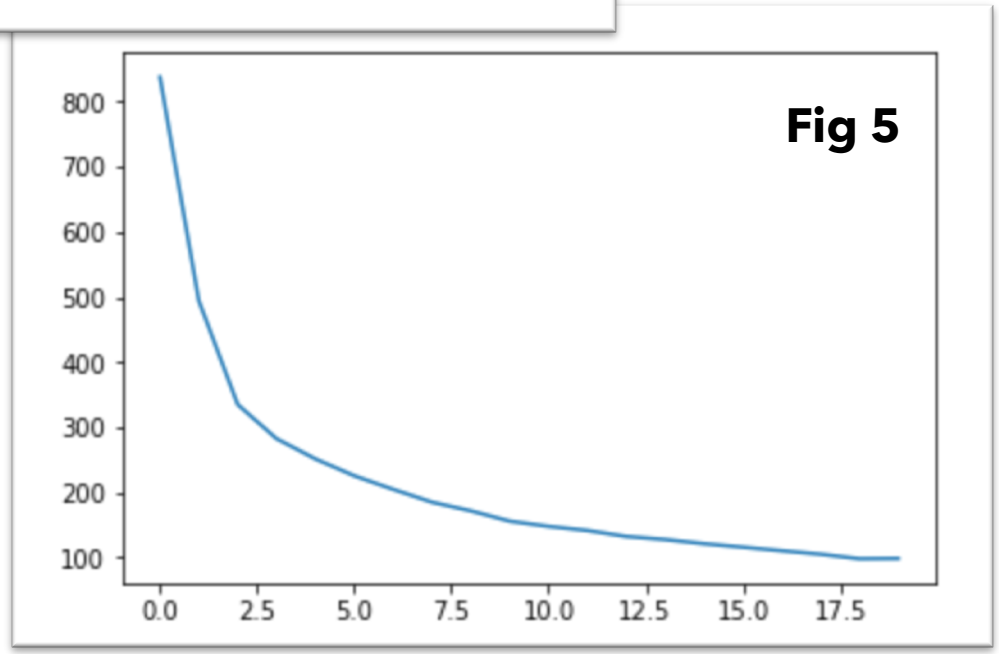
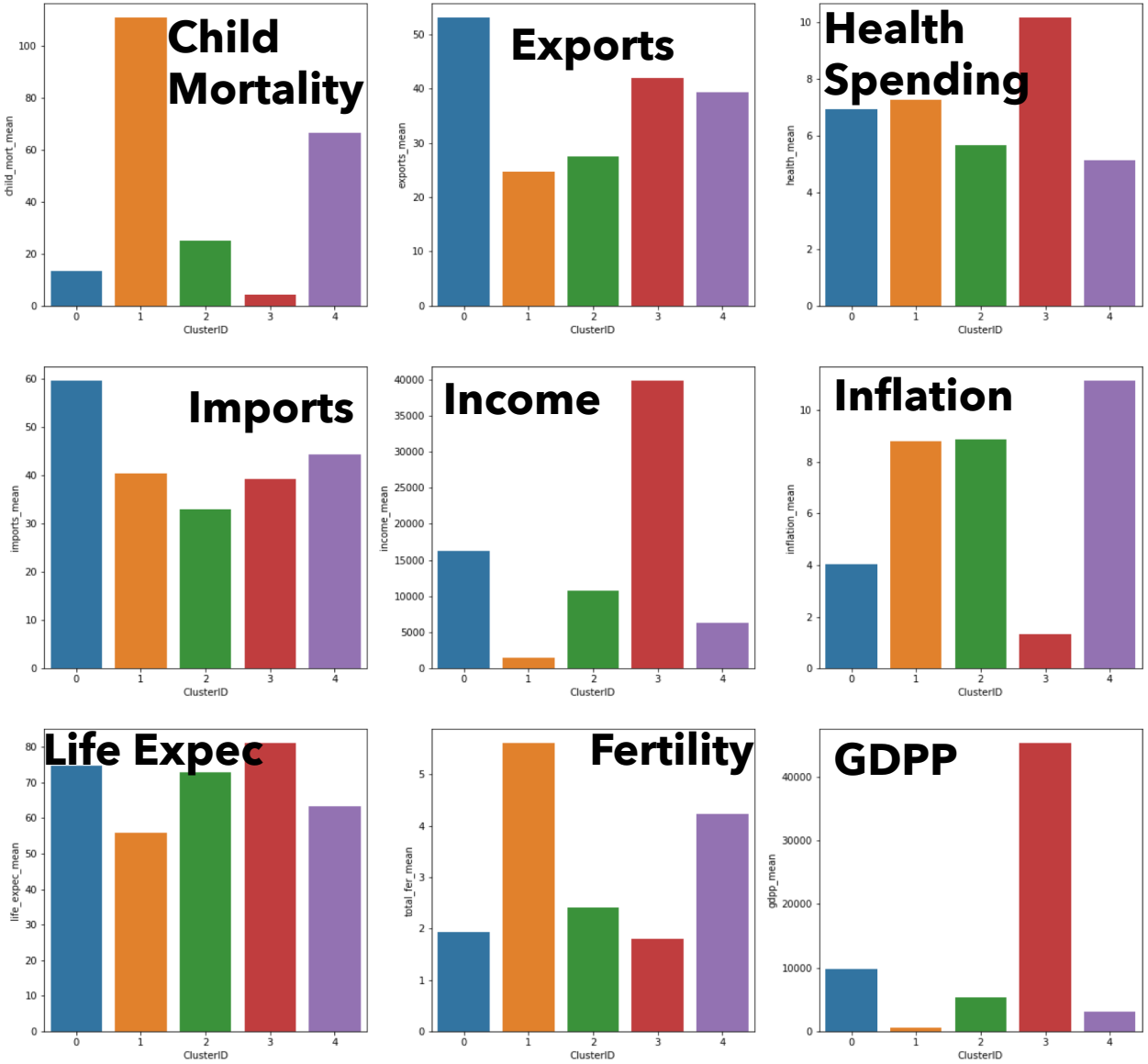


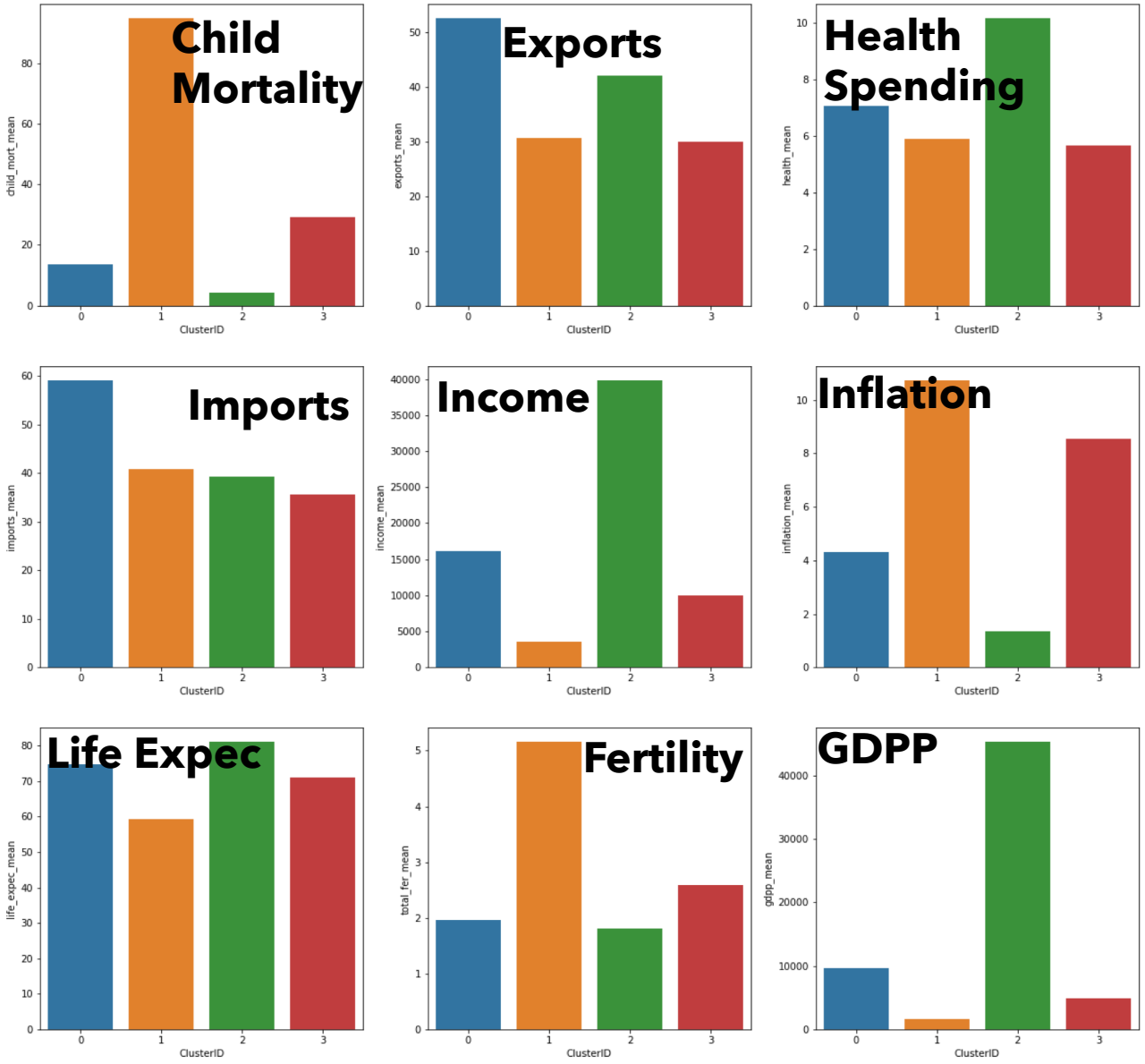
Fig 5

K-Means Clustering (K=5)



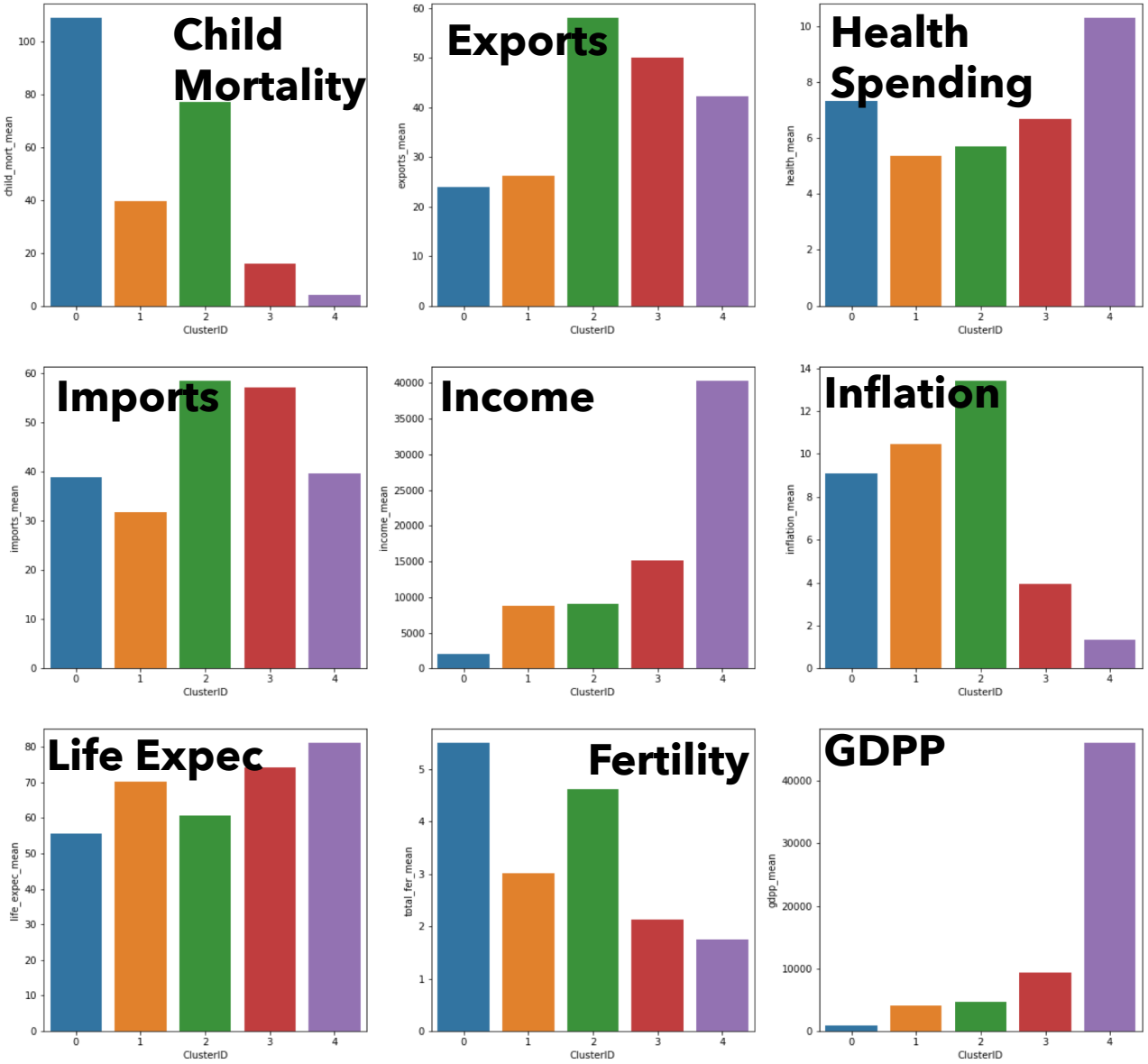
- Cluster 1 countries can be the most effective deployment of investment, as:
  - Socio-economic:
    - Low exports, High imports, Low GDPP
    - Low Income
    - High Inflation
  - Health
    - Low Life expectancy
    - High Child Mortality
    - High Fertility

K-Means Clustering (K=4)



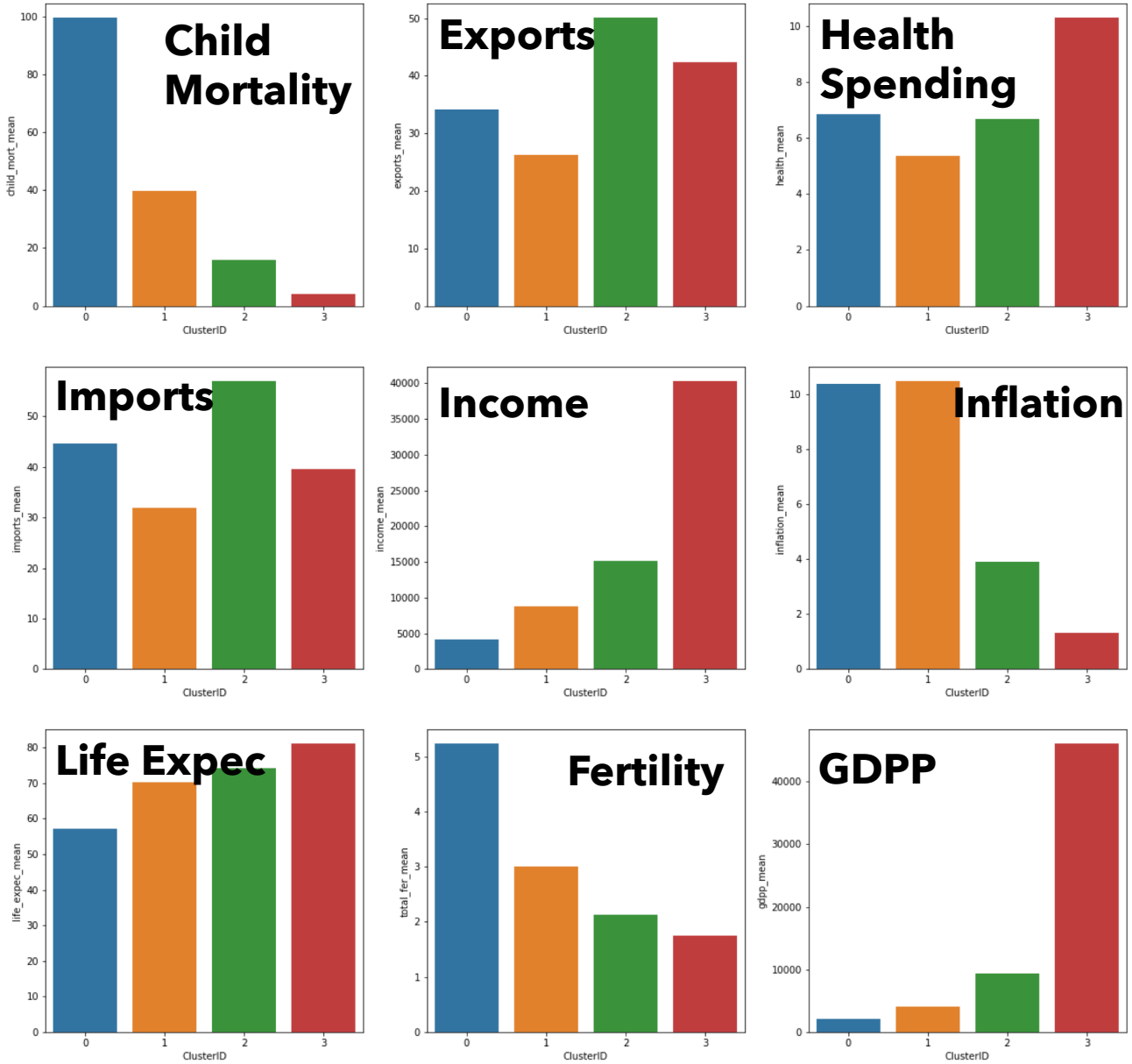
- Cluster 1 countries can be the most effective deployment of investment, as:
  - Socio-economic:
    - Low exports, High imports, Low GDPP
    - Low Income
    - High Inflation
  - Health
    - Low Life expectancy
    - High Child Mortality
    - High Fertility
    - Low Health Spending

Hierarchical Clustering (n_clusters=5)



- Cluster 0 countries can be the most effective deployment of investment, as:
  - Socio-economic:
    - Low exports, High imports, Low GDPP
    - Low Income
    - High Inflation comparatively
  - Health
    - Low Life expectancy
    - High Child Mortality
    - High Fertility

Hierarchical Clustering (n_clusters=4)



- Cluster 0 countries can be the most effective deployment of investment, as:
  - Socio-economic:
    - Low exports, High imports, Low GDPP
    - Low Income
    - High Inflation
  - Health
    - Low Life expectancy
    - High Child Mortality
    - High Fertility

K-means Clustering - Countries

K=5

1. Afghanistan
2. Burkina Faso
3. Burundi
4. Chad
5. Congo, Dem. Rep.
6. Cote d'Ivoire
7. Guinea
8. Guinea-Bissau
9. Haiti
10. Malawi
11. Mali
12. Mozambique
13. Niger
14. Rwanda
15. Sierra Leone
16. Tanzania
17. Togo
18. Uganda
19. Zambia

K=4

1. Afghanistan
2. Angola
3. Burkina Faso
4. Burundi
5. Chad
6. Comoros
7. Congo, Dem. Rep
8. Congo, Rep.
9. Cote d'Ivoire
10. Equatorial Guinea
11. Eritrea
12. Gabon
13. Gambia
14. Ghana
15. Guinea
16. Guinea-Bissau
17. Haiti
18. Kenya
19. Lao
20. Madagascar
21. Malawi
22. Mali
23. Mauritania
24. Mozambique
25. Namibia
26. Niger
27. Pakistan
28. Rwanda
29. Senegal
30. Sierra Leone
31. Sudan
32. Tanzania
33. Togo
34. Uganda
35. Yemen
36. Zambia

Hierarchical Clustering - Countries

n_cluster=5

1. Afghanistan
2. Burkina Faso
3. Burundi
4. Chad
5. Congo, Dem. Rep.
6. Cote d'Ivoire
7. Guinea
8. Guinea-Bissau
9. Haiti
10. Malawi
11. Mali
12. Mozambique
13. Niger
14. Rwanda
15. Sierra Leone
16. South Africa
17. Tanzania
18. Uganda
19. Zambia

n_cluster=4

1. Afghanistan
2. Angola
3. Botswana
4. Burkina Faso
5. Burundi
6. Chad
7. Congo, Dem. Rep
8. Congo, Rep.
9. Cote d'Ivoire
10. Equatorial Guinea
11. Guinea
12. Guinea-Bissau
13. Haiti
14. Malawi
15. Mali
16. Mauritania
17. Mozambique
18. Namibia
19. Niger
20. Rwanda
21. Sierra Leone
22. Solomon Islands
23. South Africa
24. Tanzania
25. Togo
26. Uganda
27. Zambia

Conclusion

- From all the four clusters, we select the countries which are common across all clusters
- We get 17 countries in total which match the criteria
- Socio-Economic factors:
  - Export, Import, GDPP
  - Income
  - Inflation
- Health Factors:
  - Fertility
  - Child Mortality
  - Health Spending
  - Life Expectancy

1. Afghanistan
2. Burkina Faso
3. Burundi
4. Chad
5. Cote d'Ivoire
6. Guinea
7. Guinea-Bissau
8. Haiti
9. Malawi
10. Mali
11. Mozambique
12. Niger
13. Rwanda
14. Sierra Leone
15. Tanzania
16. Uganda
17. Zambia