Help International Aid Allocation

Name: VIKUL AGGARWAL

Background

Mission of Help International: To fight poverty and provide the people of backward countries with basic amenities and relief during the time of disasters and natural calamities.

Task at hand: To categorize the given countries using socio-economic and health factors that determine the overall development of the countries. This categorization is then to be used to suggest the countries which are in direst need of aid. These suggestions will be used for strategic and efficient spending of the around $10 Million that have been raised.

Business objective: To identify the countries that are in the direst need of aid so that funds can be allocated for welfare.

Process used for the problem statement: Principal Component Analysis and Clustering.

DATA UNDERSTANDING

Types of variables:

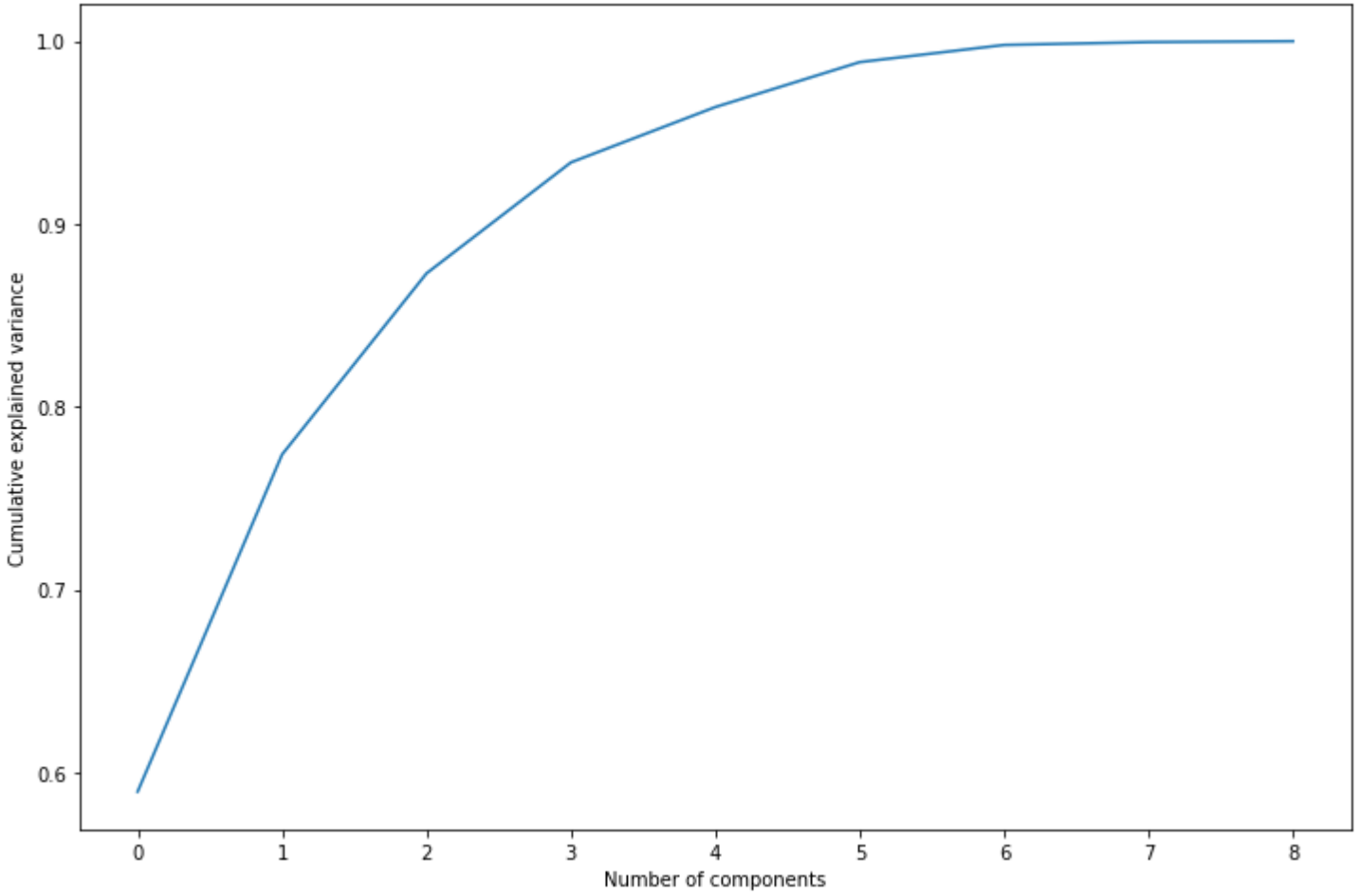
1. Variables related to economic state of a country
2. Variables related to health factors

Socio-Economic Indicators	Health Factors
GDP per Capita	Child Mortality Rate
Income	Life Expectancy
Inflation	Total Fertility
Exports	Health Spending
Imports	

Spread of the Data

	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
count	167.00	167.00	167.00	167.00	167.00	167.00	167.00	167.00	167.00
mean	38.27	7420.62	1056.73	6588.35	17144.69	7.78	70.56	2.95	12964.16
std	40.33	17973.89	1801.41	14710.81	19278.07	10.57	8.89	1.51	18328.70
min	2.60	1.08	12.82	0.65	609.00	-4.21	32.10	1.15	231.00
25%	8.25	447.14	78.54	640.21	3355.00	1.81	65.30	1.79	1330.00
50%	19.30	1777.44	321.89	2045.58	9960.00	5.39	73.10	2.41	4660.00
75%	62.10	7278.00	976.94	7719.60	22800.00	10.75	76.80	3.88	14050.00
max	208.00	183750.00	8663.60	149100.00	125000.00	104.00	82.80	7.49	105000.00

Cumulative Explained Variation vs No of PCs

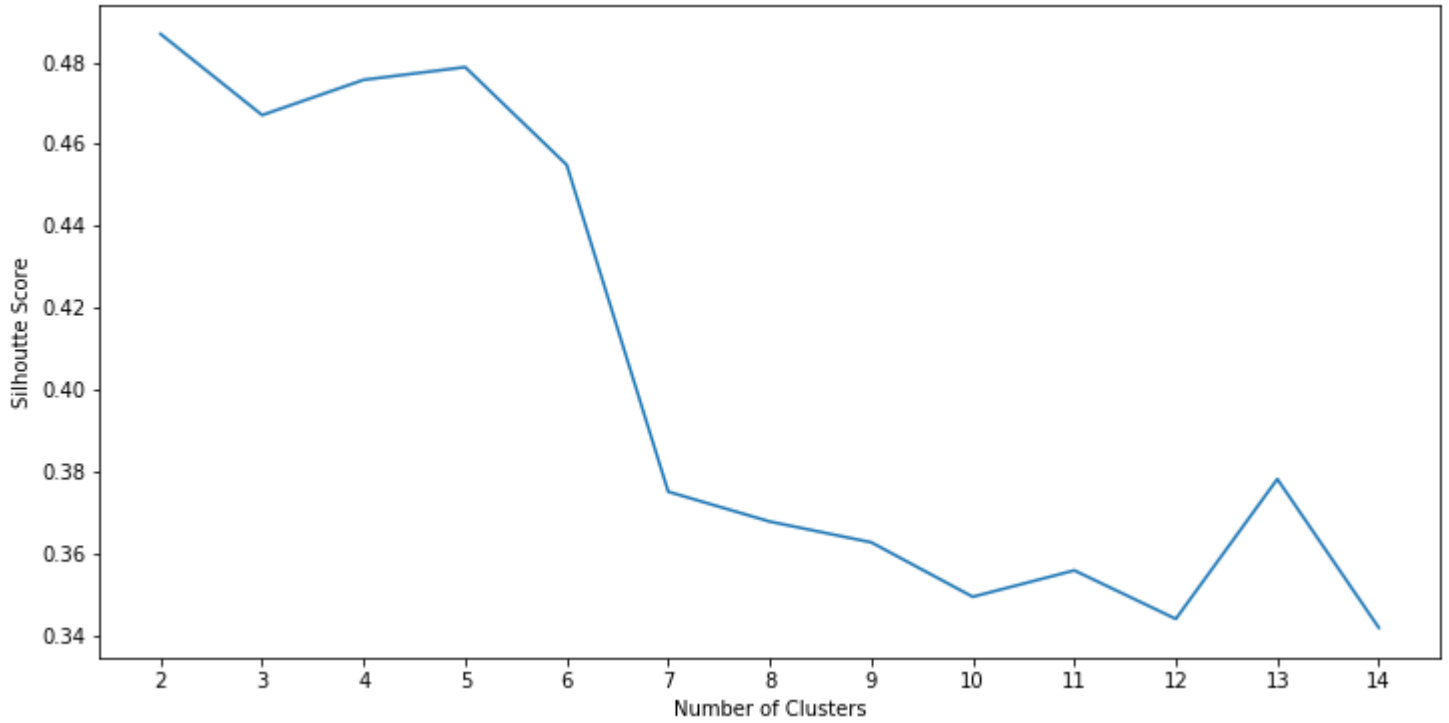


4 Principal Components are enough to explain ~95% of variance in the data.

Hopkins Statistic

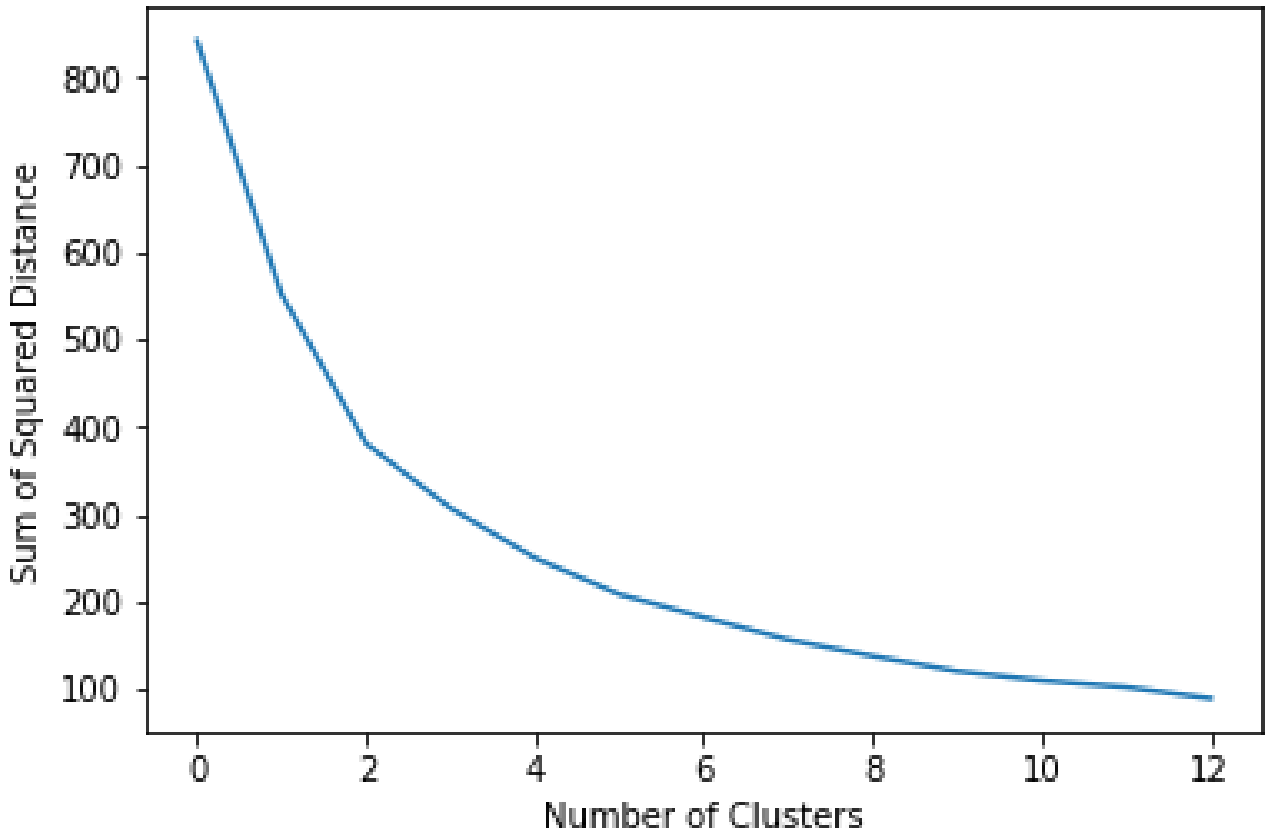
The Hopkin’s Statistic for PC Modified data was found to be more than **0.9** showing a high clustering tendency.

Silhouette Analysis for K-Means Clustering



Silhouette score is high for k between 2 to 5, and then drops drastically beyond 5.

Elbow Curve Analysis



Points 2 to 5 look like the elbow of the curve, i.e. the appropriate number of clusters that can be taken.

K-Means with K=3

Following are the number of countries in each cluster when K-Means is performed with K=3:

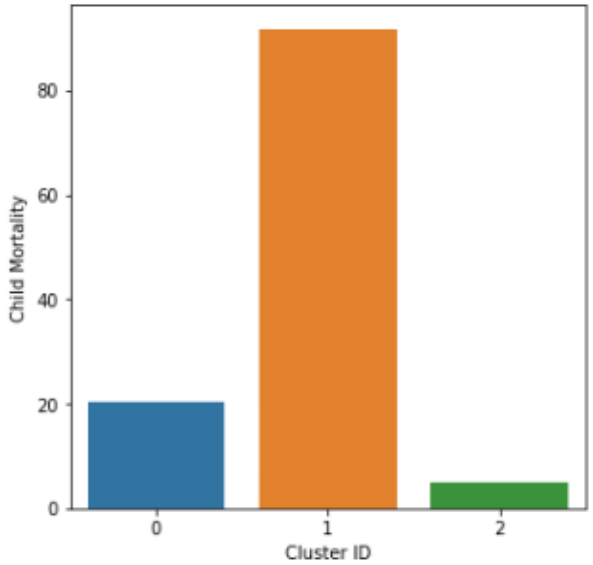
- Cluster 0 – 91
- Cluster 1 – 48
- Cluster 2 – 28

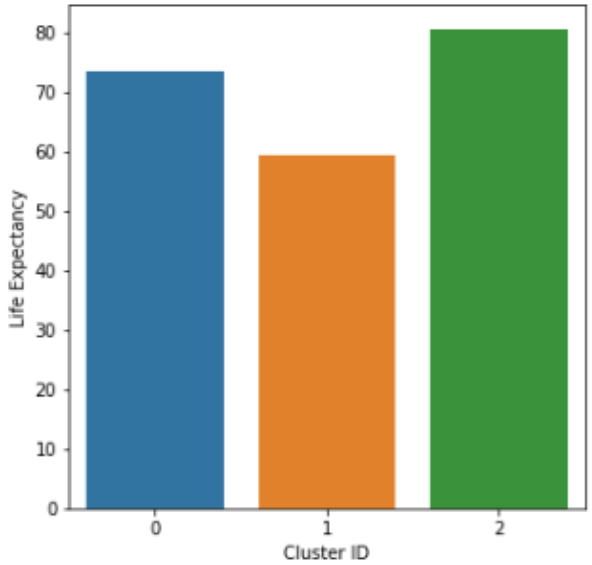
Following are the cluster-wise mean values for the features:

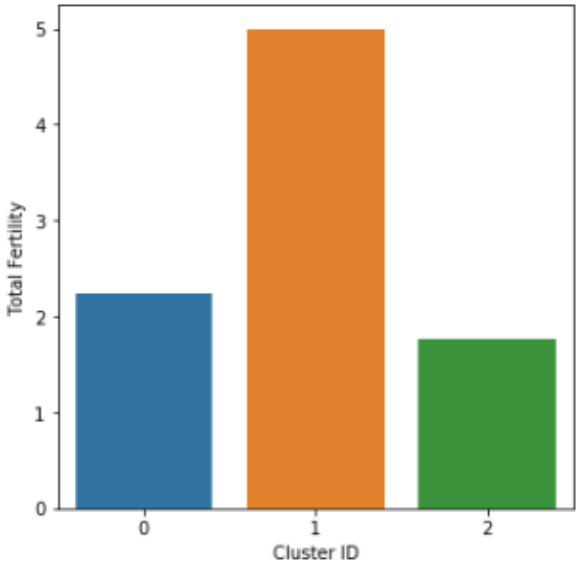
	cluster_id	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	0	20.36	3604.15	547.28	3710.45	13968.02	7.07	73.46	2.24	7979.91
1	1	91.61	879.06	114.82	827.03	3897.35	11.91	59.24	4.99	1909.21
2	2	5.05	31038.24	4327.16	25818.10	50178.57	3.01	80.51	1.76	48114.29

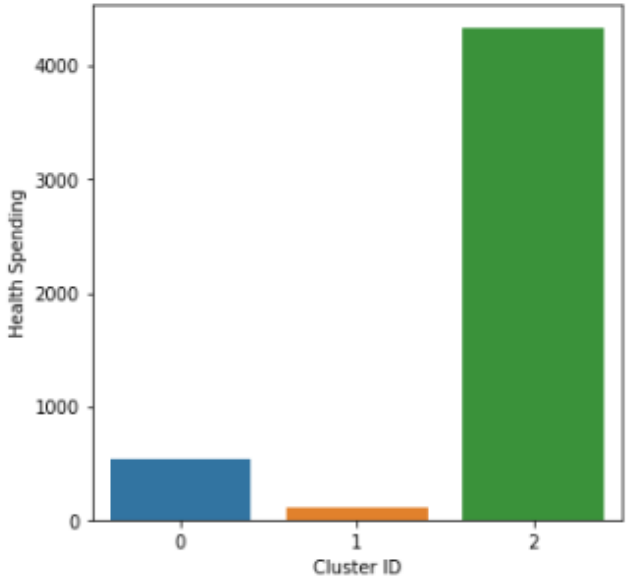
It is evident from the above values that the cluster representation is as follows:
Developed - Cluster 2 (High values for GDPP, Income, Life Expectancy; Very low values for child mortality etc.)
Developing - Cluster 0 (Moderate values for GDPP, Income, Life Expectancy, child mortality etc.)
Under-Developed - Cluster 1 (Low values for GDPP, Income, Life Expectancy; high values for child mortality etc.)

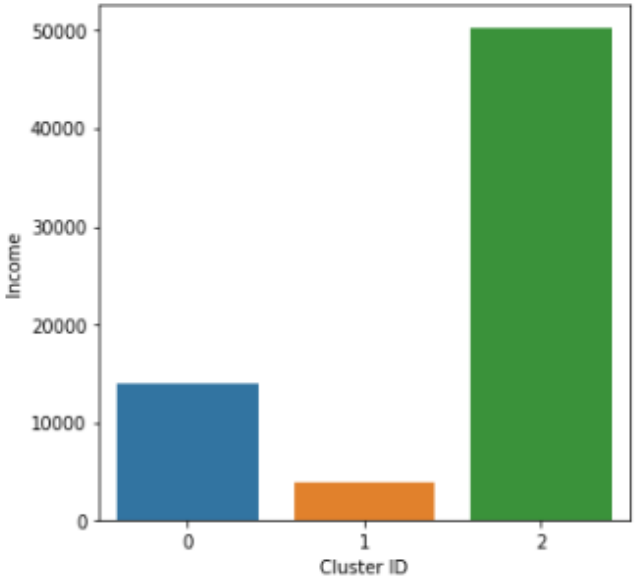
Cluster-wise mean values - Plots

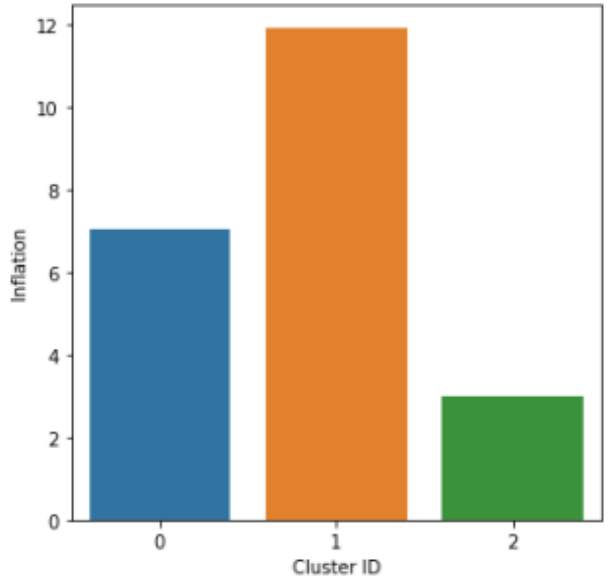




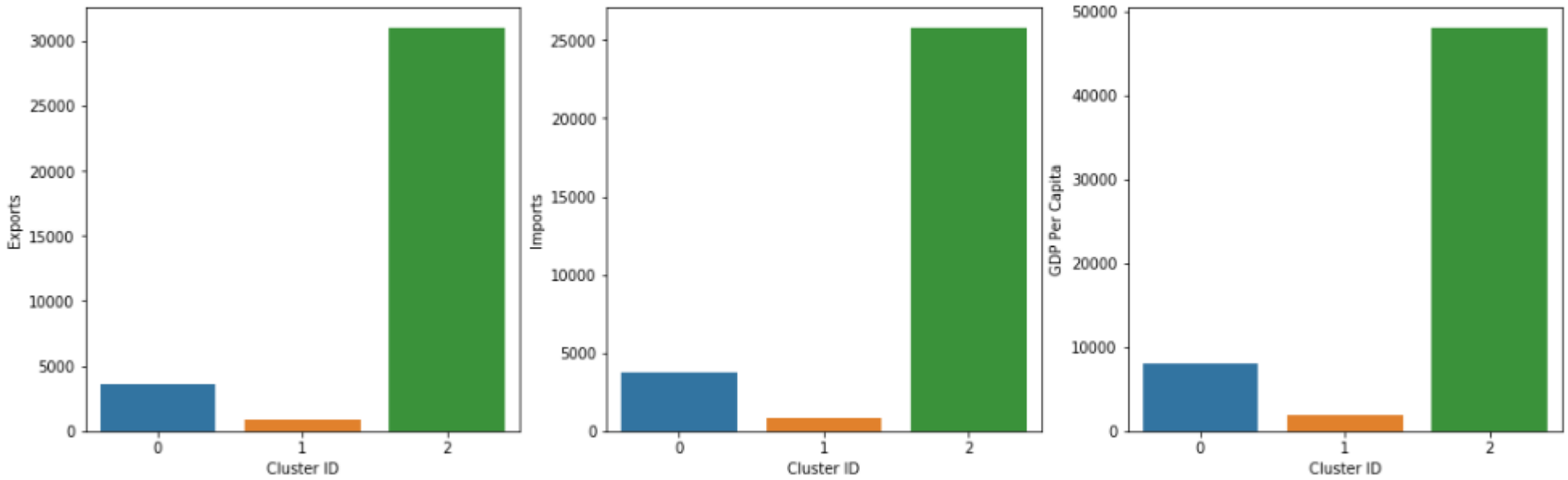






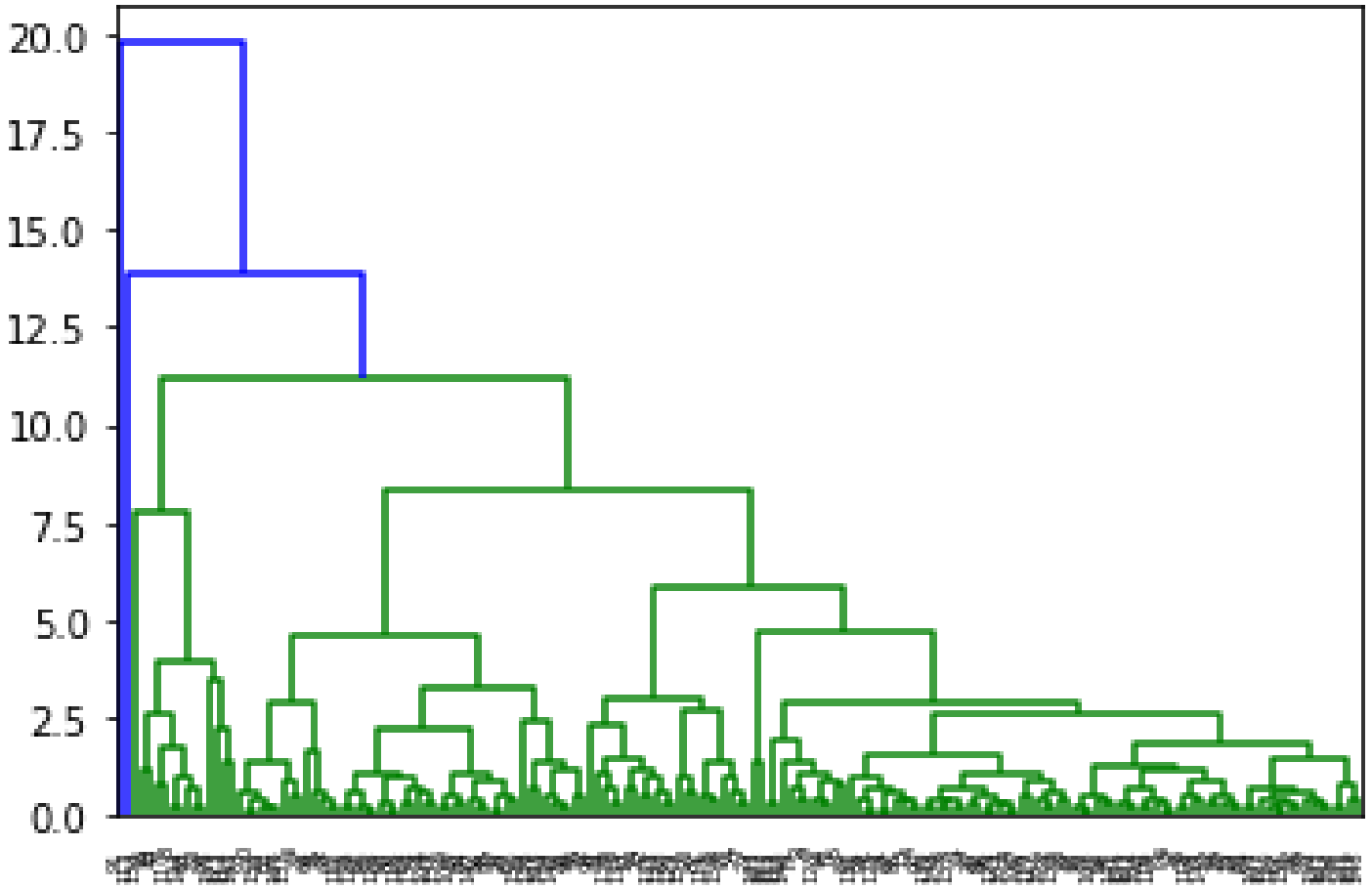


Cluster-wise mean values – Plots (Contd..)



Looking at the cluster-wise mean values, it is evident that **the countries in cluster with cluster ID 1 need to be analyzed to find out the countries in direst need of aid.**

Hierarchical Clustering – The Dendogram



Cluster-wise mean values for features after Hierarchical Clustering

Following are the cluster-wise mean values for the features after using n_clusters = 5:

	cluster_id	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	0	90.79	885.22	114.75	836.00	3870.70	9.95	59.21	4.97	1900.26
1	1	18.56	5115.82	857.67	4943.14	17679.71	6.72	74.32	2.17	11638.19
2	2	4.37	34375.21	5257.60	28383.62	53264.29	1.83	80.63	1.79	54142.86
3	3	2.80	183750.00	8158.50	149100.00	91700.00	3.62	81.30	1.63	105000.00
4	4	130.00	589.49	118.13	405.42	5150.00	104.00	60.50	5.84	2330.00

Looking at the cluster-wise mean values, it is evident that **the countries in cluster with cluster ID 0 and 4 need to be analyzed to find out the countries in direst need of aid.**

Countries in dire need of aid

Following are the countries in dire need of aid. Chosen clusters were analyzed and countries with child mortality more than the 95th percentile, life expectancy less than the 5th percentile, health spending less than the 5th percentile and total fertility more than the 95th percentile were picked.

Haiti is in dire need of aid since it has very high child mortality rate and very low expectancy. It's possible that this country is prevalant with diseases, or lack of resources that lead to children dying in early stages of life, and adults not being able to stay alive for long.

Sierra Leone and Chad are in dire need of aid to tackle very high child mortality rate by spending more in child-care

Lesotho and Central African Republic are in dire need of aid to tackle very low life expectancy by spending more in healthcare programs.

Niger and Mali can be alloted some aid to spend in awareness programs to educate people about birth-control since Niger has very high total fertility rate.

Mozambique and Congo, Dem. Rep. are two other countries which can be alloted aid. These countries have very low spendings on health, high child mortality rate and low life expectancy.

Nigeria, Angola, Burkina Faso and Guinea-Bissau are other countries which need aid since they have high child mortality rate and low life expectancy.