

FINAL PROJECT:

Clustering the Countries by using K-Means for HELP
International

Auliya Tri Meliyani - 2024



OBJECTIVE

Untuk mengkategorikan negara menggunakan faktor sosial ekonomi dan kesehatan yang menentukan pembangunan negara secara keseluruhan

TENTANG ORGANISASI

HELP International adalah LSM kemanusiaan internasional yang berkomitmen untuk memerangi kemiskinan dan menyediakan fasilitas dan bantuan dasar bagi masyarakat di negara-negara terbelakang saat terjadi bencana dan bencana alam.

PERMASALAHAN

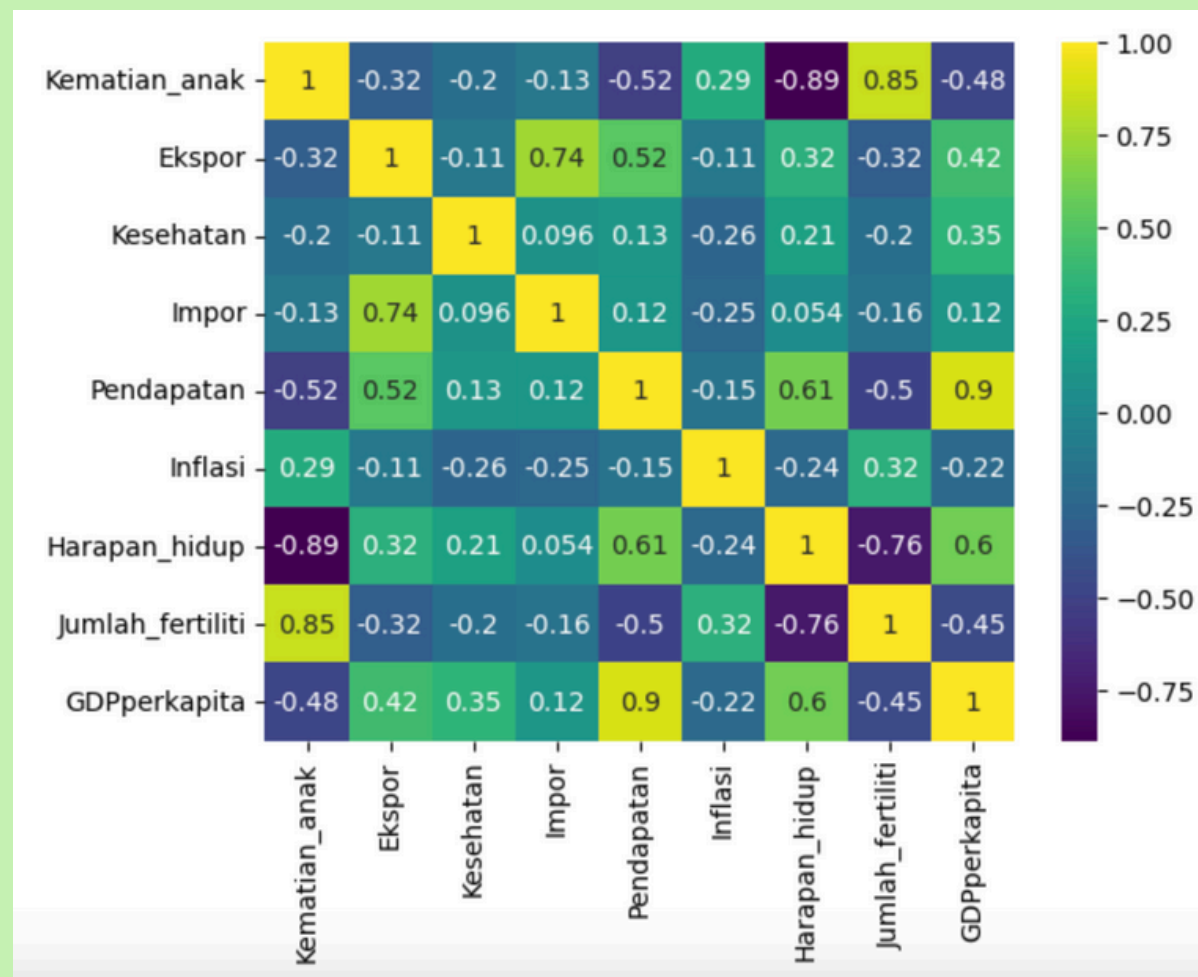
HELP International telah berhasil mengumpulkan sekitar \$ 10 juta. Saat ini, CEO LSM perlu memutuskan bagaimana menggunakan uang ini secara strategis dan efektif. Jadi, CEO harus mengambil keputusan untuk memilih negara yang paling membutuhkan bantuan. Oleh karena itu, Tugas teman-teman adalah mengkategorikan negara menggunakan beberapa faktor sosial ekonomi dan kesehatan yang menentukan perkembangan negara secara keseluruhan. Kemudian kalian perlu menyarankan negara mana saja yang paling perlu menjadi fokus CEO

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

Terdapat berbagai data yang akan dianalisis dalam tabel diantaranya yaitu Negara, Kematian Anak, Ekspor, Kesehatan, Impor, Pendapatan, Inflasi, Harapan Hidup, Jumlah Fertiliti, dan GDP perkapita

	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000
mean	38.270060	41.108976	6.815689	46.890215	17144.688623	7.781832	70.555689	2.947964	12964.155689
std	40.328931	27.412010	2.746837	24.209589	19278.067698	10.570704	8.893172	1.513848	18328.704809
min	2.600000	0.109000	1.810000	0.065900	609.000000	-4.210000	32.100000	1.150000	231.000000
25%	8.250000	23.800000	4.920000	30.200000	3355.000000	1.810000	65.300000	1.795000	1330.000000
50%	19.300000	35.000000	6.320000	43.300000	9960.000000	5.390000	73.100000	2.410000	4660.000000
75%	62.100000	51.350000	8.600000	58.750000	22800.000000	10.750000	76.800000	3.880000	14050.000000
max	208.000000	200.000000	17.900000	174.000000	125000.000000	104.000000	82.800000	7.490000	105000.000000

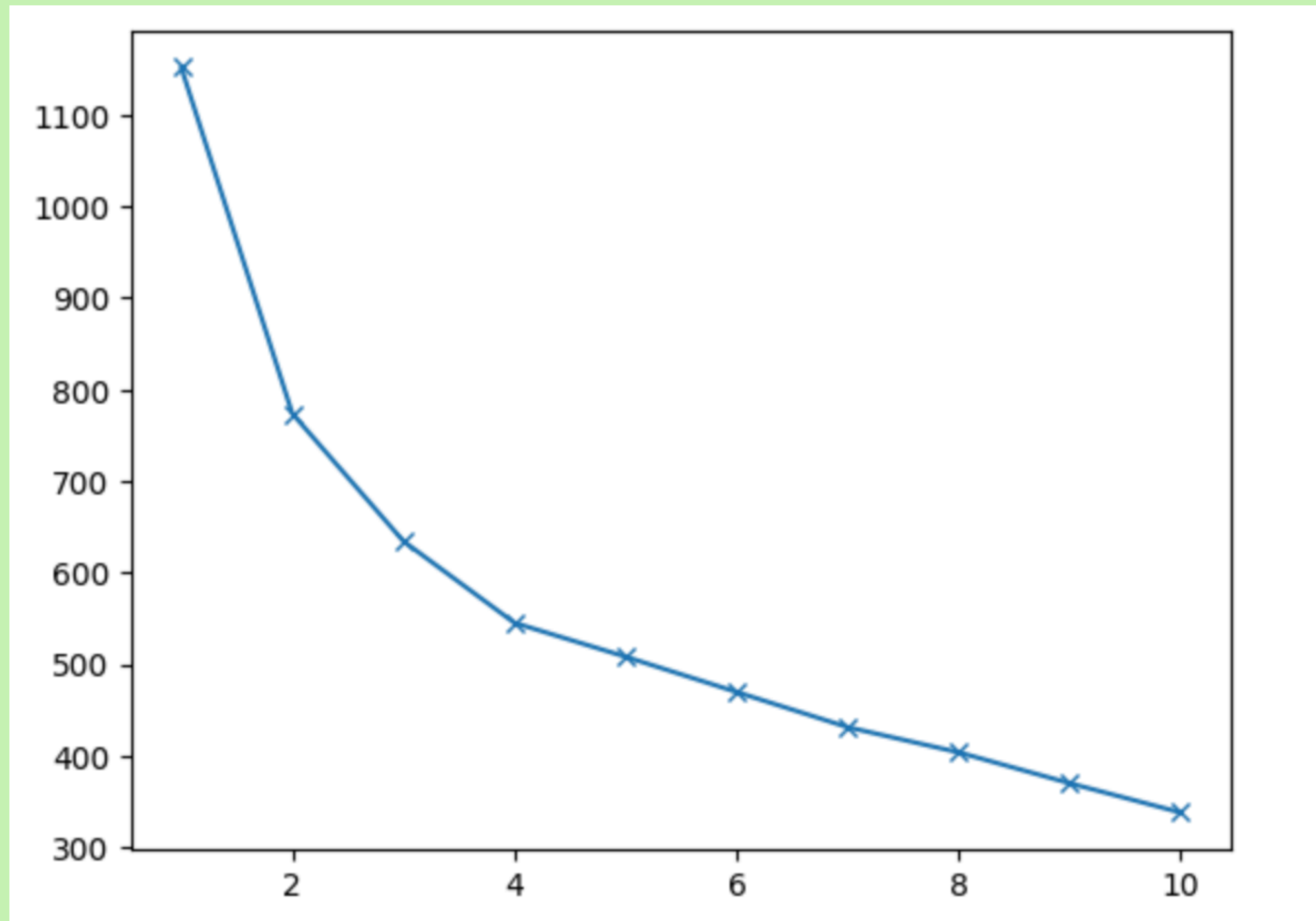
Berdasarkan data tersebut didapatkan nilai max dan min yang cukup signifikan untuk setiap kategori. Hal ini menunjukkan bahwa terdapat adanya kesenjangan di beberapa negara terhadap kategori yang akan dianalisis.



Berdasarkan data korelasi didapatkan bahwa adanya korelasi yang tinggi antara GDP perkapita dengan pendapatan. GDP juga berkaitan dengan ekspor, kesehatan, dan harapan hidup. Kemudian, ada juga korelasi yang tinggi antara kematian anak dengan jumlah fertiliti.

Kemudian, terdapat juga korelasi yang dapat dikategorikan sebagai hubungan ekonomi dan kesehatan yaitu adanya hubungan yang tidak signifikan antara kematian anak dan pendapatan. Hal tersebut berarti apabila pendapatan rendah, maka akan terjadi kematian anak yang tinggi, begitu juga sebaliknya.

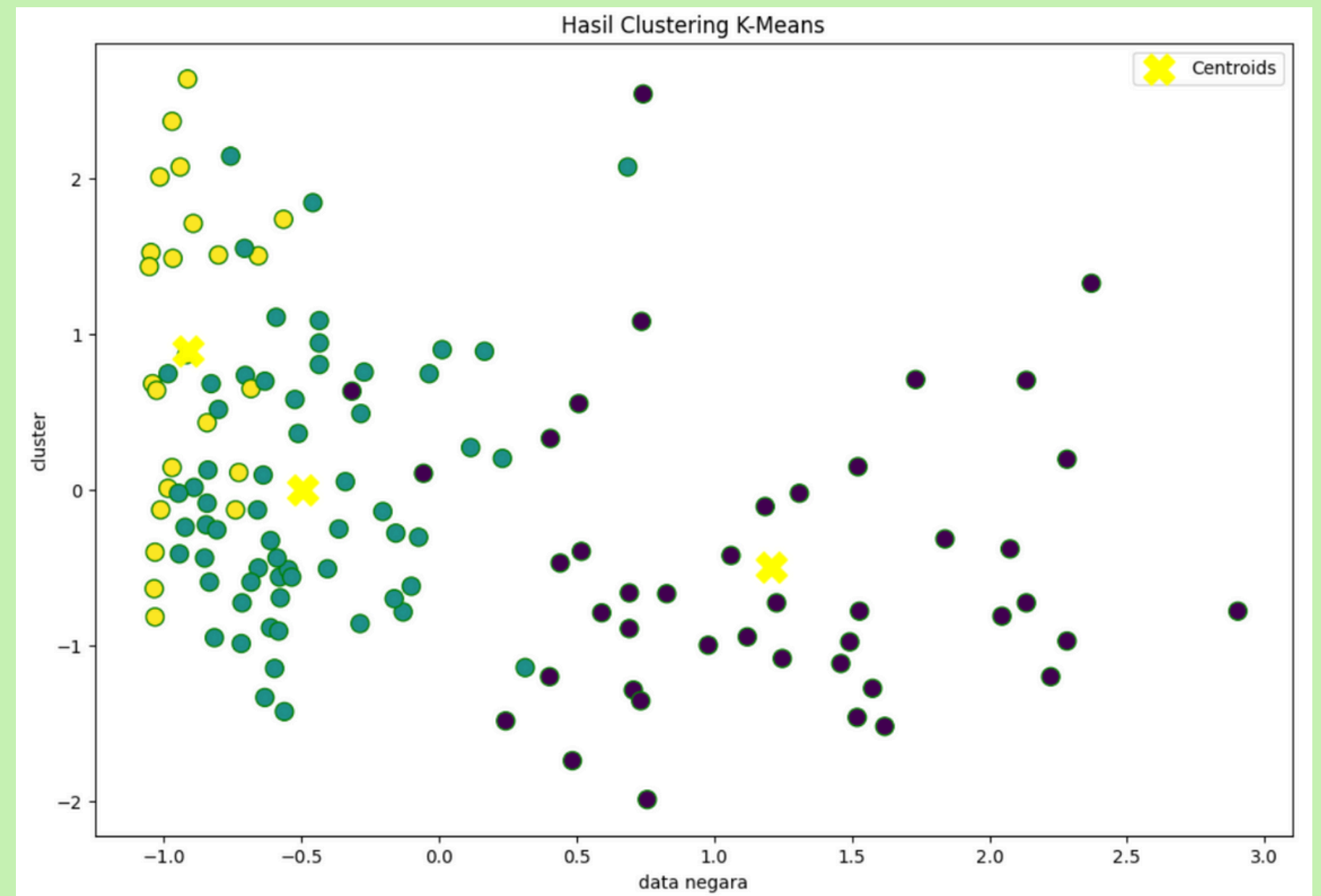
Maka dari itu secara ideal data utama yang akan dilihat untuk clustering adalah GDP dan pendapatan untuk kategori ekonomi, serta kematian anak untuk kategori kesehatan.



Nilai K akan digunakan untuk clustering pada metode k-means sehingga nilai ini sangat penting untuk hasil akhir analisis data

Berdasarkan metode Elbow didapatkan bahwa nilai K yaitu 3

Berdasarkan grafik scatter didapatkan hasil clustering K-Means dengan nilai $K=3$ yaitu seperti gambar disamping. Data dengan lingkaran berwarna kuning merupakan data dengan nilai cluster 2, lingkaran berwarna hijau merupakan data dengan nilai cluster 1, dan lingkaran berwarna hitam merupakan data dengan nilai 0.

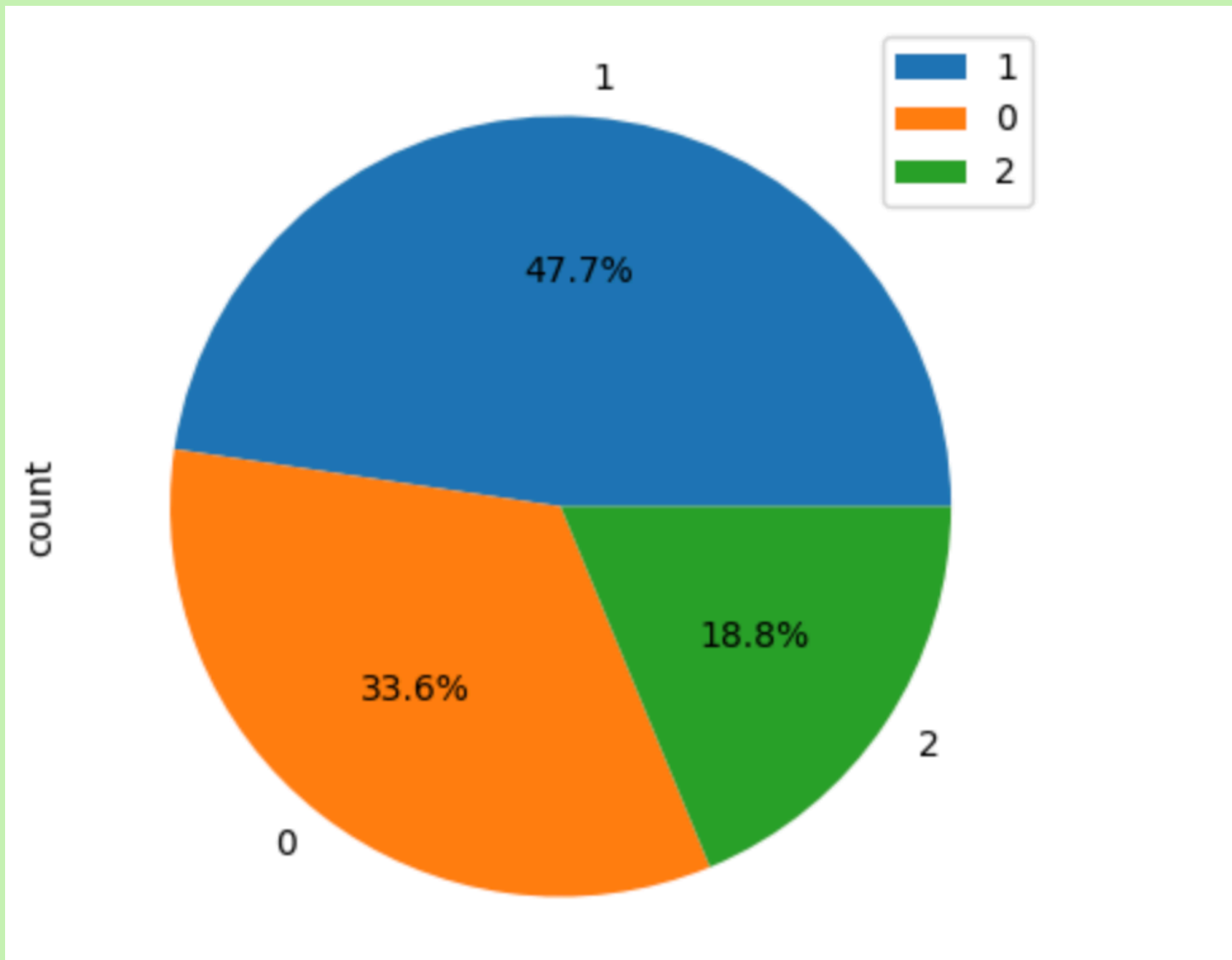


	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	label	kmeans_clusters
0	90.2	10.0	7.58	44.9	1610.0	9.440	56.2	5.82	553.0	0	0
1	16.6	28.0	6.55	48.6	9930.0	4.490	76.3	1.65	4090.0	1	1
2	27.3	38.4	4.17	31.4	12900.0	16.100	76.5	2.89	4460.0	1	1
3	119.0	62.3	2.85	42.9	5900.0	22.400	60.1	6.16	3530.0	0	0
4	10.3	45.5	6.03	58.9	19100.0	1.440	76.8	2.13	12200.0	2	2
5	14.5	18.9	8.10	16.0	18700.0	20.900	75.8	2.37	10300.0	1	1
6	18.1	20.8	4.40	45.3	6700.0	7.770	73.3	1.69	3220.0	1	1
9	39.2	54.3	5.88	20.7	16000.0	13.800	69.1	1.92	5840.0	1	1
10	13.8	35.0	7.89	43.7	22900.0	-0.393	73.8	1.86	28000.0	2	2
11	8.6	69.5	4.97	50.9	41100.0	7.440	76.0	2.16	20700.0	2	2

Analisis jenis kategori berdasarkan grafik scatter dapat dilengkapi berdasarkan nilai tabel clustering yaitu dapat dilihat bahwa cluster 0 dapat di kategorikan sebagai negara dengan ekonomi dan kesehatan rendah. Cluster 1 merupakan negara dengan ekonomi dan kesehatan menengah. Cluster 2 merupakan negara dengan ekonomi dan kesehatan tinggi.

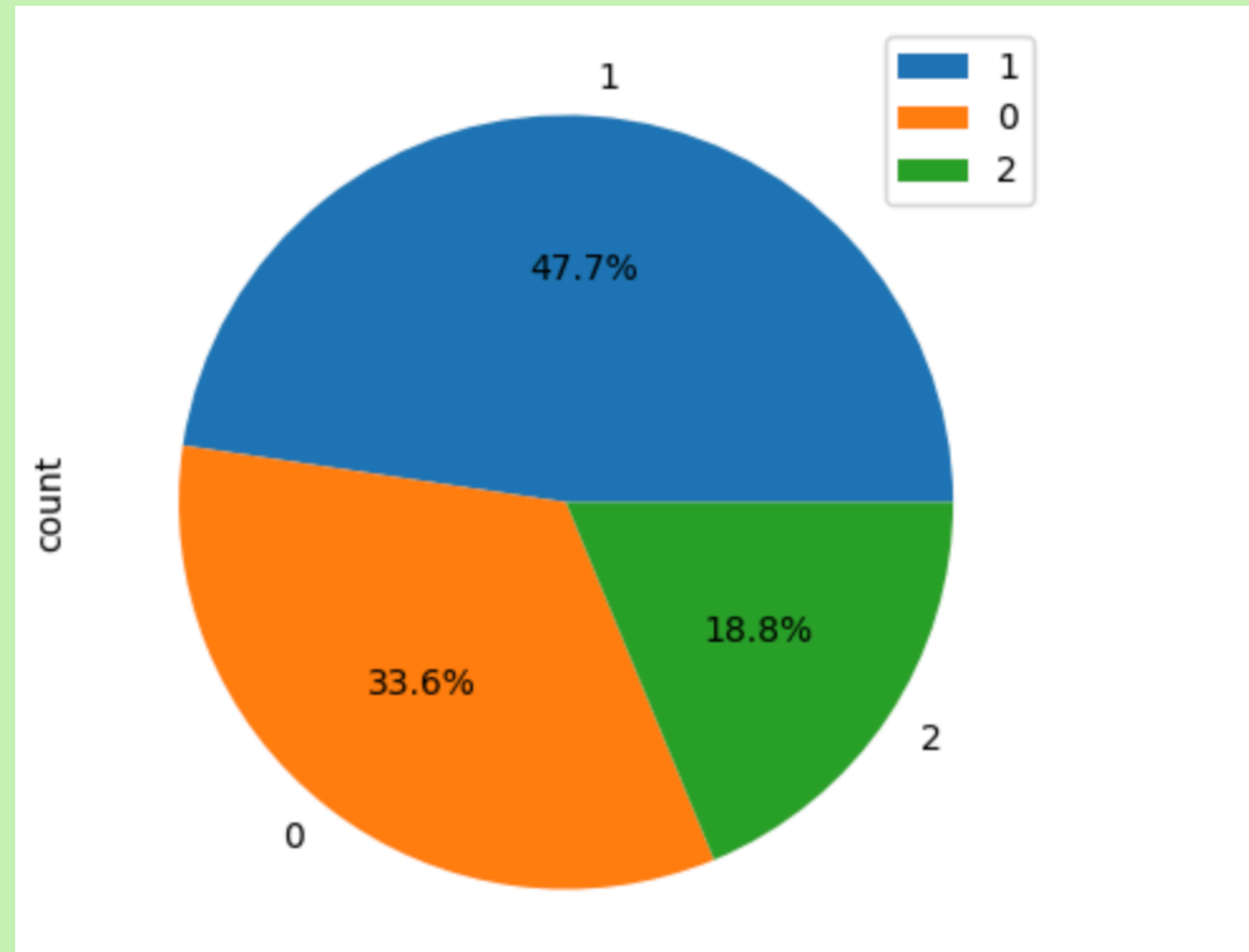
cluster	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	label	kmeans_clusters
0	79.688372	28.162767	6.107209	41.143393	3485.372093	9.818721	61.058140	4.705814	1589.720930	0.0	0.0
1	22.095082	37.378689	6.221475	45.190164	10349.344262	7.164082	72.903279	2.297049	5064.655738	1.0	1.0
2	7.995833	54.150000	7.085000	52.104167	26916.666667	3.344625	77.012500	1.841250	18552.083333	2.0	2.0

Kemudian, data tabel clustering juga didukung oleh data rata-rata nilai cluster yaitu semakin kecil (0) nilai kmeans_clusters, maka semakin besar tingkat kematian anak, sedangkan nilai lainnya semakin kecil. Sebaliknya semakin besar (2) nilai kmeans_clusters, maka semakin kecil tingkat kematian anak, sedangkan nilai lainnya semakin besar. Terakhir, nilai kmeans_clusters menengah (1) berada diantaranya.



Berdasarkan grafik pie disamping, jumlah negara dengan kluster 1 (menengah) menjadi nilai tertinggi, diikuti dengan jumlah negara kluster 0 (rendah), dan terakhir kluster 2 (tinggi).

Berdasarkan grafik pie tersebut maka, jumlah negara dengan kluster 0 (terendah) menempati nilai ke-2 sehingga perlu dilakukan kembali pengurutan negara agar dana yang diberikan tepat guna



Clustering dilakukan dengan mengurutkan GDP, pendapatan, dan kematian anak

	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	label	kmeans_clusters	cluster
26	93.6	8.92	11.60	39.2	764.0	12.30	57.7	6.26	231.0	0	0	0
88	89.3	19.10	11.80	92.6	700.0	5.47	60.8	5.02	327.0	0	0	0
37	116.0	41.10	7.91	49.6	609.0	20.80	57.5	6.54	334.0	0	0	0
106	101.0	31.50	5.21	46.2	918.0	7.64	54.5	5.56	419.0	0	0	0
94	90.5	22.80	6.59	34.9	1030.0	12.10	53.1	5.31	459.0	0	0	0
150	90.3	40.20	7.65	57.3	1210.0	1.18	58.7	4.87	488.0	0	0	0
64	114.0	14.90	8.50	35.2	1390.0	2.97	55.6	5.05	547.0	0	0	0
0	90.2	10.00	7.58	44.9	1610.0	9.44	56.2	5.82	553.0	0	0	0
56	80.3	23.80	5.69	42.7	1660.0	4.30	65.5	5.71	562.0	0	0	0
25	116.0	19.20	6.74	29.6	1430.0	6.81	57.9	5.87	575.0	0	0	0

Data tabel tersebut diambil berdasarkan faktor ekonomi (pendapatan dan GDP) serta kesehatan (kematian anak) dari nilai rata-rata dari setiap kluster. Kemudian, diurutkan kembali berdasarkan nilai GDP karena nilai GDP mempengaruhi banyak faktor lain dalam ekonomi. Maka dapat diambil kesimpulan bahwa 10 negara tersebut merupakan negara yang paling membutuhkan bantuan dari HELP organization berdasarkan data ekonomi dan kesehatan.

Clustering dilakukan dengan mengurutkan GDP, pendapatan, dan kematian anak

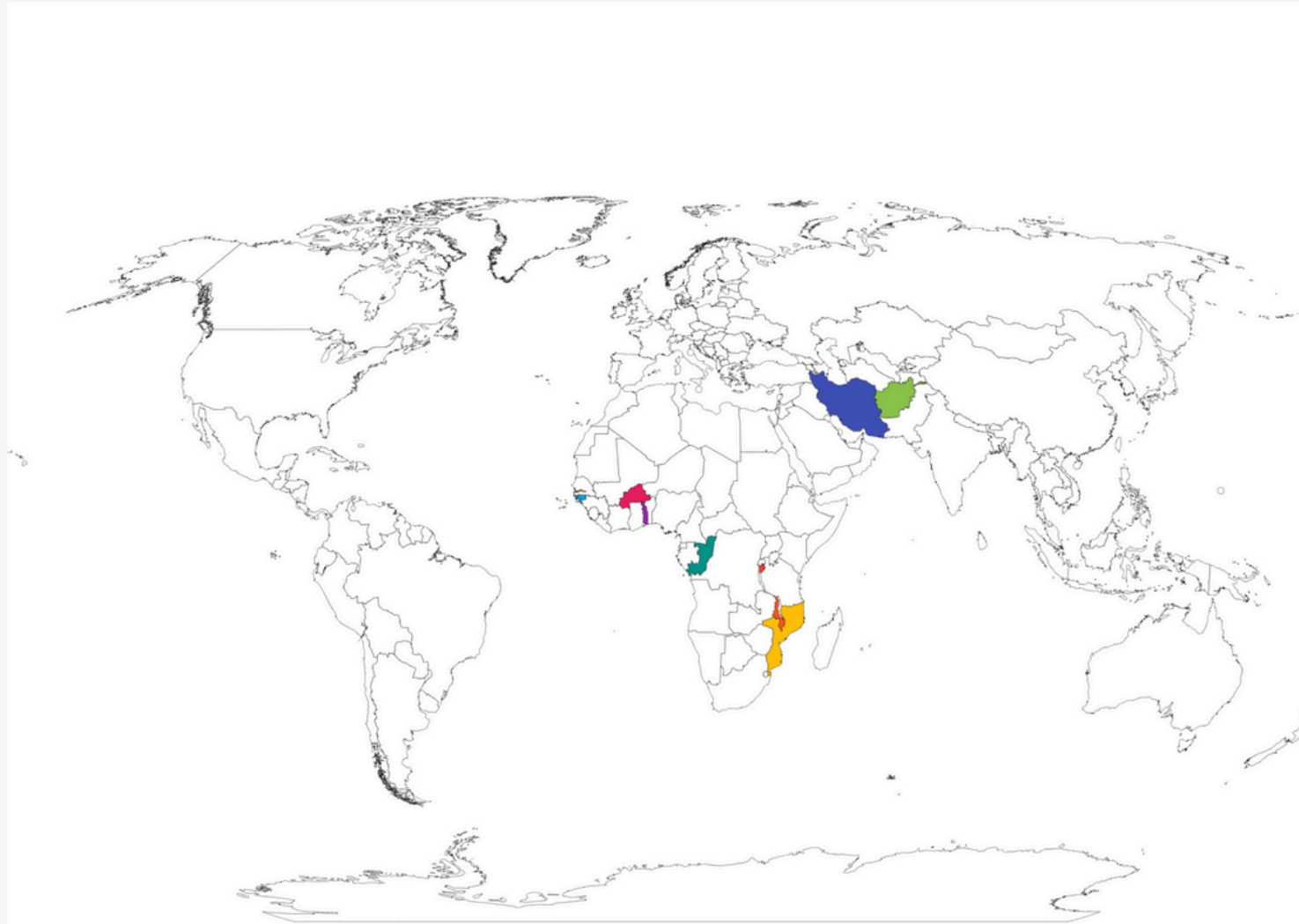
	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	label	kmeans_clusters	cluster
26	93.6	8.92	11.60	39.2	764.0	12.30	57.7	6.26	231.0	0	0	0
88	89.3	19.10	11.80	92.6	700.0	5.47	60.8	5.02	327.0	0	0	0
37	116.0	41.10	7.91	49.6	609.0	20.80	57.5	6.54	334.0	0	0	0
106	101.0	31.50	5.21	46.2	918.0	7.64	54.5	5.56	419.0	0	0	0
94	90.5	22.80	6.59	34.9	1030.0	12.10	53.1	5.31	459.0	0	0	0
150	90.3	40.20	7.65	57.3	1210.0	1.18	58.7	4.87	488.0	0	0	0
64	114.0	14.90	8.50	35.2	1390.0	2.97	55.6	5.05	547.0	0	0	0
0	90.2	10.00	7.58	44.9	1610.0	9.44	56.2	5.82	553.0	0	0	0
56	80.3	23.80	5.69	42.7	1660.0	4.30	65.5	5.71	562.0	0	0	0
25	116.0	19.20	6.74	29.6	1430.0	6.81	57.9	5.87	575.0	0	0	0

10 negara secara berurutan yaitu Burundi, Liberia, Congo, Dem. Rep., Mozambique, Malawi, Togo, Guinea-Bissau, Afghanistan, Gambia, Burkina Faso.

Visualisasi Peta Negara

Bantuan HELP untuk 10 Negara

- Burundi
- Liberia
- Congo
- Mozambique
- Malawi
- Togo
- Guinea-Bissau
- Afghanistan
- Gambia
- Burkina Faso



Visualisasi peta untuk negara yang dapat dipertimbangkan oleh CEO HELP agar segera mendapat bantuan baik dari segi ekonomi maupun kesehatan.

data svg secara live(https://drive.google.com/file/d/lrHHqYjHYvAqgUbfLQVrV6alB6uKW7tBx/view?usp=drive_link)

TERIMA KASIH