

SANBERCODE FINAL PROJECT

CLUSTERING THE COUNTRIES BY USING
K-MEANS FOR HELP INTERNATIONAL

INTRODUCTION

LATAR BELAKANG

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.

OBJEKTIF

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

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.

THE DATA

167 ROWS
10 COLUMNS

	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
...
162	Vanuatu	29.2	46.6	5.25	52.7	2950	2.62	63.0	3.50	2970
163	Venezuela	17.1	28.5	4.91	17.6	16500	45.90	75.4	2.47	13500
164	Vietnam	23.3	72.0	6.84	80.2	4490	12.10	73.1	1.95	1310
165	Yemen	56.3	30.0	5.18	34.4	4480	23.60	67.5	4.67	1310
166	Zambia	83.1	37.0	5.89	30.9	3280	14.00	52.0	5.40	1460

Negara : Nama negara

Kematian_anak: Kematian anak di bawah usia 5 tahun per 1000 kelahiran

Ekspor : Ekspor barang dan jasa perkapita

Kesehatan: Total pengeluaran kesehatan perkapita

Impor: Impor barang dan jasa perkapita

Pendapatan: Penghasilan bersih perorang

Inflasi: Pengukuran tingkat pertumbuhan tahunan dari Total GDP

Harapan_hidup: Jumlah tahun rata-rata seorang anak yang baru lahir akan hidup jika pola kematian saat ini tetap sama

Jumlah_fertiliti: Jumlah anak yang akan lahir dari setiap wanita jika tingkat kesuburan usia saat ini tetap sama

GDPperkapita: GDP per kapita. Dihitung sebagai Total GDP dibagi dengan total populasi.

EDA

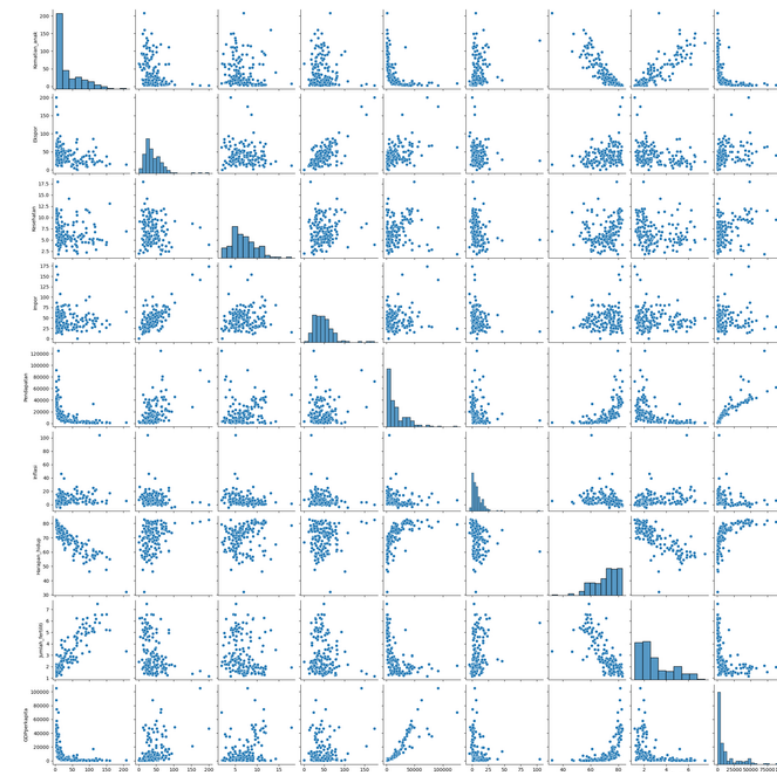
Dari beberapa variabel diatas dapat dikelompokkan sebagai berikut:

Faktor Sosial Ekonomi:

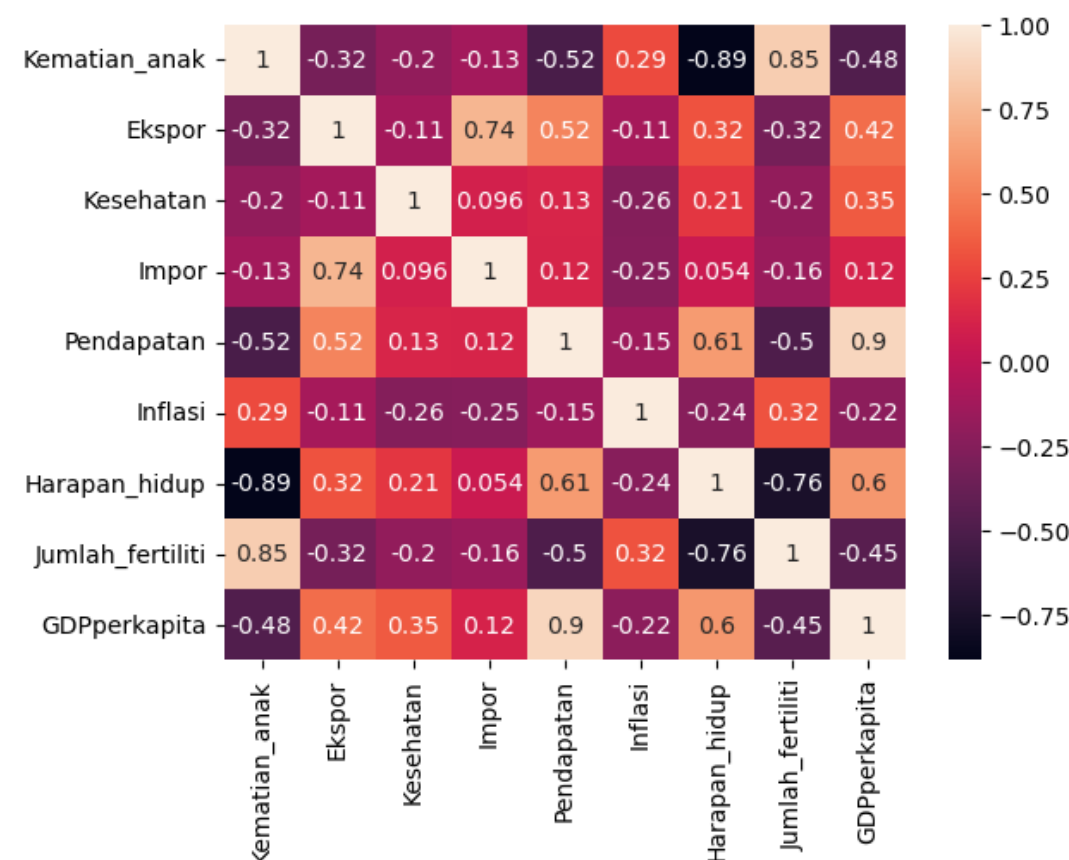
- * Ekspor
- * Kesehatan
- * Impor
- * Pendapatan
- * Inflasi
- * GDPperkapita

Faktor Kesehatan:

- * Kematian_anak
- * Harapan_hidup
- * Jumlah_fertiliti



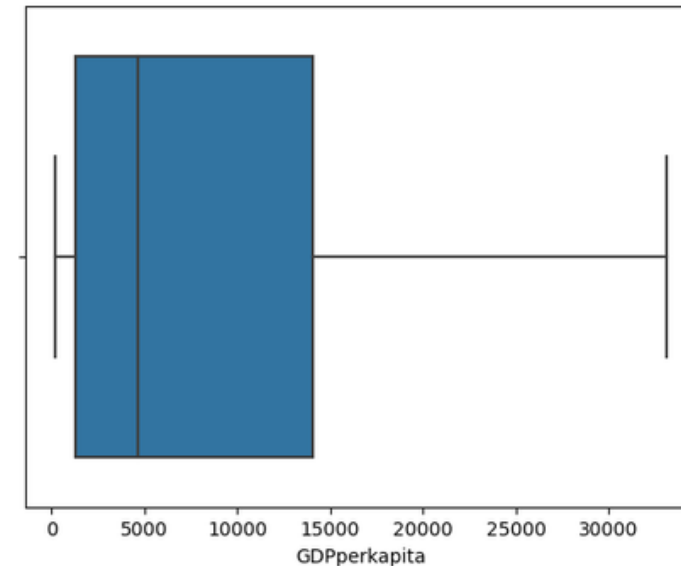
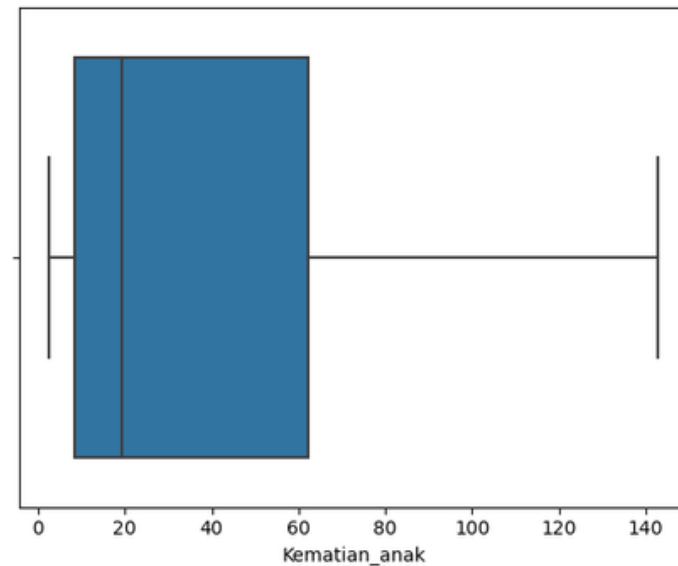
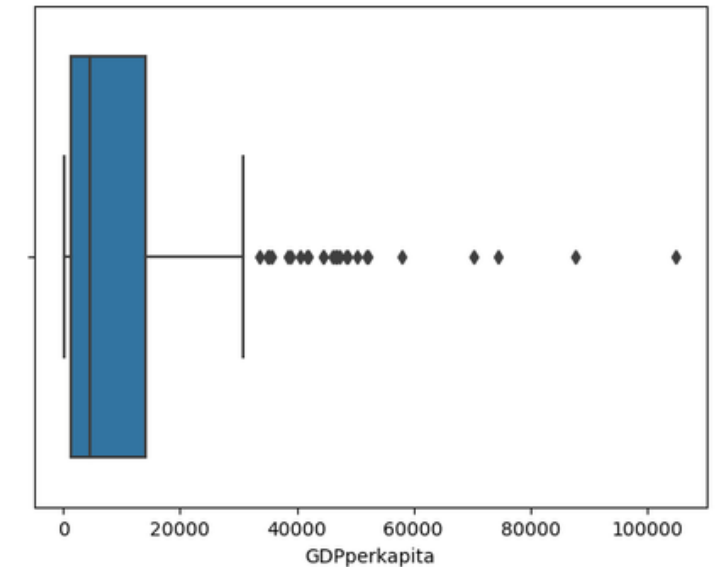
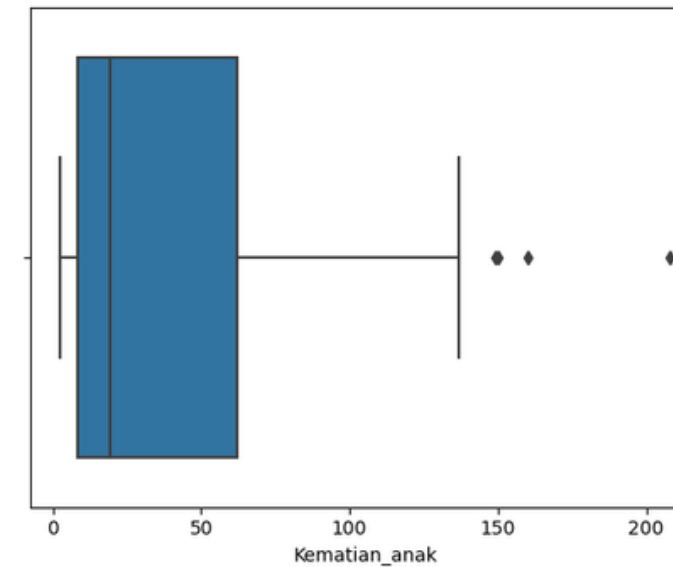
	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000
mean	38.270060	41.108976	6.815689	46.890215	17144.688623	7.781832	70.555689	2.947964	12964.155689
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



Dari kedua faktor tersebut, pada proyek ini akhirnya memutuskan untuk menggunakan features GDPperkapita dan juga Kematian_anak sebagai 2 buah faktor yang nantinya digunakan sebagai acuan dalam menentukan prioritas bantuan terhadap suatu negara. Alasan dipilihnya kedua features tersebut adalah karena kedua features tersebut memiliki korelasi yang tinggi serta seperti yang kita ketahui bahwa nilai GDP pada suatu negara menggambarkan tingkat pembangunan ekonomi pada negara itu sendiri, dan GDP memiliki pengaruh terhadap pertumbuhan serta kesehatan dari sumber daya manusia di negara itu sendiri, karena kualitas kesehatan dari SDM suatu negara akan mempengaruhi terhadap tingkat produktivitas serta pertumbuhan ekonomi negara tersebut. Kualitas Kesehatan dari suatu negara juga dapat dilihat dari tingkat kematian anak di negara tersebut, karena tingkat kematian anak yang tinggi berasal dari tingkat kualitas kesehatan serta fasilitas kesehatan yang kurang memadai.

HANDLING OUTLIERS

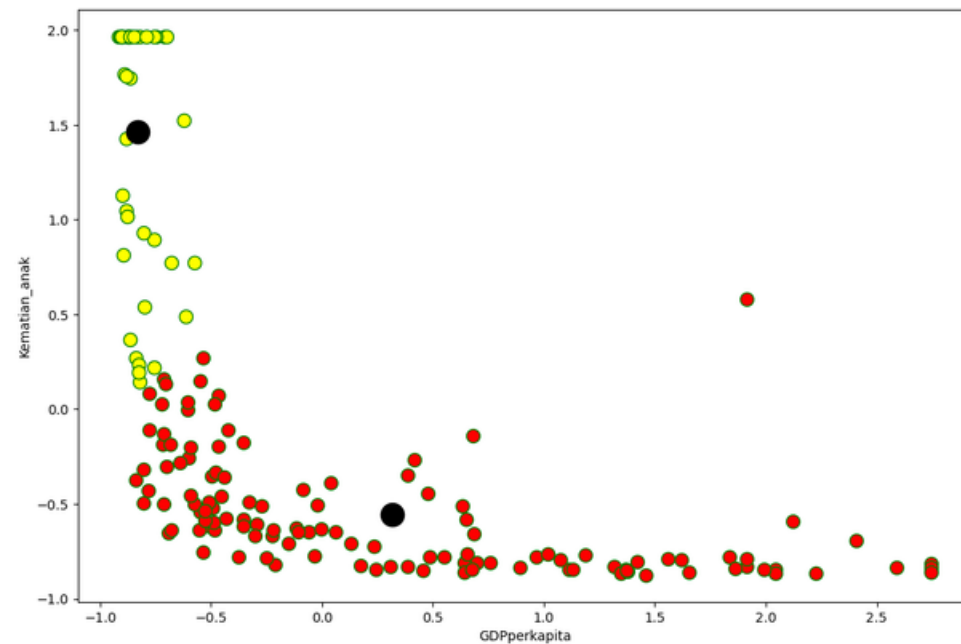
BEFORE



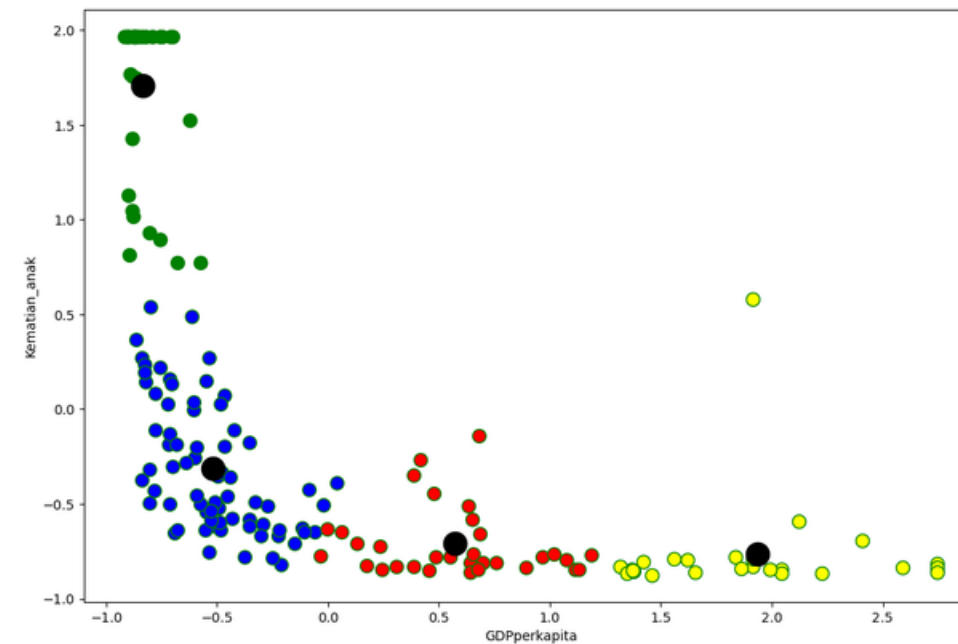
AFTER

Handling Outliers dilakukan dengan mengisi nilai outliers dengan nilai batas atas maupun batas bawah dari data. Karena kasus yang dihadapi adalah membuat rekomendasi negara yang membutuhkan bantuan dengan asumsi data yang dimiliki adalah real maka tidak bisa dilakukan penghapusan data, maka untuk menghindari penghapusan data pada proyek ini dilakukan dengan mengubah nilai nilai yang berada diluar batas dengan nilai batasan itu sendiri.

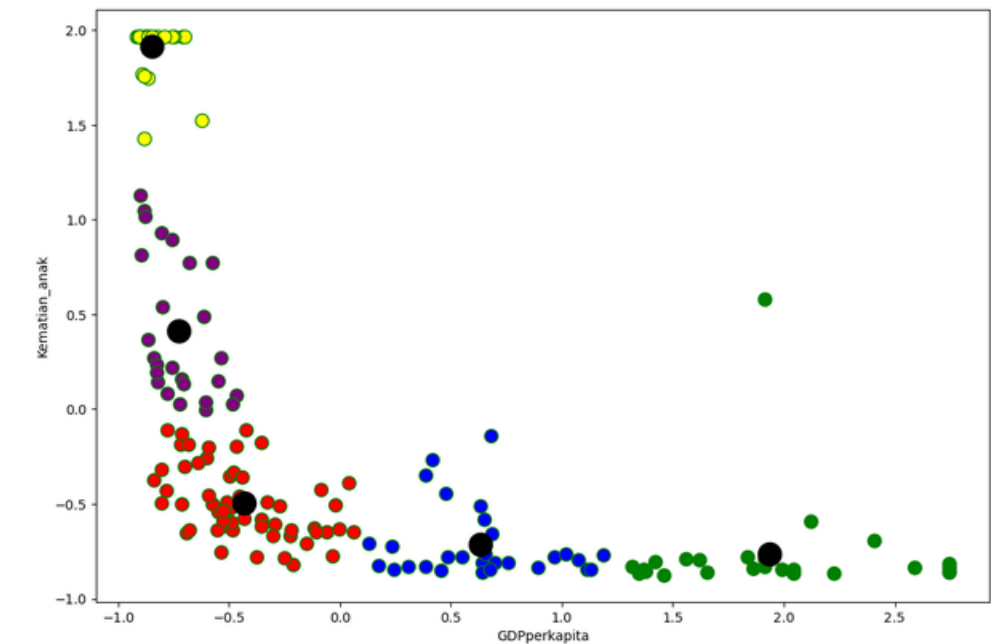
CLUSTERING



SILHOUETTE SCORE N=2 : 0.531169750080884



SILHOUETTE SCORE N=4 : 0.5724557123177565



SILHOUETTE SCORE N=5 : 0.5630058686530804

Dari ketiga model tersebut didapatkan bahwa jumlah cluster dengan score silhouette tertinggi adalah dengan 4 buah cluster.

RESULT

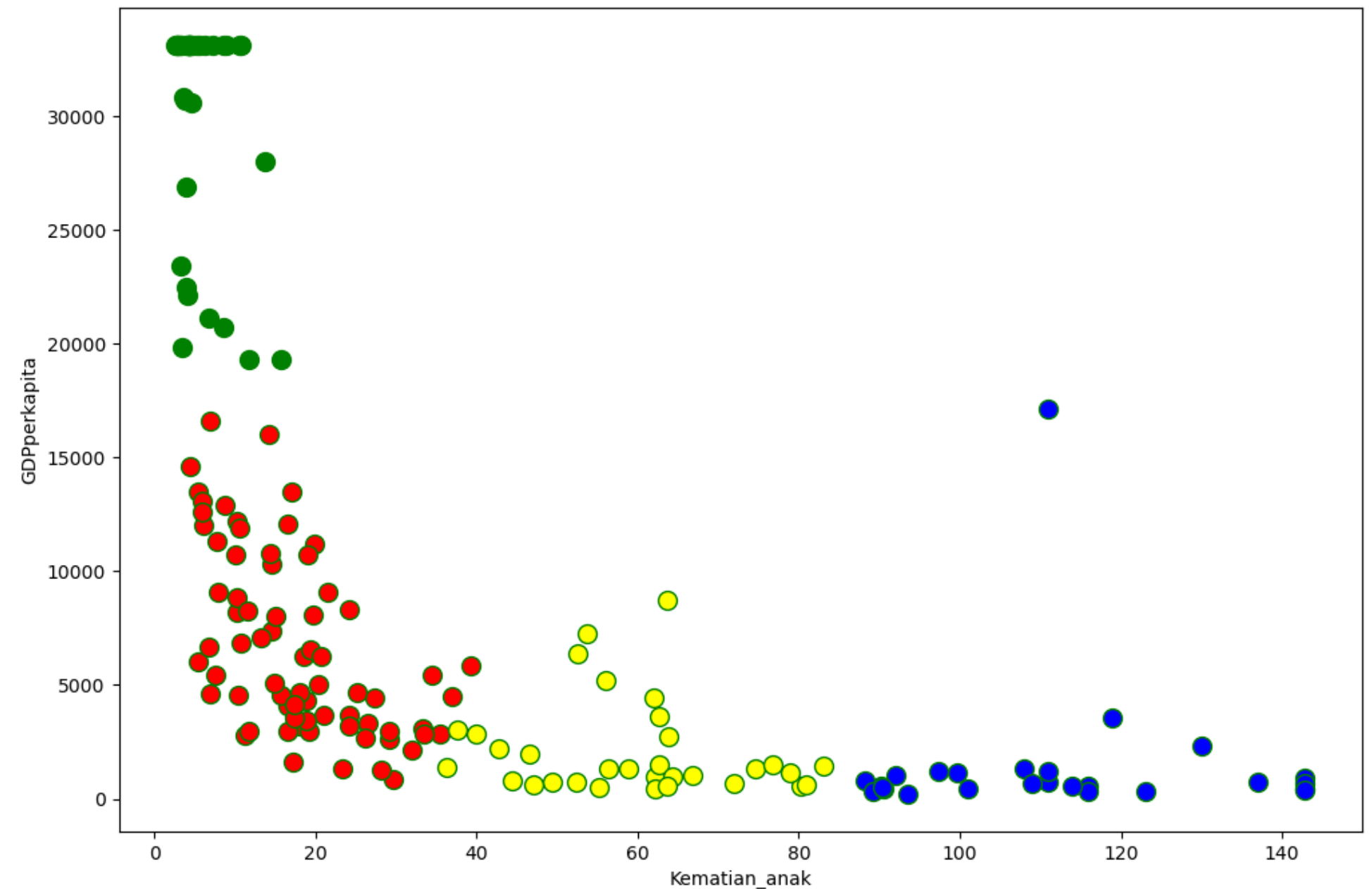
Berdasarkan hasil clustering didapatkan 4 buah cluster yang terdiri dari:

- Red Cluster
- Yellow Cluster
- Green Cluster
- Blue Cluster

Dari ke-4 cluster tersebut didapatkan 'Blue Cluster' sebagai cluster yang lebih di prioritaskan untuk menerima bantuan dari HELP INTERNATIONAL.

Dan karena biaya yang dibutuhkan oleh HELP INTERNATIONAL terbatas maka perlu di prioritaskan kepada negara top 3 yang paling membutuhkan bantuan yaitu:

- 1.Haiti
- 2.Sierra Leone
- 3.Chad



	index	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	Cluster
0	66	Haiti	208.0	15.3	6.91	64.7	1500	5.45	32.1	3.33	662	3
1	132	Sierra Leone	160.0	16.8	13.10	34.5	1220	17.20	55.0	5.20	399	3
2	32	Chad	150.0	36.8	4.53	43.5	1930	6.39	56.5	6.59	897	3
3	31	Central African Republic	149.0	11.8	3.98	26.5	888	2.01	47.5	5.21	446	3
4	97	Mali	137.0	22.8	4.98	35.1	1870	4.37	59.5	6.55	708	3



MUHAMMAD N RAFI

PYTHON – DATA SCIENCE

THANK YOU

BATCH 39 – 2022