

ANALISIS DATA NEGARA

Objective:

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

Permasalahan:

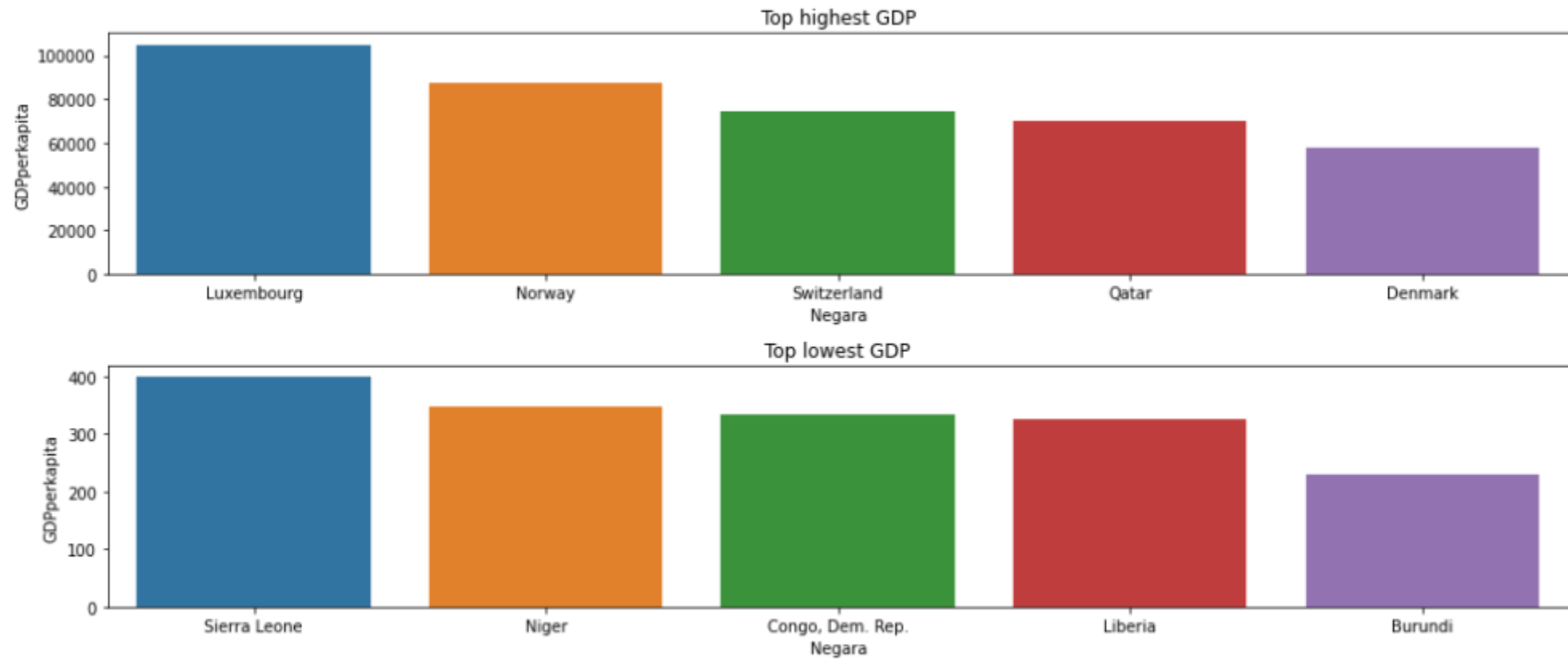
- HELP International telah berhasil mengumpulkan sekitar \$ 10 juta. Saat ini, CEO LSM perlu memutuskan bagaimana menggunakan uang ini secara strategis dan efektif. Jadi, CEO harus mengambil keputusan untuk memilih negara yang paling membutuhkan bantuan. Oleh karena itu, Tugas teman-teman adalah mengkategorikan 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.

HASIL ANALISIS

DATA_NEGARA_HELP.CSV

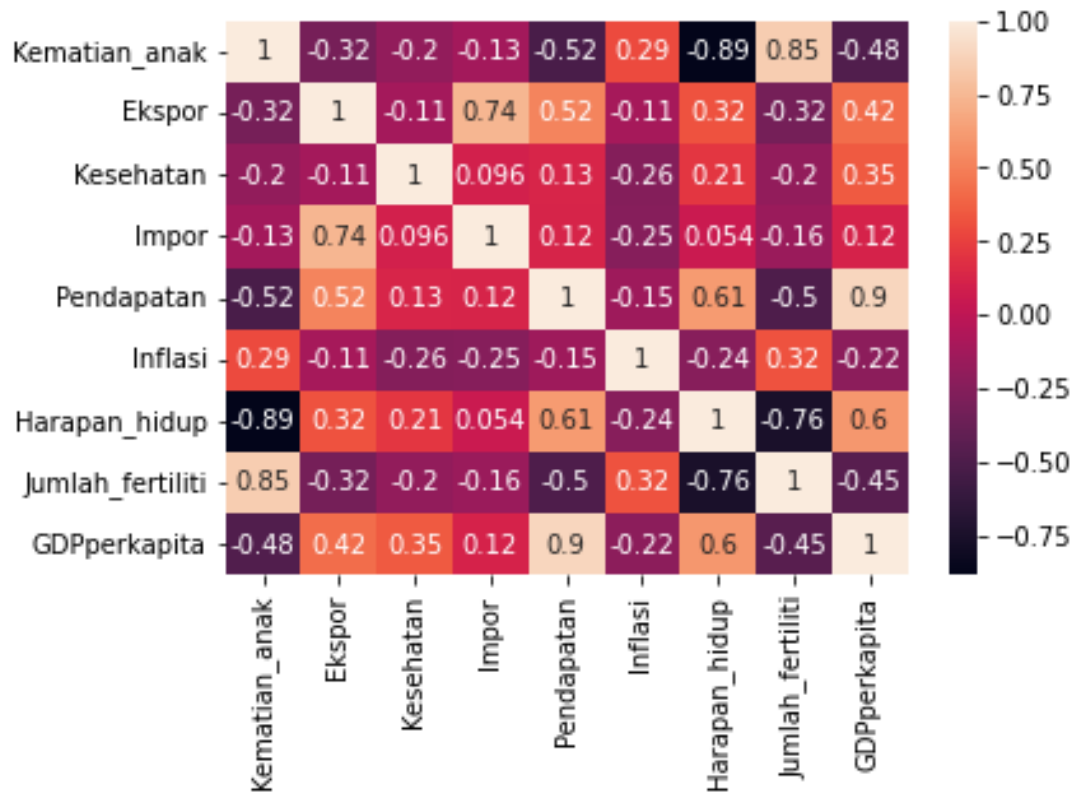
- 167 rows x 10 columns
- Tidak ada Nilai Null

	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 Grafik diatas dapat disimpulkan bahwa 5 teratas negara dengan GDP terbesar yakni Luxembourg, Norway, Switzerland, Qatar, dan Denmark
- Sedangkan 5 Negara terbawah dengan GDP terendah yakni Sierra Leone, Niger, Congo, Liberia, dan Burundi

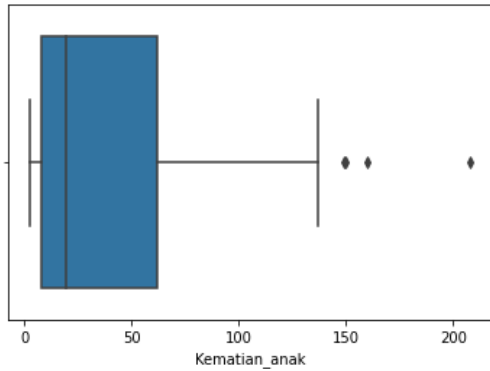
ANALISIS HEATMAP



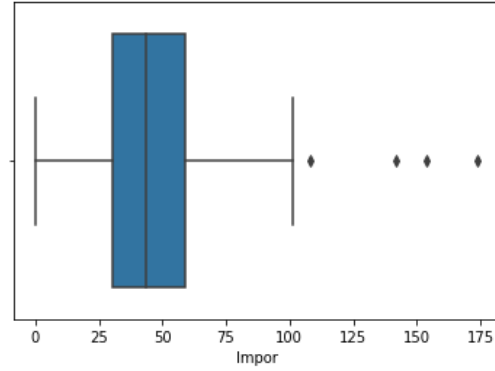
- Heatmap disamping menunjukkan bahwa semakin gelap warna heatmap maka semakin besar hubungan antar 2 variable tersebut.
- Hubungan Variabel terkuat
 - Kematian anak & Harapan Hidup
 - Jumlah Fertility & Harapan Hidup
 - GDP perkapita & Kematian Anak

OUTLIERS ATAU DATA PENCILAN DI SETIAP COLUMNS

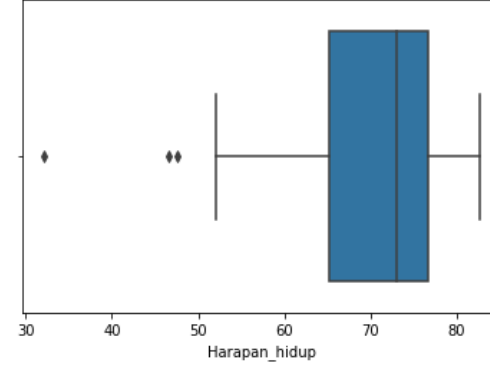
Kematian_anak



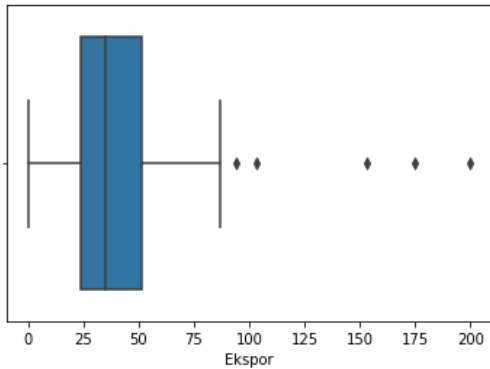
Impor



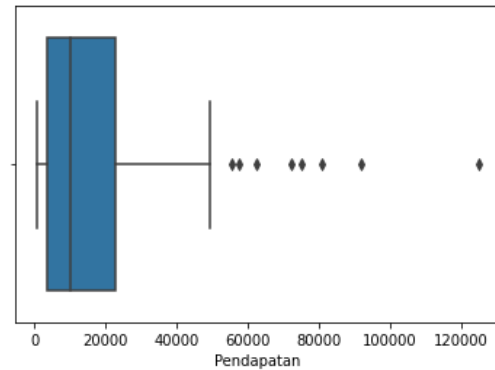
Harapan_hidup



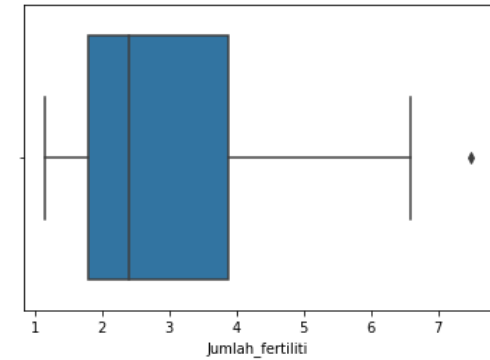
Ekspor



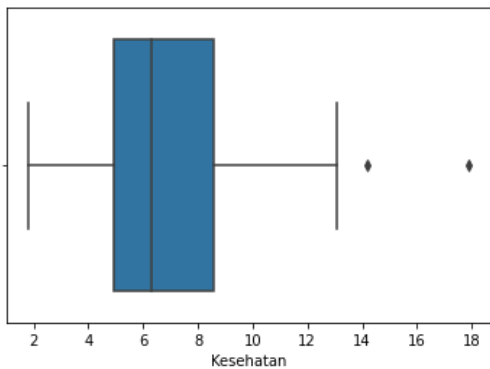
Pendapatan



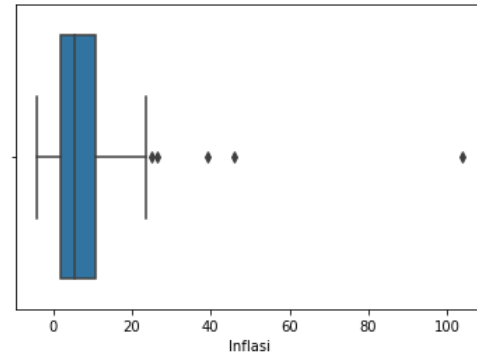
Jumlah_fertiliti



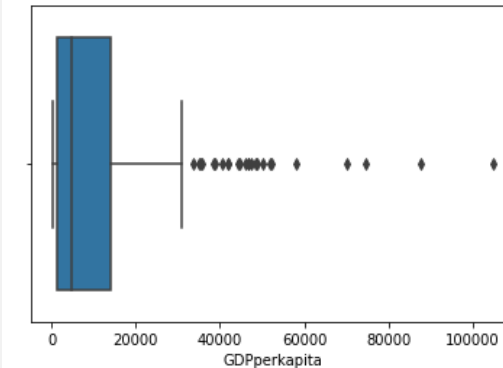
Kesehatan



Inflasi



GDPperkapita



Pada setiap Kolom terdapat Outliers atau data pencilan. Sebelum melakukan analisis kita harus melakukan data cleansing terlebih dahulu untuk menghilangkan outliers

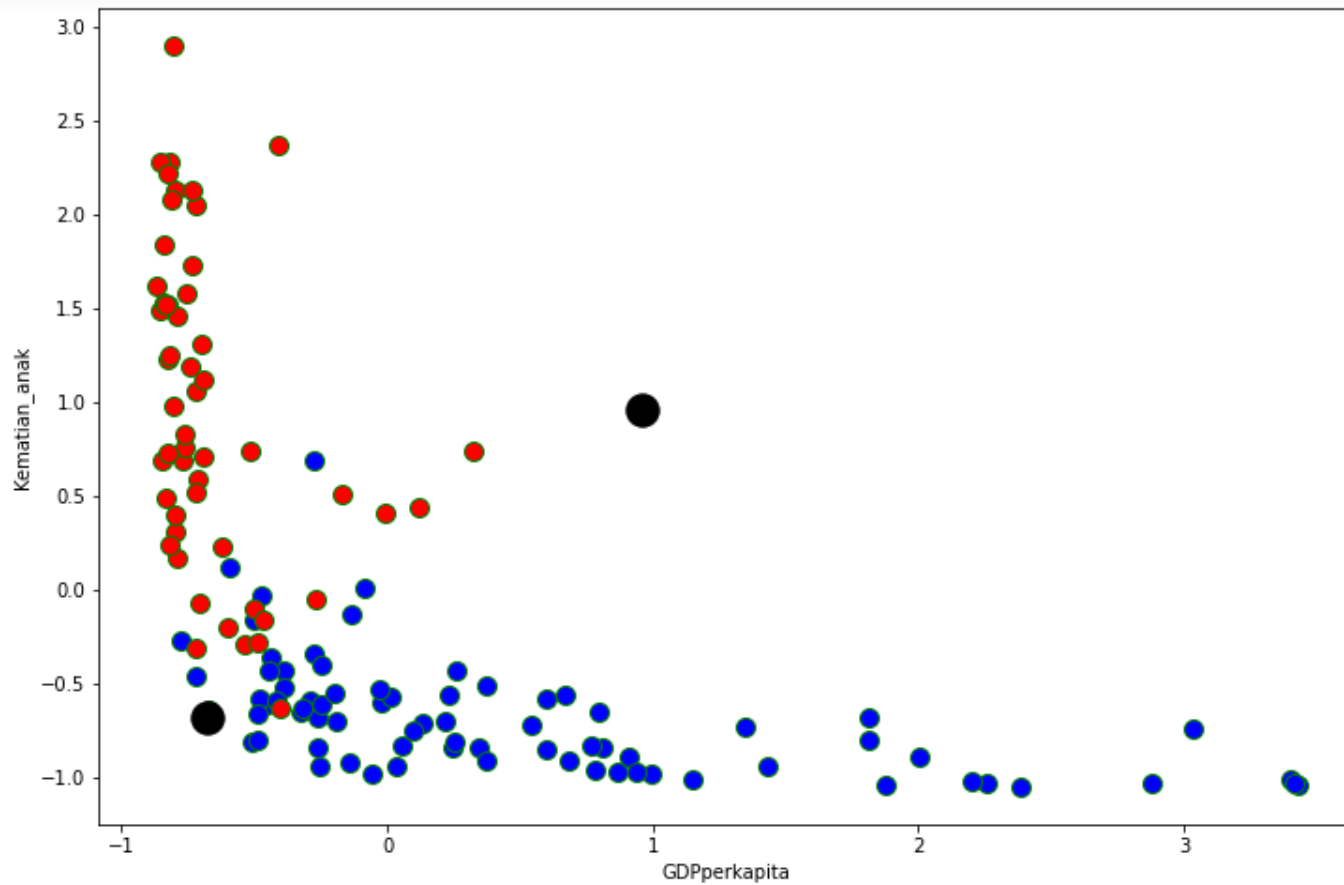
REMOVE OUTLIERS

	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
0	90.2	10.0	7.58	44.9	1610.0	9.44	56.2	5.82	553.0
1	16.6	28.0	6.55	48.6	9930.0	4.49	76.3	1.65	4090.0
2	27.3	38.4	4.17	31.4	12900.0	16.10	76.5	2.89	4460.0
3	119.0	62.3	2.85	42.9	5900.0	22.40	60.1	6.16	3530.0
4	10.3	45.5	6.03	58.9	19100.0	1.44	76.8	2.13	12200.0
...
161	36.3	31.7	5.81	28.5	4240.0	16.50	68.8	2.34	1380.0
162	29.2	46.6	5.25	52.7	2950.0	2.62	63.0	3.50	2970.0
164	23.3	72.0	6.84	80.2	4490.0	12.10	73.1	1.95	1310.0
165	56.3	30.0	5.18	34.4	4480.0	23.60	67.5	4.67	1310.0
166	83.1	37.0	5.89	30.9	3280.0	14.00	52.0	5.40	1460.0

128 rows × 9 columns

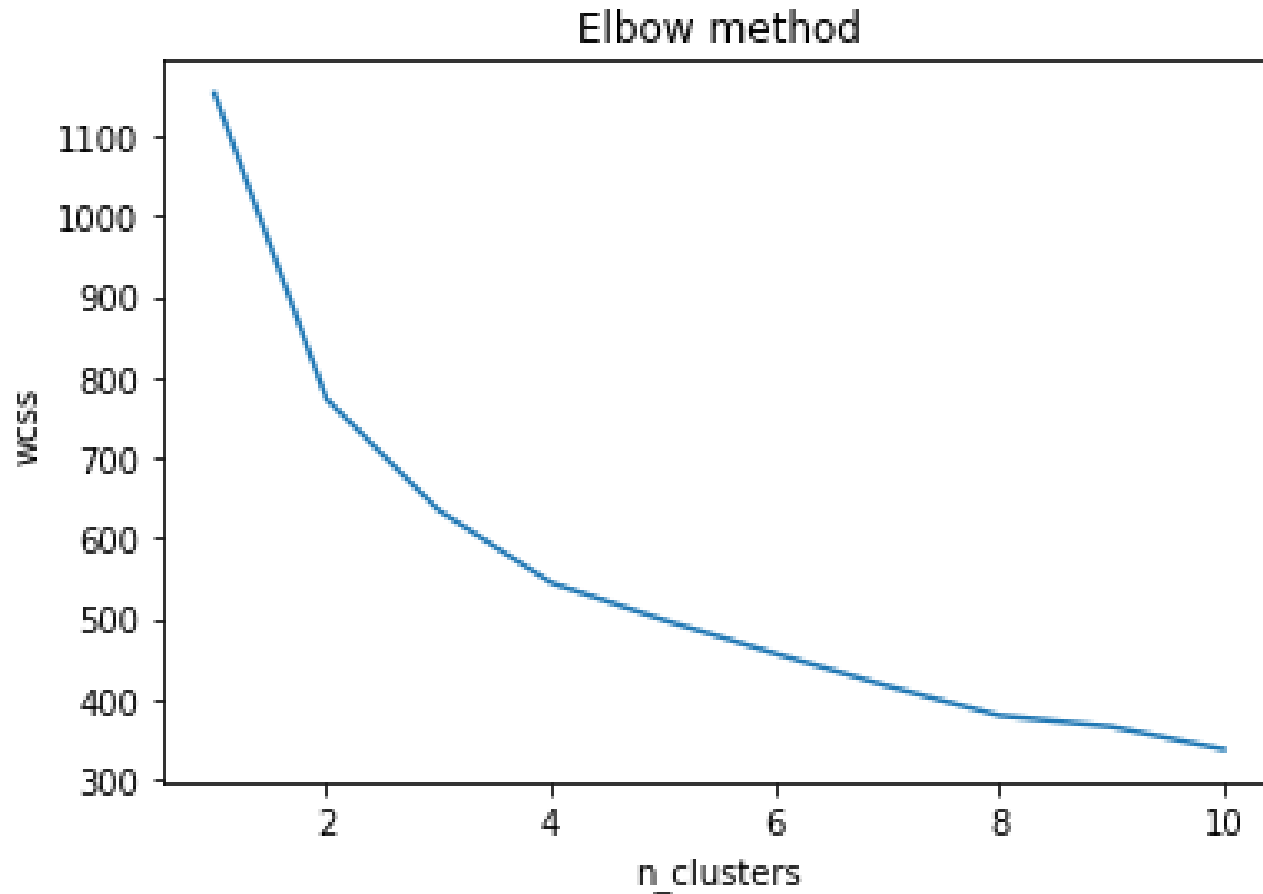
Setelah menghilangkan outliers, Data menjadi 128 rows yang sebelumnya 167 rows.

K-MEAN CLUSTERING GDP DAN KEMATIAN_ANAK



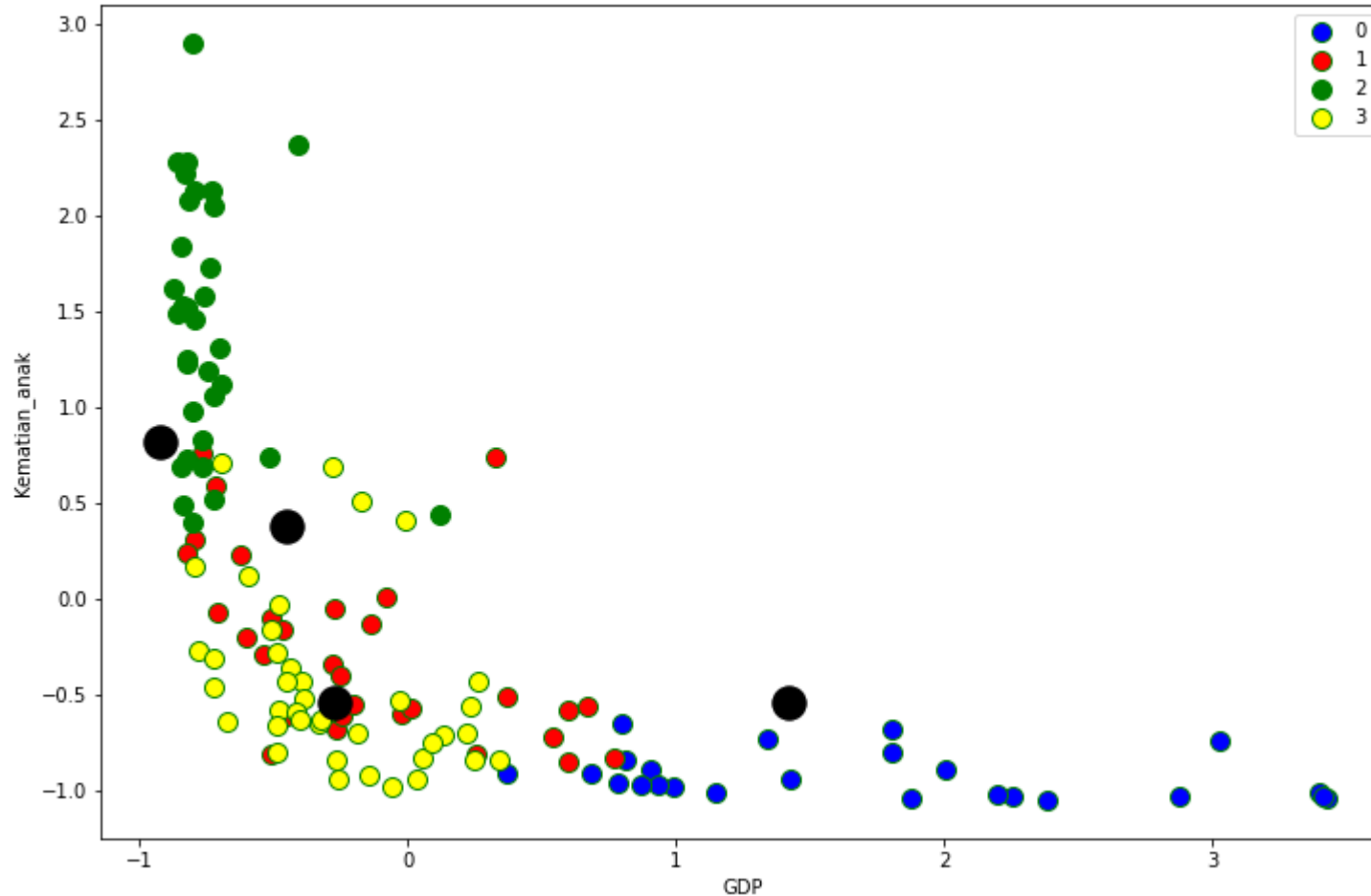
- Dari grafik disamping dapat disimpulkan bahwa
 - Biru : jika GDP besar maka angka kematian anak akan tinggi juga
 - Merah : jika GDP kecil maka angka kematian anak juga kecil

ELBOW METHOD



- Jika menggunakan elbow method dapat dilihat bahwa cluster yang direkomendasikan adalah 4 atau menggunakan cluster sebanyak 4 agar data yang dihasilkan lebih akurat

4 CLUSTER DENGAN VAR GDP DAN KEMATIAN_ANAK



- Dapat dilihat disamping bahwa masing-masing cluster mendapatkan titik centroid yang berbeda-beda dengan penjelasan :
 - Biru : semakin besar GDP suatu negara maka angka kematian anak akan semakin kecil sekitar 0.5 point
 - Merah : jika GDP suatu negara kurang dari 1 point maka angka kematiannya akan sedikit lebih kecil dari cluster biru
 - Kuning : jika GDP lebih sedikit dari GDP cluster merah maka tingkat kematian anaknya akan tinggi melebihi 0
 - Hijau : jika GDP suatu negara sedikit maka dapat dipastikan angka kematian anak akan tinggi

SILHOUETTE_SCORE

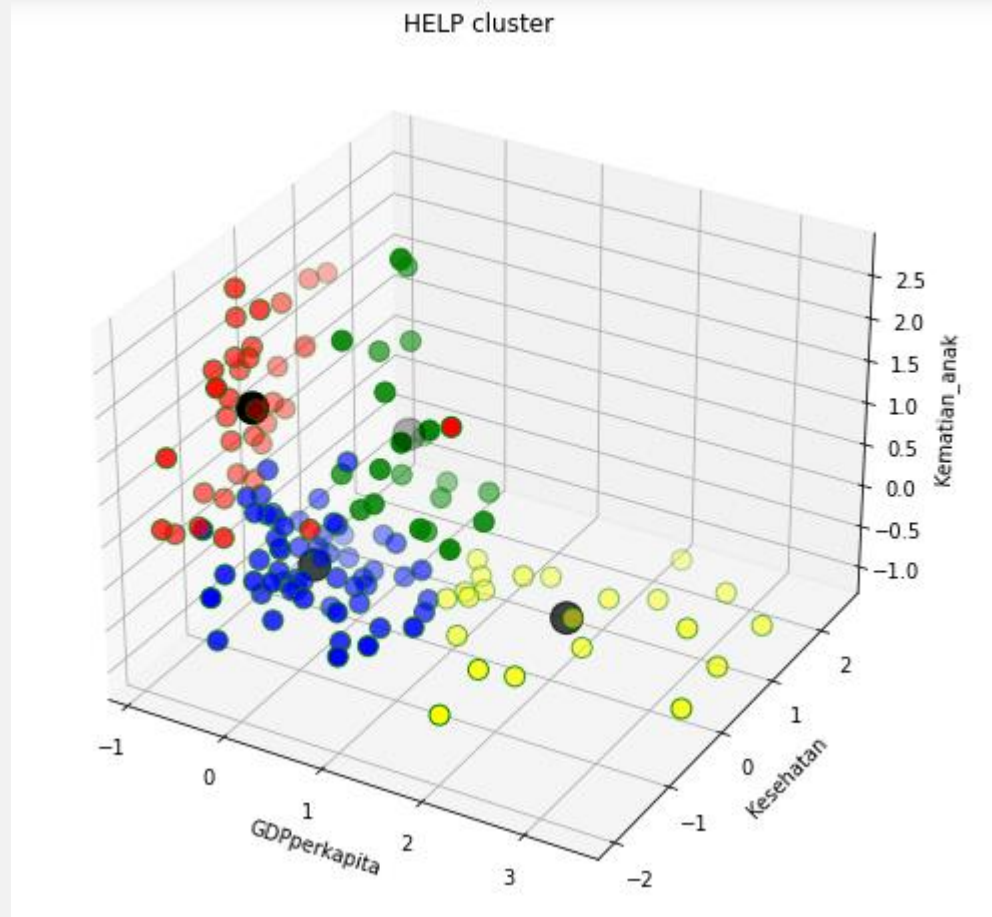
```
print(silhouette_score(df_std, labels= labels1))  
print(silhouette_score(df_std, labels= labels2))
```

```
0.28074249512402516
```

```
0.2208489669981804
```

- Silhouette score disamping menandakan bahwa grafik pertama memiliki jarak antar cluster yang jauh, dan grafik kedua memiliki jarak antar cluster yang dekat

HELP CLUSTERING DENGAN 4 CLUSTER



- Dari grafik disamping dapat diambil data dari setiap cluster yakni :
 - Cluster Kuning : semakin besar GDP suatu negara dan total anggaran Kesehatan suatu negara maka tingkat kematian anak sangat rendah.
 - Cluster biru : sama seperti cluster kuning namun GDP dan total anggaran Kesehatan negara tersebut lebih rendah tapi angka kematian anak juga rendah.
 - Cluster Hijau : GDP negara rendah dan total anggaran Kesehatan negara tersebut sangat besar, namun tingkat kematian anak juga lebih banyak dibanding dengan cluster biru dan kuning.
 - Cluster Merah : GDP negara sangat rendah, total anggaran Kesehatan juga rendah, maka Kematian anak sangat tinggi dibanding cluster lain.

KESIMPULAN

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	K_means_labels
26	Burundi	93.6	8.920	11.60	39.2000	764	12.30	57.7	6.26	231	0
93	Madagascar	62.2	25.000	3.77	43.0000	1390	8.79	60.8	4.60	413	0
106	Mozambique	101.0	31.500	5.21	46.2000	918	7.64	54.5	5.56	419	0
94	Malawi	90.5	22.800	6.59	34.9000	1030	12.10	53.1	5.31	459	0
36	Comoros	88.2	16.500	4.51	51.7000	1410	3.87	65.9	4.75	769	0
107	Myanmar	64.4	0.109	1.97	0.0659	3720	7.04	66.8	2.41	988	0
87	Lesotho	99.7	39.400	11.10	101.0000	2380	4.15	46.5	3.30	1170	0
99	Mauritania	97.4	50.700	4.41	61.2000	3320	18.90	68.2	4.98	1200	0
59	Ghana	74.7	29.500	5.22	45.9000	3060	16.60	62.2	4.27	1310	0
102	Moldova	17.2	39.200	11.70	78.5000	3910	11.10	69.7	1.27	1630	0

- Negara yang sangat perlu dibantu oleh HELP yakni Burundi karena memiliki GDP yang paling rendah dibandingkan dengan 167 negara lain, ditambah tingkat pendapatan yang rendah, angka kematian anak dan angka inflasi yang tinggi menyebabkan negara ini menjadi negara perlu dibantu.