

Sanbercode

Final Project – BDS 28

Oleh Muhammad Nur Arifin

Reading Data & Understanding

	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

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.

Objective

Mengategorikan 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,

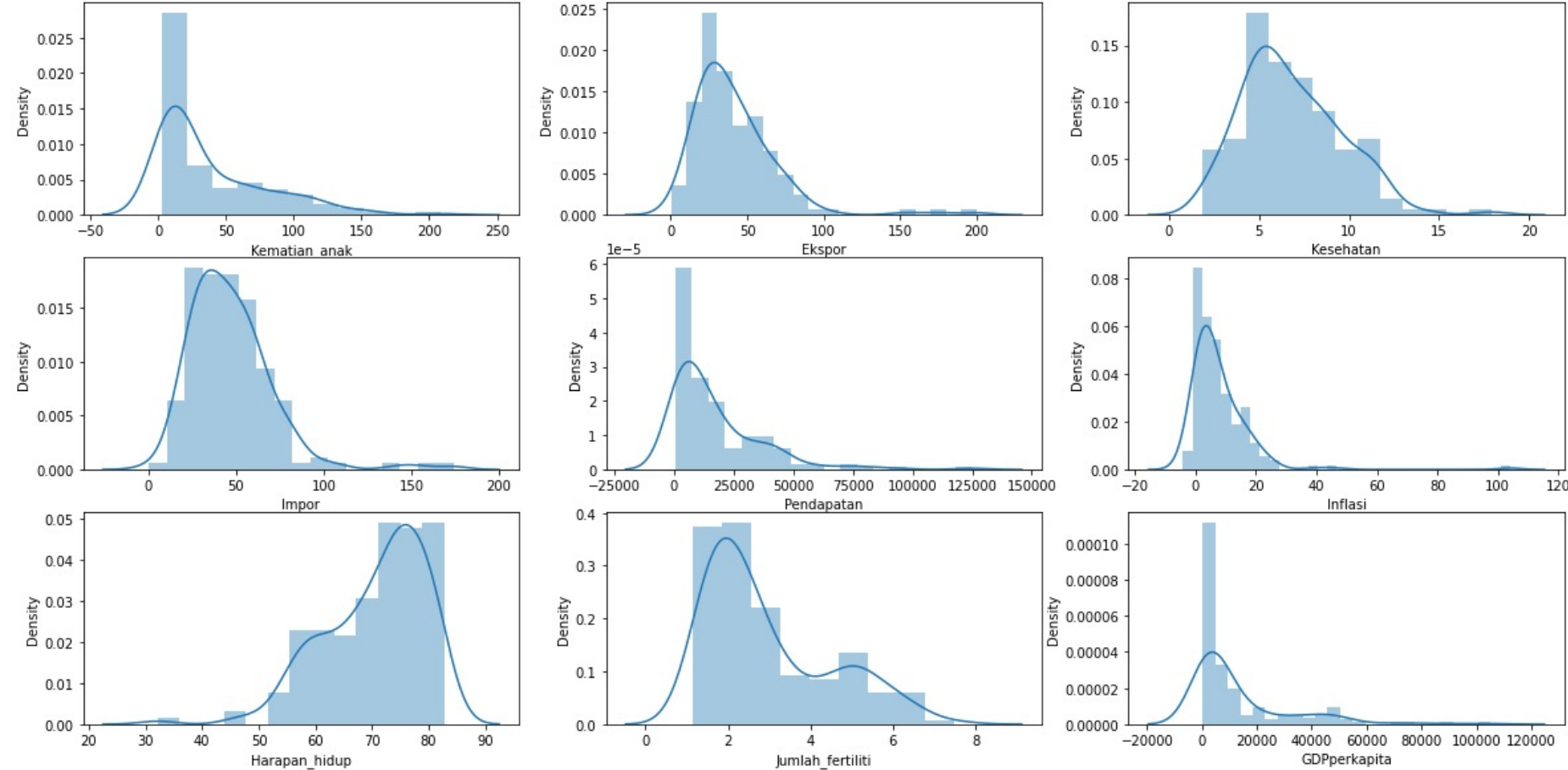
Tugas teman-teman adalah mengategorikan negara menggunakan beberapa faktor sosial ekonomi dan kesehatan yang menentukan perkembangan negara secara keseluruhan. Kemudian kalian perlu menyarankan negara mana saja yang paling perlu menjadi fokus CEO.

Goals :
5 Negara Sebagai Penerima Bantuan

(Berdasarkan faktor Ekonomi-Sosial dan Kesehatan)

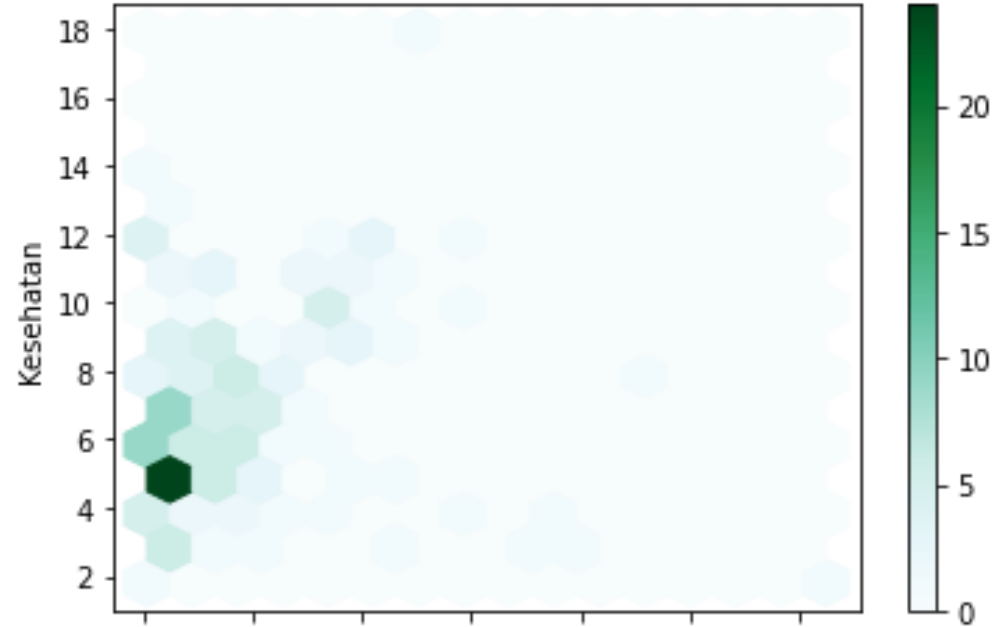
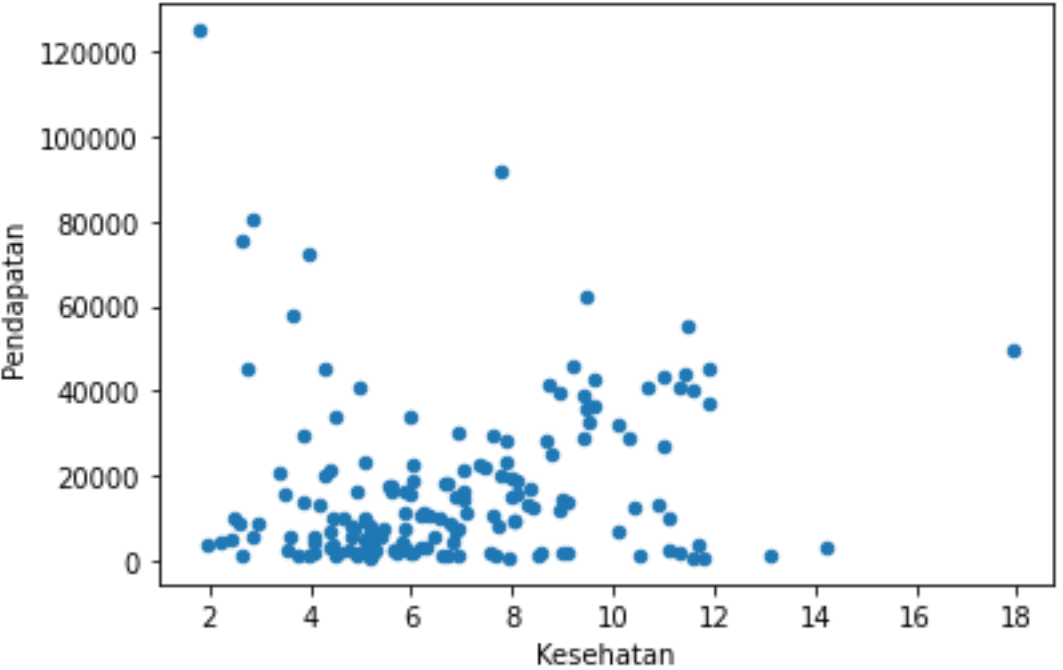
Exploratory Data Analysis

A. Univariate



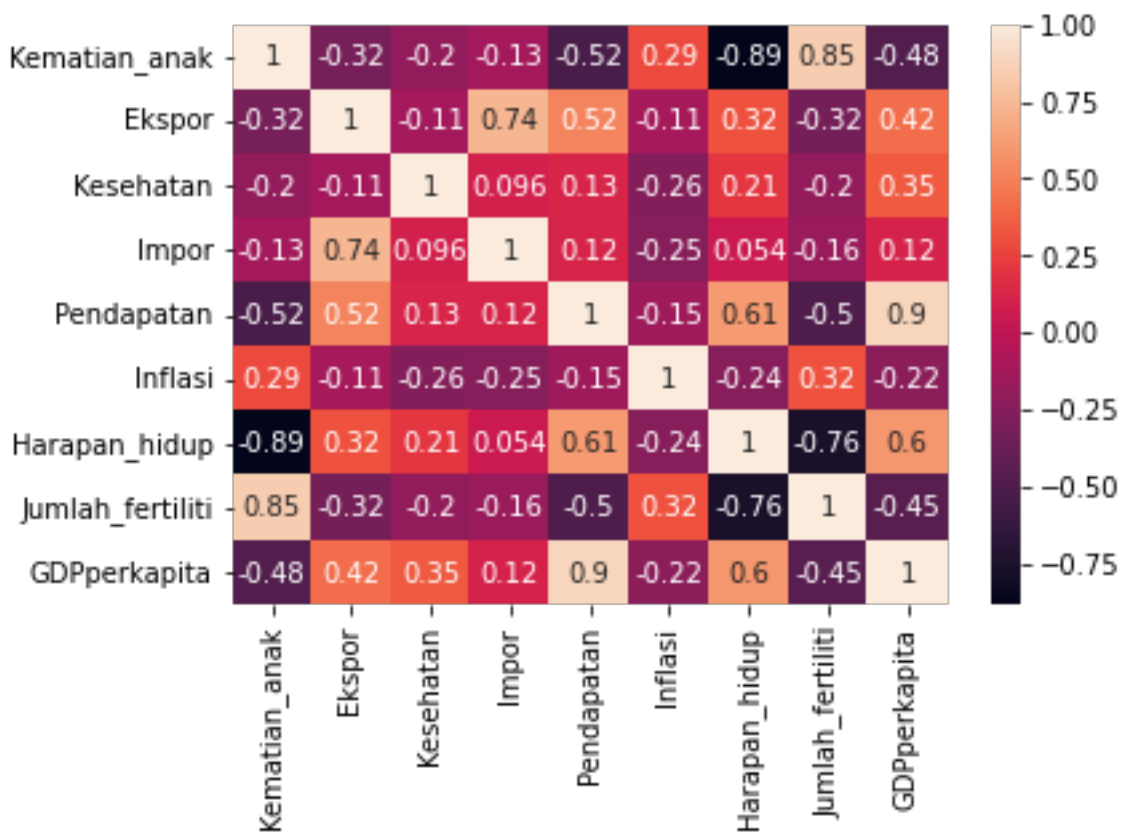
Exploratory Data Analysis

B. Bivariate



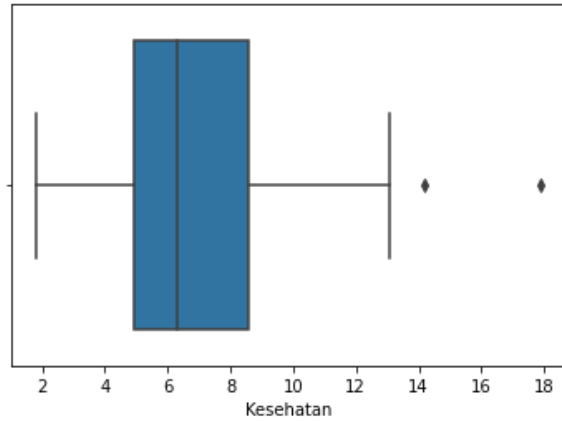
Exploratory Data Analysis

C. Multivariate

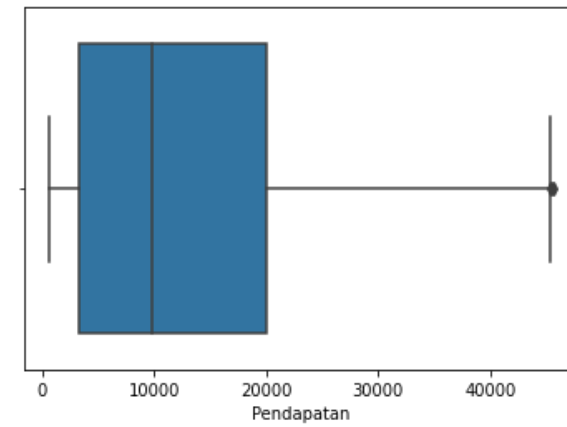
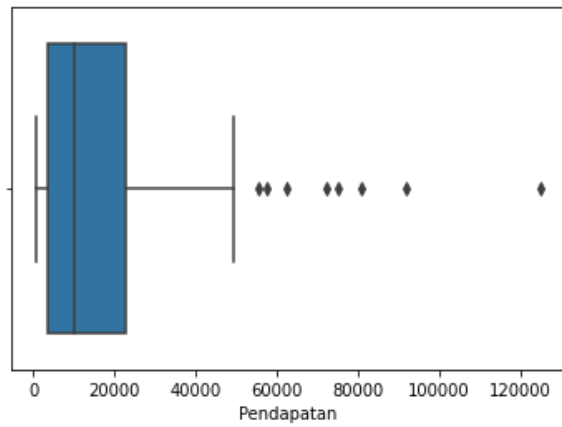
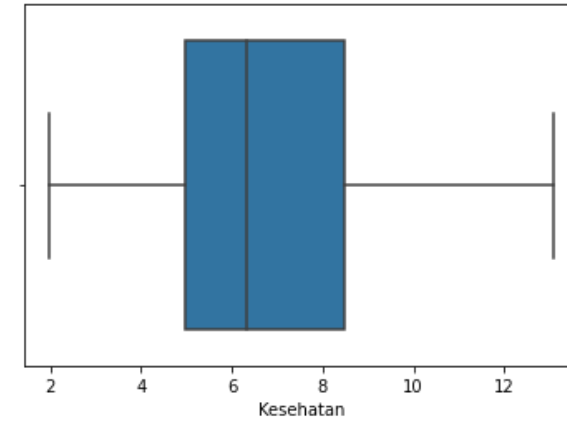


Outliers Treatment

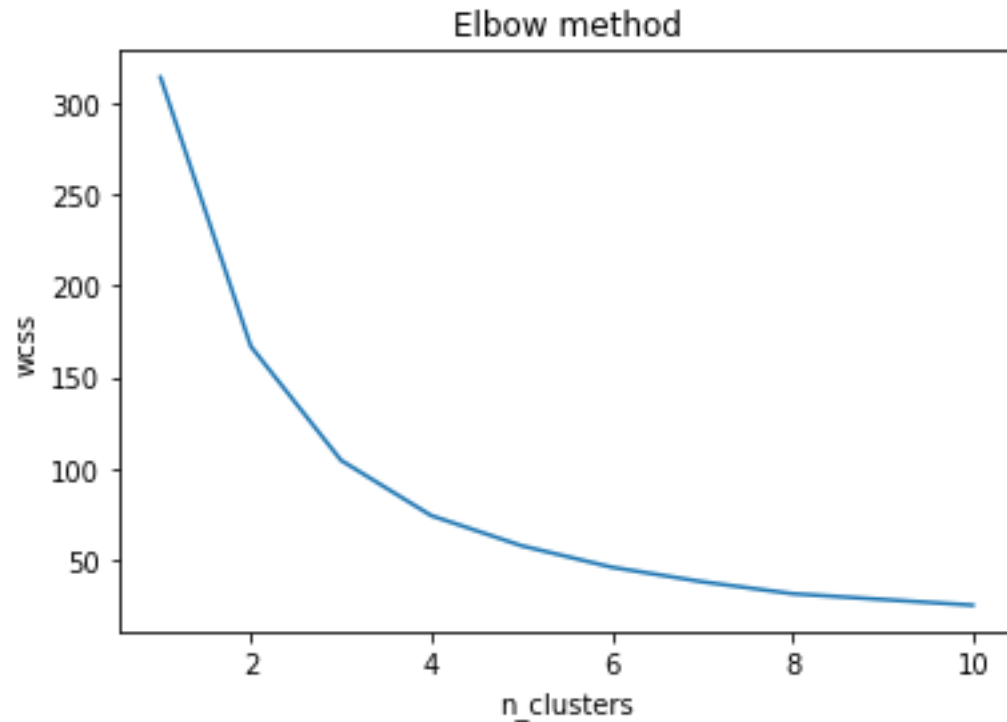
Mendeteksi Outlier



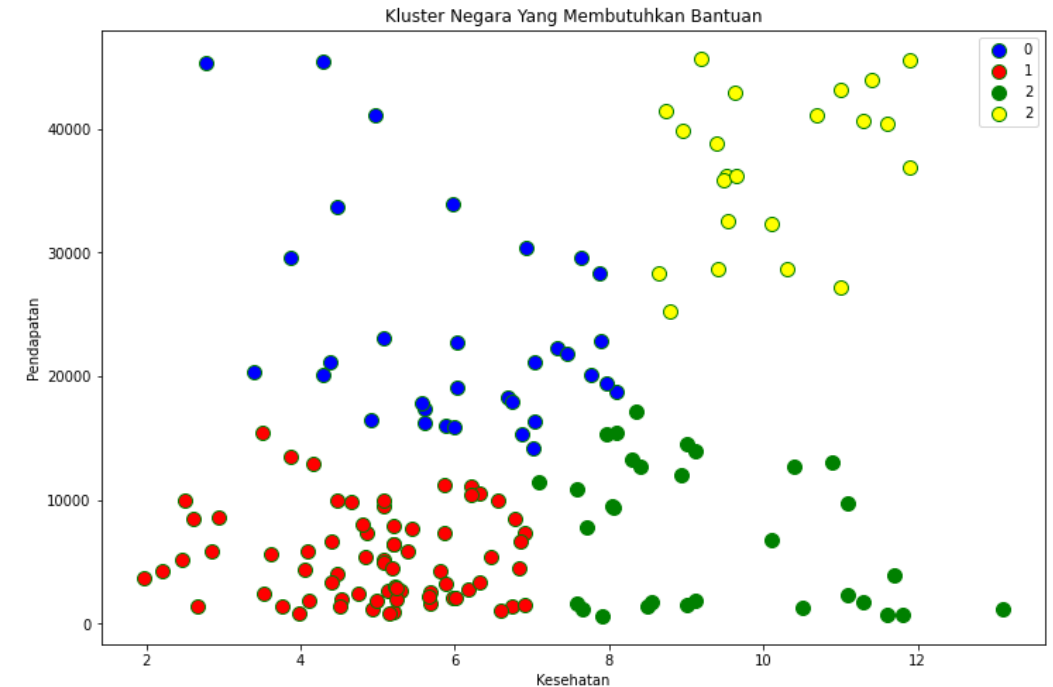
Menghapus Outlier



K-Means Clustering & Visualizing Clusters Formed

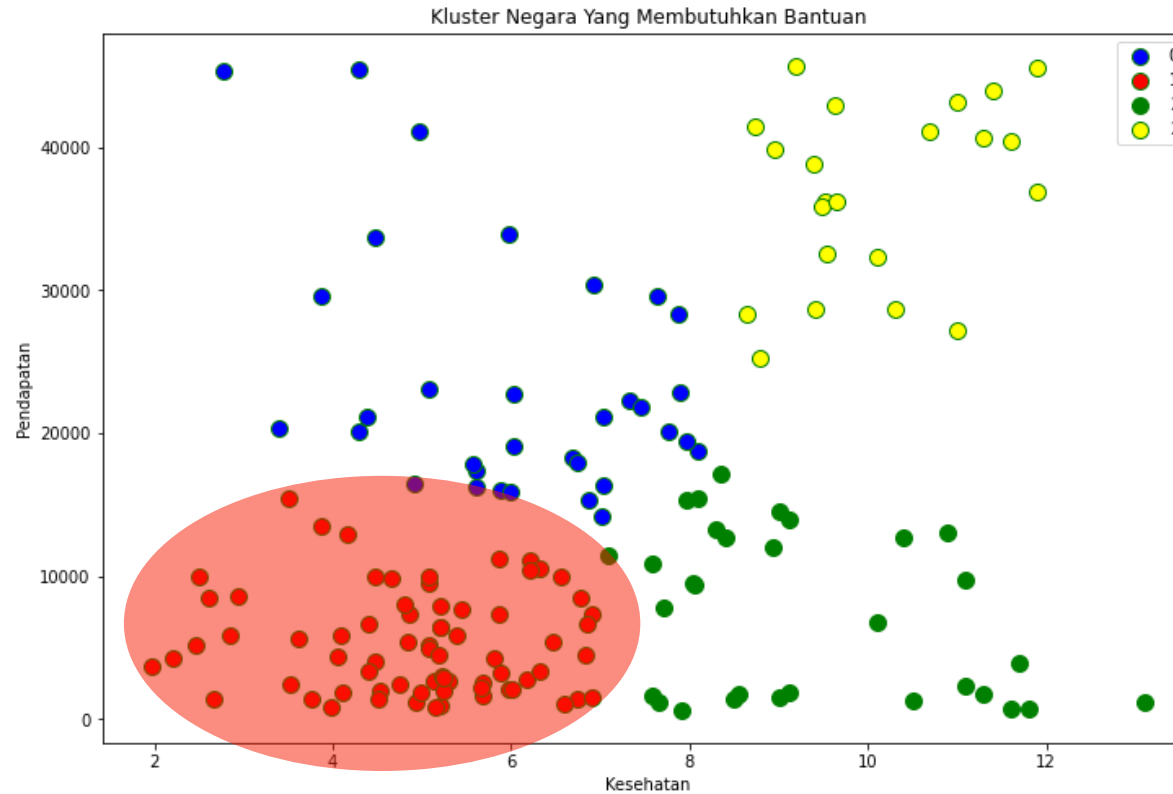


Setelah dilakukan tahap **scaling data dengan metode standarisasi**, kemudian dilanjutkan dengan **Elbow Method** untuk mendapatkan jumlah cluster yang tepat. Dalam kasus ini jumlah yang dianjurkan adalah **4 cluster**



Dari clustering diatas negara-negara yang ada pada dataset telah ter-clustering menjadi 4 cluster

K-Means Clustering & Visualizing Clusters Formed



Berdasarkan permasalahan yang ingin di-solving dan objektif yang ingin dicapai maka pada hasil akhir clustering, Cluster merah dengan label 1 merupakan Negara-Negara yang dicari. Dalam visualisasi diatas cluster tersebut adalah negara dengan Tingkat Kesehatan dan Tingkat Pendapatan terendah

Report Countries



	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	label1_kmeans	label2_kmeans
176	NaN	NaN	NaN	1.97	NaN	3720.0	NaN	NaN	NaN	NaN	0.0	1
116	Pakistan	92.1	13.50	2.20	19.4	4280.0	10.90	65.3	3.85	1040.0	0.0	1
38	Congo, Rep.	63.9	85.10	2.46	54.7	5190.0	20.70	60.4	4.95	2740.0	0.0	1
154	Turkmenistan	62.0	76.30	2.50	44.5	9940.0	2.31	67.9	2.83	4440.0	0.0	1
172	NaN	NaN	NaN	2.61	NaN	8430.0	NaN	NaN	NaN	NaN	0.0	1
50	Eritrea	55.2	4.79	2.66	23.3	1420.0	11.60	61.7	4.61	482.0	0.0	1
167	NaN	NaN	NaN	2.85	NaN	5900.0	NaN	NaN	NaN	NaN	0.0	1
140	Sri Lanka	11.2	19.60	2.94	26.8	8560.0	22.80	74.4	2.20	2810.0	0.0	1
55	Gabon	63.7	57.70	3.50	18.9	15400.0	16.60	62.9	4.08	8750.0	0.0	1
169	NaN	NaN	NaN	3.52	NaN	2440.0	NaN	NaN	NaN	NaN	0.0	1

Iniilah daftar negara cluster 1 berlabel merah setelah dilakukan sort values berdasarkan 'Kesehatan'.

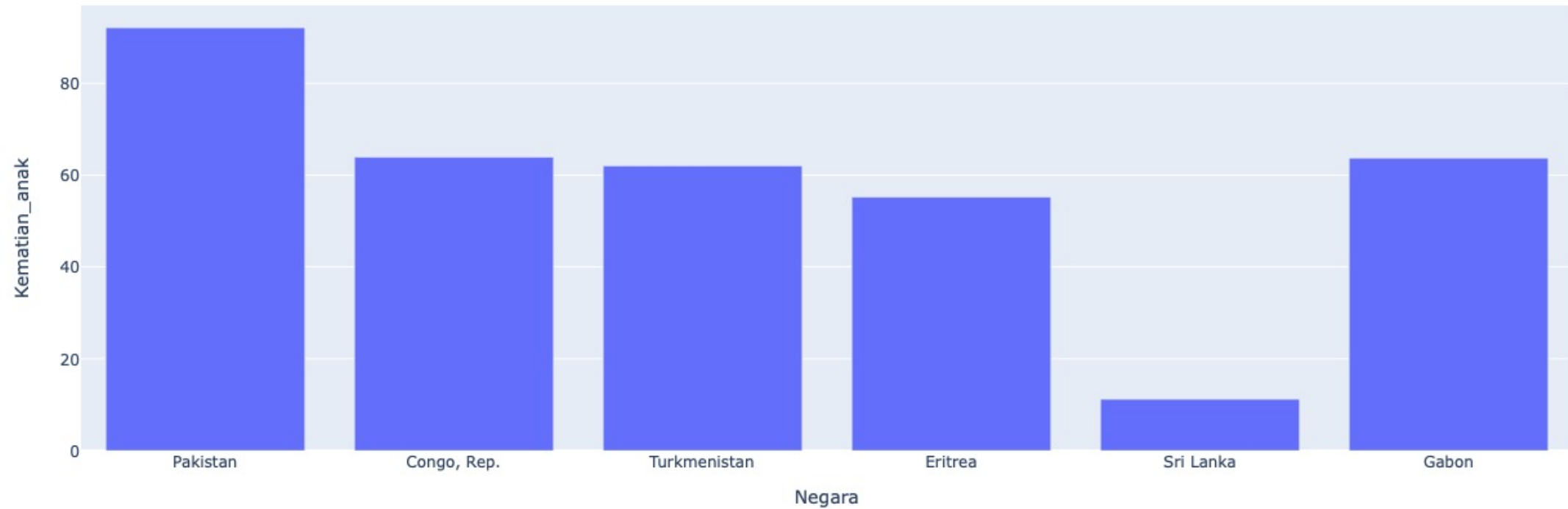
Terlihat diatas negara seperti ***Pakistan, Congo Rep., Turkmenistan, Eritrea, Sri Lanka dan Gabon.***

Negara-Negara yang hasil proses diatas selanjutnya akan dilihat lagi data dari variable lainnya yang mewakili sektor Sosial-ekonomi dan Kesehatan selain variable Pendapatan & Kesehatan tetapi tetap mengacu cluster 1

Report Countries

Variabel lain yang merepresentasikan faktor Kesehatan adalah Tingkat Kematian Anak
Berikut visualisasinya:

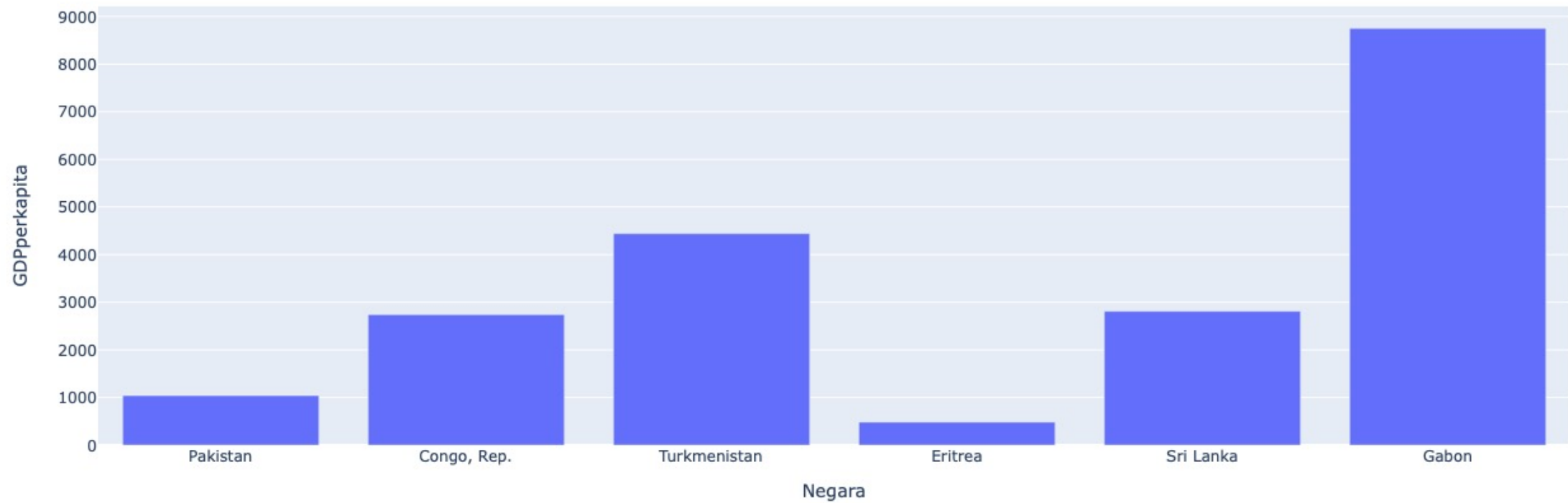
Urutan Negara Yang Kesehatannya Rendah Berdasarkan Tingkat Kematian



Report Countries

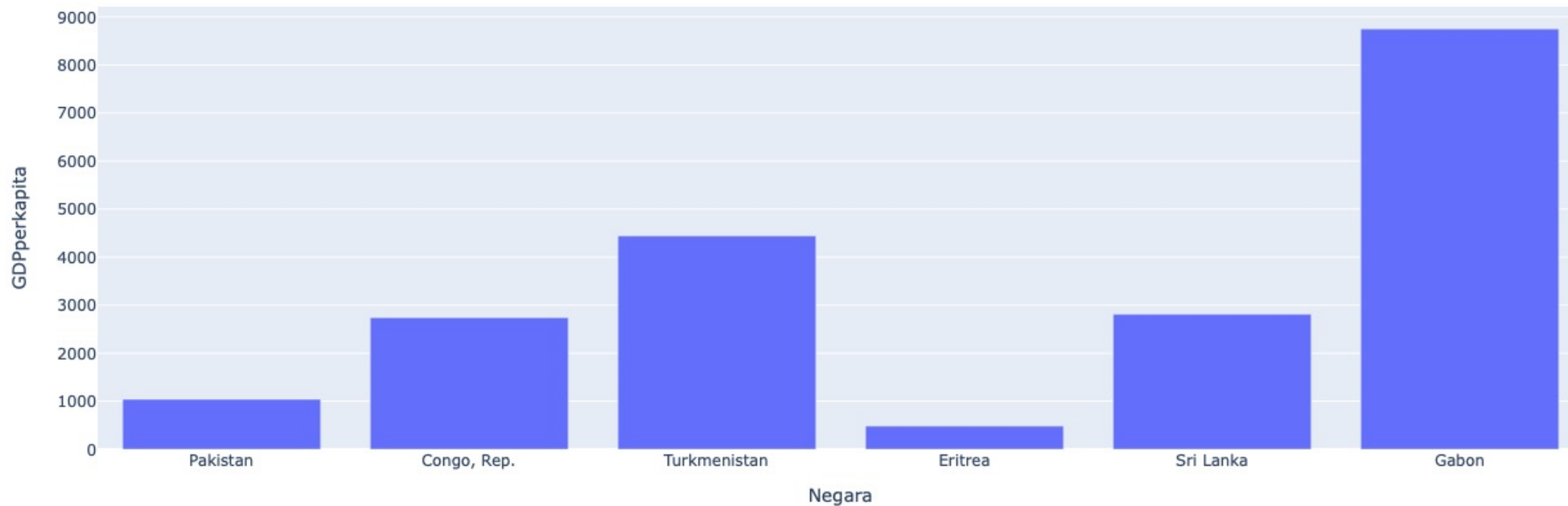
Variabel lain yang merepresentasikan faktor Sosial-Ekonomi adalah GDP Perkapita
Berikut visualisasinya:

Urutan Negara Yang Kesehatannya Rendah Berdasarkan GDP Perkapita



Negara Yang Membutuhkan Bantuan (Final Result)

Urutan Negara Yang Kesehatannya Rendah Berdasarkan GDP Perkapita



Secara urutan:

1. Eritrea 2. Pakistan 3. Congo Republic 4. Sri Lanka 5. Turkmenistan