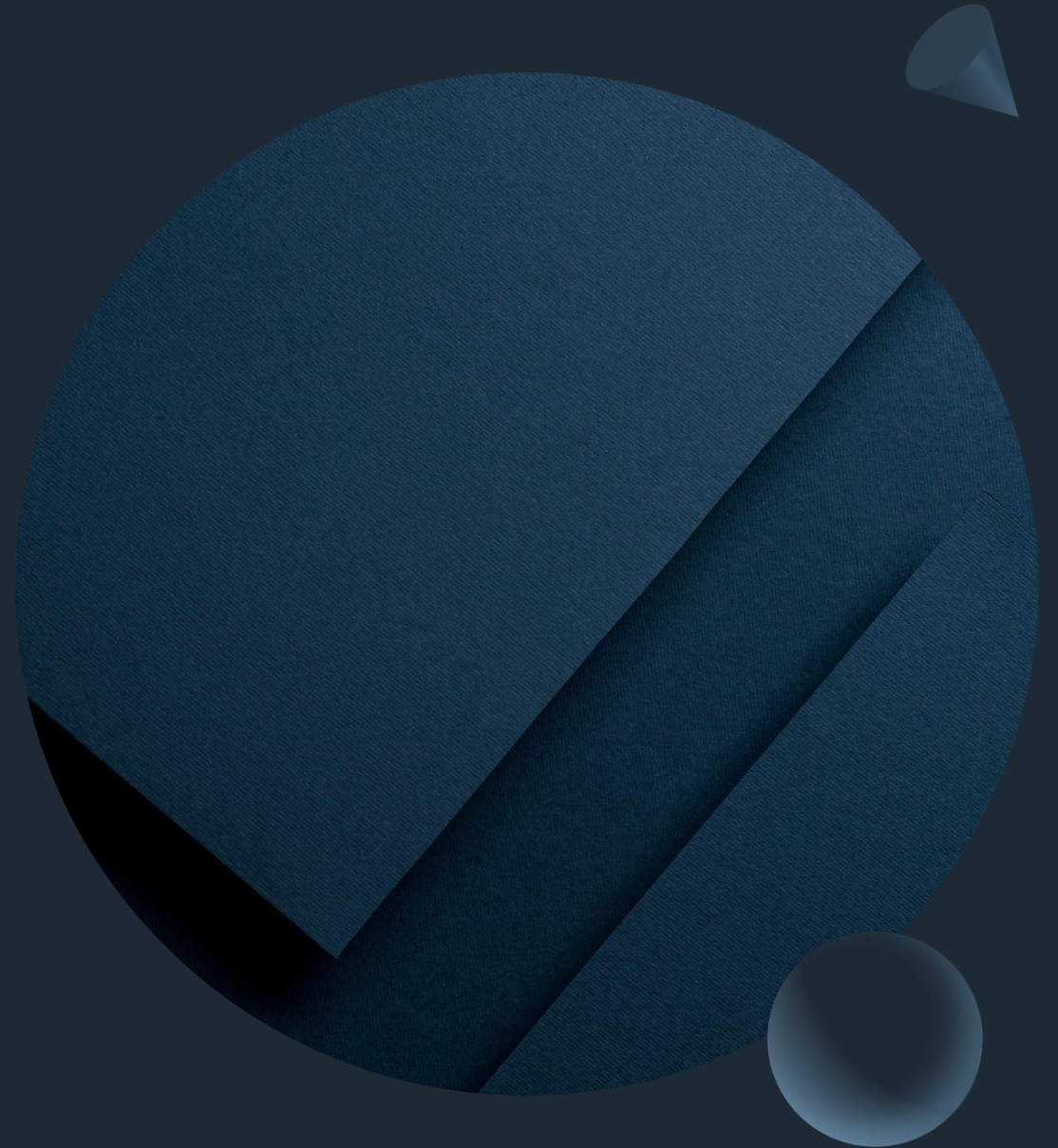


Final Project for HELP International

By Benidictus Galih Mahar Putra



Tentang HELP International

- 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.





Latar Belakang

- HELP International telah berhasil mengumpulkan sekitar \$ 10 juta. CEO harus mengambil keputusan untuk memilih negara yang paling membutuhkan bantuan. Diberikan data negara dari faktor sosial ekonomi dan kesehatan yang menentukan perkembangan negara secara keseluruhan. Kemudian dari data perlu merekomendasikan negara mana yang paling perlu menjadi fokus CEO.

KOLOM	PENJELASAN
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


Data HELP Negara

- Terdiri dari 167 baris dan 10 kolom
- Setiap kolom bertipe float kecuali kolom Negara dengan tipe object dan Inflasi serta GDPperkapita dengan tipe integer



Statistik Deskriptif

	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



- Summary dengan count atau banyak data, mean atau rata-rata, std atau standar deviasi, min atau data minimal, 25% atau data kuartil pertama, 50% atau kuartil kedua, 75% atau kuartil ketiga, dan max atau data maksimal

Missing Value

Dari gambar, tidak ditemukan nilai NULL atau NaN karena banyak data setiap kolom = banyak data awal (167)

Data columns (total 10 columns):

#	Column	Non-Null Count	Dtype
0	Negara	167 non-null	object
1	Kematian_anak	167 non-null	float64
2	Ekspor	167 non-null	float64
3	Kesehatan	167 non-null	float64
4	Impor	167 non-null	float64
5	Pendapatan	167 non-null	int64
6	Inflasi	167 non-null	float64
7	Harapan_hidup	167 non-null	float64
8	Jumlah_fertiliti	167 non-null	float64
9	GDPperkapita	167 non-null	int64

dtypes: float64(7), int64(2), object(1)

Univariate Analysis (1)

- **GDP per kapita:**

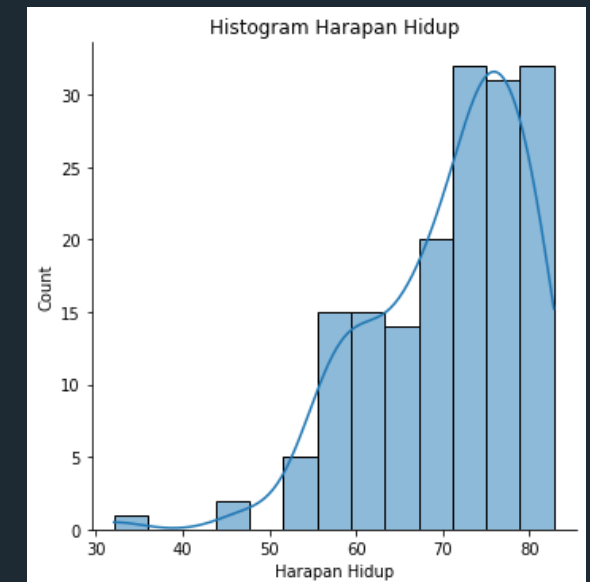
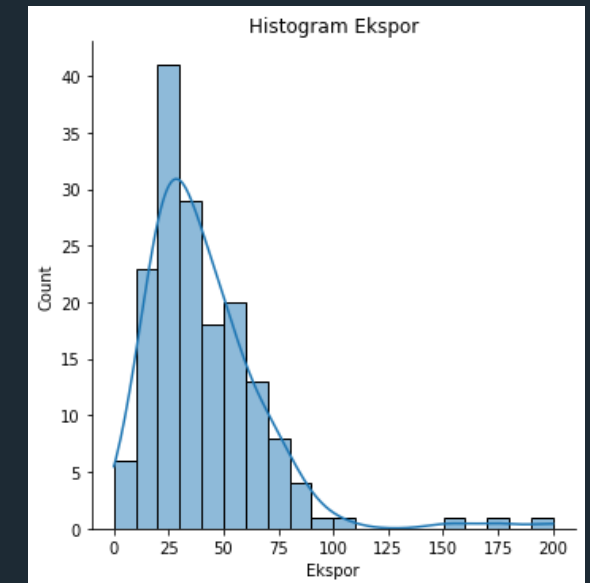
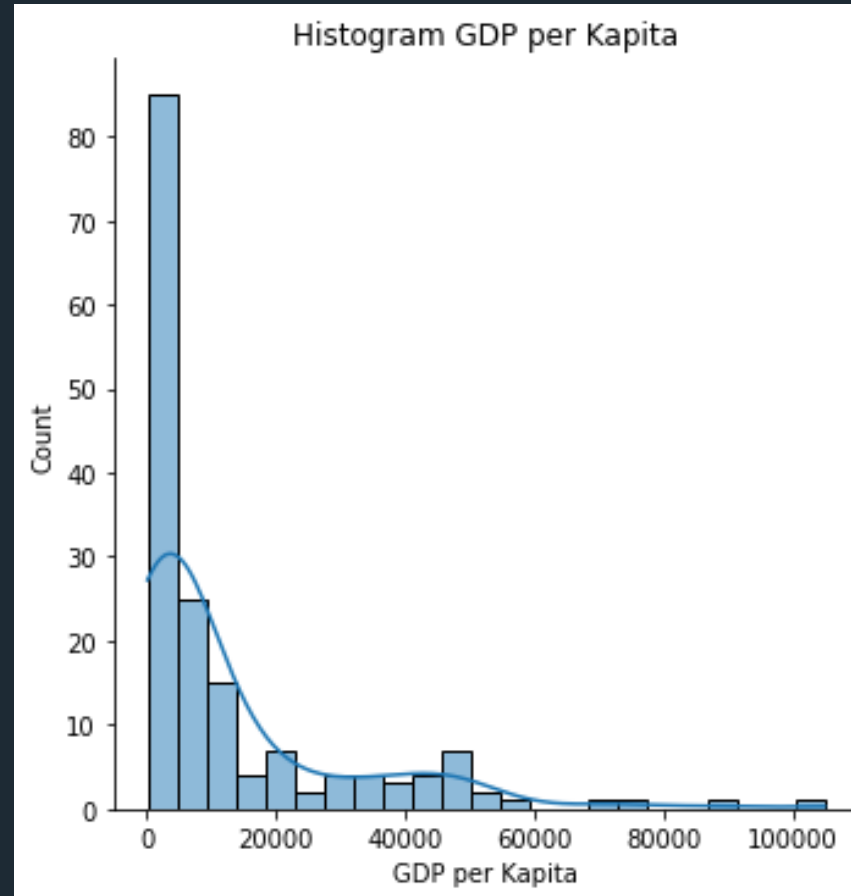
1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 0-5000
3. Jumlah tertinggi rentang 0-5000 lebih dari 80
4. Jumlah terendah rentang 60000-120000 mendekati 0

- **Ekspor:**

1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 12-75
3. Jumlah tertinggi pada rentang 25 sebanyak 40
4. Jumlah terendah pada rentang 100-200 sebanyak mendekati 0

- **Harapan Hidup:**

1. Persebaran skew ke kiri
2. Persebaran sebagian besar tersebar pada rentang 70-85
3. Jumlah tertinggi pada rentang 70-85 lebih dari 30
4. Jumlah terendah tpada rentang 30-50 sebanyak mendekati 0



Univariate Analysis (2)

- **Inflasi:**

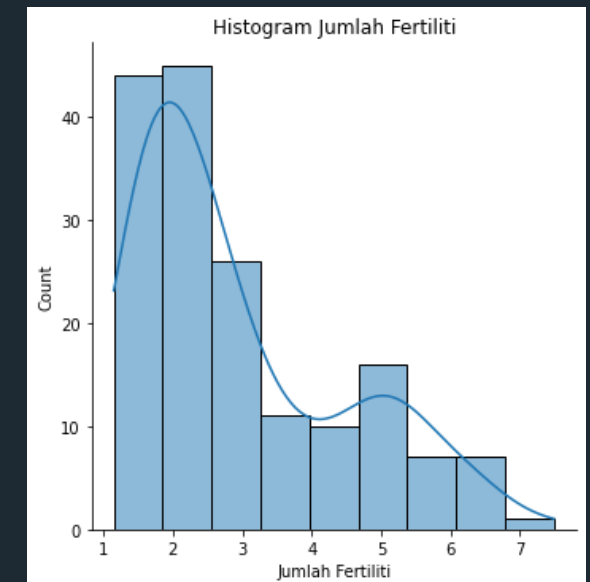
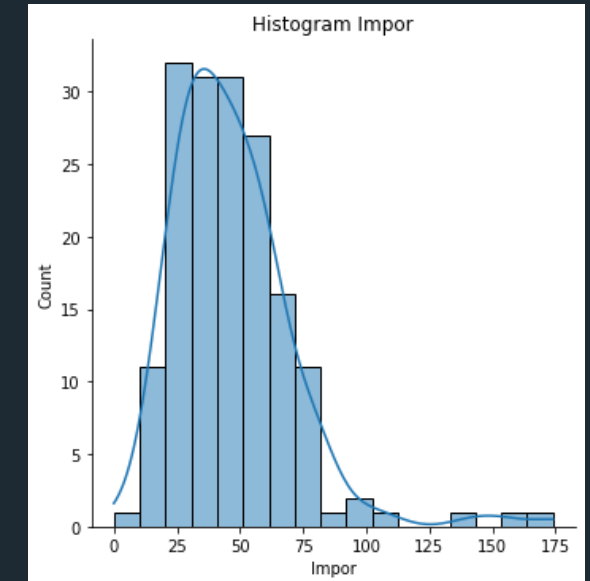
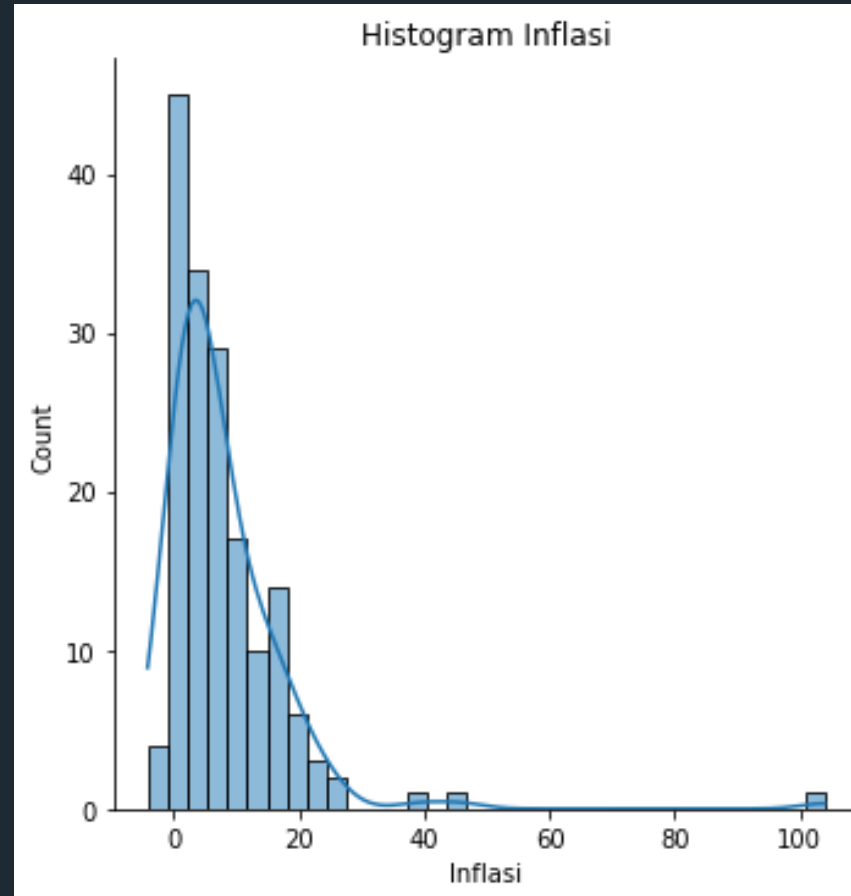
1. . Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 0-10
3. Jumlah tertinggi pada rentang 5 sebanyak lebih dari 40
4. Jumlah terendah pada rentang 20-110 sebanyak mendekati 0

- **Impor:**

1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 12-75
3. Jumlah tertinggi pada rentang 25 – 50 sebanyak lebih dari 30
4. Jumlah terendah pada rentang 0-12 dan 80-175 sebanyak mendekati 0

- **Jumlah Fertiliti:**

1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 1-3
3. Jumlah tertinggi pada rentang 1-3 lebih dari 40
4. Jumlah terendah pada rentang 7 sebanyak mendekati 0



Univariate Analysis (3)

- **Pendapatan:**

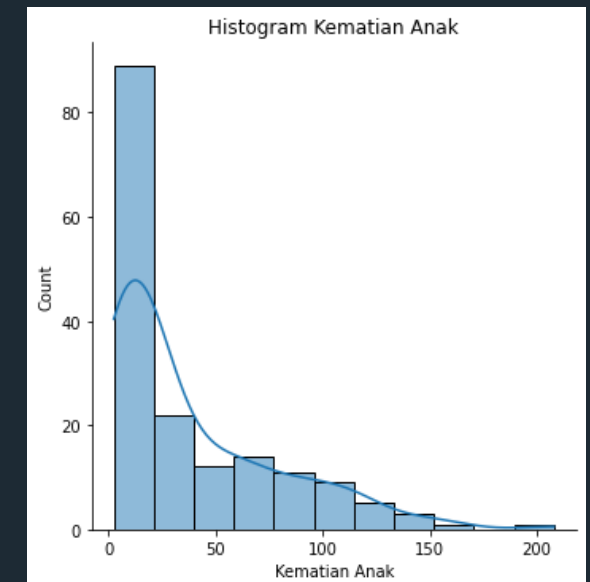
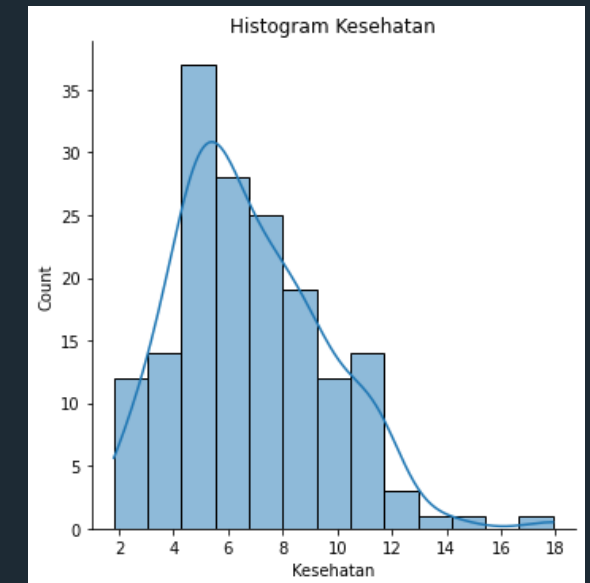
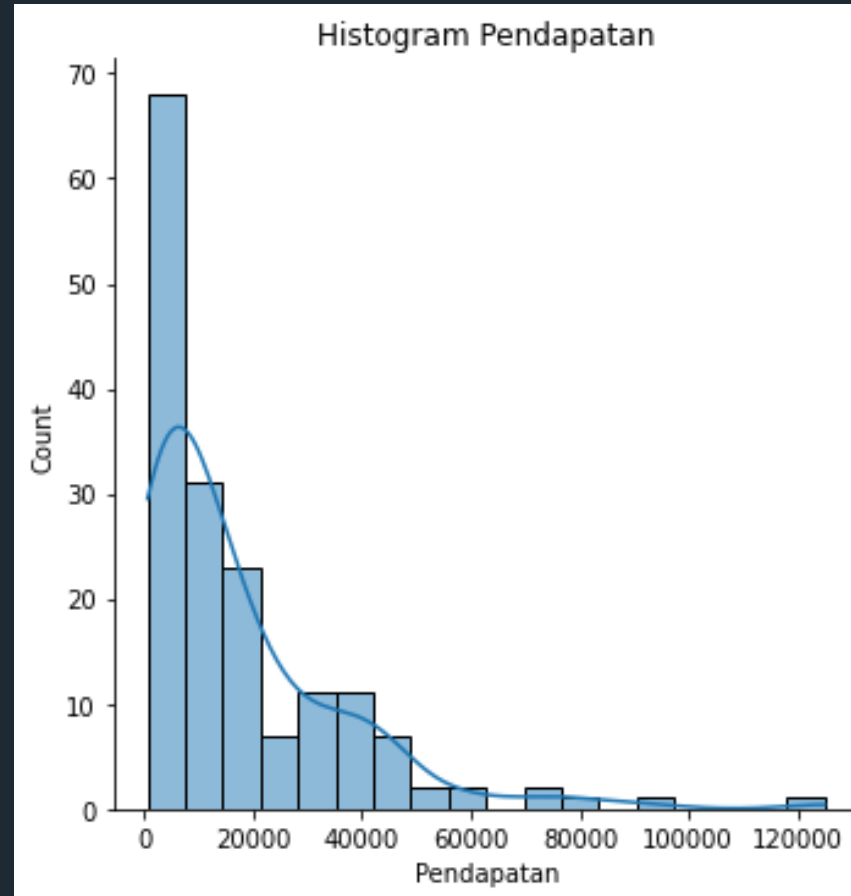
1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 0-20000
3. Jumlah tertinggi pada rentang 0-10000 sebanyak lebih dari 60
4. Jumlah terendah pada rentang 50000-120000 sebanyak mendekati 0

- **Kesehatan:**

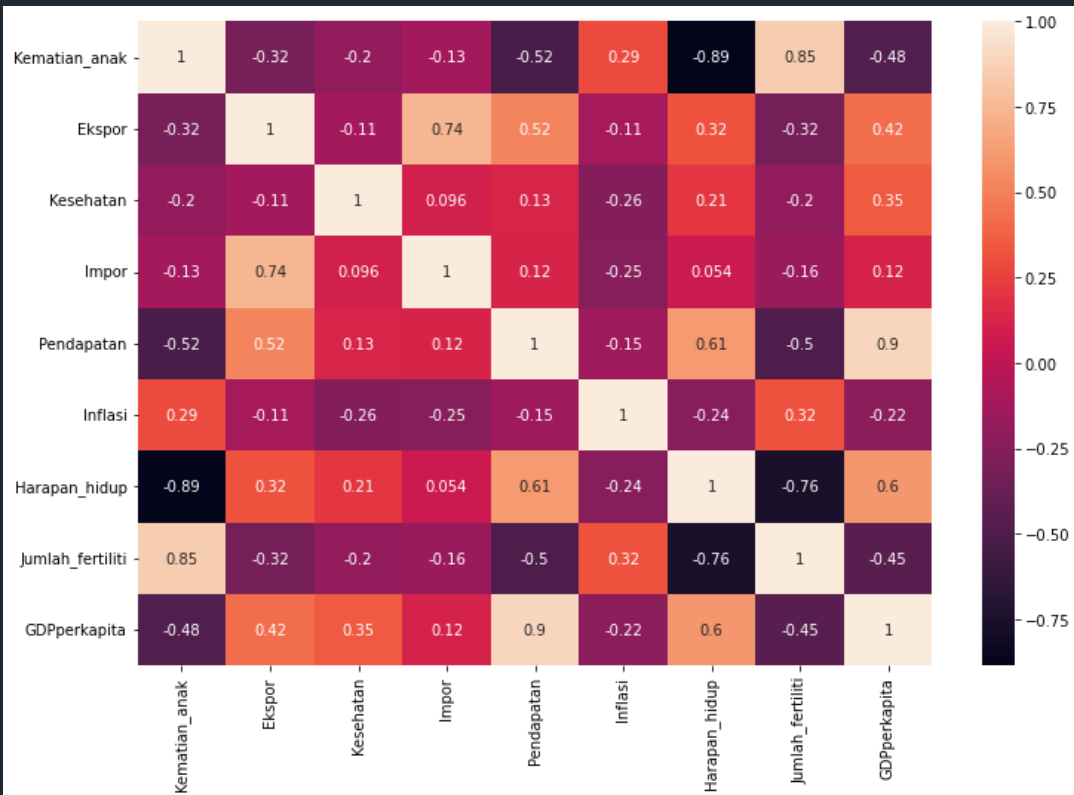
1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 5-9
3. Jumlah tertinggi pada rentang 5 sebanyak lebih dari 35
4. Jumlah terendah pada rentang 13-18 sebanyak mendekati 0

- **Kematian Anak:**

1. Persebaran skew ke kanan
2. Persebaran sebagian besar tersebar pada rentang 0-50
3. Jumlah tertinggi pada rentang 0-50 sebanyak lebih dari 80
4. Jumlah terendah pada rentang 150-200 sebanyak mendekati 0



Multivariate Analysis



KOLOM

Kematian_anak

Ekspor

Kesehatan

Impor

Pendapatan

Inflasi

Harapan_hidup

Jumlah_fertiliti

GDPperkapita

PASANGAN KORELASI YANG KUAT*

Harapan_hidup**, Jumlah_fertility

Impor, Pendapatan

Korelasinya yang cukup lemah dengan kolom lain

Ekspor

Kematian_anak, Ekspor, Harapan_hidup, Jumlah_fertility, GDPperkapita

Korelasinya yang cukup lemah dengan kolom lain

Kematian_anak**, Pendapatan, Jumlah_fertility**, GDPperkapita

Kematian_anak, Pendapatan, Harapan_hidup

Pendapatan, Harapan_hidup

*) Correlation coefficient ≥ 0.5 or ≤ -0.5

**) Berbanding terbalik

Columns for Clustering

Kematian_anak –
Harapan_hidup (-0.89) –
Jumlah_fertility (0.85)

Ekspor – Impor (0.74) –
Pendapatan (0.52)

GDPperkapita –
Pendapatan (0.9) –
Harapan_hidup (0.6)

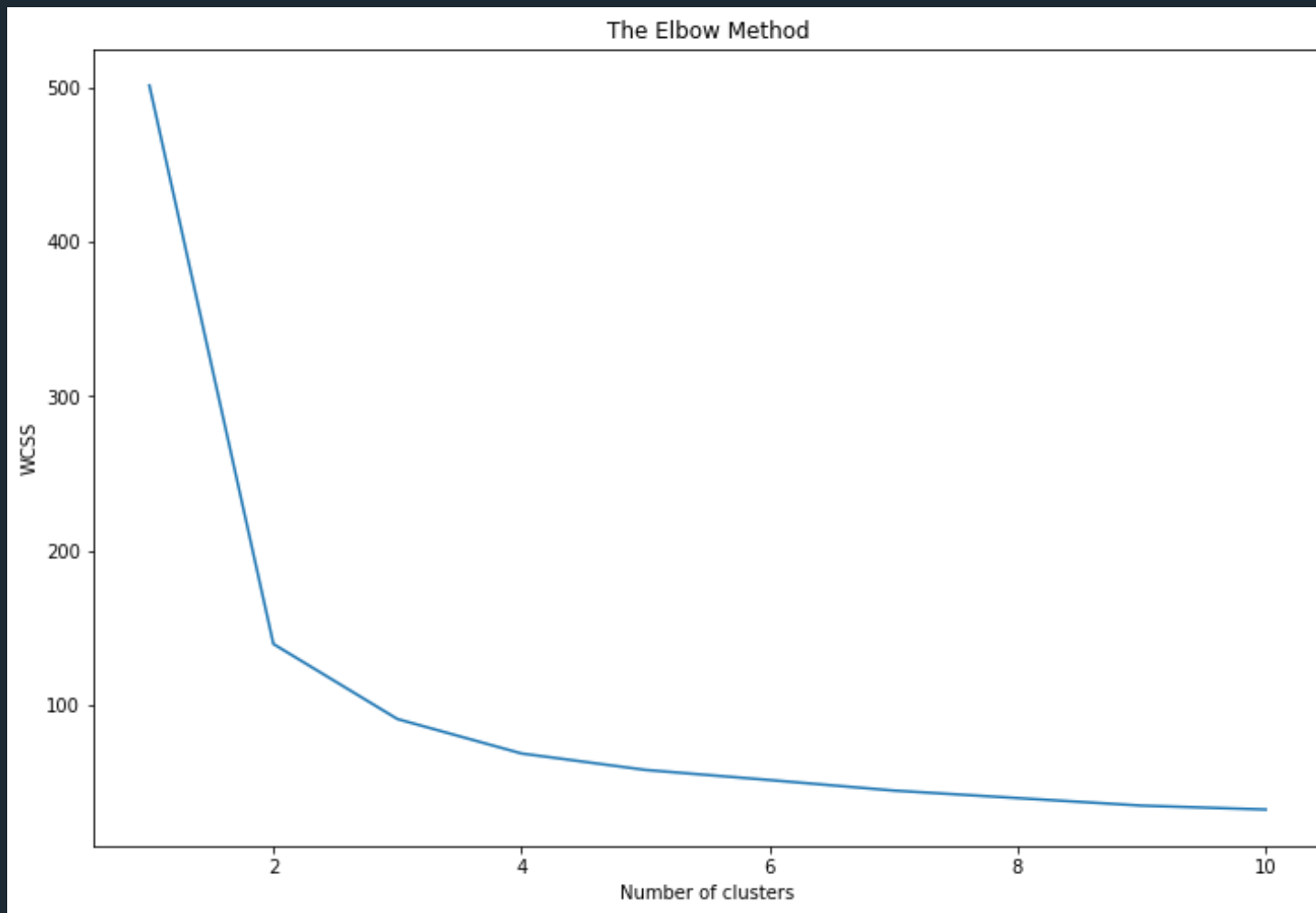
Handling Outliers

- Handling outliers dilakukan dengan menggunakan quantile based flooring and capping
- Quantile based flooring and capping merupakan metode menggantikan outliers dengan batas atas atau batas bawah atau nilai kuartil
- Dipakai karena nilai data terjaga integritasnya dan ukuran data menyesuaikan dengan kelebihan atau kekurangan pada batas - batas.
- Tidak menggunakan remove/trimming karena dapat berpotensi menghilangkan data kandidat satu negara (Kolom negara unik)





Data Clustering –
Kematian_anak,
Harapan_hidup, dan
Jumlah Fertiliti



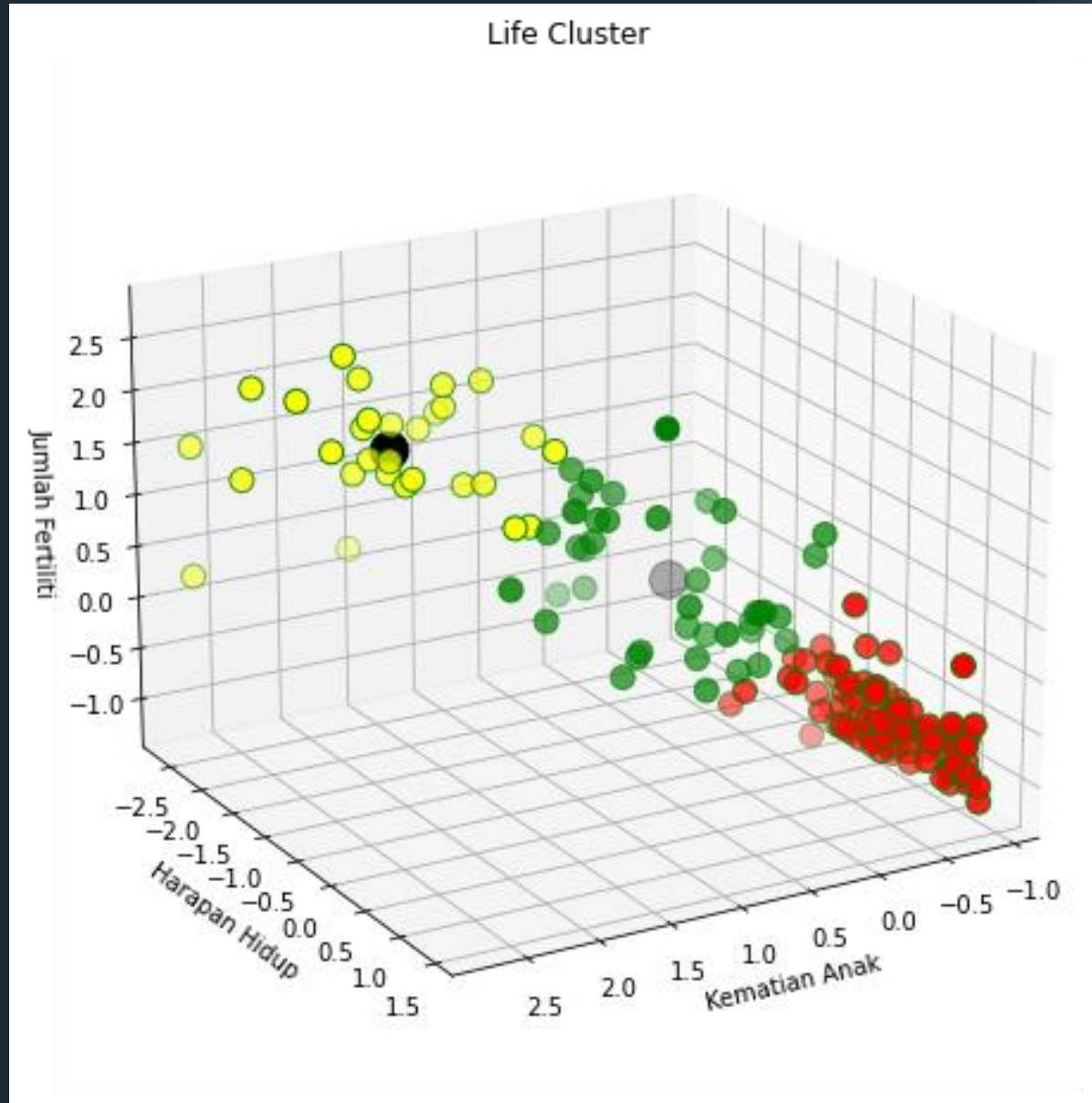
Kematian Anak Elbow Method

Analysis: Nilai K yang
direkomendasikan yaitu $x = 3$

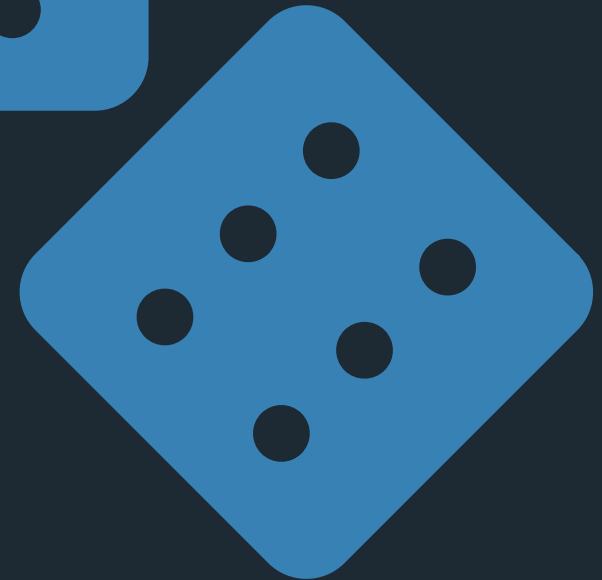
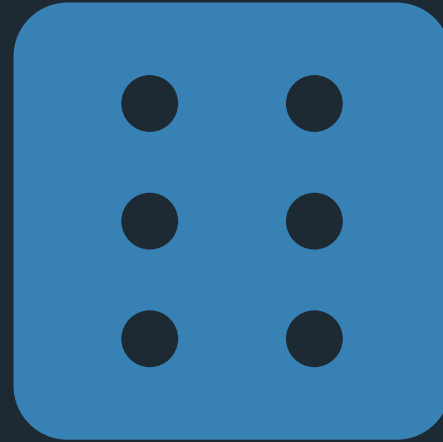


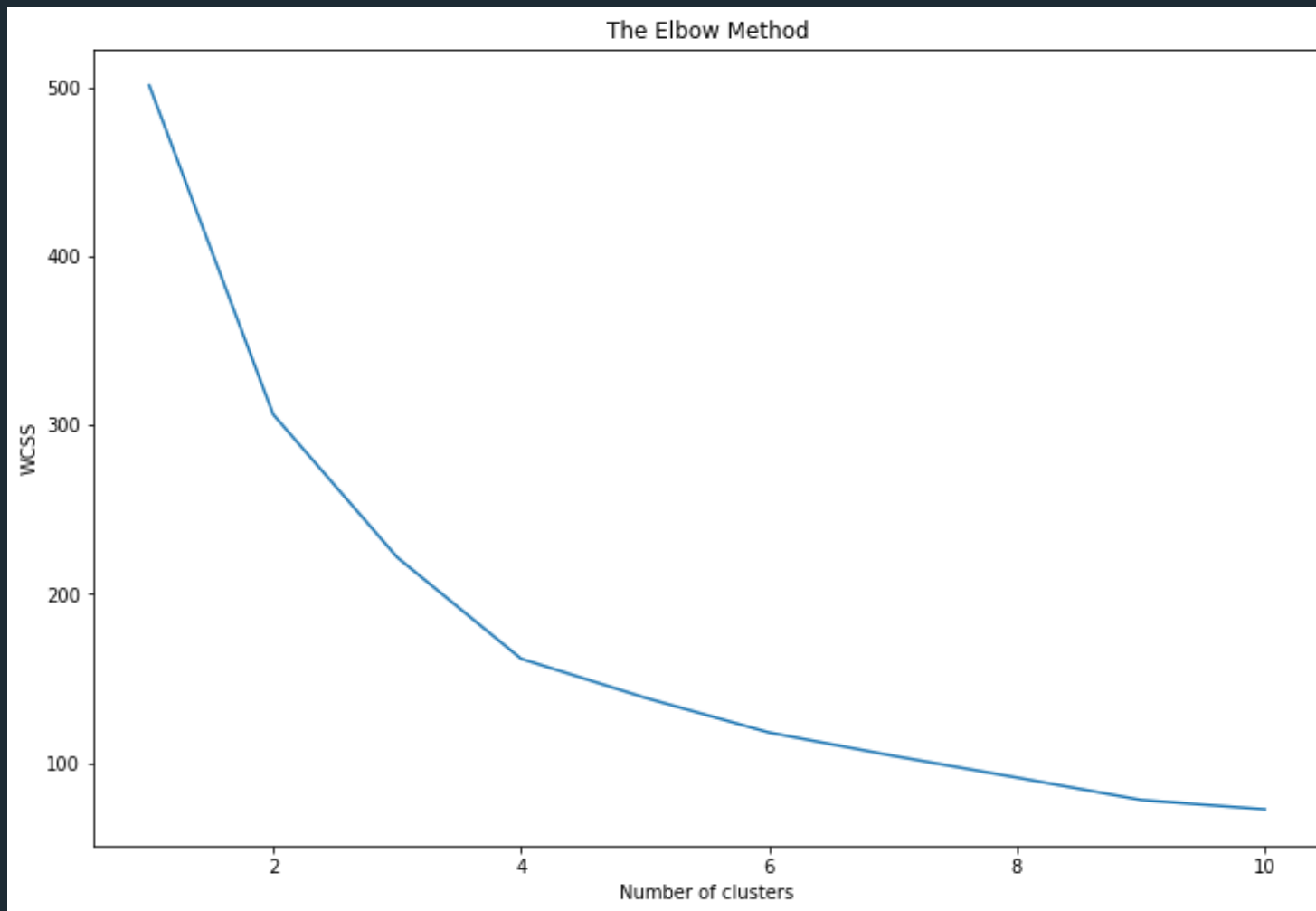
3-Means Cluster

- Analysis:
- Pada cluster 1 (hijau), data memiliki data rentang yang normal (median dari setiap kategori) dari kategori kematian anak, harapan hidup, dan jumlah fertiliti.
- Pada cluster 2 (red), data memiliki angka kematian anak yang rendah, harapan hidup yang tinggi, dan jumlah fertiliti yang rendah.
- Pada cluster 3 (yellow), data memiliki angka kematian anak yang tinggi, harapan hidup yang rendah, dan jumlah fertiliti yang tinggi.



Data Clustering – Ekspor, Impor, dan Pendapatan





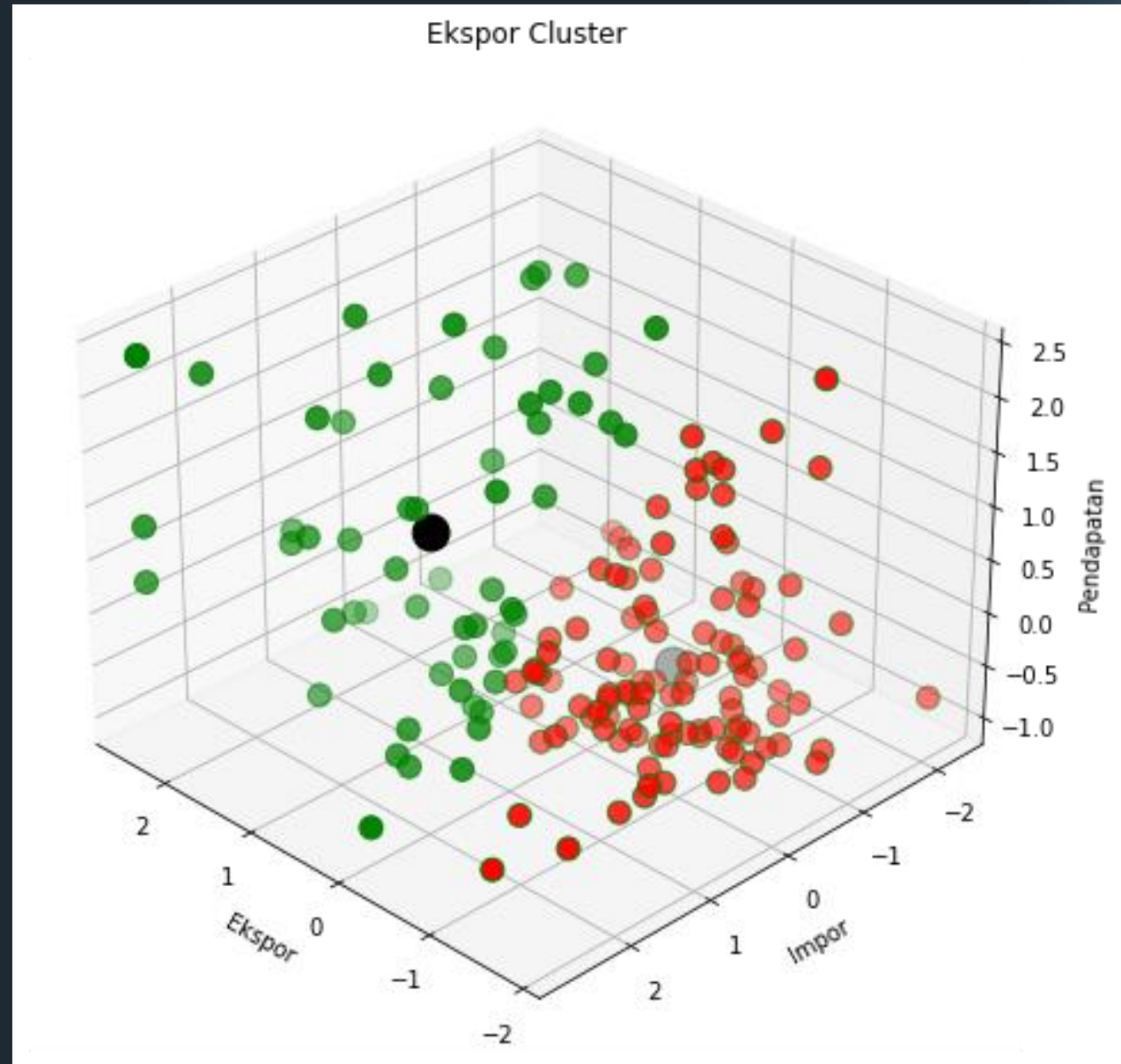
Ekspor Elbow Method

- Analysis: Tidak dapat ditemukan nilai optimal K untuk KMeans

Saya menduga kemungkinan ini karena data kolom Pendapatan yang sangat besar sehingga mempengaruhi skala setiap jarak pada saat dilakukan scaling.

2-Means Clustering

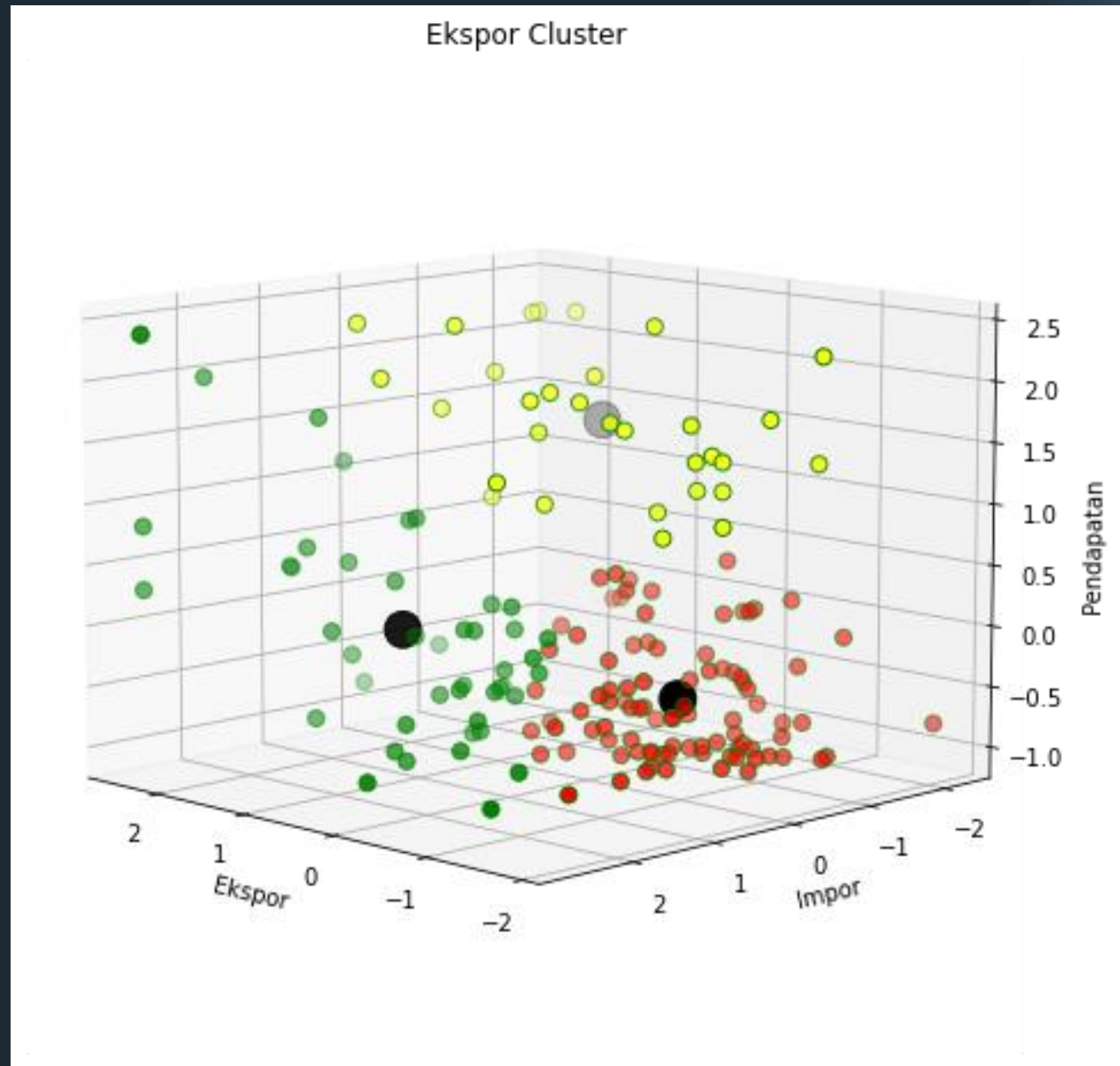
- Analysis: Persebaran data terlalu general sehingga tidak dapat diambil kesimpulan dari data.



3-Means Clustering

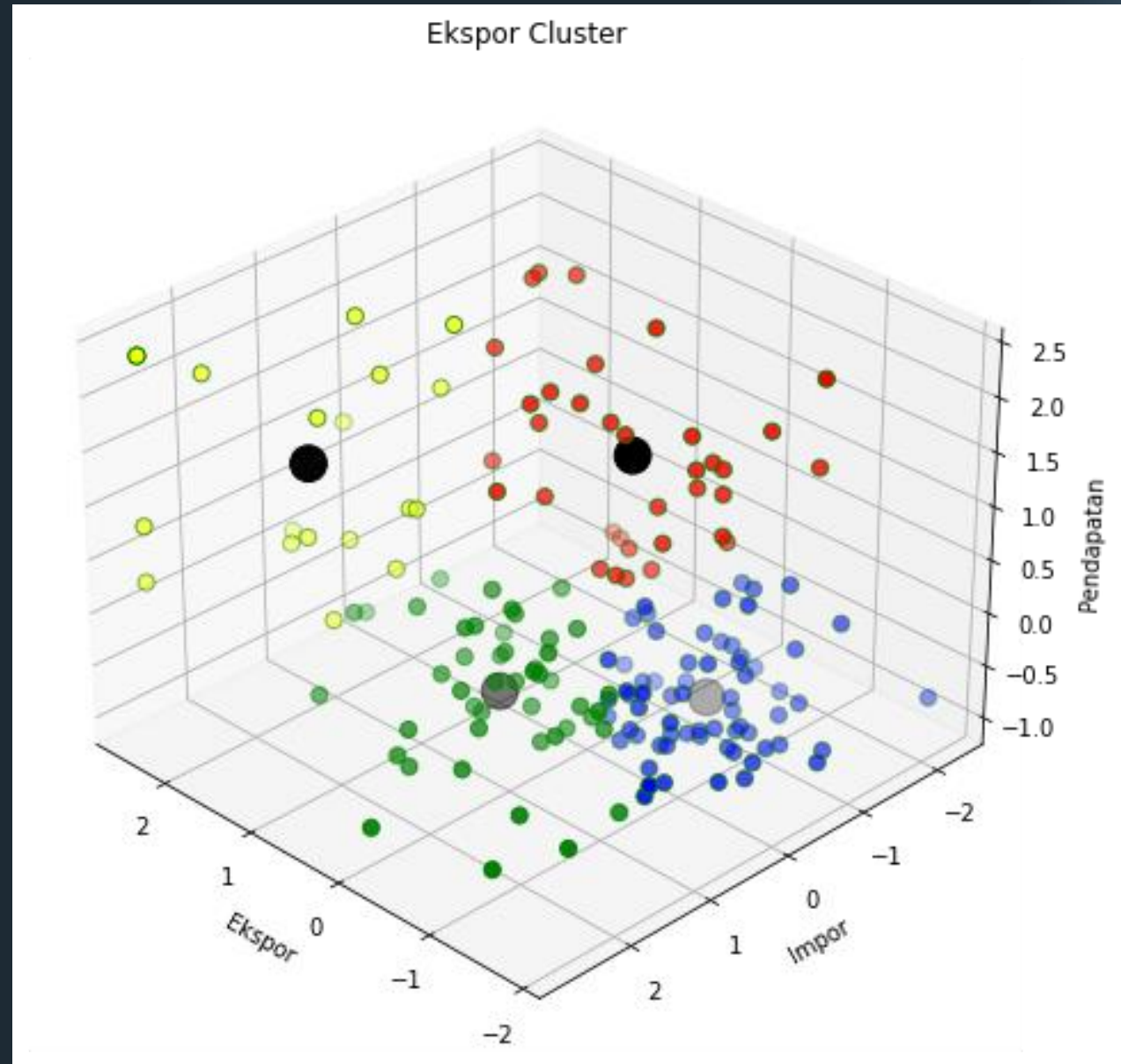
- Analysis:
- Cluster 1 (green), memiliki angka impor yang lebih tinggi
- Cluster 2 (red), memiliki angka pendapatan yang rendah dan impor yang rendah Cluster 3 (yellow), memiliki angka impor, ekspor dan pendapatan yang tinggi


Untuk cluster 1 dan 3 masih terdapat beberapa data yang general dan tidak dapat diambil kesimpulan terutama untuk kolom ekspor.



4-Means Clustering

- Analysis:
- Cluster 1 (green), memiliki angka impor yang tinggi, ekspor rentang 2 sampai -1, dan pendapatan yang rendah
- Cluster 2 (red), memiliki angka impor yang rendah, ekspor yang tinggi, dan pendapatan yang tinggi
- Cluster 3 (yellow), memiliki angka import yang tinggi, ekspor yang tinggi, dan pendapatan yang tinggi
- Cluster 4 (blue), memiliki angka impor yang rendah, ekspor yang rendah, dan pendapatan yang rendah

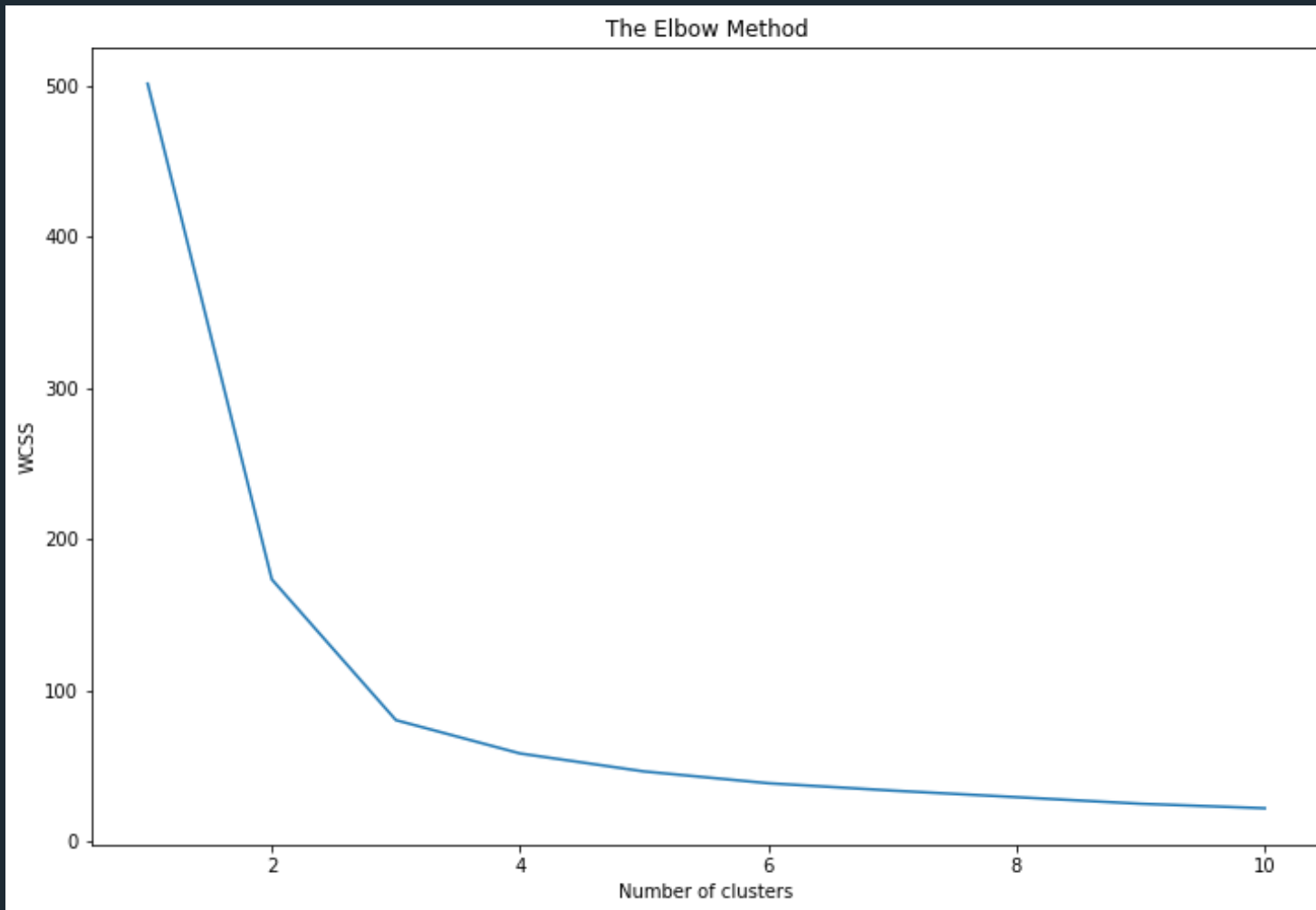




Data Clustering – GDPperkapita, Harapan_hidup, dan Pendapatan

GDP per kapita Elbow Method

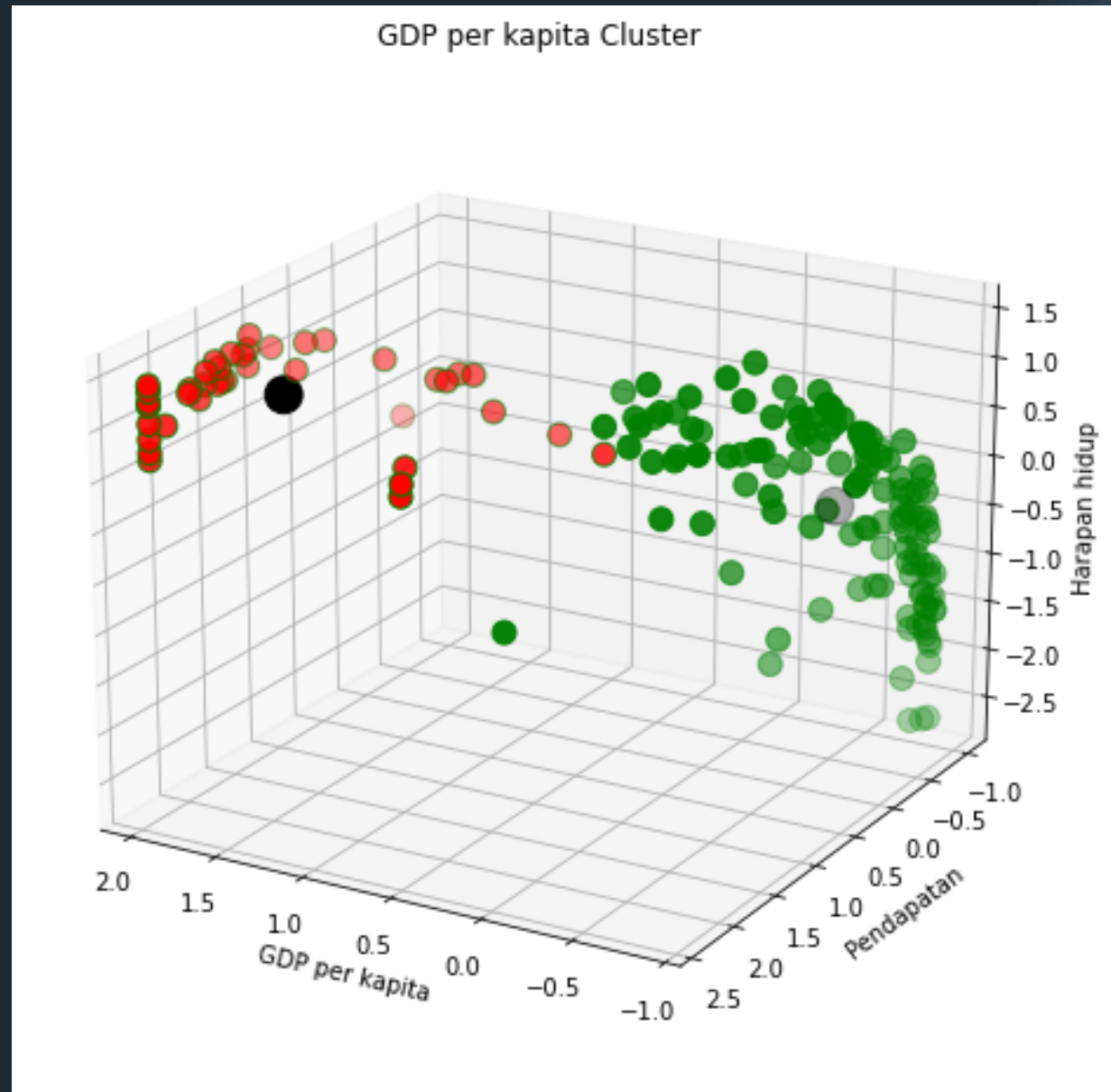
- Terdapat titik elbow di $x = 2$ atau 3



2-Means Clustering

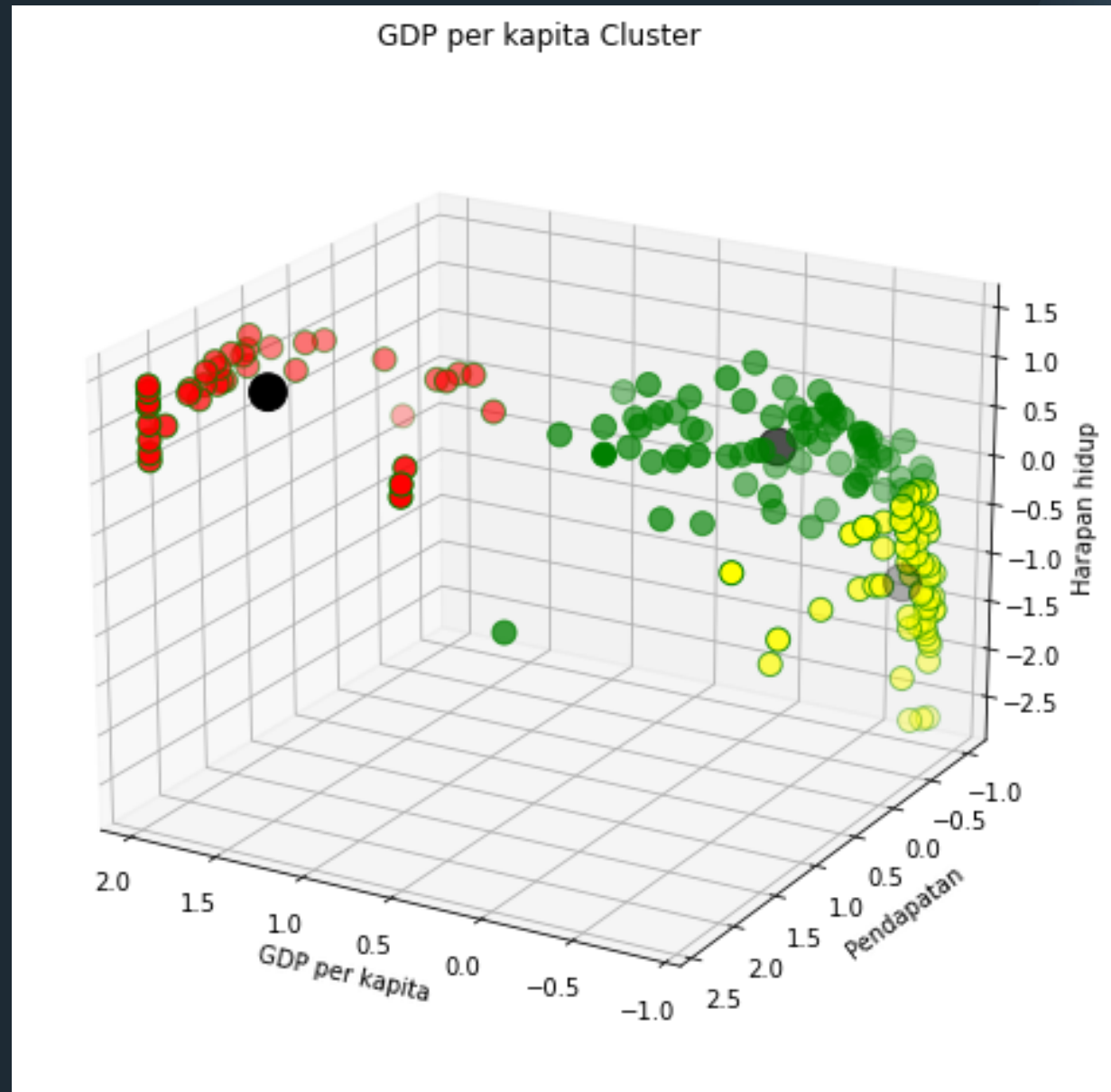
- Analysis:
- Cluster 1 (green), memiliki value harapan hidup yang bervariasi, pendapatan yang rendah, dan GDP per kapita yang rendah
- Cluster 2 (red), memiliki value harapan hidup yang tinggi, pendapatan yang tinggi, dan GDP per kapita yang tinggi

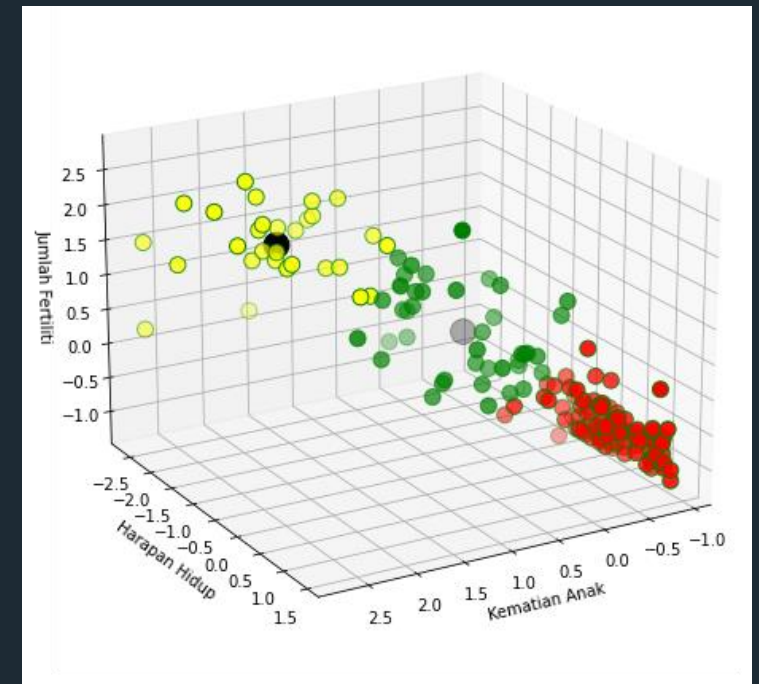
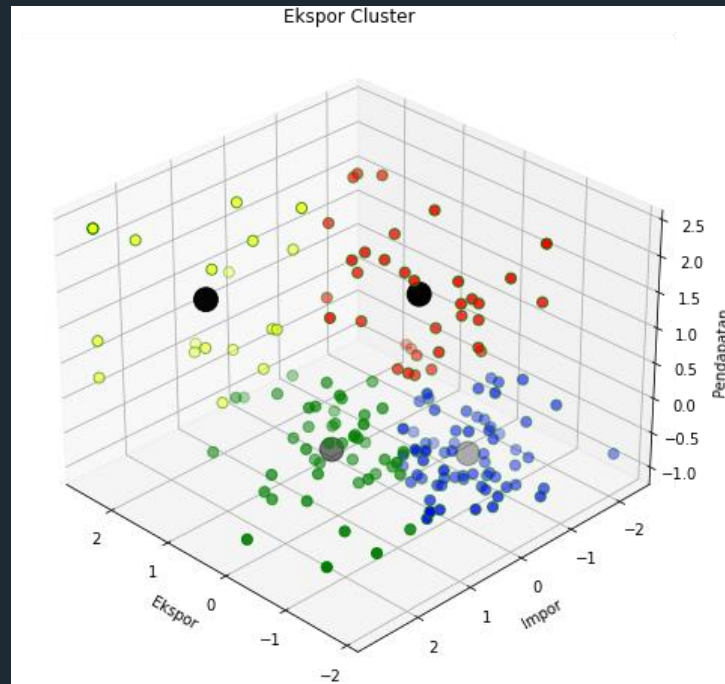
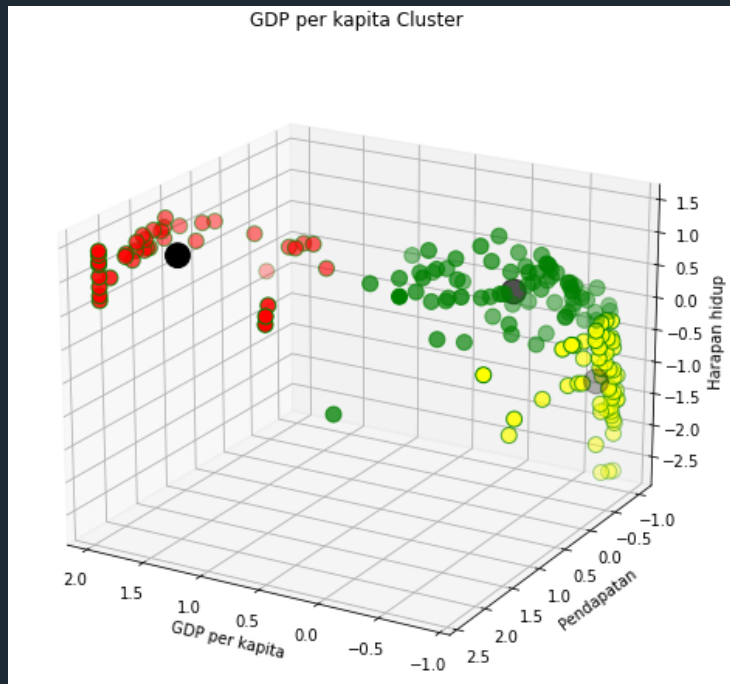
Karena value harapan hidup pada cluster 1 masih bervariasi, ditambahkan satu cluster lagi supaya dapat mendapatkan perbedaan dari harapan hidup



3-Means Clustering

- Analysis:
- Cluster 1 (green), memiliki value harapan hidup yang tinggi, pendapatan yang rendah, dan GDP per kapita pada rentang 1 sampai -1
- Cluster 2 (red), memiliki value harapan hidup yang tinggi, pendapatan yang tinggi, dan GDP per kapita yang tinggi
- Cluster 3 (yellow), memiliki value harapan hidup yang rendah, pendapatan yang rendah, dan GDP per kapita yang rendah





Data Insight 3

Columns Cluster

- Dari gambar GDP per kapita Cluster, diambil cluster warna kuning karena memiliki harapan hidup, pendapatan, dan GDP per kapita yang rendah
- Dari gambar Ekspor Cluster, diambil cluster warna biru karena memiliki pendapatan, impor, dan ekspor yang rendah
- Dari gambar Life Cluster, diambil cluster warna kuning, karena memiliki tingkat kematian anak dan jumlah fertility yang tinggi namun harapan hidup yang rendah



Summary

- Dari data insight sebelumnya dan telah diambil data berdasarkan cluster tadi, saya merekomendasikan 10 negara berikut untuk diberikan bantuan

Negara	Kematian_anak	Ekspor	Impor	Pendapatan	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
Afghanistan	90.2	10.00	44.9	1610	56.2	5.82	553
Benin	111.0	23.80	37.2	1820	61.8	5.36	758
Burkina Faso	116.0	19.20	29.6	1430	57.9	5.87	575
Burundi	93.6	8.92	39.2	764	57.7	6.26	231
Cameroon	108.0	22.20	27.0	2660	57.3	5.11	1310
Central African Republic	149.0	11.80	26.5	888	47.5	5.21	446
Chad	150.0	36.80	43.5	1930	56.5	6.59	897
Comoros	88.2	16.50	51.7	1410	65.9	4.75	769
Gambia	80.3	23.80	42.7	1660	65.5	5.71	562
Guinea	109.0	30.30	43.2	1190	58.0	5.34	648

