

# CLUSTERING COUNTRIES

[www.helpinternational.com](http://www.helpinternational.com)





# INTRO- DUCTION

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.



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.

**UNTUK ITU,**

Dibutuhkan pengkategorian negara menggunakan beberapa faktor sosial ekonomi dan kesehatan yang menentukan perkembangan negara secara keseluruhan yang nantinya akan menyimpulkan negara mana saja yang paling perlu menjadi fokus CEO



# SOURCE DATA

DATASET'S FEATURE

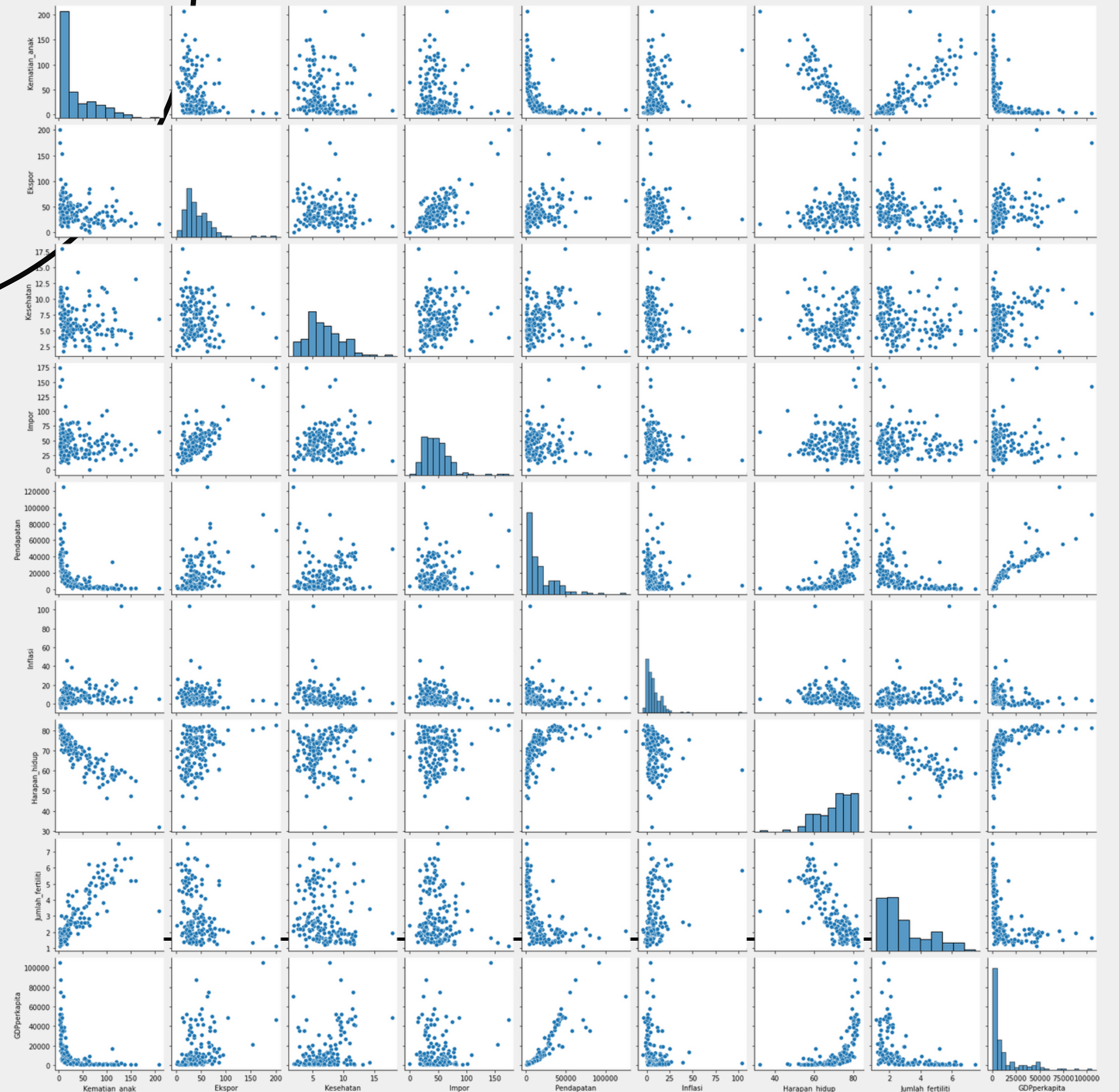
	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
0	Afganistan	90.2	10.0	7.58	44.9	1610	9.44	50.2	5.82	555
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	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

167 rows × 10 columns

AMOUNT OF DATA



# Multivariate



	Harapan_hidup	GDPperkapita
0	56.2	553.0
1	76.3	4090.0
2	76.5	4460.0
3	60.1	3530.0
4	76.8	12200.0
...	...	...
162	63.0	2970.0
163	75.4	13500.0
164	73.1	1310.0
165	67.5	1310.0
166	52.0	1460.0
167 rows × 2 columns		



Feature ini akan digunakan sebagai basis untuk analisis dan clustering

Pemilihan feature ini didasari dari perhitungan korelasi antar kolom yang menyangkut kesehatan dan sosial ekonomi tersendiri yang mana Harapan\_hidup dapat mewakili semua kolom yang menyangkut kesehatan dan GDPperkapita dapat mewakili semua kolom yang menyangkut sosial ekonomi

# Missing Values

Setelah di cek, tidak terdapat missing values pada kedua feature sehingga tidak diperlukan penanganan lebih lanjut.

```
# Find missing values  
nation[['Harapan_hidup', 'GDPperkapita']].isnull().sum()
```

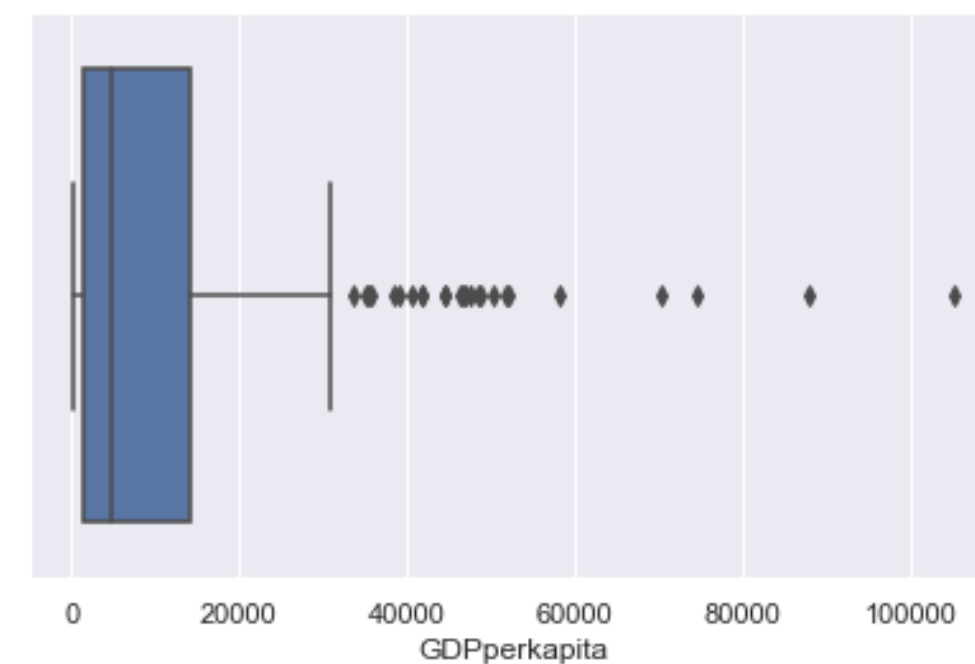
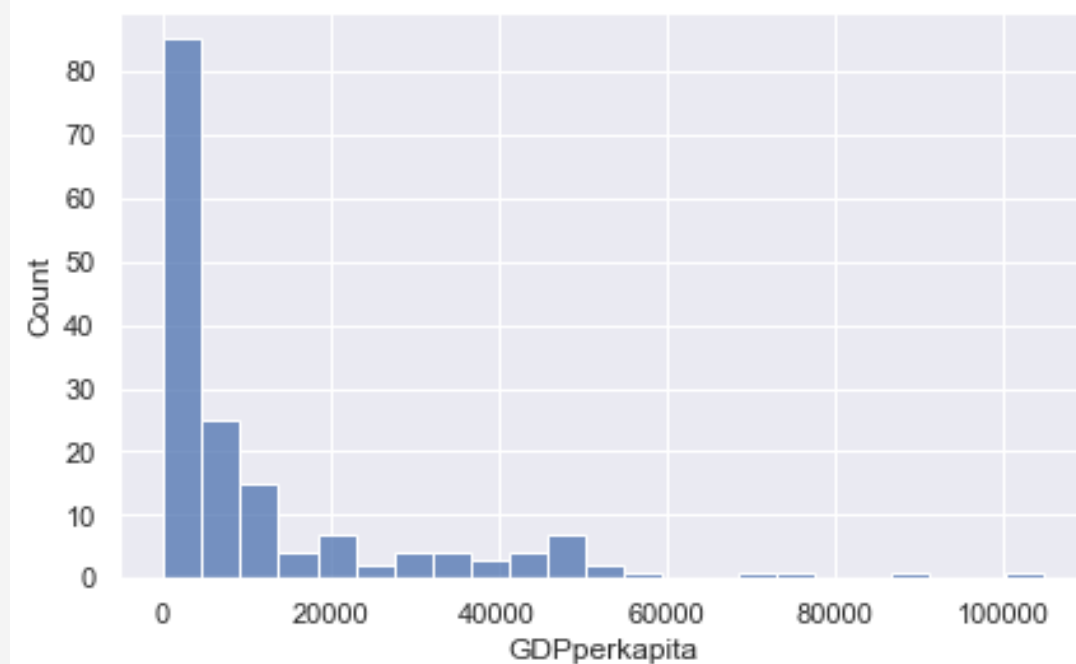
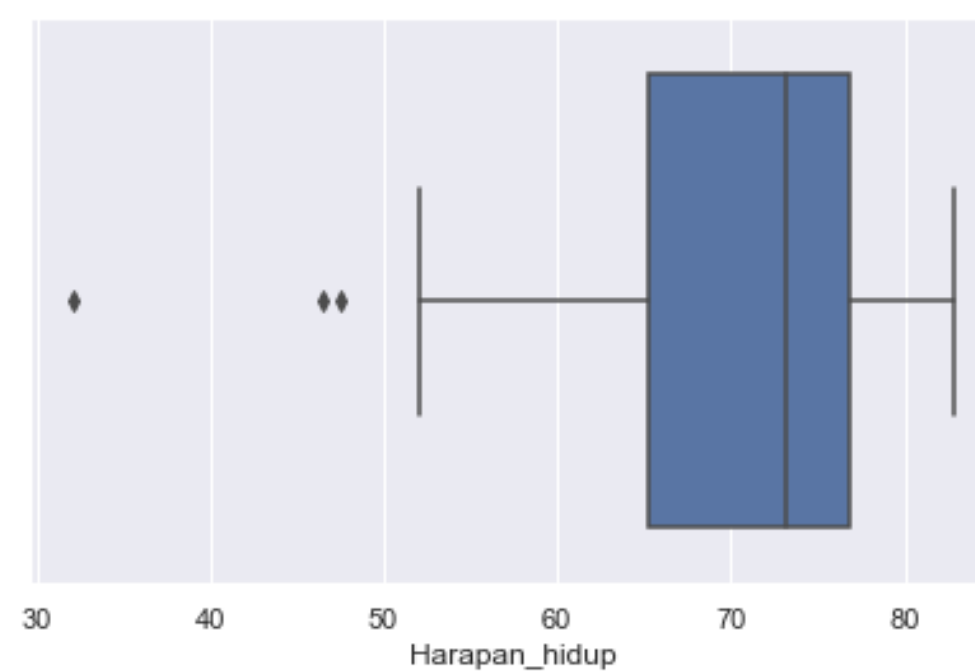
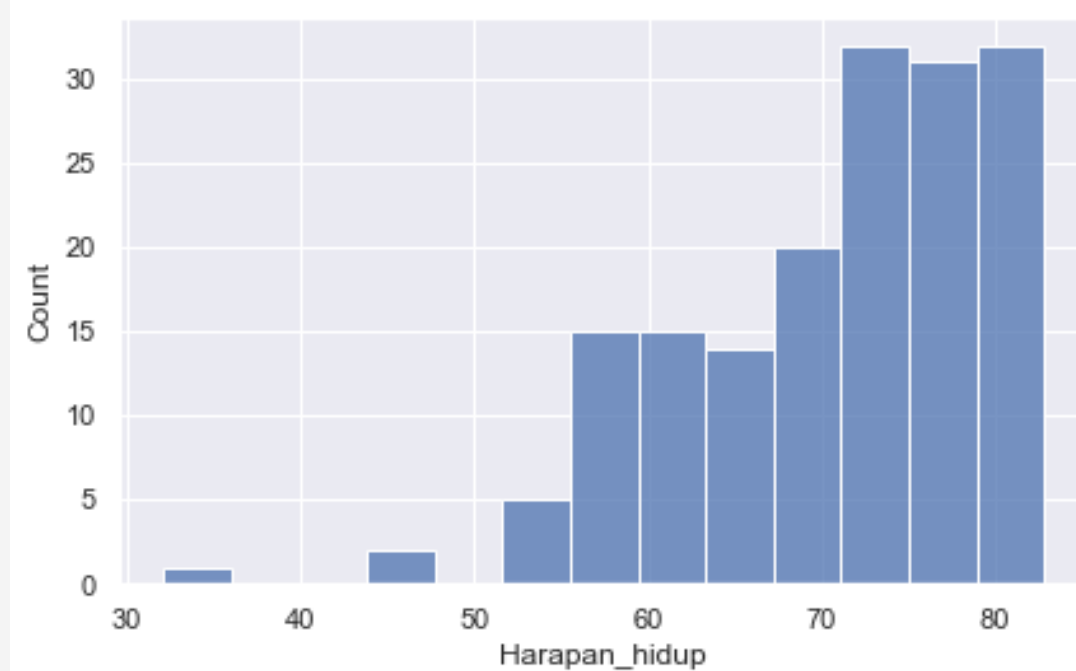
✓ 0.1s

```
Harapan_hidup    0  
GDPperkapita     0  
dtype: int64
```





# OUTLIERS



**Terdapat outliers pada kedua feature sehingga membutuhkan penanganan lebih lanjut.**



# ***HANDLING OUTLIERS***



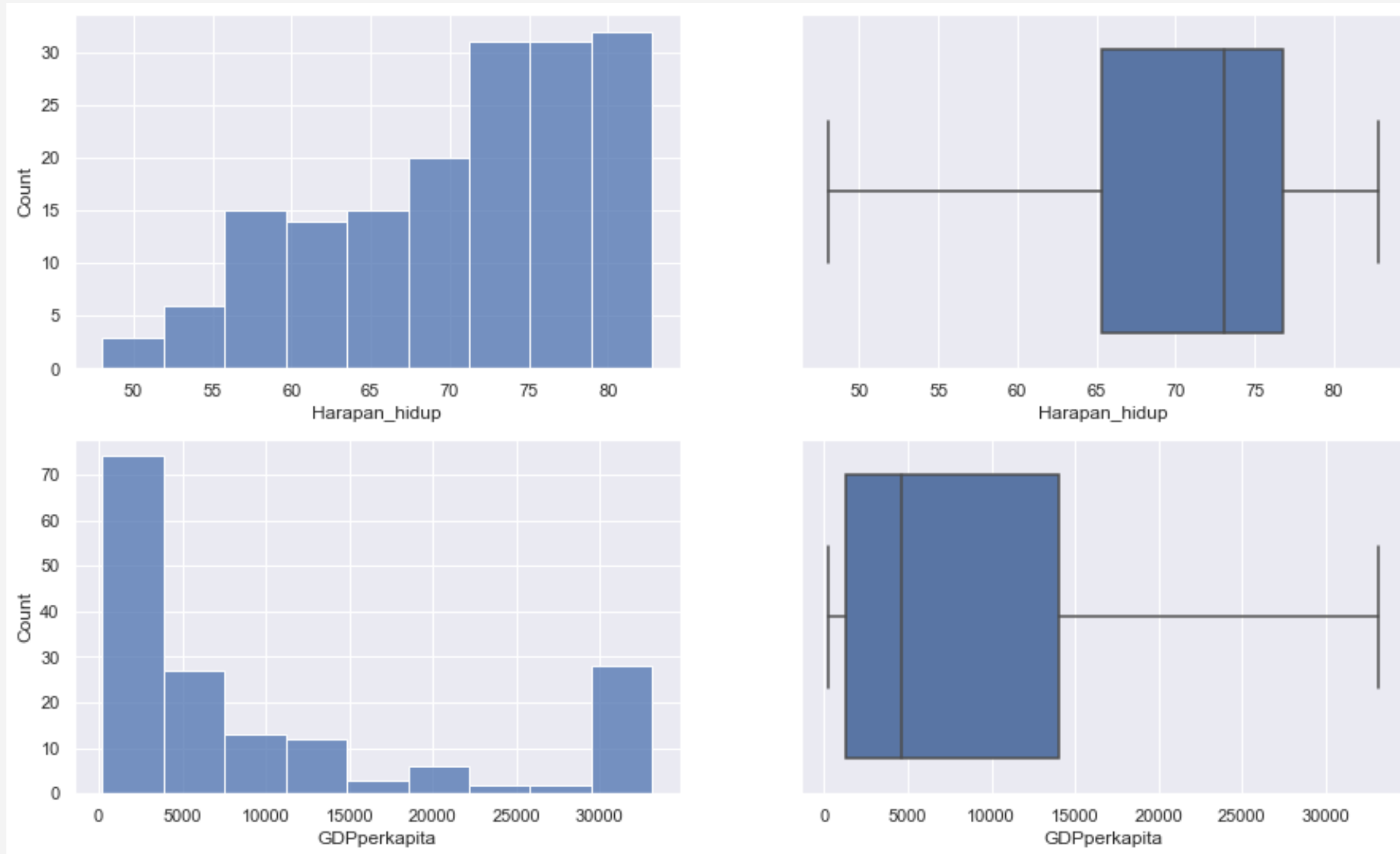
Metode yang digunakan untuk menangani outliers adalah dengan mengubah semua outliers menjadi batas yang telah ditentukan

Metode ini digunakan karena tidak mengurangi/mengubah data yang sudah ada

Batas diperoleh dari interquartile range method. Hal ini didasarkan dari skewness data yang diperoleh adalah right-skewed/left-skewed (bukan normal distribution)



# OUTLIERS CHECK

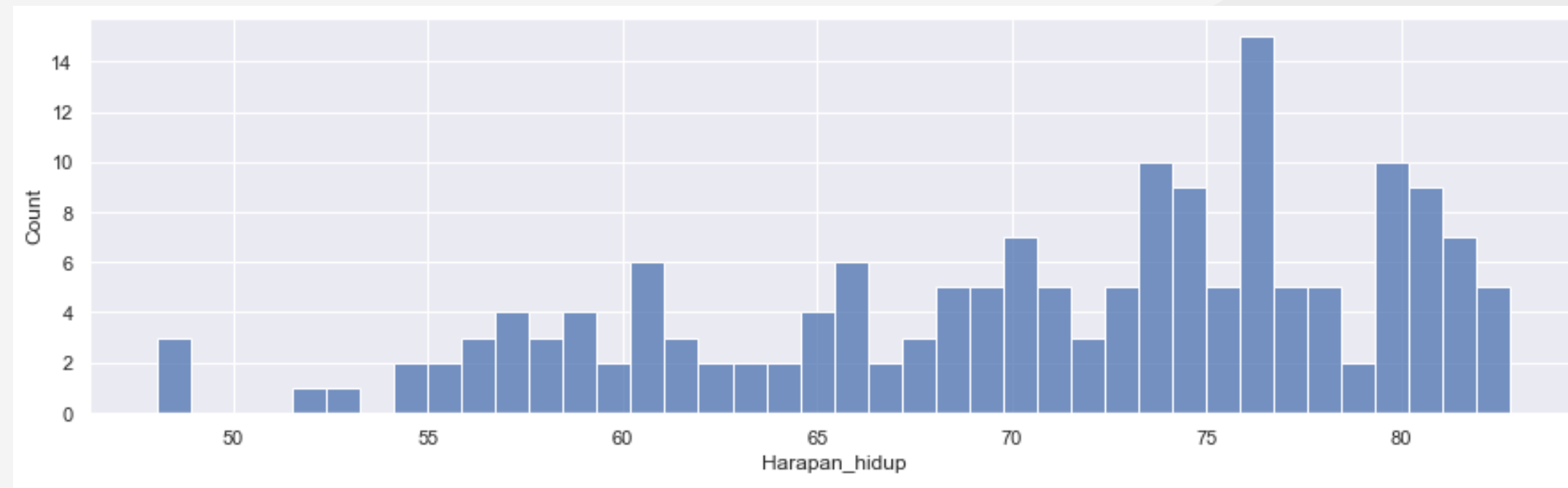


**Grafik setelah  
difilter  
menunjukkan  
sudah tidak ada  
outliers pada  
data**



# ANALISIS FEATURE

## Harapan\_hidup



71-80 tahun  
46.1%

Persebaran harapan hidup cenderung kearah kanan (right-skewed) yang menandakan banyak negara yang memiliki harapan hidup tinggi. Untuk banyaknya sendiri, 46.1 % negara memiliki harapan hidup diantara 71-80 tahun. Sedangkan yang lain terdapat tiga negara yang memiliki harapan hidup dibawah 50 tahun.

61-70 tahun  
25.7%

80-100 tahun  
13.8%

51-60 tahun  
12.6%

0-50 tahun  
1.8%

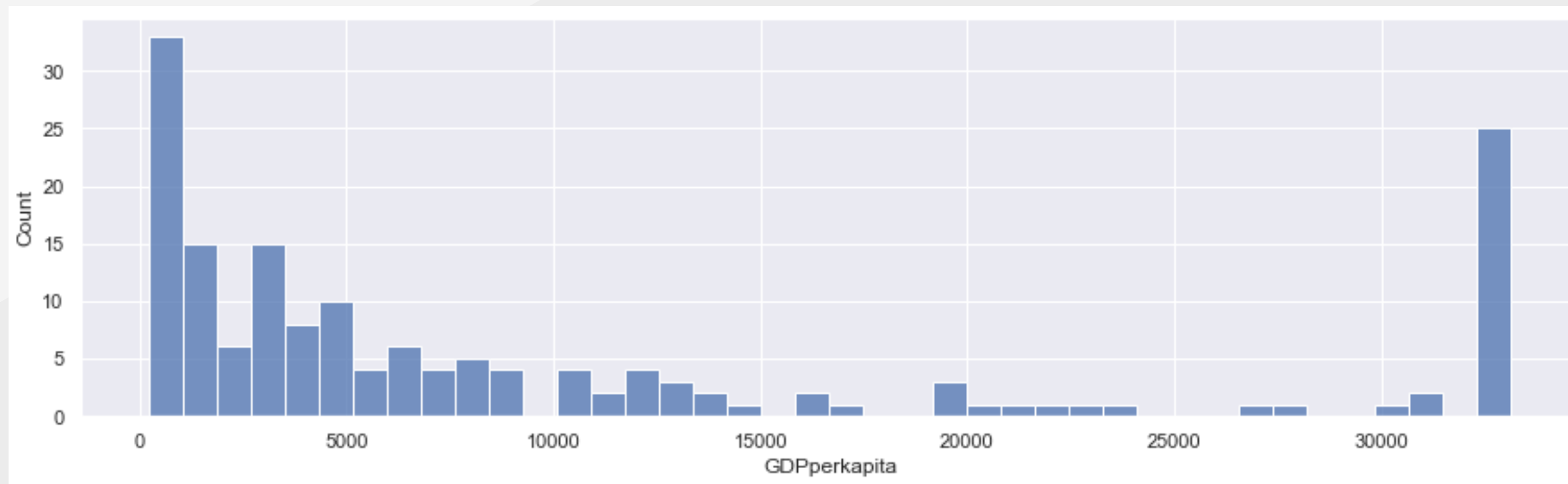
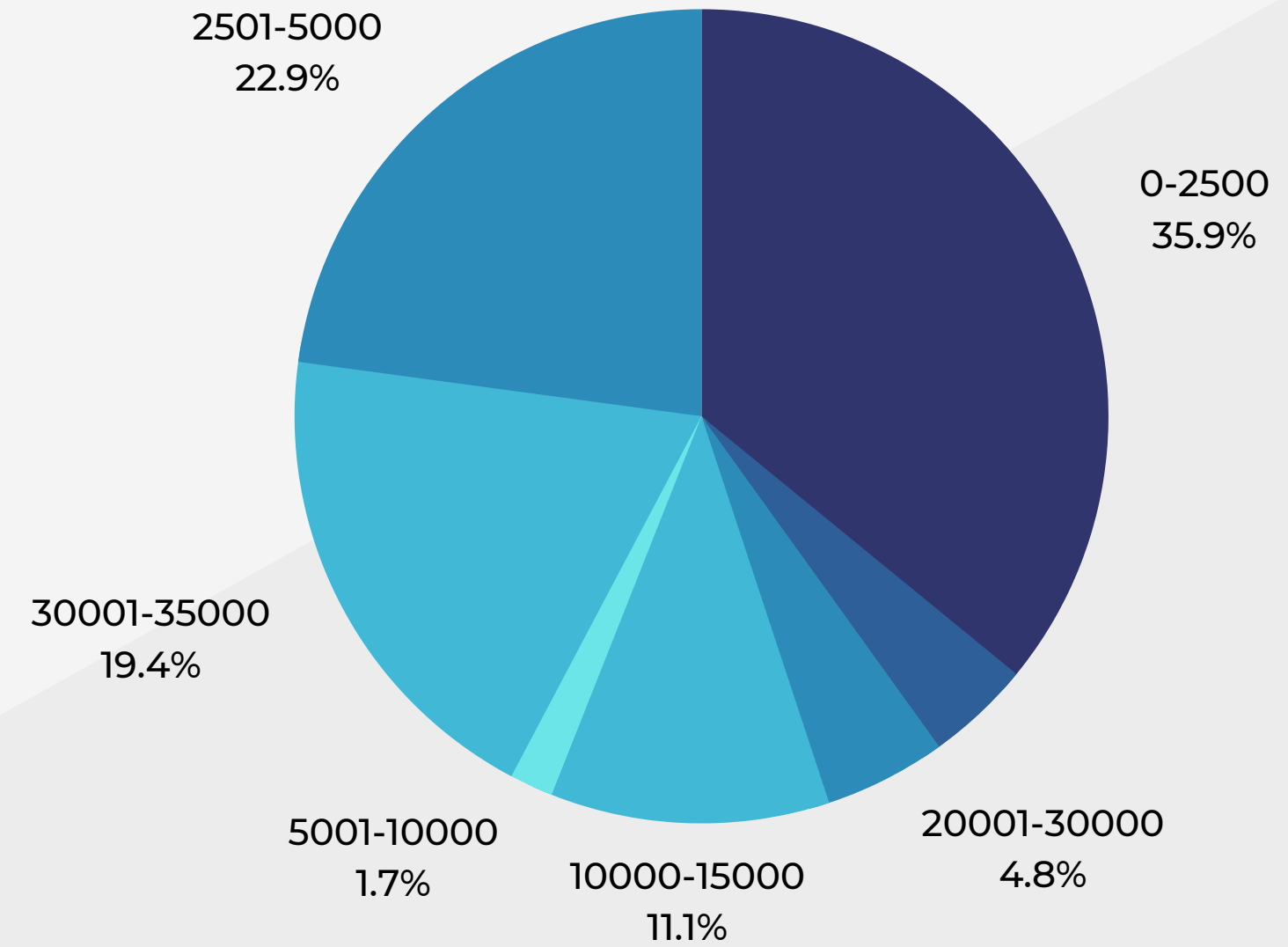




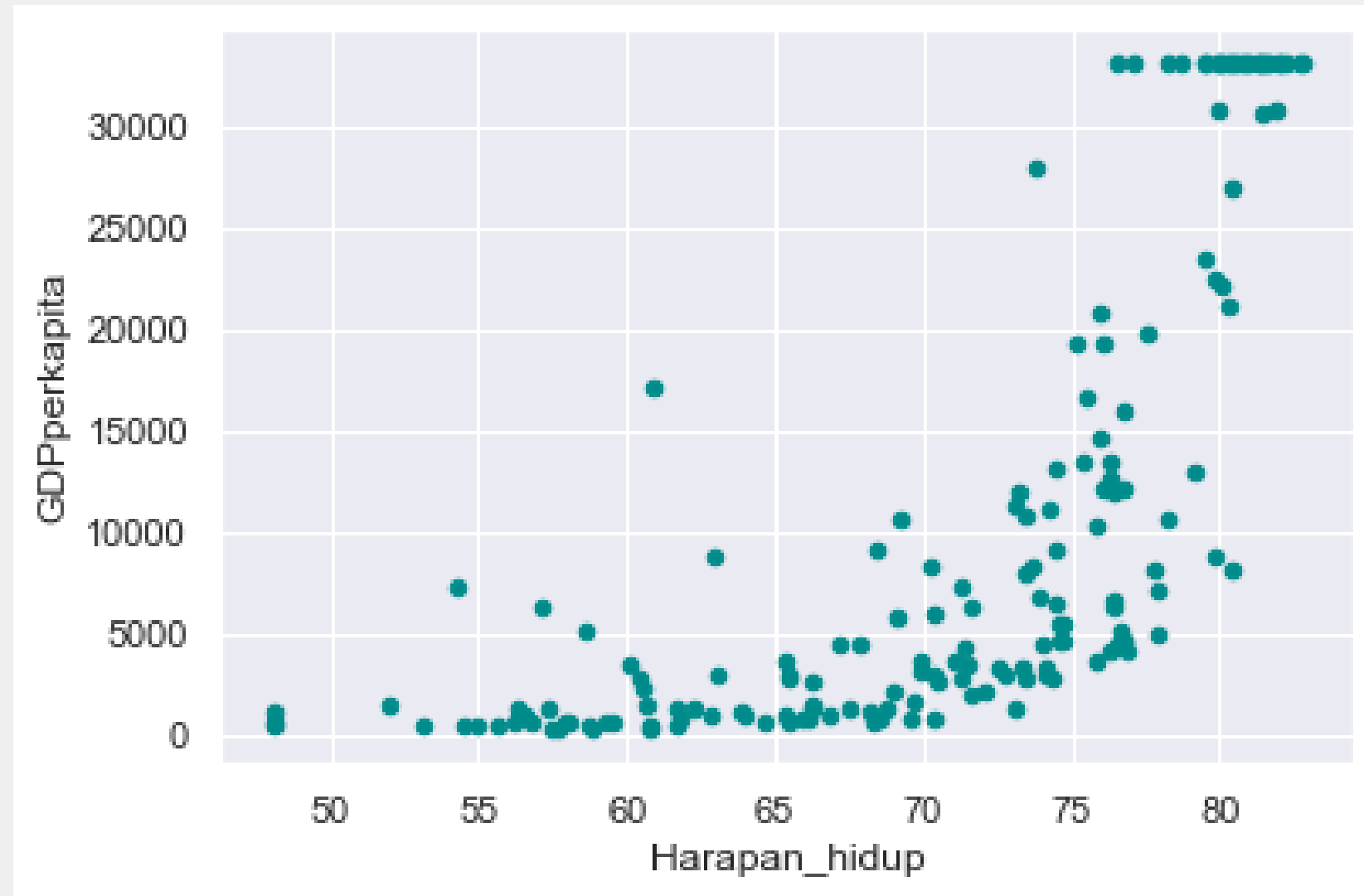
# ***ANALISIS FEATURE***

## **GDPperkapita**

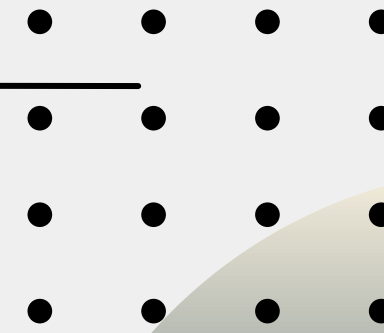
Persebaran data GDPperkapita cenderung ke kiri (left-skewed) yang menandakan banyak negara yang memiliki angka GDP yang rendah yang didominasi diantara 0 sampai 2500 sebanyak 31.1%.



# Analisis Bivariate



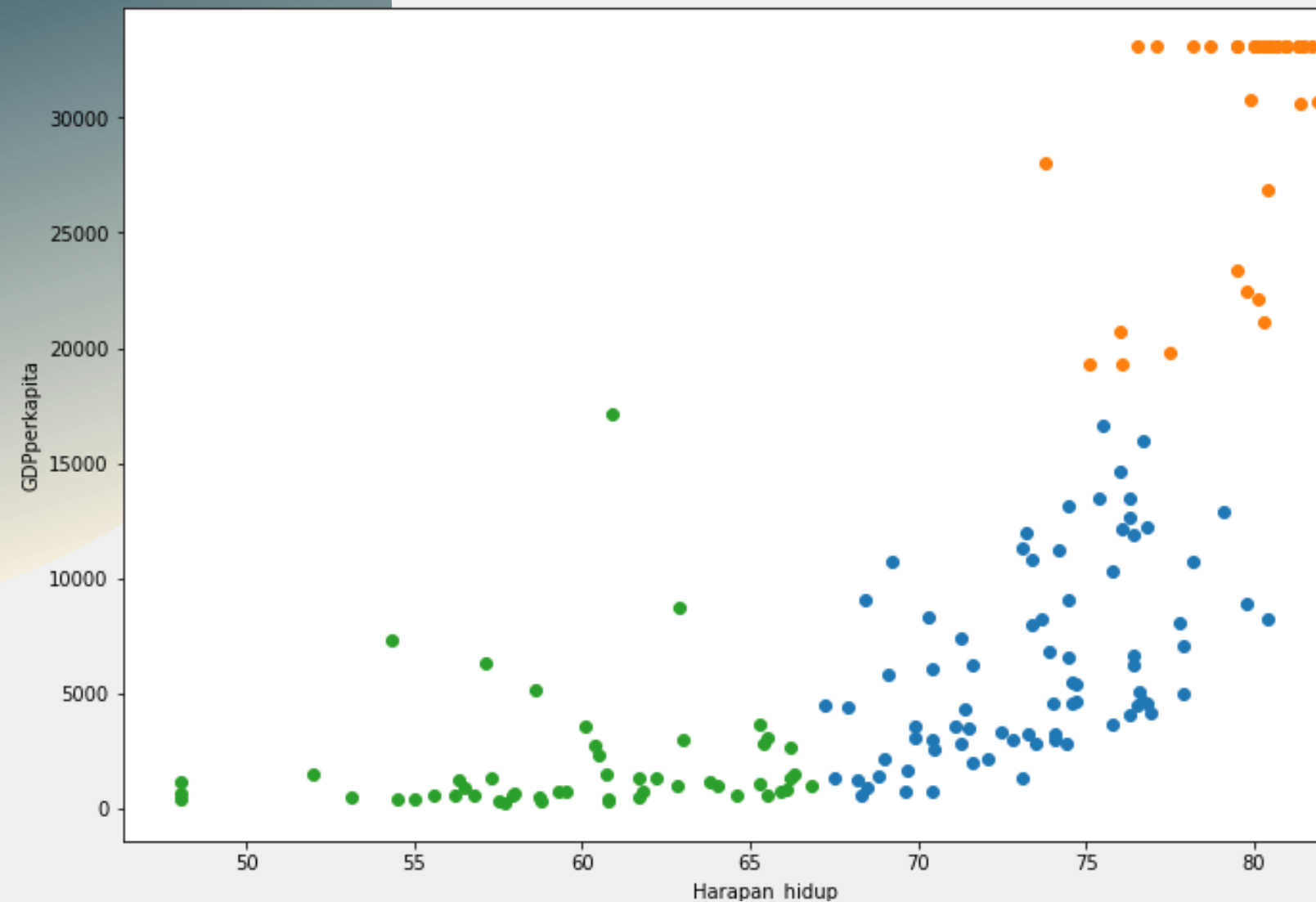
Data tersebar dengan merata dan berbanding lurus terlihat pada grafik berbentuk garis keatas.



Koefisien korelasi antara Harapan hidup dan GDP mendekati 1 yang berarti saling berkaitan berbanding lurus

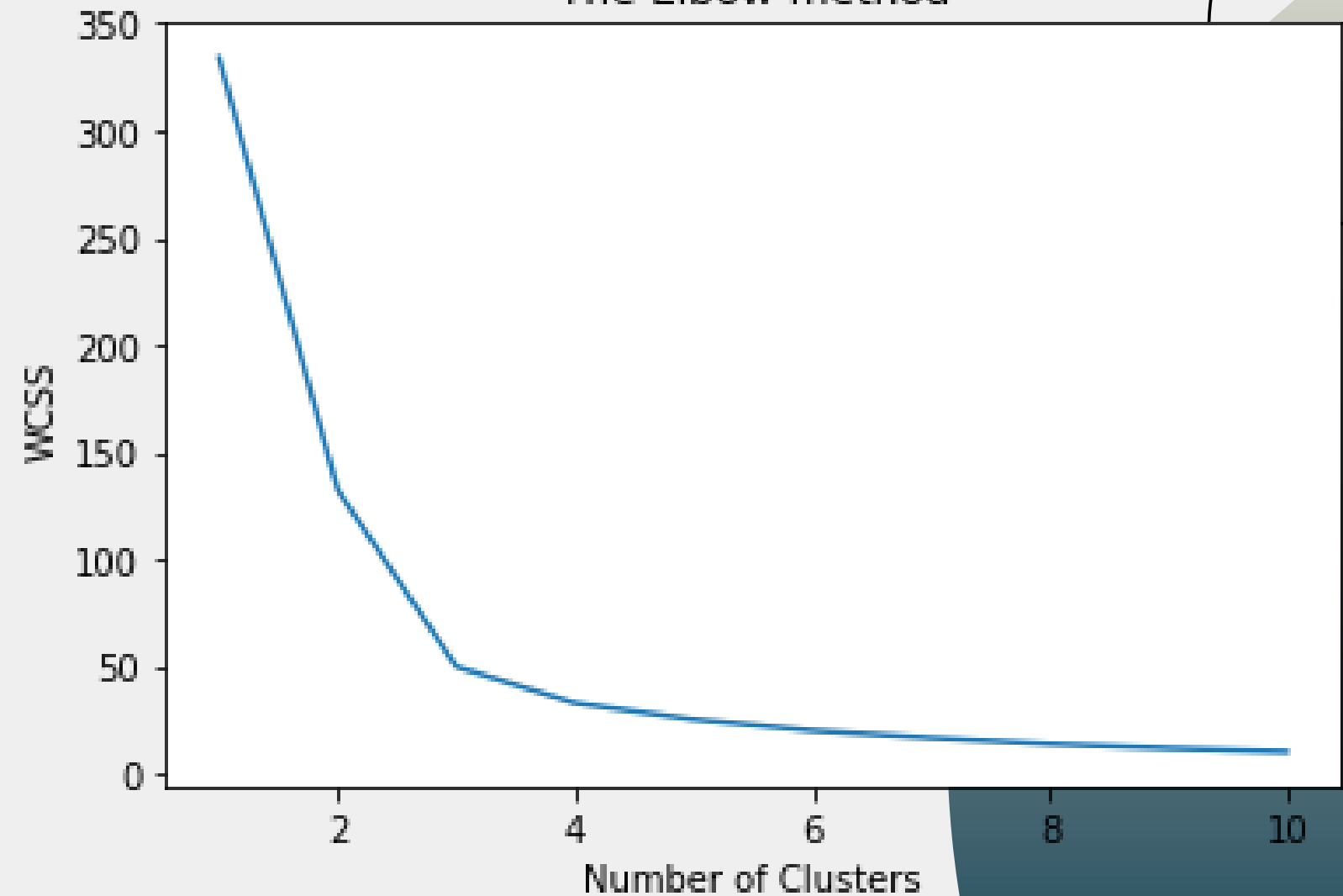
0.7038276968521354

# CLUSTERING



🔍 Dari analisis bivariate antara GDP dengan harapan hidup dapat dibagi menjadi tiga cluster (bawah, menengah, dan atas).

The Elbow method

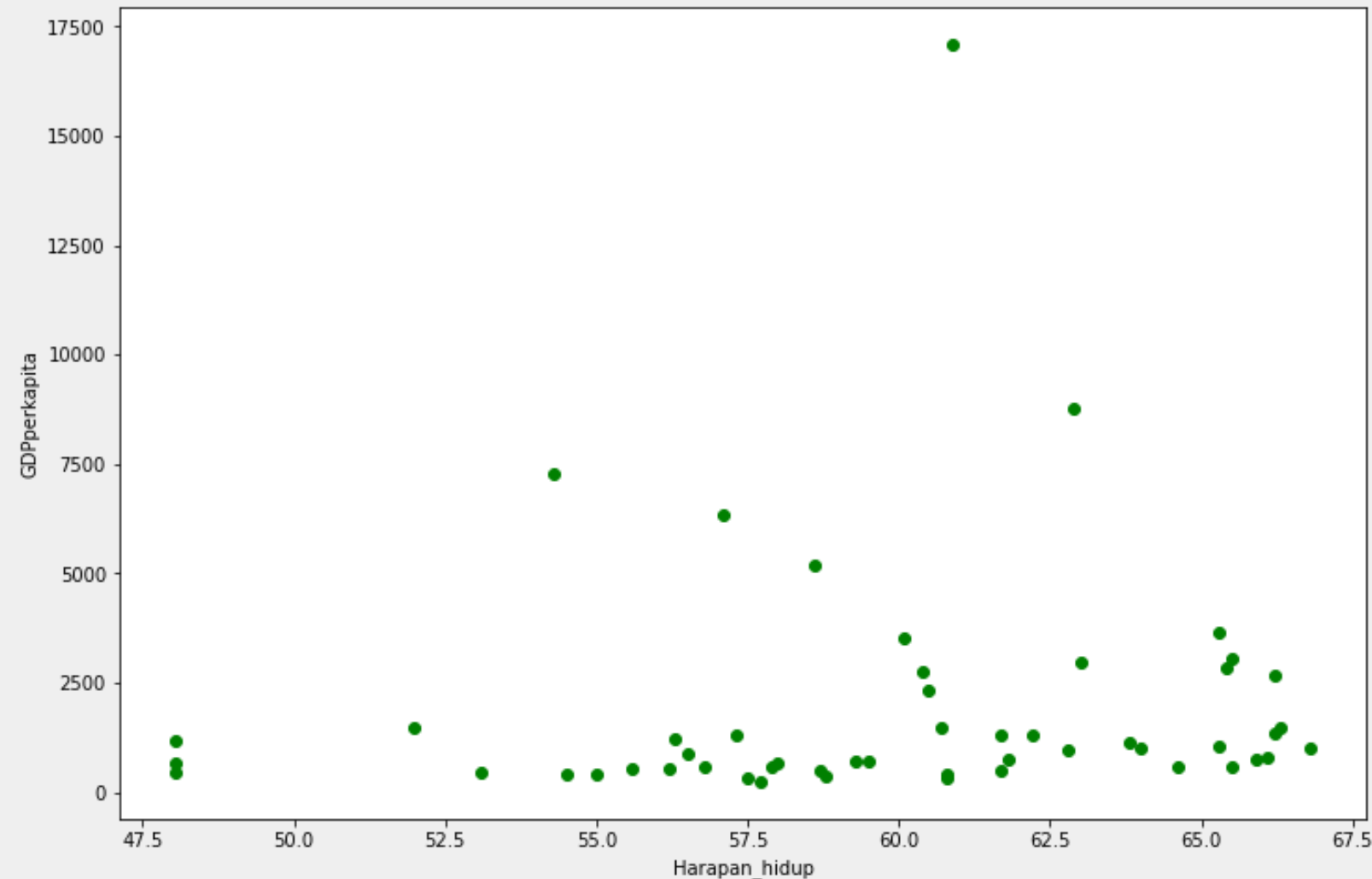


🔍 Banyak cluster didapatkan dari elbow method dan dicek kembali dengan silhouette score method

# RECOMMEN- DATION

Cluster yang digunakan sebagai pertimbangan adalah lower cluster dimana harapan hidup dan GDP keduanya rendah

Untuk Negara yang masuk dalam lower cluster adalah

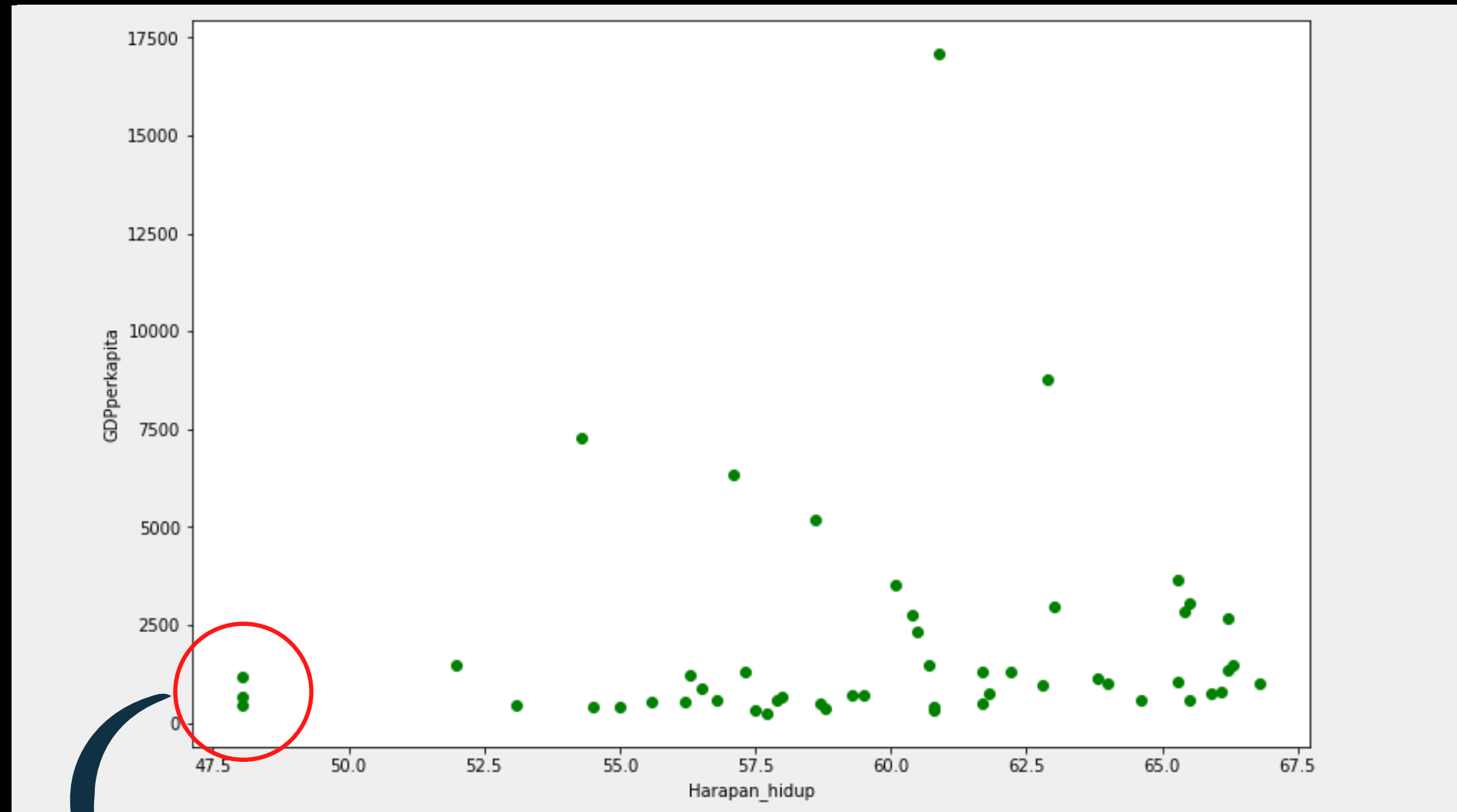


Negara klasifikasi lower cluster

```
['Afghanistan', 'Angola', 'Benin', 'Botswana', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon', 'Central African Republic', 'Chad', 'Comoros', 'Congo, Dem. Rep.', 'Congo, Rep.', 'Cote d'Ivoire', 'Equatorial Guinea', 'Eritrea', 'Fiji', 'Gabon', 'Gambia', 'Ghana', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti', 'India', 'Kenya', 'Kiribati', 'Lao', 'Lesotho', 'Liberia', 'Madagascar', 'Malawi', 'Mali', 'Micronesia, Fed. Sts.', 'Mongolia', 'Mozambique', 'Myanmar', 'Namibia', 'Niger', 'Nigeria', 'Pakistan', 'Rwanda', 'Senegal', 'Sierra Leone', 'Solomon Islands', 'South Africa', 'Sudan', 'Tanzania', 'Togo', 'Uganda', 'Vanuatu', 'Zambia']
```



# RECOMMENDATION



**NEED HELP**

Terdapat tiga negara yang paling membutuhkan bantuan. Yaitu, **Central African Republic, Haiti, dan lesotho**

Hal ini dapat terlihat dari grafik bivariate antara harapan hidup dan GDP pada cluster bawah yang menunjukkan terdapat tiga negara yang berada jauh dibawah kiri dibandingkan negara lainnya