

06211940000059

BRAMANDIKA

EURUSD (Bid), Ticks, # 300 / 300

23:35 23:40 23:45

Gold, spot - 1.276,820 - 23:00:00 13 giu (CEST)
Gold, spot (Bid), 1 minute, # 159 / 300, Logarithmic, Heiken Ashi



LATAR BELAKANG

Ada banyak sekali kekurangan yang dialami oleh berbagai negara, seperti kekurangan finansial, kekurangan pelayanan kesehatan, dll. Hal ini tentu akan memengaruhi pertumbuhan ekonomi suatu negara, salah satunya GDP. GDP adalah salah satu indikator penting untuk mengetahui kondisi ekonomi di suatu negara dalam suatu periode tertentu. Untuk itu, kita melakukan sebuah penelitian dengan membagi seluruh negara menjadi 3 cluster dengan melihat setiap kekurangan yang ada di negara tersebut.

“

TUJUAN PENELITIAN

TUJUAN DARI KITA MELAKUKAN PENELITIAN
INI ADALAH UNTUK MENGETAHUI CLUSTER
MANA YANG MENGALAMI KEKURANGAN DI
BERBAGAI BIDANG YANG SIGNIFIKAN DAN
NEGARA TERSEBUT MEMBUTUHKAN
BANTUAN DARI KEKURANGAN TERSEBUT.

Keterangan Data

Data berasal dari Data_Negara_HELP.csv terdiri atas 10 kolom dan 167 row dengan rincian sebagai berikut:

- 1 Negara 167 non-null object
- 2 Kematian_anak 167 non-null float64
- 3 Ekspor 167 non-null float64
- 4 Kesehatan 167 non-null float64
- 5 Impor 167 non-null float64
- 6 Pendapatan 167 non-null int64
- 7 Inflasi 167 non-null float64
- 8 Harapan_hidup 167 non-null float64
- 9 Jumlah_fertiliti 167 non-null float64
- 10 GDPperkapita 167 non-null int64

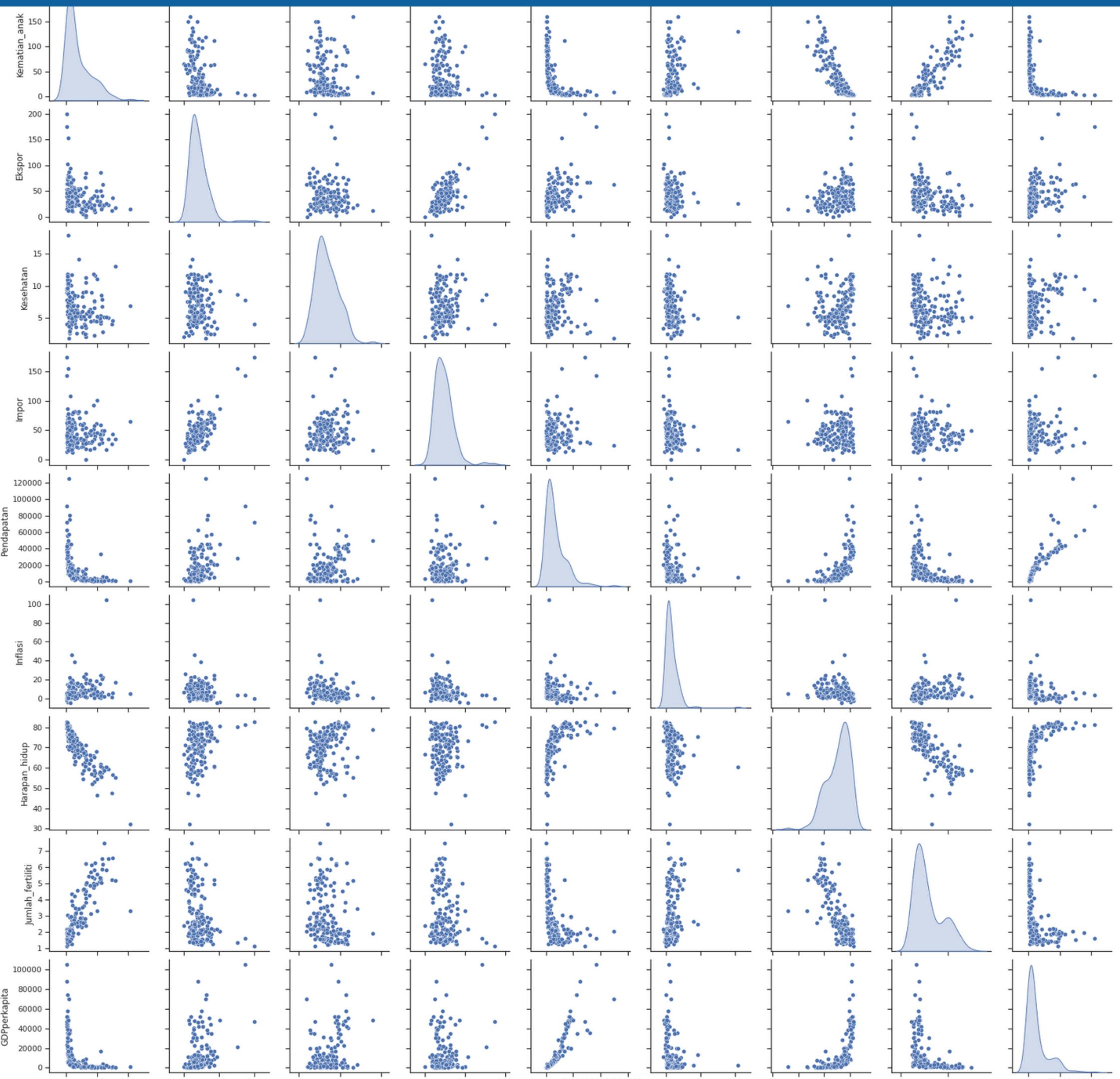
Keterangan Variabel

- 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.

Pre- processing data

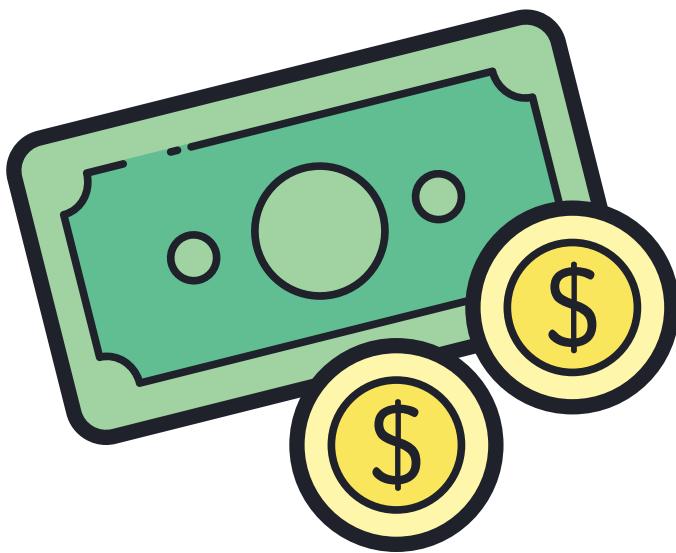
- reading data
- memeriksa outlier dan missing value
- menghapus outlier dari setiap kolom kecuali gdp per kapita dan pendapatan karena dianggap dapat mempengaruhi terhadap cluster
- membuat pair plot
- melakukan cluster terhadap negara-negara menggunakan k-means 2 parameter yaitu gdp per kapita dan pendapata
- mencari insight tiap cluster melalui visualisasi data

PAIR PLOT



INTERPRETASI

- Hampir semuanya memiliki skew yang positif kecuali untuk variabel angka harapan hidup
- Hanya variabel pendapatan dan gdp per kapita yang memiliki hubungan linier selain itu memiliki pola yang acak

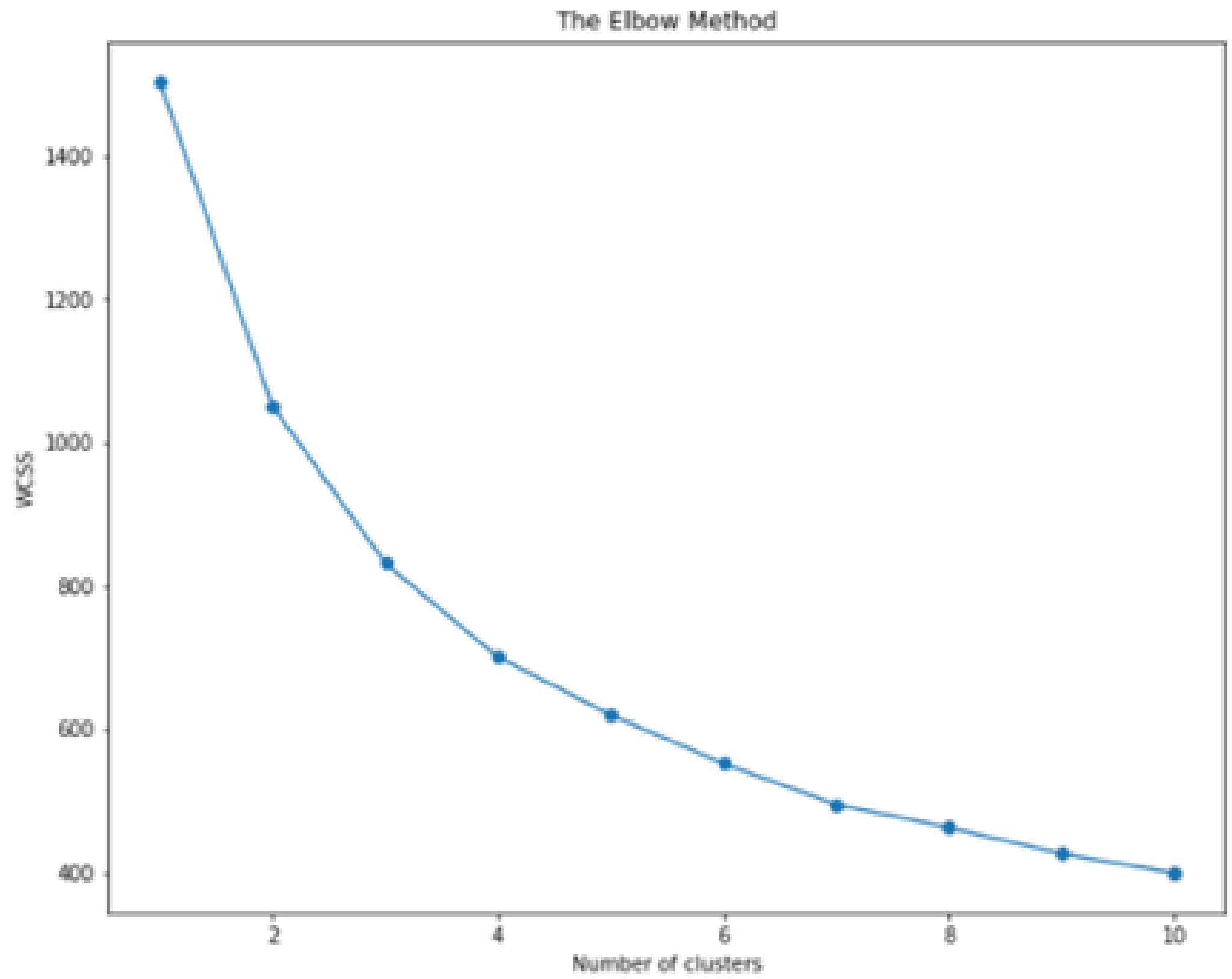


**Mengcluster Menggunakan GDP
perkapita dan pendapatan
sebagai parameternya**

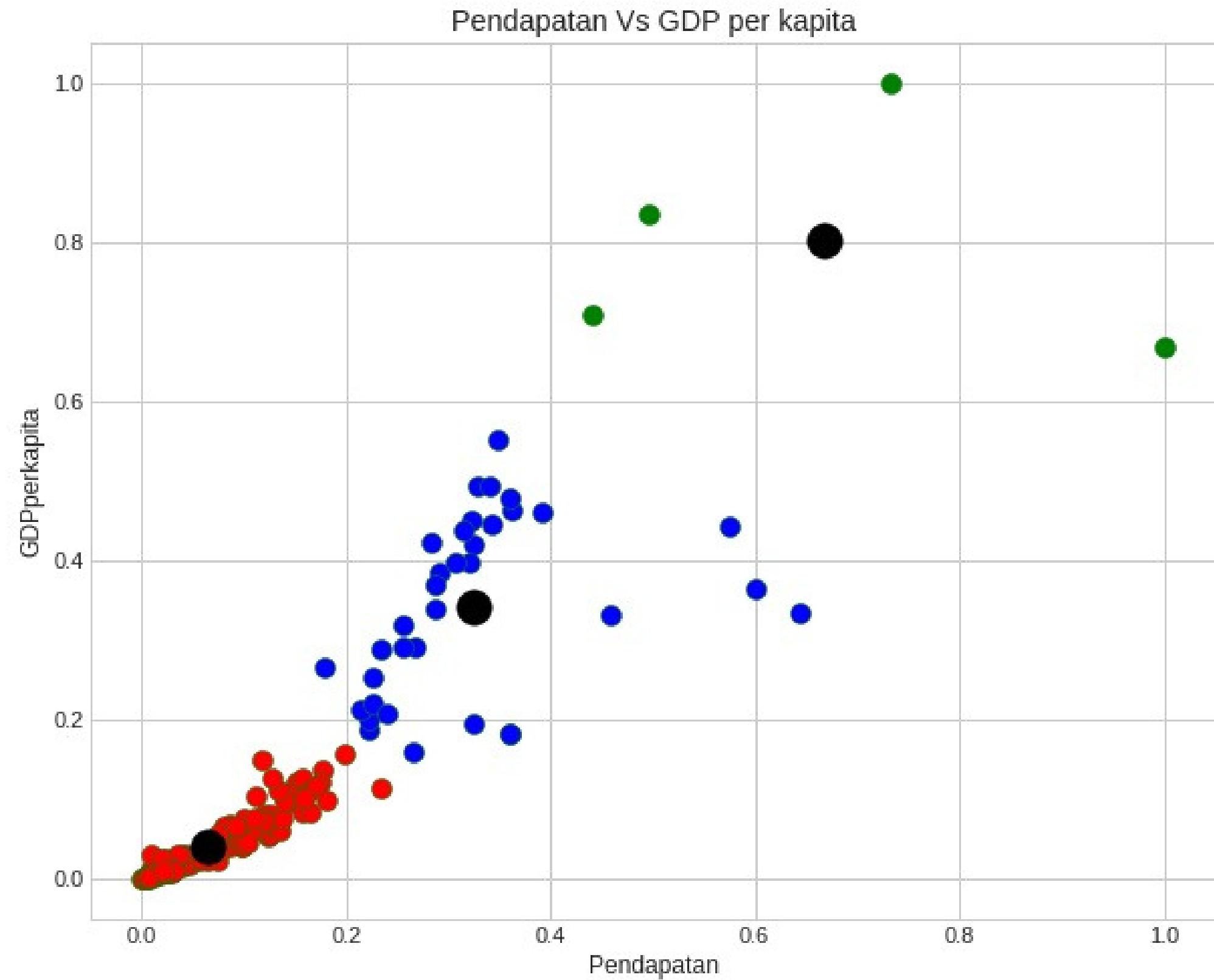
MENGAPA GDP DAN PENDAPATAN

Karena dari Pair plot kita mengatahui bahwa dia berbentuk linear dan mengarah ke kanan atas yang berarti memiliki korelasi positif yang kuat. Selain itu, GDP dan Pendapatan menurut penelitian juga merupakan salah satu faktor penentu negara berkembang dan negara maju dari segi ekonominya sehingga kita pada akhirnya dapat menentuan target yang tepat untuk diberikan bantuan.

ELBOW METHOD



SCATTER PLOT

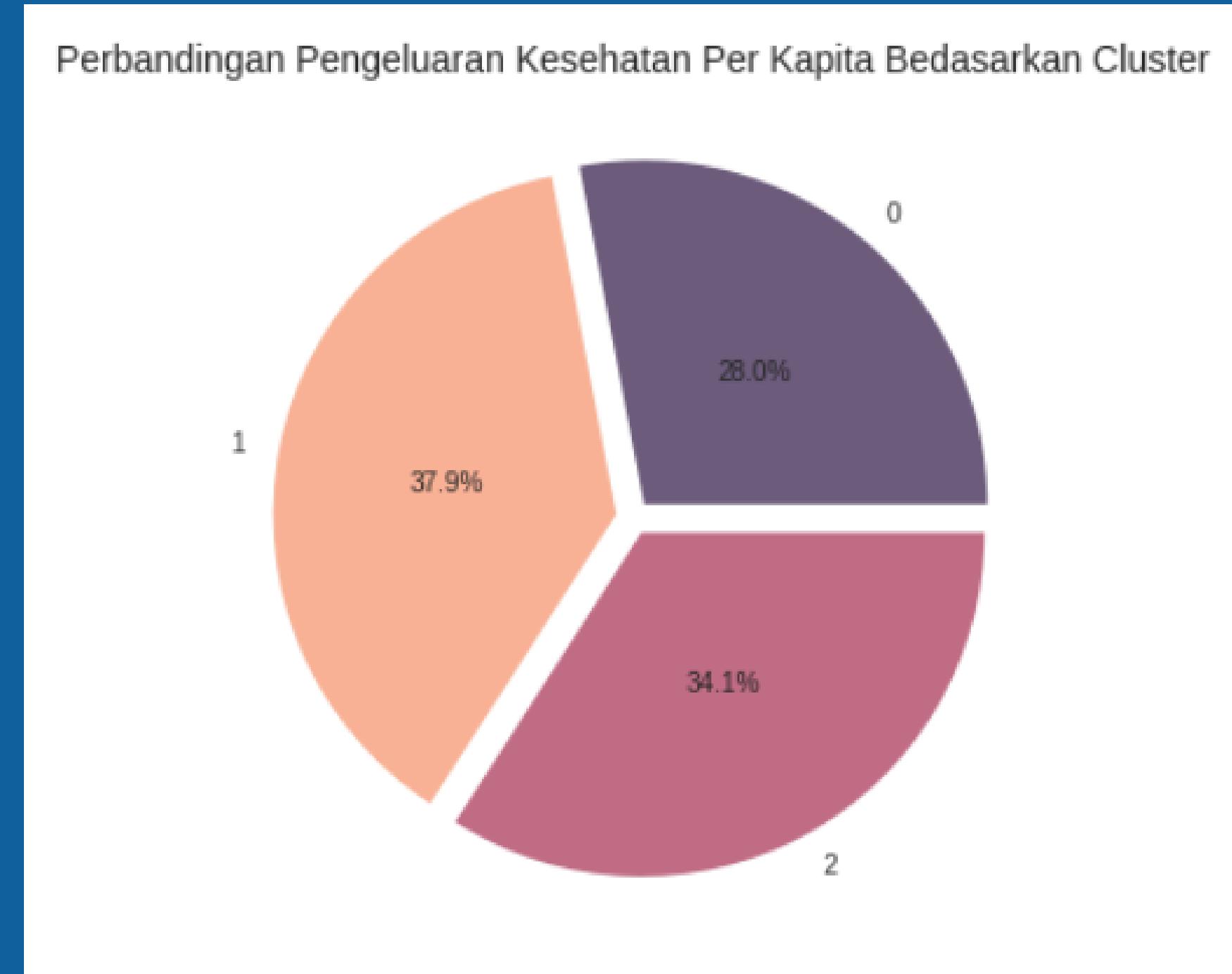


INTERPRETASI

Dengan Elbow Method kita dapat melihat jumlah kluster yang dibutuhkan dalam hal ini berjumlah 3.

Selanjutnya menggunakan metode K-means untuk mengcluster GDP dan pendapatan didapatkan secara visual: cluster 0 ditandai dengan warna merah yang merupakan negara terbelakang, cluster 1 ditandai dengan warna biru yang dimana merupakan negara berkembang dan cluster dua ditandai dengan warna hijau yang merupakan negara maju. Dengan menggunakan silhouette score didapatkan bahwa pembagian kluster ini sudah akurat sebesar 89,84%

PIE CHART



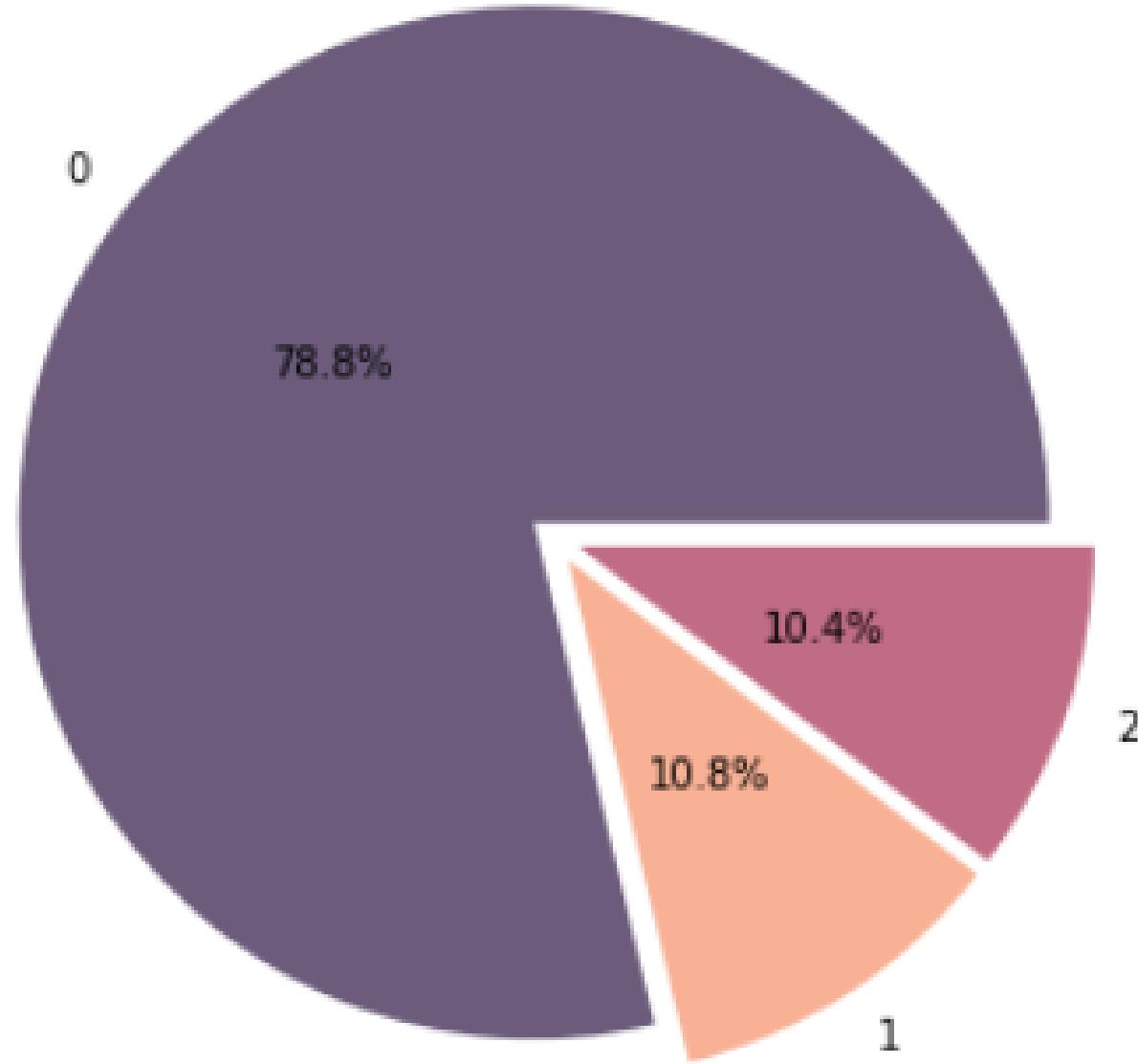
ket: Untuk piechart nilai persen untuk cluster berasal dari total rata-rata variabelnya

INTERPRETASI

Hasil interpretasi dari gambar pertama adalah pengeluaran kesehatan per kapita berdasarkan cluster yang berbeda. Dalam hal ini, di cluster 0 memiliki pengeluaran kesehatan perkapita sebesar 28% dari total rata-rata, untuk negara berkembang sebesar 37,9%, dan untuk negara maju sebesar 34,1%. Artinya, dalam hal ini, pengeluaran kesehatan negara berkembang lebih besar dibanding negara terbelakang dan maju. Negara berkembang, membutuhkan bantuan pelayanan kesehatan lebih banyak karena pengeluaran yang dikeluarkan pun lebih tinggi dibanding yang lain.

PIE CHART

Kematian Anak bedasarkan cluster



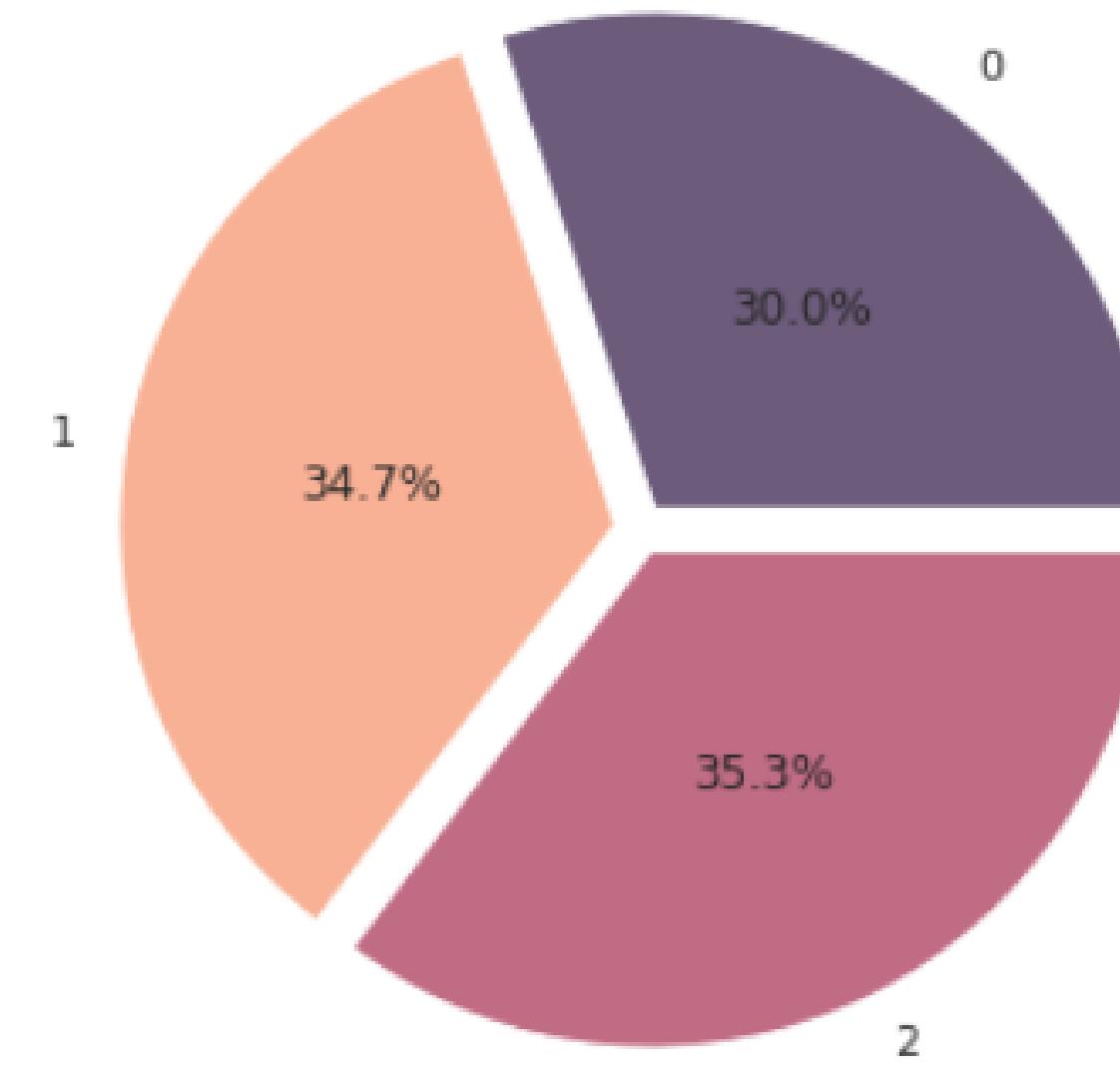
ket: Untuk piechart nilai persen untuk cluster berasal dari total rata-rata variabelnya

INTERPRETASI

HASIL INTERPRETASI DARI GAMBAR KEDUA ADALAH KEMATIAN ANAK BERDASARKAN CLUSTER YANG BERBEDA. DALAM HAL INI, UNTUK NEGARA TERBELAKANG MEMILIKI TINGKAT KEMATIAN ANAK SEBESAR 78,8% DARI TOTAL RATA-RATA, UNTUK NEGARA BERKEMBANG SEBESAR 10,8%, DAN UNTUK NEGARA MAJU SEBESAR 10,4%. ARTINYA, DALAM HAL INI, TINGKAT KEMATIAN ANAK NEGARA TERBELAKANG LEBIH BESAR DIBANDING NEGARA BERKEMBANG DAN MAJU. NEGARA TERBELAKANG, MEMBUTUHKAN BANTUAN MENGENAI PELAYANAN KESEHATANBAIK FISIK MAUPUN MENTAL UNTUK ANAK USIA DINI ATAU HARUS DITELITI LEBIH BANYAK KARENA TINGKAT KEMATIAN ANAK PUN LEBIH TINGGI DIBANDING YANG LAIN.

PIE CHART

Perbandingan Angka Harapan hidup tiap clusternya



ket: Untuk piechart nilai persen untuk cluster berasal dari total rata-rata variabelnya



INTERPRETASI

Hasil interpretasi dari gambar pertama adalah pangka harapan hidup berdasarkan cluster yang berbeda. Dalam hal ini, di cluster 0 memiliki angka harapan hidup sebesar 30% dari total rata-rata, untuk negara berkembang sebesar 34,7%, dan untuk negara maju sebesar 35,3%. Artinya, dalam hal ini, angka harapan hidup negara maju lebih besar dibanding negara terbelakang dan berkembang. Artinya, di negara maju, jumlah tahun rata- rata seorang anak yang baru lahir akan hidup jika pola kematian saat ini tetap sama berada pada persentase 35,5%.

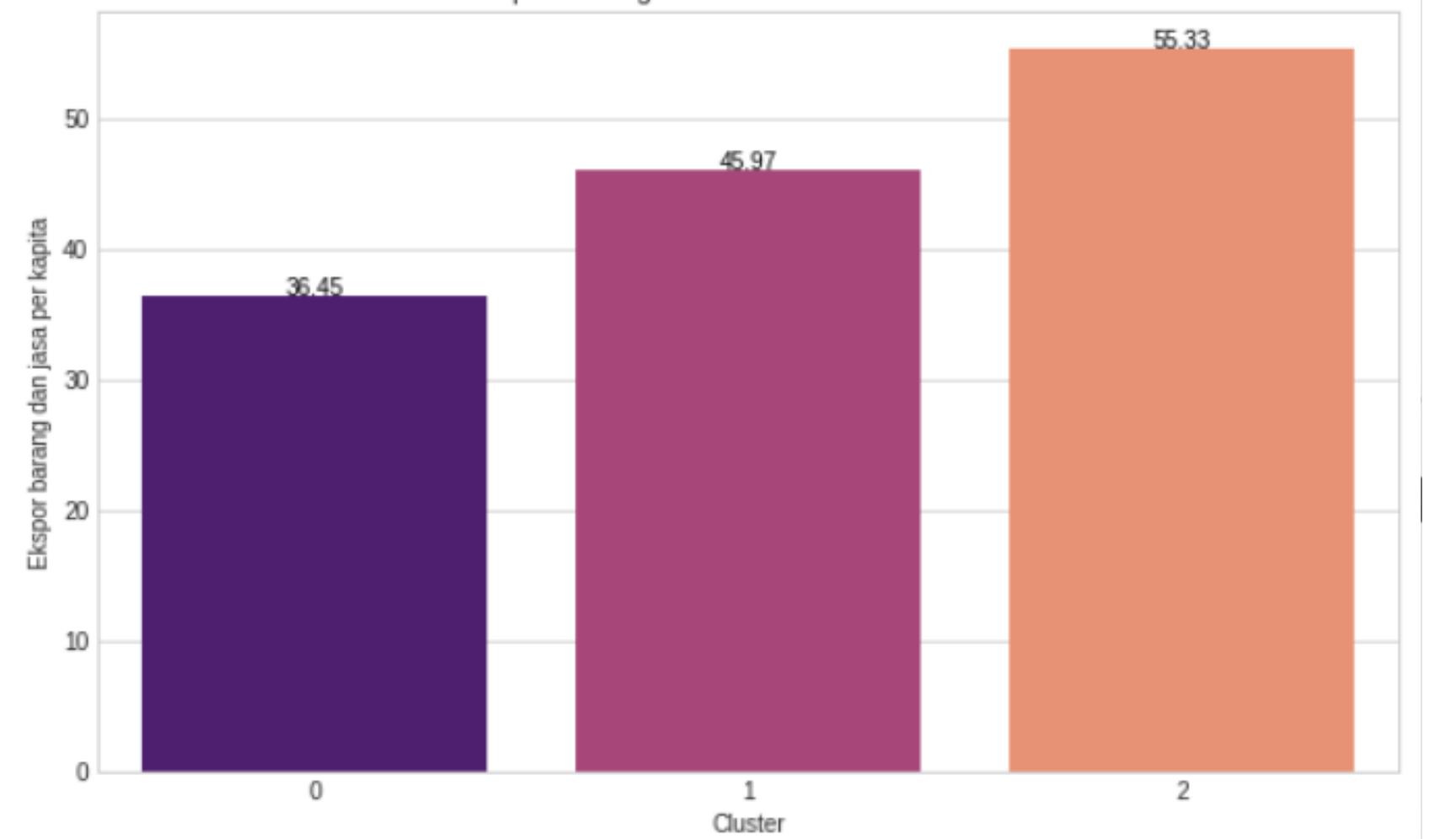
HASIL DATA

C	Kematian_anak	Eksport	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
Cluster									
0	47.389844	35.865617	6.332031	45.921609	8569.242188	8.957648	67.873437	3.263359	4438.390625
1	8.734286	55.240000	8.490286	48.717143	40908.571429	3.889143	79.171429	1.926571	35977.142857
2	4.875000	85.250000	7.640000	61.900000	83625.000000	4.216750	81.000000	1.792500	84425.000000

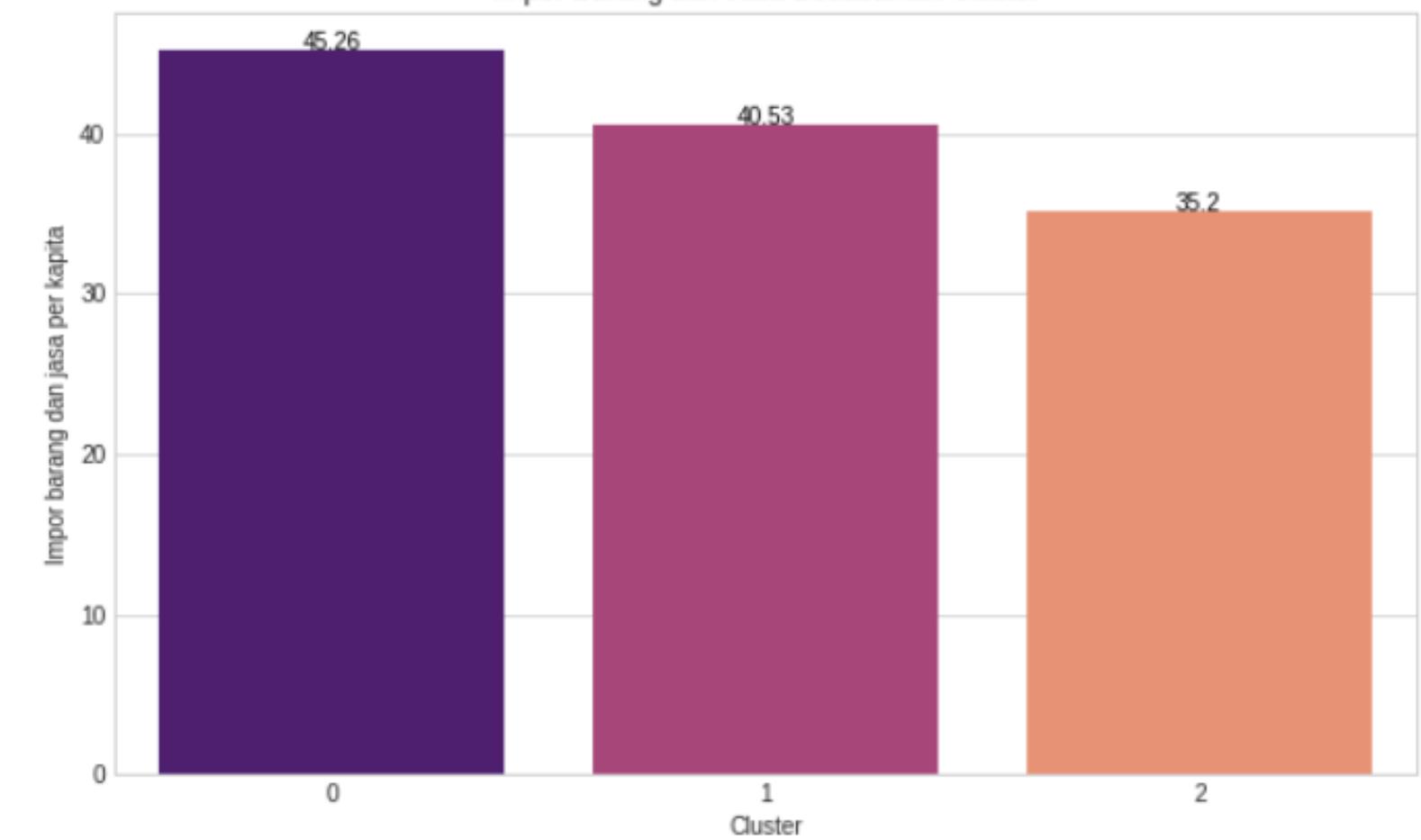
BAR PLOT



Ekspor Barang dan Jasa Bedasarkan Cluster



Impor Barang dan Jasa Bedasarkan Cluster



<Figure size 432x288 with 0 Axes>

INTERPRETASI

Dari barplot tersebut didapatkan bahwa secara eksport seakin tinggi nilai pendapatan dan gdpnya maka nilai dari eksport tersebut semakin besar sedangkan untuk impor berbanding terbalik, yang artinya empat variabel tersebut memiliki hubungan yang dimana enurut penelitian ke 3 variabel tersebut mempengaruhi gdp perkapita dengan rumus sebagai berikut:

Komponen-komponen dari PDB (Mankiw:2003) adalah:

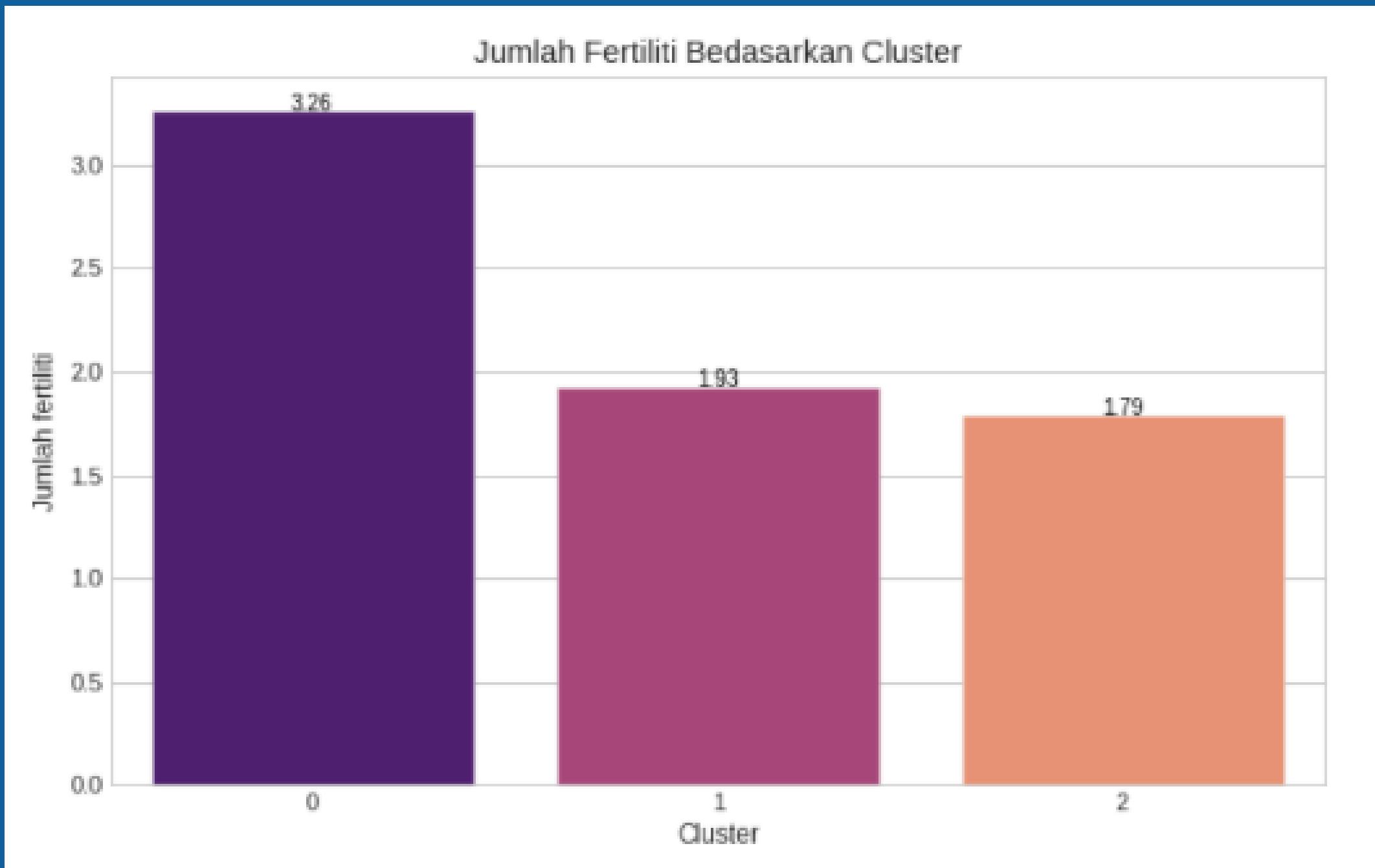
- A. konsumsi
- B. Investasi/tabungan
- C. Pengeluaran pemerintah
- D. Ekspor-impor (NX)

Sehingga dapat ditulis sebagai berikut:

$$PDB = C + I + G + NX \dots \dots \dots \quad (2)$$

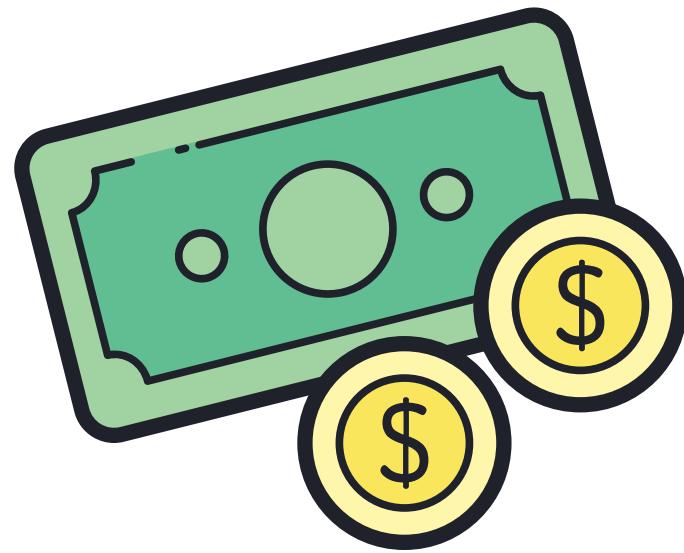


BAR PLOT



INTERPRETASI

JUMLAH ANAK YANG AKAN LAHIR APABILA TINGKAT KESUBURAN WANITA USIA SAATINI SAMA MAKA CLUSTER 0 AKAN MEMILIKI JUMLAH ANAK TERBANYAK KEMUDIAN DISUSUL OLEH CLUSTER 1 DAN 2



Clustering Menggunakan Seluruh Variabel Sebagai Parameternya

RADAR PLOT



INTERPRETASI

TERDAPAT PERBEDAAN YANG SANGAT JAUH APABILA DIBANDINGKAN DENGAN SAAT ENGGUNAKAN DUA PARAETER YANG DIMANA SEBELUMNYA CLUSTER O YANG DILABEL SEBAGAI NEAGARA TERBELAKANG MALAH MENJADI NEGARA MAJU TETAPI NEGARA YANG MAJU SEHARUSNYA MEMILIKI ANGKA INFLASI RENDAH TETAPI DISINI MEMILIKI ANGKA INFLASI YANG TINGGI MMAKA DARI ITU MENGGUNAKAN SELURUH VARIABEL SEBAGAI PARAETER UNTUK CLUSTERING DINGGAP TIDAK REPRESENTATIF DAN BISA MENIMBULKAN SALAH SASARAN TERHADAP BANTUAN

KESIMPULAN

- Dengan menggunakan elbow method diapatkan jumlah cluster yang sesuai adalah 3
- Dengan menggunakan metode K-means didapatkan 3 cluster tersebut yang masing-masing clusternya terdiri sebagai berikut: cluster 0 yang dimana dalam penelitian ini saya beri label ya itu sebagai negara terbelakang terdiri atas 116 negara, untuk cluster 1 diberi label negara berkembang dan terdiri atas 30 negara.Untuk cluster 2 diberi label sebagai negara maju dan terdiri atas 3 negara untuk rincian negaranya dapat dilihat dalam google colab
- Adapun akurasi dalam pengclusteran ini sebesar 89,84% yang artinya model sudah cukup baik

KESIMPULAN

- Darii visualisasi barplot dan pie chart sebelumnya diapatkan bahwa cluster 0 yang merupakan negara terbelakang yang paling membutuhkan bantuan secara segi ekonomi maupun kesehatan
- Kita tidak bisa menggunakan semua parameter sebagai cluster dikarenakan variabel-variabel yang memiliki korelasi yang tidak kuat dapat menimbulkan hasil yang jurang serta jawaban yang tidak logis

SARAN

- Dalam penelitian ini saya masih kurang mengetahui apakah outlier harus dibuang atau dibiarkan dikarenakan dalam hal ini saya membuangnya, meskioun tidak mempengaruhi hasil clustering sebagai contoh outlier yang memiliki angka kematian tinggi dihapus dari data sehingga yuang seharusnya sebelumnya ia dikategorikan sebagai cluster 0 menjadi tidak mendapat bantuan dana kesehatan sama sekali dikarenak tidak termasuk dalam data

,

