



CLUSTERING COUNTRIES USING K-MEANS

01

Angela Gloria Cahya Dewi





02

Table of Contents

Latar Belakang
Tujuan
Data Understanding
Exploratory Data Analysis
Outliers Treatment
K-means Clustering
Summary

...

03

Latar Belakang

HELP International

HELP International adalah LSM Kemanusiaan internasional yang berkomitmen memerangi kemiskinan bagi negara terbelakang dengan data terkumpul sebanyak \$ 10 juta.





04

Tujuan

HELP International

Mengkategorikan negara menggunakan faktor sosial ekonomi dan kesehatan yang menentukan pembangunan negara secara menyeluruh.



...

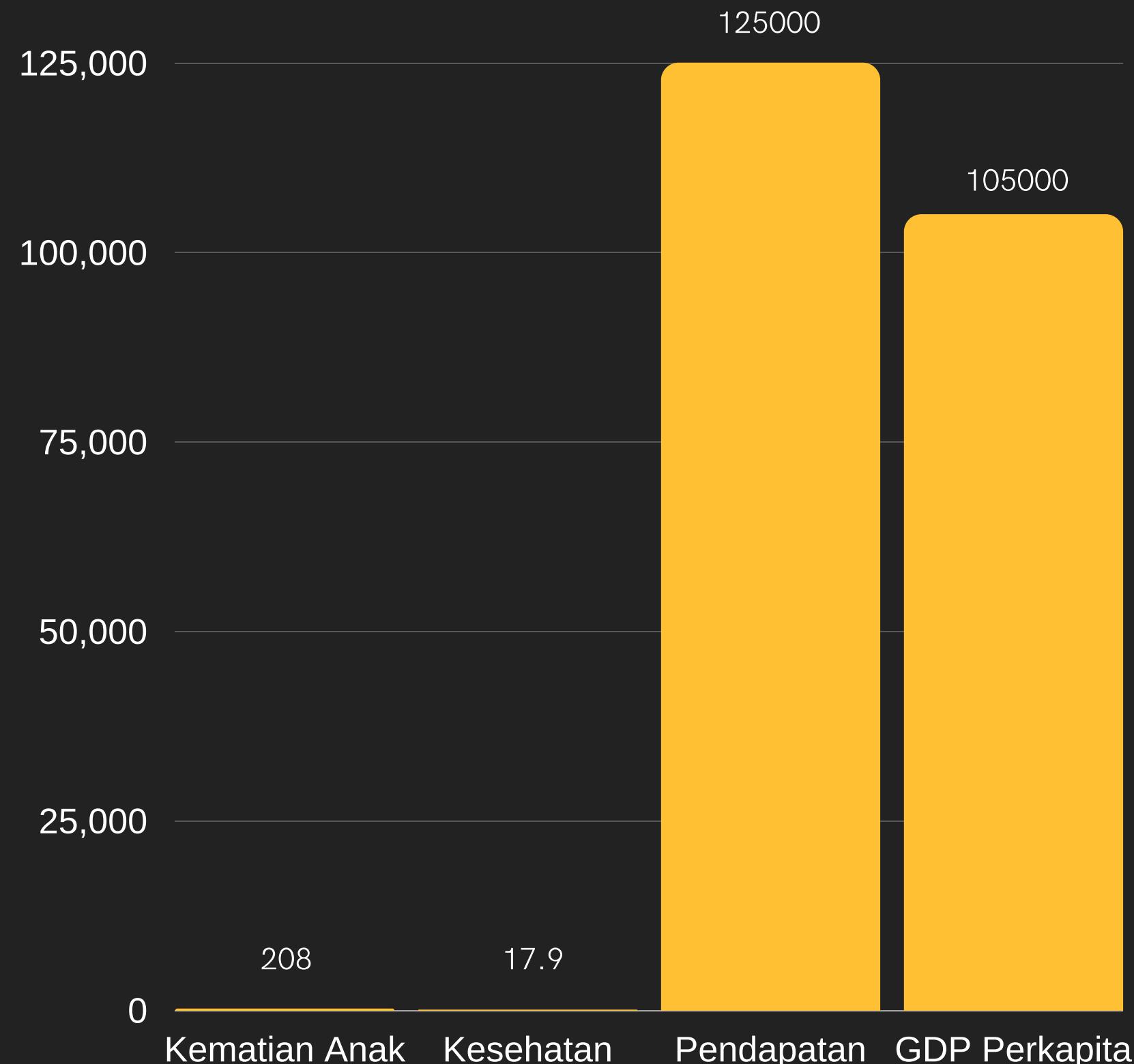


Data understanding

05

Terdapat 167 data dan 10 kolom, dimana terdapat kolom Negara, Kematian_anak, Ekspor, Kesehatan, Impor, Pendapatan, Inflasi, Harapan_hidup, Jumlah_fertiliti, dan GDPperkapita.

Data yang saya gunakan untuk menganalisis faktor kesehatan adalah kolom Kematian_anak dan kesehatan, sementara untuk faktor ekonomi adalah kolom Pendapatan dan GDPperkapita.



*berdasarkan summary data sebelum data diolah

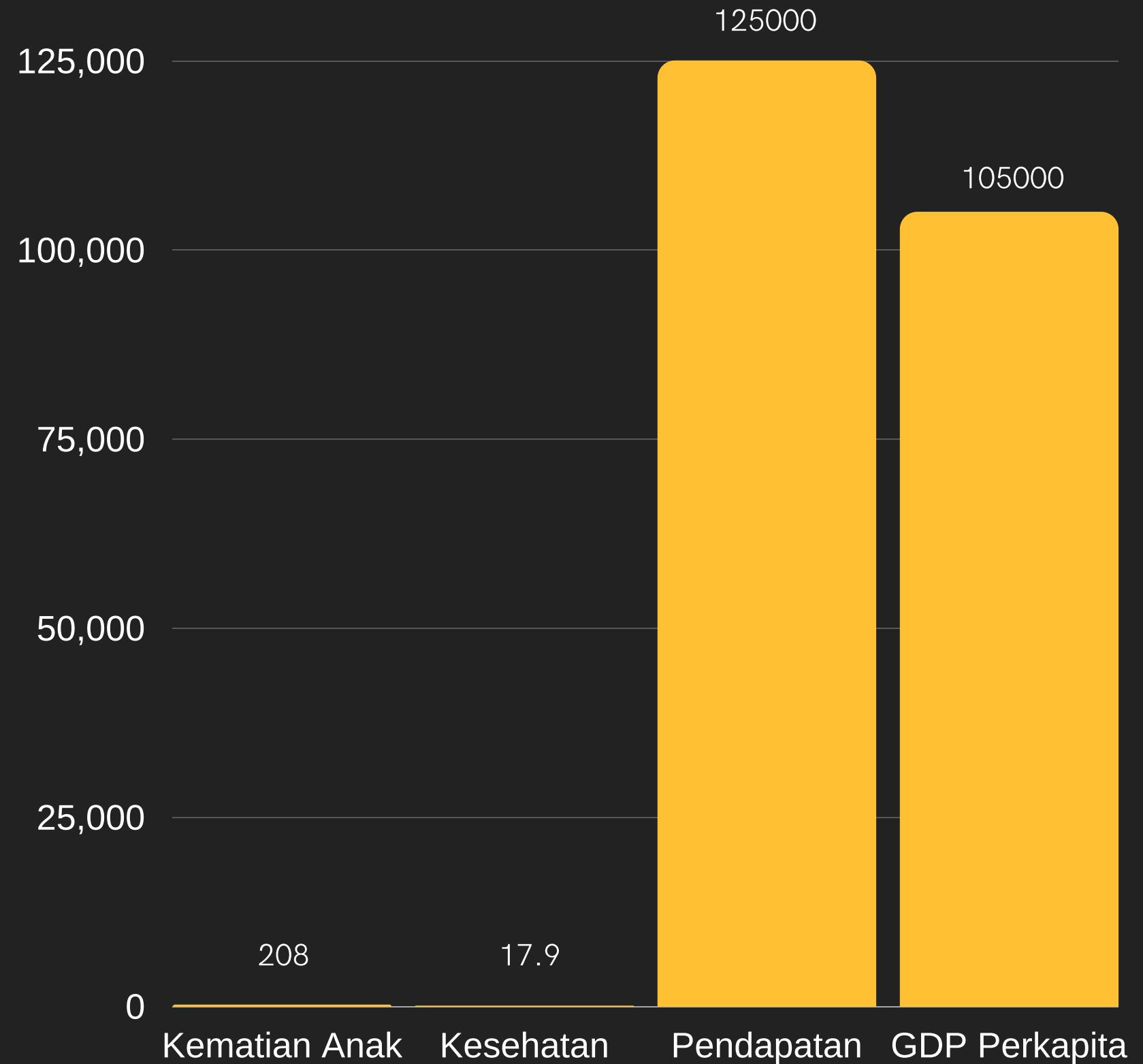


Data understanding

06

Pemilihan kolom faktor *kesehatan* didasari pada alasan bahwa semakin tinggi kematian anak, maka akan semakin rendah faktor kesehatan suatu negara. Hal itu dapat kita bandingkan dengan berapa jumlah dana kesehatan yang masuk.

Pemilihan kolom faktor *ekonomi* didasari pada alasan bahwa semakin rendah GDP (kualitas hidup), maka akan semakin rendah juga faktor ekonomi suatu negara. Hal itu dapat kita bandingkan dengan berapa jumlah pendapatan yang masuk.



*berdasarkan summary data sebelum data diolah

Hasil Deskriptif

21

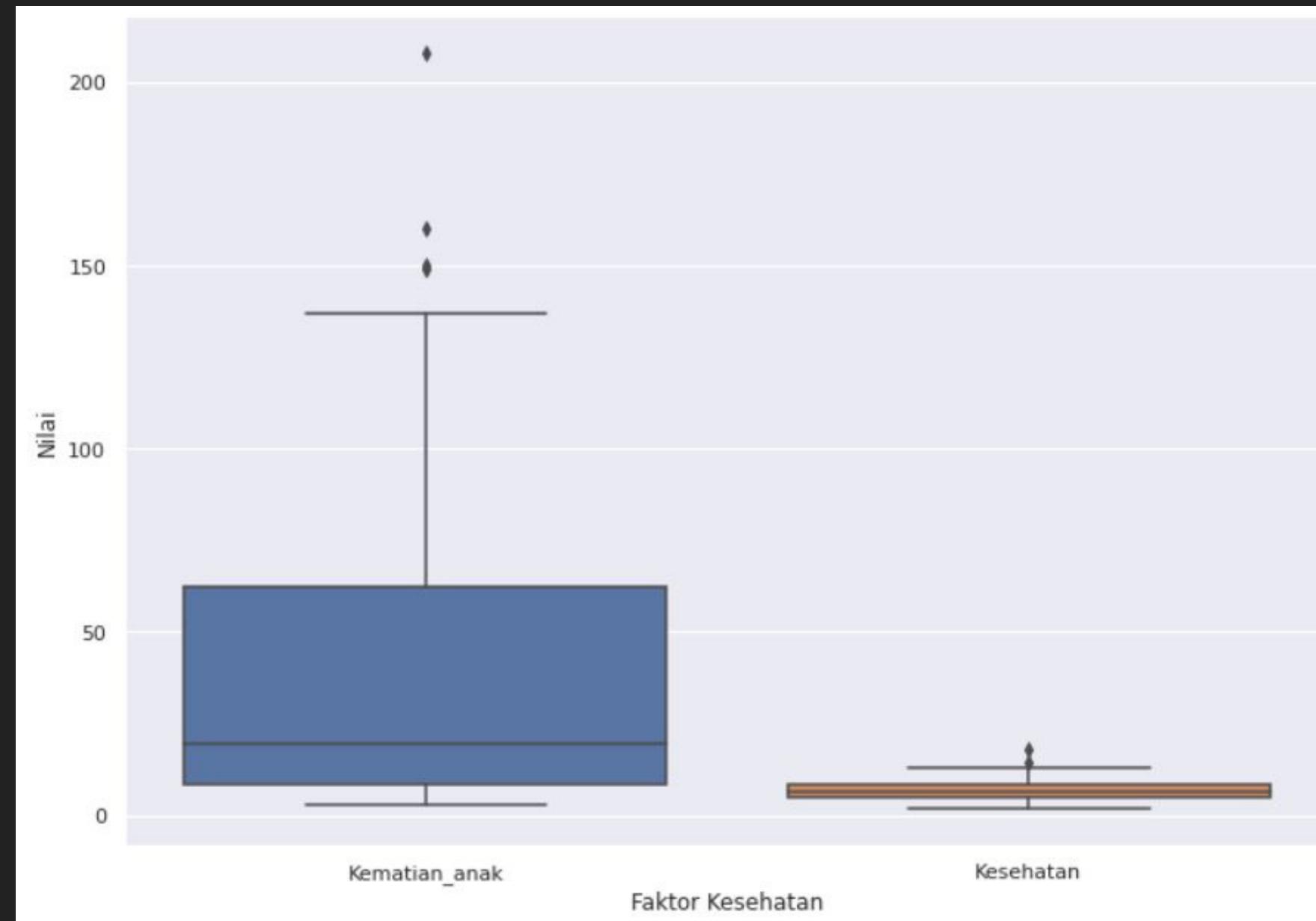
	Kematian_anak	Kesehatan	Pendapatan	GDPperkapita
count	167.000000	167.000000	167.000000	167.000000
mean	38.270060	6.815689	17144.688623	12964.155689
std	40.328931	2.746837	19278.067698	18328.704809
min	2.600000	1.810000	609.000000	231.000000
25%	8.250000	4.920000	3355.000000	1330.000000
50%	19.300000	6.320000	9960.000000	4660.000000
75%	62.100000	8.600000	22800.000000	14050.000000
max	208.000000	17.900000	125000.000000	105000.000000

Tabel disamping menunjukkan hasil statistik deskriptif dari data 167 negara terbalakang.



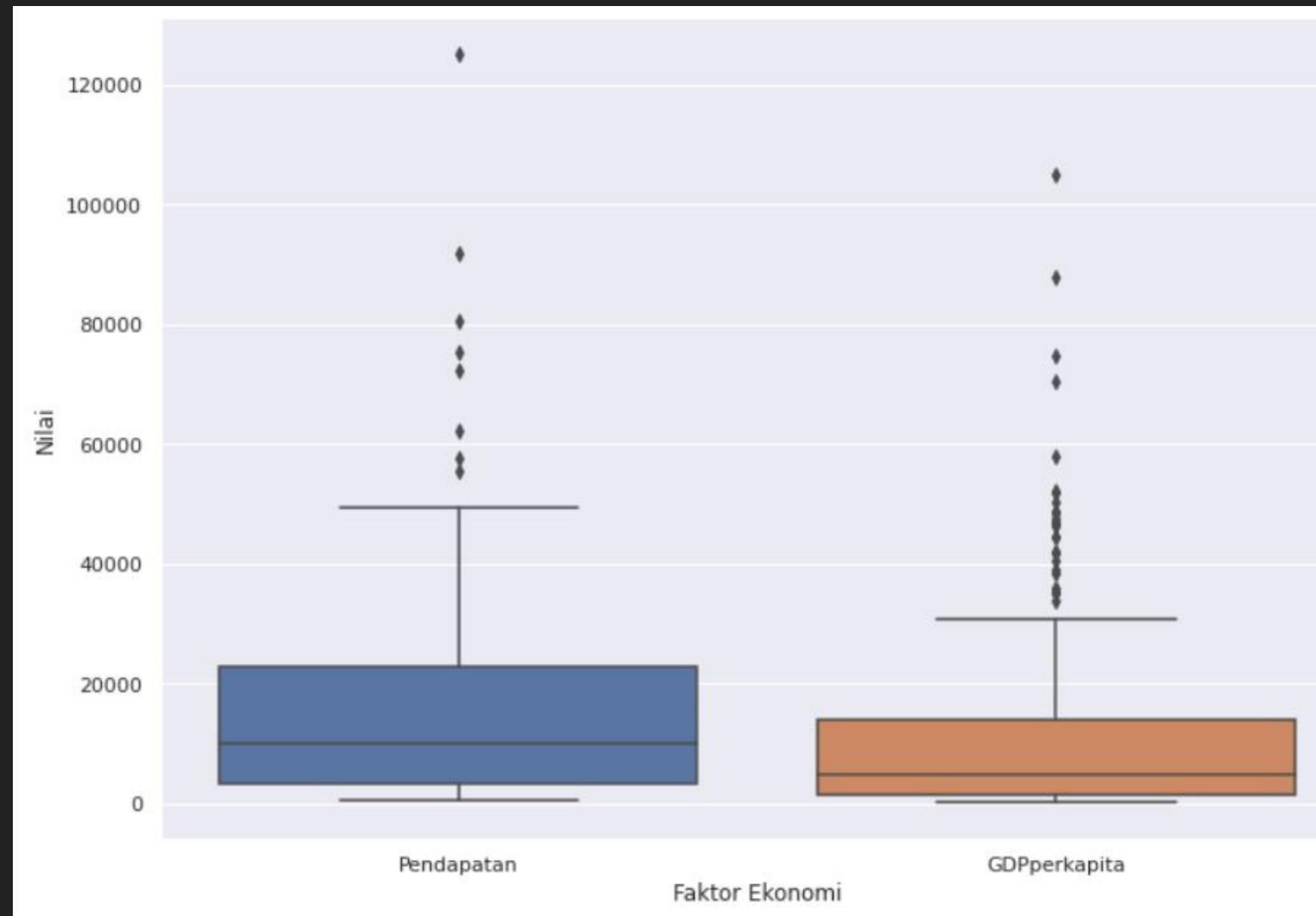


09



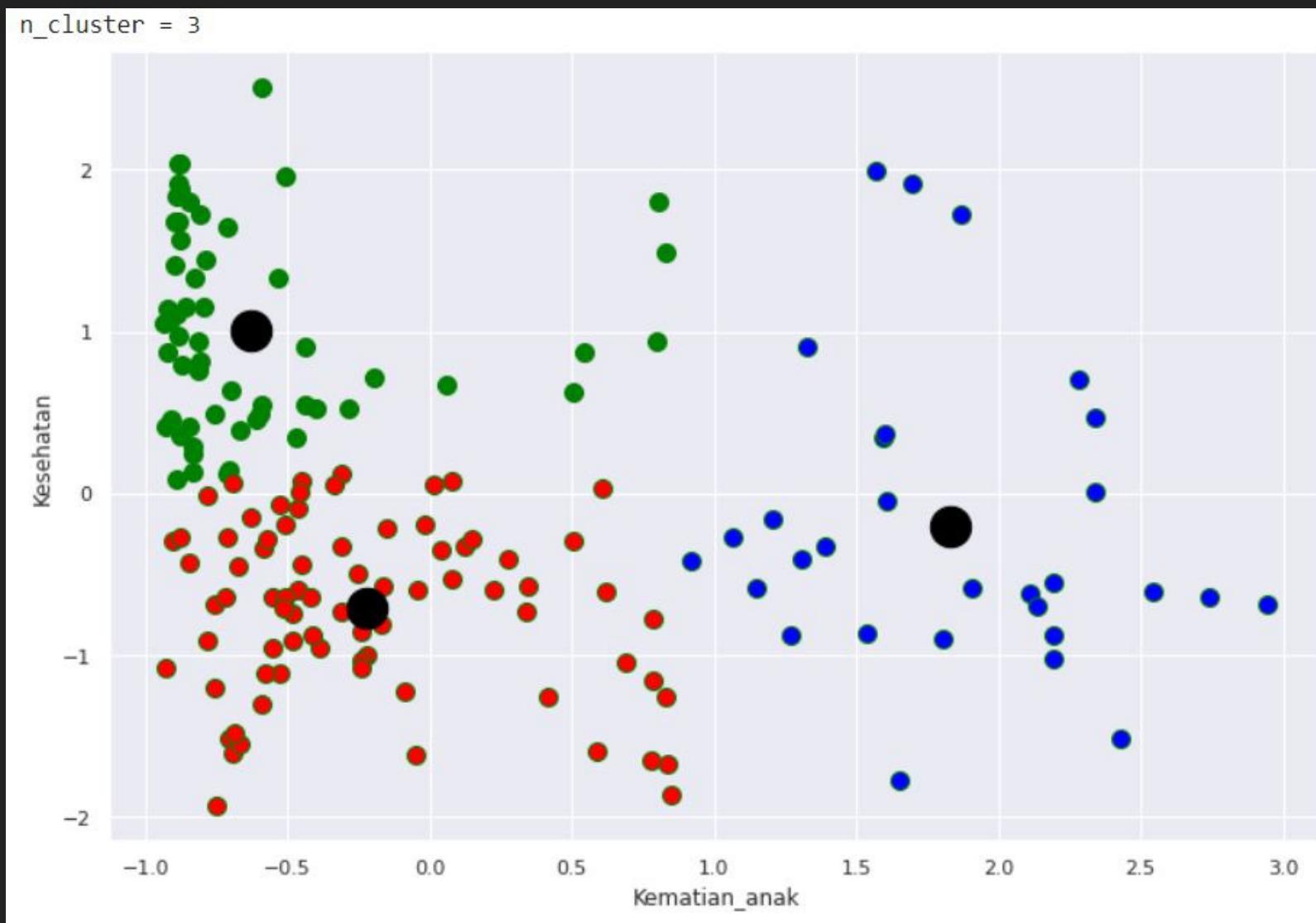
Boxplot Faktor Kesehatan

Terdapat beberapa data pencilan pada kolom Kematian_anak dan Kesehatan. Maka dari itu saya menghilangkan outlier dan melakukan data cleaning dengan mengisi nilai pada data sebelumnya



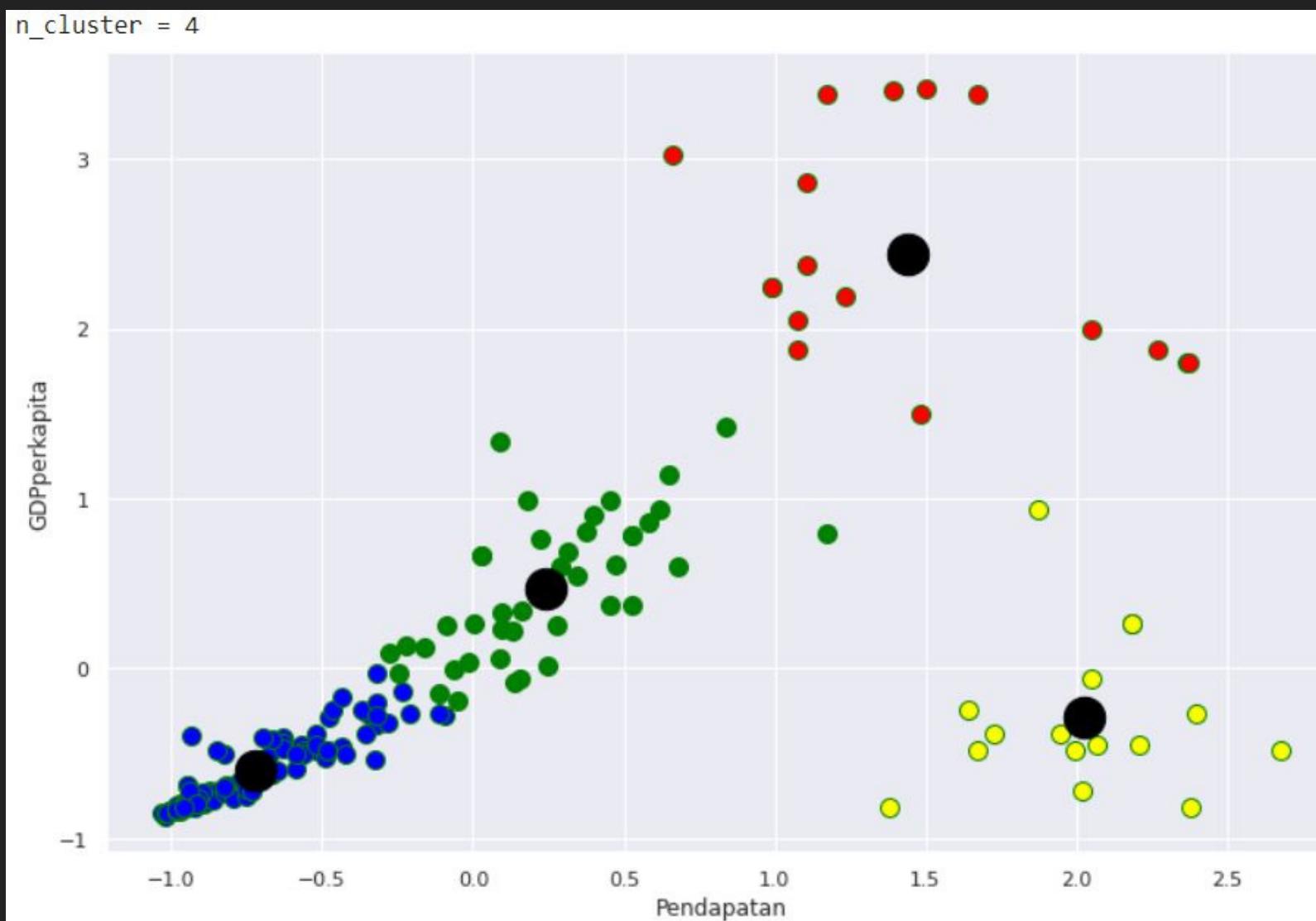
Boxplot Faktor Ekonomi

Terdapat beberapa data pencilan pada kolom Pendapatan dan GDPperkapita. Maka dari itu saya menghilangkan outlier dan melakukan data cleaning dengan mengisi nilai pada data sebelumnya.



Elbow Method Faktor Kesehatan

Setelah menghitung cluster dengan Elbow Method, saya mendapatkan bahwa cluster yang paling tepat adalah 3. Maka dari itu saya membuat 3 kategori data. Dimana data pada warna biru menunjukkan negara dengan kematian anak tinggi dan pemasukan dana kesehatan rendah.



Elbow Method

Faktor Ekonomi

Setelah menghitung cluster dengan Elbow Method, saya mendapatkan bahwa cluster yang paling tepat adalah 4. Maka dari itu saya membuat 4 kategori data. Dimana data pada warna biru menunjukkan negara dengan GDP (kualitas hidup) rendah dan pendapatan rendah.



Solusi

14

1

Melakukan pengelompokan clustering map dengan index 0 sebagai negara-negara yang tinggi akan kematian anak dan memiliki GDP rendah

2

Melakukan sort data berdasarkan kematian anak dengan tipe ascending dan GDP perkapita dengan tipe descending

3

Melakukan penggabungan data yang memiliki index 0 (memenuhi syarat) dengan memilih negara yang ada di kedua tabel.

4

Mendapatkan hasil negara-negara terbelakang yang membutuhkan bantuan dana dari HELP International

...



Summary

19

Berdasarkan hasil yang diperoleh, terdapat beberapa negara terbelakang, yaitu Nigeria, Lao, Timor-Leste, dan Eritrea. Maka dari itu HELP International akan membantu negara-negara tersebut dengan masing-masing mendapatkan \$2.5 juta.



25

Terima
Kasih



...