

HELP INTERNATIONAL

Mochamad Rizal Prasetyo

Table of Contents

01 **Alur Pengerjaan**

How we solve the problem

02 **Data Exploration**

Understanding the current data situation

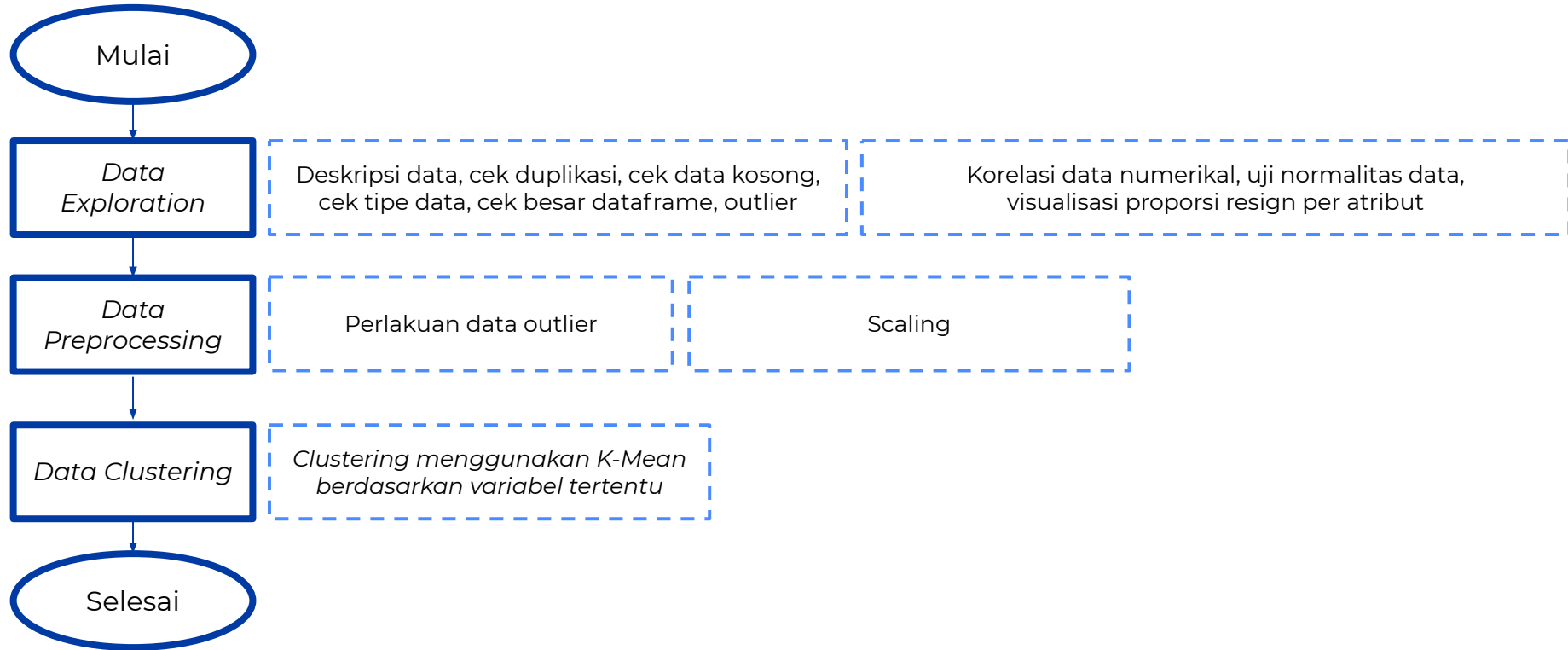
03 **Data Preprocessing**

Preparing the data for further process

04 **Data Clustering**

Clustering by data

01 - Alur Pengerjaan



02 - Data Exploration

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

Tipe Data

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 167 entries, 0 to 166  
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)  
memory usage: 13.2+ KB
```

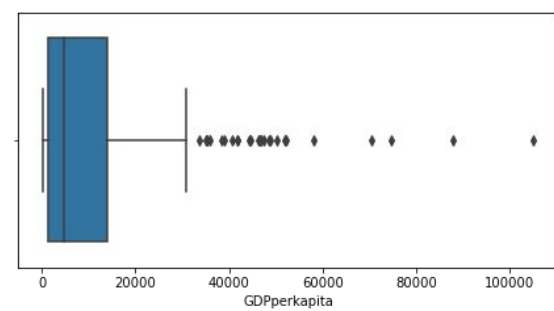
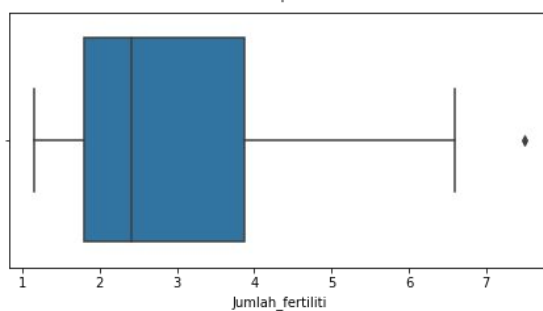
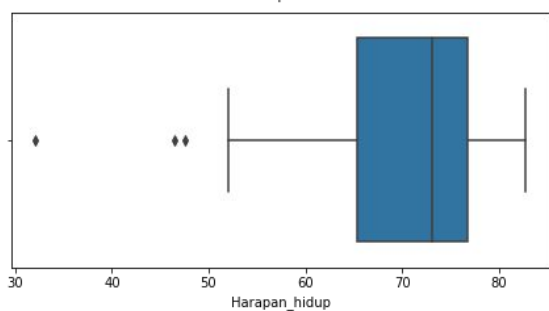
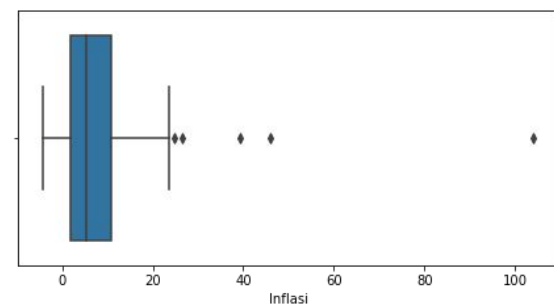
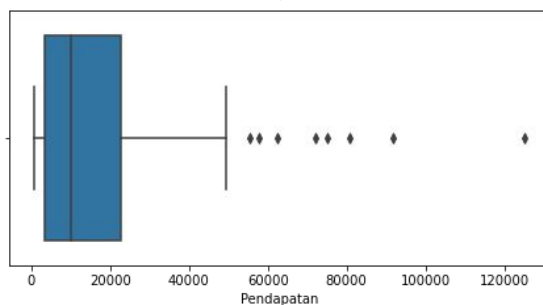
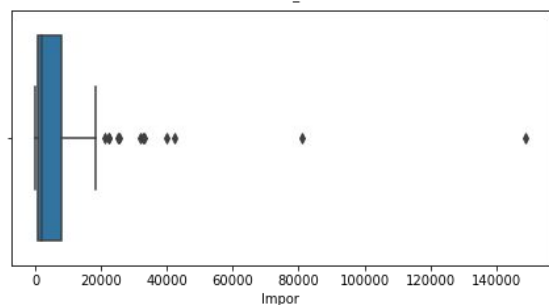
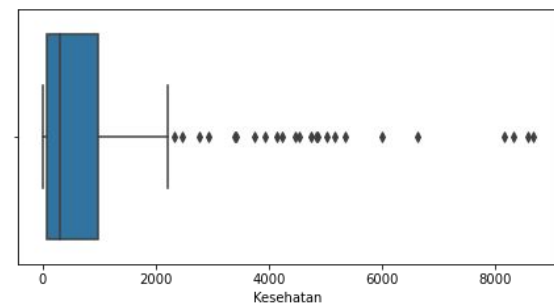
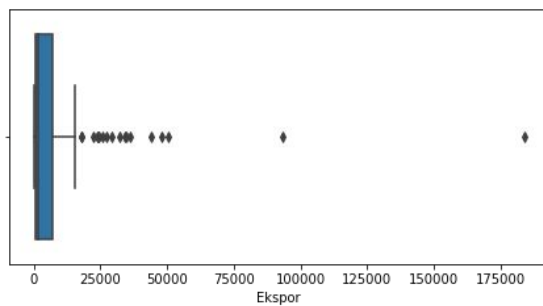
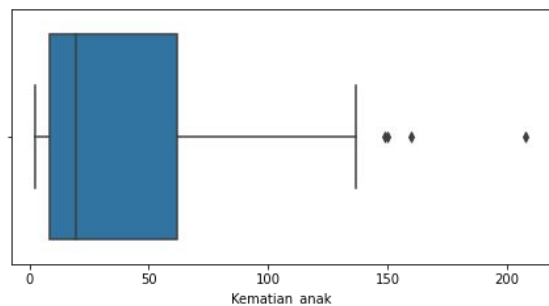
```
In [4]: #cek duplikasi data  
df.duplicated().any()  
  
Out[4]: False
```

Dataset yang ada tidak memiliki *missing values* dan data duplikat

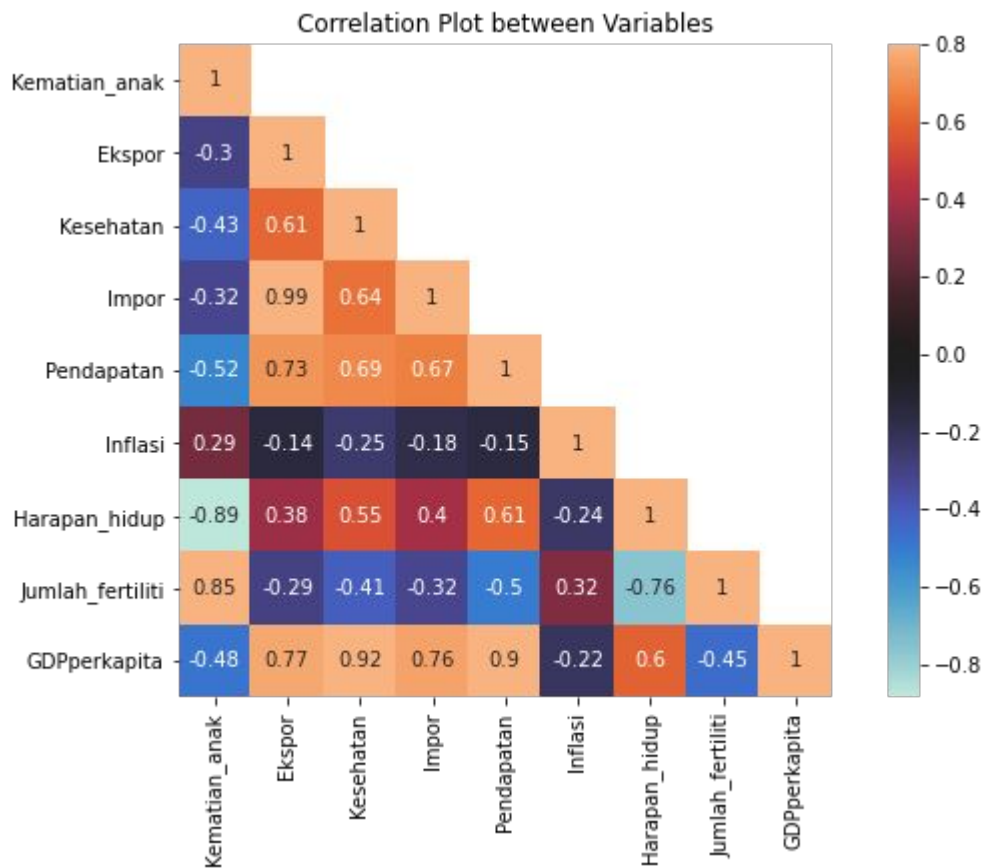
```
In [5]: #cek missing value  
df.isna().sum()  
  
Out[5]: Negara                0  
Kematian_anak              0  
Ekspor                    0  
Kesehatan                  0  
Impor                      0  
Pendapatan                 0  
Inflasi                    0  
Harapan_hidup              0  
Jumlah_fertiliti           0  
GDPperkapita               0  
dtype: int64
```

```
: #ekspor, kesehatan, dan impor merupakan nilai satuan per kap  
df['Ekspor'] = (df['Ekspor']*df['GDPperkapita'])/100  
df['Kesehatan'] = (df['Kesehatan']*df['GDPperkapita'])/100  
df['Impor'] = (df['Impor']*df['GDPperkapita'])/100  
df.head()
```

Mengubah Ekspor, kesehatan, dan impor dari satuan perkapita menjadi suatu nilai tertentu



Semua data numerik pada dataset memiliki outlier

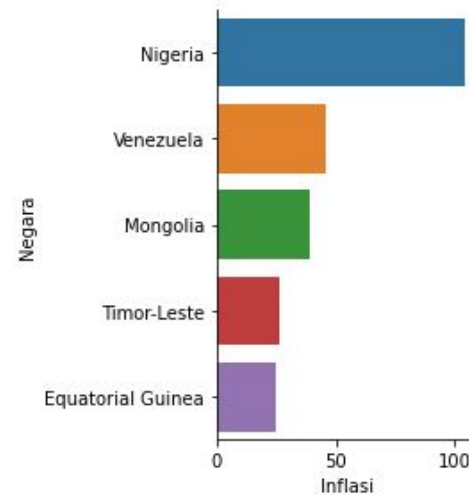
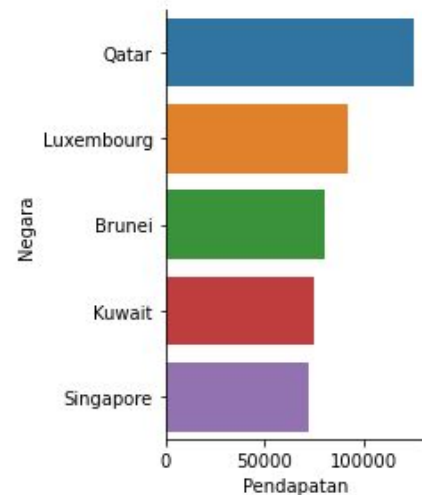
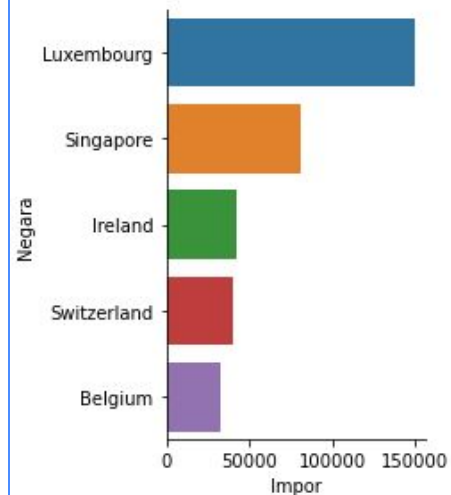
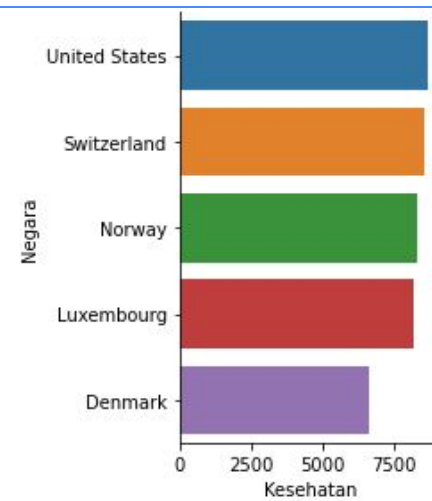
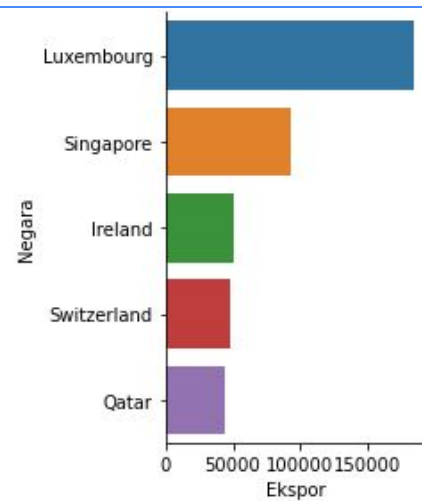
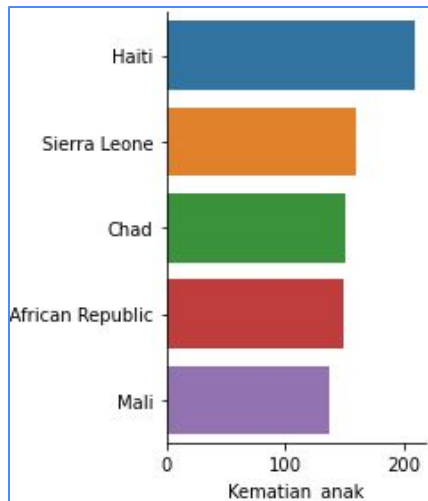


Uji Korelasi Data Numerik

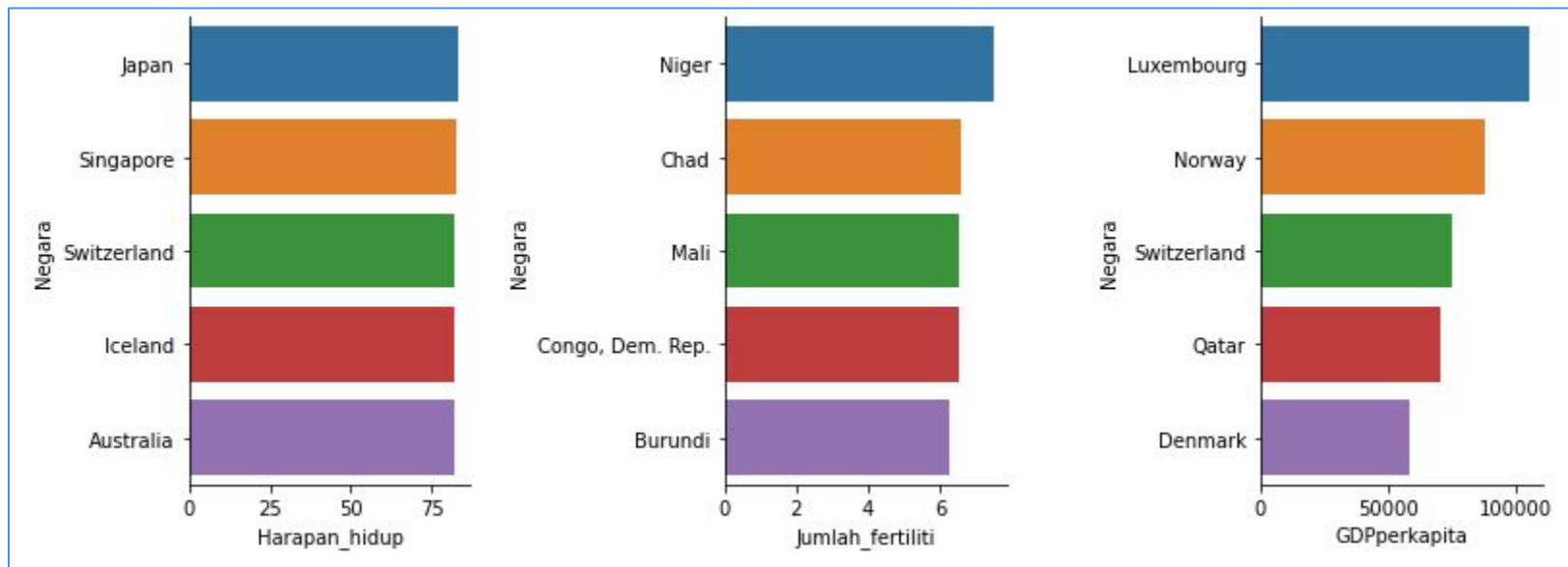
Terdapat multikolinearitas
pada dataset *train*.
(*cut-off* > 0.8)

Sehingga data yang akan digunakan
pada clustering adalah

1. GDPperkapita
2. Inflasi
3. Kematian_anak
4. Jumlah_fertiliti

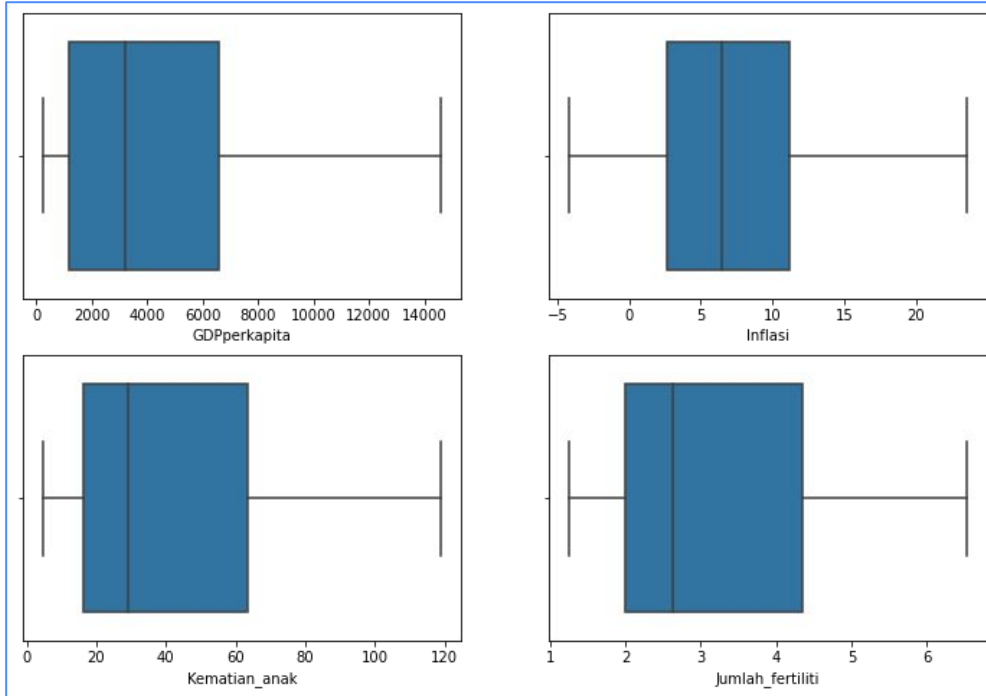


Top 10
Untuk setiap
variabel



Top 10 Untuk setiap variabel

03 - Data Preprocessing



Metode Penanganan Outlier

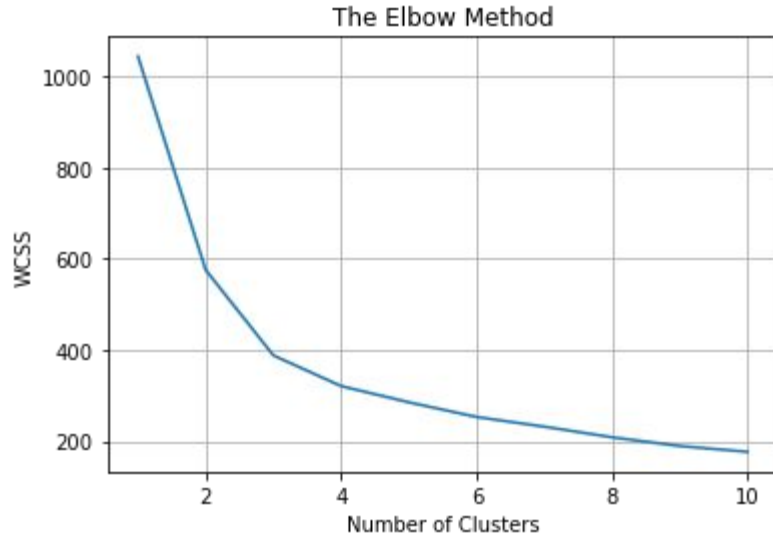
Metode penanganan menggunakan metode persentil di setiap variabel. Variabel yang digunakan disesuaikan dengan hasil korelasi, dimana yang digunakan adalah GDPperkapita, Inflasi, Kematian_anak, dan Jumlah_fertiliti.

```
sc = StandardScaler()  
df4 = sc.fit_transform(df2.drop('Negara',axis=1))
```



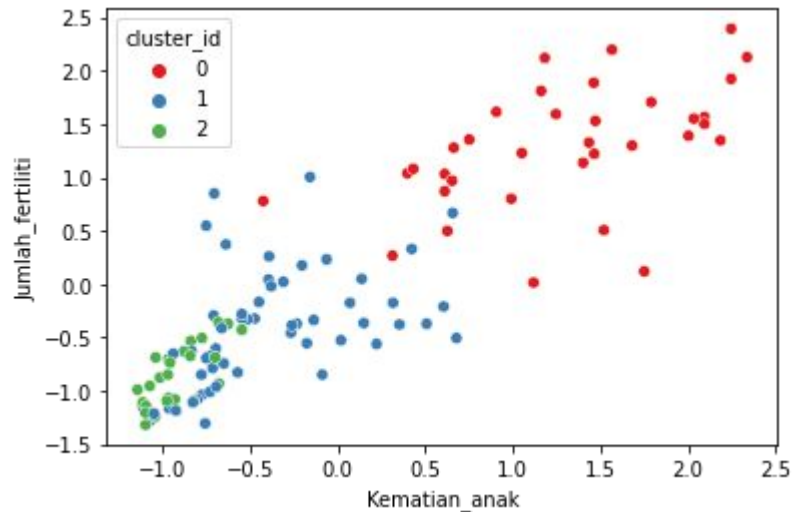
Menggunakan metode standarscaler untuk ke-4 variabel

04 - Data Clustering

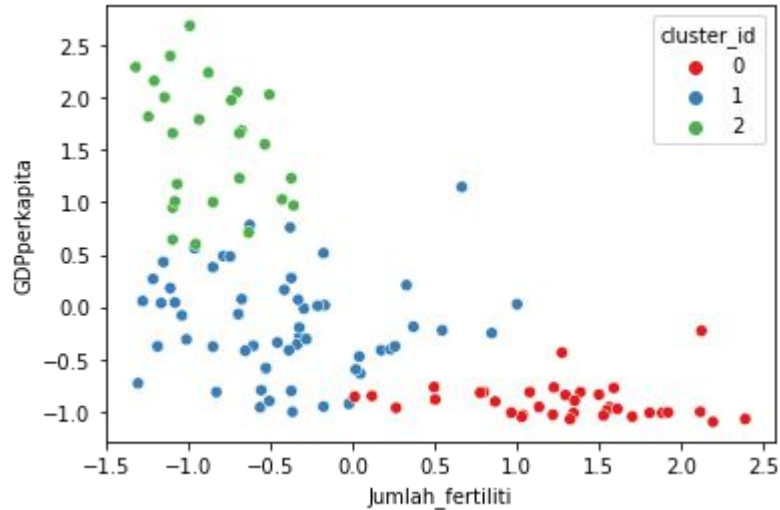


```
For n_clusters=2, the silhouette score is 0.36360372349602926
For n_clusters=3, the silhouette score is 0.34913612441344155
For n_clusters=4, the silhouette score is 0.3069455287209281
For n_clusters=5, the silhouette score is 0.26718720886877273
For n_clusters=6, the silhouette score is 0.2903349498727447
For n_clusters=7, the silhouette score is 0.2586057856340755
For n_clusters=8, the silhouette score is 0.23698426362843616
```

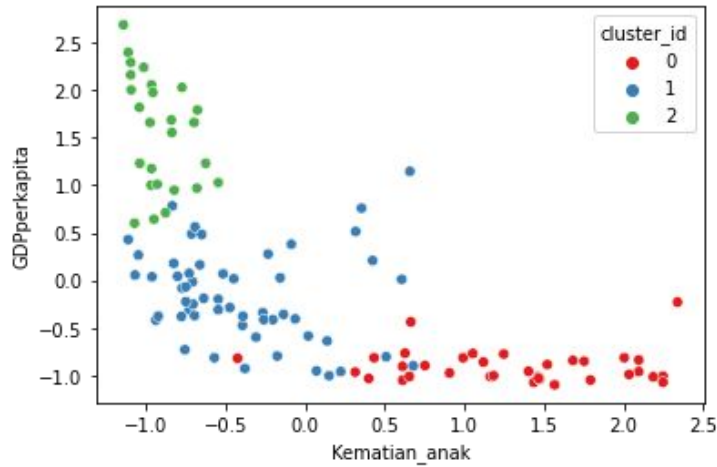
Penggunaan metode Elbow dan Silhouette menghasilkan n cluster optimal = 3



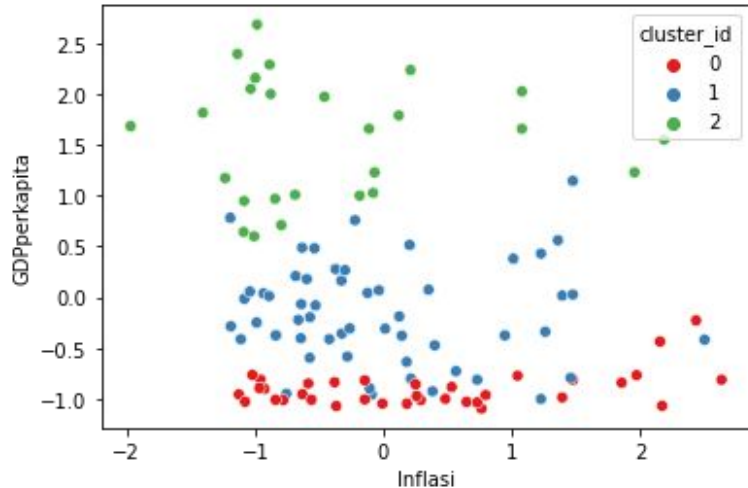
Cluster 0 atau berwarna merah menunjukkan bahwa kondisi pada negara tersebut memiliki kondisi kesehatan dan sosial yang buruk ditunjukkan dengan tingkat kematian_anak dan jumlah_fertiliti yang tinggi



Cluster 2 atau berwarna hijau menunjukkan bahwa dengan tingkat GDPperkapita yang tinggi akan berdampak pada jumlah_fertiliti yang rendah. Ekonomi berpengaruh terhadap jumlah_fertiliti sehingga butuh perhatian lebih pada cluster 0.



Cluster 2 atau berwarna hijau menunjukkan bahwa dengan tingkat GDPperkapita yang tinggi akan berdampak pada kematian_anak yang rendah. Ekonomi berpengaruh terhadap Kematian_anak sehingga butuh perhatian lebih pada cluster 0.



Cluster 0 atau berwarna merah menunjukkan negara dengan tingkat GDPperkapita rendah dan inflasi rendah. Perlu perhatian lebih terhadap negara negara pada cluster tersebut untuk menciptakan GDPperkapita tinggi dengan inflasi yang rendah.

	Negara	cluster_id
0	Afghanistan	0.0
3	Angola	0.0
18	Bhutan	0.0
19	Bolivia	0.0
21	Botswana	0.0
26	Burundi	0.0
27	Cambodia	0.0
28	Cameroon	0.0
30	Cape Verde	0.0
36	Comoros	0.0
40	Cote d'Ivoire	0.0
45	Dominican Republic	0.0
46	Ecuador	0.0
56	Gambia	0.0
57	Georgia	0.0
59	Ghana	0.0
62	Guatemala	0.0

63	Guinea	0.0
67	Hungary	0.0
71	Iran	0.0
81	Kiribati	0.0
89	Libya	0.0
94	Malawi	0.0
100	Mauritius	0.0
101	Micronesia, Fed. Sts.	0.0
108	Namibia	0.0

Negara Rekomendasi

Negara-negara berikut tergolong negara dengan cluster 0. CEO HELP International perlu memfokuskan donasi terhadap negara-negara tersebut.

Implikasi dan Rekomendasi

Implikasi

1. HELP International perlu berfokus dalam meningkatkan GDPperkapita atau ekonomi negara tertentu untuk negara yang benar membutuhkan bantuan
2. Hasil clustering bisa dijadikan acuan oleh HELP International untuk mulai mengalokasikan dana mereka dalam membantu tingkat kesehatan di negara tersebut

Rekomendasi

1. Negara termasuk cluster 0 pada clustering kematian_anak dan jumlah_fertiliti perlu perbaikan dari segi kesehatan dan juga ekonomi.
2. GDPperkapita dan kesehatan berkorelasi tinggi sehingga kedua variabel tersebut berbagi informasi yang sama, dengan kata lain GDPperkapita tinggi, maka tingkat kesehatan juga tinggi atau baik.
3. Ekonomi berdampak langsung terhadap kesehatan khususnya kematian_anak dan jumlah_fertiliti, sehingga dengan meningkatnya ekonomi akan berdampak pada menurunnya tingkat kematian_anak dan jumlah_fertiliti
4. Negara dengan GDPperkapita dan inflasi pada cluster 0 perlu diperhatikan dengan baik karena GDPperkapita sangat berpengaruh terhadap kesehatan