



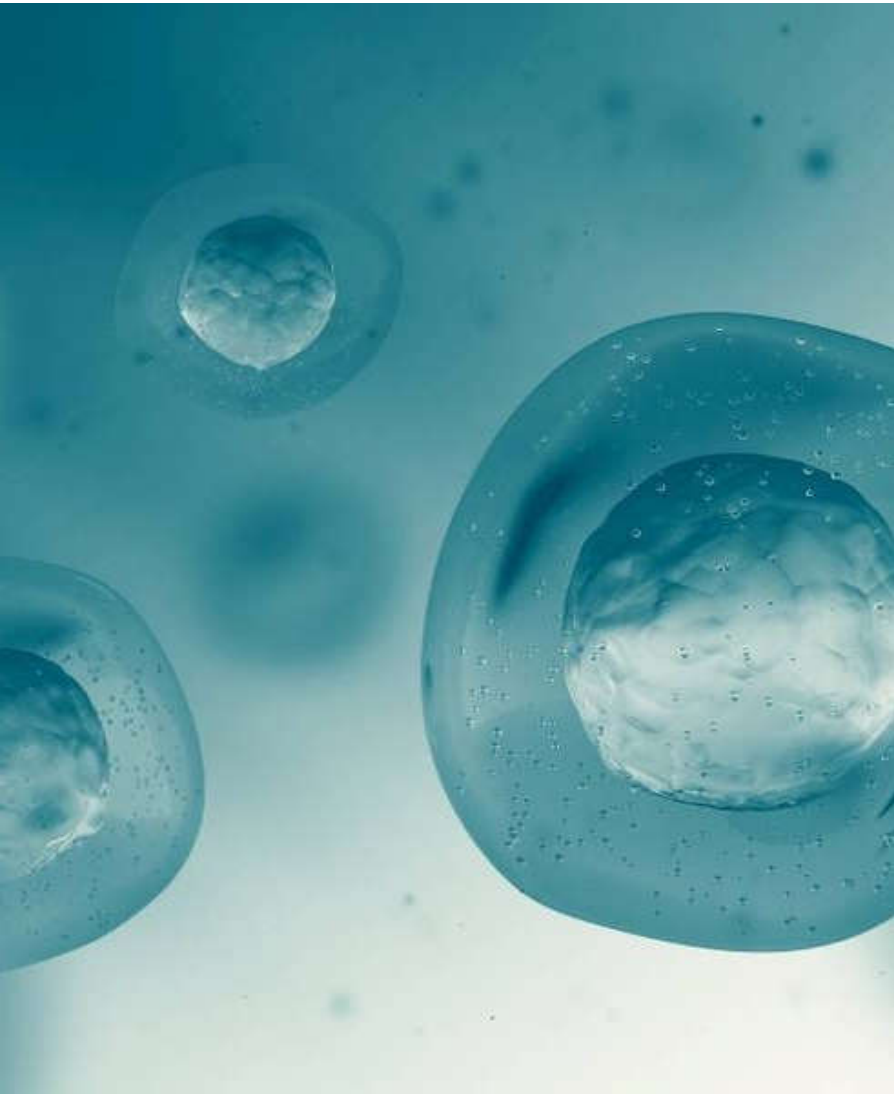
RELIEF FUNDS

Prakoso Adi Bagaskara



ABOUT US

HELP International is an international humanitarian NGO committed to fighting poverty and providing basic facilities and assistance to people in underdeveloped countries when disasters and natural disasters occur.



PROBLEM

UNALLOCATED RELIEF FUNDS

We has managed to raise relief funds around \$ 10 million.

WHICH COUNTRIES NEED THE MOST

The CEO needs to decide how to use this money strategically and effectively for which countries need it the most.

THE DATA

	Country	Child Mortality	Export	Health Expenditure	Import	Net Income	Inflation	Life Expentancy	Fertility Rate	GDP/capita
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
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

WE HAVE 167 ROWS WITH 10 COLUMNS

Which mean we have 167 candidate countries with 10 features

FEATURE

COUNTRY

The name of candidate country that will be given the relief funds.

CHILD MORTALITY

Deaths of children under 5 years of age per 1000 births in the country.

HEALTH EXPENDITURE

Total health spending per capita.

LIFE EXPECTANCY

The average number of years a newborn child would live if the current pattern of death remained the same.

FERTILITY RATE

The number of children that would be born to each woman if the current age fertility rate remained the same.

IMPORT

Imports of goods and services per capita.

EXPORT

Exports of goods and services per capita.

INFLATION

Measurement of the annual growth rate of Total GDP.

NET INCOME

Net income per person, in the country.

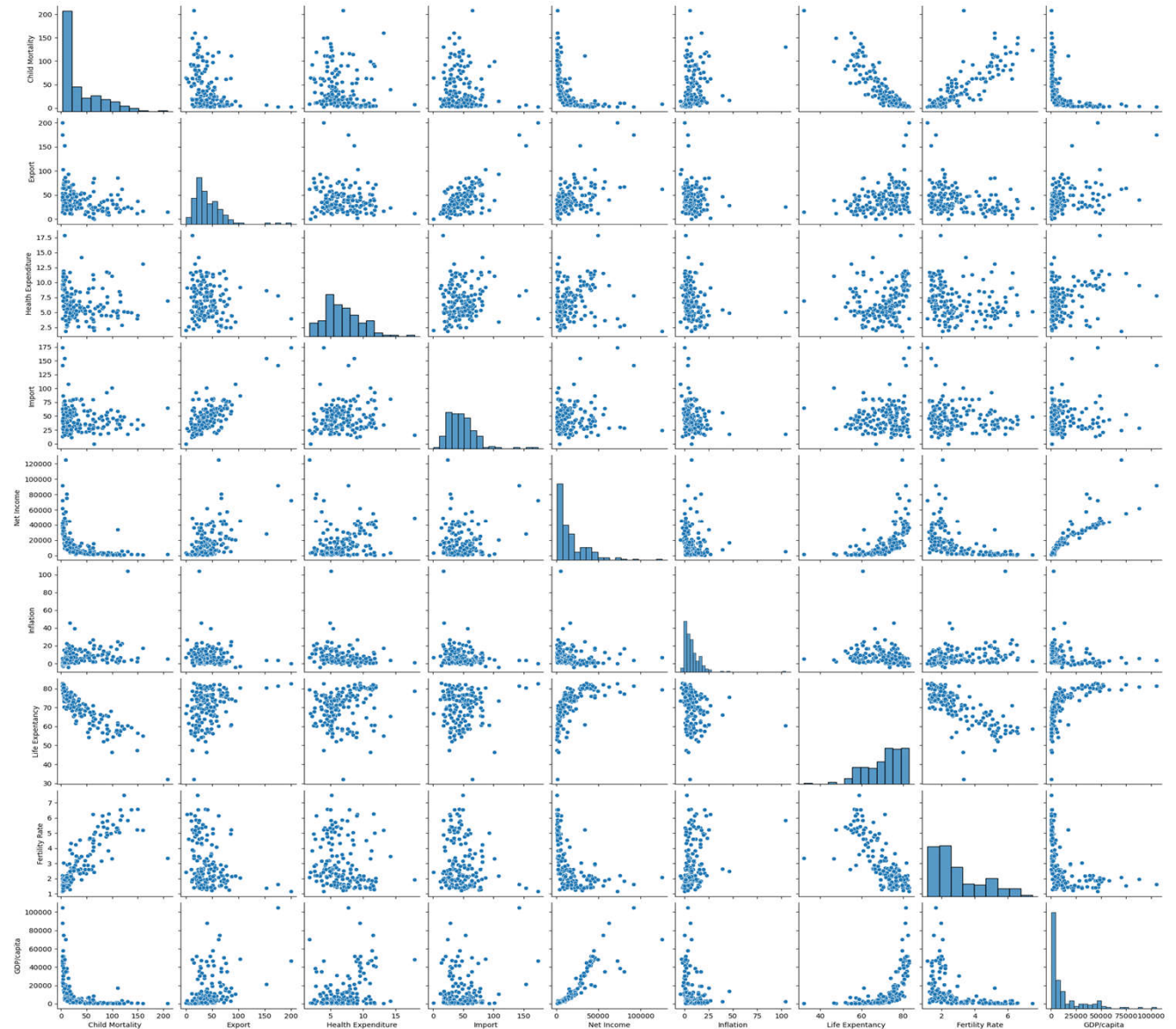
GDP/CAPITA

Calculated as Total GDP divided by the total population.

MULTIVARIATE

Jika dilihat dari chart disamping variable yang berkorelasi adalah sebagai berikut:

- Net Income dengan GDP berkorelasi positif.
- Import dengan ekspor berkorelasi positif.
- Child Mortality dengan life expectancy berkorelasi negative.
- Child Mortality dengan Fertility Rate berkorelasi positif.
- Life Expectancy dengan Fertility Rate Berkorelasi negatif



FEATURE FOCUS



INFLATION

Inflasi dipilih dikarenakan dengan adanya inflasi yang tinggi pasti akan menyebabkan kesulitan untuk masyarakatnya, apalagi masyarakat kecil dalam membeli kebutuhan sehari-hari. Dengan dana bantuan ini diharapkan bisa membantu masyarakat kecil dan juga meningkatkan *buying power* untuk menekan laju inflasi.

CHILD MORTALITY

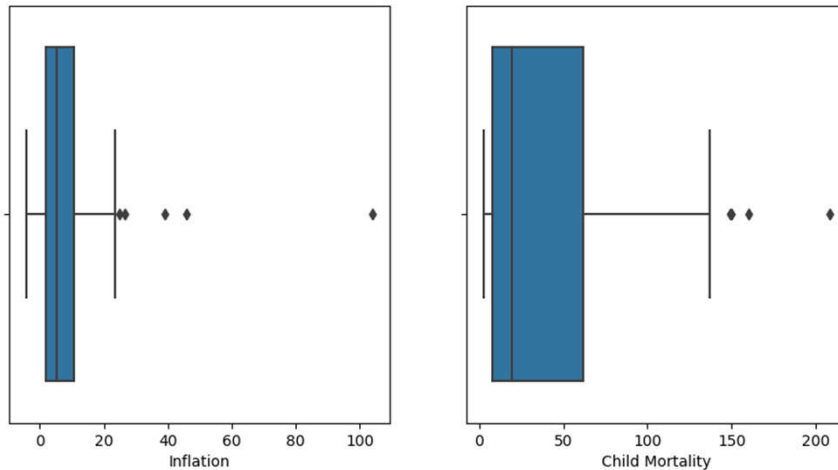
Anak merupakan asset masadepan dari negara dan siapa yang tidak sedih jika tingkat child mortality di negaranya tinggi, dengan dana bantuan ini semoga bisa membantu negara tersebut untuk bisa membuat program ataupun fasilitas khusus untuk menangani hal ini untuk menekan child mortality rate pada negara tersebut.

DATA CLEANING

```
RangeIndex: 167 entries, 0 to 166
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Country                167 non-null   object  
1   Child Mortality         167 non-null   float64  
2   Export                 167 non-null   float64  
3   Health Expenditure      167 non-null   float64  
4   Import                 167 non-null   float64  
5   Net Income              167 non-null   int64  
6   Inflation               167 non-null   float64  
7   Life Expentancy         167 non-null   float64  
8   Fertility Rate          167 non-null   float64  
9   GDP/capita              167 non-null   int64  
dtypes: float64(7), int64(2), object(1)
```

CHECK MISSING VALUE

Cek missing value dengan menggunakan method `.info()`. Hasilnya menunjukkan dari 167 baris di semua column mengeluarkan hasil 167 non null (tidak ada yang kosong).



CHECK OUTLIERS

Cek outliers dilakukan dengan membuat boxplot untuk melihat apakah ada data outlier yang digambarkan dengan bentuk diamond diluar boxplot.

Pada column **Inflation** terdapat 5 data outliers, sedangkan **Child Mortality** terdapat 4 data outliers yang semuanya diatas dari upper bound.

HANDLE OUTLIERS

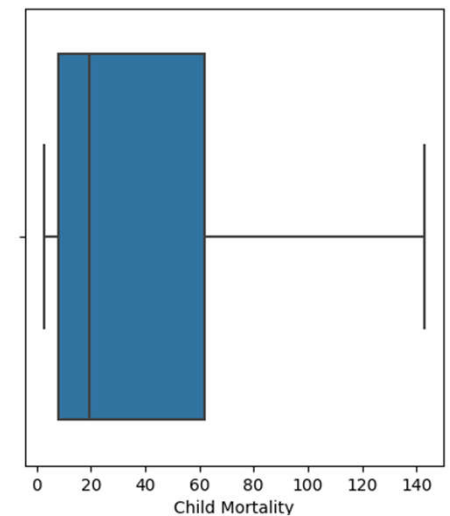
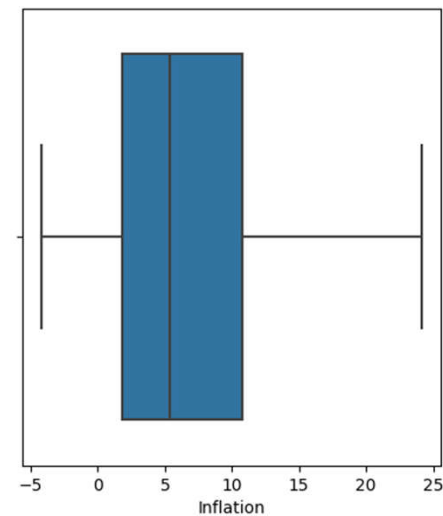
HANDLE DILAKUKAN DENGAN CARA:

- Menemukan nilai IQR, lower bound, upper bound dari masing-masing column.
- Melakukan cek data mana saja yang melebihi upper bound/lower bound.
- Setelah dicek menghasilkan hasil seperti berikut:

```
Outliers Inflation pada [] [ 24.9  39.2 104.  26.5  45.9]
Outliers Child Mortality pada [] [149. 150. 208. 160.]
```

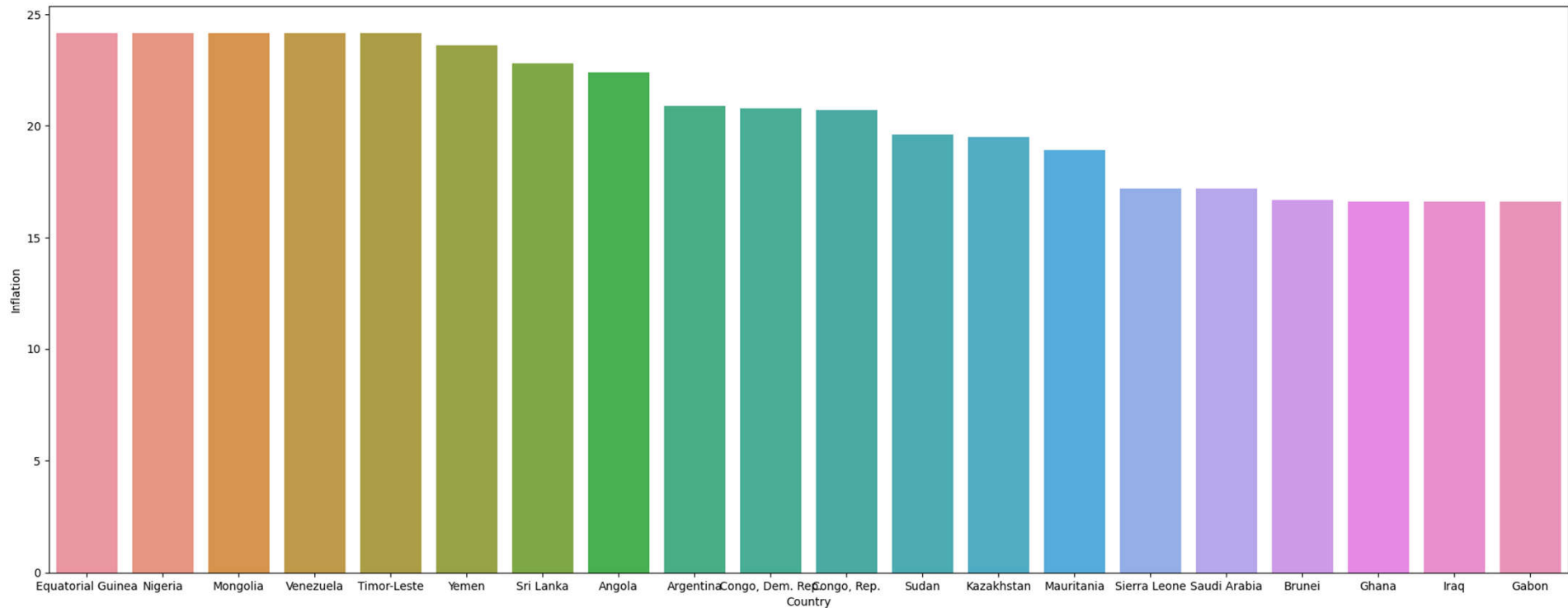
- Pada Lower Bound tidak ditemukan outlier pada kedua column, sama hasilnya dengan boxplot pada slide sebelumnya.
- Setelah itu nilai dari upper bound diganti valuenya dengan nilai upper bound menggunakan method `.replace()`.
- Lalu dilakukan pengecekan kembali menggunakan boxplot.
- Semua outlier berhasil direplace.

Alasan menggunakan method Interquartile Range (IQR) yaitu karena nilai yang didapatkan berdasarkan pada nilai yang berasal dari bagian tengah distribusi, kecil kemungkinannya dipengaruhi oleh outlier.



UNIVARIATE ANALYSIS

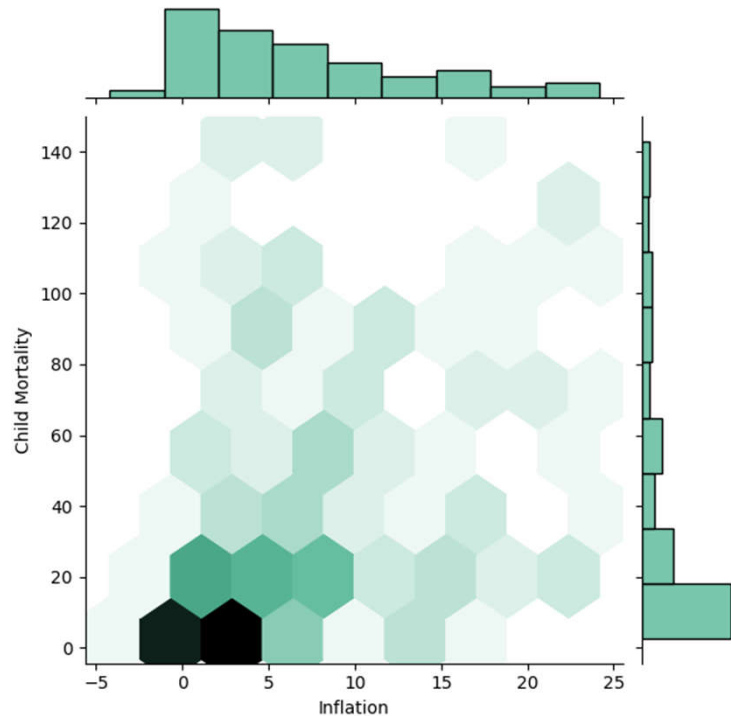
UNIVARIATE ANALYSIS DENGAN COLUMN INFLATION (TOP 20)



Analysis univariate dilakukan dengan membuat boxplot untuk agar lebih mudah melihat angka rate inflasi dimana. Dari hasil boxplot diatas ada 5 negara dengan Tingkat Inflasi diatas 24 (bar menunjukkan max 24 dikarenakan sudah direplace dengan nilai upper bound). Negara tersebut adalah: Equatorial Guinea, Nigeria, Mongolia, Venezuela, dan Timor Leste.

BIVARIATE ANALYSIS

COLUMN INFLATION DENGAN CHILD MORTALITY



- Bivariate dilakukan dengan cara membuat Hexbins dari library Seaborn.
- Hexbins dipilih agar lebih mudah melihat tingkat Inflation rate dan Child Mortality pada mayoritas Negara.
- Bisa diambil insight dari Hexbins bahwa dari 167 Negara kebanyakan Negara Inflation Rate-nya berada di sekitar area -2 sampai 5 dan tingkat Child Mortality berada di rate 20 kebawah.

DATA CLUSTERING

STEP DATA CLUSTERING :



Data Clustering dilakukan untuk melihat cluster negara-negara dengan berdasarkan tingkat Inflation Rate dan juga Child Mortality. Hasil dari data clustering nanti kita bisa melihat cluster negara-negara dengan tingkat Inflation Rate & Child Mortality yang tinggi, yang nantinya akan memudahkan kita untuk menentukan negara mana saja yang perlu dibantu dengan uang bantuan yang kita punya.

RESCALING DATA

- Awal melakukan rescale yaitu memisahkan data yang mau direscale dan menampungnya pada dataframe baru. Lalu pastikan indexnya sama dengan dataframe asli, agar setelah dilakukan clustering dapat dengan mudah menemukan negara mana saja yang masuk dalam cluster yang dipilih.
- Setelah itu dilakukan rescale data dengan menggunakan fungsi StandarScaler dari library sklearn.

Dataframe rescale

	Inflation	Child Mortality
0	9.44	90.2
1	4.49	16.6
2	16.10	27.3
3	22.40	119.0
4	1.44	10.3
...
162	2.62	29.2
163	24.16	17.1
164	12.10	23.3
165	23.60	56.3
166	14.00	83.1

167 rows × 2 columns

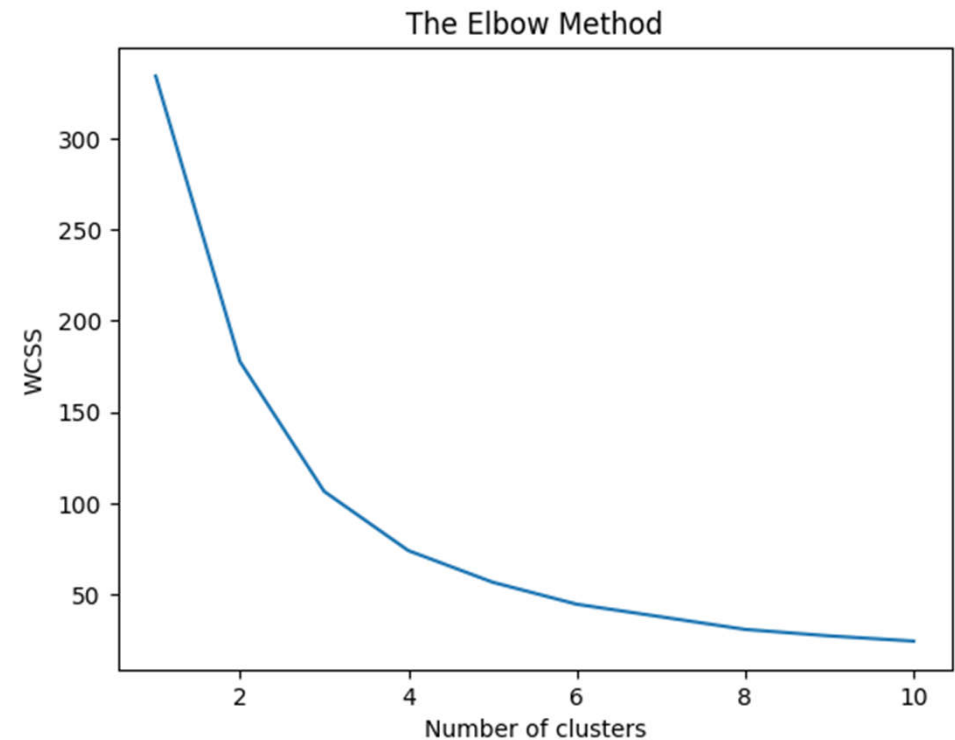
Dataframe asli

	Country	Inflation	Child Mortality
0	Afghanistan	9.44	90.2
1	Albania	4.49	16.6
2	Algeria	16.10	27.3
3	Angola	22.40	119.0
4	Antigua and Barbuda	1.44	10.3
...
162	Vanuatu	2.62	29.2
163	Venezuela	24.16	17.1
164	Vietnam	12.10	23.3
165	Yemen	23.60	56.3
166	Zambia	14.00	83.1

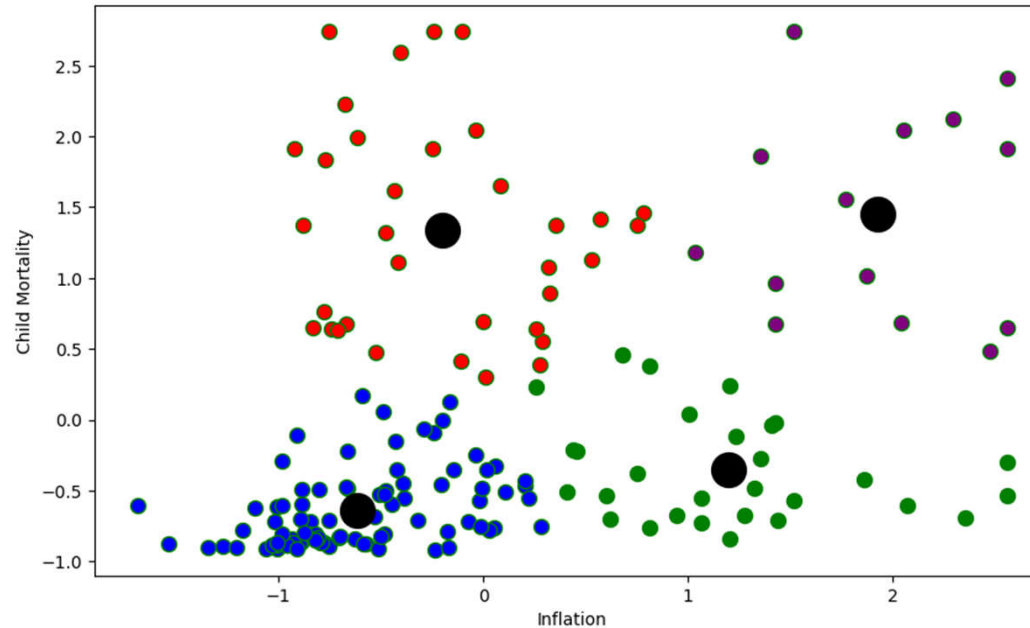
167 rows × 3 columns

ELBOW METHOD

- Untuk memilih berapa cluster yang tepat untuk nanti dilakukan pada algoritma K-Means, disini kita menggunakan cara Elbow Method.
- Pada Elbow Method kita mencari jarak antar titik dengan pusat klusternya atau bisa disebut dengan WCSS(Within-Cluster-Sum of Squared), nilai kluster terbaik yaitu jika nilai dari WCSS sudah mulai stabil.
- Pada hasil Elbow Method bisa dilihat nilai WCSS mulai stabil pada angka 4. Jadi, kita akan membuat 4 kluster pada algoritma K-Means.



K-MEANS



Hasil dari K-means dengan 4 kluster menunjukkan cluster yang bisa diklasifikasikan sebagai berikut:

- Cluster Biru: Cluster Negara-Negara dengan Tingkat Inflasi dan Child Mortality rendah, cluster ini paling banyak anggotanya sesuai hasilnya dengan analisis Bivariate sebelumnya, menunjukkan bahwa dari 167 Negara, mayoritas negara tingkat Inflasi dan Child Mortality-nya tergolong rendah.
- Cluster Hijau: Cluster Negara dengan tingkat Inflasi tinggi namun Child Mortality nya rendah.
- Cluster Merah: Cluster Negara dengan tingkat Inflasi rendah namun Child Mortality nya tinggi.
- Cluster Ungu: Nah, pada cluster inilah yang menjadi calon terbaik untuk menerima uang bantuan, karena pada cluster ini tingkat Inflasi dan Child Mortality-nya tinggi.

K-MEANS

Berikut adalah hasil dari anggota Cluster Ungu, lalu kita cocokan dengan data dari dataframe asli berdasarkan dari nilai indexnya.

Data Cluster Ungu

	Inflation	Child Mortality	label_kmeans2
3	2.293979	2.121210	3
37	2.054632	2.042939	3
38	2.039673	0.683620	3
49	2.557260	1.912486	3
55	1.426347	0.678402	3
59	1.426347	0.965398	3
63	1.351551	1.860305	3
99	1.770408	1.557654	3
113	2.557260	2.408207	3
132	1.516102	2.744123	3
142	1.875122	1.017579	3
149	2.557260	0.649702	3
165	2.473489	0.485332	3
166	1.037409	1.184559	3

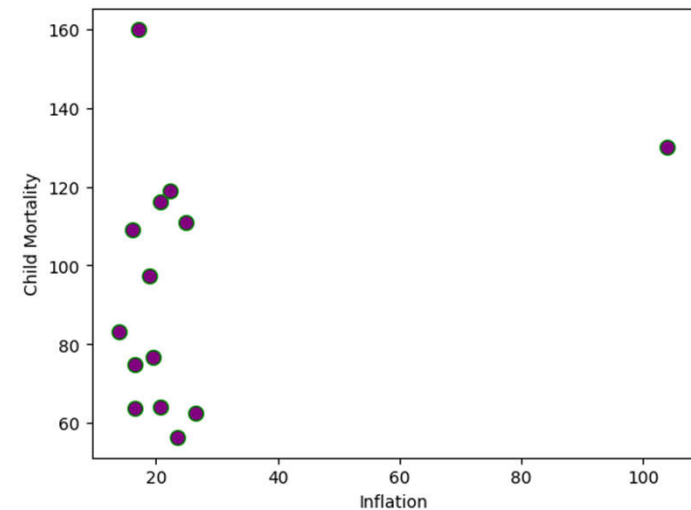
Data Asli

	Country	Child Mortality	Export	Health Expenditure	Import	Net Income	Inflation	Life Expentancy	Fertility Rate	GDP/capita
3	Angola	119.000	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
37	Congo, Dem. Rep.	116.000	41.1	7.91	49.6	609	20.80	57.5	6.54	334
38	Congo, Rep.	63.900	85.1	2.46	54.7	5190	20.70	60.4	4.95	2740
49	Equatorial Guinea	111.000	85.8	4.48	58.9	33700	24.16	60.9	5.21	17100
55	Gabon	63.700	57.7	3.50	18.9	15400	16.60	62.9	4.08	8750
59	Ghana	74.700	29.5	5.22	45.9	3060	16.60	62.2	4.27	1310
63	Guinea	109.000	30.3	4.93	43.2	1190	16.10	58.0	5.34	648
99	Mauritania	97.400	50.7	4.41	61.2	3320	18.90	68.2	4.98	1200
113	Nigeria	130.000	25.3	5.07	17.4	5150	24.16	60.5	5.84	2330
132	Sierra Leone	142.875	16.8	13.10	34.5	1220	17.20	55.0	5.20	399
142	Sudan	76.700	19.7	6.32	17.2	3370	19.60	66.3	4.88	1480
149	Timor-Leste	62.600	2.2	9.12	27.8	1850	24.16	71.1	6.23	3600
165	Yemen	56.300	30.0	5.18	34.4	4480	23.60	67.5	4.67	1310
166	Zambia	83.100	37.0	5.89	30.9	3280	14.00	52.0	5.40	1460

REKOMENDASI NEGARA

Sebelum kita menentukan rekomendasi negara mana saja yang mendapatkan dana bantuan, kita buat kembali scatterplot dengan data yang asli.

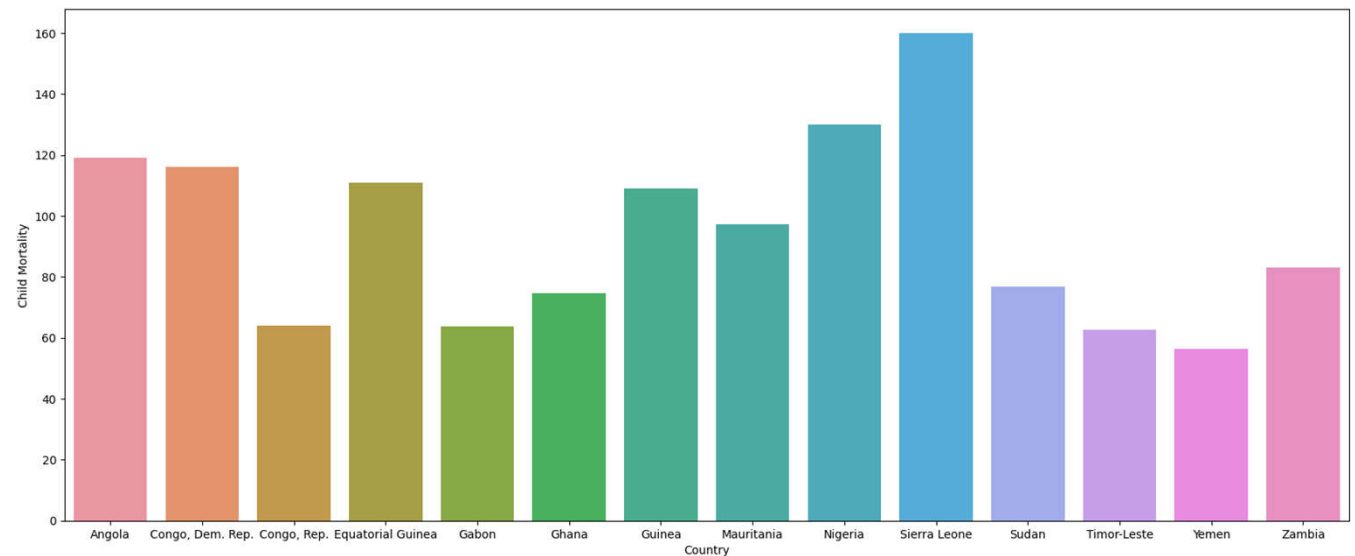
Jika dilihat dari scatterplot terdapat 2 Negara dengan yang nilainya jauh lebih tinggi dari negara-negara lain di cluster Ungu, satu Negara dengan tingkat Child Mortality yang nilainya tertinggi (nilainya jauh dibanding negara lain), dan satu negara lagi dengan Inflation Rate tertinggi dan juga Child Mortality tertinggi kedua.



Untuk mengetahui negara mana yang harus menjadi perhatian kita bisa tampilkan barplot dengan column Child Mortality saja, karena jika kita lihat dari scatterplot, terdapat 2 negara dengan child mortality jauh bedanya dari yang negara lain sedangkan pada tingkat inflasi hanya satu negara yang bedanya jauh dan negara tersebut yaitu terletak pada urutan ke-2 dengan child mortality tertinggi.

Dilihat dari barplot Negara tersebut adalah **Nigeria**.

Jadi, rekomendasi saya untuk negara yang mendapatkan bantuan adalah **Nigeria** dan juga satu lagi **Sierra Leone**, yaitu negara dengan tingkat child mortality yang tinggi.



THANK YOU



AKHIR KATA

Semoga dengan dana bantuan ini kedua negara tersebut bisa mengurangi tingkat Child Mortality, dan juga pada Nigeria bisa juga untuk membantu masyarakat kecil untuk bisa membeli kebutuhan sehari-hari dan meningkatkan *buying power* pada masyarakat dan selanjutnya bisa mengurangi tingkat inflasi menjadi lebih kecil.