

A cluster of overlapping, semi-transparent geometric shapes in shades of blue, green, and red, located in the top-left corner of the slide.

Clustering the Countries by using K-Means for HELP International

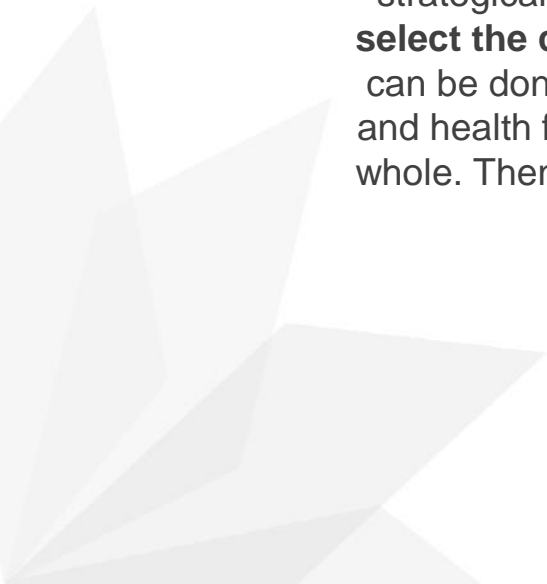
A vertical bar composed of four thin, parallel lines in blue, green, red, and orange, positioned to the left of the author's name.

RENI FATARINA HANDAYANI

A cluster of overlapping, semi-transparent geometric shapes in shades of light gray, located in the bottom-right corner of the slide.



Problem



HELP International has managed to raise around \$10 million. Right now, the NGO CEO needs to decide how to use this money strategically and effectively. So, the CEO must **make a decision to select the country that needs** the most assistance. Decision making can be done by categorizing countries using several socio-economic and health factors that determine the development of the country as a whole. Then we'll suggest which countries the CEO needs to focus on the most.



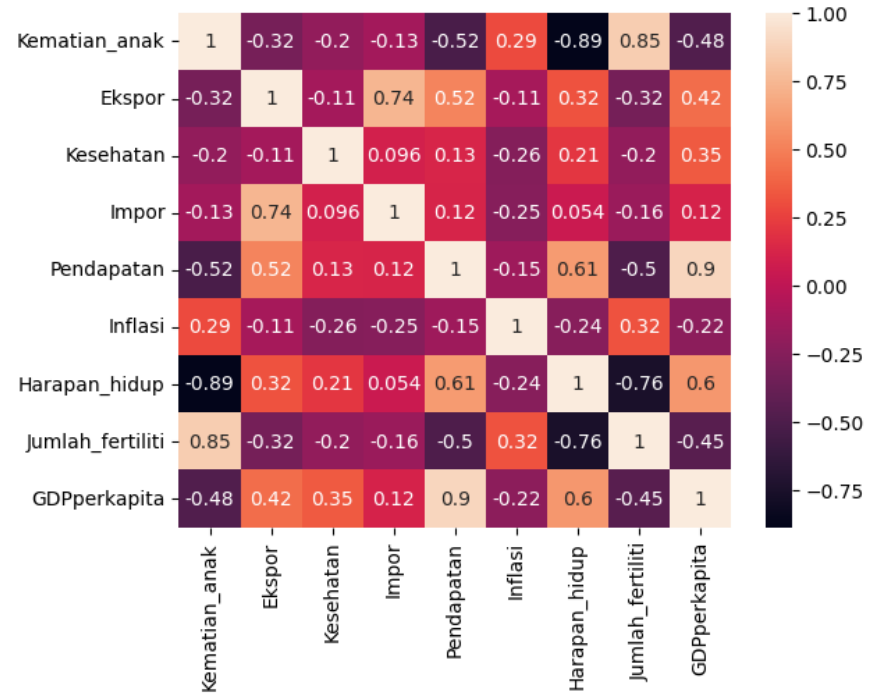
DETAIL DATASET

There are 167 rows of data and 10 columns. The following columns are available:

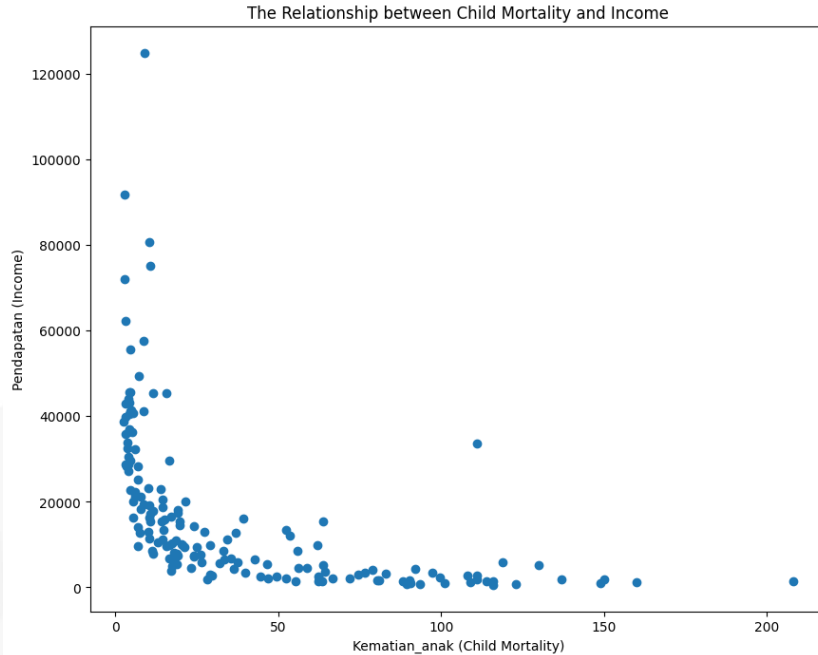
- **Negara** : Country name
- **Kematian_anak** : Death of child under 5 years old per 1000 births
- **Ekspor** : Export of goods and services per capita
- **Kesehatan** : Total health expenditure per capita
- **Impor** : Imports of goods and services per capita
- **Pendapatan** : Net income per person
- **Inflasi** : Measurement of the annual growth rate of Total GDP
- **Harapan_hidup** : The average number of years a newborn child would live if current mortality patterns remained the same
- **Jumlah_fertiliti** : The number of children that would be born to each woman if the current age fertility rate remained the same
- **GDPperkapita** : GDP per capita. It is calculated as Total GDP divided by the total population.

EDA 1

The heatmap uses colored cells to represent the **correlation between two variables**. Dark red indicates a strong positive correlation (close to 1), dark blue indicates a strong negative correlation (close to -1), and colors close to neutral (white or yellow) indicate a correlation close to 0



Feature Selection



The features selected for socio-economic and health factors are **Child Mortality (Kematan_anak)** and **Income (Pendapatan)**. It means child mortality shows the death rate of children under 5 per 1000 births. So, the higher the death rate, the lower the health level. While income is net income per person, the lower the income value, the lower the economic level.

Data Cleaning

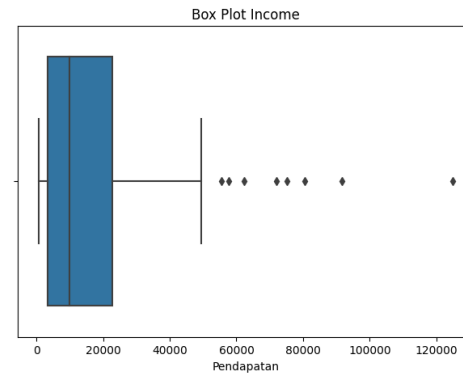
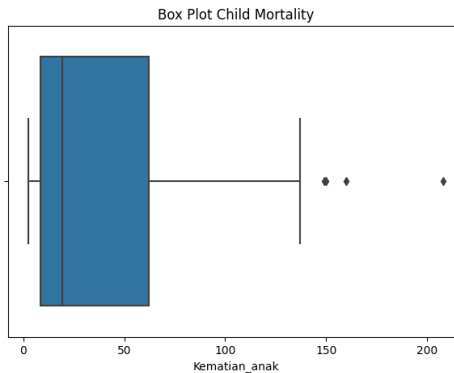


```
Kematian_anak    0  
Pendapatan       0  
dtype: int64
```

01

Check Data Null

There is **no data null** in the data frame.

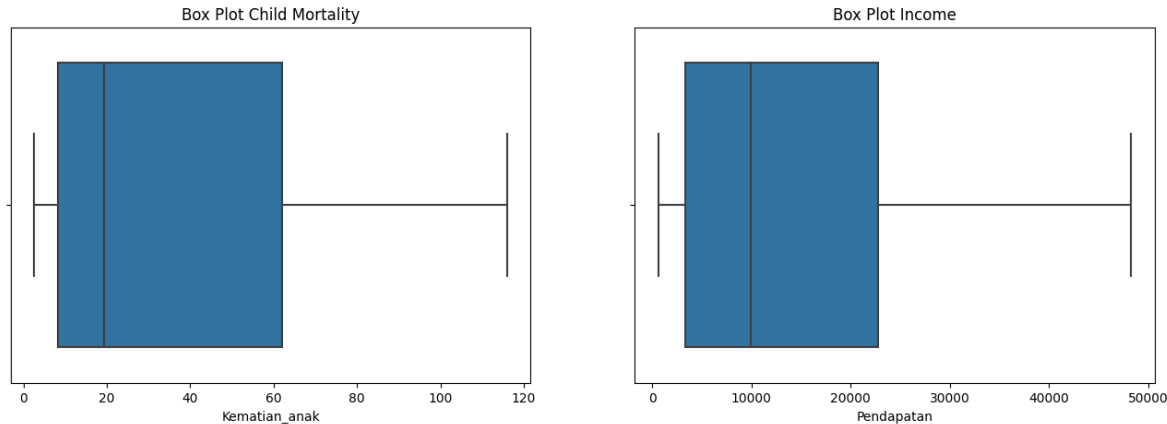


02

Check Data Outlier

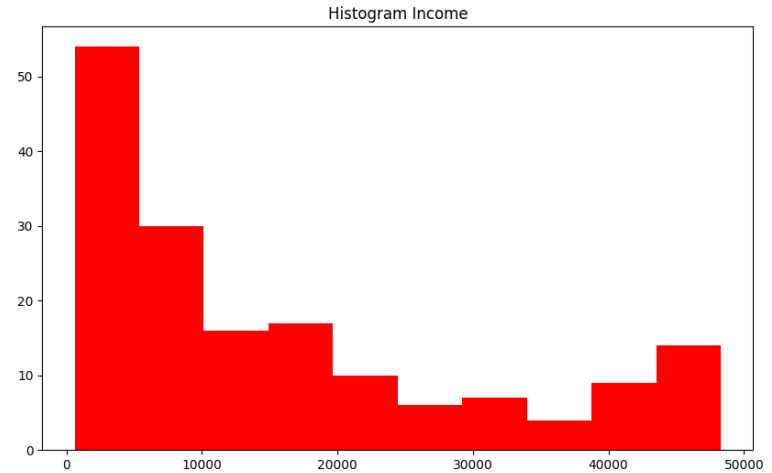
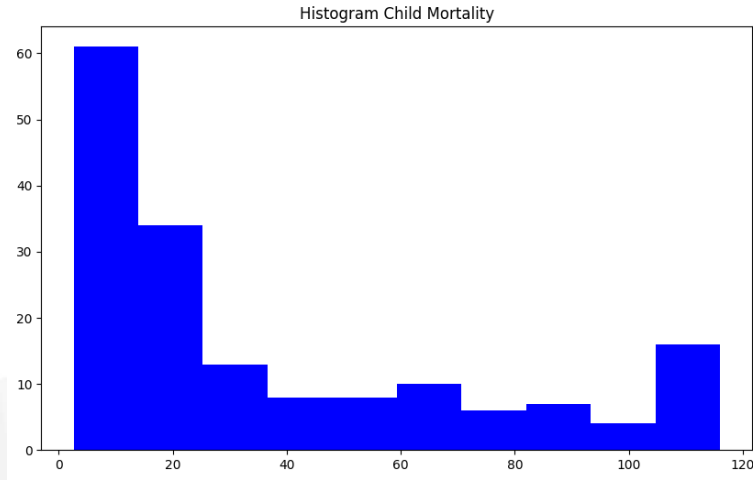
There are **outliers in both columns**.

Handling Outlier



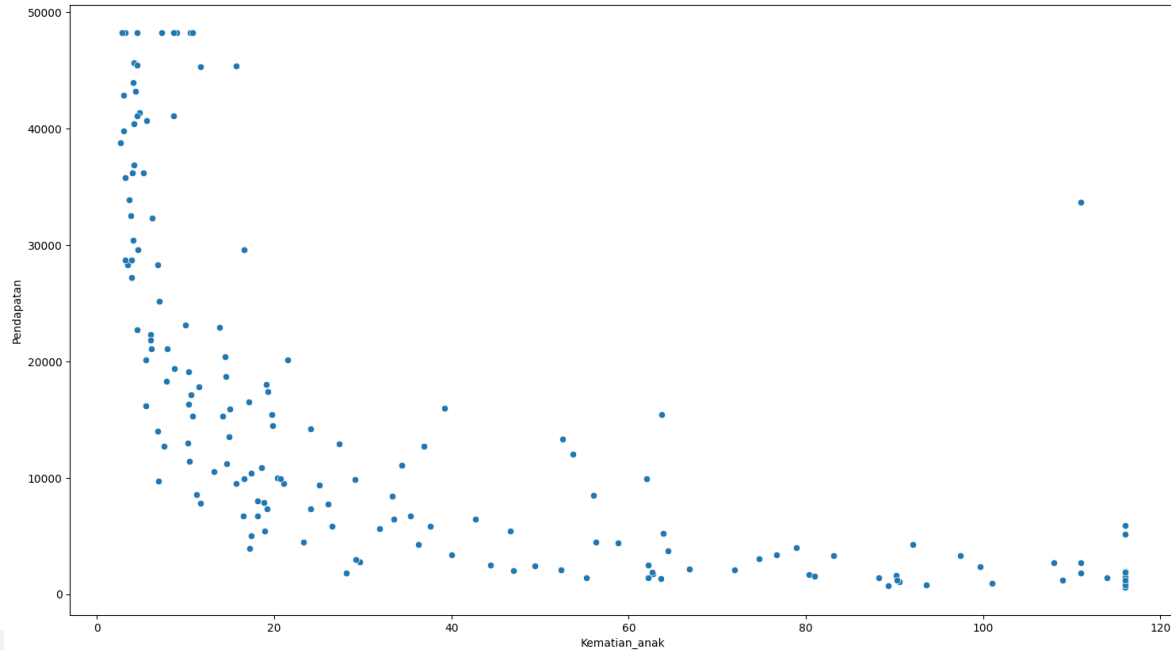
This outlier handling is done by the **Winsorization** Method. It is because Winsorization replaces outlier values with some cut-off values (e.g., a specific percentile value), thus retaining the information from the original data. Additionally, Winsorization can help reduce their influence on the analysis if the data has very extreme outliers without completely losing those extreme values. Both data handle using 95 percentile.

EDA 2



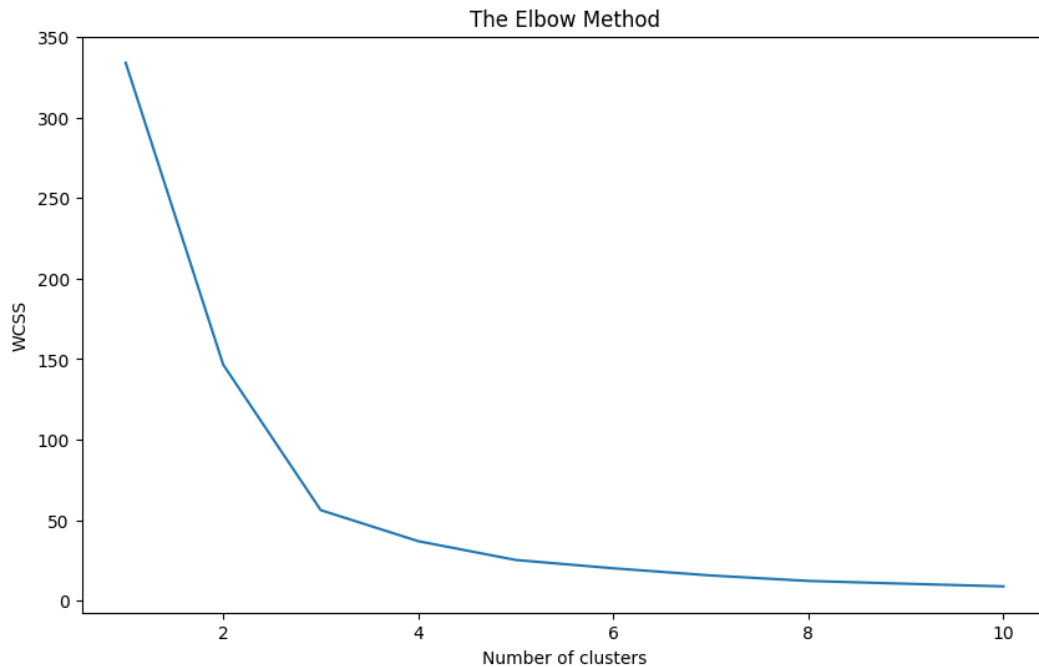
From the graph it can be seen that the number of child mortality is around 10 very high. For income, most countries have an income of around 5000.

EDA 2



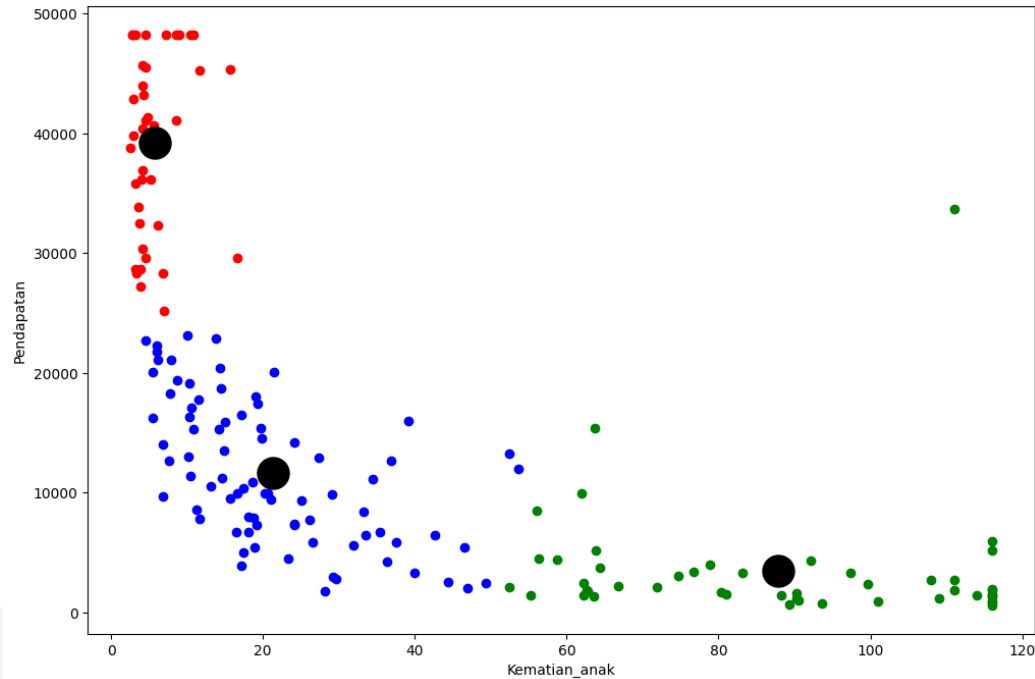
This graph shows that the higher the income, the lower the number of child deaths. Vice versa, the higher the child mortality rate, the lower the income rate.

Clustering



The data frame was scaled to equalize the value to make it easier to do clustering. Determining the number of clustering is done with elbows, so 3 clusters are found.

Clustering





Recommendation County for Providing Financial Assistance

1. Congo, Dem. Rep.
2. Liberia
3. Burundi
4. Niger
5. Central African Republic



The country has a low income and a high child mortality rate. So that, the 5 country recommendations deserve assistance



Thank you