# Summary

The "Clustering of Countries" assignment, undertaken on behalf of HELP International, aimed to prioritize aid distribution to nations based on socio-economic and health parameters. The core problem addressed was to determine which countries, among all, required urgent humanitarian aid. After obtaining a commendable fund of $10 million, the NGO sought to deploy these resources effectively and strategically.

My approach began with a comprehensive Exploratory Data Analysis (EDA). I initiated data cleaning to handle anomalies and missing values, followed by univariate and bivariate analyses to understand the data's distribution and relationships between variables. Through this, I could identify critical factors that influence a country's developmental status.

Given the data's nature, I employed two clustering methodologies: K-means and Hierarchical Clustering. Both methods, though distinct, aim to group data into clusters based on similarity. My analysis revealed that while both methods were insightful, K-means produced more consistent and actionable clusters for this dataset.

A crucial aspect of my solution was the outlier analysis. Recognizing that some outliers could represent countries in severe distress, I made a strategic choice to retain them, ensuring no nation in genuine need was overlooked.

By examining variables like GDP per capita, child mortality, and income levels, I was able to distinguish between developed and underdeveloped nations. The clusters formed provided a clear picture of countries in dire need, facilitating a data-driven recommendation for the NGO's aid distribution.