# Unsupervised Learning Technique of World Happiness Data

The dataset we worked with includes a wide range of indicators that represent the economic, social, and well-being status of various countries. The columns in the dataset are: Country name (identifying the country), year (indicating the time period of the data), Life Ladder (a measure of subjective well-being or life satisfaction), Log GDP per capita (economic strength of the country measured in GDP per capita and transformed logarithmically), Social support (availability of support networks for individuals), Healthy life expectancy at birth (years a newborn is expected to live in good health), Freedom to make life choices (level of autonomy individuals feel in making decisions), Generosity (perception of generosity in the country), Perceptions of corruption (perceived levels of corruption in governance), Positive affect (extent to which positive emotions are experienced), and Negative affect (frequency of negative emotions like sadness or anger). These indicators offer a comprehensive view of how citizens in each country experience life, in terms of both economic and social well-being. For the purpose of analysis we only limit our data to year 2023.

The objective of the analysis was to identify clusters of countries that share similar characteristics across these well-being indicators. Using clustering algorithms, such as KMeans, we sought to group countries based on their performance in areas like life satisfaction, health, GDP, and social support. By reducing the dimensionality of the data with Principal Component Analysis (PCA), we aimed to visualize these clusters more effectively, simplifying the complex multivariate data into two principal components for easy interpretation.

After performing the clustering analysis and using the elbow method, we identified 3 and 5 as the optimal numbers of clusters. When applying the KMeans clustering algorithm, we grouped countries based on the similarity of their economic and well-being indicators. The scatter plot of the clusters provided insight into how countries were grouped, though the individual country labels were also plotted to allow for a more detailed analysis. Countries within the same cluster demonstrated similar levels of economic development, social support, life expectancy, and subjective well-being. For instance, countries with higher Log GDP per capita, Life Ladder, and Healthy life expectancy tended to cluster together, while others with lower values in these categories formed different clusters. Additionally, the variance in attributes like Perceptions of corruption and Negative affect was a strong factor in the formation of certain clusters.

In conclusion, by clustering countries based on these indicators, we can better understand how different nations compare in terms of economic, social, and emotional well-being, and identify patterns that could inform global health and economic policies.