

Name: Saher Saeed

Student No: 23095056

Github Repository: <https://github.com/saeedsahar/clustering-and-fitting.git>

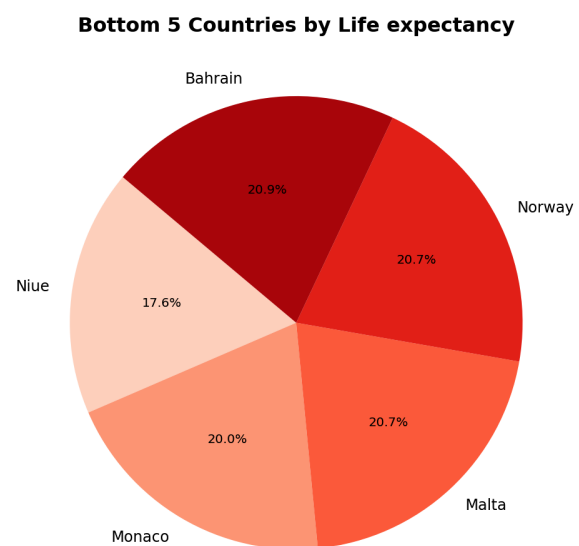
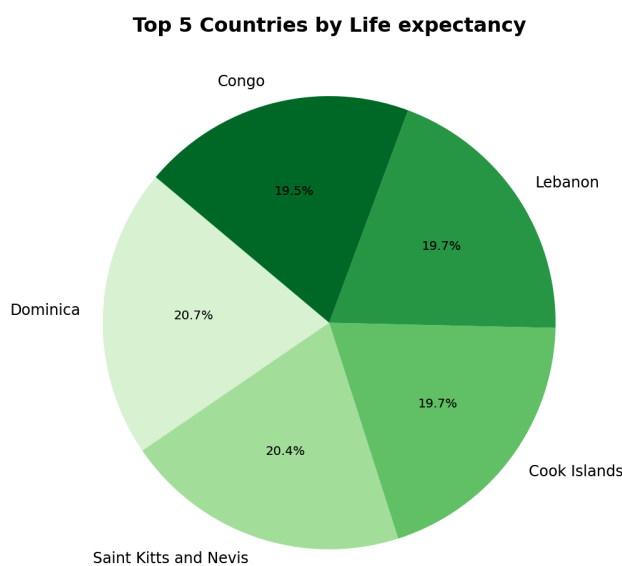
Life Expectancy Analysis Report

Introduction

This report takes a deep dive into life expectancy across the globe, examining factors such as economic capacity (GDP), mortality, education levels and distribution of resources. We visualized our analysis on multiple charts to show trends so, revealing the inequalities and relationships between these factors.

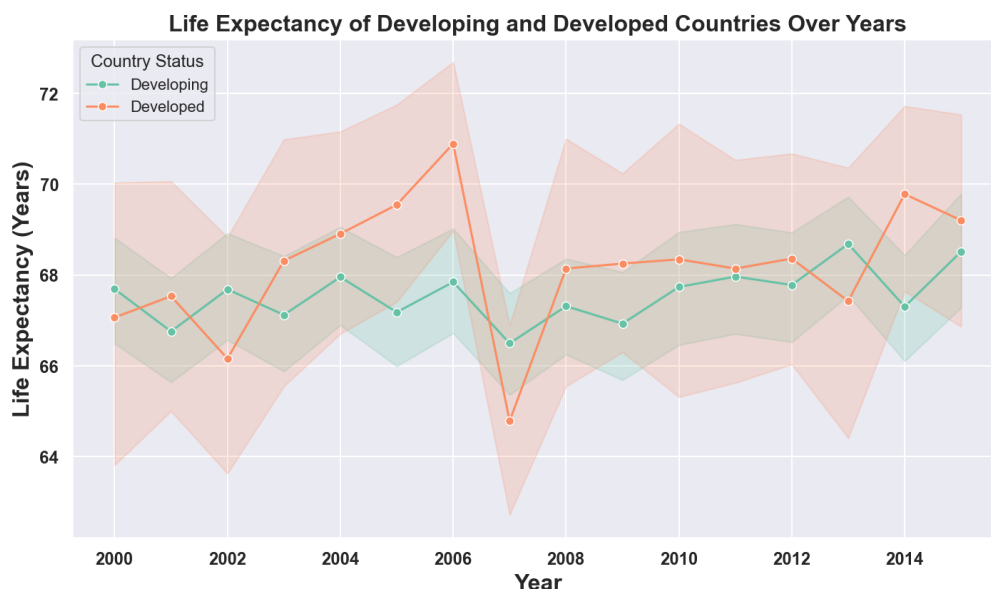
Top and Bottom 5 Countries by Life Expectancy

This pie chart shows the top and bottom 5 countries based on male life expectancy. Countries with the highest standards are shaded in green, while countries with the lowest standards are in red. Dominica tops the list, followed by Guinea-Bissau and Yemen at the bottom. This picture highlights the differences in global life expectancy.



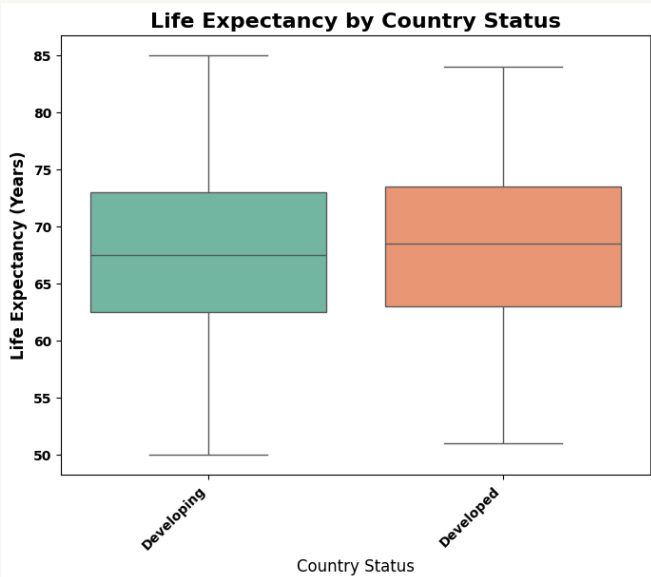
Line plot shows the life Expectant for developing and developed countries over time

Life expectancy in developed countries is consistently higher than in developing ones. Both groups show a general upward trend, indicating improvements in global healthcare and living standards. Developing countries exhibit more fluctuations, reflecting inconsistent progress and challenges in maintaining healthcare advancements.



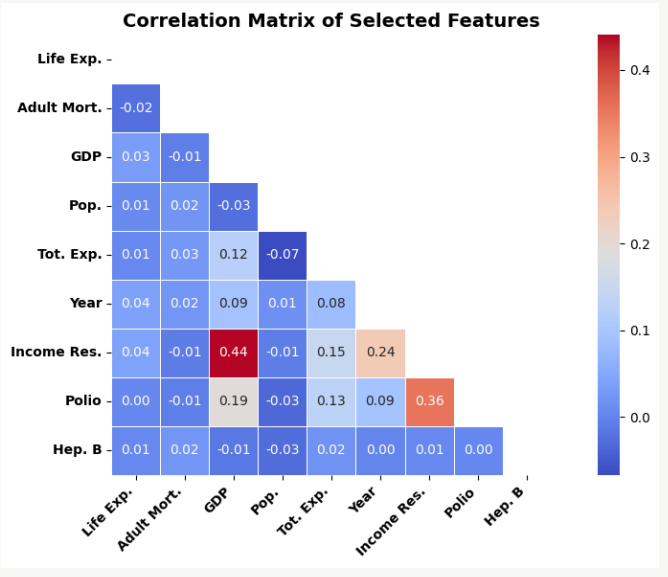
Box Plot of Life Expectancy by Country Status

- The box plot compares life expectancy for men between developed and developing countries.
- Developed countries have a higher median life expectancy for men, with fewer outliers.
- Developing countries have a lower median life expectancy, and the spread indicates significant variability, hinting at uneven healthcare and infrastructure development.



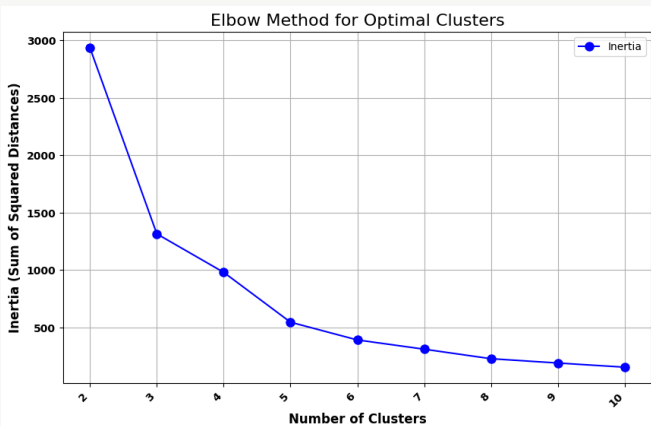
The heatmap visualizes correlations between all numerical variables.

- Life Expectancy is negatively correlated with Adult Mortality, meaning higher mortality leads to lower life expectancy.
- GDP and Income Composition of Resources show positive correlations with life expectancy, highlighting the role of economic development in improving healthcare and living standards.
- Strong correlations between health indicators like Polio, and Hepatitis B vaccination rates reveal their combined impact on life expectancy.



Elbow Method for Optimal Clusters

The Elbow Method plot shows the optimal number of clusters for K-Means. The inertia (sum of squared distances) decreases as clusters increase, but the drop slows significantly at $k=3$, the "elbow point." This suggests that 3 clusters best capture the natural grouping in the data without overfitting.



Silhouette Analysis for Optimal Clustering

The best cluster counts based on evaluation are 2 (with a silhouette score of 0.86), 3 (with a score of 0.82), and 4 (with a score of 0.75).

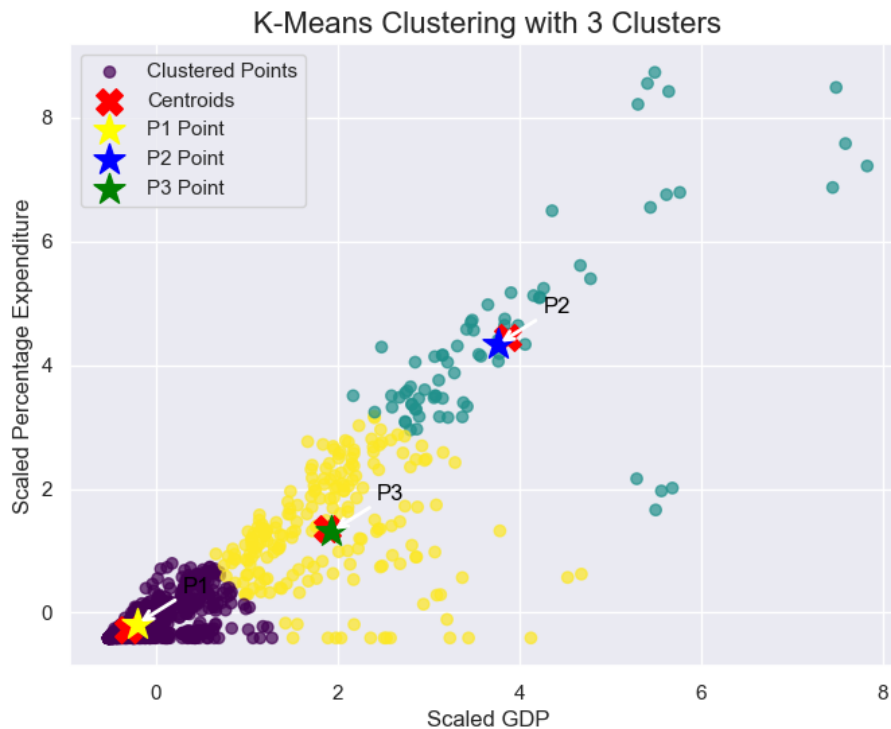
- **2 Clusters:** Provides the highest quality clusters with clear distinctions.
- **3 Clusters:** Introduces more detail while slightly reducing clustering quality.
- **4 Clusters:** Further granularity but with a noticeable drop in quality.

K-means Clustering

This scatter plot shows K-Means clusters of countries based on scaled GDP (x-axis) and scaled percentage expenditure (y-axis). Points are color-coded by cluster, with red X markers as centroids.

- **Cluster 1 (Purple):** Developing countries with low GDP and expenditure.
- **Cluster 2 (Yellow):** Transitional economies with moderate GDP and expenditure.
- **Cluster 3 (Green):** Developed countries with high GDP and expenditure.

The clusters illustrate economic progression from developing to developed nations



Linear Regression: Predicted vs Actual Values

This scatter plot compares actual values (x-axis) with predicted values (y-axis). Points near the diagonal line indicate high accuracy and minimal errors, showing the model's strong predictive performance.

