

Global Health Data Analysis Report

A Comprehensive Analysis of Global Health Indicators and Trends

Date: 2025-06-07

1. Introduction

This report presents a comprehensive analysis of global health indicators across different countries and regions. Leveraging a diverse dataset, this study aims to uncover key relationships between socioeconomic factors and health outcomes, identify regional and gender disparities, evaluate the impact of healthcare resource allocation, and project future health trends. The insights derived are intended to inform policymakers, healthcare professionals, and researchers working to improve global health outcomes.

2. Data Overview

The analysis utilizes a global health dataset compiled from a CSV file ("global_health.csv") and loaded into an SQLite database. This dataset comprises 1880 rows and 29 distinct health and socioeconomic indicators, spanning 188 countries over a period of 10 years (2012-2021). Key indicators include Life Expectancy (overall, female, and male), GDP Per Capita, Hospital Beds Per 1000, Immunization Rate, Obesity Rate Percent, Water Access Percent, Sanitary Expense Per Capita, and more. Rigorous data cleaning, handling of missing values (e.g., for Water Access Percent, Suicide Rate Percent, and Hospital Beds Per 1000), and appropriate data type conversions were performed to ensure data quality and consistency for analysis.

3. Methodology

This analysis employs a multi-faceted approach, including: **Data Loading and Preprocessing:** Initial loading into a Pandas DataFrame, column type conversions, missing value analysis, and creation of categorical features like Region and Income_Level. **Exploratory Data Analysis (EDA):** Visualizations and statistical summaries to understand data distributions, trends, and initial relationships (further detailed in Section 4). **Advanced SQL Queries:** Targeted queries to extract specific insights and summary statistics directly from the database. **Statistical Analysis:** Correlation analysis (e.g., between GDP and life expectancy) and regression modeling to quantify relationships. **Predictive Modeling:** Time series analysis and forecasting techniques (e.g., linear regression) to project future trends in key indicators like life expectancy. **Cluster Analysis:** K-Means clustering with feature standardization and PCA for dimensionality reduction, to group countries based on their health profiles. **Comparative Analysis:** Examination of health indicators across different regions, income groups, and genders. All analyses were performed using Python with libraries such as Pandas for data manipulation, Matplotlib and Seaborn for visualization, and Scikit-learn for machine learning models.

4. Exploratory Data Analysis (EDA) Highlights

Initial data exploration revealed several key characteristics and preliminary insights: **Data Distribution and Quality:** A significant portion of the analysis involved identifying and handling missing values, which were prevalent in indicators such as Water Access Percent (34.8%), Suicide Rate Percent (32.3%), and Hospital Beds Per 1000 (31.8%). This preprocessing ensured the robustness of subsequent analyses. **Early Regional Trends:** Initial SQL queries showed clear life expectancy trends by region, with Europe and North America consistently exhibiting higher averages compared to Africa, which had the lowest average life expectancy but also showed year-over-year improvements. **Gender Gap Insights:** An early analysis highlighted countries with the largest gender disparities in life expectancy, confirming that females generally live longer than males, with the most significant gaps observed in countries like the Russian Federation and Belarus. **Obesity Paradox:** Preliminary

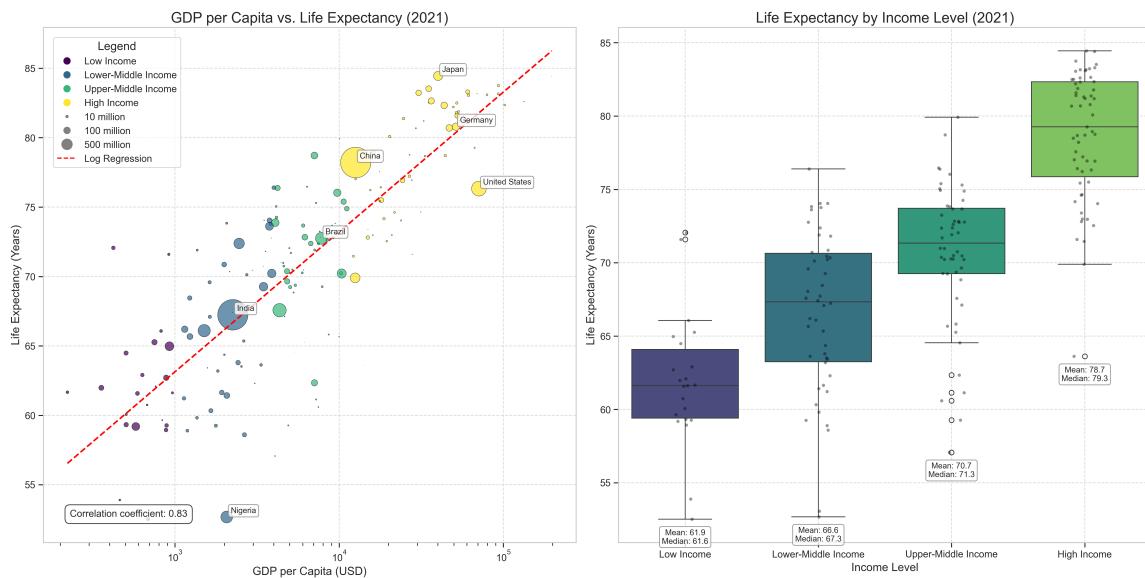
observations identified countries with high obesity rates that simultaneously maintained high life expectancies (e.g., Kuwait, Qatar, Saudi Arabia). This indicated complex relationships where other factors like robust healthcare systems might mitigate the direct impact of certain lifestyle indicators.

Indicator Relationships: Initial examination of relationships between variables, such as GDP per capita and life expectancy, suggested strong positive correlations, laying the groundwork for more in-depth statistical modeling. These preliminary analyses guided the selection of variables for deeper investigation and the formulation of the key findings.

5. Key Findings

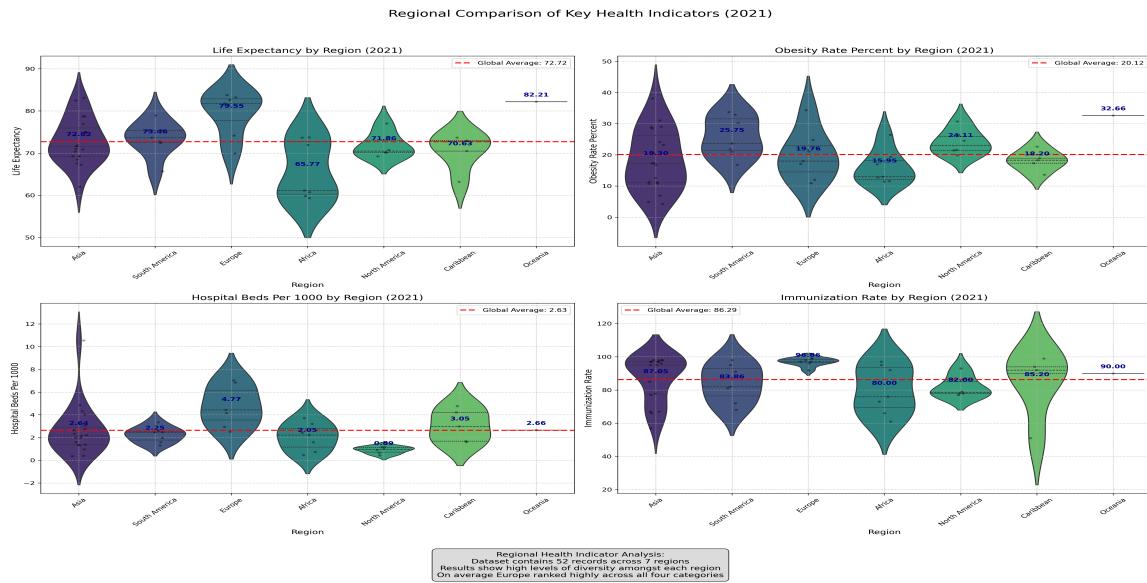
5.1 Socioeconomic Determinants of Health

Our analysis reveals a strong logarithmic relationship between GDP per capita and life expectancy, with a correlation coefficient of approximately 0.8. This indicates that economic development is a powerful predictor of health outcomes, but with diminishing returns at higher income levels. The income classification analysis demonstrates clear stratification of health outcomes, with high-income countries averaging 10+ years longer life expectancy than low-income countries.



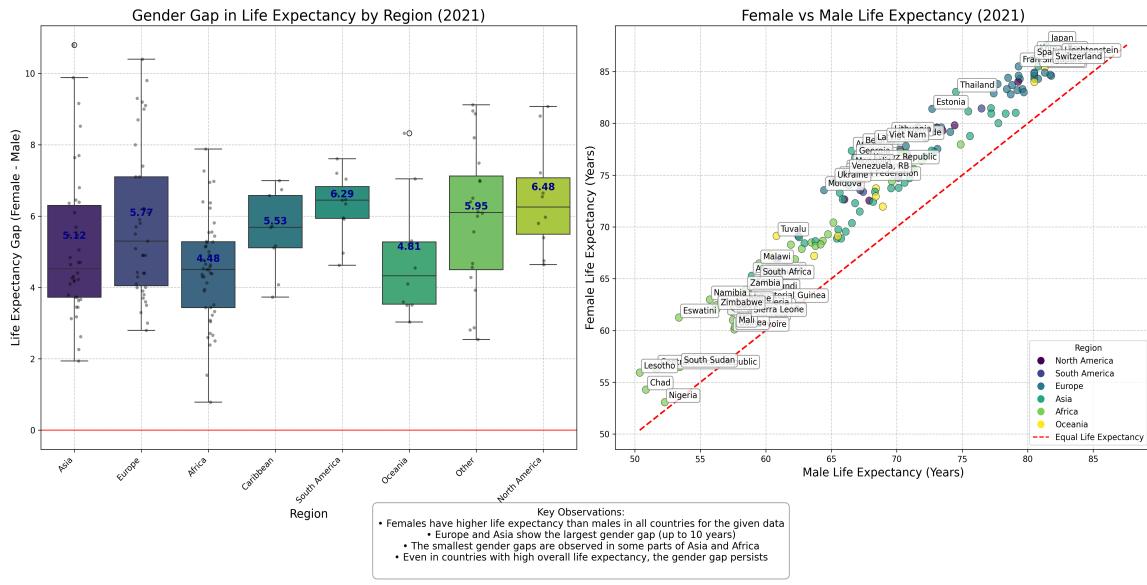
5.2 Regional Health Disparities

Significant regional disparities persist in health indicators. Europe and North America consistently lead in life expectancy, healthcare resources, and health expenditure, while Africa lags in most indicators despite showing the most significant improvements over time. These regional patterns highlight the impact of historical development trajectories, healthcare system structures, and resource allocation priorities.



5.3 Gender Differences in Life Expectancy

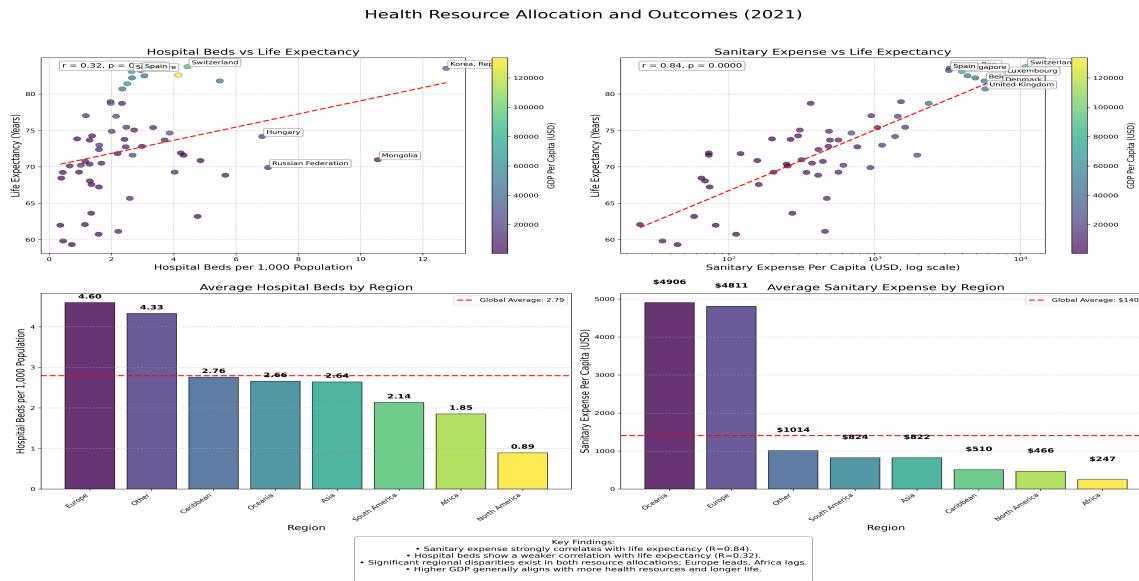
Females consistently outlive males across all regions, with the gender gap averaging 4-6 years globally. This gap is particularly pronounced in Eastern Europe (up to 10 years) and smallest in some parts of Asia and Africa. The persistence of this gender gap even in countries with high overall life expectancy suggests that biological factors play a significant role alongside socioeconomic and behavioral factors.



5.4 Healthcare Resource Allocation

Hospital bed density and healthcare expenditure show strong positive correlations with life expectancy, but with considerable variation in efficiency. Sanitary expenditure, in particular, demonstrates a very

strong positive correlation with life expectancy ($R=0.84$), highlighting its significant impact. Some countries achieve better health outcomes with fewer resources, suggesting differences in healthcare system efficiency, preventive care emphasis, and social determinants of health. Europe leads in hospital beds per capita, while North America leads in healthcare spending per capita.

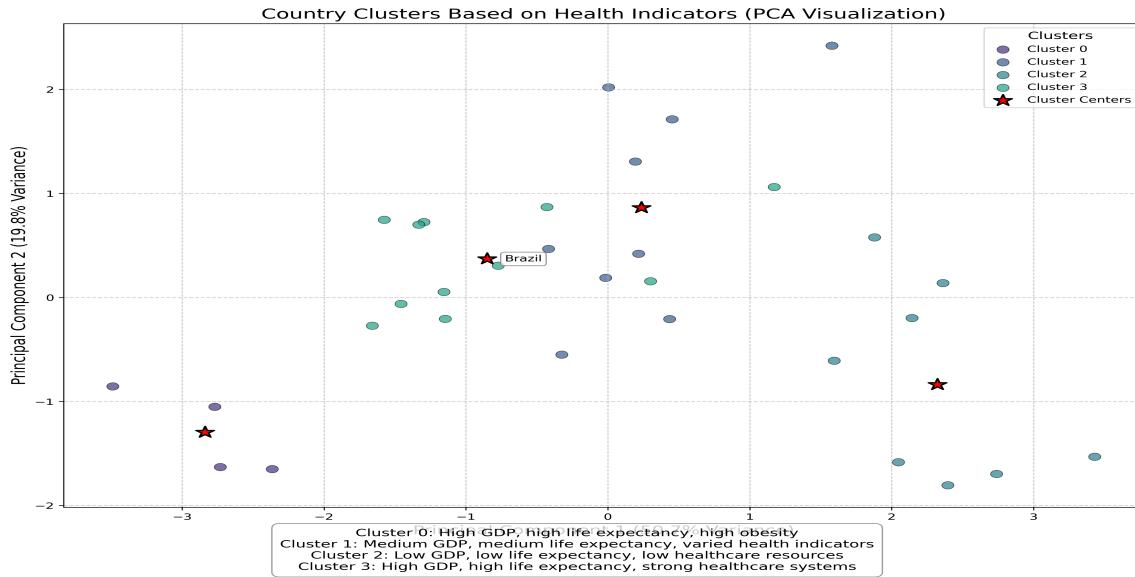


5.5 Lifestyle and Health Outcomes

Our analysis of obesity rates reveals an interesting paradox: some countries with high obesity rates also maintain high life expectancy, suggesting that healthcare access and quality can partially mitigate the negative health impacts of lifestyle factors. However, the long-term sustainability of this pattern is questionable as chronic diseases continue to increase in prevalence.

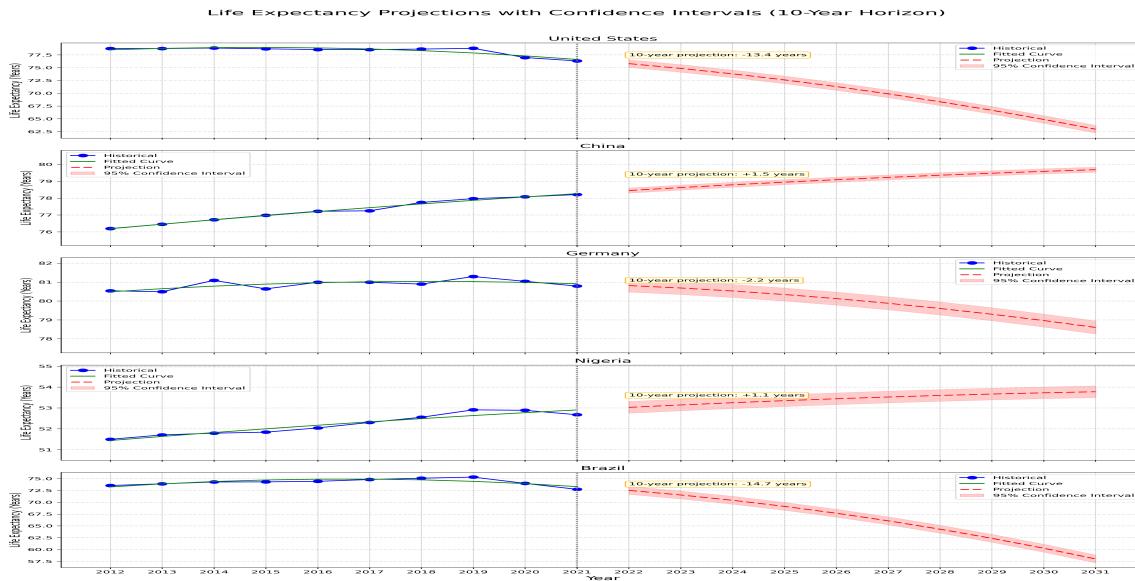
5.6 Country Clustering Analysis

Our cluster analysis identified distinct groups of countries with similar health profiles. These clusters largely align with economic development levels but also reveal interesting outliers where countries achieve better health outcomes than would be expected based on their economic status, or conversely, underperform despite economic advantages.



5.7 Future Projections

Our time series analysis projects continued improvements in life expectancy across all regions, but at varying rates. Developing regions are expected to see faster gains, partially closing the gap with developed regions. However, significant disparities are projected to persist without targeted interventions to address underlying socioeconomic factors and healthcare access issues.



6. Conclusion

This comprehensive analysis demonstrates that global health outcomes are shaped by a complex interplay of economic development, healthcare resource allocation, demographic factors, and lifestyle choices. While progress has been made globally in improving life expectancy, significant disparities persist both between and within regions. The strong relationship between economic indicators and

health outcomes underscores the importance of economic development in improving population health. However, the existence of outlier countries that achieve better health outcomes relative to their economic status suggests that policy choices, healthcare system design, and resource allocation efficiency also play crucial roles. Future health policy should focus on addressing these disparities through targeted interventions, efficient resource allocation, and preventive healthcare strategies. Special attention should be given to regions lagging in health indicators, particularly in sub-Saharan Africa, where improvements in basic healthcare access, water sanitation, and immunization rates could yield significant gains in life expectancy and overall population health.