Analysing the Impact of Life Expectancy on Healthcare Expenditure, Age Dependency Ratio, and Economic Productivity

By

Dhriti Makhija

Student ID: 2795390

Supervisor: Dr. Mubashir Ali

A thesis submitted to the University of Birmingham

For the degree of MSc in Data Science

School of Computer Science

University of Birmingham, Birmingham, UK

September 2025

DECLARATION

I hereby declare that the work presented in this project is my own, except where otherwise acknowledged.

Generative AI tools were used in a transparent and responsible manner to support my work. Specifically:

Code development was assisted by Claude Sonnet 3.7.

Debugging and optimisation were supported by ChatGPT 4.0.

Report structuring and editing for clarity, grammar, and formatting were supported by ChatGPT 4.0.

All outputs from these tools were reviewed, tested, and edited by me to ensure accuracy and to reflect my own understanding. The project's direction, structure, and contributions remain my own.

A record of Generative AI usage, including example prompts and my critical reflections, is provided in the Appendix.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor, Dr. Mubashir Ali, for his patient guidance, constructive feedback, and constant encouragement throughout this project. His insights into data science methodology and careful attention to detail shaped every stage of this dissertation.

I am grateful to the academic and support staff of the School of Computer Science, University of Birmingham, for providing an engaging learning environment and timely assistance across the MSc in Data Science programme.

My heartfelt thanks go to my friends and MSc cohort for their camaraderie, late-night debugging sessions, and thoughtful discussions that improved this work. Above all, I am deeply indebted to my family for their unwavering support, patience, and belief in me.

Any errors that remain are entirely my own.

ABSTRACT

This study investigates the socio-economic determinants of age dependency, health expenditure, and GDP per capita across high-income and low-income countries from 2004 to 2023. Using panel data from the World Bank, the analysis employs both classical econometric models—Multiple Linear Regression, Fixed Effects, Random Effects, and Hausman tests—as well as modern machine learning techniques, including XGBoost regression and LSTM neural networks, to uncover patterns and predictive relationships. The panel regression results highlight significant effects of variables such as life expectancy, fertility rate, and labour participation on the dependent indicators. The Hausman test guides model selection between fixed and random effects, strengthening model validity. Results reveal clear disparities between income groups, suggesting that policy interventions should be context-specific. Overall, this multi-method approach demonstrates the value of combining traditional econometrics with machine learning to produce robust and interpretable insights for global development planning.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

People all over the world have begun to live much longer lives not like that in the past century. The average life expectancy has increased by more than 30 years since 1950, primarily due to improvements in general living conditions, healthcare, nutrition, and medications [1]. Nowadays, it is typical for people to live into their 70s or 80s in many parts of the world, including East Asia, Europe, and Latin America. Societies have changed as a result of this significant advancement, with an increasing proportion of the population now being older. By 2050, one in six people worldwide will be over 65, up from just one in eleven in 2019 This trend is predicted to continue [2].

Rising life expectancy today represents both victory and complexity. Longer lives promise richer opportunities for later stages—expanded seasons for work, study, and mentoring across generations—but they also challenge the structures that hold societies together. Older populations tend to need denser health services, face escalating levels of chronic illness, and claim pensions over lengthening horizons. Governments, therefore, confront mounting anxieties over the durability of public health funding, particularly in regions where the age curve is inclined most steeply. Some contend that projected pressures are exaggerated, while others warn that the impending demand for in-home and institutional long-term care, linked to increasing frailty and disability, could outstrip even the most carefully envisioned budgets [3].

In addition to pressuring health systems, rising life expectancy contributes to changes in the age dependency ratio – the ratio of dependents (young and elderly) to working-age individuals. This ratio is a key determinant of a country's fiscal health and labour productivity. As the proportion of elderly rises, fewer workers are available to support public services and welfare systems, raising questions about the intergenerational equity of current economic models [4]. Traditional dependency measures may even underestimate the true burden if declining fertility and premature labour market exits reduce the effective workforce. Yet some researchers advocate a shift to "prospective" aging measures, which adjust the definition of old age in light of improving longevity and health status, suggesting that aging may be less economically burdensome than raw population ratios [5].

Beyond health and dependency, the economic effects of longer life expectancies also affect the paths of national growth. There is a complex relationship between longevity and economic performance. Longer lifespans can result in higher population growth and resource demands, particularly in developing nations, even though better health and survival rates are linked to higher labor productivity and human capital. The overall impact is contingent upon policy responses, demographic transition timing, and institutional capacity. A brief "demographic dividend" has fueled rapid growth in nations that have been able to lower fertility while maintaining high life expectancy; nevertheless, as populations age and the proportion of retirees increases, this window becomes smaller [6].

The global policy discourse is changing as countries of all income levels struggle with the effects of aging. Development strategies are becoming more and more focused on issues like how to create age-resilient economies, prolong working lives, and finance health care. It is more crucial than ever to comprehend how life expectancy, healthcare requirements, demographic pressures, and economic expansion interact in this setting. Longer lifespans are expected in the future, but their social and economic viability is not assured, so policymakers must navigate this reality.

1.1 Aim of the Study

To examine how increasing life expectancy, healthcare costs, and public concern intersect with demographic and economic indicators, with a comparative focus on low-income and high-income countries.

1.2 Objective

Primary Objective: Investigate the demographic and economic implications of rising life expectancy across countries.

Comparative Focus:

- Compare low-income and high-income nations.
- Highlight disparities and patterns.

Key Areas of Analysis:

- Healthcare costs
- Dependency ratios
- Economic productivity
- Public awareness and perception

Methodology:

• Use econometric analysis

• Analyze trend data over time

Intended Outcome:

• Generate insights to inform policy responses

• Address challenges of aging populations

• Understand shifting development dynamics due to demographic changes.

1.3 Research Questions

1. How does life expectancy impact the dependency ratio in low-income vs high-income countries?

2. What is the relationship between life expectancy and healthcare costs across income groups?

3. How does GDP relate to key demographic and socioeconomic factors such as fertility rate, life expectancy, and labour participation across different income levels?

4. How does public concern or awareness measured via Google Trends relate to demographic and economic productivity in low- and high-income countries?

CHAPTER 2
LITERATURE REVIEW

Life Expectancy and Healthcare Expenditure

Many people believe that longer life expectancies could lead to higher healthcare costs, particularly in aging societies. In general, older populations require more medical care, including long-term care and the management of chronic diseases. As the population ages in high-income countries, the percentage of GDP allocated to healthcare has been gradually rising. Research, however, indicates that the direct impact of aging on medical expenses might be less severe than is frequently thought. While population aging moderately raises acute care costs, it significantly raises long-term care expenditures because of the increasing number of older people with disabilities, according to a critical review by Wouterse et al [3]. Medical technology, rather than aging, remains the main driver of healthcare spending growth.

Moreover, the literature highlights the significance of proximity to death as a more reliable predictor of healthcare costs than age alone. Raitano's survey of end-of-life spending patterns revealed that a substantial portion of lifetime medical expenses is concentrated in the final years of life, regardless of chronological age [7]. This implies that increases in life expectancy might not necessarily lead to proportional increases in health expenditures—especially if longevity gains are accompanied by better health in old age.

In developing economies, the relationship between health expenditure and life expectancy can differ significantly. A study in Bangladesh by Zaman et al. found no statistically significant relationship between increased life expectancy and total health expenditure over a 10-year period. Instead, GDP was the primary driver of health spending, suggesting that economic development, rather than demographic shifts, governs public health budgets in lower-income countries [8].

Life Expectancy and Age Dependency Ratios

Global demographic structures are changing as a result of rising life expectancy and falling fertility rates. Because of this, the old-age dependency ratio (OADR), which measures the proportion of people 65 and older to people of working age, is continuously increasing. The United Nations projects that by 2050, one in six people worldwide will be 65 years of age or older, up from one in eleven in 2019 [1].

Social services, labor markets, and pension systems are all under a lot of strain as a result of this demographic shift.Nonetheless, some academics contend that traditional measures of dependency might exaggerate the impact of aging populations. Instead of using a set chronological threshold, Sanderson and Scherbov's Prospective Old-Age Dependency Ratio (POADR) defines "old age" dynamically based on remaining life expectancy [9]. According to related research, POADR can stay constant or even decrease with longevity, whereas conventional OADR increases, suggesting that people are not only living longer but also continuing to be healthier and more productive for longer periods of time [5].
However, developing nations face unique challenges. Many people are rapidly aging at relatively low income levels, which means they have less time and money to adjust. The need for structural changes to healthcare and social support systems is increased by the World Health Organization's prediction that 80% of older people will live in low- and middle-income nations by 2050 [2]. In order to sustain growth momentum as dependency ratios increase, Lee and Mason stress that nations going through demographic transition must make early adjustments to their pension systems and invest in human capital [4].

Life Expectancy and Economic Growth

Another important factor in the economy's long-term growth is life expectancy. Increased investment in labor force participation, savings, and human capital—all crucial factors in productivity and growth—usually results from longer life expectancy. A ten-year increase in life expectancy would increase annual per capita GDP growth by roughly 0.5 to 0.6 percentage points, according to Barro's groundbreaking cross-country analysis[10].
However, the effects vary by location. Although many international health initiatives from the middle of the 20th century increased life expectancy, Acemoglu and Johnson demonstrated that these efforts did not result in a corresponding rise in per capita income. This was partly because the improvements in longevity caused the alleged Malthusian effects of population growth, which crowded out capital accumulation, particularly in nations that had not yet completely moved away from the demographic state Pon Jeremy discussed [11].
A more sophisticated framework is provided by Cervellati and Sunde, who suggest that the growth impact of life expectancy is contingent upon a nation's demographic transition stage. Increased life expectancy may slow per capita growth in pre-transition societies with high fertility; in post-transition societies with lower fertility, longevity encourages

saving and education, which boosts economic growth[12].

The economies of East Asia are prime examples of the advantages of matching economic strategy with demographic trends. A brief period of low dependency ratios brought about by rising life expectancy and declining fertility allowed nations like South Korea and Singapore to enjoy rapid economic growth thanks to an increase in labor supply and savings—known as the "demographic dividend"[13].

Lastly, life expectancy's behavioral effects have an impact on economic development as well. Increased longevity encourages forward-thinking behaviors that support long-term growth, while higher adult mortality shortens planning horizons and deters investments in savings, entrepreneurship, and education [14].

CHAPTER 3
METHODOLOGY

Using a panel dataset of nations from 2004 to 2023, this study uses a multi-method approach to analyze the effects of rising life expectancy on important socioeconomic outcomes. In order to compare demographic trends across various levels of development, the analysis separates high-income and low-income nations (based on GNI per capita). The World Bank World Development Indicators (WDI) were used to compile the dataset, guaranteeing uniformity and cross-national comparability. The key variables in the dataset include:

- GDP per Capita: Gross domestic product per person, measured in constant international dollars (purchasing power parity). GDP per capita serves as the dependent variable (after log transformation) and an overall indicator of economic development.
- Life Expectancy at Birth (years): The average number of years a newborn is expected to live if current mortality patterns persist. This reflects health outcomes and is treated as a main independent variable.
- Fertility Rate (births per woman): The total fertility rate, i.e. the number of children a woman would bear over her lifetime given age-specific fertility rates of the period. This demographic variable is included as a predictor, as fertility can influence economic growth and age structure.
- Labor Force Participation Rate (percent): The proportion of the working-age population (typically ages 15 and above) that is economically active (either employed or actively seeking work). These variable captures labor market activity and human capital utilization.
- Health Expenditure (percent of GDP): Total health care expenditure (public and private) expressed as a percentage of GDP. Health spending is a social development indicator that may correlate with economic performance and life expectancy.

These variables were chosen to represent diverse aspects of development: economic output (GDP per capita), health outcomes (life expectancy), demographic trends (fertility), labor supply (participation rate), and public investment in health (health expenditure). The combination provides a comprehensive view of socioeconomic factors in the panel dataset.

3.1 Data Preprocessing

The data were meticulously preprocessed before modeling to guarantee consistency and analysis suitability. Among the crucial feature engineering and preprocessing steps were:
• Data Cleaning: To maintain panel balance, we eliminated country-year observations with insufficient data or, when necessary, used imputation for sporadic gaps. To make analysis easier, variable names were standardized (e.g., by using consistent lower-case and converting spaces to underscores). Additionally, we checked the scales and units of each indicator, converting currency to constant dollars or percentages to fractional form as necessary for consistency.

3.2 Experimental Evaluation: EDA and Feature Engineering

To better understand the underlying patterns and prepare the dataset for modelling, the following experimental evaluations were carried out:

- Exploratory Data Analysis (EDA):
  - Correlation matrices and scatter plots helped identify initial associations between variables.
  - Time-series plots of life expectancy and health expenditure revealed trends and outliers over time.
- Feature Engineering:
  - Creation of `log_gdp_per_capita` to stabilize variance.
  - Standardization was considered but ultimately omitted to preserve interpretability in regression coefficients.

3.3 Modelling

We used a range of modelling approaches to gain a deeper understanding of the intricate relationship between socioeconomic and demographic indicators and development outcomes in both high-income and low-income nations. These included sophisticated machine learning algorithms and conventional econometric panel data models. We were able to triangulate insights and evaluate both linear relationships and more complex, data-driven patterns by employing a variety of methodologies.

1. Multiple Linear Regression (MLR)

As a starting point, we used Multiple Linear Regression (MLR). It assumes that all observations are independent and that there is no unobserved heterogeneity across nations, treating the dataset as pooled cross-sectional data. This simple model aids in establishing preliminary correlations between variables. However, MLR may produce biased estimates because it ignores country-specific factors like geography or cultural norms that may not change over time.In order to determine important predictors, the model's performance was evaluated using R^2 to gauge explanatory power and statistical significance of coefficients. The outcomes offered a standard by which more sophisticated models could be evaluated.

2. Fixed Effects (FE) Model

We used the Fixed Effects model to account for unobserved, time-invariant heterogeneity between nations. FE makes the assumption that every nation has distinct traits that are constant over time and may have an impact on results. This model effectively lets each nation act as its own control by estimating the impact of the explanatory variables using within-country variation. In many policy-oriented studies, FE is a more robust approach because it is especially helpful when those unobserved country-specific factors are correlated with the predictors.We evaluated the significance and cross-country stability of the coefficients as well as the model's performance using R^2 (within variation).

3. Random Effects (RE) Model

The Random Effects model, as opposed to FE, makes the assumption that the explanatory variables and unobserved country-specific factors are uncorrelated. When its assumptions are met, this model is statistically more efficient and takes into account both within- and between-country variation. When we wish to incorporate time-invariant variables into the model, RE is particularly useful. The estimates could become skewed, though, if the RE assumptions are broken. R^2 and the statistical significance of the predictors were used to evaluate the model fit.

4. Hausman Test

To determine whether Fixed Effects or Random Effects is more appropriate, we conducted the Hausman test. This statistical test checks whether the unobserved country-specific effects are correlated with the explanatory variables.

5. XGBoost Regression

We employed XGBoost, a potent gradient boosting algorithm, to capture possible non-linear relationships and interactions. This model was used independently for both income groups and works especially well with tabular data. We used R^2, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to assess the model after training it on historical data. Partial Dependence Plots (PDPs) gave us information about the marginal effect of each feature, and feature importance rankings helped us determine which variables had the greatest predictive influence.

6. LSTM Neural Network

Lastly, we used a neural network with Long Short-Term Memory (LSTM), a deep learning model designed specifically for sequence data. LSTM enabled us to take into consideration long-term trends and temporal dynamics because our dataset covers several years for every nation. A series of historical economic and demographic factors were used to train the model to forecast life expectancy. R^2 and RMSE were used to assess LSTM, which offered a sophisticated perspective on patterns that conventional models might overlook, particularly those pertaining to time-based dependencies.

CHAPTER 4
EXPERIMENTAL EVALUATION

4.1 Overview of Dataset

The dataset compiled by the World Bank contains annual socioeconomic and demographic indicators for both high- and low-income countries from 2004 to 2023. The main variables used in the analysis are:

- Life Expectancy
- Age Dependency Ratio
- Fertility Rate
- Labour Participation
- Healthcare Expenditure
- GDP per Capita
- Population

During the data cleaning process, missing values were removed entirely to ensure the model input remained consistent throughout all stages of analysis.

4.2 Summary Statistics

Based on a dataset with 2,179 observations from high- and low-income countries between 2004 and 2023, Table 1 shows the summary statistics for the key variables used in the analysis. These variables include GDP, life expectancy, healthcare spending, fertility rate, labor force participation, and age dependency ratio.

Variable	**Mean**	**Std Dev**	**Min**	**Median**	**Max**
Life Expectancy	70.01	17.76	0	74.83	85.53
GDP	19,475	23,180	0	8,908	133,711
Health Expenditure (%)	5.76	3.37	0	5.66	21.83
Age Dependency	54.96	17.09	17.3	51.22	110.29
Fertility Rate	2.39	1.38	0.59	1.88	7.78
Labour Participation	61.57	10.32	33.71	61.74	88.35
Population	59.4M	∼181M	∼0.27M	∼15.9M	1.4B

Table 4.1: Summary of key variables in the dataset.

4.3 Correlation Analysis

A correlation heatmap was created to identify potential multicollinearity among the predictors.

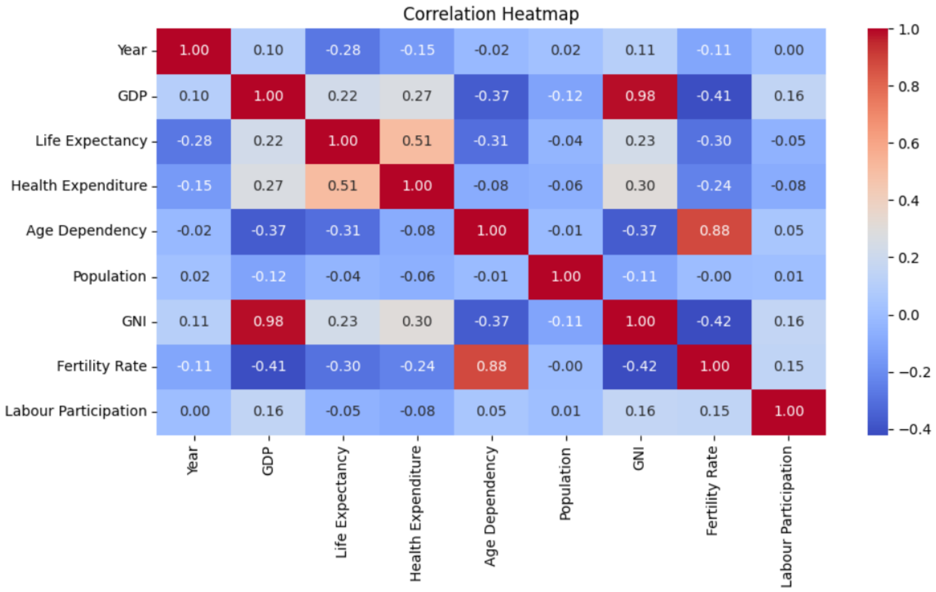


Figure 4.1: Correlation heatmap between various variables

Strong positive correlations were observed between:

- Fertility Rate and Age Dependency
- Health Expenditure and Life Expectancy

Strong negative correlations were observed between:

- Fertility Rate and Life Expectancy
- GDP and Age Dependency
- Fertility Rate and GDP

4.4 Visual Exploration of Key Trends

The notebook included multiple visualizations of line plot over time

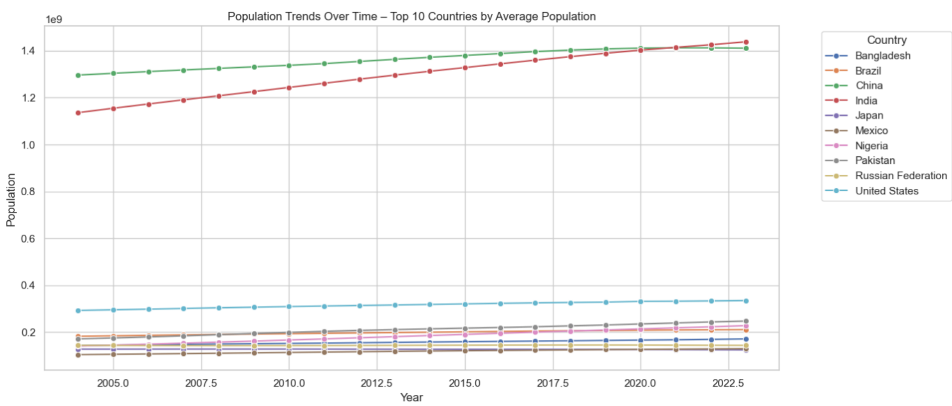


Figure 4.2: Population trend for top 10 countries

The two countries with the largest populations were China and India.In 2022, India surpassed China. Nigeria's population grew quickly, while the United States grew steadily and Russia and Japan remained stable.

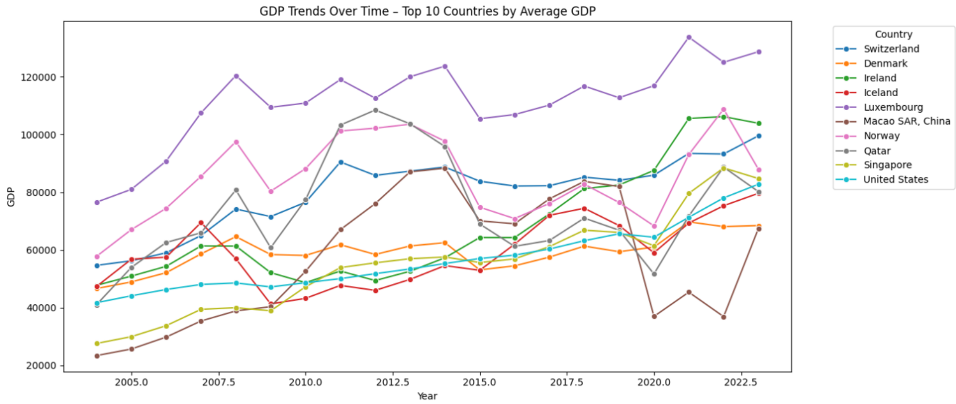


Figure 4.3: GDP trend for top 10 countries over time

Monaco's GDP per capita was consistently the highest and continued to rise in 2022. Despite earlier robust growth, Luxembourg experienced an unusual decline in 2022, falling to the bottom among top performers.

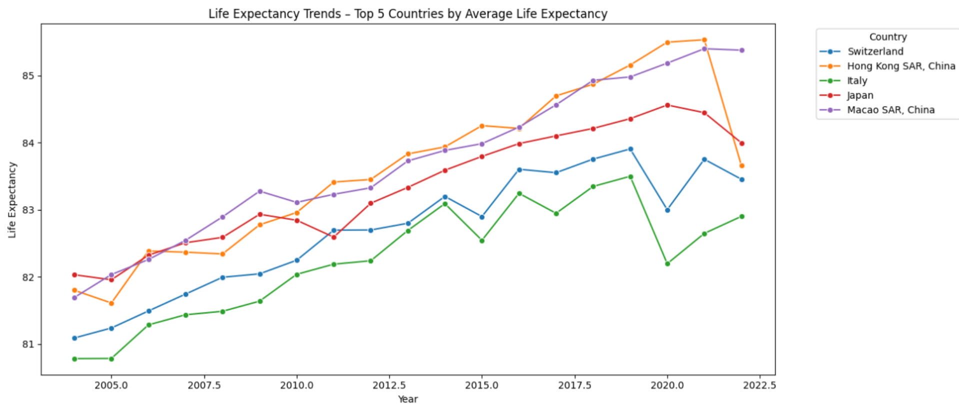


Figure 4.4: Life Expectancy trend for top 5 countries

Between 2004 and 2019, life expectancy increased steadily, with Macao SAR and Hong Kong SAR leading the way. Although long-term trends are still positive, a post-2019 decline, particularly in Japan and Hong Kong SAR, is probably due to COVID-19.

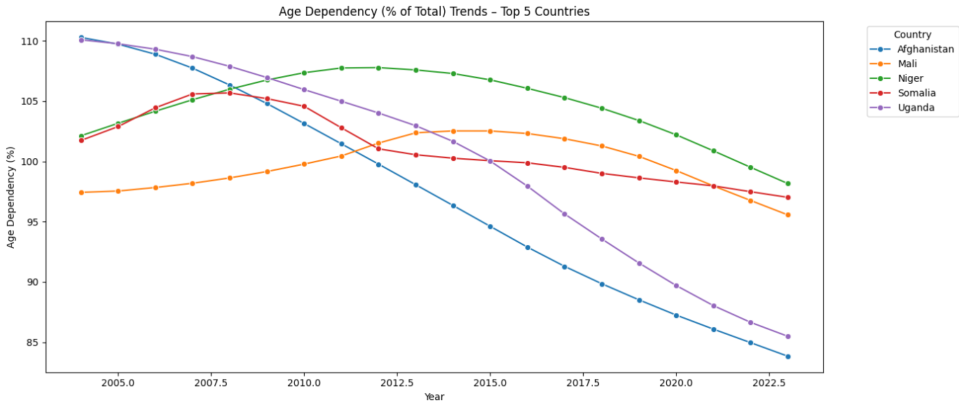


Figure 4.5: Age Dependency trend for top 5 countries

With fewer dependents compared to the workforce, age dependency generally decreased in the top 5 countries.Mali initially rose, then gradually declined, maintaining a high dependency level. Afghanistan showed the steepest decline, moving from the highest to the lowest by 2022.

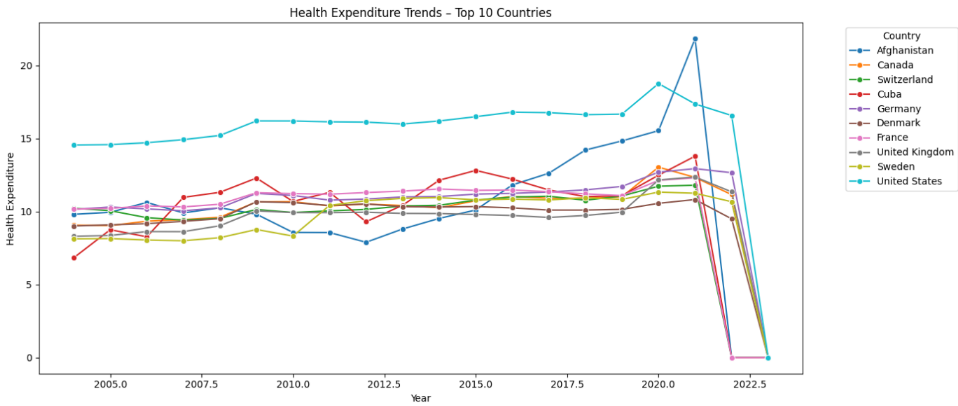


Figure 4.6: Health Expenditure Trend for top 10 countries

Prioritizing health spending, the United States reached its highest level in 2021. Afghanistan briefly increased in 2021 before declining. Prior to the sharp post-COVID declines in 2022–2023, most countries experienced steady growth until 2021.

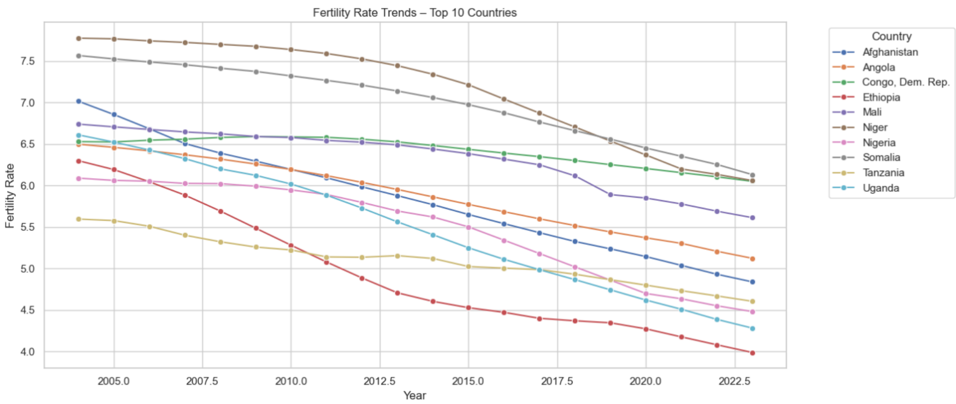


Figure 4.7: Fertility Rate trends for top 10 countries

Between 2004 and 2023, fertility rates in the top 10 countries decreased. Ethiopia saw the largest decline, reflecting a rapid demographic shift, while Niger remained at the top but is also seeing a decline. Most other countries improved gradually due to better family planning and healthcare.

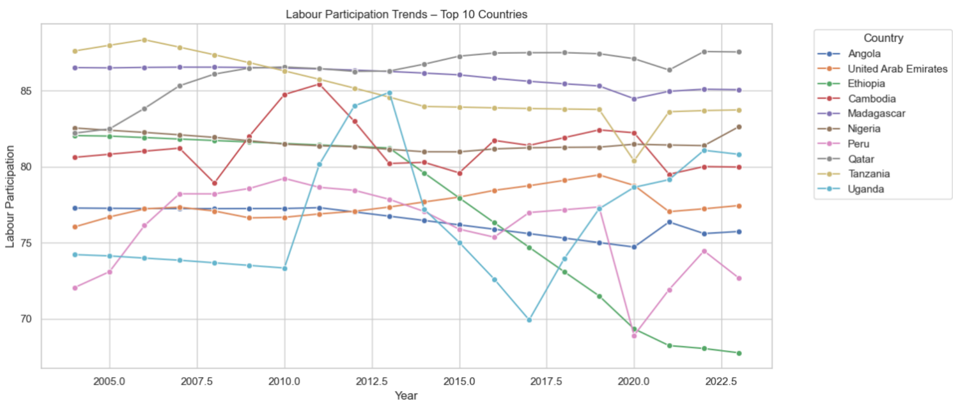


Figure 4.8: Labour Participation trend for top 10 countries

Madagascar and Qatar maintained labor participation above 85% between 2004 and 2023.In response to labor market shocks and demographic changes, Ethiopia dropped from 80% to just under 70%, while Uganda and Cambodia peaked around 2010 before declining.

4.5 Feature Engineering

To enhance model interpretability and performance, a range of domain-informed feature engineering techniques were applied. These steps helped transform raw variables into more meaningful predictors and allowed the study to capture key demographic and economic dimensions affecting health expenditure and economic development.

A. Constructed Variables

• Total Health Expenditure: Calculated as a percentage of GDP, representing the total health spending at the macroeconomic level:

Total Health Expenditure = (Health Expenditure / 100) × GDP

• Health Expenditure per Capita: Normalized health spending across countries by population size:

Health Expenditure per Capita = Health Expenditure / Population

• GDP Growth: Year-over-year percentage growth was derived for country-level GDPs using pct_change() grouped by country.

• Log Transformations: Applied log_gdp and log_health_expenditure using natural logarithms to reduce skewness and improve model robustness.

D. Dependency Ratios and Population Structure

Derived variables using estimated shares of population age groups:

• Youth Dependency Ratio = Population aged 0–14 ÷ Population aged 15–64

This measures the number of young dependents per working-age person.

• Old Age Dependency Ratio = Population aged 65+ ÷ Population aged 15–64

This shows how many elderly dependents there are for every working-age individual.

• Total Dependency Ratio: Sum of youth and old-age ratios.

• Share Variables: Computed shares of youth, working-age, and elderly cohorts with respect to total population.

These demographic constructs offer a refined understanding of how population age structure affects national resource distribution and dependency burdens.

E. Income Group Classification

Countries were classified as "High Income" or "Low Income" based on the 2024 World Bank GNI threshold of $12,265. A binary indicator was generated accordingly:

```
if gni >= 12265:
return "High income"
else:
return "Low income"
```

This segmentation was crucial in modelling heterogeneous effects across income tiers and aligning with the comparative objectives of the study.

CHAPTER 5
RESULTS AND DISCUSSIONS

The results of our empirical analysis are presented in this section, which is organized around the three main research questions and a Google Trends analysis. We used appropriate econometric models, Multiple Linear Regression, and panel regression techniques, such as Fixed Effects (FE) and Random Effects (RE) models, for each question in order to assess the connections between economic and demographic indicators in high- and low-income nations. The Hausman test helped choose the model by determining whether FE or RE was a better fit. To increase prediction accuracy and identify possible non-linear patterns, we also used XGBoost regression. In addition, we used neural networks with Long Short-Term Memory (LSTM).

5.1 Q1. How does increasing life expectancy impact the dependency ratio in low- and high-income countries?

Regression Equation:

To explore this relationship, we specified the following panel regression model:

$$\begin{aligned} \text{DependencyRatio}_{it} &= \beta_0 + \beta_1 \times \text{LifeExpectancy}_{it} \\ &+ \beta_2 \times \log(\text{GDP}_{it}) + \beta_3 \times \text{FertilityRate}_{it} \\ &+ \beta_4 \times \text{LabourParticipation}_{it} + \alpha_i + \varepsilon_{it} \end{aligned} \tag{5.1}$$

Where:

- i and t index country and time, respectively
- α_i represents unobserved country-specific effects.
- ε_{it} is the idiosyncratic error term.

Variables	MLR	Fixed Effects	Random Effects	XGBoost	LSTM
Intercept	53.8986(0.000)***	35.099 (6.9285)***	39.342 (6.4274)***	–	–
Life Expectancy	-0.0523	-0.0350 (0.0061)***	-0.0367 (0.0061)***	✓	✓
Log GDP per Capita	4.7393(0.000)***	4.6564 (0.6336)***	4.5303 (0.5978)***	✓	✓
Fertility Rate	5.1448(0.000)***	1.0001 (0.6916) (ns)	1.1687 (0.6609)*	✓	✓
Labour Participation	-0.9718	-0.5688 (0.0591)***	-0.6111 (0.0551)***	✓	✓
R-squared		0.1591	0.2851	0.742	
RMSE		–	–	0.3745	
MSE		–	–	21.2948	
Hausman Test	$\chi^2 = 9.13$, p = 0.1039 → Random Effects preferred			–	–

Table 5.1: Econometric Parameters wrt. Age Dependency for High-income countries

Variables	MLR	Fixed Effects	Random Effects	XGBoost	LSTM
Intercept	45.3434(0.000)***	54.006 (3.2055) ***	50.254 (2.8504)***	–	–
Life Expectancy	-0.0288	-0.0128 (0.0057)***	-0.0165 (0.0056) ***	✓	✓
Log GDP per Capita	-1.8994	-1.7951 (0.2273) ***	-1.7953 (0.2190) ***	✓	✓
Fertility Rate	10.4589(0.000)***	8.9880 (0.2786)***	9.4967 (0.2319) *	✓	✓
Labour Participation	0.0273(0.024)***	-0.0753 (0.0386) *	-0.0321 (0.0298) (ns)	✓	✓
R-squared		0.5717	0.6884	0.979	
RMSE		–	–	0.3855	
MSE		–	–	7.1346	
Hausman Test	$\chi^2 = 16.80$, p = 0.0049 → Fixed Effects preferred			–	–

Table 5.2: Econometric Parameters wrt. Age Dependency for Low-income countries

- *** indicates significance at 1%, ** at 5%, * at 10%, and (ns) = not significant.
- ✓ in XGBoost and LSTM columns indicates the variable was included.
- LSTM results are detailed in the appendix, as per model type.

Econometric Results

High-Income Countries

Dependency was significantly and consistently negatively affected by life expectancy.With a coefficient of -0.037, which indicates that dependency decreases by 0.04 units for every additional year of life expectancy, the Random Effects model (Hausman test p = 0.10) was selected. This aligns with data showing that longer-living populations help maintain economic balance by delaying retirement and increasing labor force participation.

Low-Income Countries

The impact was less pronounced. While improvements in survival reduce dependency, fertility patterns continue to be the main driver of demographic pressure, according to the Fixed Effects model (preferred, Hausman p = 0.005), which estimated a coefficient of approximately -0.013 to -0.017. Overall, the effect of longevity is more noticeable in wealthier economies, but due to ongoing demographic changes, it remains relatively small in poorer nations.

Overall, in richer economies, the impact of longevity on dependency is more significant, while in poorer settings, it remains modest due to continuous demographic shifts.

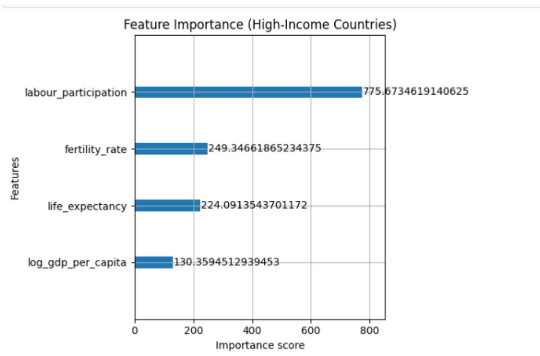
Machine Learning Insights

XGBoost outperformed econometric models, especially in low-income areas, where the R^2 value reached 0.98. Feature importance indicated that:

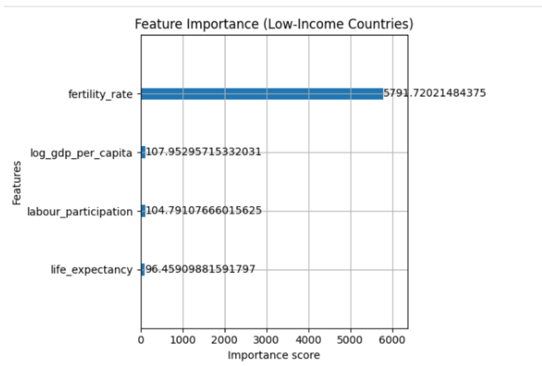
High-income: Labour participation was the most significant factor, followed by fertility rates and life expectancy.

Low-income: Fertility had the strongest influence on dependency, while life expectancy had the least impact.

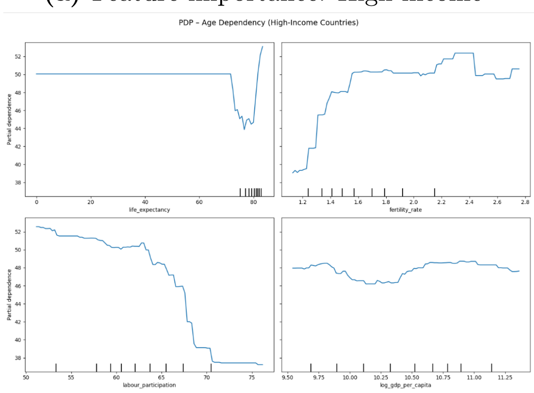
Partial dependence plots supported these findings: in high-income groups, longer life expectancy slightly reduced dependency, although this effect decreased with age. In contrast, life expectancy had little effect on dependency in low-income settings.



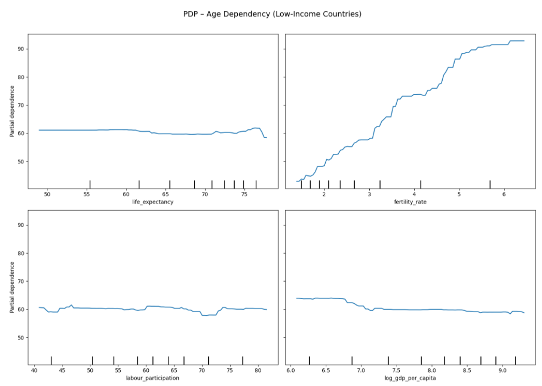
(a) Feature importance: High-income



(b) Feature importance: Low-income



(c) PDP: High-income



(d) PDP: Low-income

Figure 5.1: Feature importance and partial dependence plots for Q1 across high- and low-income countries

Scatter and Temporal Trends

Scatterplots showed variation: while poorer countries exhibited more pronounced negative slopes as early survival gains reduced child dependency, richer countries showed weaker, policy-driven variations. This difference was confirmed by time-series visualizations, which showed that while low-income countries' demographic burden more closely followed fertility declines than gains in longevity, high-income countries steadily increased life expectancy without increasing dependency.

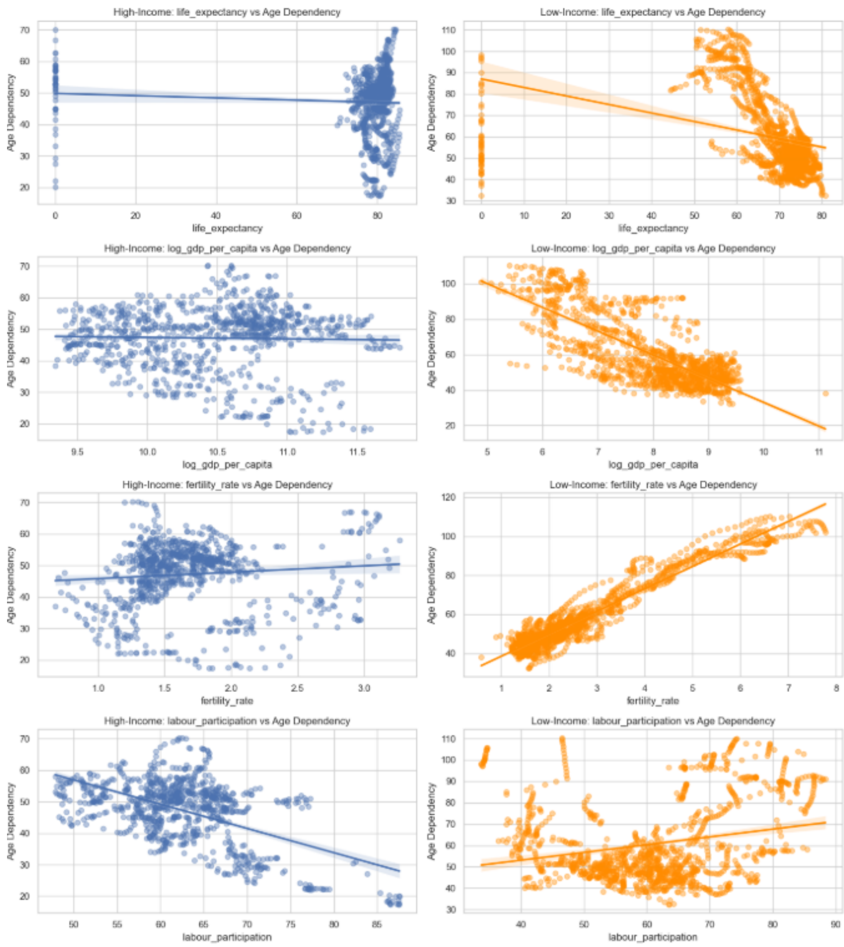


Figure 5.2: Scatter plot of age dependency with other economic factors

Long Short-Term Memory (LSTM) neural network

The LSTM model performed poorly (negative R^2 across groups) and was unable to account for inter-country variation and policy effects while still capturing broad temporal patterns. This suggests that in the absence of richer contextual variables, sequential deep learning techniques are less effective.

Discussion

Overall, life expectancy reduces age dependency; however, this effect is stronger in high-income nations and weaker in low-income ones, where fertility remains the main factor.This result highlights how demographic dividends vary by stage of development and supports the first research hypothesis.

5.2 Q2. Relationship between life expectancy and health expenditure in high- and low-income countries?

The purpose of this analysis is to assess the relationship between health spending and life expectancy across income brackets. Health spending serves as a stand-in for national investments in healthcare services and infrastructure, which have a direct effect on population longevity.

We modelled the relationship using the following panel regression structure:

$$\begin{aligned} \text{Life Expectancy}_{it} = \beta_0 + \beta_1 \times \text{Health Expenditure}_{it} \\ + \beta_2 \times \log(\text{GDP}_{it}) + \beta_3 \times \text{FertilityRate}_{it} \\ + \beta_4 \times \text{LabourParticipation}_{it} + \alpha_i + \varepsilon_{it} \end{aligned} \tag{5.2}$$

Variables	MLR	Fixed Effects	Random Effects	XGBoost	LSTM
Intercept	-6.552	8.5686 (3.7199)**	–	–	–
Life Expectancy	0.0935(0.000)***	0.0902 (0.0034)***	0.0924 (0.0033)***	✓	✓
Log GDP per Capita	0.6094(0.002)***	-0.8276 (0.3353)**	-0.0483 (0.1806) (ns)	✓	✓
Fertility Rate	0.0638(0.798)	0.1575 (0.3800) (ns)	0.2807 (0.3344) (ns)	✓	✓
Labour Participation	–	–	-0.0045 (0.0283) (ns)	✓	✓
R-squared	0.225	0.5019	0.5591	0.6348	0.0667
RMSE	–	–	–	2.308	0.1757
MSE	–	–	–	5.3269	–
Hausman Test	–	$\chi^2 = 8.27$, p = 0.0407 → Fixed Effects preferred	–	–	–

Table 5.3: Econometric Parameters wrt. Health Exp. For Low Income Countries

Variables	MLR	Fixed Effects	Random Effects	XGBoost	LSTM
Intercept	-0.9464	-2.8625 (1.3146)**	–	–	–
Life Expectancy	0.0775(0.000)***	0.0672 (0.0033)***	0.0700 (0.0032)***	✓	✓
Log GDP per Capita	0.104(0.268)**	0.0859 (0.1319) (ns)	0.0013 (0.0742) (ns)	✓	✓
Fertility Rate	-0.0422	0.8953 (0.1576)***	0.2883 (0.1082)**	✓	✓
Labour Participation	–	–	-0.0073 (0.0120) (ns)	✓	✓
R-squared	0.222	0.2989	0.4531	0.4992	0.083
RMSE	–	–	–	1.6782	0.128
MSE	–	–	–	2.8165	–
Hausman Test	–	$\chi^2 = 30.86$, p = 0.0000 $\rightarrow$ Fixed Effects preferred	–	–	–

Table 5.4: Econometric Parameters wrt. Health Exp. For High Income Countries

Econometric Results

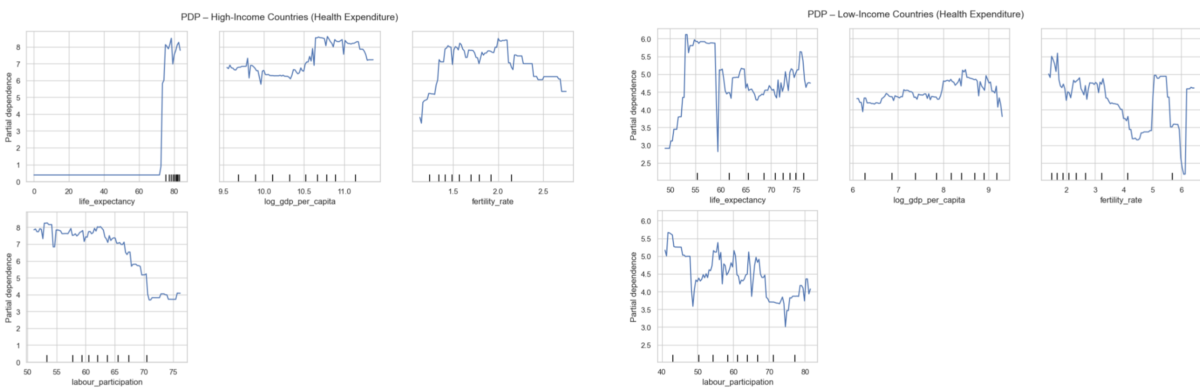
Health spending was positively and significantly correlated with longer life expectancy in both income groups.

High-income countries

A coefficient of 0.09 was estimated by the Fixed Effects model (preferred, Hausman $\chi^2 = 30.9,\ p < 0.001$), indicating that life expectancy increases by 0.09 years for every unit increase in health spending. Nevertheless, the effect reaches a plateau at high levels, indicating that advanced systems have diminishing returns.

Low-income countries

A smaller but still positive coefficient (0.07) was discovered by the Fixed Effects model ($\chi^2 = 8.3,\ p = 0.04$). Here, gains persisted consistently without showing signs of saturation.



(a) PDP: High-income countries **(b)** PDP: Low-income countries

Figure 5.3: PDPs for Q2 (health spend and life expectancy).

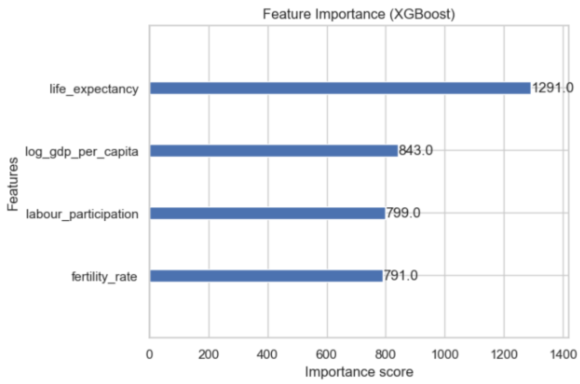


Figure 5.4: Feature importance for Q2.

Machine Learning Insights

In terms of predictive accuracy, XGBoost performed better than regressions ($R^2 = 0.63$ for high-income, 0.50 for low-income). Life expectancy was the most important feature in both groups, followed by GDP, fertility, and labor participation.

Partial dependence plots confirmed these dynamics:

Benefits flattened after 7–8 units of expenditure in high-income nations.

The slope stayed upward throughout the range in low-income nations, indicating consistent returns on investment.

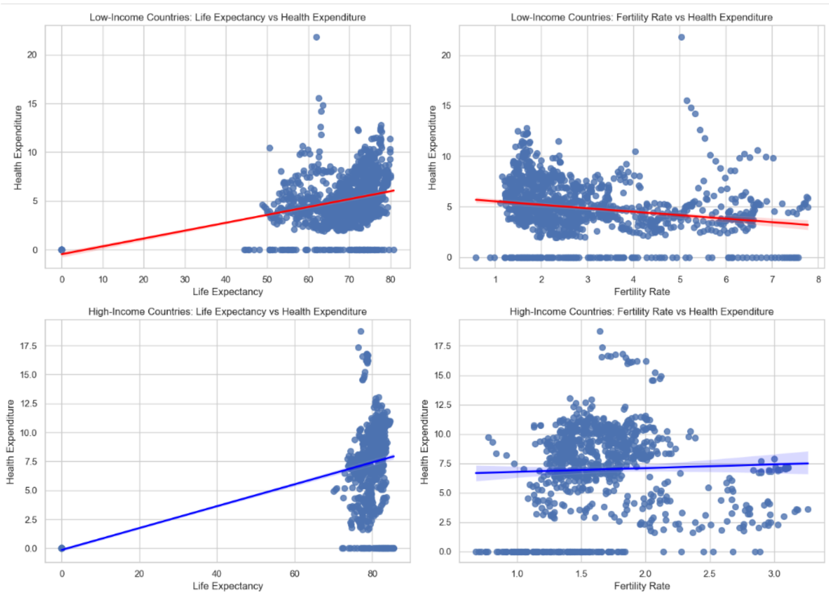


Figure 5.5: Scatter plot of health expenditure with other economic factors

Scatter and Temporal Trends

Scatterplots reinforced these results:

With only slight increases in life expectancy, high-income nations tended to spend heavily.

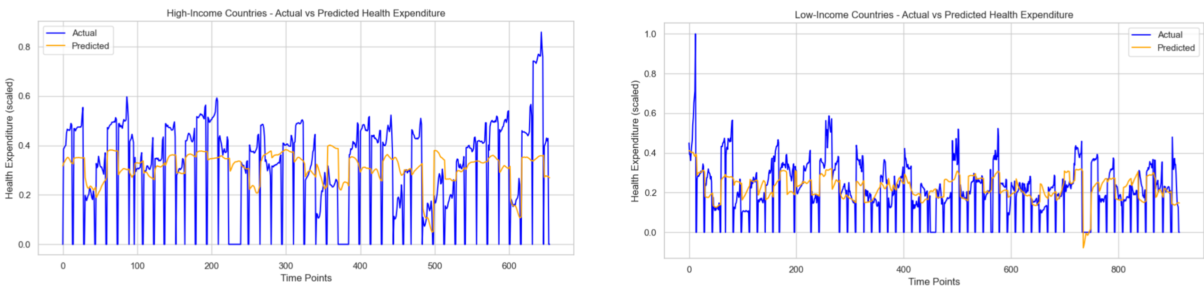
Although some outliers attained relatively high longevity at modest cost, low-income countries displayed more pronounced linear increases.

Long Short-Term Memory (LSTM) neural network

Although it was only able to explain a moderate amount of variance (R2 = 0.12–0.24), the LSTM was able to capture broad spending patterns. High-income nations with erratic spending patterns had larger prediction errors, whereas low-income nations displayed more consistent alignment. According to forecasts, India, Nigeria, and South Africa will continue to have lower spending levels than the US, UK, and Germany.

Dataset	**RMSE**	R^2
High-Income Countries	0.170	0.124
Low-Income Countries	0.117	0.236

Table 5.5: Performance Metrics for LSTM



(a) High-Income Countries **(b)** Low-Income Countries

Figure 5.6: Actual vs Predicted plot for health expenditure across high-income and low-income countries

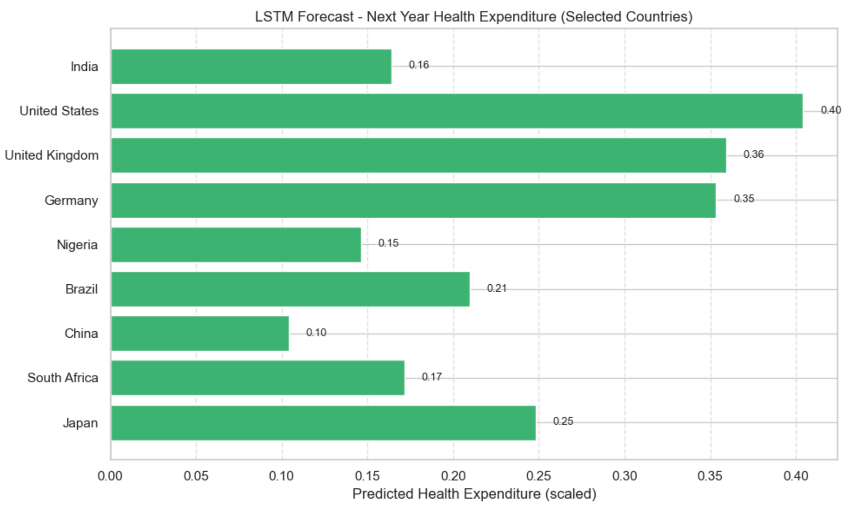


Figure 5.7: Country-Level Forecast Plot

Discussion

Although the results vary, the evidence consistently shows that health spending increases longevity everywhere.

High-income nations, like the US, have diminishing marginal benefits; spending more money does not always translate into longer life expectancies.

Low-income nations: substantial, linear benefits; even modest increases in investment result in notable advancements (e.g., Afghanistan, Nigeria).

This research points to a crucial policy implication: while efficiency gains may be more important than higher spending in wealthier systems, incremental health spending is very cost-effective in extending life expectancy in resource-constrained settings.

5.3 Q3. How does GDP relate to key demographic and socioeconomic factors such as fertility rate, life expectancy, and labour participation?

This analysis looks into how much a nation's economic performance, as indicated by log GDP per capita, is influenced by important demographic and socioeconomic indicators, such as life expectancy, labor force participation, and fertility rate. Together, these variables show workforce participation, population dynamics, and public health elements that are essential to sustained economic growth.

In order to account for both cross-country and over-time variation, we model this relationship using panel data regression, controlling for unobserved country-specific characteristics:

$$\begin{aligned}\log(\text{GDP per Capita})_{it} = {} & \beta_0 + \beta_1 \times \text{Fertility Rate}_{it} \\ & + \beta_2 \times \text{Life Expectancy}_{it} \\ & + \beta_3 \times \text{Labour Participation}_{it} \\ & + \alpha_i + \varepsilon_{it} \qquad (5.3)\end{aligned}$$

This framework allows us to assess how demographic structure and labour market engagement jointly influence economic performance, and how these relationships differ across income groups.

Variables	MLR	Fixed Effects	Random Effects	XGBoost	LSTM
Intercept	8.5726(0.000)***	9.0061 (0.2119)***	–	–	–
Fertility Rate	-0.1376	-0.1342 (0.0370)***	0.4582 (0.0548)***	✓	✓
Life Expectancy	0.0015(0.085)	-0.0018 (0.0003)***	-0.0010 (0.0005)***	✓	✓
Labour Participation	0.0319(0.000)***	0.0289 (0.0030)***	0.1328 (0.0023)***	✓	✓
R-squared	0.170	0.1814	0.8337	0.668	-0.088
RMSE	–	–	0.3200	0.2507	–
MSE	–	–	0.1024	–	–
Hausman Test	–	χ^2 = 2119.99, p = 0.0000 → Fixed Effects preferred	–	–	–

Table 5.6: Econometric Parameters wrt. Log GDP per Capita. For High Income Countries

Variables	MLR	Fixed Effects	Random Effects	XGBoost	LSTM
Intercept	9.3431(0.000)***	9.9073 (0.2917)***	–	–	–
Fertility Rate	-0.4770	-0.4396 (0.0333)***	-0.4337 (0.0421)***	✓	✓
Life Expectancy	0.0019(0.160)	-0.0014 (0.0007)*	0.0018 (0.0010)†	✓	✓
Labour Participation	-0.0022	-0.0096 (0.0049)†	0.1157 (0.0034)***	✓	✓
R-squared	0.529	0.1608	0.5090	0.8317	0.0570
RMSE	–	–	0.4159	0.2210	–
MSE	–	–	0.1730	–	–
Hausman Test	–	χ^2 = 1117.62, p = 0.0000 → Fixed Effects preferred	–	–	–

Table 5.7: Econometric Parameters wrt. Log GDP per Capita. For Low Income Countries)

Econometric Results

High-income countries

FE preferred, χ^2 = 2119.99, p < 0.001

Fertility rate had a significant negative ($\beta = -0.1342$, $p < 0.01$).

Labour participation showed a positive impact ($\beta = 0.0289, p < 0.01$).

Life expectancy had a small negative coefficient ($\beta = -0.0018$, $p < 0.01$), reflecting aging population pressures.

Low-income countries

FE preferred, χ^2 = 1117.62, p < 0.001

Fertility rate was again a strong negative predictor ($\beta = -0.4396$, p < 0.01).

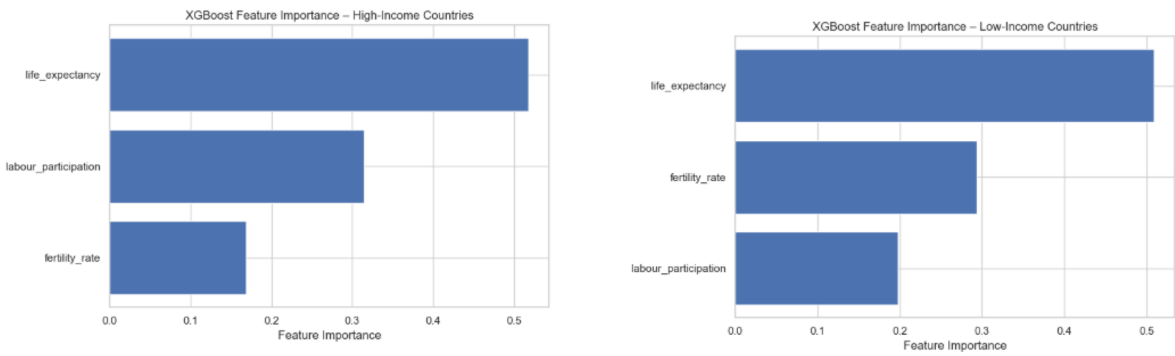
Life expectancy showed a marginally negative effect ($\beta = -0.0014$, $p \approx 0.05$).

Labour participation had a slight negative impact ($\beta = -0.0096$, $p \approx 0.05$), possibly due to underemployment/informal sectors.

Machine Learning Insights

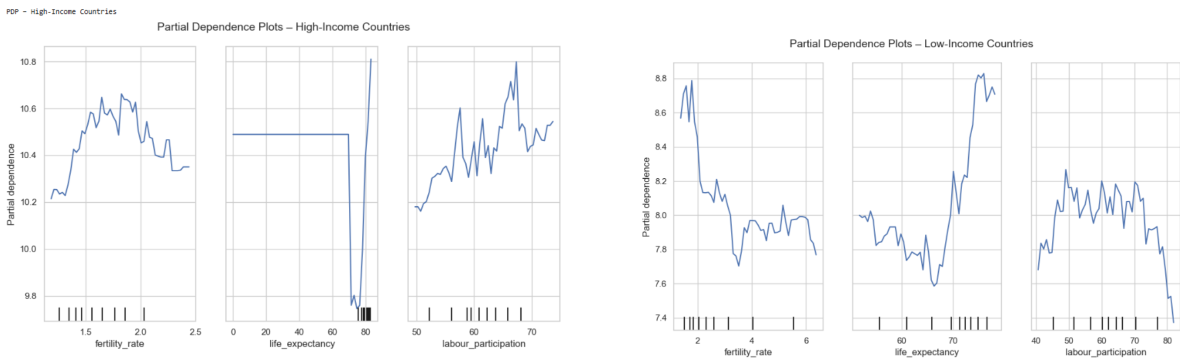
Metric	High-Income	Low-Income
R-squared	0.668	0.8317
RMSE	0.3200	0.4159
MSE	0.1024	0.1730

Table 5.8: XGBoost Performance



(a) High-income countries **(b)** Low-income countries

Figure 5.8: Feature importance plots for Q3 across income groups



(a) High-income countries **(b)** Low-income countries

Figure 5.9: PDP plots of log GDP per capita with other variables across income groups

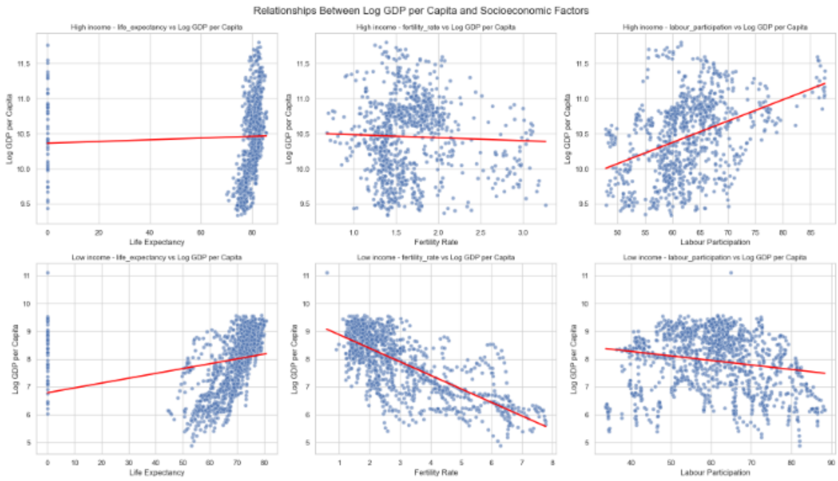


Figure 5.10: Scatter plot of log gdp per capita with other economic factors

Scatter and Temporal Trends

High-income: fertility was negatively linked to GDP, labour positively linked, life expectancy flat/negative.

Low-income: fertility again negative, life expectancy more positive, labour weakly negative.

Long Short-Term Memory (LSTM) neural network

- Full Dataset: RMSE = 0.2338, R^2 = –0.0099
- High-Income Countries: RMSE = 0.2507, R^2 = –0.0885
- Low-Income Countries: RMSE = 0.2228, R^2 = 0.0394

Low-income countries: modest predictive value ($R^2 = 0.04$).

High-income countries: near-zero explanatory power ($R^2 < 0$), suggesting GDP depends on broader macroeconomic/technological drivers not captured here.

Discussion

In both groups, the fertility rate continuously limited GDP, but this effect was most pronounced in low-income countries. In high-income economies, labor participation undoubtedly increased growth; however, because of structural employment problems, it was less successful in low-income contexts. The effects of life expectancy varied, being somewhat positive in poorer countries (health and productivity gains) and negative in wealthy ones (aging burden).

5.4 Q4. How does public concern or awareness measured via Google Trends relate to demographic and economic productivity?

This section explored the connection between economic and demographic factors and public awareness, as measured by Google Trends. The pytrends API was used to collect the normalized search interest index (ranging from 0 to 100) from Google Trends. We monitored search terms such as "retirement age," "elderly care," "healthcare costs," and "aging population" across different countries and years. Only countries with at least one full year of complete trend data were included in the preprocessed dataset to ensure consistency in comparisons. For correlation analysis, these search interest indices were paired with socioeconomic indicators, including GDP, fertility rate, life expectancy, labor participation, age dependency, and healthcare spending. Only correlations with a magnitude of more than |0.4| were shown to focus on meaningful patterns.

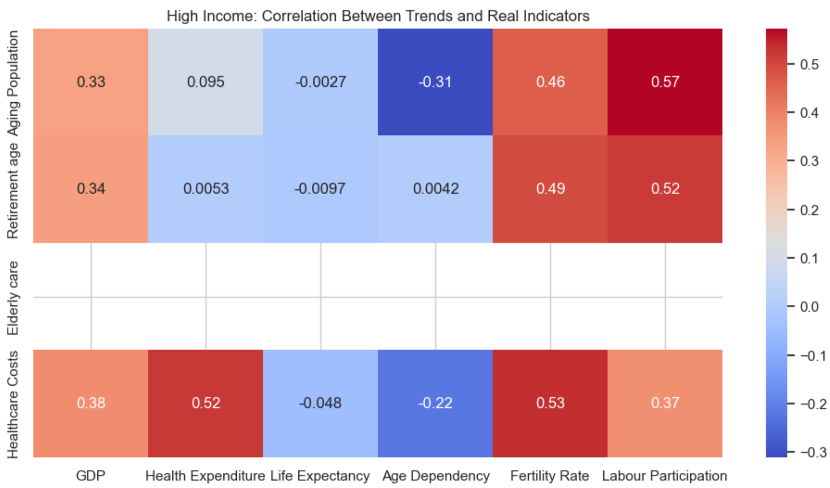


Figure 5.11: Correlation plot of Google trend for high income

High-Income Countries

- **Healthcare Costs:**
  - Positive correlation with both Health Expenditure ($r \approx 0.52$) and Fertility Rate ($r \approx 0.53$).
  - This implies that growing public anxiety reflects real increases in family planning and health care expenditures.
- **Aging Population & Retirement Age:**

  - There is a moderate correlation between GDP ($r \approx 0.33$) and Labour Participation ($r \approx 0.57$).
  - Retirement age and labour participation have a comparable positive correlation ($r \approx 0.52$).
  - These trends show that policy-driven workforce engagement—like extended retirement policies in aging societies—corresponds with increased public concern.

- **Interpretation:**
  - Digital awareness appears to be proactive and policy-aligned in wealthy countries; public concern is not merely reactive but rather reflects planned demographic and economic shifts.

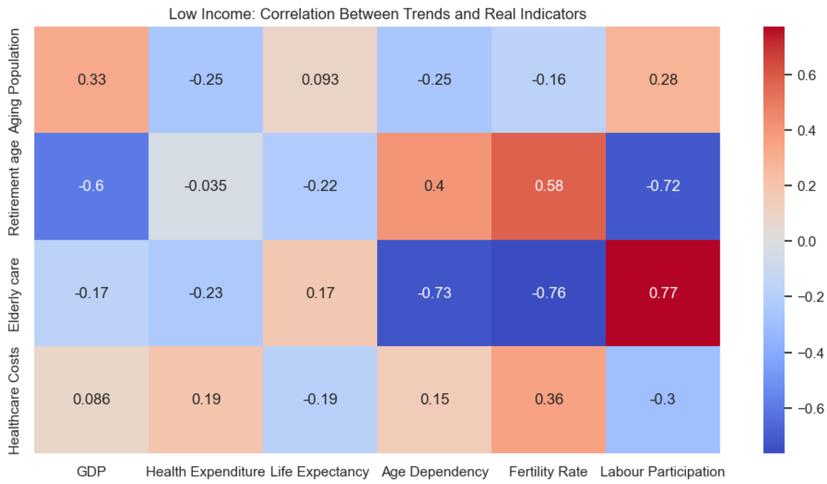


Figure 5.12: Correlation plot of Google trend for low income

Low-Income Countries

- **Elderly Care:**
  - Fertility Rate ($r \approx -0.76$) and Age Dependency ($r \approx -0.73$)have strong negative correlations with elder care.
  - Strong positive correlation with Labour Participation ($r \approx 0.77$).
  - This reflects changing family dynamics, including declining fertility and an increase in concerns about elder care.

- **Retirement Age:**
  - Shows negative correlations with GDP ($r \approx -0.60$) and Labour Participation ($r \approx -0.72$).
  - Instead of institutional retirement systems, public anxiety in these contexts may be a reflection of shaky economic security and informal employment arrangements.
- **Healthcare Costs:**
  - Weak correlations with economic indicators.
  - Demonstrates that although there is concern, healthcare spending and access are still either restricted or dependent on outside sources (such as aid).
- **Interpretation:**
  - Reactive patterns are seen in low-income countries in contrast to wealthier ones; social pressures (urbanization, deteriorating family networks) drive public concern rather than policy alignment.

5.5 Comparing Models: Econometrics, XGBoost, and LSTM

The project struck a balance between performance and interpretability. While XGBoost produced strong predictive power, capturing non-linearities and achieving high R2, econometric models offered causal insights and variable-level interpretation. LSTM added temporal analysis, performing well for gradual demographic shifts in low-income countries but poorly in high-income contexts where short-term dynamics depend on more intricate, non-demographic factors.

5.6 Bringing It All Together

A number of themes come to light when combining the findings from visual analysis, machine learning, and regression:

• The most important demographic factor influencing age dependency, GDP, and health is fertility rate.

• The role of life expectancy is more complicated; it is positively correlated with longevity and, to a lesser extent, with GDP, but its influence lessens as other development thresholds are reached.

CHAPTER 6
FUTURE SCOPE

This study investigated the relationships between population, health, and economic factors in various nations. Even though the insights were significant, the analysis could still be expanded and new perspectives could be investigated. Future research could expand on this study in the following ways:

A) Add More Indicators

The focus was mainly on key variables like GDP, life expectancy, fertility, labour force participation, and health spending. But countries are also influenced by:

- Education levels
- Unemployment rates
- Inflation
- Political stability
- Environmental and climate-related events

Including these could provide a more comprehensive understanding of what propels or hinders national development.

B) Try More Advanced Models

Although some models, such as LSTM, did not perform well in this instance, their accuracy may increase with improved tuning or more thorough data. Future initiatives might attempt:

- Better-designed neural networks
- Time series analysis using ARIMA models
- Models of causal inference to more clearly examine cause-and-effect relationships

C) Break Down by Region or Income

The analysis could be broadened beyond merely contrasting "rich vs. poor" nations by examining:

- Regions (e.g., Africa, Asia, Europe)
- More income categories (low, lower-middle, upper-middle, high)

This would make it easier to find more focused tactics for various groups.

CHAPTER 7
CONCLUSION

The purpose of this study was to investigate the relationship between a nation's economic prosperity and important health and demographic variables, including life expectancy, the number of children born, employment rates, and healthcare spending.

Main Takeaways:

• Investments in health are crucial, particularly in low-income nations where even modest advancements can result in longer lifespans.

• In poorer nations, high fertility puts pressure on the economy. Birth rates can be lowered to promote growth and population balance.

• In wealthier nations, aging populations present a problem. It will be crucial to maintain the effectiveness of healthcare systems and assist older adults in continuing to work.

• Productivity is increased by labor participation, particularly when paired with robust educational and infrastructure systems.

In order to describe current trends, comprehend deeper patterns, and provide future predictions, this project integrated machine learning tools with conventional regression models. These studies can contribute to the development of policies that promote long-term, sustainable development for people of all income levels with improved data and careful analysis.

CHAPTER A

APPENDIX A: GENERATIVE AI PROMPTS

Table A.1: Examples of Generative AI Prompts Used in This Dissertation

Stage of Work	Example Prompt	How It Was Used
Literature Search	"Summarise key studies on the relationship between life expectancy and healthcare expenditure in low and high-income countries."	Helped me identify themes and references. Verified sources independently from the World Bank and academic papers.
Writing(Clarity)	"Rephrase this paragraph to improve clarity while keeping technical accuracy."	Improved grammar and readability. Edited outputs before including.
Code Debugging	"Why am I getting a key error in Python when merging two dataframes by country and year?"	AI suggested likely syntax fixes. I tested and modified code myself to ensure understanding.
Report Structuring	"Suggest a logical chapter structure for a dissertation analysing life expectancy, healthcare expenditure, and GDP."	Provided ideas for organisation. I adapted and finalised the structure myself.
Formatting	"Generate a proper APA reference for this World Bank dataset."	Quick formatting check. Double-checked all references manually.

CHAPTER B
APPENDIX B: GITLAB REPOSITORY

The full project (code, data, and dashboards) is hosted on GitLab: `https://git.cs.bham.ac.uk/dxm491/impact_of_increasing_life_exp_on_diff_factors`

Repository contents

1. `Final_Project.ipynb` — main Jupyter notebook with the analysis and figures.
2. `final_dataset.xlsx` — primary dataset used in the analysis.
3. `dashboard.twb` — Tableau workbook containing the interactive dashboards developed for visualization..
4. `README.md` — basic project instructions.

(Open `dashboard.twb` in Tableau Desktop to view the interactive dashboards.)

Runtime notes

- Python 3 environment (e.g., 3.8+). Use Jupyter Notebook/Lab to run `Final_Project.ipynb`.
- Install the libraries listed in the notebook/README before execution.
- Tableau Desktop required to open `dashboard.twb`.

References

[1] United Nations. *World Population Prospects 2019*. United Nations, 2019. URL: https://population.un.org/wpp/.

[2] World Health Organization. *Decade of Healthy Ageing: Baseline report*. WHO, 2021. URL: https://www.who.int/publications/i/item/9789240017900.

[3] B. Wouterse, J. Polder, and M. Koopmanschap. "The effect of population aging on health expenditure growth: a critical review". In: *Health Economics* (2017). URL: https://link.springer.com/article/10.1007/s10433-013-0280-x.

[4] R. Lee and A. Mason. "Fertility, human capital, and economic growth over the demographic transition". In: *European Journal of Population* 26.2 (2010), pp. 159–182. DOI: 10.1007/s10680-009-9186-x.

[5] W. Sanderson and S. Scherbov. "Are we overly dependent on conventional dependency ratios?" In: *Population and Development Review* 41.4 (2015), pp. 687–708. DOI: 10.1111/j.1728-4457.2015.00091.x.

[6] D. E. Bloom, D. Canning, and J. Sevilla. *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change*. RAND Corporation, 2003. URL: https://www.researchgate.net/publication/322427863_The_Demographic_Dividend_A_New_Perspective_on_the_Economic_Consequences_of_Population_Change.

[7] M. Raitano. *The Impact of Death-Related Costs on Healthcare Expenditures: A Survey*. Tech. rep. ENEPRI Research Report No. 17. European Network of Economic Policy Research Institutes, 2006. URL: https://www.ceps.eu/ceps-publications/impact-death-related-costs-health-care-expenditure-survey/.

[8] S. B. Zaman et al. "An association of total health expenditure with GDP and life expectancy". In: *Journal of Medical Research and Innovation* 1.2 (2017), AU7–AU12. DOI: 10.5281/zenodo.576546.

[9] D. M. Ediev, W. C. Sanderson, and S. Scherbov. "The inverse relationship between life expectancy-induced changes in the old-age dependency ratio and the prospective old-age dependency ratio". In: *Theoretical Population Biology* 125 (2019), pp. 1–10. DOI: 10.1016/j.tpb.2018.10.001.

[10] R. J. Barro. *Determinants of Economic Growth: A Cross-Country Empirical Study*. Tech. rep. NBER Working Paper No. 5698. National Bureau of Economic Research, 1996. URL: https://www.nber.org/papers/w5698.

[11] D. Acemoglu and S. Johnson. "Disease and development: The effect of life expectancy on economic growth". In: *Journal of Political Economy* 115.6 (2007), pp. 925–985. DOI: 10.1086/529000.

[12] M. Cervellati and U. Sunde. "Life expectancy and economic growth: The role of the demographic transition". In: *Journal of Economic Growth* 16.2 (2011), pp. 99–133. URL: https://www.iza.org/publications/dp/4160/life-expectancy-and-economic-growth-the-role-of-the-demographic-transition#:~:text=Improvements%20in%20life%20expectancy%20do,and%20after%20the%20demographic%20transition..

[13] D. E. Bloom and J. G. Williamson. "Demographic transitions and economic miracles in emerging Asia". In: *The World Bank Economic Review* 12.3 (1998), pp. 419–455. DOI: 10.1093/wber/12.3.419.

[14] P. Lorentzen, J. McMillan, and R. Wacziarg. "Death and development". In: *Journal of Economic Growth* 13.2 (2008), pp. 81–124. DOI: 10.1007/s10887-008-9029-3.