

Hanoi University of Science and Technology
School of Information and Communication Technology



Introduction to Business Analytics

Life Expectancy Analysis

Nguyen Quang Minh 20204884

Le Van Hoang 20204911

Ly Nhat Nam 20204886

Luong Hoang Anh 20204868

Supervisor: TS. Nguyen Binh Minh

Contents

I. Introduction	3
II. Problem Statement	4
III. Data Preparation	4
1. Data overview	4
2. Data preprocessing	5
IV. Analysis and Findings	7
1. GDP - Life expectancy correlation	7
1.1 Overview of GDP per capita – life expectancy correlation:	7
1.2 GDP per capita – Life expectancy correlation for developed countries:	8
1.3 GDP per capita – Life expectancy correlation for developing countries:	9
2. General overview for each GDP value and life expectancy	9
2.1 Top 5 countries with the highest average GDP value:	9
2.2 Top 5 countries with the highest average life expectancy value:	10
2.3 Average and median GDP and life expectancy of countries by year (2000-2015):	11
2.4 Distribution of Average GDP between developed and developing countries (2000-2015).	12
2.5 World map showing the average value of GDP and life expectancy of countries around the world (2000-2015):	14
3. Life Expectancy Influence	14
3.1. Overview	14
3.2. Alcohol	15
3.3. Schooling	15
3.4. Total Expenditure	16
3.5. HDI on income resource	16
3.6. BMI	17
3.7. Children deaths	18
3.8. Vaccination	18
3.9. Health Expenditure	19
4. Life Expectancy Trends Across Countries and Over Time	20
4.1. Overall Global Trend:	20
4.2. Country-Specific Trends in Life Expectancy	21
V. Conclusion	23

I. Introduction

An essential indicator of a country's health, life expectancy considers a confluence of social, political, economic, and environmental aspects in addition to the standard of medical care provided. Policymakers, healthcare providers, and analysts must all have a thorough

understanding of the dynamics of life expectancy as the global landscape changes.

This report aims to investigate the complex relationship between economic success and life expectancy, with a particular focus on the relationship between GDP per capita and life expectancy in various geographic regions. It is predicted that life expectancy will rise with the development of modern medicine and the internationalization of health programs, possibly coinciding with economic expansion. However, given the wide range of variables that can affect health outcomes, this premise merits careful examination.

Through the utilization of the extensive dataset made available by the World Health Organization (WHO), which includes a multitude of data on GDP per capita and life expectancy in different nations and areas, our goal is to visualize and analyze these metrics in order to identify trends, patterns, and insights. Our goal in conducting this study is to provide important insights that will help future business analysts create plans that advance global prosperity and health.

In this report, we use PowerBI as a tool to visualize and analysis about life expectancy data, the original file can be viewed at : <https://github.com/katakad/Intro-to-BA>

II. Problem Statement

This analysis's main goal is to understand how life expectancy and economic variables relate to one another, with a focus on GDP per capita. This link, which is frequently taken to be positively connected, needs to be empirically validated in order to fully comprehend its subtleties. The following questions guide our exploration:

- For each geographic unit, is there a clear correlation between GDP per capita and life
- Which geographic units boast the highest and lowest life expectancies, and how does their economic standing compare?
- What other potential issues could take place in the geographic units with a lower life
- More broadly, has life expectancy increased globally in the modern era, and does this trend correspond with global economic growth?

By addressing these issues, we may improve our understanding of the relationship between health and the economy and identify possible areas for foreign assistance and policy change. This analysis has far-reaching consequences that go well beyond the scope of a single indicator, providing an insight into the health and economic trends that influence our global community.

III. Data Preparation

1. Data overview

The dataset includes information on life expectancy, health factors, and economic data for 193 countries, collected from 2000 to 2015. Noteworthy progress in the health sector, particularly in developing nations, has been observed over the past 15 years. Initial data inspection revealed missing values, primarily for population, Hepatitis B, and GDP, with most gaps in less-known countries. Using R software and the Missmap command, missing data was addressed. Countries with persistent missing data, such as Vanuatu, Tonga, Togo, Cabo Verde, were

excluded. The final dataset comprises 22 columns, 2938 rows, and 20 predicting variables, categorized into Immunization, Mortality, Economic, and Social factors.

The attributes of the dataset are:

- Country: Country
- Year: Year
- Status: Developed or Developing status
- Life expectancy: Life Expectancy in age
- Adult Mortality: Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
- infant deaths: Number of Infant Deaths per 1000 population
- Alcohol: Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
- percentage expenditure: Expenditure on health as a percentage of Gross Domestic Product per capita (%)
- Hepatitis B: Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
- Measles: Measles - number of reported cases per 1000 population
- BMI: Average Body Mass Index of entire population
- under-five deaths: Number of under-five deaths per 1000 population
- Polio: Polio (Pol3) immunization coverage among 1-year-olds (%)
- Total expenditure: General government expenditure on health as a percentage of total government expenditure (%)
- Diphtheria: Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
- HIV/AIDS: Deaths per 1 000 live births HIV/AIDS (0-4 years)
- GDP: Gross Domestic Product per capita (in USD)
- Population: Population of the country
- Thinness 1-19 years: Prevalence of thinness among children and adolescents for Age 10 to 19 (%)
- Thinness 5-9 years: Prevalence of thinness among children for Age 5 to 9(%)
- Income composition of resources: Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
- Schooling: Number of years of Schooling (years)

2. Data preprocessing

There are numerous null values in this dataset with different approaches to handle based on the context of the data. We only choose a certain number of features to replace null values as it is easier to handle non-missing value. For 'life expectancy' attribute, we simply remove all null values from the dataset as it is crucial that this attribute is examined later in this project.

'Population' null values will not be replaced and will remain untouched to keep the authentication of the data. The remaining missing values of attribute such as GDP, BMI, mortality, alcohol consumption, HepB, polio, diphtheria, thinness, income, schooling will simply be replaced by 0.

Country	0
Year	0
Status	0
Life expectancy	10
Adult Mortality	10
infant deaths	0
Alcohol	194
percentage expenditure	0
Hepatitis B	553
Measles	0
BMI	34
under-five deaths	0
Polio	19
Total expenditure	226
Diphtheria	19
HIV/AIDS	0
GDP	448
Population	652
thinness 1-19 years	34
thinness 5-9 years	34
Income composition of resources	167
Schooling	163

Figure 1. Total number of missing values in each attribute.

The next task is to rounding numeric values in this dataset. To support the analysis task, attributes such as alcohol, percentage expenditure, income composition values will be rounded to 2 decimal places. BMI, GDP, thinness, schooling will be rounded to only 1 decimal place. The remaining numerical values will be rounded to integer. Finishing this task, our dataset is ready for analyzing

IV. Analysis and Findings

1. GDP - Life expectancy correlation

1.1 Overview of GDP per capita – life expectancy correlation:

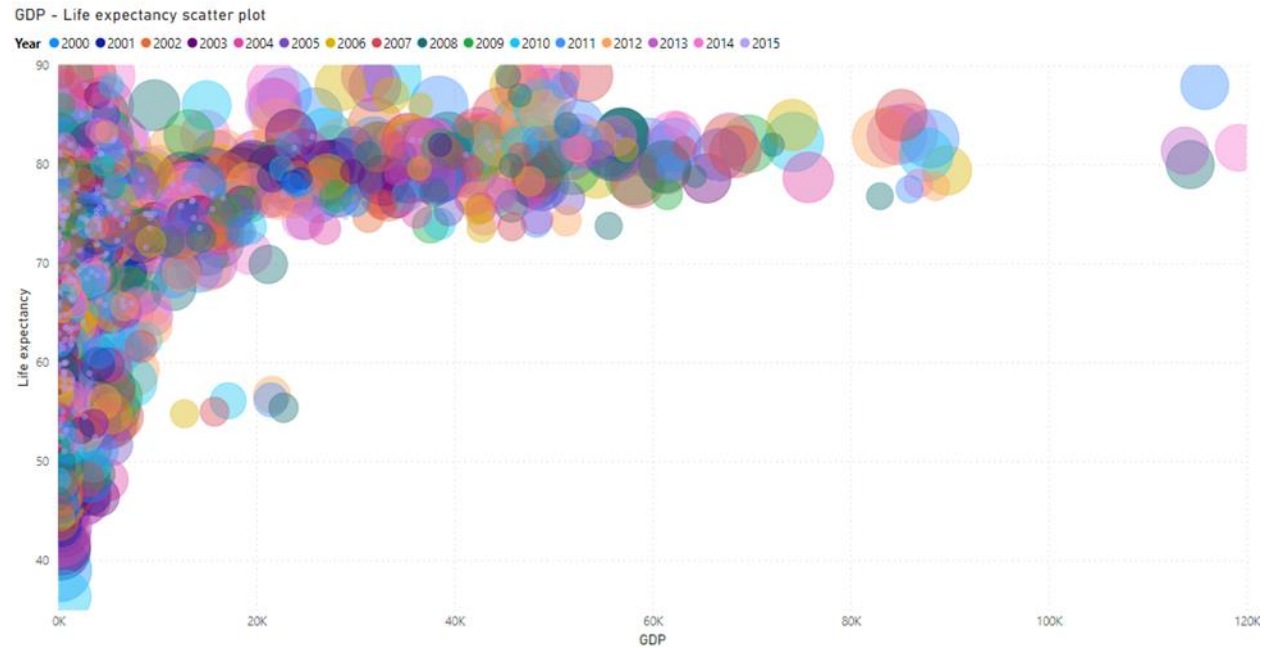


Figure 2. GDP – life expectancy correlation plot for all countries.

The above bubble plot shows the correlation of GDP per capita in the x-axis with Life expectancy in the y-axis of the country. The size of the bubble represents the country's total population and different bubble color represent the year. In general, for all countries listed in the dataset, there is a strong correlation between GDP per capita and life expectancy. Countries with high GDP tend to have high life expectancy.

1.2 GDP per capita – Life expectancy correlation for developed countries:

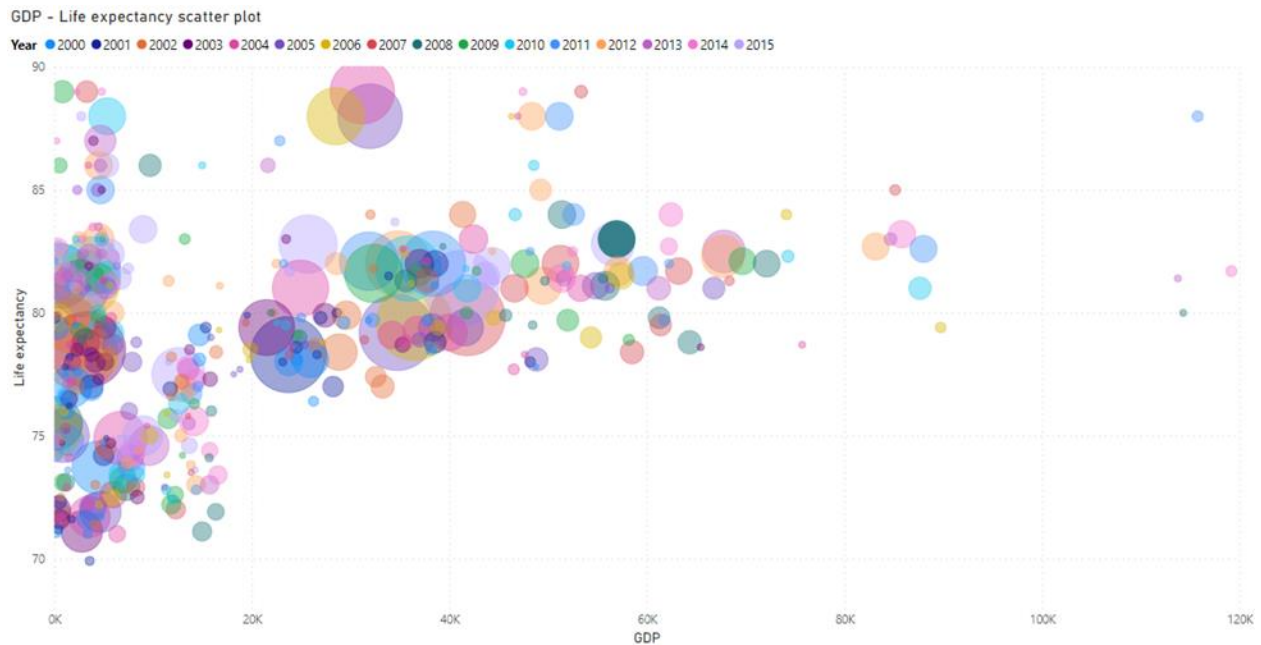


Figure 3. GDP – life expectancy correlation plot for developed countries.

For the developed countries, it can be clearly seen that there are a lot of countries with high GDP per capita, whose numbers could go up to over 100k USD. With regards to life expectancy, these countries are also among the highest at over 80 years of age. However, among the 0 to 20k GDP range there are still a lot of countries with high age expectancy due to the fact that they have high expenditure on health, a significant characteristic of developed countries.

1.3 GDP per capita – Life expectancy correlation for developing countries:

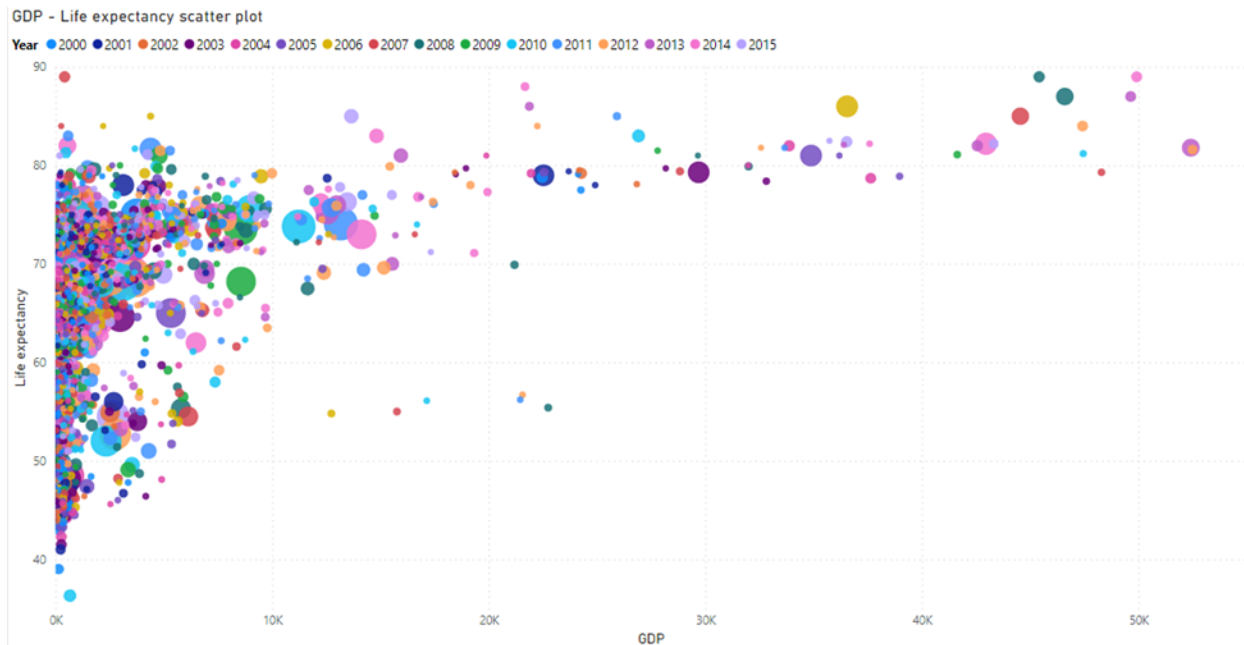
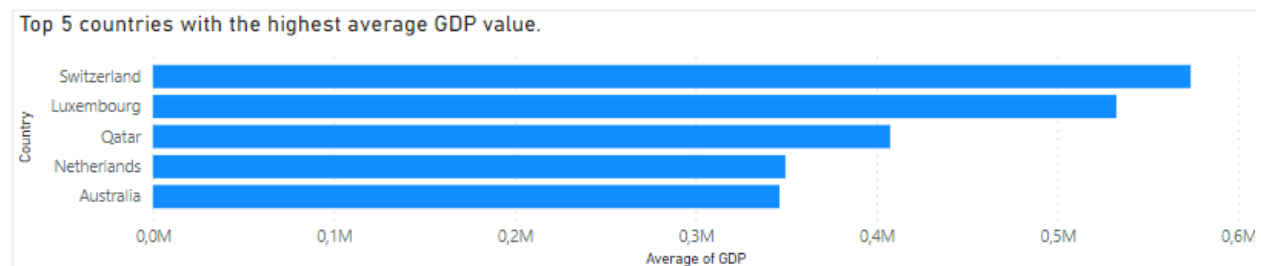


Figure 2. GDP – life expectancy correlation plot for developing countries.

For the developing countries, it can be inferred that the GDP per capita isn't as high as the number of developed countries, as the highest only reaches over 50K USD. However, they share the same characteristics as the high life expectancy strongly correlated with high GDP per capita. As expected of developing countries with low GDP per capita, the duration of a person living in these countries' life are among the lowest with some could go even below 50 years of age, something we never see in developed countries, which is alarming.

2. General overview for each GDP value and life expectancy.

2.1 Top 5 countries with the highest average GDP value:



Examining the stacked bar chart presented above, we've delved into the data representing the top 5 countries with the highest average GDP per capita on a global scale, spanning the years from 2000 to 2015. The conspicuous standout is Switzerland, positioned as the foremost country in terms of average GDP per capita, standing at an impressive 917,806 USD. This figure marks a substantial deviation from the positions held by other nations in this economic metric.

Notably, Luxembourg follows closely behind, securing the second-highest position with an average GDP per capita of 852,112 USD. The third spot is occupied by Qatar, with a noteworthy figure of 651,975 USD per capita. The ensuing countries in this distinguished top 5 list are the Netherlands and Australia, both demonstrating noteworthy figures in terms of average GDP per capita. Their values, 559,436 USD and 554,201 USD per capita respectively, indicate a relative proximity to each other in the global economic landscape.

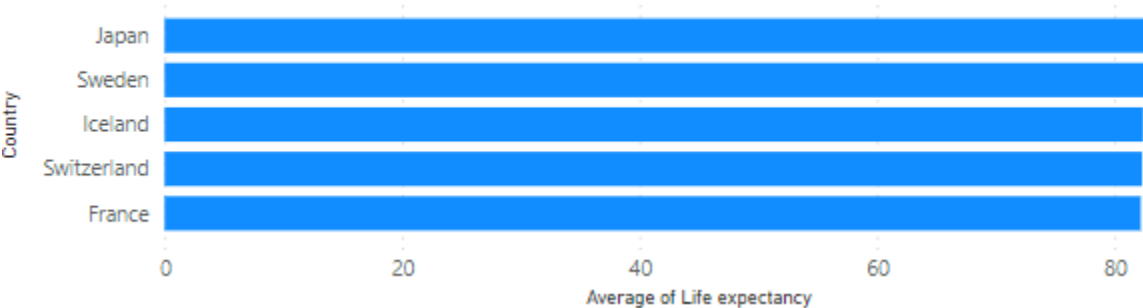
The insights garnered from this chart underscore the economic disparities and strengths among nations during the specified timeframe. Switzerland's commanding lead suggests a unique economic prowess, while the closely following countries contribute to the nuanced global economic landscape, each with its distinct position in the hierarchy of average GDP per capita. This analysis provides valuable perspectives on the economic trajectories of these nations and the intricate dynamics that shape their financial standings on the world stage.

Notably, the top 5 countries highlighted predominantly comprise developed nations, specifically Switzerland, Luxembourg, Netherlands, and Australia. This emphasizes a concentration of economic prosperity within this subset of well-established and economically advanced countries. What stands out as particularly noteworthy is the inclusion of Qatar, the lone developing country in this elite grouping, securing the third position in terms of average GDP per capita.

This juxtaposition highlights a significant economic dichotomy within the top-performing nations. While Switzerland, Luxembourg, Netherlands, and Australia showcase the economic strength and stability typically associated with developed nations, the presence of Qatar introduces a compelling dynamic. As a developing nation, Qatar's notable position suggests a unique and robust economic trajectory, challenging conventional expectations about the distribution of economic prowess among nations.

2.2 Top 5 countries with the highest average life expectancy value:

Top 5 countries with the highest average Life expectancy value



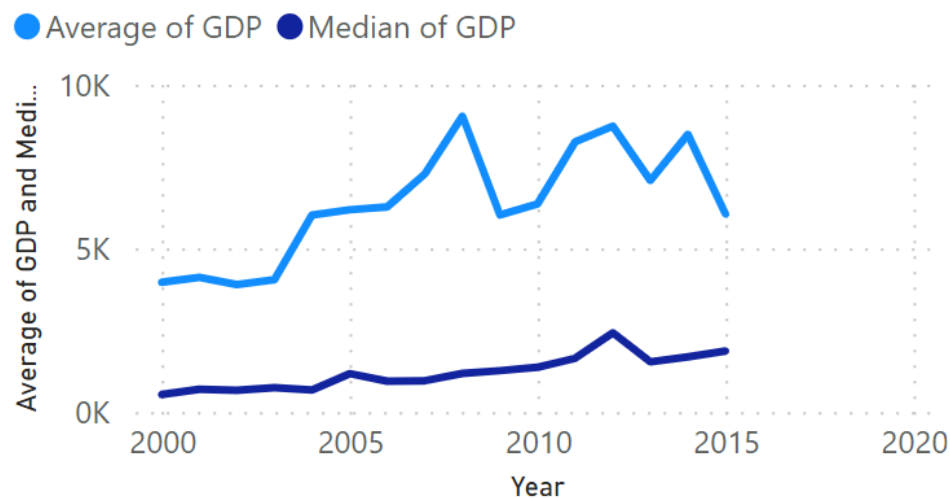
Based on the aforementioned observations, the provided stacked bar chart illustrates the top 5 countries with the highest average life expectancy globally between 2000 and 2015. Initially, it is apparent that these countries

exhibit minimal differences in their average life expectancy values. Upon closer examination, Japan stands out with the highest average life expectancy at 82.54 years, followed closely by Sweden at 82.52, Iceland at 82.44, Switzerland at 82.33, and France at 82.22 years.

Upon scrutinizing the data, a noticeable trend emerges, indicating that, in general, developed countries tend to have a higher average life expectancy compared to developing countries. However, it's noteworthy that within the top 5 countries listed above, France is categorized as a developing country based on the data spanning the years 2000 to 2015.

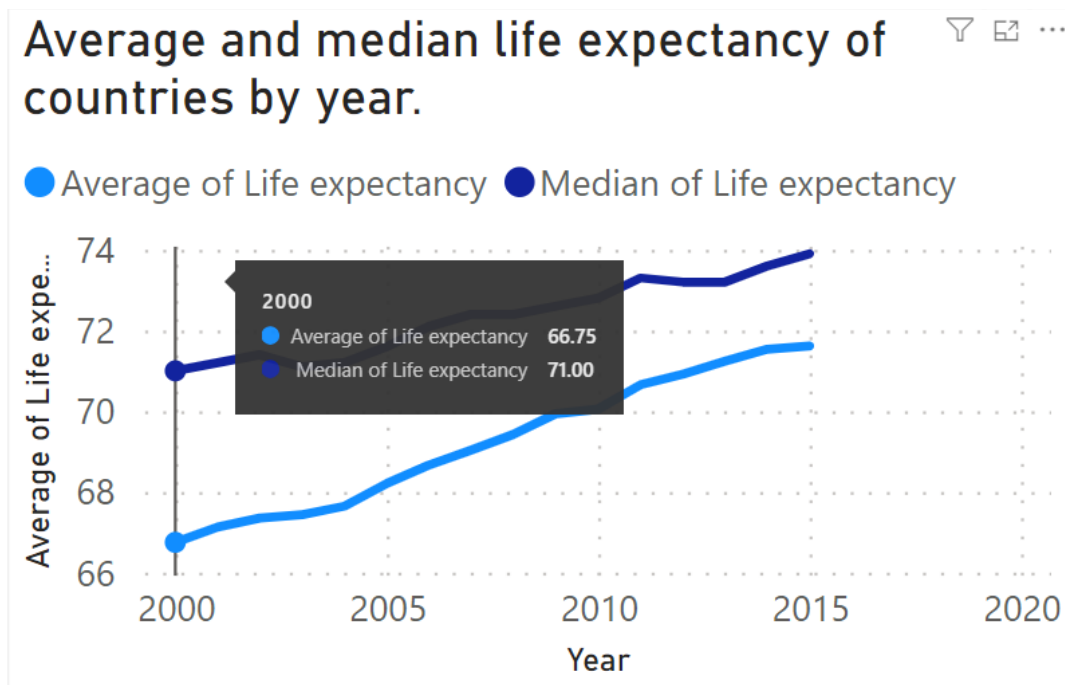
2.3 Average and median GDP and life expectancy of countries by year (2000-2015):

Average and median GDP of countries by year.



Examining the initial and final points on the line chart depicting the average GDP, we can observe a slight overall increase from 2000 to 2015. However, the interim period exhibits intricate fluctuations. The average GDP steadily rose from 2000 to 2008, followed by irregular oscillations until 2015. The pinnacle occurred in 2008, marking the highest average GDP value at approximately 9040 USD per capita. Subsequently, post-2008, there was a sharp decline to the lowest value of 6019 USD per capita within the 2008-2015 timeframe.

In contrast, the median GDP value undergoes more stable fluctuations. The nadir was observed in 2000, with a value of around 534 USD per capita, while the zenith was reached in 2012, registering 2423 USD per capita.



Contrary to the line chart of GDP, the life expectancy of countries tends to increase strongly and clearly, there does not seem to be any obvious change, the value only shows signs of a slight decrease in 2003. then continued to increase strongly. Looking right at the chart, we can see that from 2000 to 2015, the average life expectancy increased from 67 years to nearly 72 years, the median value of life expectancy increased from 71 to 74 years. This shows that development in improving life is increasingly enhanced.

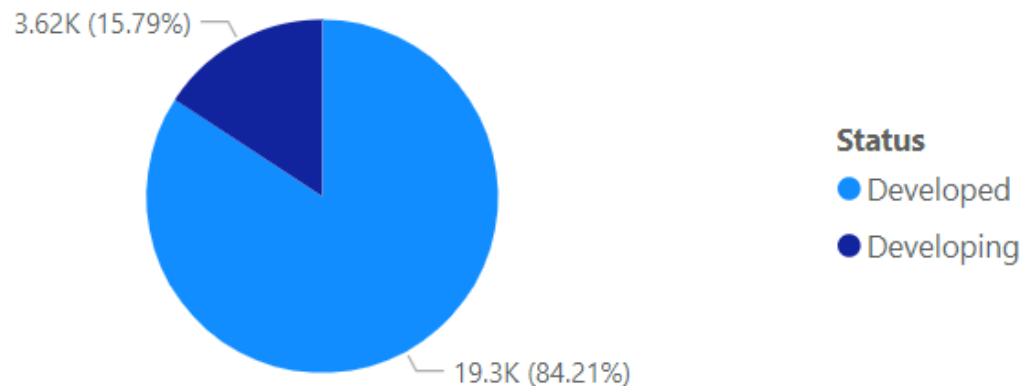
2.4 Distribution of Average GDP between developed and developing countries (2000-2015).

This dashboard displays the average Gross Domestic Product (GDP) and life expectancy for two groups: developed and developing countries.

The data is presented in two pie charts, with each chart section representing a country's contribution to the average for its respective group.

GDP Insights:

Average GDP distribution between developed and developing countries

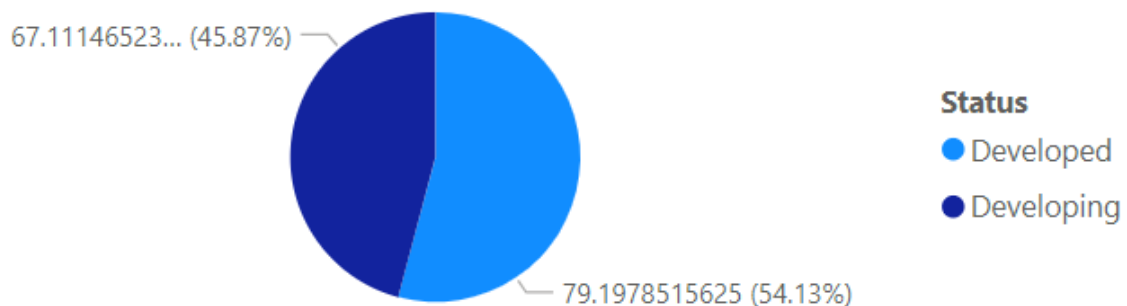


The pie chart on the left shows that developed countries have a significantly higher average GDP compared to developing countries. This is indicated by the larger size of the blue pie slice representing developed countries, compared to the dark blue slice for developing countries.

The specific average GDP values are also displayed next to each pie chart. As of the date captured in the dashboard, developed countries have an average GDP of around \$19,300, while developing countries have an average of around \$3,620. This translates to a roughly fivefold difference in average GDP between the two groups.

Life Expectancy Insights:

Average life expectancy distribution between developed and developing countries



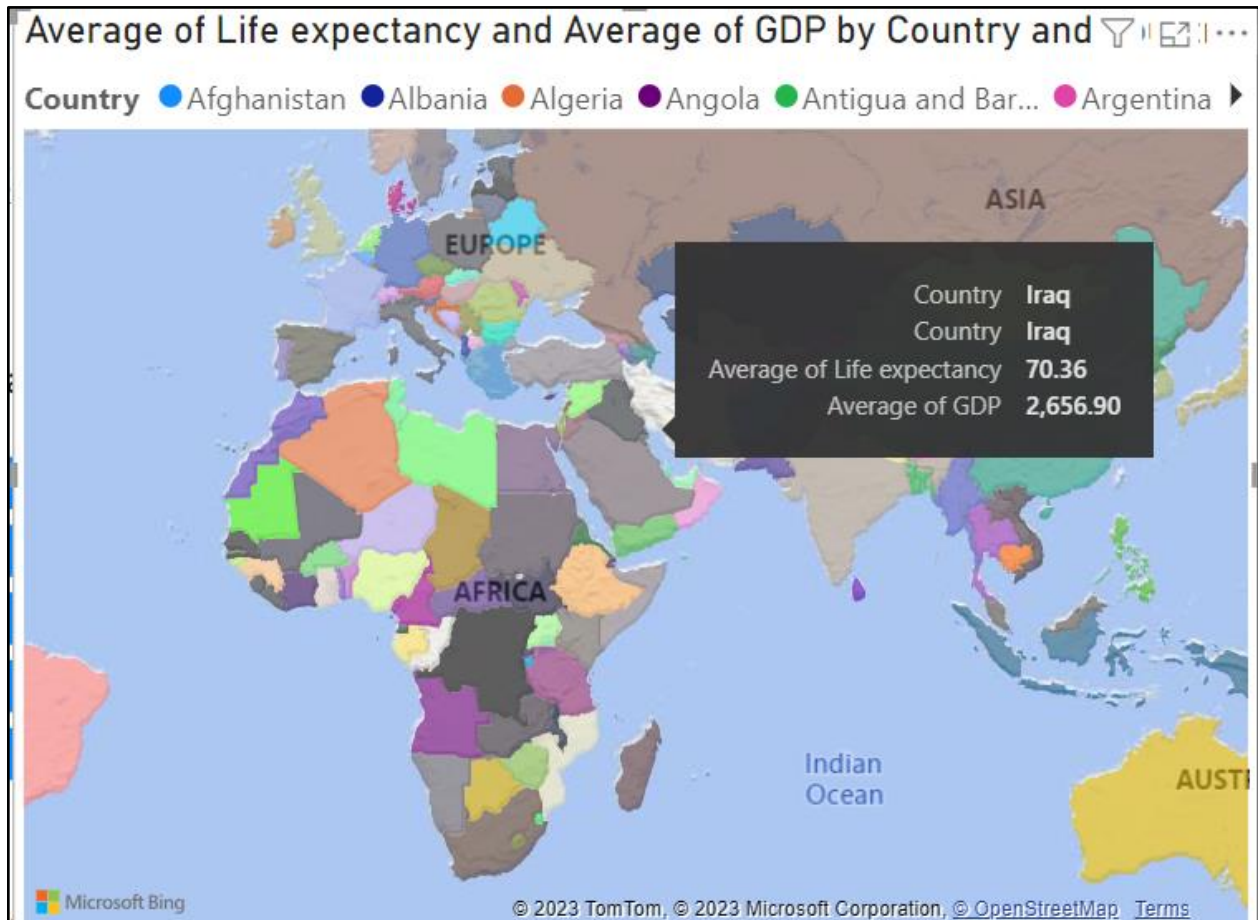
The pie chart on the right presents data on average life expectancy. Although developed countries still hold an edge, the disparity is less pronounced compared to GDP.

The blue slice representing developed countries accounts for about 45.87% of the average life expectancy, while the dark blue slice for developing countries makes up roughly 54.13%.

The specific average life expectancy values are also shown. Developed countries have an average life expectancy of around 79.2 years, while developing countries have an average of around 67.1

years. This translates to a difference of about 12 years between the two groups.

2.5 World map showing the average value of GDP and life expectancy of countries around the world (2000-2015):



The comprehensive Filled map visually encapsulates data on average life expectancy and average GDP per capita spanning the years 2000 to 2015. By utilizing this map, you gain a panoramic insight into global trends, allowing for an extensive exploration of countries across the world and the valuable information we offer. The geographical representation serves as a powerful tool for understanding the variations and nuances in life expectancy and economic prosperity on a global scale. This map is designed to provide an accessible overview, facilitating a deeper comprehension of the intricate interplay between demographics and economic factors across diverse nations during the specified timeframe.

3. Life Expectancy Influence

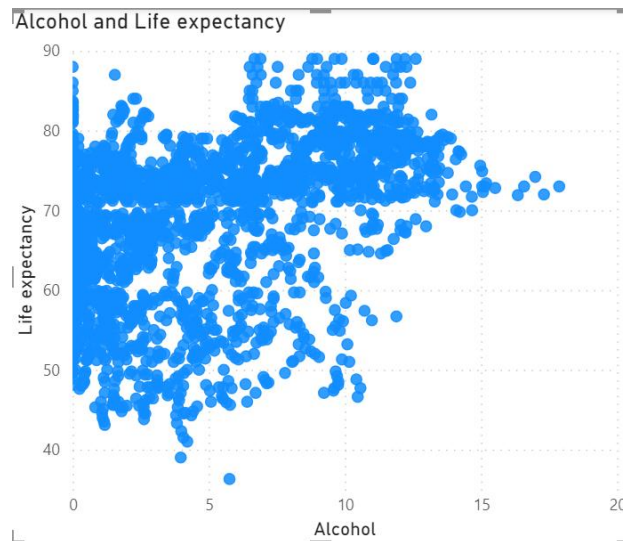
3.1. Overview

In this part, we define the 3 labels for the life expectancy which include: High life expectancy (> 70 years), Middle life expectancy (70-50 year) and low life expectancy (< 50 year). According to the database, developing countries have all kinds of labels, but developed

countries only have high and middle life expectancy.

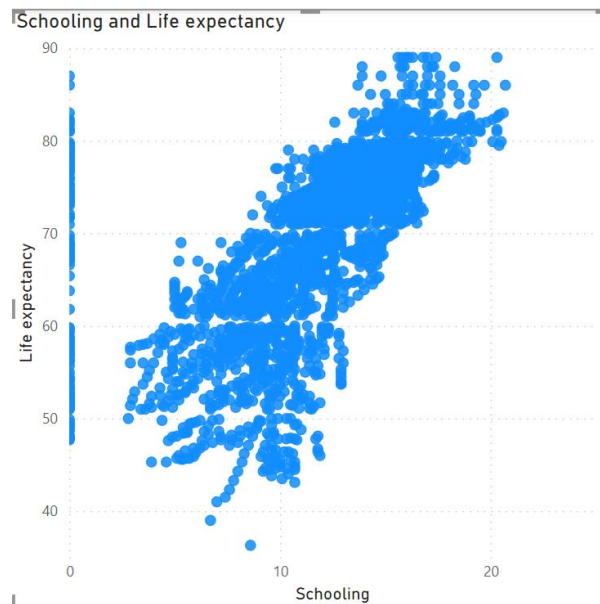
In this project, we consider 8 aspects that have much influence on life expectancy most, which include: Alcohol, Schooling, Total expenditure, HDI(Human Development Index) on income resource, BMI(Body mass Index), Children deaths, Vaccination and health expenditure percentage.

3.2. Alcohol



Alcohol, recorded per capita (15+) consumption (in liters of pure alcohol), witnesses no difference between high and low life expectancy as all the countries consume alcohol, however, the high life expectancy have a higher consumption of alcohol than the low ones.

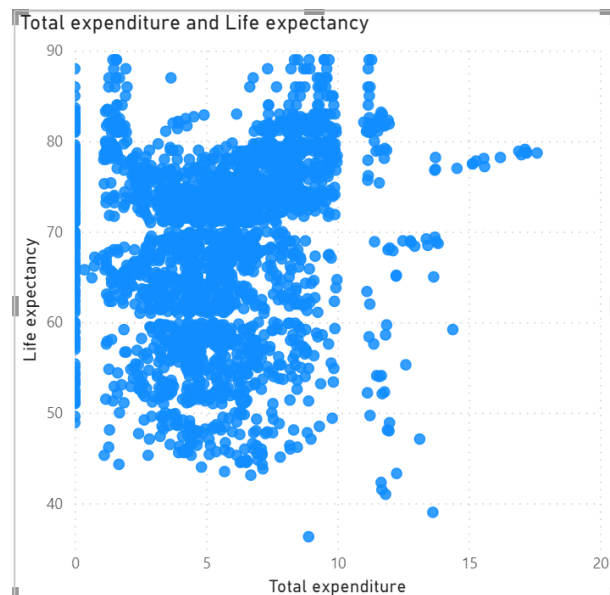
3.3. Schooling



Schooling is the average number of years that people in a country do their schooling. We can see the correlation of this feature to life expectancy. The number of schooling years is in direct proportion to life expectancy. And as a result, the low life expectancy is corresponding with the number of schooling years less than 10, which means the education system can affect life

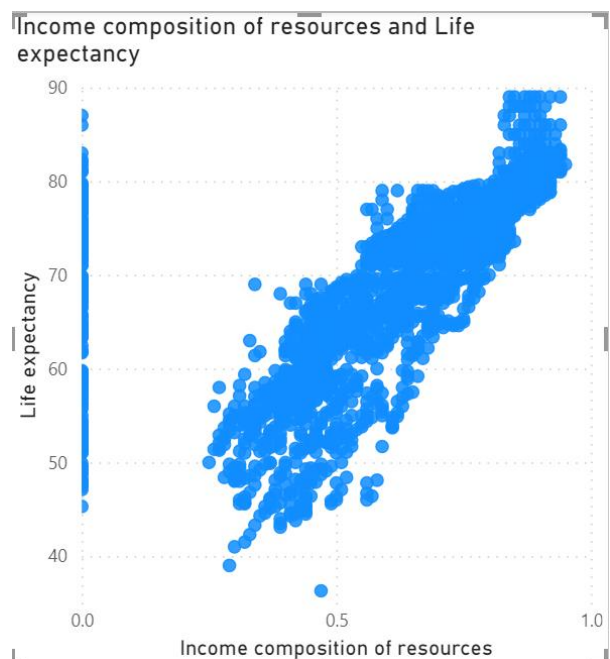
expectancy.

3.4. Total Expenditure



The total expenditure or general government expenditure on health as a percentage of total government expenditure (%). As we can see, there is no difference between high life expectancy and low one, which proves that in every country, healthcare plays an important role in government strategy. It indicates that both developed or developing countries still pay the same amount of money on health as these money from developed countries may come from the high cost for medical support and medicine while in developing countries this cost may come from the number of patients cured.

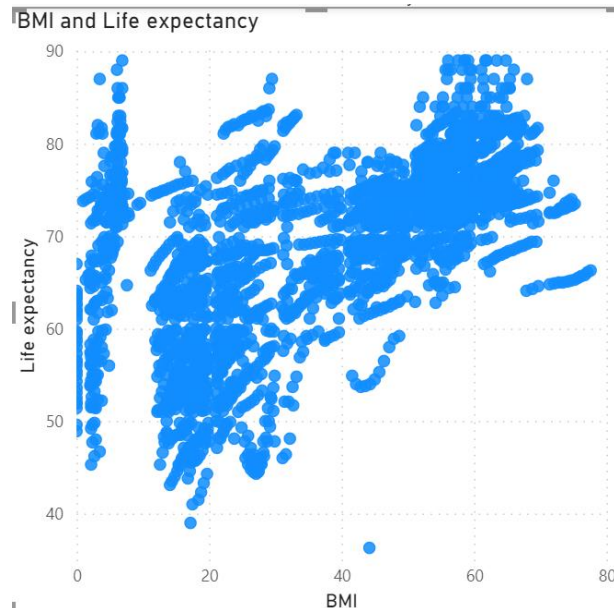
3.5. HDI on income resource



Human Development Index in terms of income composition of resources (index ranging from 0

to 1) shows a linear covariation with the life expectancy as the higher HDI, the higher expectancy. Moreover, for low life expectancy countries the HDI is only approximately around 0.5 while in the high life expectancy, the index is in the range from 0.7 to 1.0 .

3.6. BMI



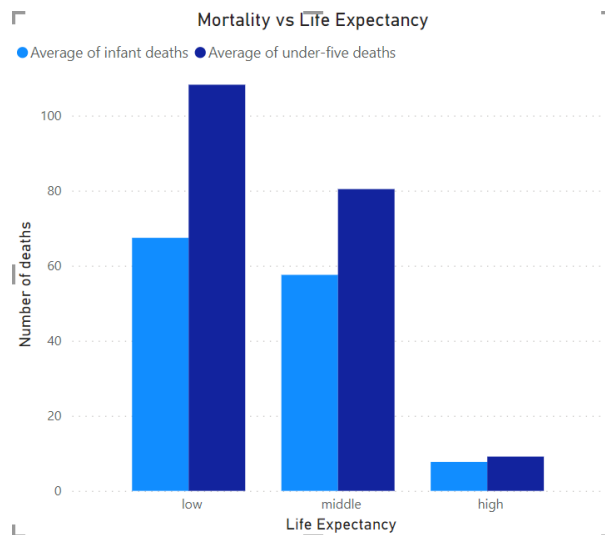
Average Body Mass Index of the entire population. This index illustrates the index of a person's weight in kilograms (or pounds) divided by the square of height in meters (or feet). It was divided into 5 interval

- + below 18.5 – you're in the underweight range
- + between 18.5 and 24.9 – you're in the healthy weight range
- + between 25 and 29.9 – you're in the overweight range
- + 30 or over – you're in the obese range

The diagram shows that, both in high life expectancy countries, the BMI filled with each group but gathered much with the BMI index > 40, which in real life means high life expectancy witness that many of them get obesity. This may be due to the advance in technology and medical systems of developed countries that make people live longer.

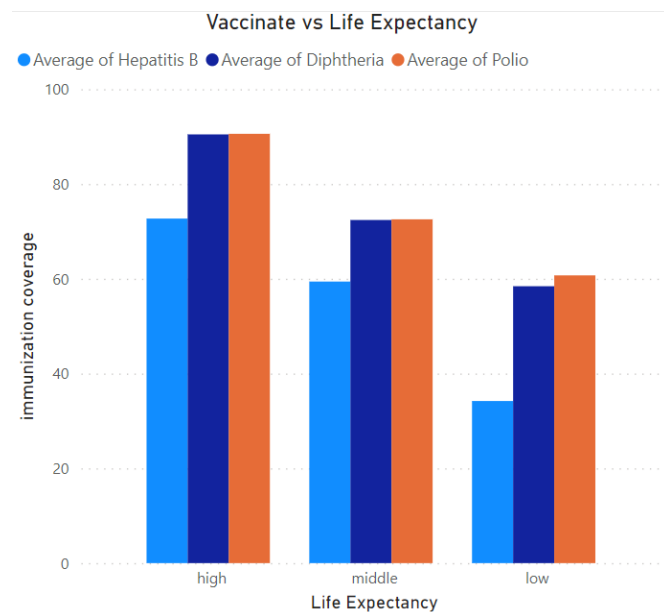
On the other hand low life expectancy countries of have a underweight body maybe because of poverty, starvation and wars.

3.7. Children deaths



This chart shows the number of children deaths per 1000 population. We easily see that, with low life expectancy the number of children deaths especially with under-five is high (up to over 100 cases average per 1000 population) while in high expectancy countries is only 15 for average.

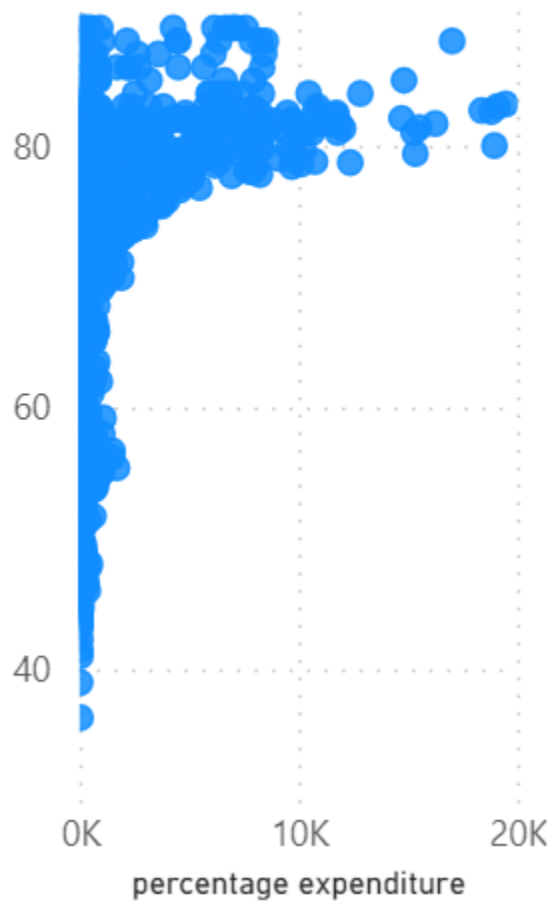
3.8. Vaccination



This graph illustrates the immunization coverage among 1-year-olds (%). The high life expectancy country get the highest immunization coverage among 1-year-olds while these in low countries is only around 60% with Polio, Diphtheria and 35% with Hepatitis B.

3.9. Health Expenditure

percentage expenditure
and Life expectancy



Expenditure on health as a percentage of Gross Domestic Product per capita(%). It clearly shows that only high life expectancy countries have pay a lot for healthcare service while in low countries, the many compared to GDP is very small (below 1000%)

4. Life Expectancy Trends Across Countries and Over Time

4.1. Overall Global Trend:

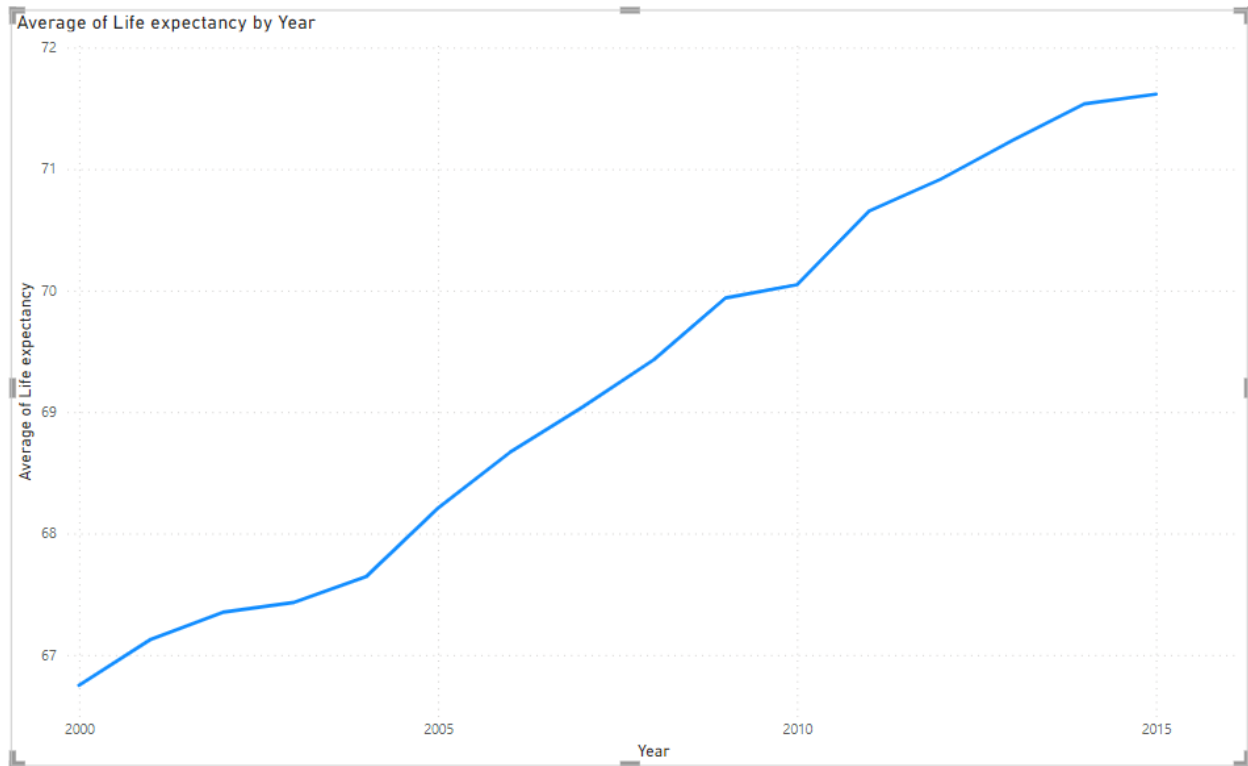


Figure 1: Global Trend of Life Expectancy from 2000 to 2015.

The enclosed line graph (Figure 1) shows the average life expectancy trend during the 15-year period. The graphic depiction makes it evident that there has been an overall rising tendency. The average life expectancy in the world in 2000 was almost 66.75 years. This is the lowest point in the time range of our dataset. A continuous rise is noted throughout the ensuing years, reaching its maximum average life expectancy of 71.62 years in 2015.

Statistical Highlights:

- Initial Point (2000): The graph indicates the starting average life expectancy at 66.75 years.
- Final Point (2015): A notable increase is seen by the end of the period, with an average life expectancy of 71.62 years.
- Increment: The increase in life expectancy does not appear to be linear, with some years showing more significant growth than others. However, the overall progression suggests improvement in global health standards, medical advancements, or other socio-economic developments contributing positively to life spans across the globe.

Through the 15-year window, the data does not show any significant declines in average life expectancy, indicating no global regressions in health outcomes during these years. The consistent upswing aligns with global efforts to improve healthcare access, disease prevention, and socio-economic conditions.

4.2. Country-Specific Trends in Life Expectancy

Overview and Identification of Trends:

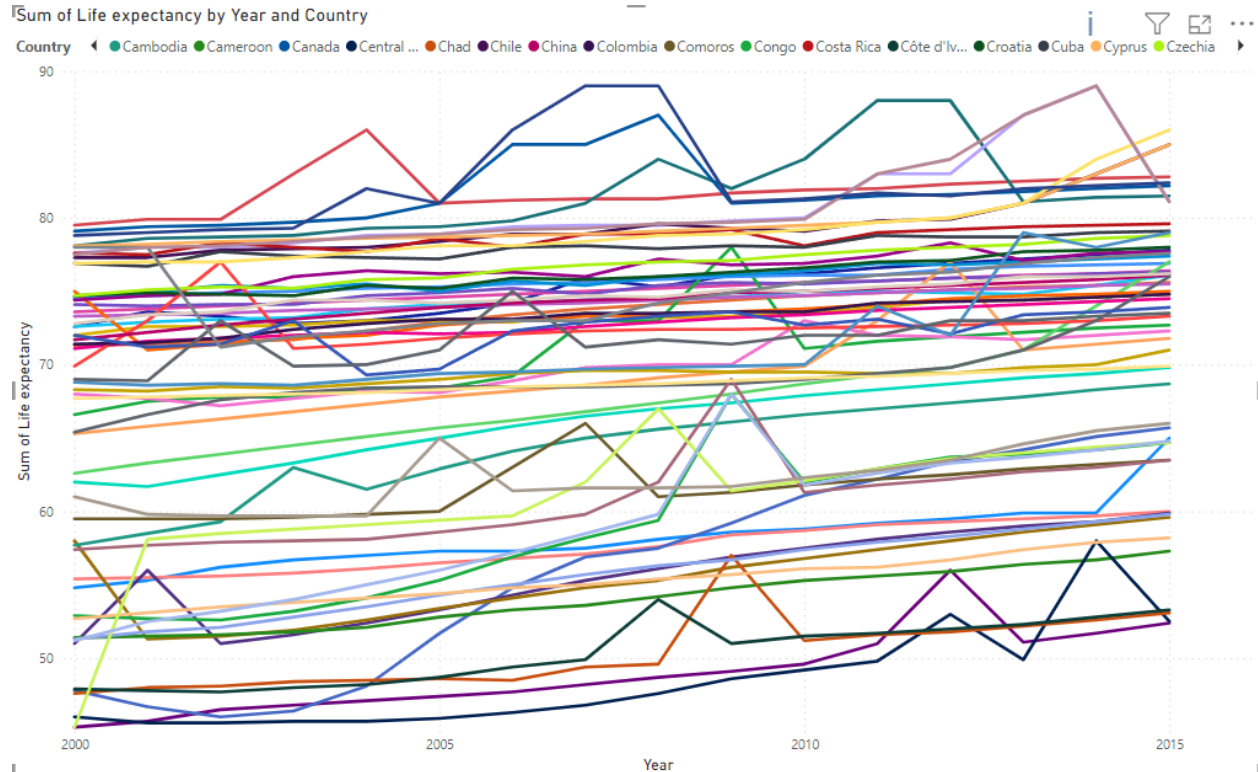


Figure 2: Overall of Life Expectancy each country from 2000 to 2015.

The life expectancy landscape from 2000 to 2015 is a mosaic of difficulties and advancements in health. Two main categories of patterns across many nations are revealed by our data analysis: stable and unstable. Little annual fluctuations in life expectancy are indicative of a stable trend, which points to a steady improvement in health outcomes. On the other hand, the erratic pattern is characterized by notable fluctuations in life expectancy, which may be a symptom of more serious health issues or notable advancements at different intervals over the duration.

Stable Life Expectancy Trends:

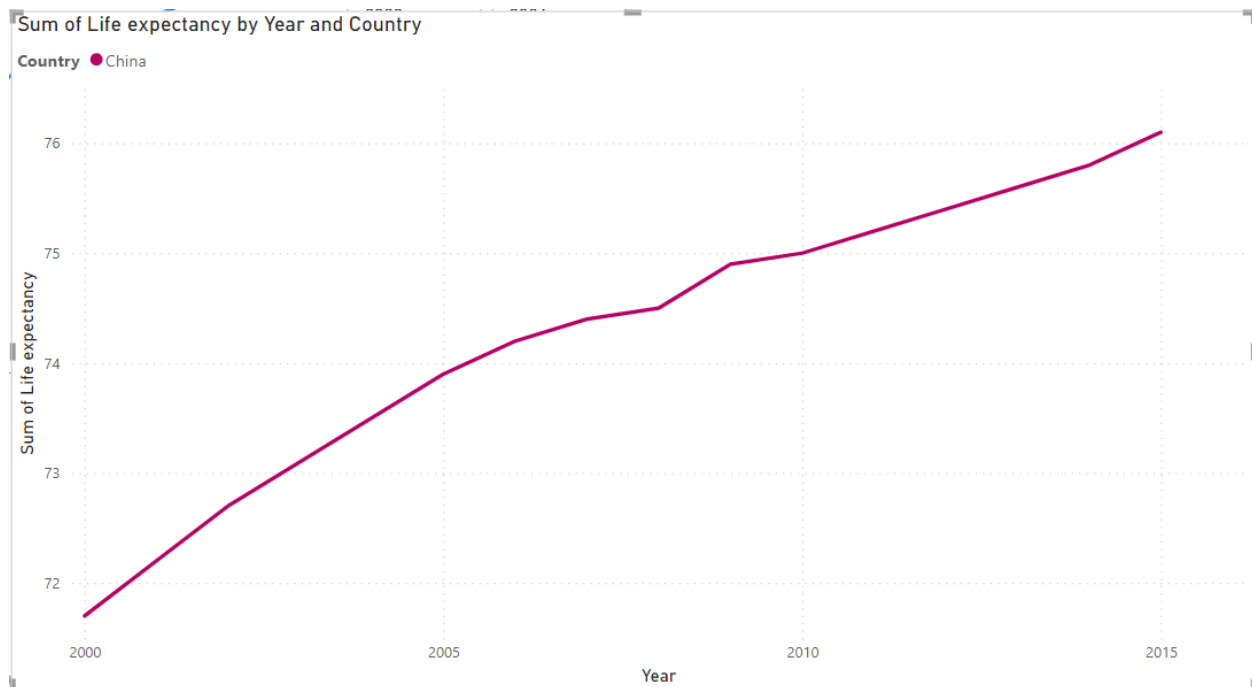


Figure 3: Life Expectancy of China compare other countries.

Over the course of the 15-year period, several nations have demonstrated a consistent trajectory for life expectancy. Some countries, like China, have had very slight variations in life expectancy, ranging from slightly higher to slightly lower compared to last year. This consistency is probably a result of ongoing improvements in healthcare accessibility, the success of public health initiatives, and general social stability.

Unstable Life Expectancy Trends:

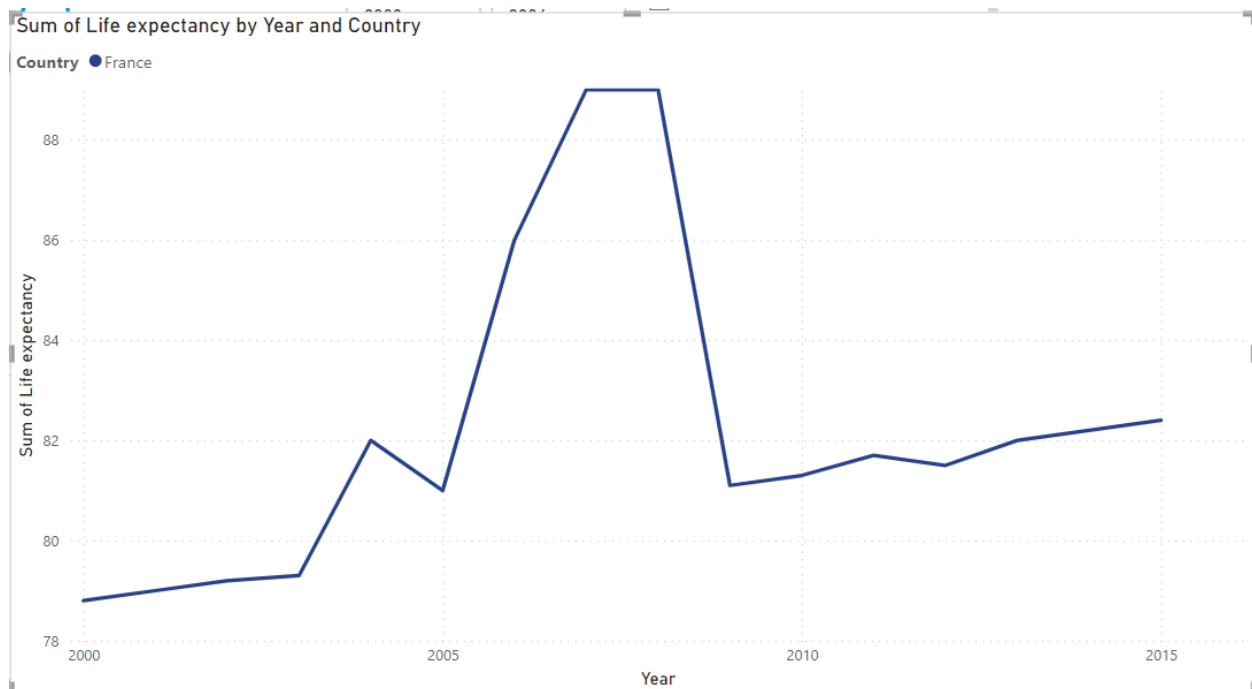


Figure 4: Life Expectancy of France compare other countries.

On the other hand, certain country have experienced significant variations in life expectancy, which are frequently associated with significant occurrences or changes. Life expectancy has fluctuated dramatically in places like France, specially from 2005 to 2009, presumably because of political unrest, major healthcare reforms, economic hardship, or epidemics. These sudden shifts demonstrate how susceptible health outcomes are to both internal and environmental factors.

In summary, data from 2000 to 2015 indicates a general rise in life expectancy worldwide, pointing to a global improvement in health outcomes. We found that a portion of the nations in this general trend had steady life expectancy rates, which is a sign of stable economic and health circumstances. On the other hand, a number of nations showed erratic patterns in life expectancy, which might were a result of shifting socioeconomic conditions and medical regulations. These results point to a complex relationship—where economic growth is a major element, but not the only one—between national wealth and health and the trends in life expectancy.

V. Conclusion

In conclusion, our comprehensive analysis has illuminated a global rise in life expectancy, indicative of progress in health outcomes. Through meticulous data scrutiny, we have been able to identify stable patterns in certain countries and volatile patterns in others, which are associated with distinct health and economic factors. These tendencies have provided a more comprehensive picture of the evolution of life expectancy and highlighted the role that the economic environment plays in influencing these patterns. This paper provides historical context for health trends and establishes a standard for future research on the relationship between global health and economic development.

