**EMMANUEL AKPE**

|  |  |
| --- | --- |
|  | LIFE EXPECTANCY DATA EXPLORATORY REPORT |

Table of Contents

[Introduction 1](#_Toc47812753)

[Problem statement 2](#_Toc47812754)

[Project Hypothesis. 2](#_Toc47812755)

[Questions 3](#_Toc47812756)

[METHODOLOGY 3](#_Toc47812757)

[Data Understanding 3](#_Toc47812758)

[Data Cleaning 5](#_Toc47812759)

[Data Exploration 6](#_Toc47812760)

[Analysis and Discussion of Findings 6](#_Toc47812761)

[Trend of Life Expectancy over the Years 6](#_Toc47812762)

[Correlation between Key Indicators 6](#_Toc47812763)

[How Immunization Factors Affect Life Expectancy. 7](#_Toc47812764)

[How Economic Factors Affect Life Expectancy. 9](#_Toc47812765)

[Life Expectancy & Mortality Factors. 9](#_Toc47812766)

[Life Expectancy & Social Factors 10](#_Toc47812767)

[Life Expectancy and Health Factors 10](#_Toc47812768)

[Conclusion. 11](#_Toc47812769)

[Recommendations. 12](#_Toc47812770)

[Reference 13](#_Toc47812771)

# Introduction

Life expectancy is a statistical measure of the average time an individual is expected to live, based on the year of its birth, its current age and other demographic factors.

Life expectancy at birth reflects the overall mortality level of a population. That is, Life Expectancy summarizes the mortality pattern that prevails across all age groups each year (children and adolescents, adults and the elderly). The global average life expectancy changes (increase or decrease) each year and this could be attributed to a lot of factors. This Work seeks to analyze the various factors that affects life Expectancy and hence recommend ways by which a country can increase their respective Life Expectancy.

## Problem statement

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that the effect of immunization and human development index was not considered in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization factors like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations in this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting to a country which area should be given importance in order to efficiently improve the life expectancy of its population.

## Project Hypothesis.

This project focuses on how immunization factors, mortality factors, economic factors, social factors and other health related factors impact life expectancy. In view of this objective, the project seeks to evaluate the following hypothesis.

1. : there is a general downtrend in life expectancy over the years

: there is a general uptrend of life expectancy over the years

1. : Immunization against diseases such as Polio, Measles, Diphtheria and Hepatitis B leads to increase in Life Expectancy

: Immunization against diseases such as Polio, Measles, Diphtheria and Hepatitis B leads to decrease in Life Expectancy.

1. : Economic factors such as GDP of a country, Income and Total expenditure of citizens contributes to higher Life Expectancies.

: Economic factors such as GDP of a country, Income and Total expenditure of citizens contributes to lower Life Expectancies.

1. : Adult mortality, Infant mortality and Under-five mortality are some types of mortality factors that result in lower Life Expectancies among countries.

: Adult mortality, Infant mortality and Under-five mortality are some types of mortality that results in higher Life Expectancies amount countries.

1. : Health issues and conditions such as HIV/AIDS and BMI leads to lower levels of Life Expectancies among countries.

: Health issues and conditions such as HIV/AIDS and BMI leads to higher levels of Life Expectancies among countries.

1. : Alcohol consumption and schooling are some social factors that affects higher Life Expectancies among both Developed and Developing countries.

: Alcohol consumption and schooling are some social factors that affects higher Life Expectancies among both Developed and Developing countries.

## 

## **Questions**

1. What is the general trend of life expectancy over the years?
2. Does immunization against diseases such as polio, diphtheria, measles and hepatitis B increase life expectancy?
3. Does economic factors such as GDP, income and total expenditure of citizens impact life expectancy?
4. What is the relationship between mortality factors (such as adult mortality, infant mortality, and under five mortality) and life expectancy?
5. Does health conditions such as HIV/AIDS and BMI lead to lower life expectancy among countries?
6. How does social factors such as alcohol consumption and schooling affect life expectancy?

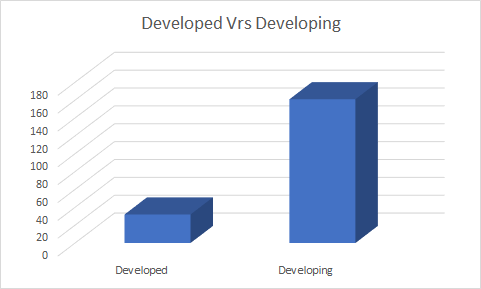
# METHODOLOGY

To fully understand how immunization factors, mortality factors, economic factors, social factors and other health related factors impacts life expectancy, several steps were taken to ensure that an accurate and efficient relationship is developed from the given dataset. The steps taken have been subdivided into three parts namely, Data understanding, Data cleaning and finally Data Exploration.

The two main tools used for this data exploratory was Python 3.8.6 and Microsoft Excel

## Data Understanding

The Life Expectancy dataset is made of data from 193 countries around the world. The data is an aggregate of many indicators for a country in a particular year. That is, the data has multiple indicators in a time series separated by country. The countries have been categorized into developed and developing based on the economic factors and standard of living in the respective countries. It was realized from the dataset that we have 161 countries from developing countries and 32 countries from developed countries.



The dataset was collected for a period of 15 years across these 193 countries, that is from the year 2000 to the year 2015. The dataset has a lot of key indicators which makes up the column of the dataset. These columns are listed and explained briefly below.

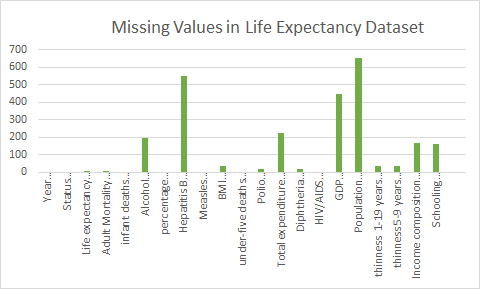
* Country (Nominal): Indicates the country in which the indicators are from (i.e. Afghanistan, United States of America etc.)
* Year (Ordinal) - Indicates the year from which the dataset was collected. It ranges from 2000 to 2015.
* Status (Nominal) - This explains whether a country is classified as 'Developing' or 'Developed' by standards of WHO.
* Life expectancy (Ratio) - Shows the life expectancy of people living in a country in a respective year.
* Adult mortality (Ratio) - the adult mortality rate per 1000 population (i.e. number of people dying between 15 and 60 years per 1000 population); if the rate is 263 then that means 263 people will die out of 1000 between the ages of 15 and 60; another way to think of this is that the chance an individual will die between 15 and 60 is 26.3%
* Infant deaths (Ratio) - number of infant deaths per 1000 population.
* Alcohol (Ratio) - The alcohol consumption rate measured as liters of pure alcohol consumption per capita in a country.
* Percentage expenditure (Ratio) - expenditure on health as a percentage of Gross Domestic Product (GDP)
* Hepatitis (Ratio) - number of 1-year old with Hepatitis B immunization over all 1-year old in population
* Measles (Ratio) - number of reported Measles cases per 1000 population
* BMI (Interval/Ordinal) - average Body Mass Index (BMI) of a country's total population
* under-five deaths (Ratio) - number of people under the age of five deaths per 1000 population
* polio (Ratio) - number of 1-year old with Polio immunization over the number of all 1-year old in population
* Total expenditure (Ratio) - government expenditure on health as a percentage of total government expenditure
* Diphtheria (Ratio) - Diphtheria tetanus toxoid and pertussis (DTP3) immunization rate of 1-year old
* HIV/AIDS (Ratio) - deaths per 1000 live births caused by HIV/AIDS for people under 5; number of people under 5 who die due to HIV/AIDS per 1000 births
* GDP (Ratio) - Gross Domestic Product per capita
* Population (Ratio) - population of a country
* thinness\_1-19\_years (Ratio) - rate of thinness among people aged *10-19* (Note: variable should be renamed to *thinness\_10-19\_years* to more accurately represent the variable)
* Thinness\_5-9\_years (Ratio) - rate of thinness among people aged 5-9
* Schooling (Ratio) - average number of years of schooling of a population

We realized that all the dataset is numeric except for Country Name and Status (Developed or Developing).

Now that we fully understand our dataset and what each variable(indicator) means, lets dive into cleaning of our dataset.

## Data Cleaning

Data cleaning is the process of detecting and correcting corrupt or inaccurate records from a dataset. We cleaned our dataset to ensure that our dataset conforms to standards in order to facilitate the smooth running of our analyses and to obtain accurate relationships between variables(indicators). To achieved this, we first checked for the number of missing values in our dataset. The graph below shows the number of missing values in the dataset.



Clearly, there exists several missing values in the dataset and they are mostly situated in indicators like Population and Hepatitis B, and this could be due to several factors. Since we will need perform basic statistics like finding the mean, median, variance with columns involving missing values, we must find a way of dealing with these missing values in order to facilitate the smooth analyses on the variables. To achieve this, we came up with different ideas like deleting the rows or columns involving these missing values, populating the missing values with the mean of each column or interpolate those columns or rows involving the missing values. We decided to populate the missing fields with the mean of their respective columns. After doing that, we realized that there are no more missing fields in our dataset and hence we can proceed with our appropriate analyses.

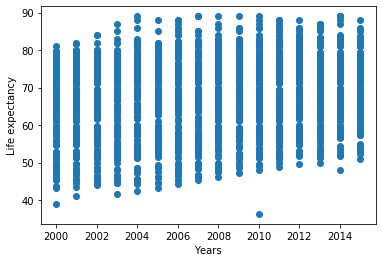
## **Data Exploration**

To fully understand the relationship between Life Expectancy and the various indicators, we used visual exploration to achieve this goal. This is performed based on how various key indicators affect Life Expectancy in a country. These various indicators have been grouped into Immunization factors, Health factors, Mortality factors, Economic factors and social factors. To begin with the analyses, we first look at the general trend of the Life expectancy across the world in progressive years.

# Analysis and Discussion of Findings

### Trend of Life Expectancy over the Years

The graph below shows the general trend of Life Expectancy in all the countries present in the dataset from 2000 to 2015.



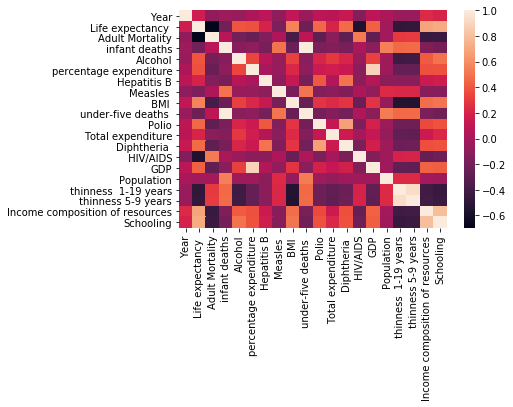
It can be observed from the graph that, life expectancy has been increasing steadily from the year 2000 to the year 2004. it started fluctuating (Increase or decrease) from 2004 to 2014.

Why is this relationship so?

To answer this question, we need to analyse the various key indicators with Life expectancy over time. We started by looking at the general correlation between the various key indicators. We achieved this by plotting the correlation matrix.

### Correlation between Key Indicators

The diagram below shows the correlation between the various key indicators within the dataset. The legend tells that the warmer colours show higher and positive correlation whiles the colder colours show low or negative correlation.

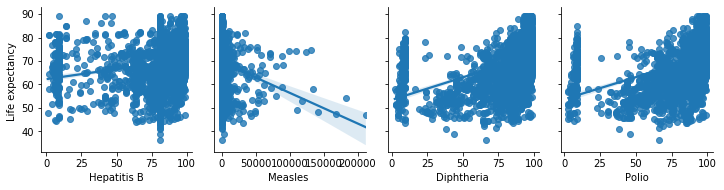


It can therefore be observed from the diagram above that there exists very high correlation between Life expectancy and Schooling, Life expectancy and Total expenditure, Life Expectancy and GDP, Life expectancy and percentage expenditure just to mention a few. That is as these indicators increases in value, it will result in a corresponding increase in the Life expectancy and whiles they decrease in value, they will yield a resulting decrease in Life Expectancy.

Also, there exists very low correlation between Life Expectancy and Thinness (thinness 1-19, thinness 5-9), Life Expectancy and HIV/AIDS, Life expectancy and Adult mortality and infant deaths. That is, whiles these indicators increase in value, it will yield a corresponding decrease in the Life Expectancy and whiles they decrease in value, it will result in a corresponding increase in Life Expectancy.

### How Immunization Factors Affect Life Expectancy.

Immunization is defined by the WHO as the process whereby a person is made immune or resistant to an infectious disease, typically by the administration of a vaccine. That is, Vaccines are used to boost an individual’s immune system in order to prevent serious life-threatening diseases. Examples includes diphtheria, polio, Hepatitis B and the like. We will look at the relationship between Life expectancy and the immunization factors like diphtheria, Polio, Measles, Diphtheria and Hepatitis B. The graph below shows the relationship between Life expectancy and the other factors of Immunization.



#### Life Expectancy & Measles

The measles column in the dataset shows the number of children who were infected with Measles in the respective countries.

It is observed that Life Expectancy and Measles are negatively related. This means that, whiles the rate of Measles increases, the Life Expectancy falls accordingly. In the same way, whiles the rate of Measles decreases, the Life Expectancy will increase accordingly, all other things been constant. A possible cause of this negative relationship is that, measles as one of the six killer disease contributes to early deaths, as result, immunizing against this disease means that all other things being equal, people who are immunized can live for long and this results in higher levels of life expectancies.

#### Life Expectancy &Hepatitis B

The Hepatitis B column shows the number of 1-year old with Hepatitis B immunization over all 1-year old in a population. It is observed from the graph above that there exists a mild positive relationship between Hepatitis B (Immunization) and Life Expectancy. That is, the more people are immunized against Hepatitis B, the higher the life expectancy. In a similar way, the less people who are immunized against the Hepatitis B the lesser the life expectancy.

#### Life Expectancy & Diphtheria

The Diphtheria column shows the Diphtheria tetanus toxoid and pertussis (DTP3) immunization rate of 1-year old in various countries. It is observed from the graphs above that there is a positive relationship between Life Expectancy and Diphtheria (Immunization). That is when more people are immunized against Diphtheria, it will yield a corresponding increase in the Life Expectancy rate and vice versa.

#### Life Expectancy and Polio

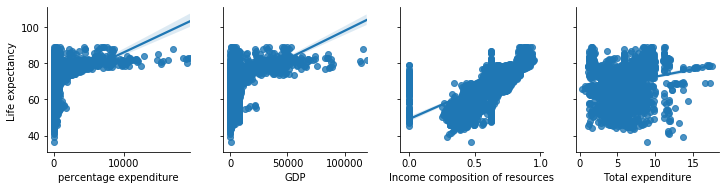
The Polio column shows the number of 1-year old with Polio immunization over the number of all 1-year old in population(country). It is observed from the graph above that there exists positive relationship between Polio(immunization) and the Life Expectancy. That is, when more people are immunized against polio, Life Expectancy will increase and vice versa.

It can therefore be concluded that, Immunization is positively related to Life Expectancy. This means that when more people are immunized against diseases like Polio, Measles Diphtheria and Hepatitis B, the Life Expectancy level will increase, and the Life Expectancy will fall if less people are not immunized against such diseases. This is because immunization protects the immune system against these diseases which are results in early mortality and contributes to reductions in life expectancies.

### How Economic Factors Affect Life Expectancy.

Economic factors refer to the factors that tend to influence the economic growth and sustainability. Examples include GDP of a country, Income and Total expenditure of citizens just to mention a few. We will look at the relationship between Life expectancy and Economic factors like total expenditure, GDP, Income composition of resources. The graphs below shows the relationship between Life Expectancy and the Economic factors.

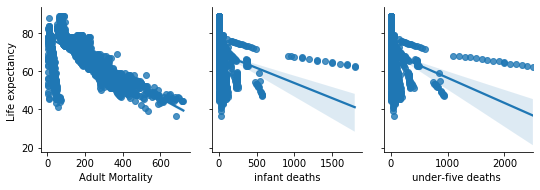
Economic factors such as GDP of a country, Income and Total expenditure of citizens (on health and health care) contributes to higher Life Expectancies.



It is observed that Life expectancy is related positively to the Economic factors like percentage expenditure, GDP, Income composition of resources and the Total expenditure. This means that, whiles the values of the economic factors increase, the level of Life Expectancy will also increase and vice versa. This is mostly because, an increase in the economic factors is likely to increase the well-being of the general citizens which in turn could lead to a corresponding increase in the Life expectancy of such country. For instance, an increase in expenditure on health and health care means that the country invests more in health care and this is likely to improve the health of its citizens and subsequently resulting in increase in life expectancy. Increase in the GDP of an economy translates into increased per capita income, all other factors being held constant, a higher per capita income means that citizens can now spend more on goods and services including health and health care, other things being equal. This is likely to increase life expectancies in such economies. With health being a normal good in the human capital expenditure index, an increase in income means individuals will demand more of healthcare and this positively translates into higher levels of life expectancies. There for increase in income composition of resources leads to increase in life expectancy.

### Life Expectancy & Mortality Factors.

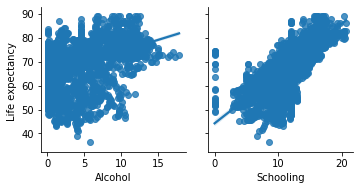
Mortality is a measure of the number of deaths within a country over a period. Examples include under-five deaths, adult mortality and infant deaths. The relationship between Life expectancy and Mortality factors are analyzed below.



It is observed that Life expectancy is related negatively to the Mortality Factors like under-five deaths, adult mortality and infant deaths. This means that, whiles the values of the Mortality factors increase, the level of Life Expectancy will decrease and vice versa. As a country records more deaths (adult mortality, infant deaths and under-five deaths), especially, infant deaths and under-five deaths, the average number of years that a person is expected to die decreases, and this translates into less life expectancies among countries

### Life Expectancy & Social Factors

Social factors refer to the conditions that influence our health and well-being like schooling, relationships, alcohol consumption just to mention a few. The relationship between Social factors like schooling and alcohol and Life Expectancy is explored in the graph below.

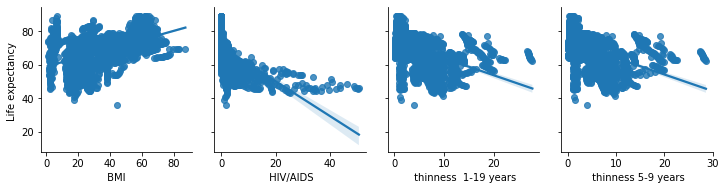


It is observed from the graph above that the social factors like schooling and alcohol consumption has a strong positive relationship with Life expectancy. This implies that when there is an increase in the social factors like schooling, it will lead to a greater increase in the level of Life expectancy. A country with an educated population is more likely to consume healthy foods which would improve their health and it is assumed that the population with higher percentage being educated are more likely to earn higher income. Increase in income means higher demand for health and health care since health is normal good, all other things being equal. The effect is higher life expectancy.

### Life Expectancy and Health Factors

The Health factors explains the general health state of an individual in a country at a period.

Examples include HIV/AIDS status, the BMI levels just to mention a few. The graph below examines the relationship between Life Expectancy and Health factors.



#### BMI & Life Expectancy

BMI stands for body Mass index and it indicates whether a person’s weight is normal, underweight or overweight. It can be observed from the graph that the BMI is positively related with Life Expectancy. This means that when the number of people with normal BMI increase, Life expectancy will also increase. However, if the number of people with normal BMI falls, Life expectancy will fall.

#### Life Expectancy and other Health factors

It can also be observed that there exists strong negative relationship between the other Health factors like HIV/AIDS, thinness in people, and Life Expectancy.

This means that when there is an increase in the Health factors like HIV/AIDS, thinness in people, it will lead to a corresponding fall in Life Expectancy and vice versa. This is because with HIV/AIDS for instance, a country with a population with higher number of HIV/AIDS patients are more likely to have its population experiencing high levels of mortality at early stages which results in lower life expectancy.

Overall, we can conclude that when an individual life a healthy life, it will increase his or her Life expectancy.

# Conclusion.

The following conclusion can be drawn from the various analysis above.

1. Life expectancy among countries assumed an upward trend over the years from 2000 to 2015. This means that Life expectancy has been increasing over the years.
2. Immunization against diseases such as such as Polio, Measles, Diphtheria and Hepatitis B leads to increase in Life Expectancy. When the population of a country is immunized, the rate of mortality from these diseases is reduced which results in higher Life expectancy.
3. Economic factors such as GDP of a country, percentage of expenditure on health, income from resources and total expenditure individually accounts for higher life expectancy.
4. Our findings revealed that mortalities such as adult mortality, infant mortality and under-five mortality negatively impacts Life expectancy. This means that when there is an increase in mortalities in a country, it will lead to a proportionate fall in the Life expectancy of such country and vice versa.
5. In addition, we explored the relationship between life expectancy and some social factors such as schooling. We found that these social factors are positively related to Life expectancy. That is, the Life expectancy of a country will increase if its citizenry attains higher levels of education and vice versa.
6. Finally, the analysis reveals that life expectancy is negatively impacted by other health issues such as HIV/AIDS and thinness while BMI directly affects Life expectancy. This means that, when the health standards are compromised in a country, it is likely that its Life expectancy will fall proportionately.

# Recommendations.

Based on our findings, we recommend that the following.

1. Governments of various countries should ensure that their populations are immunized against diseases such as Polio, Measles, Diphtheria and Hepatitis B. The immunizations must be carried out by the appropriate bodies and agencies in individual countries in accordance with WHO recommendations.
2. Governments should implement sound economic policies that leads to reduced interest rates (to boost investments), reduced inflation among others that are necessary to enhance economic growth. Economic growth leads to increased income which is necessary for ensuring and maintain higher Life expectancies.
3. To improve Life expectancy, schooling (education) especially basic education should be made accessible to all. Higher education should be affordable and easily attainable by the average class in the society.
4. Countries should strengthen their health care systems to provide improved health care to its citizens. This is to ensure that there is a reduction in the different types of mortalities. There should be increased maternal health education to help reduce infant and under-five mortalities.
5. Governments should invest in quality health systems to help manage other health related issues and conditions in the country. In addition, there should be enough education to sensitive citizens on preventive measures against diseases as such as HIV/AIDS.

# Reference

<https://www.kaggle.com/ecedolen/detailed-regression-project-on-life-expectancy>

<https://ourworldindata.org/lifeexpectancy#:~:text=The%20term%20%E2%80%9Clife%20exectancy%E2%80%9D%20refers,will%20be%20when%20they%20die.>

<https://en.wikipedia.org/wiki/Life_expectancy>

<https://www.britannica.com/science/mortality-demography>

<https://www.who.int/gho/mortality_burden_disease/life_tables/situation_trends_text/en/>

<https://www.who.int/data/gho/publications>

<https://www.worldometers.info/population/>

<https://data.worldbank.org/indicator/SP.DYN.LE00.IN>

<https://www.disabledworld.com/fitness/longevity/#:~:text=Significant%20factors%20in%20life%20expectancy,factors%2C%20genetics%20and%20lifestyle%20choices.>

<https://www.fingent.com/blog/a-simple-guide-on-understanding-exploratory-data-analysis/>