EDA:

**Goal:**

The aim of the project is to understand how life expectancy is dependent upon different factors like economical, immunization rate, diseases, lifestyle and health care expenditure in a country. The data is collected from 2000 to 2015 for 193 countries.

We analysed the four major segments which affect Life expectancy using the following factors:

**Healthcare & Immunization**- expenditure on healthcare as a percent of GDP and polio immunization

**Education**- schooling years

**Economic factors**- Developed or Developing countries

**Geographic location**- Continents the country belongs to

**Data Description**:

The dataset is available at <https://data.worldbank.org/indicator/SP.DYN.LE00.IN> . The data-set related to life expectancy and health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from the United Nation website. In total, there are 8 columns and 3000 rows in the final merged file. An initial look at the data revealed some missing values. Since the data sets come from WHO, we found no obvious errors. Based on the results, most of the missing data were related to population, Hepatitis B, and GDP. The variable descriptions:

* **Life Expectancy:** Measured in Age (Years)
* **Country**
* **Year**: Ranging from 2000-2014
* **Status**: Developed/Developing
* **Polio**: Percent of 1 year-old population vaccinated Polio
* **Diphtheria**: Percent of 1 year-old population vaccinated against Diphtheria
* **Total** **Expenditure**: General government expenditure on health as a percentage of total government expenditure (%)
* **Schooling**: Number of years of Schooling(years)

Exploratory Data Analysis:

Over the span of 15 years, the general trend of life expectancy is in the increasing direction. As the years go by, the life expectancy increases. The rate of growth of life expectancy is somewhat constant across all the continents except Africa.

Chart

Description automatically generated with low confidence

In the continent of Africa, the percentage change in the life expectancy over the years is comparatively higher. We see that there is almost a 20% increase in the life expectancy in Africa over the 15 years while there is only a 5-8% increase in life expectancy in the other continents.

We will look into three major factors which might effect life expectancy:

1. Percentage of expenditure
2. Immunization
3. Schooling

Life expectancy vs Percentage of expenditure:

As the Percentage of expenditure increases, the general trend is that the life expectancy also increases. Usually developed countries spend more on healthcare and thus we can see that more developed continents like Europe and North America have a positive trend to the percentage expenditure. However, in the case of Africa even with the increase in the percentage expenditure there is a decline in the life expectancy.

Chart, scatter chart

Description automatically generated

We wanted to dig deeper and understand the reason behind this:

Looking further into Africa, we divided the 15 years African data into buckets of 5 years to see the trend and we observed something interesting. Between the years 2000-2005 life expectancy was strongly negatively correlated i.e., as the percentage of expenditure increased life expectancy decreased which was surprising. However, as the years increases this trend goes away.

Chart, scatter chart

Description automatically generated

After some research we noticed that most of the deaths which occurred in the years 2000-2010 were mainly due to the two diseases HIV and tuberculosis.

Our hypothesis is that, due to the deaths which occurred from HIV there would have been a lot of hospitalizations which would have resulted in the increase in the health expenditure to those countries affected by it.  Hence the life expectancy kept declining even as health expenditures increased in these countries most affected by HIV and tuberculosis from 2000-2010.

A screenshot of a game

Description automatically generated with medium confidence

To further investigate, we took a cohort of those countries from Africa which had percentage of expenditure greater than 7.5 and life expectancy lesser than 50 between 2000-2005 and saw how it varied across 15 years.

As this graph illustrates, even with the increase in percentage expenditure over the years, life expectancy has remained the same in the majority of the countries like Zimbabwe and Bostwana.

We did some research and determined that these are the countries that were most affected by HIV, as you can see on the map below from the World Bank.

Map

Description automatically generated

Zambia

Zimbabwe

Botswana

Prevalence of HIV/AIDS in Africa, total (% of population ages 15–49), in 2011 ([World Bank](https://en.wikipedia.org/wiki/World_Bank))

Life expectancy vs Polio:

Chart, scatter chart

Description automatically generated

Over the years we can see that, with the increase in the Polio vaccinations, the life expectancy grew. This effect is more evident in the continents of Africa and Asia where we have a higher number of developing countries. In most of the developed countries like those present in Europe and North America we already have a Polio vaccination percentage of above 75 and hence we don’t see much of a trend here.

Life expectancy vs Schooling:

Chart, scatter chart

Description automatically generated

Every continent showed a positive trend with respect to schooling and as we can see schooling is highly correlated to life expectancy.

Correlation of the variables:

Diagram, engineering drawing

Description automatically generated

* Life expectancy has a high correlation with schooling.

* Polio and diphtheria are equally correlated with life expectancy. They both have a high correlation among themselves too. Hence, it’s better to only use one of the two columns for our modelling purpose.

We also tried to look into another relationship between the percentage change in the life expectancy and the percentage change in % expenditure and the other variables as well. But we did not get any useful insights out of it.