

WHO LIFE EXPECTANCY REPORT

Essential links

1. [Dataset used](#)
2. [Trello board](#)
3. [Python notebook](#)
4. [Tableau dashboard](#)
5. [Powerpoint presentation](#)

INTRODUCTION

Life expectancy is the number of years that an individual is expected to live as determined by statistics. It is the age to which an average person is expected to live, based on the person's sex, health factors, and other demographic data. Life expectancy at birth is one of the most frequently used health status indicators. Gains in life expectancy at birth can be attributed to a number of factors, including rising living standards, improved lifestyle and better education, as well as greater access to quality health services.

The complexity of the meaning of the concept of ageing, the nature of age-related health problems, the current trend of population ageing, the social impact of HIV and AIDS, and the prevailing socio-economic situation of a country pose new challenges in the care and well being of the people. These factors might have negative or positive impact on the quality of life of the people and influence their perceptions about the ageing phenomenon.

Population ageing has become an important developmental issue that requires urgent action. The World Health Organisation (WHO) states that "Investing in health and promoting it throughout the lifespan is the only way to ensure that more people will reach old age in good health and capable of contributing to society, intellectually, spiritually and physically" (WHO 1998a: 6).

Life expectancy is the key matrix in determining the success of the healthcare system of a country, a region and overall the world.

1) BUSINESS UNDERSTANDING

a) Problem statement

What factors contribute to life expectancy as observed from 2000-2015 in the world. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well.

b) Objectives

Specific objectives/Research questions

- a. Is there a difference between life expectancy in developing countries and developed? What contributes to this difference?
- b. How has life expectancy changed over time in the different countries of the world and in particular eastern Africa?
- c. What is driving this change in life expectancy?
- d. What is the impact of immunization coverage on life expectancy?
- e. Do countries need to improve their health care to reduce mortality rate or increase life expectancy?
- f. Which factors contribute highly to life expectancy?
- g. What is the relationship between the overall health of citizens of a particular country and the life expectancy?
- h. What factors need to be improved in the countries in Eastern Africa with low life expectancy?...i.e schooling, immunization, healthcare expenditure?

c) Approach

Python notebook: Data cleaning and tidying up

Analysis: EDA

Hypothesis testing

Reduction techniques using factor analysis

Analysis and conclusion

Report

Tableau

2) DATA UNDERSTANDING

Data collection

Data source was sourced from [Kaggle website](#). The original datasets were collected by Global Health Observatory under WHO. The datasets were merged into one dataset and provided on the kaggle website.

Data description

Description of the data variables:

Column name	Column description	Data type
Country	Country name	String
Year	Year of data collection	Integer
Status	Developing or developed	String
Life expectancy	Life expectancy age in years	Integer
Adult mortality	Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)	Integer
Infant deaths	Number of Infant Deaths per 1000 population	Integer
Alcohol	Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)	Integer
Percentage expenditure	Expenditure on health as a percentage of Gross Domestic Product per capita(%)	Integer
Hepatitis B	Hepatitis B (HepB) immunization coverage among 1-year-olds (%)	Integer
Measles	Measles - number of reported cases per 1000 population	Integer
BMI	Average Body Mass Index of entire population	Integer
Under-five deaths	Number of under-five deaths per 1000 population	Integer
Polio	Polio (Pol3) immunization coverage among 1-year-olds (%)	Integer
Total	General government expenditure on health	Integer

expenditure	as a percentage of total government expenditure (%)	
Diphtheria	Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)	Integer
HIV/AIDS	Deaths per 1 000 live births HIV/AIDS (0-4 years)	Integer
GDP	Gross Domestic Product per capita (in USD)	Integer
Population	Population of the country	Integer
Thinness 10-19 years	Prevalence of thinness among children and adolescents for Age 10 to 19 (%)	Integer
Thinness 5-9 years	Prevalence of thinness among children for Age 5 to 9(%)	Integer
Income composition	Human Development Index in terms of income composition of resources (index ranging from 0 to 1)	Integer
Schooling	Number of years of Schooling(years)	Integer

Data quantity

The final available data contained 2938 rows and 22 columns and was collected from 193 countries over a period of 15 years from 2000 to 2015.

Data quality verification and Data exploration

We previewed the head and tail of the data as well the data types and we had data types such as integers, floats and string format.

3) DATA PREPARATION

Data cleaning(validity, accuracy, completeness,consistency, uniformity)

Validity/Accuracy

The data provided was valid enough to meet the specific objectives and answer the problem statement.

We checked for outliers and anomalies and failed to drop the outliers as dropping them would leave the data frame with negligible items to analysis. Furthermore the different variables were varied between countries and over the years.

Completeness

We checked for null and missing values in the different columns. The null values found in the columns:life expectancy, adult mortality,alcohol, hepatitis-B,BMI, polio, total expenditure,diphtheria, GDP, population,thinness 10-19, thinness 5-9, income composition of resources and schooling.The null values were interpolated and then remaining null values that failed to interpolate were dropped along the row axis.

Consistency

We checked for duplicate entries and found none.

Uniformity

We dropped the trailing white spaces in the column names, standardized the column names to lowercase and replaced the white spaces with underscore. We used the strip(), lower() and replace() functions of Python Pandas library.

Data exploration

The resultant dataset had 1987 rows and 22 columns were exported to a new dataframe named 'cleaned.csv' which was used for further analysis.

4) ANALYSIS

Univariate analysis

Several visualizations tools were used and the following findings were established:

- There were more developing countries than developed countries.
- Most of the data variables appeared to be relatively normally distributed.

- The columns life expectancy, hepatitis B, BMI, polio, diphtheria, income composition of resources and schooling are negatively skewed to the left while the rest are positively skewed to the right.

Bivariate analysis

Correlation heatmap showing the pearson's coefficients for the variables was plotted and the

findings were as follows:

- Life expectancy has a negative correlation with the following fields: adult mortality, hiv/aids and thinness of both 10-19 and 5-9 years. Therefore suggesting that if more adults die in a population, the number of deaths of children aged 0-4 years increases and the more children and adolescents(age 0-19) are thin(indicator of the nutritional status), life expectancy is expected to be lower.
- Life expectancy has positive correlations with BMI, schooling and income composition of resources. This suggests that the higher the education level, the better the BMI of an individual and the higher the human development index the higher the life expectancy of the individual in the country.
- There is a positive correlation between adult mortality and hiv/aids suggesting that most deaths which occur from hiv/aids in infants aged 0-4 years and the rate of deaths of adults from 15-60 years is related. The specific relationship is not fully understood.
- There is a positive correlation between infant deaths and measles, population and thinness of both 10-19 years and 5-9 years. This suggests the higher the number of reported measles cases, the higher the population and the poorer the nutritional status of children aged 0-19, the higher the number of infant deaths.
- The positive high correlation between alcohol and income composition of resources and schooling suggests that the more people are educated and have a higher level of education the more prone to higher alcohol consumption.
- The very high positive correlation between percentage expenditure on health and the GDP of a country suggests that the countries with a higher Gdp are expected to direct more funding and resources towards their health care sector.

- Hepatitis B is positively correlated with polio and diphtheria. These are all immunizable diseases suggesting that if a child under 1 year is immunized against one they are likely to be immunised against the rest to increase the coverage of immunization of the three diseases.
- BMI is positively correlated with schooling and income compositions of the resources and negatively correlated with thinness of both 10-19 years and 0-5 years. Going to school for more years and and the higher the human development index the better the general health status.

Multivariate analysis

We performed Factor analysis as a reduction technique to investigate latent factors of life expectancy. We found 5 latent factors namely :

- Nutritional status(alcohol, BMI, thinness 5-9, thinness 10-19)
- Infant deaths per population(infant deaths, under five deaths, population)
- Immunizable disease(hepatitis-B, polio, diphtheria)
- Economic status(percentage expenditure, GDP)
- Death rates(adult mortality, HIV/AIDS)

These factors have a huge effect on a particular country's life expectancy with a combined total variance of 62%.

5) RECOMMENDATION

- Since hiv/aids related deaths in children aged 0-4 years affects life expectancy, countries increase awareness on the mother to child transmission of HIV/AIDS. The majority of children living with HIV are infected via mother-to-child transmission (MTCT), during pregnancy, childbirth or breastfeeding. Early testing of pregnant mothers to establish whether they are infected in order to start ART early and reduce risk of transmission. Saving children's life depends on early identification of those living with the infection and starting the ART early to suppress the viral load.
- Further investigation is needed to find out factors that contribute to high adult mortality rate of individuals aged 15-60 years and appropriate policies, measures and contingencies put into place to combat them in an effort to increase the life expectancy.

- Studies of human populations renowned for longevity also observe links between low calorie intake, an extended lifespan, and a lower likelihood of disease . What's more, calorie restriction may help reduce excess body weight and belly fat, both of which are associated with shorter lifespans. Nutrition also has a positive influence on life expectancy therefore countries should encourage their citizens to take nutrition into consideration. Citizens are encouraged to reduce alcohol consumption and reduce BMI in order to lead more healthier lifestyles and in the long run increase their overall life expectancy.
- Since immunization influences the life expectancy of citizens in a particular country,countries with a lower immunization rate of its children should focus or put more resources into providing and carrying out immunization in a bid to improve the life expectancy rate of its citizens in the future.
- Schooling raises the life expectancy of the citizens of a particular country. Therefore the countries with a low percentage of schooled individuals need to improve and enrol more individuals/ children into education in order to raise the life expectancy.
- Eastern African countries have seen a steady increase in life expectancy over the years with Somalia and South Sudan lagging behind. More research and studies need to go into these countries to find out why they have made little progress over the last 15 years in regard to increasing the life expectancy of its citizens.
- Countries with a higher GDP allocate more finances and resources to the health care sector therefore increasing and improving the life expectancy of its citizens. Countries are therefore encouraged to improve their economies and allocate more to their health sector and improve the health care systems. Life expectancy is dependent on healthcare. Healthcare performance is dependent on the economy and health care systems. Wealthier countries have wealthier citizens and by extension healthier citizens who lead a longer life.

6) EVALUATION

The research and analysis of factors that contribute and affect the life expectancy of a country is considered a success as the specific objectives were met. The recommended suggestions are the best given the limited data

provided/available on life expectancy from the whole world but not exhaustive. They should be taken into consideration while at the same time considering that not all recommendations apply to all countries at the same magnitude and degree.