**PSD 10**

Group Members

Eugenia Owusu Ansah

Faith Mutua

Evans Teye Matey

**EXPLORATORY DATA ANALYSIS ON LIFE EXPECTANCY (WHO)**

1. **Introduction**

Our group (PSD 10) used data from The WHO on Life Expectancy. We did an exploratory data analysis on the Life Expectancy dataset, our focus was on factors such as health care expenditure, adult mortality, alcohol consumption, past predicting factors, and immunization. We provided recommendations based on our data outputs.

1. **Analytical approach / Algorithm used**

We processed our data using the exploratory method. We used python data processing tools such as numpy, pandas and matplotlib in our data cleaning and plotting. We processed the data to ascertain the impact of factors such as health care expenditure, adult mortality, alcohol consumption and immunizations on life expectancy. The following processes were used to run the data to arrive at our results:

* We imported the required libraries to access their packages
* We imported the dataset to check the first five rows so we have an overview of the data(structure)

Data Cleaning:

* We cleaned the column/variable titles to put them in a more accessible python format, Variables such as the BMI, HIV/ AIDS has spaces from the raw dataset.
* rename thinnes\_1-19\_years column to thinnes\_10-19\_years as it was a more accurate depiction of what the variable meant.
* Missing values detection
* visualize the data to detect anomalies
* replace error values in selected columns with NaN
* check null values count/percentage
* drop the BMI column (almost 50% of data is missing)
* Outliers detection using histogram and boxplots
* count number of outliers using tukey’s method
* Winsorize the data

We later arrived at our final dataset for the analysis after our thorough data cleaning process.

1. **Hypotheses and Research Questions**

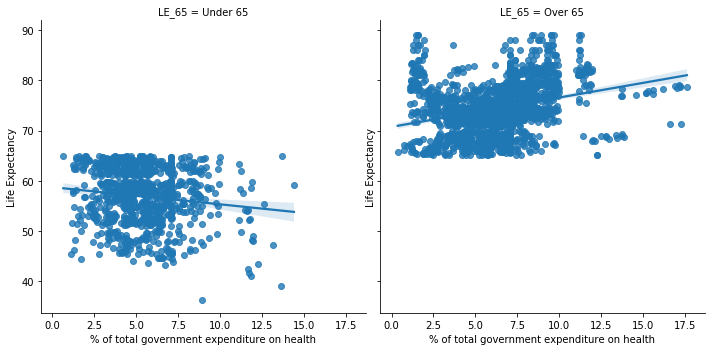
Hypotheses

* An increase in healthcare expenditure of a country having a low life expectancy value (<65) improves its average lifespan​
* Infant and Adult mortality rates have an effect on life expectancy​
* Life Expectancy has a relationship with drinking alcohol​
* Densely populated countries tend to have lower life expectancy​
* There is an impact of Immunization coverage on life Expectancy

Research Questions Raised

1. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure in order to improve its average lifespan?
2. How does adult mortality rate affect life expectancy?
3. Does life expectancy have a positive or negative relationship with drinking alcohol?
4. Do various predicting factors that has been chosen initially really affect life expectancy? What are the predicting variables affecting life expectancy?
5. What is the impact of immunization coverage on life expectancy?
6. **Results and Findings**

**“Should a country having a lower life expectancy value (<65) increase its healthcare expenditure in order to improve its average lifespan?”**



*Figure 1:* The impact of government expenditure on health correlated to Life Expectancy

*Figure 2 :* A developing country and developed country output on the impact of health expenditure on life Expectancy

The above (*figure 1)* shows that there is a positive correlation between life expectancy and government expenditure. Such that as the percentage of government healthcare increases there is an increase in life expectancy so, in increasing healthcare expenditure, a country having a lower life expectancy value (<65) can improve its average lifespan. And also **Fig 2** shows an output of a developing country Ghana’s having a lower life expectancy as a result of the government expenditure on health .

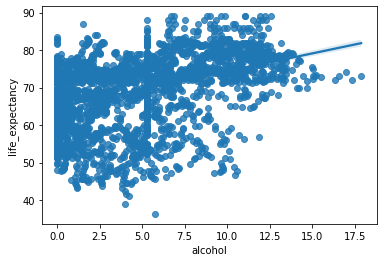
**“How does adult mortality rate affect life expectancy?”**



*Figure 3:* The effects of adult mortality on life expectancy per 1000 population

The plot above shows a negative correlation between life expectancy and adult mortality. Increase in adult mortality decreases life expectancy. Adult mortality represents the probability that a 15-year-old will die before reaching his/her 60th birthday.

**“Does life expectancy have a positive or negative relationship with drinking alcohol?”**



*Figure 4:* The impact of Alcohol Consumption on Life Expectancy

Alcohol has a positive correlation of 0.3929 with life expectancy

**“Do various predicting factors that has been chosen initially really effect life expectancy? What are the predicting variables actually affecting life expectancy?”**

Life expectancy = 0.9999999999999998),

Adult mortality = -0.6783825597940726),

Infant deaths, -0.13325091527669333),

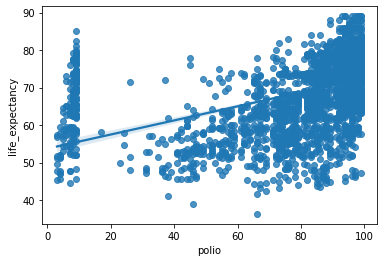
alcohol', 0.3929343679746082),

percentage expenditure', 0.3815426339299435),

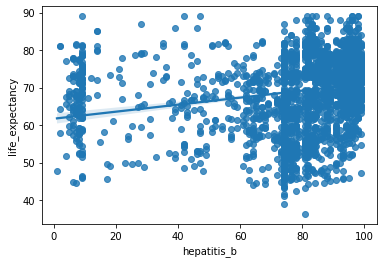
hepatitis b', 0.22338184963625946)]

The results show that initial factors chosen has effects on life expectancy, there are both positive and negative correlations between these factors and our target variable Life expectancy.

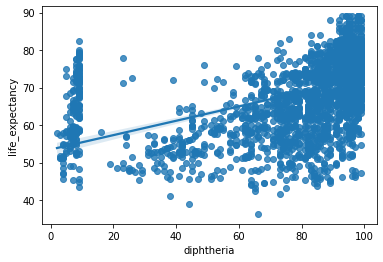
**“What is the impact of immunization coverage on life expectancy?”**



*Figure 5:* The impact of polio immunization on Life expectancy



*Figure 6:* The impact of hepatitis b immunization on life expectancy



*Figure 7:* The impact of diphtheria on life expectancy

The results in Fig 6,7,8 supports our earlier assumption that immunizations have effects on Life Expectancy. From the outputs we see a positive correlation between the type of immunization with life expectancy.

1. **Recommendation/ Implication**

Based on question One: two countries (Ghana and China) plot graph on the impact of health care expenditure on Life Expectancy.

From our results we realized that developing countries such as Ghana, Kenya, Nigeria and others had lower life expectancy due low expenditure on healthcare from their governments. Whereas developed countries such as China, Germany, France, and others have higher life expectancy rates due to the high healthcare expenditure disbursed from the government. Therefore, we recommend that governments from developing countries should increase their healthcare expenditure such as government health insurance, building more health facilities to improve the life expectancy of its citizens.

Based on the output(figure 2) for research question two we realized that there was a negative correlation between adult mortality and life expectancy, Adult mortality rate per 1000 population (ie the number of people dying between 15 – 60 years per 1000 population). Based on our output we recommend that countries or government ensure low adult mortality rate through proper health care systems. Pregnancy mortality is one of the highest contributing factors of adult mortality with 90% coming from developing countries. This implies that to increase life expectancy countries and especially developing countries should improve in maternal healthcare and general health practices.

Based on the output from **Fig 4** the results showed that there was a significant correlation between alcohol consumption and life expectancy, we run an analysis on alcohol consumption and life expectancy from the data set provided. The analysis focused on all countries provided from the dataset. The findings imply that alcohol had negative relation with life expectancy therefore going against several research that had been conducted on the impact of alcohol on life expectancy. Countries that had high level of alcohol consumption with high levels of life expectancy rates could have other influencing factors such better health facilities and social conditions influencing their life expectancy.

There are several predicting factors such as hepatitis b, life expectancy, adult mortality, infant deaths, percentage expenditure and health expenditure actually have an influence on life expectancy.

Our last research sort out to answer if immunization coverage really had an influence on life expectancy. Immunizations such as hepatitis b, polio and diphtheria have shown a positive correlation with life expectancy. We recommend that free or affordable vaccinations and immunization program be made available especially in developing countries.

1. **SUMMARY/ CONCLUSION**

In summary our group, PSD10 worked on the Life Expectancy Data using Python for data science tools such as NumPy, Pandas and matplotlib. We started with data cleaning of the dataset down to 12 variables to describe the target variable: life expectancy. It started with detecting and dealing with missing values and outliers. We converted inexplicit missing to explicit data values through the imputation method based on the means of countries.

Once the missing values were found we dealt with the outliers, we did this by using the standard box and whisker plots, each values data was winsorized on a one by one technique to remove outliers. Once all these were done data exploration on the clean data begun.

We used Bivariate analysis to understand the relationship between our dependent variable with other factors and variables. The correlation matrix in conjunction with the heatmap visual from seaborn library was used. We found out that factors such as adult mortality, government expenditure on health, immunizations, alcohol consumptions had correlations with our target variable (Life Expectancy). Our implication and recommendations centered on the impact on developing countries.

We also recommend further research into the specifiers of the factors should be done, because breaking the research into individual countries will give a clearer picture.