

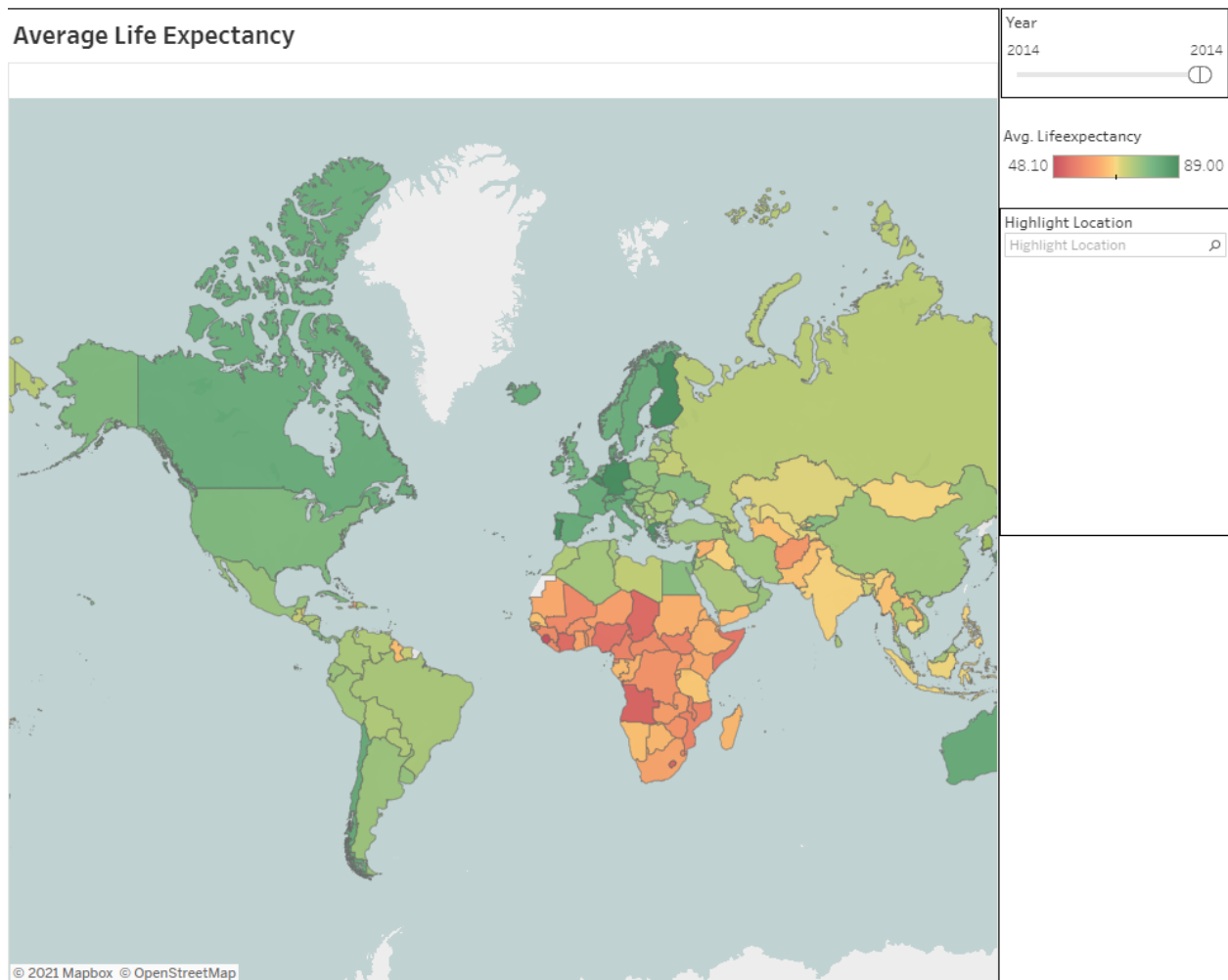
# Life Expectancy Around the World

## Executive Summary

Group 3

Abby Herrup, Mark Burton, Jim Worlein and Clarence Robinson

29 March 2021



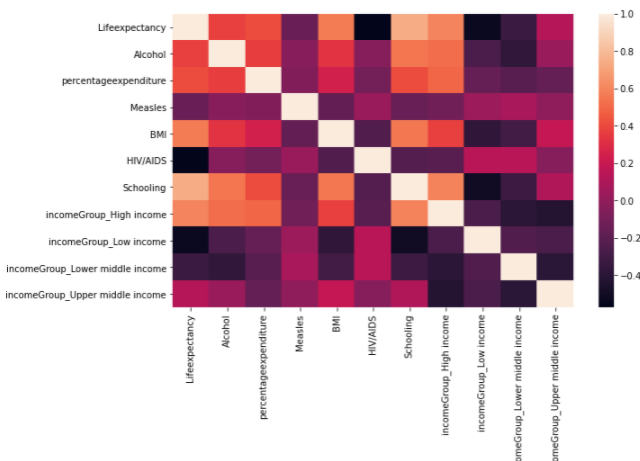
## Overview

For our capstone project, we decided we wanted to explore life expectancy around the world. We set out on a journey to unveil the data behind one of the most mysterious controversial questions we all ponder. The term “life expectancy” refers to the number of years a person can expect to live. Life expectancy is important for assessing population health. By understanding how the data is shaped by the demographics you can gain a broader more unbiased look at the world.

## Data & Modeling Approach

Our data set was posted on Kaggle and was created using health data from the World Health Organization and economic data from the United Nations. The CSV file covers 193 countries with 22 columns. The time period of the data goes from 2000 to 2015. We also found a CSV from the World Bank that had income groups for the countries over a similar timeline.

Before starting the machine learning problem to answer our questions above, we wanted to take some time to explore the data. When exploring the data, we wanted to see how correlated each of our inputs are, to get a sense of what data may be most predictive for estimating life expectancy.



## Results of Machine Learning Problem

We looked at many different machine learning models, but in the end, we selected Random Forest Regressor. This model performed very well for our data with an Out Sample R-squared factor of 93.9%. This outperformed the Gradient Boosting Regressor and XGB Regressor models which had R-squared factors of 89.4% and 92.3% respectively.

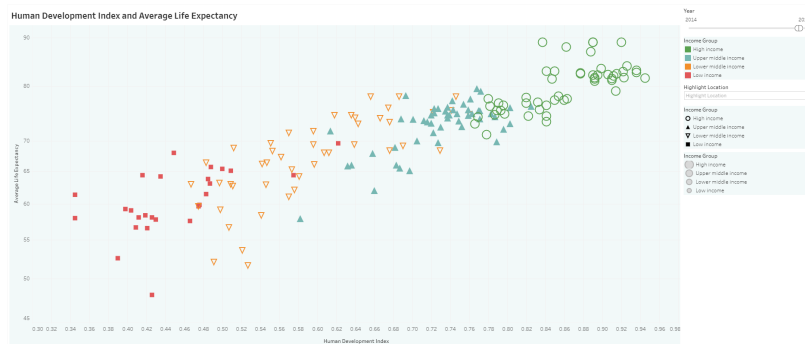
Our model found that the rate of HIV/AIDS deaths is by far the most significant health indicator in predicting life expectancy of a country. The income group of a country is also a significant indicator on the life expectancy of a country. On average, the countries grouped into the “High Income” tier of gross national income per capita had a higher life expectancy.

Input	Definition	Model Importances
Alcohol Usage	Alcohol, recorded per capita (15+) consumption (in liters of alcohol)	3.4%
Measles	Number of reported cases per 1000 population	1.7%
Percentage Expenditure on Healthcare	Expenditure on health as a percentage of Gross Domestic Product (GDP) per capita(%)	2.6%
Average BMI	Average Body Mass Index of entire population	6.7%
HIV/AIDS	Deaths per 1000 live births HIV/AIDS (0-4 years)	61.9%
Schooling	Average Number of years of schooling	10%
Income	Income group for the adjusted net national income per capita	13.7%

Key for the Machine Learning Inputs (variable definitions provided by Kaggle)

## Conclusions

There is a direct correlation between life expectancy and the Human Development Index.



Access to healthcare, economic opportunities, and education have significant impacts on life expectancies throughout the world.

When analyzing our machine learning model- most of our variables we predicted would have an effect on life expectancy did- but there was one surprise. We hypothesized that expenditure on health care would significantly impact how long one would live, however, this variable had very low significance in the predictive power of our model.

## Limitations/Bias & Future Work Recommendations

The main limitations we have encountered are (i)there are some null values in the data set. (ii) lack of data past 2015, it would be interesting to have more recent data, as socioeconomic changes happen often in developing countries.

One area that could be further looked into is the relationship between healthcare and life expectancy. Going along with healthcare, one could analyze the access to vaccines and modern medicine remedies in the country and see how differences in that could possibly affect life expectancy. Future work could also be done to see how the pandemic has shifted life expectancy in various countries, especially for countries that were especially hard hit with Coronavirus cases, such as the United States.

## Works Cited

### Data:

1. [Life Expectancy \(WHO\)](#) Kaggle, 10 February 2018
2. World Bank staff estimates based on sources and methods in World Bank's "The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium" ( 2011 ) <https://data.worldbank.org/indicator/NY.ADJ.NNTY.PC.CD>