

# Exploring Life Expectancy: An Interdisciplinary Analysis

By: Shubham Verma (22099668)

Supervisor: Ralf Napiwotzki

## Abstract

The analysis explores global life expectancy, emphasizing the positive impact of education on longer lives. The intricate relationship between adult literacy and life expectancy is directly proportional. The UK linear regression predicts life expectancy from 1990 to 2040, revealing a 10-year timeline for a 2-year increase.

## Introduction

We embark on a profound exploration into the complex mechanisms of life expectancy for year 2019. We meticulously analyze its interplay with vital elements such as GDP, adult literacy rate, and alcohol consumption. Utilizing a vast array of data from an all-encompassing dataset, we skillfully employ the technique of scatter plot visualization to unveil hidden patterns and discern emerging trends. The scatter plot and cluster analysis flawlessly portrays the intricate connection between Life Expectancy and Adult Literacy Rate, with each individual data point encapsulating the unique essence of a particular country. The proximity of the line to each distinct point provides a profound and enlightening perspective, shedding light on the subtle variations in life expectancy across diverse regions. The purpose of this visual analysis is to captivate a diverse group of three to four individuals, enabling them to grasp a more profound comprehension of the intricate and interconnected relationship between sociodemographic factors and life expectancy. Together, let us embark on an exploration to unravel the intricate tapestry of influences that mold life expectancy on a global level.

	Time	Life expectancy	Adult Literacy rate	CO2 emissions	Forest area	GDP	GDP growth	Inflation	Youth Literacy rate	Alcohol consumption
count	216.0	208.000000	27.000000	1.900000e+02	213.000000	2.090000e+02	208.000000	176.000000	41.000000	186.000000
mean	2019.0	72.986566	90.020508	1.780360e+05	31.760500	4.154190e+11	2.849367	3.828853	88.951990	5.421133
std	0.0	7.498638	15.825336	8.826307e+05	24.350862	1.862908e+12	3.502930	8.550837	18.880024	3.982776
min	2019.0	52.910000	29.663280	8.900000e+00	0.000000	5.412320e+07	-11.320755	-3.233389	20.760000	0.000000
25%	2019.0	67.262500	89.515667	2.606500e+03	10.946540	6.366000e+09	1.124295	0.830292	83.626472	1.872873
50%	2019.0	74.050171	95.736076	1.338245e+04	30.269977	2.594703e+10	2.687464	2.186902	98.409531	4.968620
75%	2019.0	78.986750	98.276428	6.418130e+04	50.255959	1.820000e+11	4.783657	3.705587	99.334396	8.521270
max	2019.0	85.180488	99.999947	1.076282e+07	97.490577	2.140000e+13	23.412017	87.241364	99.999924	16.989640

Fig 1: Data summary

## The Dataset

The dataset under review is a rich repository of sociodemographic data, including life expectancy, country names, GDP, GDP growth, inflation, alcohol consumption and adult literacy rates. The dataset consists of data from many countries, enabling a comprehensive examination of the complex relationships between life expectancy and adult literacy rates. Each row in the dataset represents a unique country and provides a complete picture of the global landscape.

The cornerstone of our analysis is life expectancy, a basic metric that captures the average life expectancy of individuals in each population. Together with the adult literacy rate, a key indicator of the education level of a population, the dataset facilitates a multifaceted exploration of possible country groupings based on common socio-demographic characteristics. The inclusion of country names provides a geospatial context to our analysis, enabling a nuanced understanding of regional differences. We aim to use cluster analysis techniques on this data set to reveal hidden patterns and similarities between countries based on their life expectancy and adult literacy rates. With clustering, we try to classify countries into distinct groups and shed light on possible correlations and differences.

This dataset and its variables are poised to provide valuable information about the interconnected nature of life expectancy and literacy, paving the way for a deeper understanding of the global socio-demographic landscape. The dataset is a treasure trove for uncovering hidden structures and trends that can impact population health and education worldwide.

Scatter Plot on Map: Alcohol Consumption and Life Expectancy

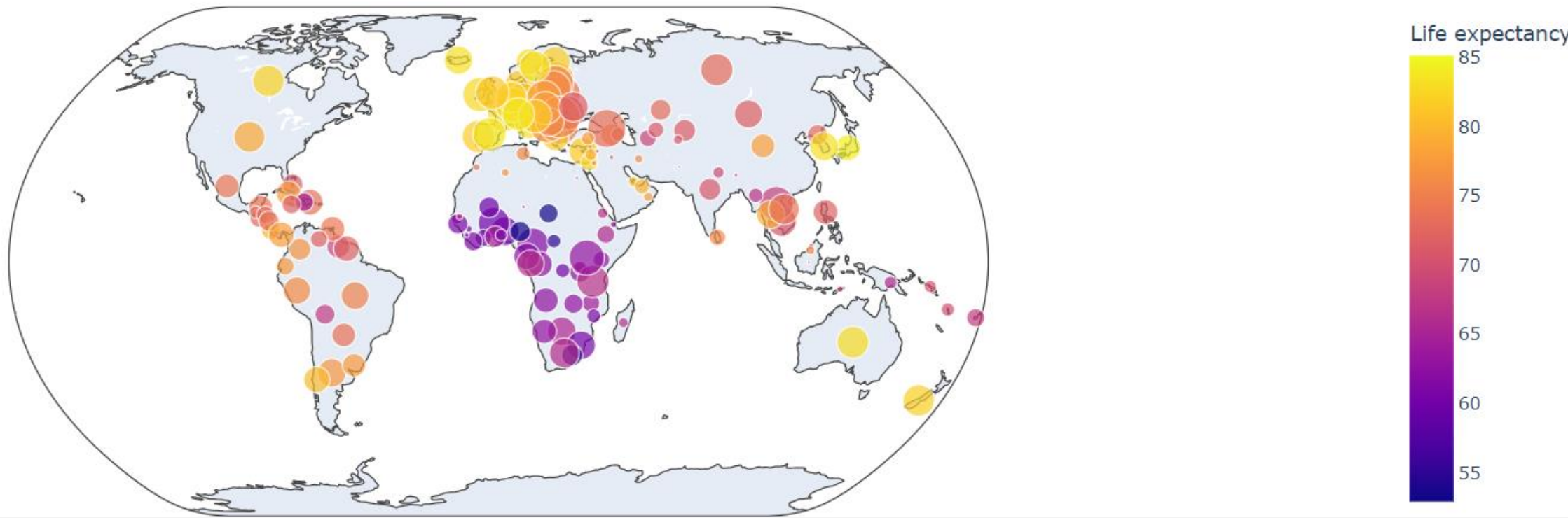


Fig 2: Life expectancy on alcohol consumption in 2019 (Bubble size: Alcohol consumption)

## Methodology

When conducting a cluster analysis of the life expectancy of different countries in 2019, a structured approach is used in the methodology, which includes the evaluation of silhouette points. The dataset, which includes variables such as GDP, GDP growth, inflation, alcohol consumption and adult literacy, undergoes careful preparation, including data cleaning and normalization. The selection of important characteristics is critical and focuses on factors that significantly affect life expectancy.

Using the K-means clustering algorithm, the silhouette points are used to determine the optimal number of clusters. The algorithm iterates through different cluster numbers, and the silhouette points help identify a configuration that maximizes intra-cluster coherence and minimizes inter-cluster divergence. Once the optimal number of clusters is determined, the model is fitted to the data. This process involves assigning each country to a specific cluster based on its life expectancy and other relevant characteristics. The silhouette score acts as a quantitative measure of cluster validity, ensuring the quality and coherence of the clusters formed. The methodology concludes with a comprehensive analysis of cluster characteristics that interprets the changing patterns of life expectancy across countries. Insights gathered from clusters can inform targeted actions and policy recommendations. This method, which combines silhouette scoring, improves the accuracy and robustness of cluster analysis to provide a nuanced understanding of lifespan dynamics.

## Preliminary Results and Conclusion

The analysis presented a comprehensive view of global life expectancy by examining the complex relationship between life expectancy and key factors (literacy level and alcohol consumption). The two cluster charts clearly show the effect of education on life expectancy, showing groups of countries where literacy rates and life expectancy are correlated. In contrast, the relationship between alcohol consumption and life expectancy is less clear, reflecting the nature of this relationship. Subsequent confirmation graphs reinforce the observed trend, confirming the positive relationship between literacy and life expectancy, and also showing a weak, negative relationship with alcohol consumption. This demonstrates the important role of education in creating life outcomes and demonstrates the broad positive impact of education in diverse regions. The Global life Index shows global disparities in life outcomes. Higher life expectancy in Europe, North America and East Asia reflects the impact of socio-economic conditions, while lower life expectancy in parts of Africa and Asia reflects the challenges facing those regions. Indeed, these analyzes show the multifaceted nature of global life outcomes. It shows how education has lifelong benefits, exemplified by adult literacy. Meanwhile, the diverse effects of alcohol consumption and the geographic and economic context add complexity to the global life structure. This broad understanding fosters important insights into health interventions, policy making, and the pursuit of equitable life outcomes worldwide.

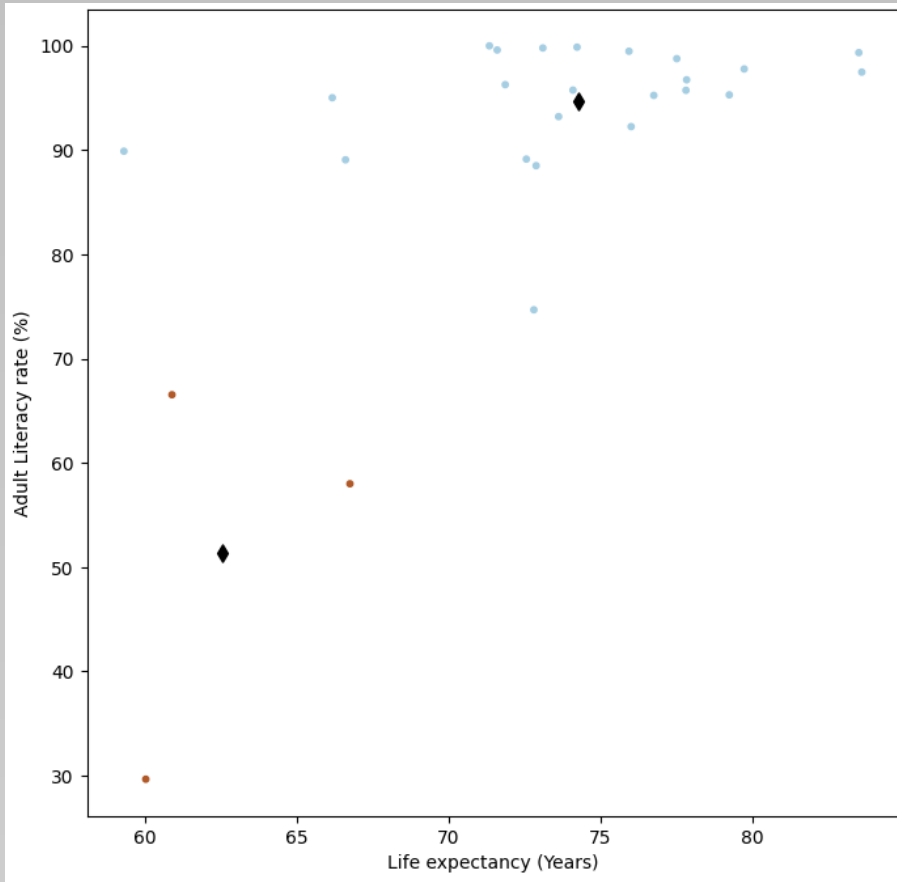


Fig 3: Life vs Literacy Cluster

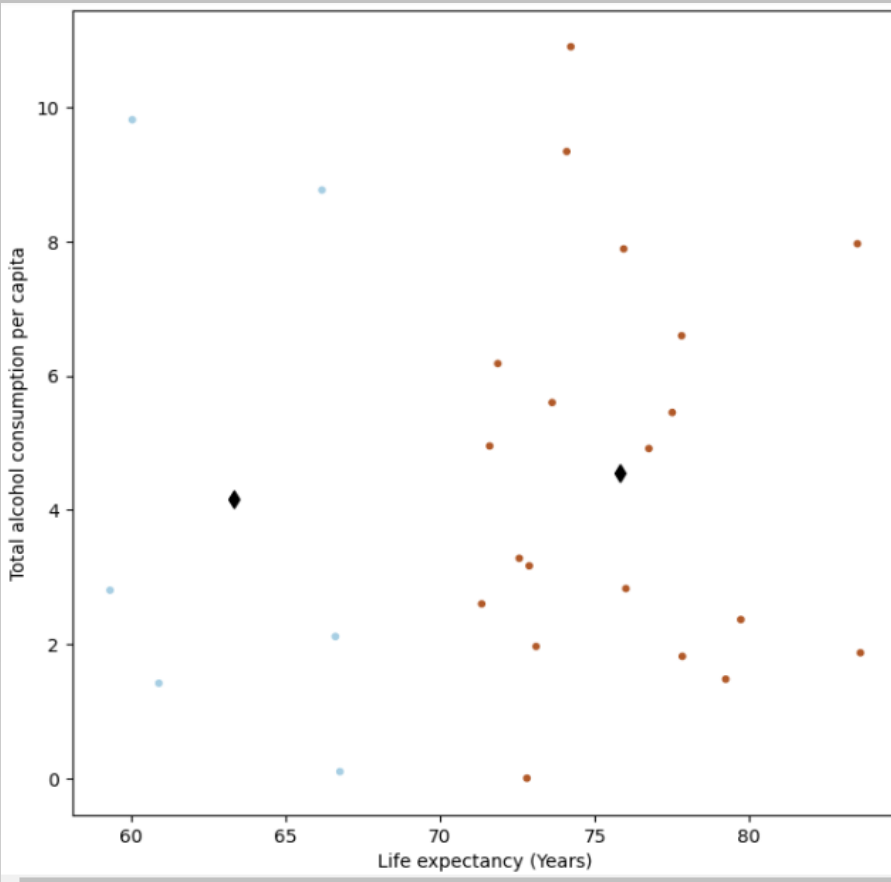


Fig 4: Life vs Alcohol Cluster

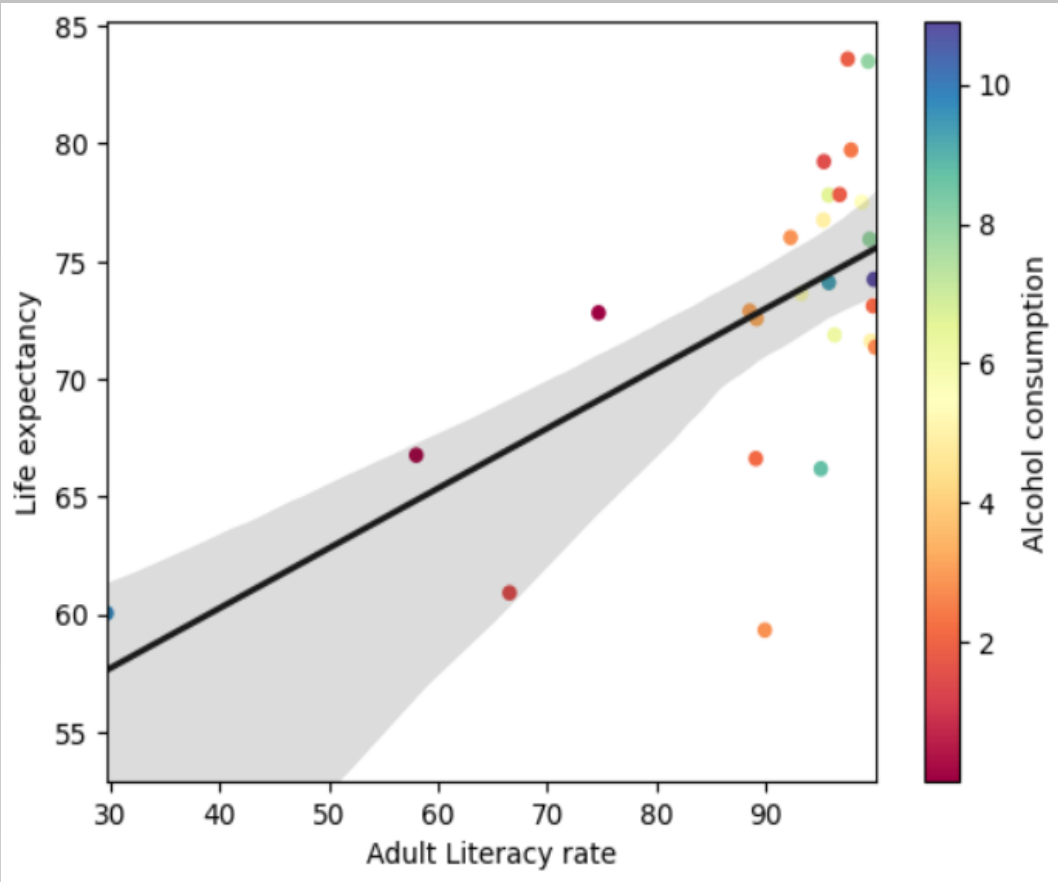


Fig 5: regression line and confidence intervals

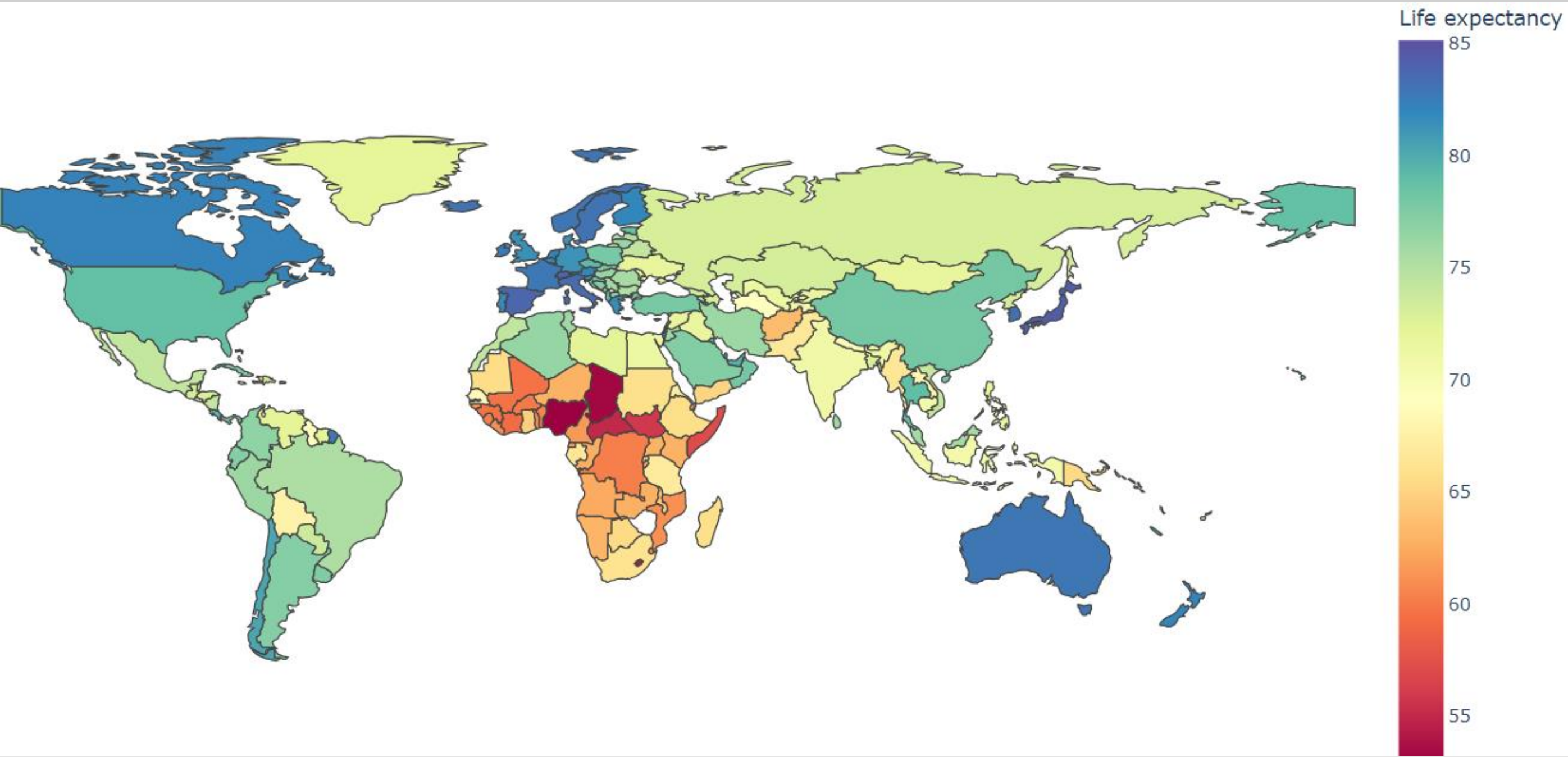


Fig 6: Worldwide life expectancy in 2019

## Prediction

The analysis involves using linear regression to examine the historical life expectancy data for the United Kingdom, covering various years. The dataset includes information on life expectancy corresponding to each year. After necessary data preparation, a linear regression model is trained to understand the relationship between time and life expectancy. The trained model is then applied to predict life expectancy for the years 1990 to 2040. The predictions are presented with confidence intervals, acknowledging the inherent uncertainty in forecasting future life expectancy. The resulting visualization combines actual and predicted values, providing a comprehensive view of potential outcomes. Also, it shows that it would takes 10 years to increase 2 years of life expectancy.

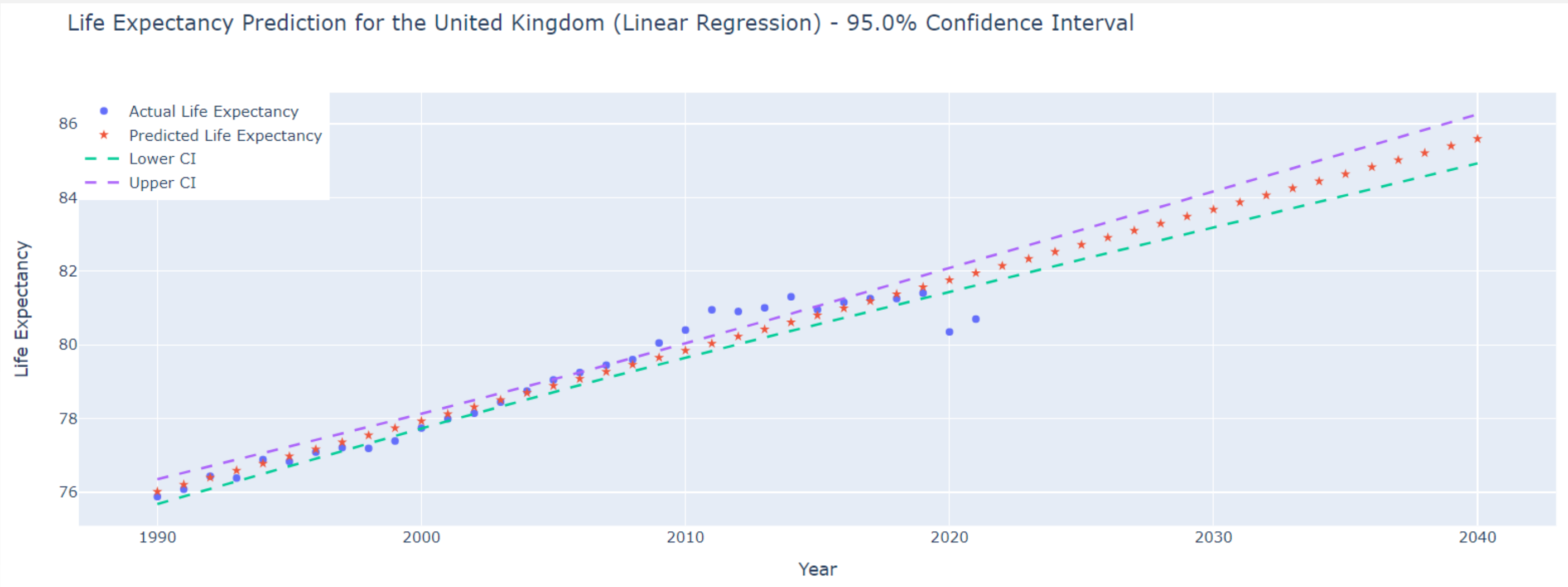


Fig 7: Life Expectancy for UK Real vs Prediction

## References

Github links: [Github](#), [Github repository](#), [Github code link](#), [Github poster pdf](#)  
Data source files: [World Data Bank](#)  
Useful resources: [Stack Overflow](#), [Google](#), [GeeksForGeeks](#), [Datacamp](#), [Hackerrank](#), [Anaconda Learning](#)