Life Expectancy and Population Correlation

Caleb Trimble, Michael Speer, Elijah Taber

May 4, 2025

# **Introduction**

As global populations shift (World Population Review, 2025), sustaining quality of life and human survival demands analyzing life expectancy and death rates. Can we generate enough life for stability, or do declining life expectancy and rising death rates threaten long-term viability?

Immigration has stabilized the U.S. population, but life expectancy trends raise important questions: Have they improved or declined? How do they correlate with population rates? Examining these patterns offers insights into our past and future.

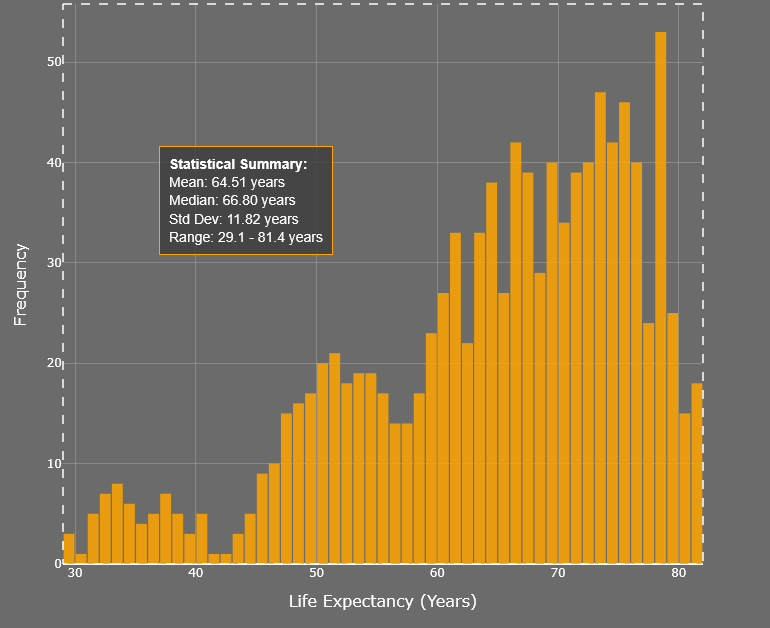
Major events—wars, epidemics, socioeconomic shifts—shape population dynamics, but is life expectancy truly tied to these disruptions? While accounting for all external factors is complex, data may reveal unexpected trends that challenge conventional assumptions.

This project explores the interplay between population size, life expectancy, and death rates, establishing an empirical foundation for predicting future trends.

# **Business Problem/Hypothesis**

Population shifts driven by life expectancy and mortality trends require proactive responses from government agencies, healthcare providers, and policymakers. As shown in Figure 1, U.S. life expectancy has generally ranged from 55 to 75 years, with a tail dipping into the 30s and 40s—a legacy of early-20th-century disease, poor sanitation, and limited medical care. Over time, rightward shifts reflect the impact of vaccines, antibiotics, and expanded healthcare access, while sharp declines in 1918 and 2020 highlight influenza and COVID-19 pandemics.

Understanding this broad span of outcomes clarifies future challenges: an aging population and slower growth may trigger labor shortages, rising healthcare costs, and economic strain. Building on Figure 1’s patterns, we next examine how age-adjusted death rate fluctuations align with these longevity trends

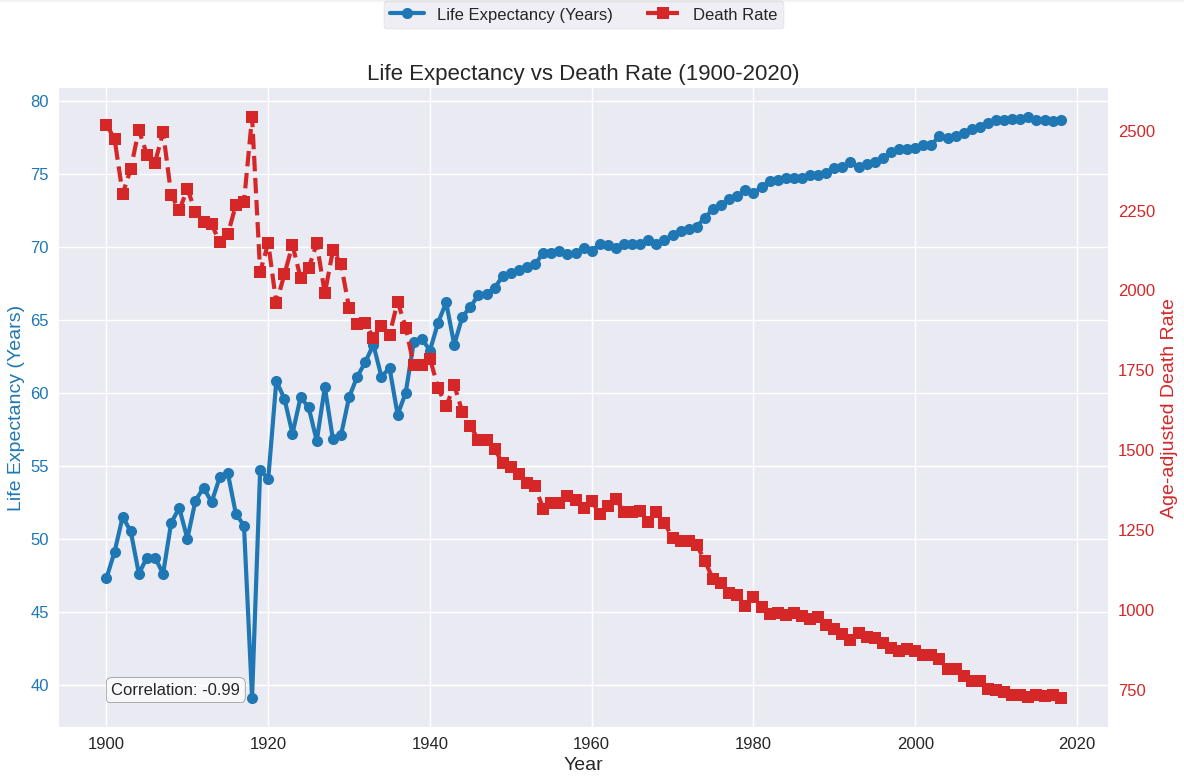
  
**Figure 1: Distribution of U.S. Life Expectancy**

# **Methods/Analysis**

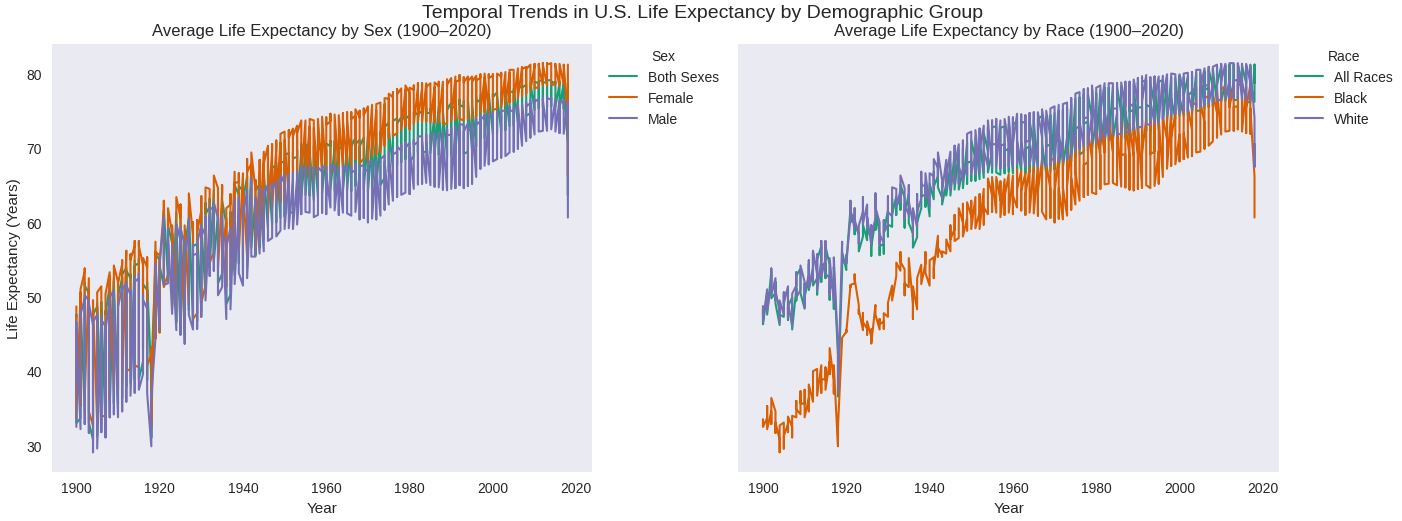
We obtained U.S. life expectancy and age-adjusted death rates from the CDC (n.d.) and decennial population counts from the Census Bureau (2022). After loading these into a Pandas DataFrame, we standardized numeric fields, converted relevant values to floats/integers, and imputed missing death-rate data (≈1.2%) using the column median to maintain distribution integrity.

To facilitate time-series comparisons, decennial population records were mapped to midpoint years (e.g., 1910→1915) and merged with annual life expectancy data. We verified data types, removed residual string entries, and conducted exploratory summaries (means, standard deviations, outlier checks) across sex and race subgroups for quality control.

Exploratory Data Analysis (EDA) examined the bivariate relationship between mortality and longevity. A scatter plot of death rate vs. life expectancy showed a strong inverse correlation (Pearson’s r ≈ –0.99), indicating reductions in mortality nearly linearly increase lifespan. Complete-case correlations were used to ensure robustness and avoid spurious imputation effects

  
**Figure 2: Life Expectancy vs Death Rate (1900 -2020)**

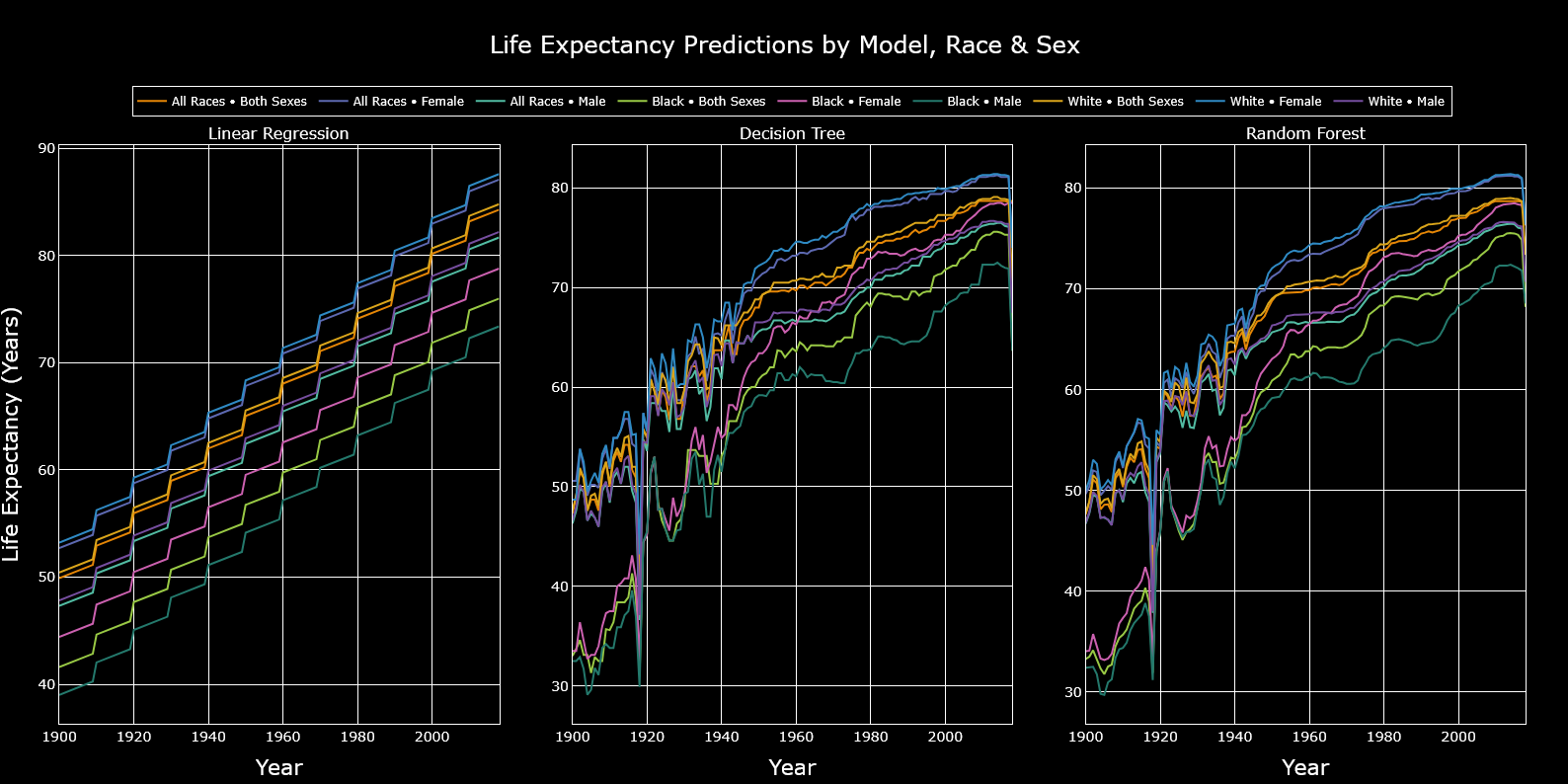
Temporal trends stratified by sex and race reveal that females consistently outlive males and White populations maintain higher average life expectancies than Black populations, yet all groups have achieved substantial gains since 1900. The line graphs (Figure 3) also show the life-expectancy downturns in 1918 and again in 2020, corresponding to influenza and COVID-19 pandemics, before each group rebounds in subsequent years. These observed patterns directly informed our selection of demographic predictors and shaped the assumptions underpinning the predictive modeling phase.

  
**Figure 3: Temporal Trends in U.S. Life Expectancy by Demographic Group**

# **Results**

Our first finding is that life expectancy is not normally distributed. It is left-skewed with most of the values occurring between the 60-80 year mark and some outliers falling toward the 20-30 year mark. This is consistent when both race and sex are held constant.

Additionally, leveraging predictive modeling we forecasted that overall, we can expect life expectancy to approach 90 by 2037. We also can see based on the steep increase shown on the Male graphs that the discrepancy between Male and Female life expectancies is expected to narrow. Additionally, because the slope shown in the Black Male graphic is the sharpest increase, the disparity between Black and White life expectancies is expected to narrow as well.



**Figure 4: Life Expectancy Prediction By Model, Grouped By Race and Sex**

The Linear Regression model achieved a Mean Absolute Error (MAE) of 2.78 years and an R² of 0.897, indicating moderate predictive accuracy. While the high R² suggests it explains ~90% of variance in life expectancy, the larger MAE reflects sensitivity to outliers and an inability to capture non-linear trends. This aligns with the model’s simplicity, as linear assumptions may overlook complex demographic interactions.

The Decision Tree Regression outperformed Linear Regression, yielding an MAE of 0.69 and R² of 0.987. Its ability to model non-linear relationships and segment data by demographic features (e.g., race, sex) reduced error significantly. However, slight overfitting is possible, as decision trees can become overly complex. The high accuracy suggests demographic-specific mortality trends are highly structured and hierarchical.

Random Forest Regression delivered the best performance (MAE: 0.54, R²: 0.991), leveraging ensemble learning to mitigate overfitting while capturing intricate interactions (e.g., pandemic impacts across subgroups). The marginal improvement over Decision Trees shows the value of aggregating multiple models to stabilize predictions. This robustness likely stems from the dataset’s temporal and demographic variability, which ensemble methods are uniquely suited to address.

# **Recommendations/Ethical Considerations**

Our recommendation would be continued, if not increased, investment in equity amongst varied demographics. While we expect to see results in line with ethical motivations, drastic changes to ongoing investments will have drastic impacts on these forecasts that were not accounted for. As we work to reduce carbon-footprints and create more walkable cities, food deserts will continue to decrease and available nutrition will continue to rise. Worker safety regulations and data-driven automation in often dangerous male-dominated fields will likely continue to be a driving force in increasing life expectancy among males.

# **Conclusion**

This report analyzed U.S. life expectancy and mortality trends to forecast demographic shifts and inform policy responses. By integrating CDC mortality data (Centers for Disease Control and Prevention, n.d.) with Census Bureau population records, we identified inverse correlations between death rates and longevity, pandemic-driven disruptions, and disparities across race and sex. Predictive modeling revealed narrowing gaps in life expectancy, projecting a rise to 90 years by 2037, emphasizing the role of healthcare advancements and socioeconomic equity in shaping future outcomes.

The superior performance of ensemble methods like Random Forest Regression (MAE: 0.54) clearly shows the complexity of demographic interactions, particularly pandemic impacts and systemic inequities. These findings urge policymakers to prioritize targeted investments in healthcare access, workplace safety, and nutrition equity to sustain projected gains. As global populations age, proactive strategies, rooted in data-driven insights, will be critical to mitigating economic strain and ensuring long-term societal stability.

# **References**

Bureau, U. C. (2022, August 6). *Historical Apportionment Data (1910-2020)*. Census.gov. https://www.census.gov/data/tables/time-series/dec/apportionment-data-text.html

Centers for Disease Control and Prevention. (n.d.). Mortality trends in the United States, 1900–2018. <https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm>

Countries with Declining Population 2025. (2025-04-30). World Population Review. https://worldpopulationreview.com/country-rankings/countries-with-declining-population

# **Appendix**

**Data Source 1**

* **Dataset**: Death Rates and Life Expectancy (1900-2018)
* **Access Date**: April 2025
* **URL**: <https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm>

**Data Source 2**

* **Dataset:** Historical Apportionment Data (1910-2020)
* **Access Date:** April 2025
* **URL:** <https://www.census.gov/data/tables/time-series/dec/apportionment-data-text.html>

**Key Variables Used in Analysis**

* **Year**: Placed in decade buckets
* **Race**: Two primary races (Black and White), polar due to data limitations in the early days
* **Sex**: Male and female sexes for better categorization and finite insight
* **Average Life Expectancy (Years)**: Average number of years expected to live
* **Age-adjusted Death Rate**: Statistical measure used to compare mortality rates between populations with different age distributions
* **Resident population**: Total population in the United States

**Tools & Libraries**

* Python: pandas, matplotlib, seaborn, scikit-learn, numpy, urllib, plotly
* Jupyter Notebook environment