Unveiling the Secrets of Global Life Expectancy

Team Members

Rahel Tekle

Saron Haile

Snit Zerea

Contents

Introduction

- Life Expectancy (LE) measures the average number of years a person is expected to live, based on current mortality rates.

- It reflects a society's health, well-being, and access to healthcare, education, and economic resources.
- So, how can we determine life expectancy?

Objective

- Identify and rank the key factors influencing life expectancy, particularly in low-income countries.
- In order to prioritize resource allocation and develop targeted, cost-effective interventions that can significantly improve health outcomes in these regions.
- Use predictive modeling to provide insights for targeted investments and cost-effective strategies in health, education, and economic development.

Data Cleaning

Removing Irrelevant or Redundant Columns:

- Dropped "thinness 5-9 years"
- Removed "Total expenditure"
- Dropped "infant deaths"
- Eliminated duplicate entries from the dataset.

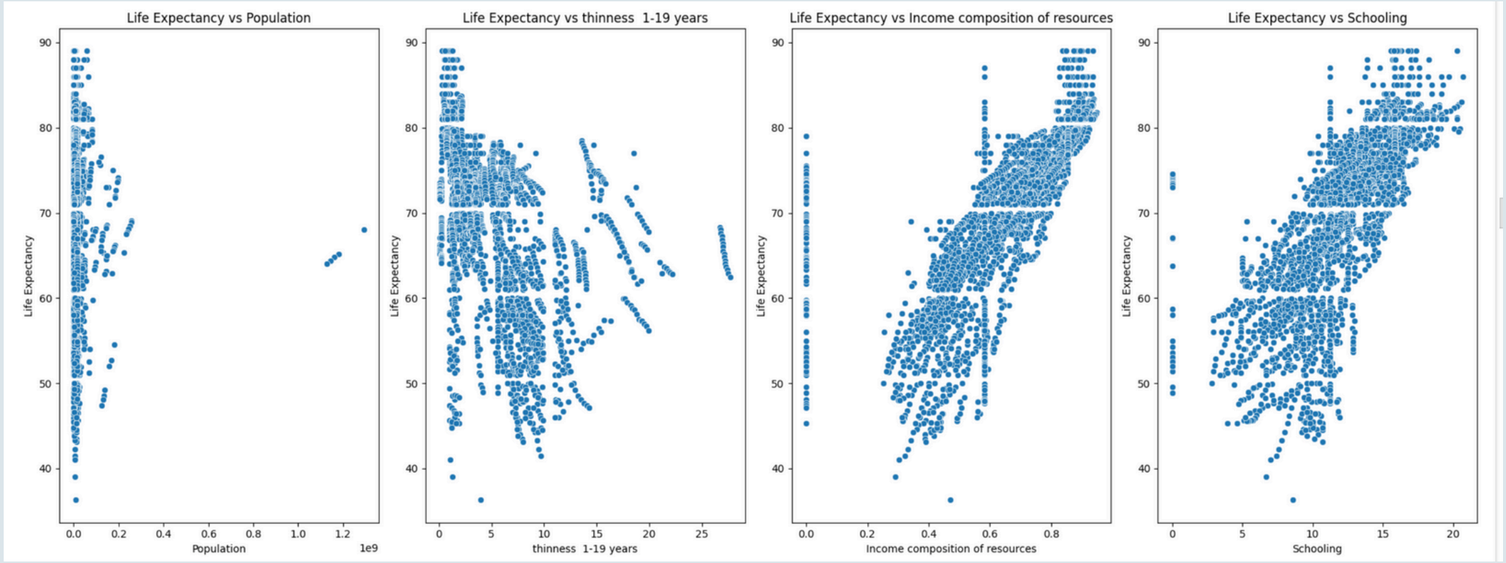
Handling Missing Data:

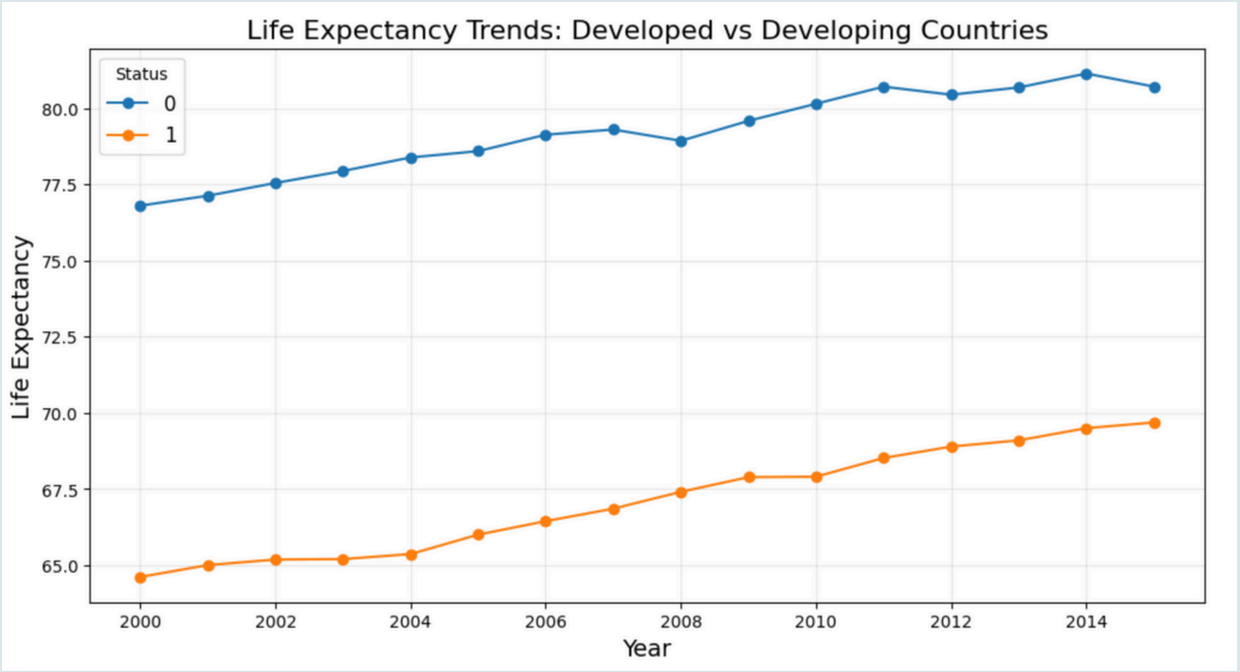
- Three-Year Average: Filled missing values using the average of the closest 3 years.
- Regional Average: Filled missing values for entire countries using the regional average based on development status.
- Excluded Countries with Excessive Missing Data: Omitted countries with more than 4 missing columns

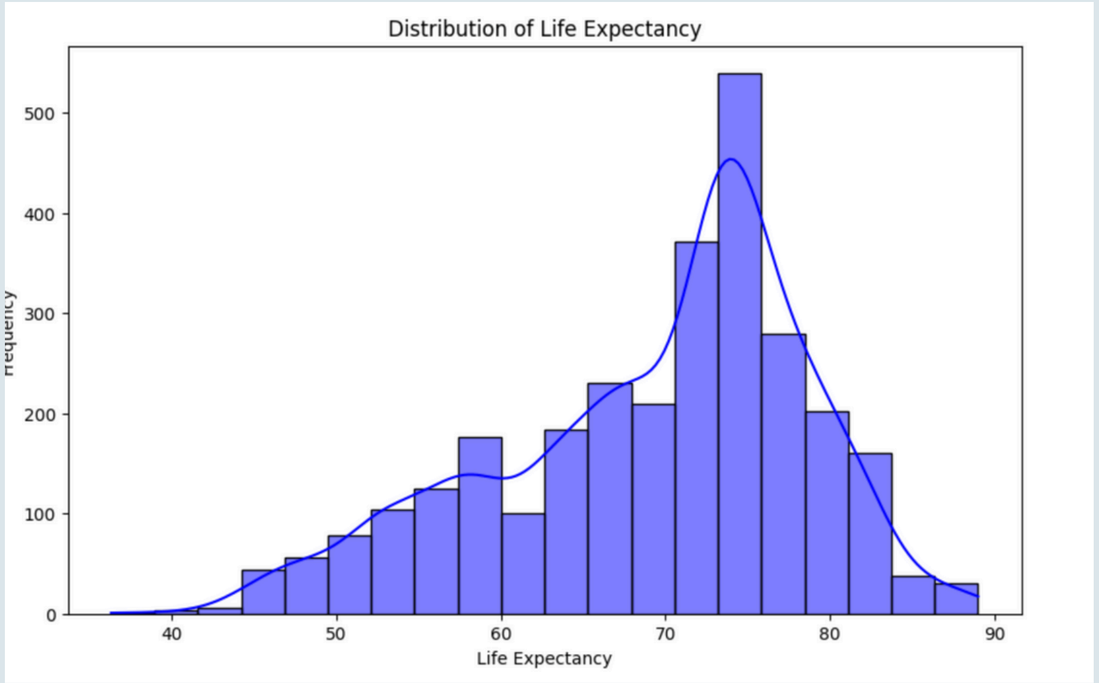
Mapping Categorical Values:

- "Status" Column: Mapped "Developed" to 0 and "Developing" to 1.

Data Exploration







Feature Selection

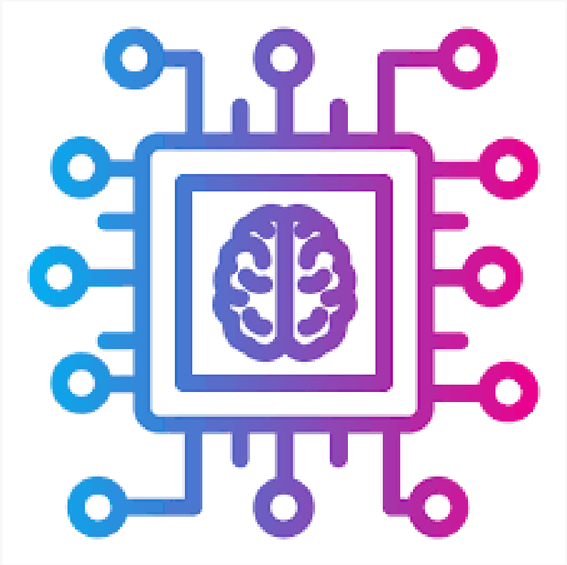
- **Most Impactful Factors:** Adult Mortality, Income Composition, Thinness (1-19 years), Schooling, and BMI, HIV/AIDS, GDP.

- **Least Impactful Factors:** Alcohol, Polio, Diphtheria, Hepatitis B, Population, Measles, and Status.

	Feature	Mutual Information
1	Adult Mortality	1.283136
14	Income composition of resources	0.945096
13	thinness 1-19 years	0.790033
15	Schooling	0.700664
6	BMI	0.573949
10	HIV/AIDS	0.528749
7	under-five deaths	0.424062
11	GDP	0.373392
2	Alcohol	0.365574
8	Polio	0.332584
9	Diphtheria	0.305194
3	percentage expenditure	0.301132
4	Hepatitis B	0.271846
0	Status	0.214052
12	Population	0.162112
5	Measles	0.113964

Machine Leaning Model Creation

Key Factors Selected
Key features were selected based on Mutual Information.



Model Created
The dataset model is built using the selected features and the best-performing algorithm.

Best Model Identified
RandomForestRegressor was selected using a function that compared evaluation metrics such as MAE, RMSE, CV-RMSE and R2

	Model	MAE	RMSE	CV-RMSE	R^2
0	GradientBoostingRegressor	1.590441	2.193252	2.868015	0.944484
1	LinearRegression	3.073920	4.278392	4.668643	0.788746
2	RandomForestRegressor	1.077722	1.693480	2.875190	0.966902
3	XGBRegressor	1.154643	1.771269	2.985875	0.963791
4	DecisionTreeRegressor	1.580612	2.666528	3.979277	0.917939
5	SVR	5.642254	7.574186	7.897013	0.337911

Conclusion

- **Key Influencers on Life Expectancy (LE):** The most significant factors impacting LE are Adult Mortality, Income Composition of Resources, Thinness (1-19 years), Schooling, BMI, HIV/AIDS, under-five deaths and GDP.
- **Targeted Investments & Cost-Effective Strategies**: Findings provide insights for guiding focused investments in health, education, and economic development to improve life expectancy (LE).
- **Optimizing Resource Allocation:** Results emphasize the importance of efficient resource allocation, ensuring better health outcomes and promoting sustainable global development.

Thank you