

LIFE EXPECTANCY

Lia Margarita Garcia Henao 20230600, Burcu Yesilyurt 20230763, Arda Kaykusuz 20230766

PROJECT AIM

The primary objective of this research initiative is to discern and analyze the pivotal determinants that exert a substantial influence on life expectancy. The overarching aim is to systematically investigate the factors that a health organization can potentially modify to augment life expectancy within a specific geographic locale.^[1] Through rigorous examination and statistical analysis, this study seeks to identify and prioritize the key variables that play a critical role in shaping life expectancy outcomes, thereby providing actionable insights for targeted interventions and healthcare strategies.

DATA SET

In this project, the data set which is ensured by the World Health Organization is used. The data set includes the data from the years 2000-2015 belonging to 193 countries. Mainly variables are come from 2 main sources. Health factors, sourced from the Global Health Observatory (GHO), and Economic factors, compiled by the United Nations (UN).

The data set contains 22 different variables as well.^[2]

The variables and their descriptions can be found here.

Variables	Description
Country	Country
Year	Year
Status	Developed or Developing status
Life expectancy	Life Expectancy of age
Adult Mortality	Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
infant deaths	Number of Infant Deaths per 1000 population
Alcohol	Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
percentage expenditure	Expenditure on health as a percentage of Gross Domestic Product per capita(%)
Hepatitis B	Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
Measles	Measles - number of reported cases per 1000 population
BMI	Average Body Mass Index of entire population
under-five deaths	Number of under-five deaths per 1000 population
Polio	Polio (PolB) immunization coverage among 1-year-olds (%)
Total expenditure	General government expenditure on health as a percentage of total government expenditure (%)
Diphtheria	Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
HIV/AIDS	Deaths per 1 000 live births HIV/AIDS (0-4 years)
GDP	Gross Domestic Product per capita (in USD)
Population	Population of the country
thinness 1-19 years	Prevalence of thinness among children and adolescents for Age 10 to 19 (%)
thinness 5-9 years	Prevalence of thinness among children for Age 5 to 9(%)
Income composition of resources	Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
Schooling	Number of years of Schooling(years)

Table 1. Variables and their descriptions

METHODOLOGY

I. Exploring Data and Preprocessing

The all variables are explored and the data distributions are examined. It is seen that the target variable is a little left skewed. (Figure 1)
According to data types, the variables having missing values were imputed by median and mode. After preprocessing, final data set is used for the analysis.

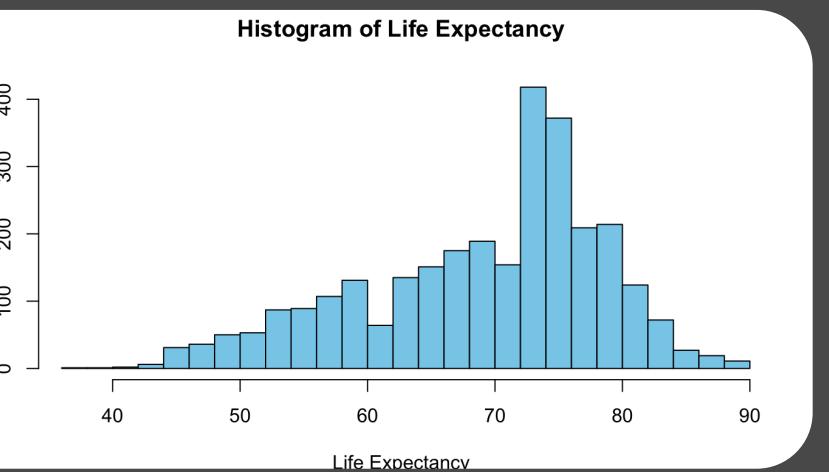


Figure 1. Changes in life expectancy of a developing and developed country over the years

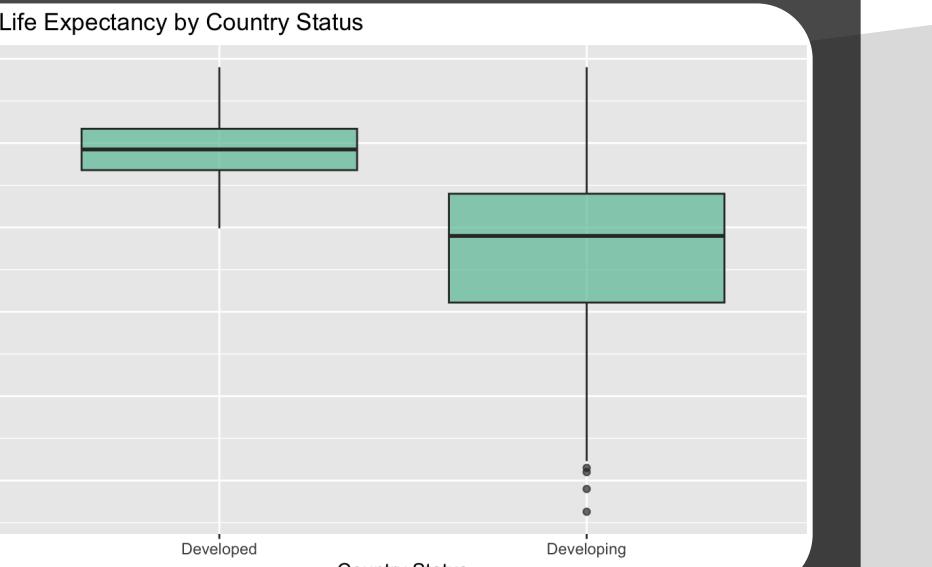


Figure 2. Life expectancy by country status

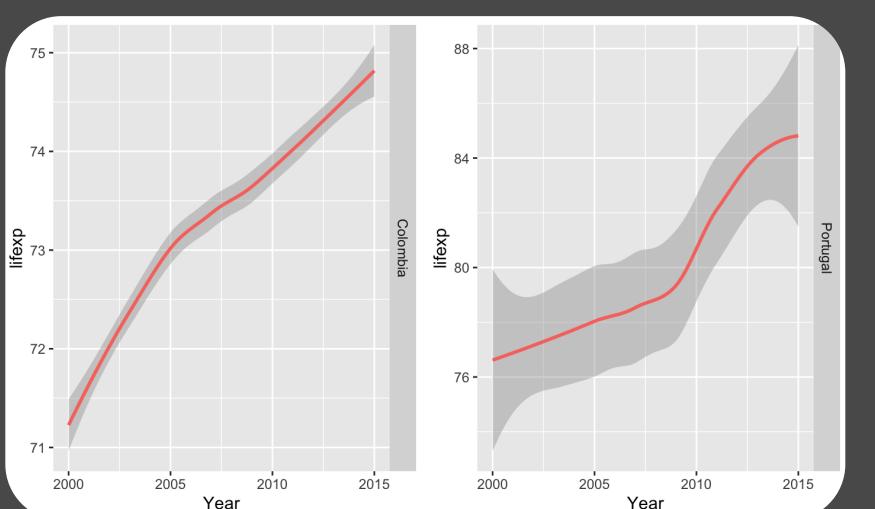


Figure 3. Changes in life expectancy of a developing and developed country over the years

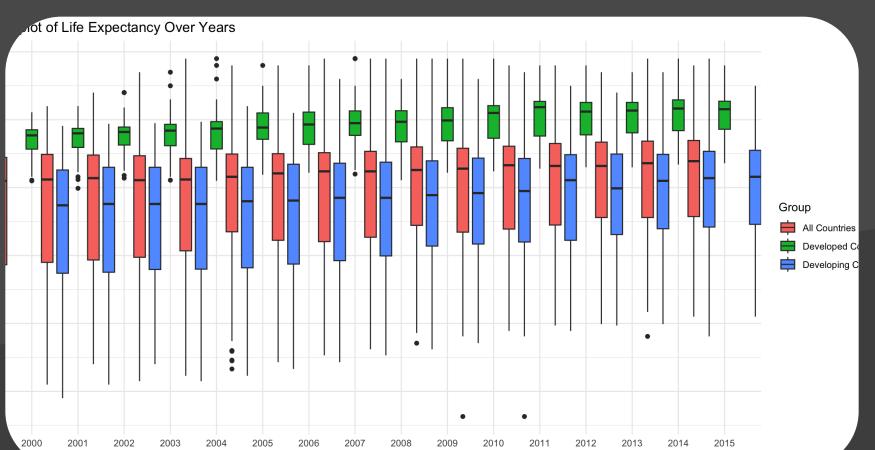


Figure 4. Life expectancies of developed and developing countries

II. Correlation & Feature & Model Selection
Initially, the correlation among non-dummy variables was assessed. Certain variables displayed moderate correlation with the variables, indicating their potential as strong predictors. Subsequently, 2D scatter plots illustrated the relationship between each variable and the target, providing insight beyond linear correlation. High correlation is observed between percepexp - GDP and infmort - BMI.

BIC is a statistical criterion used in model selection and feature selection. It is a penalized likelihood criterion that balances the goodness of fit of the model with the number of parameters used in the model in the context of feature selection.^[3] 12 features was selected for the model.

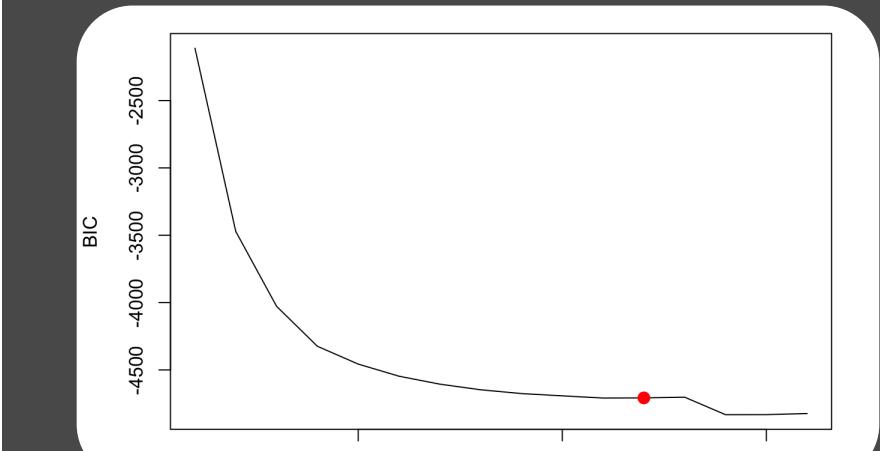


Figure 5. BIC changes according to number of features

According to the results lifexp, admort, Diphtheria, GDP, hepb, HIV, incomp_res, Measles, Polio, Schooling, Status, thinness, 5.9.years, totexp variables were used for the model.

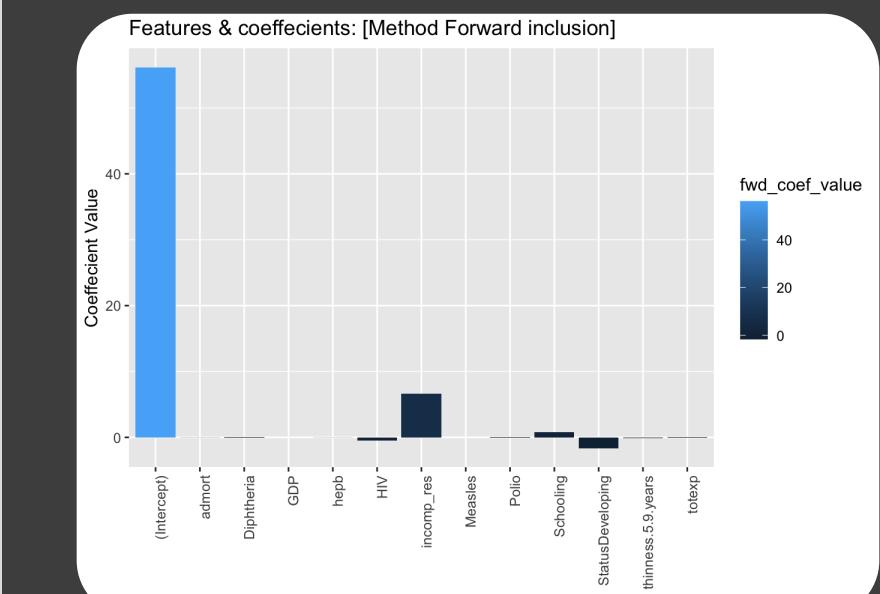


Figure 6. Features and Their Coefficients

Linear Regression model was set up and was gotten this RMSE and R-squared values respectively: 4.086219, 0.8026391

III. Conclusion

Using the life expectancy data set key factors are identified that significantly influence life expectancy. In the light of analysis, many questions have been answered, and a detailed analysis has been made on which variables affect life expectancy and how, with the data set provided by the World Health Organization. After doing analysis, it is seen that it is clearly evident that life expectancy in developed countries is growing at a faster rate than in developing countries.(Figure 2-3) Furthermore, in general, the life expectancy have been increasing over years.(Figure 4) Also, the question of should a country having a lower life expectancy value(<5) increase its healthcare expenditure in order to improve its average lifespan has been answered and it is seen that There is a significant difference in life expectancy between the two groups. Moreover, the analysis has showed that infant and adult mortality rates significantly affect life expectancy. Lastly, the analysis has revealed that Densely populated countries tend to have low life expectancy.

IV. References

- [1] Van den Heuvel, W. J. A., & Olaroiu, M. (2017). How Important Are Health Care Expenditures for Life Expectancy? A Comparative, European Analysis. *Journal of the American Medical Directors Association*, 18(3), 276.e9-276.e12. <https://doi.org/10.1016/j.jamda.2016.11.027>.
- [2] Kumar, A. (n.d.). Life Expectancy (WHO) Dataset. Retrieved from Kaggle:<https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who/data>
- [3] Bauldry, S. (2015). Structural Equation Modeling. In J. D. Wright (Ed.), International Encyclopedia of the Social & Behavioral Sciences (2nd ed., pp. 615-620). Elsevier. ISBN 9780080970875. <https://doi.org/10.1016/B978-0-08-097086-8.44055-9>.