WHO Life Expectancy Project

Description:

This project aims to develop a comprehensive regression model, incorporating mixed effects and multiple linear regression, using data spanning from 2000 to 2015 for 193 countries, to predict life expectancy for each nation. The project particularly emphasizes immunization factors, mortality rates, economic indicators, social elements, and other health-related variables. By scrutinizing these facets collectively, the research intends to provide nuanced insights into the determinants of life expectancy, guiding countries to prioritize specific areas for effective population health improvements.

Data Source:

The data originates from Kaggle, the world's largest data science community, uniting over one million registered users globally for discussions, model development, data exploration, and networking across 194 countries. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who/data

Interest in the Data Set:

My involvement in this project stems from my commitment to personal and public health. The issues addressed in the project merit global consideration and should be subjects of conversation for nations and individuals alike. This is why I am eager to conduct data analysis on this particular topic.

Contents:

The precision of our findings relies on the dependability of data from the World Health Organization (WHO). This extensive repository offers crucial health-related information for 193 countries. Life expectancy data and economic indicators from the United Nations, covering the years 2000 to 2015 were gathered, providing insight into the dynamic evolution of the health sector over 15 years. The dataset has 22 columns and 2938 rows, representing 20 predicting variables. Python software was employed for meticulous data handling, addressing missing values and other inconsistencies.

Limitations & Ethics:

Data comes from the UN and WHO, which should avoid bias. Data requires meticulous column review and cleaning for precise formatting. Temporal constraints exist (2000-2015), limiting real-time insights. Caution is advised in generalizing findings beyond the dataset, which may not fully represent the global population.

Relevance:

This dataset aligns with the project's needs due to its open-source nature, incorporation of geospatial elements, and adherence to size and variable specifications. Although over 3 years old, it meets all other requirements.
Data Profile:

Variable	Description	Time Variant/ Invariant	Structured/ Unstructured	Quantitative/ Qualitative	Discrete/ Continuous	Nominal/ Ordinal/ Binary
Country	Country	Invariant	Structured	Qualitative	N/A	Nominal
Year	Year	Variant	Structured	Quantitative	Discrete	N/A
Status	Country Developed or Developing status	Variant	Structured	Qualitative	N/A	Binary
Life Expectancy	Life expectancy in age	Variant	Structured	Quantitative	Continuous	N/A
Adult Mortality	Number of adult deaths (15-60 years) per 1000 people	Variant	Structured	Quantitative	Discrete	N/A
Infant Deaths	Number of Infant Deaths per 1000 people	Variant	Structured	Quantitative	Discrete	N/A
Alcohol	Alcohol, consumption per capita (in liters of alcohol)	Variant	Structured	Quantitative	Continuous	N/A
Percentage Expenditure	Expenditure on health as a percentage of GDP per capita	Variant	Structured	Quantitative	Continuous	N/A
Hepatitis B	Hepatitis B immunization coverage among 1-year-olds (%)	Variant	Structured	Quantitative	Continuous	N/A
Measles	Number of Measles cases per 1000 people	Variant	Structured	Quantitative	Discrete	N/A
BMI	Average Body Mass Index of entire population	Variant	Structured	Quantitative	Continuous	N/A
Under-Five Deaths	Number of under-five deaths per 1000 people	Variant	Structured	Quantitative	Discrete	N/A

Polio	Polio immunization coverage among 1-year-olds (%)	Variant	Structured	Quantitative	Continuous	N/A
Total Expenditure	Total government expenditure on health as a percentage of total government expenditure	Variant	Structured	Quantitative	Continuous	N/A
Diphtheria	Diphtheria (DTP3) immunization coverage among 1-year-olds (%)	Variant	Structured	Quantitative	Continuous	N/A
HIV/AIDS	Deaths per 1 000 births of HIV/AIDS (0-4 years)	Variant	Structured	Quantitative	Discrete	N/A
GDP	Gross Domestic Product per capita (in USD)	Variant	Structured	Quantitative	Continuous	N/A
Population	Population of the country	Variant	Structured	Quantitative	Discrete	N/A
Thinness 1-19 years	Prevalence of thinness among the ages of 10 to 19 (%)	Variant	Structured	Quantitative	Continuous	N/A
Thinness 5-9 years	Prevalence of thinness among the ages of 5 to 9 (%)	Variant	Structured	Quantitative	Continuous	N/A
Income Composition of Resources	Human Development Index in terms of income composition of resources (0 to 1)	Variant	Structured	Quantitative	Continuous	N/A
Schooling	Number of years of schooling	Variant	Structured	Quantitative	Continuous	N/A

Questions to Explore:

1. Do various predicting factors that have been chosen initially really affect life expectancy? What are the predicting variables affecting life expectancy?
2. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure to improve its average lifespan?
3. How do infant and adult mortality rates affect life expectancy?
4. What is the impact of schooling on the lifespan of humans?
5. Does life expectancy have a positive or negative relationship with alcohol consumption?
6. Do densely populated countries tend to have lower life expectancy?
7. What is the impact of immunization coverage on life expectancy?
8. How does GDP per capita affect life expectancy?