

Global Life Expectancy Data

Data Source: This data is from an open source platform and can be downloaded using this link:

<https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated>

Data Collection: This data was collected by merging many different datasets from varying sources. All sources of data are listed in the description on Kaggle.

Data Content: Data contains life expectancy, health, immunization, and economic and demographic information about 179 countries from the years 2000-2015.

Data Profile: This dataset contains 21 columns and 2864 rows. After a data quality check, there were no missing values and no duplicate values found.

Index	Column	Description	Time Variant/Invariant	Data Type
1	Country	Name of country	Invariant	Qualitative
2	Region	Region of country	Invariant	Qualitative
3	Year	Year the data was collected	Variant	Quantitative
4	Infant_deaths	Infant deaths per 1000 population	Variant	Quantitative
5	Under_five_deaths	Deaths of children under five years per 1000 population	Variant	Quantitative
6	Adult_mortality	Deaths for adults ages 15-60 per 1000 population	Variant	Quantitative
7	Alcohol_consumption	Alcohol consumption in liters of pure alcohol per capita for ages 15+	Variant	Quantitative
8	Hepatitis_B	% of coverage of Hepatitis B (HepB3) immunization among 1-year olds	Variant	Quantitative
9	Measles	% of coverage of Measles containing vaccine first dose (MCV1) immunization among 1-year-olds	Variant	Quantitative
10	BMI	BMI (body mass index) is a measure of nutritional status in adults.	Variant	Quantitative

11	Polio	% of coverage of Polio (Pol3) immunization among 1-year-olds	Variant	Quantitative
12	Diphtheria	% of coverage of Diphtheria tetanus toxoid and pertussis (DTP3) immunization among 1-year-olds	Variant	Quantitative
13	Incidents_HIV	Incidents of HIV per 1000 population aged 15-49	Variant	Quantitative
14	GDP_per_capita	GDP per capita in current USD	Variant	Quantitative
15	Population_mln	Total population in millions	Variant	Quantitative
16	Thinness_ten_nineteen_years	Prevalence of thinness (BMI < -2 standard deviations below the median) ages 10-19	Variant	Quantitative
17	Thinness_five_nine_years	Prevalence of thinness (BMI < -2 standard deviations below the median) ages 5-9	Variant	Quantitative
18	Schooling	Average years that people aged 25+ spent in formal education	Variant	Quantitative
19	Economy_status_Developed	Developed country	Invariant	Boolean
20	Economy_status_Developing	Developing country	Invariant	Boolean
21	Life_expectancy	Average life expectancy of both genders in different years from 2010-2015	Variant	Quantitative

Limitations: This dataset does not have any missing values or duplicates; the only limitation here would be the number of rows. There are only 2864 entries, which is enough for this project, but if there were more entries, we would most likely have a better analysis.

Ethics: There is no personal identifying information in this dataset, so there are no ethical concerns here.

Questions to explore:

1. Which countries have the highest life expectancy?
2. Which countries have the lowest life expectancy?
3. Is there a connection between education and life expectancy?
4. Is there a connection between alcohol consumption and life expectancy?
5. Is there a connection between immunization and life expectancy?
6. Is there a connection between economic status and life expectancy?
7. Is there a connection between BMI and life expectancy?
8. Is there a connection between child and infant mortality rates and life expectancy?