**BANL-6900-02**

**Group15**

Anirudh. Sunkara and Sumukh. Ramagiri

**What is this dataset about? Discuss briefly what things you could learn from this data set?**

This dataset is about various factors that might influence life expectancy across different countries and years. It includes data on life expectancy, adult mortality, infant deaths, alcohol consumption, expenditure on health, vaccination rates (e.g., Hepatitis B, Polio, Diphtheria), diseases (e.g., HIV/AIDS, measles), economic factors (GDP, income composition of resources), and education (schooling).

From this dataset, you can learn about the relationships between health-related factors (like vaccination rates and disease prevalence), socioeconomic factors (such as GDP and education), and life expectancy. It could be used to analyze how these variables correlate with life expectancy, potentially revealing insights into public health strategies that might improve longevity.

**Write two natural questions that can be answered using this dataset. Write which columns could be used to answer these questions.**

**1. How does the vaccination rate affect life expectancy in different countries?**

This question seeks to understand the impact of vaccination coverage on the average lifespan of a country's population. Vaccinations are crucial for preventing infectious diseases, some of which can significantly lower life expectancy if widespread within a population. By examining the relationship between vaccination rates (for Hepatitis B, Polio, and Diphtheria as indicated by the respective columns in the dataset) and life expectancy, we can assess the effectiveness of vaccination programs in enhancing public health.

* Columns to use:
* Country and Year to identify the specific context of the data.
* Life expectancy as the outcome variable we are interested in.
* Hepatitis B, Polio, Diphtheria to measure the vaccination coverage rates.

Analysis could involve comparing life expectancy in countries with high vaccination coverage to those with lower rates, controlling for year to account for global health trends.

**2. What is the relationship between GDP and life expectancy?**

This question explores the economic determinants of health by looking at the relationship between a country's economic output (GDP) and its average life expectancy. Higher GDP might indicate better access to healthcare services, better nutrition, and generally higher standards of living, all of which can contribute to longer life expectancy. By correlating GDP with life expectancy, the analysis can reveal how economic prosperity influences health outcomes.

* Columns to use:
* Country and Year for contextual identification.
* GDP as an indicator of economic status.
* Life expectancy as the health outcome of interest.

To answer this question, one might conduct a regression analysis where life expectancy is the dependent variable and GDP is the independent variable, possibly including other variables for a more nuanced analysis.

These analyses can provide insights into public health priorities and economic policies that could enhance life expectancy, highlighting the importance of vaccination and economic development as key factors in improving population health.

**Write one or two unusual (proxy) questions that can be answered using this dataset. Write which columns could be used to answer these questions.**

**1. Can the level of alcohol consumption in a country serve as a proxy for its overall health or life expectancy?**

This question whether alcohol consumption levels can indicate broader health outcomes or life expectancy trends within a country. Alcohol consumption, in excessive amounts, is known to have various negative health impacts, potentially lowering life expectancy due to diseases like liver cirrhosis, cancers, and heart disease. However, moderate consumption is sometimes associated with certain health benefits. This question posits that the pattern of alcohol consumption might reflect wider health behaviors and outcomes.

* Columns to use:
* Country and Year to distinguish the data points in time and space.
* Alcohol to measure the average consumption of alcohol per capita.
* Life expectancy to assess the overall health outcome.

To investigate this, one could analyze the correlation between alcohol consumption levels and life expectancy, potentially categorizing countries by low, moderate, and high consumption to see if there are noticeable differences in life expectancy averages.

**2. Does the number of infant deaths correlate with the country's investment in health (as measured by percentage expenditure on health)?**

This question investigates if the investment in health (indicated by the percentage of expenditure on health relative to GDP) is inversely related to negative health outcomes, specifically infant mortality. A higher health expenditure percentage could indicate better healthcare services, access to medical care, and overall health infrastructure, potentially leading to lower rates of infant deaths.

* Columns to use:
* Country and Year for contextual background.
* infant deaths as the indicator of health outcomes, particularly in the most vulnerable populations.
* percentage expenditure to measure the country's investment in health.
* Life expectancy can also be looked at to see how these factors interact with overall life expectancy.

Analyzing this relationship could involve plotting health expenditure against infant mortality rates to see if higher spending is associated with lower mortality. Regression analysis might also be used to control for other variables such as GDP, vaccination rates, or education levels.