**EDA**

1. Life Expectancy

Mean: 69.22 years

Range: From 36.3 to 89 years

Distribution: The life expectancy shows a fairly broad distribution. The lower quartile (25%) is 63.1 years, the median (50%) is 72.1 years, and the upper quartile (75%) is 75.7 years. There’s a substantial difference between the minimum (36.3) and maximum (89), indicating significant global variation.

2. Adult Mortality

Mean: 164.8 per 1000 adults

Range: From 1 to 723 per 1000 adults

Distribution: The adult mortality rate has a high level of variability. The median value is 144, but the presence of extreme values (max of 723) indicates that some countries experience much higher mortality rates.

3. Infant Deaths

Mean: 30.3 per 1000 live births

Range: From 0 to 1800 per 1000 live births

Distribution: There is a stark contrast between the minimum value (0) and the maximum value (1800), suggesting that in some regions, infant mortality is a major public health issue, while in others, it is almost negligible.

4. Alcohol Consumption

Mean: 4.6 liters per capita

Range: From 0.01 to 17.87 liters per capita

Distribution: Alcohol consumption shows a fairly broad spread, with a median of 3.76 liters. Some countries have very low alcohol consumption, while others report much higher levels, with a max of 17.87 liters.

5. Percentage of Expenditure on Health

Mean: 738.25 (currency unspecified)

Range: From 0 to 19,479.91

Distribution: There is a high degree of variability in health expenditure, from very low amounts in certain countries to extremely high spending in others. The median is significantly lower (around 64.91), suggesting that most countries spend less on health than the highest-spending nations.

6. Hepatitis B Immunization Rate

Mean: 80.94%

Range: From 1% to 99%

Distribution: The median immunization rate is 92%, with most countries reporting high vaccination coverage (especially in the 75th percentile at 97%). However, a few countries have much lower vaccination rates.

7. Measles Incidence

Mean: 2419.59 cases per 100,000 population

Range: From 0 to 212,183 cases per 100,000 population

Distribution: Measles incidence is highly variable, with a very large spread (max value of 212,183), indicating that some countries experience high rates of measles outbreaks, while others have almost no cases. The median number of measles cases is 17, suggesting a significant global disparity.

8. Body Mass Index (BMI)

Mean: 38.32

Range: From 1 to 87.3

Distribution: There is a significant range in BMI, which reflects the global prevalence of obesity and underweight conditions. The median BMI is 43.5, indicating that, on average, individuals in this dataset have a relatively high BMI, but the spread includes both extreme underweight and overweight populations.

9. Under-Five Deaths

Mean: 42.04 per 1000 children under five

Range: From 0 to 2500 per 1000 children under five

Distribution: The large range in under-five deaths indicates a significant global disparity in child mortality. Some countries have very low child mortality, while others face extremely high rates.

10. Polio Immunization

Mean: 82.55%

Range: From 3% to 99%

Distribution: Similar to Hepatitis B immunization, the spread is fairly narrow, with most countries achieving high vaccination rates. The median is 93%, indicating strong global immunization coverage.

11. Total Expenditure on Health

Mean: 5.94% of GDP

Range: From 0.37% to 17.6% of GDP

Distribution: The variability in total health expenditure is also evident here, with most countries spending between 4-7% of their GDP on health, but some nations with much higher expenditure.

12. GDP

Mean: 7483.16 billion USD

Range: From 1.68 to 119,172.74 billion USD

Distribution: GDP varies widely across the countries in this dataset. The median GDP is approximately 1767 billion USD, but a few countries with extremely high GDPs (e.g., the US, China, etc.) pull up the mean.

13. Population

Mean: 12.75 million

Range: From 34,000 to over 1.2 billion

Distribution: The population range also varies greatly, reflecting the diversity of countries included in the dataset, with a global median population of around 1.39 million.

14. HIV/AIDS

Mean: 1.74%

Range: From 0.1% to 50.6%

Distribution: There’s significant variability in HIV/AIDS prevalence, with some countries experiencing very low rates of infection, while others (primarily sub-Saharan African nations) report very high prevalence rates.

15. Schooling

Mean: 11.99 years

Range: From 0 to 20.7 years

Distribution: The median years of schooling is 12.3, indicating that many countries have achieved relatively high educational levels, though some still report very low educational attainment.

16. Income Composition of Resources

Mean: 0.63

Range: From 0 to 0.95

Distribution: This metric likely measures the quality of income distribution. The majority of countries are near or above the median of 0.68, reflecting moderate to good income equality in many regions.

A graph of life expectancy

Description automatically generatedA diagram of a life expectancy by status

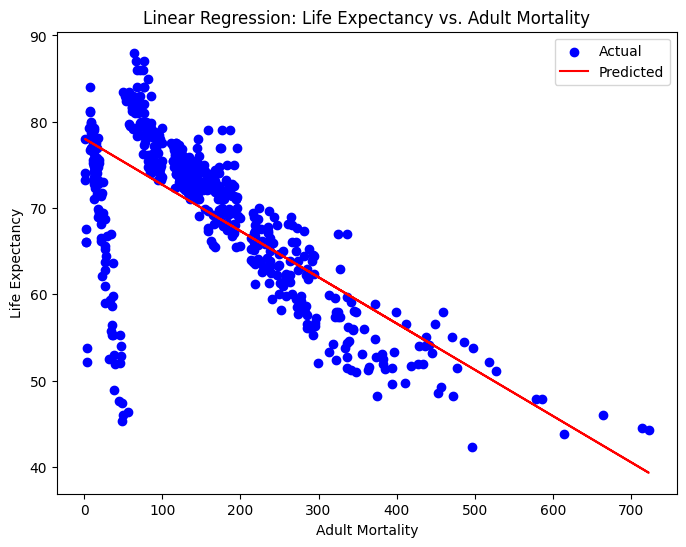
Description automatically generated

A map of the united states of america

Description automatically generatedA graph showing the growth of a number of years

Description automatically generated

**SLR**



A graph of a graph showing a line between a line and a red line

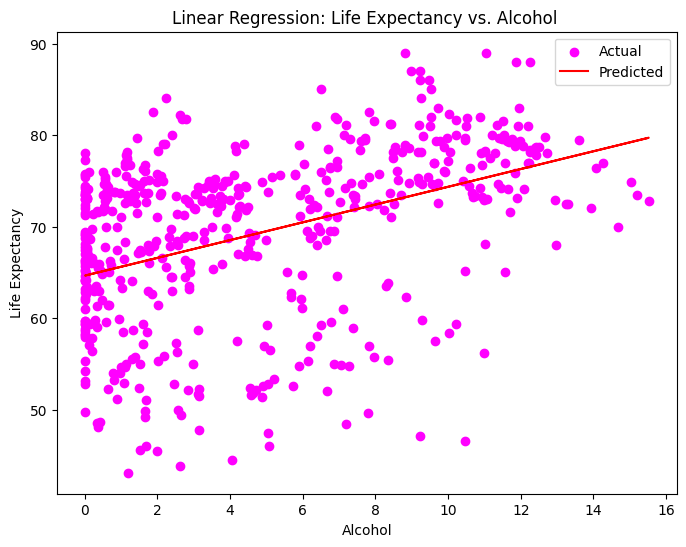
Description automatically generated with medium confidence

A graph with red dots

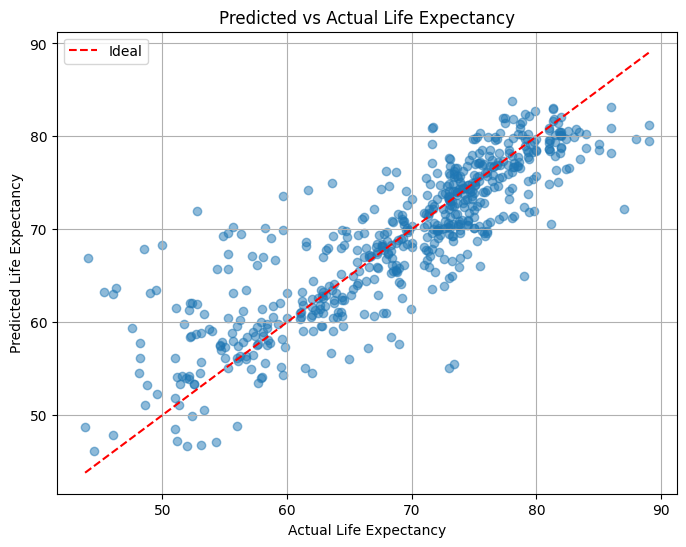
Description automatically generated

A red line with yellow dots

Description automatically generated



**MLR**



The head of the predictions DataFrame provides a side-by-side comparison of actual versus predicted life expectancy values. Here are a few examples:

| Actual | Predicted |
| --- | --- |
| 73.7 | 70.60 |
| 75.9 | 76.03 |
| 74.2 | 74.45 |
| 76.8 | 79.68 |
| 51.9 | 54.04 |

* The predicted values are fairly close to the actual values, with some deviation, especially for the lower life expectancy (e.g., 51.9 vs 54.04).

Scatter Plot Visualization:

The scatter plot compares the predicted life expectancy against the actual values, with a red dashed line representing the "ideal" scenario where the predicted values exactly match the actual values.

* Ideal Line: This line represents where the predicted values would be if the model were perfect. Points close to this line indicate better predictions.
* Scatter Distribution: Points scattered around the ideal line with some spread suggest that the model is relatively good but not perfect. The scatter allows you to see that most points fall close to the ideal line, which reflects the model's ability to predict life expectancy accurately, especially for values around the median range.

Key Insights from the Plot:

* Consistency: Most predictions are fairly close to actual values, especially in the middle range of life expectancy.
* Outliers: There might be some outliers, especially for countries with either very high or very low life expectancy values. These outliers could be due to unique factors not captured in the model or errors in the data.

Next Steps for Improvement:

* Feature Engineering: You could consider adding other relevant factors (e.g., healthcare infrastructure, public policies) to improve the model.
* Model Tuning: Try more complex models like Random Forest or Gradient Boosting, which might better capture non-linear relationships and interactions between variables.