Personal Portfolio Site

An Exploratory Data Analysis of Life Expectancy

1. Data Overview

The dataset titled "Life Expectancy" presents comprehensive information on key health and demographic indicators across 182 countries. It focuses on mortality and survival statistics, making it a valuable resource for examining the relationship between early-age mortality and average life expectancy, and for comparing overall health outcomes across world regions.

It contains 182 rows and 6 columns, each representing specific variables that together illustrate global health patterns:

- **Country:** The name of each country represented in the dataset (e.g., Japan, France, Nigeria, Brazil).
- **Region:** The broader geographical area to which each country belongs (e.g., Asia, Europe, Africa).
- **Infant_deaths:** The number of infant deaths per 1,000 live births, indicating the quality of maternal and neonatal care.
- **Under_five_deaths:** The number of deaths of children under age five per 1,000 live births, reflecting child healthcare and nutrition.
- **Adult_mortality:** The adult mortality rate per 1,000 population, showing the impact of diseases, healthcare access, and living conditions.
- **Life_expectancy:** The average number of years a newborn is expected to live, representing the overall health, environment, and economic situation of a country.

	Country	Region	Infant_deaths	Under_five_deaths	Adult_mortality	Life_expectancy
1	Country	Region	Infant_deaths	Under_five_deaths	Adult_mortality	Life_expectancy
2	Afghanistan	Asia	71.08	98.61	265.80	59.66
3	Albania	Rest of Europe	15.26	17.14	83.13	75.95
4	Algeria	Africa	26.76	31.19	113.44	73.79
5	Angola	Africa	88.77	144.16	297.84	52.83
6	Antigua and Barbuda	Central America and Caribbean	9.48	11.52	142.48	75.35
7	Argentina	South America	13.85	15.51	127.91	74.86
8	Albania	Rest of Europe	15.26	17.14	83.13	75.95
9	Armenia	Rest of Europe	19.03	21.39	137.94	72.99
10	Australia	Oceania	4.31	5.16	65.96	81.11
11	Austria	European Union	3.81	4.60	78.31	79.98
12	Azerbaijan	Asia	39.46	46.34	144.66	69.73
13	Bahamas, The	Central America and Caribbean	13.55	15.84	173.48	72.21
14	Bahrain	Middle East	8.31	9.69	75.39	75.66
15	Bangladesh	Asia	45.00		148.11	68.73
16	Barbados	Central America and Caribbean	13.90	15.11	112.07	78.10
17	Belarus	Rest of Europe	5.79	7.52	223.07	70.29
18	Belgium	European Union	3.90	4.76	85.44	79.55

Although the original source of the dataset is not explicitly mentioned, it is likely derived or compiled from official international databases such as the World Health Organization (WHO), World Bank, or the United Nations Development Programme (UNDP). These organizations

regularly publish country-level mortality and life-expectancy data that align closely with the variables presented here.

From an analytical perspective, this dataset is useful for exploring:

- The correlation between mortality indicators (infant, under-five, and adult) and average life expectancy, which helps identify how improvements in healthcare systems and economic development impact longevity.
- Regional disparities in health outcomes, illustrating global inequality in access to healthcare, sanitation, and living conditions.
- Patterns that may guide policy analysis, allowing comparisons between developed and developing regions and highlighting where public-health investment is most needed.

2. Data Cleaning

2.1. Handle missing values

1	Country	Region	Infant_deaths	Under_five_deaths	Adult_mortality	Life_expectanc
15	Bangladesh	Asia	45.00		148.11	68.73
68	Ghana	Africa	51.06	76.24		59.84
82	Iraq	Middle East	30.44	37.33		68.78

During the cleaning process, I found three missing values in the dataset: one in *Under_five_deaths* and two in *Adult_mortality*. These missing values appeared in Bangladesh, Ghana, and Iraq. Missing data can affect the accuracy of averages and comparisons, so it needed to be handled carefully before analysis.

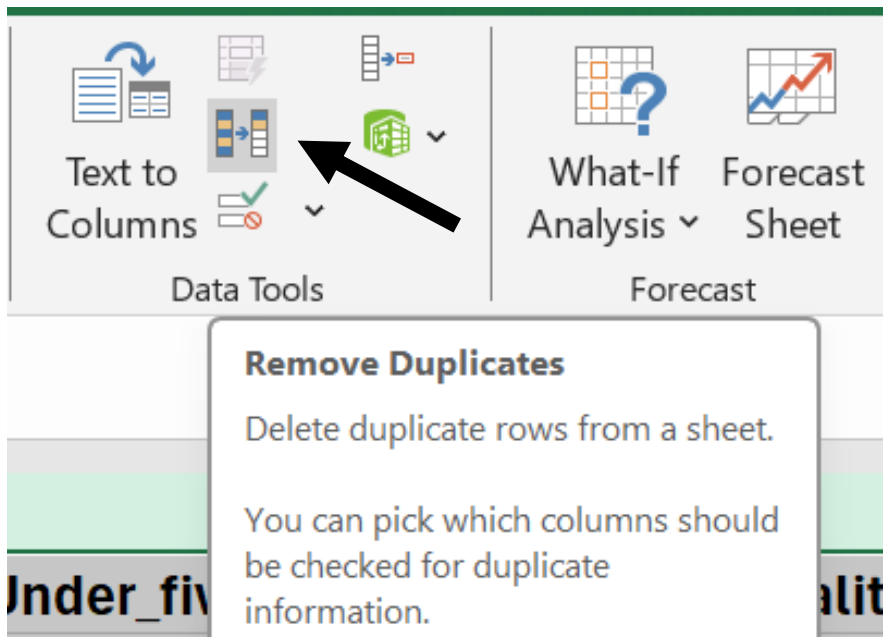| Row Labels	Average of Under_five_deaths	Average of Adult_mortality
Africa	94.03	320.03
Asia	39.17	168.45
Central America and Caribbean	25.93	163.58
European Union	5.86	104.40
Middle East	20.46	106.39
North America	11.65	100.69
Oceania	32.49	162.89
Rest of Europe	11.55	135.12
South America	23.63	155.14
Grand Total	**42.40**	**190.56**

Regional Average of Mortality

To fix this, I created a Pivot Table in Excel named *"Regional Average of Mortality"*. I selected the dataset and inserted a Pivot Table. Then, I placed Region in the *Rows* field and both *Under_five_deaths* and *Adult_mortality* in the *Values* field. I changed the calculation type from

Sum to *Average*, so the table showed the average mortality rate for each region. This made it easy to see typical values by region and use them to fill in the blanks.

	Country	Region	Infant_deaths	Under_five_deaths	Adult_mortality	Life_expectanc
15	Bangladesh	Asia	45.00	39.17	148.11	68.73
68	Ghana	Africa	51.06	76.24	320.03	59.84
82	Iraq	Middle East	30.44	37.33	106.39	68.78

Next, I referred to the Pivot Table to replace the missing data. Each blank cell was filled with the average value of the same region — for example, using the Asia average for Bangladesh, the Africa average for Ghana, and the Middle East average for Iraq. This method was chosen because countries in the same region often have similar living conditions and healthcare systems. Using regional averages helps keep the data realistic and balanced without changing the overall pattern.
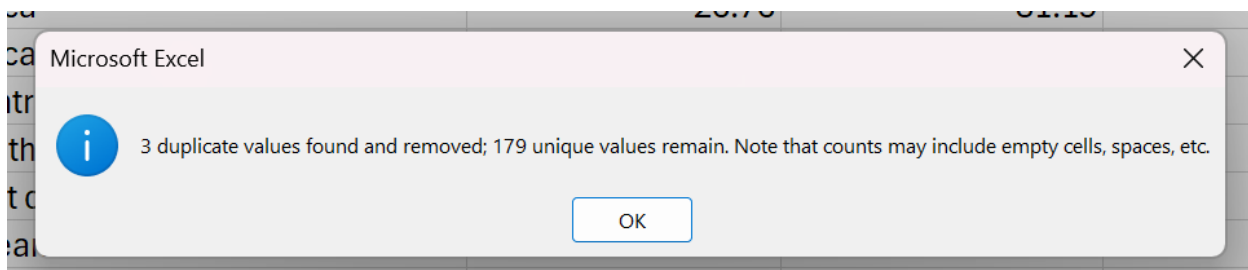
2.2. Remove Duplicates

After handling the missing data, I checked for duplicate records to ensure that each country appeared only once. I used the ***Remove Duplicates*** function in Excel by selecting all columns and confirming whether any repeated rows existed.

Data → Data Tools → Remove duplicate.

I opened the Remove Duplicates dialog box in Excel, ticked the option *"My data has headers,"* and clicked *"**Select All**"* so that every column would be compared.

Then, I pressed **OK**, and Excel displayed a message showing that **three duplicate rows were found and removed.** This step helped keep the dataset clean and prevented repeated data from affecting the analysis results.

Finally, I used the ***COUNTBLANK*** function to confirm that there were no remaining empty cells. After completing these steps, the dataset became complete and consistent.
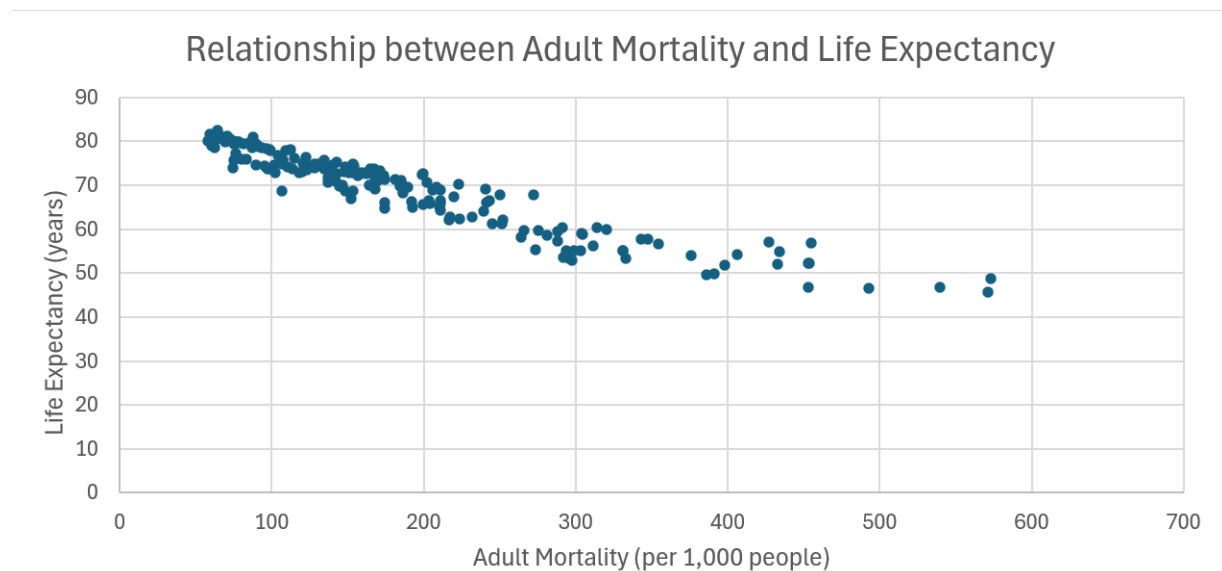
3. Descriptive Statistics

Infant_deaths		Under_five_deaths		Adult_mortality		Life_expectancy	
Mean	30.3637919	Mean	42.8307	Mean	192.151998	Mean	68.85607542
Standard Error	1.99755652	Standard Error	3.20665	Standard Error	8.37790105	Standard Error	0.687468301
Median	19.36875	Median	23.1375	Median	162.594406	Median	71.50625
Mode	#N/A	Mode	4.6	Mode	#N/A	Mode	59.65625
Standard Deviation	26.72548478	Standard Deviation	42.9021	Standard Deviation	112.088677	Standard Deviation	9.197699004
Sample Variance	714.2515369	Sample Variance	1840.59	Sample Variance	12563.8715	Sample Variance	84.59766697
Kurtosis	-0.01831644	Kurtosis	0.43931	Kurtosis	1.1637532	Kurtosis	-0.48181922
Skewness	0.982615935	Skewness	1.20376	Skewness	1.20184444	Skewness	-0.6892859
Range	113.3375	Range	175.725	Range	515.264	Range	36.85
Minimum	2.38125	Minimum	3	Minimum	57.7103125	Minimum	45.60625
Maximum	115.71875	Maximum	178.725	Maximum	572.974313	Maximum	82.45625
Sum	5435.11875	Sum	7666.7	Sum	34395.2076	Sum	12325.2375
Count	179	Count	179	Count	179	Count	179

After completing the cleaning process, I performed descriptive statistical analysis to explore the overall health trends and differences among countries. The analysis focused on four key variables: *Infant_deaths, Under_five_deaths, Adult_mortality*, and *Life_expectancy*. The descriptive statistics were calculated using Excel's **Data Analysis ToolPak**, which produced a detailed summary of the dataset's averages, variations, and ranges.

The results show that there is considerable variation in mortality and life expectancy across countries. The mean *Adult_mortality* (192.15) is much higher than the averages of *Infant_deaths (30.36)* and *Under_five_deaths (42.83)*, suggesting that adult health problems and chronic diseases contribute significantly to overall mortality differences. Meanwhile, the average Life_expectancy (68.86 years) indicates that most countries fall within the 65–75 year range, but the range of 36.85 years highlights striking inequalities between regions with advanced healthcare and those with limited medical access.

3.1. Insight 1: Relationship Between Adult Mortality and Life Expectancy A scatter plot was created to illustrate the relationship between *Adult_mortality* and *Life_expectancy.*

Figure 1: Scatter plot showing a negative relationship between adult mortality and life expectancy.

The scatter plot shows a clear negative relationship between adult mortality and life expectancy. Countries with higher adult mortality rates tend to have much lower life expectancy. This pattern means that when the number of adult deaths increases, the average lifespan of the population decreases.

The result reflects real-world differences in healthcare quality and living standards. In countries with strong healthcare systems and better access to medical services, adult mortality is lower and
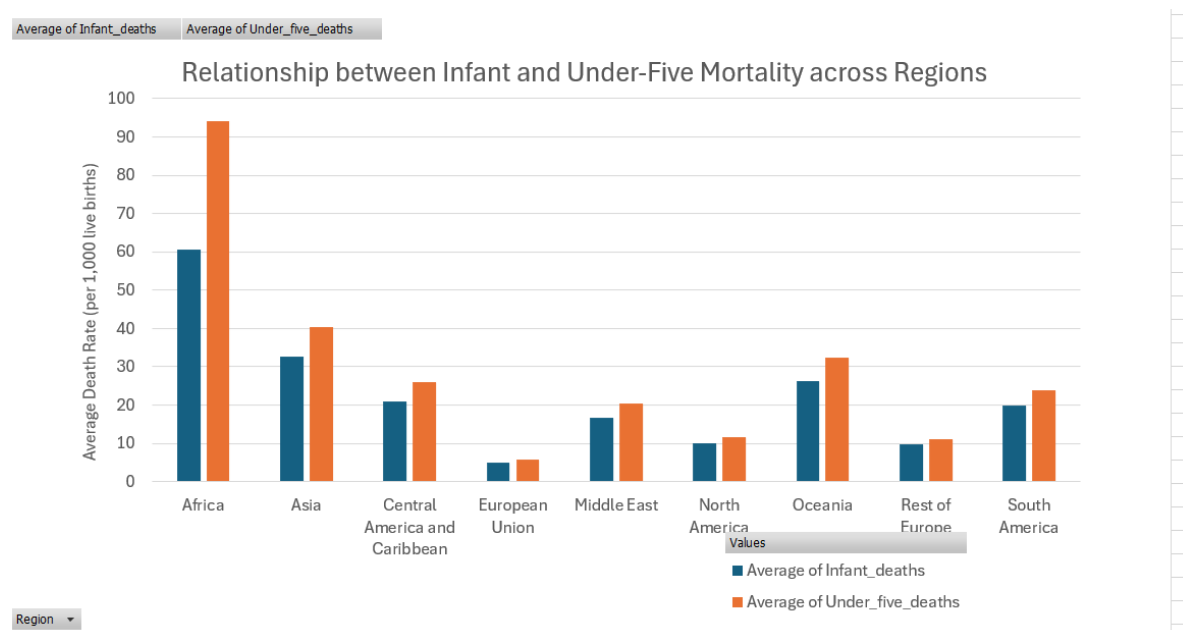
people live longer. On the other hand, in regions where healthcare is limited, adult mortality is higher and life expectancy drops significantly.

This finding highlights that improving healthcare for adults—such as disease prevention, medical accessibility, and healthy living conditions—plays a crucial role in increasing life expectancy worldwide.

3.2. Insight 2:

Region	Average of Infant_deaths	Average of Under_five_deaths
Africa	60.60	94.03
Asia	32.54	40.25
Central America and Caribbean	20.97	25.93
European Union	4.87	5.86
Middle East	16.64	20.46
North America	9.93	11.65
Oceania	26.31	32.49
Rest of Europe	9.63	11.18
South America	19.95	23.94
Grand Total	**30.36**	**42.83**

A Pivot Table and column chart were created to examine the relationship between *Infant_deaths* and *Under_five_deaths* across different regions.

Figure 2: *Column Chart – Relationship Between Infant and Under-Five Deaths Across Regions.*

The chart shows a clear positive relationship between the two variables, meaning that regions with higher infant mortality rates also have higher under-five mortality rates.

Africa records the highest average in both categories, followed by South America and some parts of Asia, where limited access to healthcare and sanitation remains a serious problem. In contrast, Europe and North America show very low death rates, reflecting better healthcare quality, maternal care, and nutrition.

This relationship suggests that child mortality is closely tied to regional healthcare development. Regions with stronger medical infrastructure, vaccination programs, and social support systems are able to reduce both infant and under-five deaths more effectively.