DATA ANALYTICS WITH POWER BI

PROJECT REPORT

August 2025 - January 2026

GLOBAL HEALTH STATISTICS

Submitted by

Rajat Rajan

Registration No. 12309843

Programme and Section - B. Tech (K23KR)

Course Code - INT374

Under the Guidance of

Dr. Mrinalini Rana

(22138)

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Rajat Rajan bearing Registration no. 12309843 has completed INT374 project titled, **"Global Health Statistics"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Dr. Mrinalini Rana
School of Computer Science
Lovely Professional University
Phagwara, Punjab.

Date:

DECLARATION

I, Rajat Rajan, student of Bachelor of Technology under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 16-12-2025 Signature

Registration No. 12309843 Rajat Rajan

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to **Dr. Mrinalini Rana Ma'am** for her valuable guidance, continuous support, and constructive feedback throughout the completion of my ***INT374*** project titled ***Global Health Statistics***. Her expertise and encouragement played a crucial role in shaping this project.

I am thankful to the **School of Computer Science**, **Lovely Professional University**, Phagwara, Punjab, for providing the necessary resources, facilities, and a supportive academic environment that made this work possible.

Finally, I would like to express my heartfelt gratitude to my family for their constant encouragement, support, and belief in me, which motivated me to successfully complete this project.

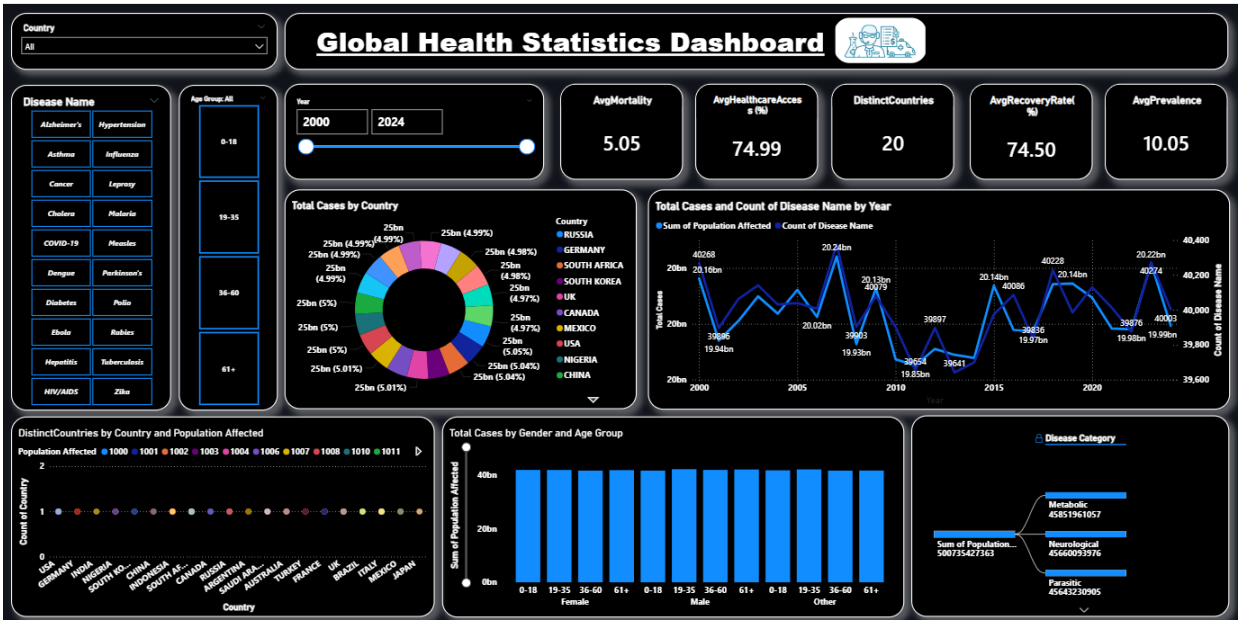
TABLE OF CONTENTS

1. Introduction

Healthcare data plays a crucial role in understanding disease patterns, improving healthcare delivery, and supporting policy decisions at both national and global levels. With the increasing availability of large-scale health datasets, analytical tools such as Power BI have become essential for transforming complex data into meaningful insights. The *Global Health Statistics Dashboard* project aims to analyse worldwide health trends using interactive data visualization techniques.

This project focuses on the analysis of global health data collected over a period from **2000 to 2024**, covering multiple countries, diseases, and demographic groups. The dashboard highlights key health indicators such as mortality rate, prevalence rate, recovery rate, healthcare access, and population affected. These indicators help in assessing the severity of diseases and the effectiveness of healthcare systems across different regions.

Microsoft Power BI is used in this project to perform data cleaning, exploratory data analysis, and visualization. The dashboard enables users to filter data dynamically using slicers and observe changes in KPIs in real time. This interactive approach enhances data understanding and supports comparative analysis across years, countries, and diseases.



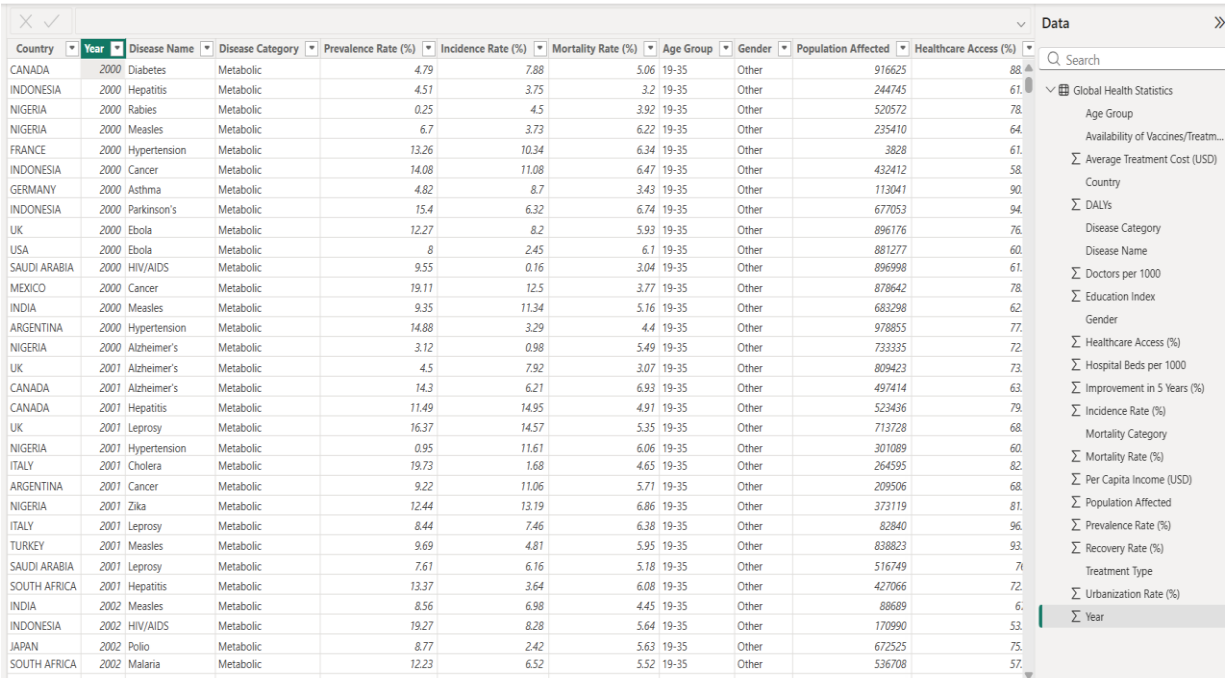
2. Source of Dataset

LINK - ***https://www.kaggle.com/datasets/malaiarasugraj/global-health-statistics***

The dataset used for this project was obtained from **Kaggle**, a widely used platform for open datasets and data science projects. The dataset is titled **"Global Health Statistics"** and contains extensive health-related data collected across different countries and years.

Parameter	Description
Source	Kaggle
Dataset Name	Global Health Statistics
Total Records	1,000,001 rows
Total Attributes	22 columns
Time Period	2000 – 2024
Countries Covered	20
Diseases Covered	20
Disease Categories	10

The dataset includes numerical indicators such as mortality rate, incidence rate, prevalence rate, recovery rate, healthcare access percentage, and population affected. It also contains categorical fields such as country, disease name, disease category, age group, gender, and year.



Country	Year	Disease Name	Disease Category	Prevalence Rate (%)	Incidence Rate (%)	Mortality Rate (%)	Age Group	Gender	Population Affected	Healthcare Access (%)
CANADA	2000	Diabetes	Metabolic	4.79	7.88	5.06	19-35	Other	916625	88.
INDONESIA	2000	Hepatitis	Metabolic	4.51	3.75	3.2	19-35	Other	244745	61.
NIGERIA	2000	Rabies	Metabolic	0.25	4.5	3.92	19-35	Other	520572	78.
NIGERIA	2000	Measles	Metabolic	6.7	3.73	6.22	19-35	Other	235410	64.
FRANCE	2000	Hypertension	Metabolic	13.26	10.34	6.34	19-35	Other	3828	61.
INDONESIA	2000	Cancer	Metabolic	14.08	11.08	6.47	19-35	Other	432412	58.
GERMANY	2000	Asthma	Metabolic	4.82	8.7	3.43	19-35	Other	113041	90.
INDONESIA	2000	Parkinson's	Metabolic	15.4	6.32	6.74	19-35	Other	677053	94.
UK	2000	Ebola	Metabolic	12.27	8.2	5.93	19-35	Other	896176	76.
USA	2000	Ebola	Metabolic	8	2.45	6.1	19-35	Other	881277	60.
SAUDI ARABIA	2000	HIV/AIDS	Metabolic	9.55	0.16	3.04	19-35	Other	896998	61.
MEXICO	2000	Cancer	Metabolic	19.11	12.5	3.77	19-35	Other	878642	78.
INDIA	2000	Measles	Metabolic	9.35	11.34	5.16	19-35	Other	683298	62.
ARGENTINA	2000	Hypertension	Metabolic	14.88	3.29	4.4	19-35	Other	978855	77.
NIGERIA	2000	Alzheimer's	Metabolic	3.12	0.98	5.49	19-35	Other	733335	72.
UK	2001	Alzheimer's	Metabolic	4.5	7.92	3.07	19-35	Other	809423	73.
CANADA	2001	Alzheimer's	Metabolic	14.3	6.21	6.93	19-35	Other	497414	63.
CANADA	2001	Hepatitis	Metabolic	11.49	14.95	4.91	19-35	Other	523436	79.
UK	2001	Leprosy	Metabolic	16.37	14.57	5.35	19-35	Other	713728	68.
NIGERIA	2001	Hypertension	Metabolic	0.95	11.61	6.06	19-35	Other	301089	60.
ITALY	2001	Cholera	Metabolic	19.73	1.68	4.65	19-35	Other	264595	82.
ARGENTINA	2001	Cancer	Metabolic	9.22	11.06	5.71	19-35	Other	209506	68.
NIGERIA	2001	Zika	Metabolic	12.44	13.19	6.86	19-35	Other	373119	81.
ITALY	2001	Leprosy	Metabolic	8.44	7.46	6.38	19-35	Other	82840	96.
TURKEY	2001	Measles	Metabolic	9.69	4.81	5.95	19-35	Other	838823	93.
SAUDI ARABIA	2001	Leprosy	Metabolic	7.61	6.16	5.18	19-35	Other	516749	7(
SOUTH AFRICA	2001	Hepatitis	Metabolic	13.37	3.64	6.08	19-35	Other	427066	72.
INDIA	2002	Measles	Metabolic	8.56	6.98	4.45	19-35	Other	88689	6.
INDONESIA	2002	HIV/AIDS	Metabolic	19.27	8.28	5.64	19-35	Other	170990	53.
JAPAN	2002	Polio	Metabolic	8.77	2.42	5.63	19-35	Other	672525	75.
SOUTH AFRICA	2002	Malaria	Metabolic	12.23	6.52	5.52	19-35	Other	536708	57.

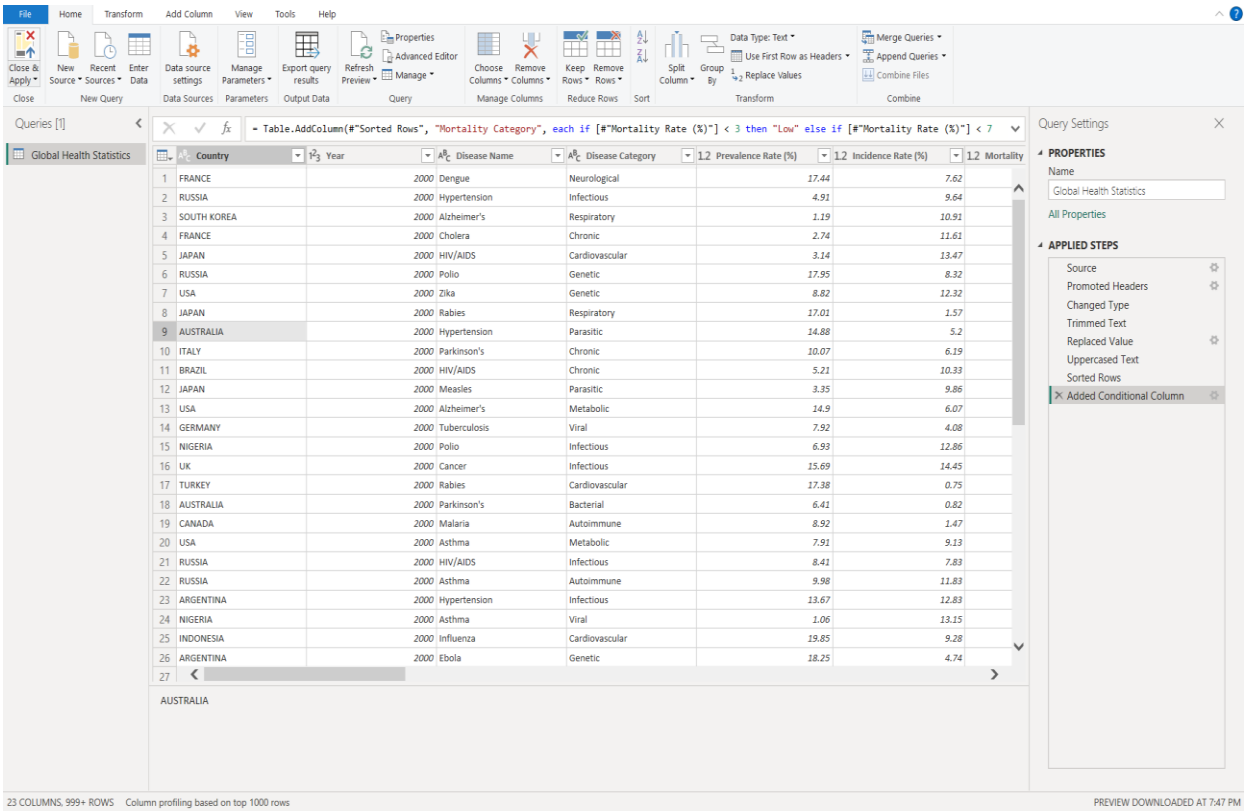
3. EDA Process (Exploratory Data Analysis)

Exploratory Data Analysis was carried out using Power Query Editor in Power BI to ensure data quality, consistency, and usability for analysis.

Data Cleaning Steps Performed

- Changed data types for numerical and percentage columns
- Trimmed text fields to remove extra spaces
- Replaced inconsistent values
- Converted country names to uppercase
- Sorted data by year
- Created a conditional column for mortality classification

Duplicate records were intentionally not removed, as each record represents a valid health observation and removing duplicates could lead to data loss.



4. Analysis on Dataset

4.1 Country-wise Disease Impact Analysis

i. Introduction

The objective of this analysis is to understand how the disease burden is distributed across different countries included in the dataset. This analysis helps in identifying countries that contribute more significantly to the overall population affected by diseases.

ii. General Description

This analysis is based on data collected from **20 countries** over a time period of **2000 to 2024**. The impact of diseases is measured using the *population affected* attribute, which represents the number of people influenced by various health conditions in each country.

iii. Specific Requirements, Functions and Formulas

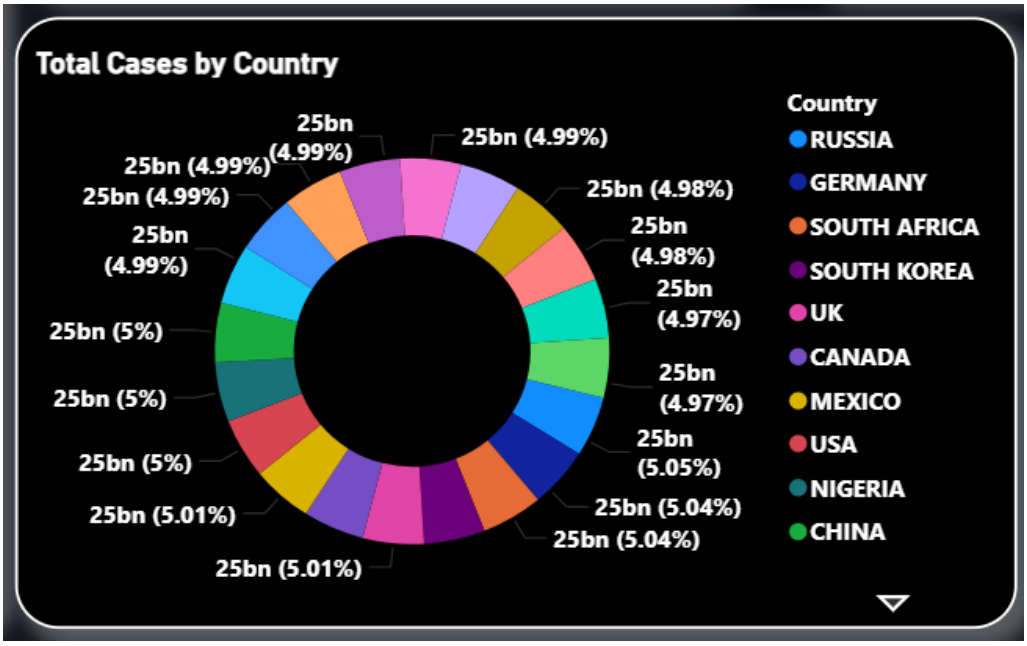
- Fields used: Country, Population Affected
- Aggregation method: Sum of Population Affected
- Filters applied: Country, Year, Disease Name, Age Group
- Nature of calculation: Dynamic aggregation based on slicer selection

iv. Analysis Results

- Total countries analysed: **20**
- The analysis shows that disease impact is not evenly distributed across countries.
- Certain countries contribute a larger proportion of the total population affected, while others contribute relatively less.
- When filters such as year or disease are applied, the country-wise contribution changes, indicating that disease impact is both time-dependent and disease-specific.

v. Visualization

This analysis is supported using a **Donut Chart** representing the proportional contribution of each country to the total population affected.



4.2 Year-wise Disease Trend Analysis

i. Introduction

The objective of this analysis is to examine how disease occurrence and population impact have changed over time. This helps in identifying long-term trends and variations in global health conditions.

ii. General Description

The dataset covers a **25-year period from 2000 to 2024**. Year-wise aggregation of disease data is used to study variations in total population affected and disease count over time.

iii. Specific Requirements, Functions and Formulas

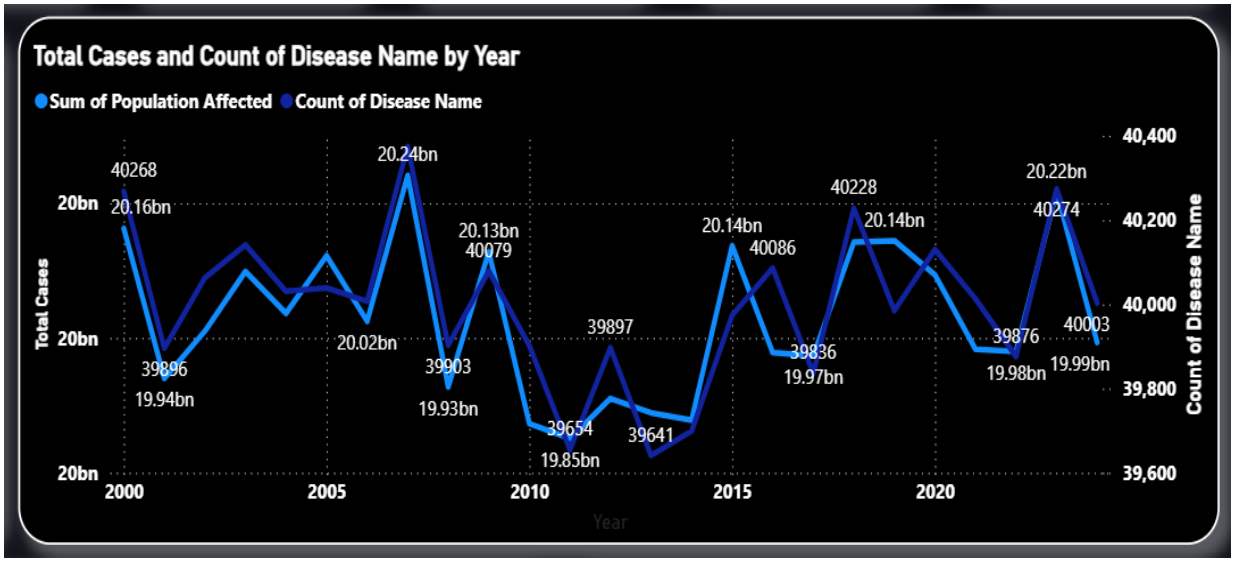
- Fields used: Year, Disease Name, Population Affected
- Measures used:
    - Total Population Affected
    - Disease Count
- Time-based filtering using year slicer

iv. Analysis Results (Numerical)

- Time period analysed: **25 years (2000–2024)**
- The results show **year-to-year fluctuations** in disease impact rather than a constant trend.
- Certain years record **higher population affected**, while others show comparatively lower values.
- These fluctuations indicate that global disease patterns **change over time** due to multiple factors such as healthcare access and disease outbreaks.

v. Visualization

A **Line Chart** is used to represent year-wise changes in disease count and population affected.



4.3 Country-wise Population Impact Comparison

i. Introduction

The objective of this analysis is to compare countries based on the population affected by diseases and identify variations in disease impact across different regions. This analysis helps in visually distinguishing countries with relatively higher and lower health burden.

ii. General Description

This analysis considers data from **20 countries** over the period **2000 to 2024**. Each country is represented as a point on the scatter plot, allowing comparison of disease impact based on the population affected metric.

iii. Specific Requirements, Functions and Formulas

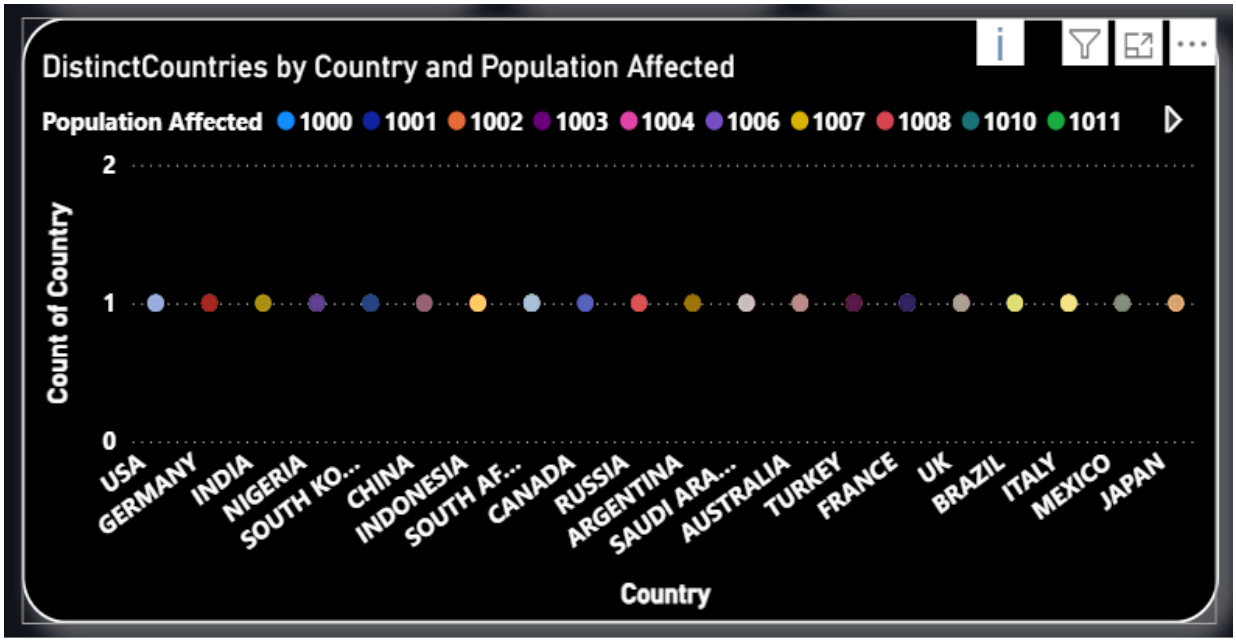
- Fields used: Country, Population Affected
- Aggregation method: Sum of Population Affected
- Filters applied: Country, Year, Disease Name
- Visualization logic: Each point represents a country's disease impact

iv. Analysis Results

- Total countries compared: **20**
- The scatter plot shows clear variation in population affected across countries.
- Some countries appear with higher population affected values, while others show comparatively lower impact.
- The spread of data points confirms that disease impact is not uniform across countries and varies significantly based on region and selected filters.

v. Visualization

A **Scatter Plot** is used to visually compare countries based on population affected and highlight differences in disease impact.



4.4 Demographic Impact Analysis (Age Group and Gender)

i. Introduction

The objective of this analysis is to study how disease impact varies across different demographic groups based on **age group and gender**. Understanding demographic patterns helps in identifying vulnerable population segments and supports targeted healthcare planning.

ii. General Description

This analysis categorizes the dataset based on multiple age groups and gender to examine differences in disease impact. The population affected metric is used to compare how diseases influence various demographic segments across the selected time period from **2000 to 2024**.

iii. Specific Requirements, Functions and Formulas

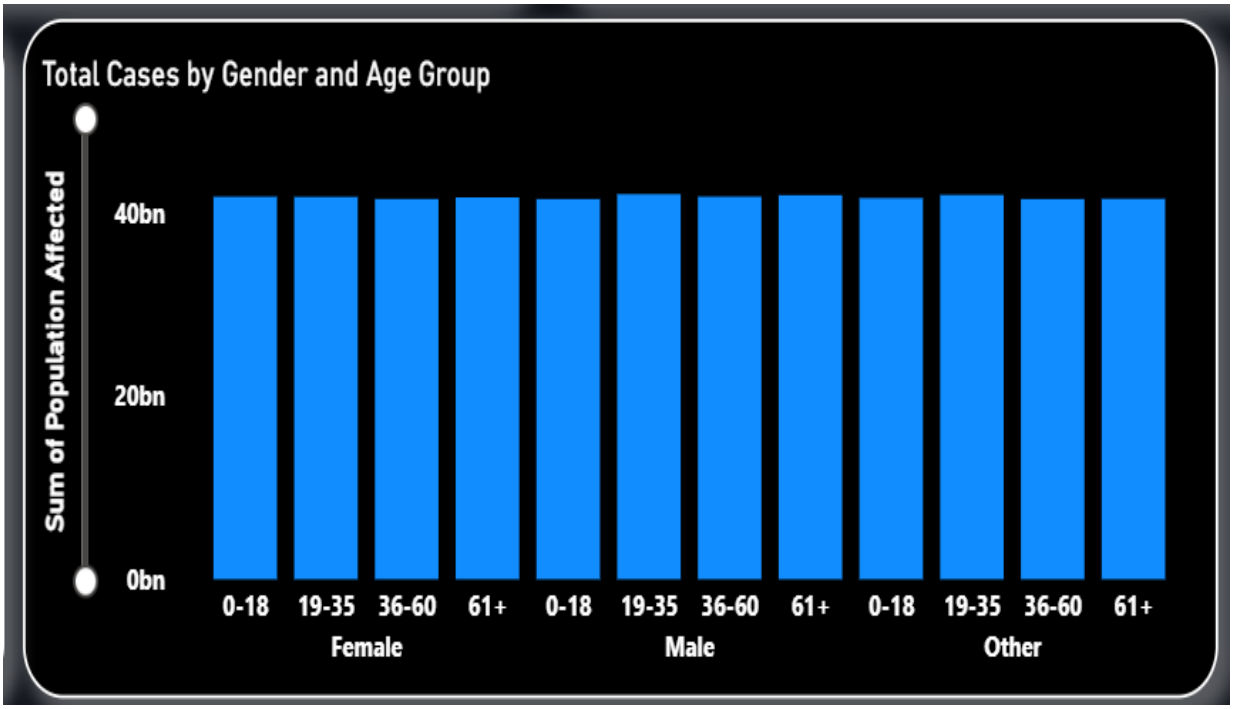
- Fields used: Age Group, Gender, Population Affected
- Aggregation method: Sum of Population Affected
- Filters applied: Age Group, Gender, Disease Name, Year
- Nature of calculation: Dynamic aggregation based on slicer selections

iv. Analysis Results (Numerical)

- Disease impact is distributed across all available age groups in the dataset.
- Certain age groups show a higher population affected, indicating increased vulnerability.
- Gender-wise comparison shows variation in disease impact between male and female populations.
- The results change dynamically when age group, gender, or disease filters are applied, confirming demographic sensitivity in disease impact.

v. Visualization

A **Bar Chart** is used to compare total population affected across different age groups and genders.



4.5 Disease Category Impact Analysis

i. Introduction

The objective of this analysis is to understand how disease impact is distributed across different **disease categories**. This analysis helps in identifying which categories contribute more significantly to the overall health burden.

ii. General Description

The dataset classifies diseases into **10 distinct disease categories**, such as infectious, chronic, metabolic, respiratory, and others. By aggregating population affected at the category level, this analysis provides a structured view of health impact across broader disease groups rather than individual diseases.

iii. Specific Requirements, Functions and Formulas

- Fields used: Disease Category, Population Affected
- Aggregation method: **Sum of Population Affected**

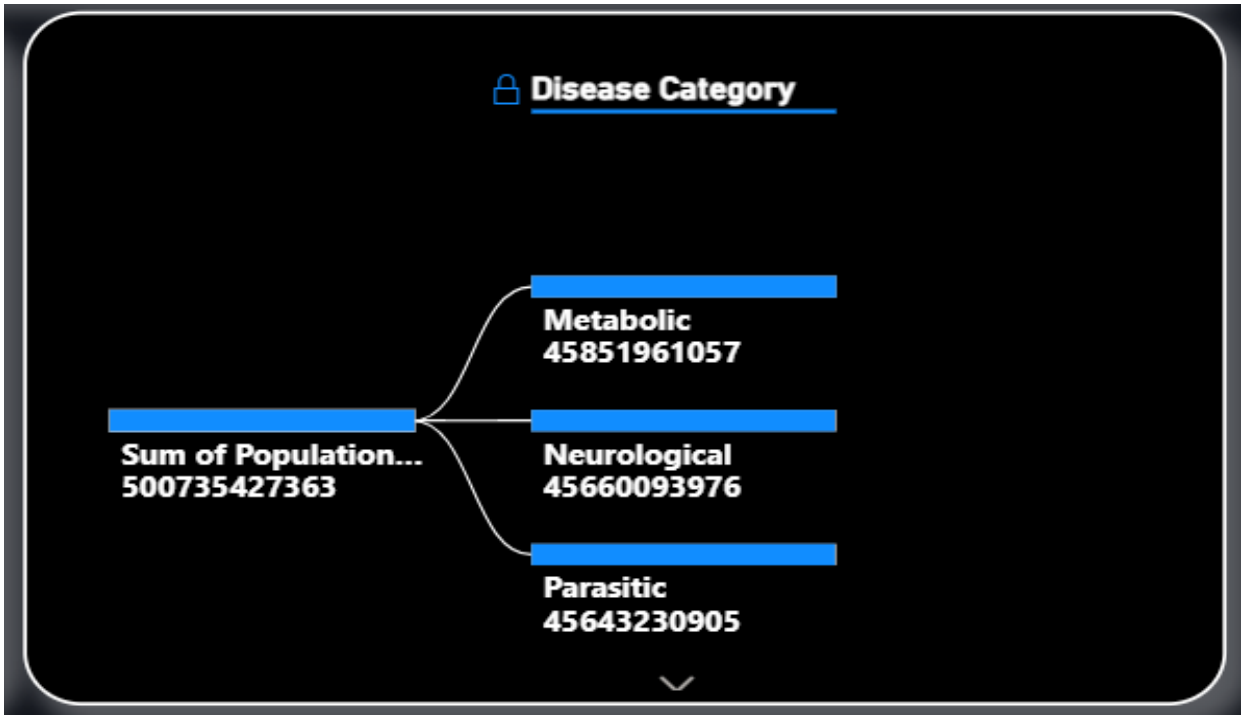
- Drill-down capability enabled to explore category-wise contribution
- Filters applied: Disease Category, Disease Name, Year

iv. Analysis Results

- Total disease categories analysed: **10**
- Disease impact is not uniformly distributed across all categories.
- Certain disease categories contribute a larger share of the total population affected, while others have relatively lower impact.
- The variation across categories highlights differences in disease severity and spread at a broader classification level.

v. Visualization

A **Decomposition Tree** is used to break down population affected by disease category and explore category-level contributions interactively.



4.6 KPI-based Global Health Indicator Analysis

i. Introduction

The objective of this analysis is to evaluate the overall global health condition using key numerical indicators derived from the dataset. KPI-based analysis provides a summarized view of health performance across all countries, diseases, and years included in the study.

ii. General Description

This analysis uses **KPI cards** to represent global average values calculated from health data covering **20 countries**, **20 diseases**, and a time span of **25 years (2000–2024)**. These indicators help in understanding the overall status of mortality, recovery, healthcare access, and disease prevalence.

iii. Specific Requirements, Functions and Formulas

The following **DAX measures** are used to compute KPI values dynamically:

- Average Mortality Rate
- Average Healthcare Access (%)
- Average Recovery Rate (%)
- Average Prevalence Rate
- Distinct Country Count

All KPI values update automatically based on slicer selections such as country, disease, age group, and year.

iv. Analysis Results

The KPI analysis produces the following global health indicators:

- Average Mortality Rate: 5.04%
- Average Healthcare Access: 74.90%
- Average Recovery Rate: 74.58%
- Average Prevalence Rate: 10.00%

- **Total Countries Analysed: 20**

These results indicate that while the average mortality rate remains moderate, healthcare access and recovery rates are relatively high, suggesting the presence of supportive healthcare infrastructure across the analysed regions.

v. Visualization

This analysis is represented using **KPI Cards**, which provide a quick and dynamic summary of key health indicators.

4.7 Slicer-based Interactive Analysis

i. Introduction

The objective of this analysis is to demonstrate how slicers enable interactive exploration of the health dataset. Slicer-based analysis allows users to dynamically filter data and observe changes in KPIs and visualizations in real time.

ii. General Description

The dashboard includes multiple slicers that allow filtering of health data across different dimensions. These slicers help users focus on specific countries, diseases, age groups, and time periods, improving analytical flexibility and clarity.

iii. Specific Requirements, Functions and Formulas

The following slicers are implemented in the dashboard:

- Country slicer (20 countries)
- Disease Name slicer (20 diseases)
- Age Group slicer
- Year slicer (2000–2024)

All slicers interact with:

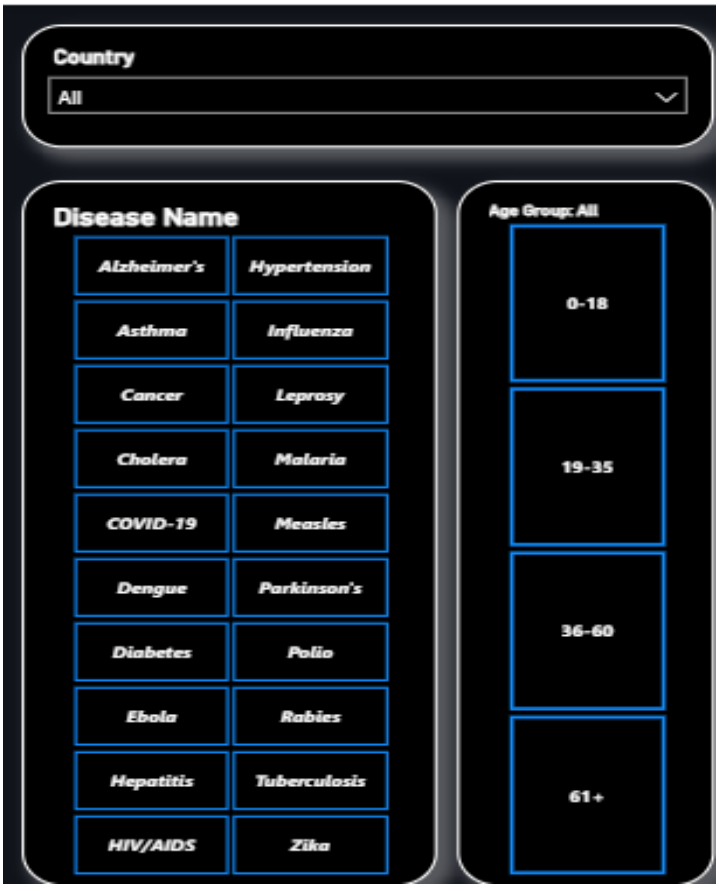
- KPI measures (Average Mortality, Healthcare Access, Recovery Rate, Prevalence)
- All visualizations in the dashboard (Donut, Line, Bar, Scatter, Decomposition Tree)

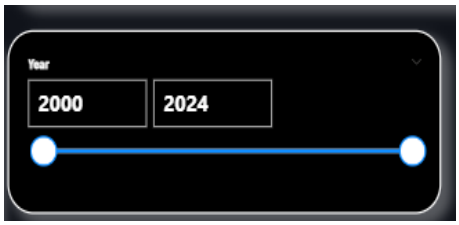
iv. Analysis Results

- KPI values such as:
    - Average Mortality Rate: 5.04%
    - Average Healthcare Access: 74.90%
    - Average Recovery Rate: 74.58%
    - Average Prevalence Rate: 10.00%
      update dynamically when slicer selections are changed.
- Filtering by year allows analysis within a specific subset of the 25-year period.
- Filtering by country or disease enables focused regional or disease-specific insights.

v. Visualization

This analysis is supported using interactive slicers connected to all KPI cards and charts.





5. Conclusion

The *Global Health Statistics Dashboard* shows how large health data can be analysed and understood easily using Power BI. In this project, a dataset with **1,000,001 records**, **22 columns**, and data from **2000 to 2024** was used to study health patterns across **20 countries** and **20 diseases**.

The analysis helped to understand that disease impact is not the same everywhere. Some countries, years, and population groups are more affected than others. The KPI results show that the **average mortality rate is 5.04%**, while **healthcare access (74.90%)** and **recovery rate (74.58%)** are relatively high. This means that although diseases affect a large population, healthcare systems are helping many people recover.

The dashboard also includes interactive slicers, which allow users to filter data by country, disease, age group, and year. When these filters are applied, all charts and KPIs change automatically, making the analysis more flexible and easy to explore.

Overall, this project clearly demonstrates how Power BI can be used to convert complex health data into meaningful insights. The dashboard makes it easier to understand global health trends and supports better analysis and decision-making using data.

6. Future Scope

The Global Health Statistics Dashboard can be further enhanced to provide deeper and more meaningful health insights. By expanding data coverage and adding advanced analytical features, the dashboard can support better understanding of global health trends and improve decision-making.

- Additional countries and diseases can be included for broader analysis.
- Real-time or frequently updated health data can be integrated.
- Predictive analytics can be added to forecast future disease trends.
- Geographical maps can be used for improved regional visualization.
- Separate dashboards can be developed for specific diseases or regions.

7. References

1. Kaggle. *Global Health Statistics Dataset*. *https://www.kaggle.com/datasets/malaiarasugraj/global-health-statistics*
2. Microsoft. *Power BI Documentation*. *https://learn.microsoft.com/power-bi/*

8. LinkedIn -

https://www.linkedin.com/posts/rajat-rajan_powerbi-datavisualization