Bharatiya Vidya Bhavan's
SARDAR PATEL INSTITUTE OF TECHNOLOGY

Advanced Data Visualization
Experiment no. 2

Submitted To

Prof. Pranav Nerurkar

Submitted By

Name: Aakriti Pathak

UID: 2021300094

Batch: BE Comps (Batch B)

1. Aim:

Create advanced charts using Tableau / Power BI / R / Python / Plotly or Chart or D3.js to be performed on the dataset - Socio economic data

- Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, Jitter, Line, Area, Waterfall, Donut, Treemap, Funnel
- Write observations from each chart

2. Procedure Description:

Step-1: Dataset:

You can view the dataset from this link.

Step-2: Description:

The dataset includes data on life expectancy and various socio-economic indicators across different countries from 2000 to 2019. The dataset provides valuable insights into how factors such as health expenditure, education, CO2 emissions, and economic status influence life expectancy across different regions and income groups.

Step-3: MetaData:

- **Country:** Name of the country.
- **Region:** Region or continent the country belongs to.
- **Income Group:** Classification of countries based on income.
- **Year:** Year of the data point (2000-2019).
- **Life Expectancy:** Average number of years a newborn is expected to live.
- **Health Expenditure (% of GDP):** Proportion of GDP spent on healthcare.
- **Education Expenditure (% of GDP):** Proportion of GDP spent on education.
- **CO2 Emissions (kilotons):** Emissions from fossil fuels and cement manufacture.
- **Prevalence of Undernourishment (%):** Percentage of the population undernourished.
- **Unemployment (%):** Unemployment rate of the total labor force.
- **Sanitation (%):** Population using safely managed sanitation services.

Step-4: Data Modeling - Star Schema:

1. Overview of the Star Schema

The star schema is a data modeling approach commonly used in business intelligence to organize data efficiently for querying and reporting. In this model, data is divided into a central fact table and surrounding dimension tables.

- **Fact Table:** Contains quantitative data such as life expectancy, health expenditure, and CO2 emissions.
- **Dimension Tables:** Includes descriptive attributes like country, region, income group, and year.

2. Data Preprocessing

During the data modeling process, duplicates, particularly within dimension tables, were identified and removed to ensure accurate analysis and establish relationships between tables.

3. Explanation of Tables

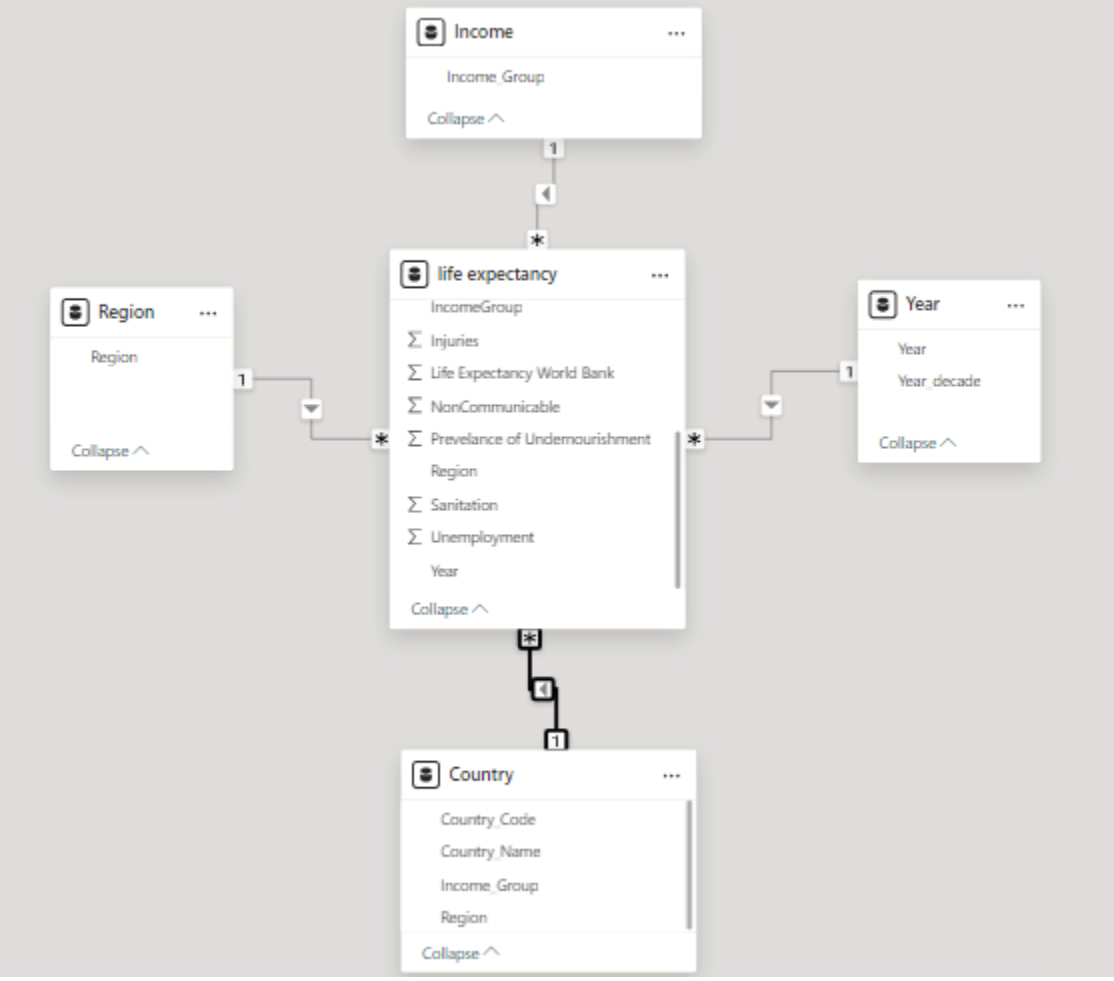
- **Fact Table:** Central table with metrics like life expectancy, health expenditure, education expenditure, CO2 emissions, etc.
- **Dimension Tables:** Include tables for Country, Region, Income Group, and Year.

4. Relationships

- **Country Dimension to Fact Table on Country Code:**
  - **Cardinality:** One-to-Many.
  - **Cross-Filter Direction:** Single.
- **Region Dimension to Fact Table on Region Code:**
  - **Cardinality:** One-to-Many.
  - **Cross-Filter Direction:** Single.

5. Star Schema Diagram

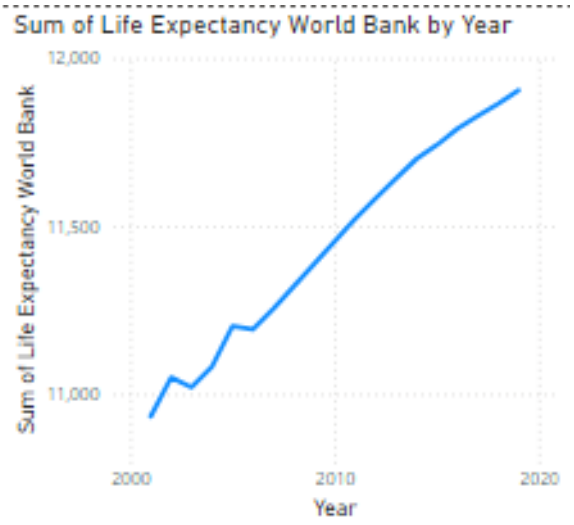
Analysis of Relationships: The relationships between the tables ensure that filtering or aggregating data in any of the dimension tables correctly reflects in the fact table, allowing for accurate and dynamic reporting.



Step-5: Data Visualization Analysis:

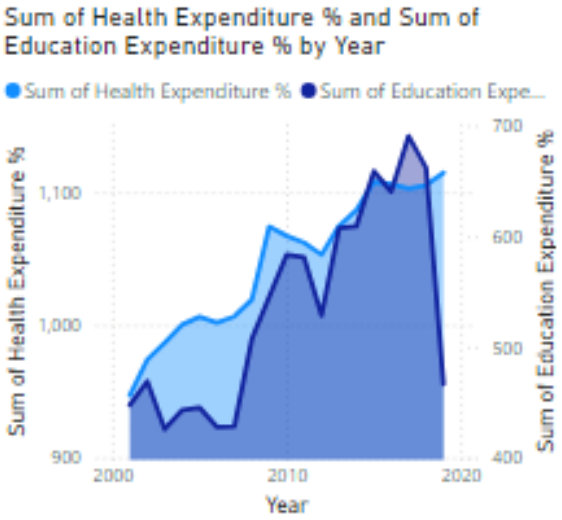
1. Line Chart: Sum of Life Expectancy by Year

- **Description:** This line chart displays the trend of life expectancy over time. The X-axis represents the years, and the Y-axis represents the sum of life expectancy.
- **Observation:** The chart shows a steady increase in life expectancy over the years, indicating improvements in healthcare, nutrition, and living conditions globally.



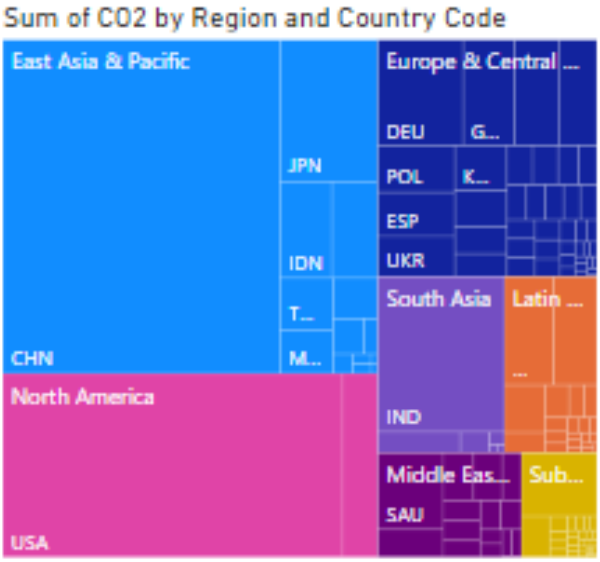
2. Area Chart: Sum of Health Expenditure % and Education Expenditure % by Year

- **Description:** This area chart plots the sum of health and education expenditures as a percentage of GDP over time, with two layers representing each metric.
- **Observation:** The chart indicates that both health and education expenditures have been increasing over time. This trend suggests a growing recognition of the importance of these sectors for improving overall life expectancy and quality of life.



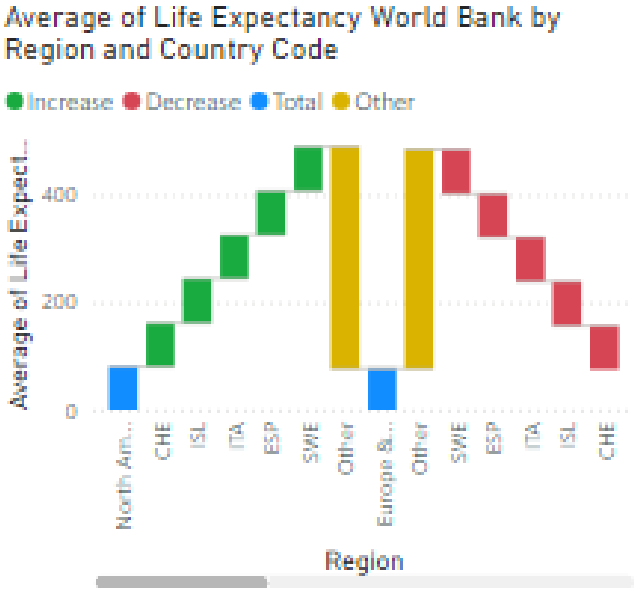
3. Treemap: Sum of CO2 Emissions by Region and Country Code

- **Description:** The treemap visualizes CO2 emissions by different regions and countries, where each rectangle's size is proportional to the emissions.
- **Observation:** The chart highlights that specific regions, such as East Asia & Pacific (e.g., China) and North America (e.g., USA), are significant contributors to global CO2 emissions. This suggests that industrialized and populous regions contribute most to environmental pollution.



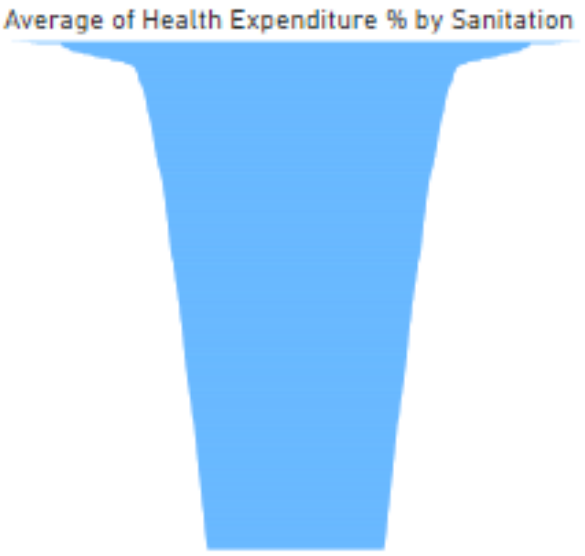
4. Waterfall Chart: Average of Life Expectancy by Region and Country Code

- **Description:** The waterfall chart shows the change in average life expectancy across different regions, with the bars indicating increases or decreases.
- **Observation:** This chart highlights the regional disparities in life expectancy. Regions such as North America and Europe generally show higher life expectancies, while other regions, like Sub-Saharan Africa, show lower values. It emphasizes the need for targeted health interventions in lower life expectancy regions.



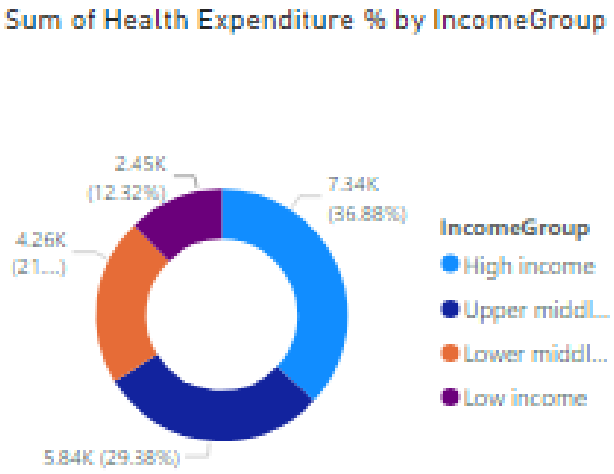
5. Funnel Chart: Average of Health Expenditure % by Sanitation

- **Description:** This funnel chart shows the relationship between sanitation levels and health expenditure as a percentage of GDP.
- **Observation:** The funnel chart reveals that higher health expenditure is associated with better sanitation services. This suggests that investments in sanitation are a critical part of public health strategies.



6. Donut Chart: Sum of Health Expenditure % by Income Group

- **Description:** This donut chart breaks down health expenditure by different income groups, with each segment representing the expenditure share of a particular income group.
- **Observation:** The chart reveals that higher-income groups allocate a larger percentage of their GDP to health expenditures. This implies that wealthier nations are able to invest more in health services, which could contribute to better health outcomes, including higher life expectancy.



Step-6: Summary of Key Insights:

- **Life Expectancy Trends:** There is a clear upward trend in global life expectancy, likely due to increased health and education expenditures.
- **CO2 Emissions:** Major industrial and populous regions are significant contributors to global CO2 emissions, raising concerns about sustainability.
- **Health and Sanitation:** Better sanitation correlates with higher health expenditures, emphasizing its role in public health.
- **Regional Disparities:** Life expectancy varies significantly by region, with wealthier regions showing better outcomes.
- **Income Group Influence:** Health expenditure is heavily influenced by a country's income group, affecting overall life expectancy and public health outcomes.