

# TB Data Visualization Dashboard

## 1. Introduction

This report provides an overview of a data visualization project focused on visualizing global TB (Tuberculosis) data. The project includes a variety of visualizations that help explore complex relationships, hierarchical data, geographic distributions, and temporal changes. The visualizations include:

- **Force-Directed Graph**
- **Map Chart**
- **Timeline Visualization**
- **Treemap**
- **Sunburst Chart**

Additionally, an interactive **Dashboard** allows users to easily navigate through these visualizations.

### Dataset Link:

<https://data.amerigeoss.org/dataset/indicator-3-3-2-tuberculosis-incidence-per-100000population/resource/c846bcb2-20a8-49d3-955a-046b52344c88>

### Preprocessing Steps:

- 1 **Data Cleaning:**
  - **Removing Columns:** The code starts by removing irrelevant columns from the dataset (like 'goal', 'target', 'indicator', etc.) which are not needed for analysis.
- 2 **Handling Missing Data:**
  - **Null Values Calculation:** It calculates the number of null values in each column and computes the percentage of missing values.
  - **Filling Null Values:** Columns with more than 10% missing values are processed:
    - For numeric columns, missing values are replaced with the **mean** of that column.
    - For non-numeric columns, missing values are filled with the **mode** (most frequent value) of the column.
- 3 **Row-wise Null Handling:**
  - **Fill with Row Mean:** It also handles missing values at the row level, where null values in specific columns (like 'value\_2000' to 'value\_2017') are filled with the row's mean for those columns.
- 4 **Warnings:** The script suppresses future warnings from pandas to avoid clutter in the output.

### Key Functionality:

- **Dropping Columns:** Unnecessary columns are removed early in the script.
- **Null Value Percentage:** It helps identify columns with a high percentage of missing data.
- **Null Imputation:** Missing data is imputed using the column's mean or mode depending on the column type.
- **Row-wise Null Handling:** For numeric columns, missing values are filled with the row's mean, preserving the row's context.

### Summary of Actions:

- The code removes irrelevant columns and fills missing data using either the column mean or mode.
- It ensures that there are no missing values left in the specific columns related to TB cases from 2000 to 2017.

## 2. GRAPHS

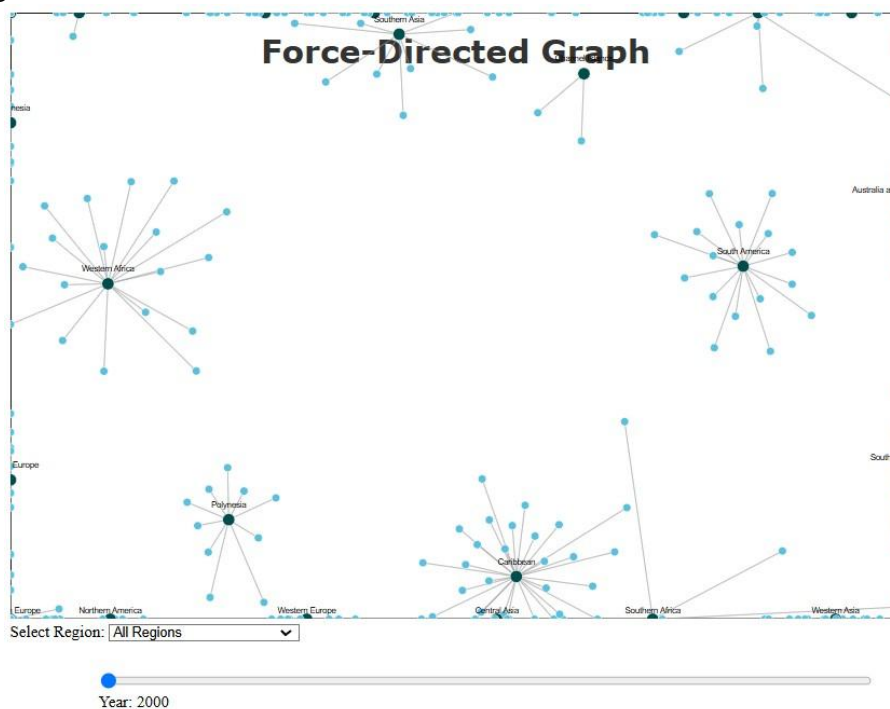
### 2.1 Force-Directed Graph for Relationship Mapping

**Overview:** The Force-Directed Graph visualizes regions and their countries in a networkstyle layout. In this graph:

- The **parent nodes** represent **regions**.
- The **child nodes** represent **countries** within those regions.
- **Edges** represent the relationships between regions and countries.
- The **edge size** is dynamically determined by the **number of cases** reported in each country.
- Users can **select a year** from a filter to view data for that specific year, with the edge sizes adjusting accordingly.
- A **dropdown** allows users to select either **individual regions** or **all regions** for focused exploration.

#### Key Features:

- **Node Representation:** Parent nodes are regions, and child nodes are countries.
- **Edge Representation:** Relationships between regions and their countries are shown as edges, with the edge size reflecting the number of cases reported in the selected year.
- **Color Coding:** Different relationship types or regions are color-coded for easy identification.
- **Interactivity:**
  - Hover over nodes to see detailed information about the region or country. ◦ Click to expand nodes for more insights.
  - Zooming and panning allow users to explore the entire network.
- **Year Filter:** A dynamic filter lets users select a year, updating the edge sizes to reflect the cases for that year.
- **Region Filter:** A dropdown allows users to filter the graph to display specific regions or show all regions.



### 2.2 Map Chart for Geographic Representation

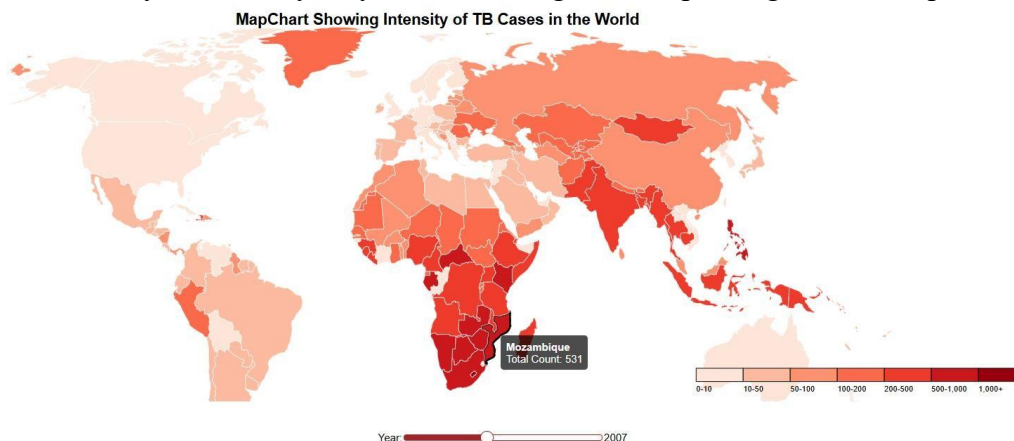
**Overview:** The Map Chart provides a geographical visualization of TB case data, displaying how TB cases are distributed across various countries and regions. The chart includes both a standard map and a heat map,

where the color intensity reflects the number of TB cases reported, offering an intuitive and engaging way to explore the global distribution of TB.

The **color intensity** on the map increases as the TB case numbers rise, helping users quickly identify regions with higher case counts. The visualization includes interactive elements such as zooming, panning, and tooltips, providing an immersive and insightful experience.

### Key Features:

- **Map and Heat Map Display:**
  - The chart shows both a standard geographical map and a **heat map**, where countries are color-coded based on their TB case intensity.
  - The **color intensity** increases as the number of reported TB cases rises, making it easy to spot countries with the highest TB burden.
- **Zoom and Pan Functionality:**
  - Users can zoom in and out to focus on specific regions or zoom out to view the global distribution of TB cases.
  - The **panning feature** allows for smooth navigation across different parts of the world, making it easier to explore specific areas of interest.
- **Dynamic Year Filter:**
  - A **slider** is included to filter the data by year. As users adjust the slider, the map updates to reflect the TB case data for the selected year.
  - The **heat map** dynamically adjusts the color intensity based on the TB cases for the selected year, providing a temporal dimension to the geographical data.
- **Interactive Legend:**
  - The **interactive legend** allows users to highlight countries within specific ranges of TB case numbers.
  - Users can click on different legend ranges to highlight countries that fall within those ranges, offering a more focused view of TB case distribution based on intensity levels.
- **Hover Highlighting:**
  - When users hover over a country, a tooltip appears with detailed information about that country's TB case data for the selected year.
  - The **highlighting effect** helps users quickly identify the country they are interacting with, improving the user experience.



## 2.3 Timeline Visualization

**Overview:** The Timeline Visualization allows users to explore the evolution of TB cases continent-wise over time. The timeline displays data from the year 2000 to 2017, providing insights into how TB case numbers have changed each year. The interactive nature of the visualization lets users select specific continents or view all continents to track the progression of cases year by year.

The **line chart** represents the total TB cases per continent, and users can interact with the chart to highlight specific years and continents. The animation feature also lets users play through the timeline to visualize changes over time, and a dropdown menu allows for dynamic filtering by continent.

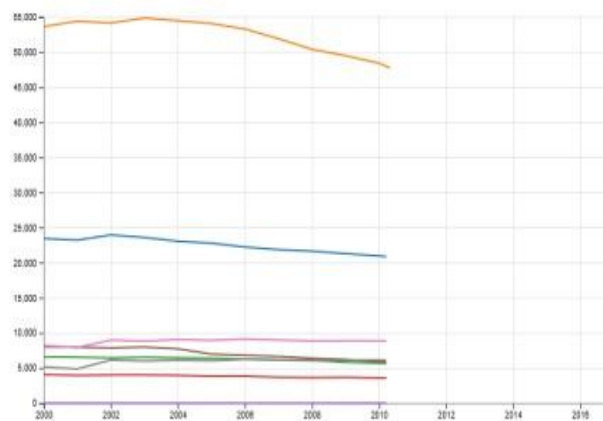
### Key Features:

- **Continent-wise Line Chart:** Each continent has its own line on the chart that displays TB case totals for each year from 2000 to 2017.
- **Dynamic Year Filtering:** A slider allows users to adjust the displayed year and see how TB cases have changed over time.
- **Play/Pause Animation:** Users can animate the timeline to observe the changes in TB cases from year to year automatically.
- **Interactive Tooltip:** Hovering over the lines shows detailed information about the selected continent, year, and number of cases.
- **Continent Dropdown:** A dropdown allows users to select individual continents (or show all continents) to filter the timeline. This enables users to focus on specific geographical regions or compare them.
- **Legend:** A legend displays the continent color scheme for easy identification of each line.

### Interactivity:

- **Hover Interaction:** As users hover over a specific part of the timeline, a tooltip shows the continent, year, and number of TB cases.
- **Play Animation:** The play button animates the timeline, showing how TB cases evolve year by year.
- **Pause Functionality:** Users can pause the animation and explore specific years manually.

Continent-wise Timeline Visualization



Pause

All Continents ▼

## 2.4 Hierarchical Tree Map for Entity Categorization

**Overview:** The Treemap Visualization displays hierarchical data using nested rectangles, representing continents, regions, and countries. Each rectangle's size corresponds to the total TB cases within that region or country, allowing users to visually compare the magnitude of TB cases across different geographical areas. The chart is interactive, providing detailed tooltips, and allowing users to drill down into specific continents, regions, or countries.

The **hierarchical structure** is clearly depicted, with continents as the top-level nodes, followed by regions and countries. The **color intensity** of each rectangle corresponds to the TB case volume, making it easy to identify regions with high or low TB burdens.

### Key Features:

- **Hierarchical Structure:**
  - The data is organized hierarchically with **continents** at the first level, **regions** at the second level, and **countries** at the third level.
  - Each level is represented by nested rectangles, with the size of each rectangle proportional to the total TB cases in that area.
- **Dynamic Data Aggregation:**
  - The data is processed to sum the TB cases for each country, region, and continent. For regions and countries, the total TB cases are computed over all available years (from 2000 to 2017).
- **Interactive Tooltips:**
  - Hovering over any rectangle reveals a tooltip with detailed information about the name of the region or country, and the number of TB cases.
- **Click-to-Drill-Down Functionality:**
  - Clicking on a **continent** will load and display the relevant regions.
  - Clicking on a **region** will display the individual countries within that region, allowing users to explore data at a deeper level.
  - Clicking on a **country** (leaf node) does not trigger further actions but provides detailed insights in the tooltip.
- **Color Coding:**
  - Each continent has its own **base color**, and each subsequent region and country is shaded based on that continent's color. The color intensity increases or decreases based on the size of the TB cases in that region or country.
  - **Color scales** for regions are adjusted to reflect the TB case intensity, helping users quickly identify areas with higher or lower case counts.
- **Legend:**
  - A **legend** is included on the side, displaying the color coding for each continent. This helps users associate the colors in the treemap with the correct geographical area.
- **Zooming and Panning:**
  - The treemap is **zoomable** and **pannable**, allowing users to focus on specific areas of the map and explore the data more interactively.
- **Responsive Design:**
  - The chart is designed to be responsive, adjusting the layout as per the screen size, ensuring a smooth experience across devices.

### Interactivity:

- **Hover Interaction:** Displays a tooltip with detailed information about the country, region, or continent's TB case data.
- **Click Interaction:** Clicking on regions or countries allows users to drill down into more detailed levels of the hierarchy.
- **Legend:** The legend helps users interpret the color coding for each continent.

### User Controls:

- **Reset Button:** A button labeled "**All**" allows users to reset the treemap and view data for all continents and regions.
- **Data Filters:** Users can filter the treemap to display specific regions or continents based on the dropdown options available.



All

## 2.5 Sunburst Chart for Hierarchical Data

**Overview:** The **Sunburst Chart** provides a circular, hierarchical representation of global TB data, with each level of the hierarchy represented by concentric rings. The chart is divided into segments, where each segment represents a geographical level—**continents**, **regions**, and **countries**. The size of each segment corresponds to the total TB cases in that area, allowing users to visually compare the relative burden of TB cases across different countries, regions, and continents.

The chart offers interactive features that allow users to **drill down** into specific continents or regions, making it possible to explore global TB data at varying levels of detail.

### Key Features:

- **Hierarchical Structure:**
  - The **root node** represents "World," with concentric rings representing **continents** (first ring), **regions** (second ring), and **countries** (third ring).
  - The size of each segment is proportional to the number of TB cases in that region or country.
- **Color Coding:**
  - Each **continent** is assigned a unique base color.
  - **Regions** and **countries** are color-coded based on their parent continent, with varying shades to represent different levels of TB case intensities within that continent.
- **Interactive Drill-Down:**
  - Users can **click** on a **continent** to reveal its regions, or **click** on a **region** to reveal the countries within it.
  - This interactive feature allows users to explore the data in a hierarchical manner, providing insights into global TB distribution from a macro (continent) to a micro (country) level.
- **Dynamic Tooltip:**
  - **Hovering** over a segment displays a tooltip showing the name of the region or country and the total TB cases in that area.
  - The tooltip helps users understand the exact value for each segment, making the chart highly informative.
- **"World" Center:**
  - The center of the chart is labeled "World" and serves as the root node. Clicking on this central node resets the visualization to show global data for all continents.
- **Legend:**



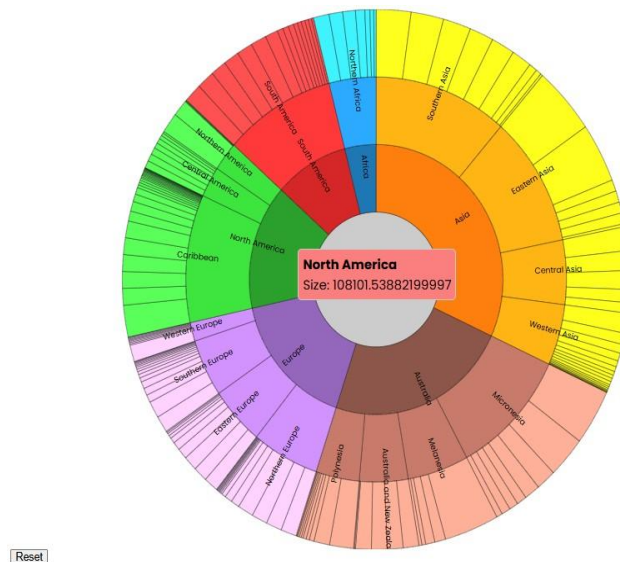
- o The **legend** displays the color coding for each continent, helping users quickly identify which colors correspond to which continents.
- **Responsive Design:**
  - o The chart is designed to be **responsive**, adjusting to different screen sizes and ensuring a seamless experience across devices.

#### Interactivity:

- **Hover Effects:** Display detailed information about the region or country in the tooltip on hover.
- **Click-to-Drill:** Clicking on a continent or region allows users to drill down and view more granular data for countries within that area.
- **Reset Button:** The "Reset" button allows users to return to the global view (World) by reloading the chart.

#### User Controls:

- **Reset Functionality:** The chart includes a button that resets the entire visualization to show data for the whole world.
- **Data Filters:** Users can filter the chart to view specific continents or regions, depending on the selection.



### 3. TB Dashboard

**Overview:** The **TB Dashboard** serves as an interactive hub where users can explore multiple visualizations related to TB data. These visualizations include:

- **Force-Directed Graph**
- **Line Chart**
- **Map Chart**
- **Treemap**
- **Sunburst**

The dashboard enables users to navigate between these visualizations through a simple and intuitive interface. Users can filter and interact with the data displayed in each visualization to gain deeper insights into the global TB burden.

## Key Features:

- **Centralized Interface:** All visualizations are embedded within the dashboard, making it easy for users to access them in one place without needing to navigate through multiple pages.
- **Responsive Layout:** The dashboard is designed to be responsive, ensuring a seamless experience across different screen sizes, including mobile devices.
  - **Two-Column Layout:** The **Force-Directed Graph** and **Sunburst** visualizations are displayed side by side.
  - **Full-Width Layout:** The **Line Chart**, **Map Chart**, and **Treemap** visualizations span the full width of the dashboard.
- **Interactive Navigation:**
  - **Navigation Buttons:** Users can toggle between different visualizations (ForceDirected Graph, Line Chart, Map Chart, TreeMap, and Sunburst) using buttons at the top of the page.
  - **Dynamic Iframe Display:** When a button is clicked, the relevant visualization is displayed in the iframe, while the others are hidden. This allows users to focus on one visualization at a time without clutter.
- **“All” Button:** A button labeled “All” resets the dashboard to show all visualizations again, giving users a comprehensive view of the entire data set.
- **Customization and Filtering:**
  - The visualizations, especially the Force-Directed Graph and Timeline, allow users to apply filters (such as selecting a region or year) to tailor the data shown according to their interests.
  - Users can interact with each visualization to zoom, pan, hover, and click to get more detailed data.
- **Embedding Visualizations:** Each visualization is embedded in an **iframe**, making it easy to load and interact with the individual charts without leaving the dashboard page.

## User Controls:

- **Toggle Between Visualizations:** The navigation buttons allow users to seamlessly switch between visualizations, ensuring a fluid exploration of the data.
- **Interactive Data Exploration:** Each visualization is interactive, with users able to zoom, pan, and hover to get more detailed information about the TB cases across different regions and years.