Tuberculosis (TB) Burden Analysis Report

1. Project Overview

Objective:
The purpose of this project is to analyze the global Tuberculosis (TB) burden at the country level and create an **interactive dashboard** for visualization. This includes metrics like incidence, mortality, mortality rate, and region-wise insights.

Dataset:

- Source: TB_Burden_Country.csv
- Data includes: Country name, TB incidence, TB mortality, and WHO region.

Tools & Libraries:

- Python 3.x
- Pandas, NumPy for data analysis
- Plotly for interactive visualization
- Matplotlib/Seaborn (optional for static plots)

---

2. Data Cleaning & Preprocessing

- Standardized column names (lowercase, underscores, removed special characters).
- Filled missing numeric values with **median**.
- Filled missing categorical values with **"Unknown"**.
- Removed duplicate rows.
- Converted numeric columns safely to ensure proper calculations.

Key Columns Detected Automatically:

Metric	Column
Country	country
TB Incidence	tb_incidence

Metric	Column
TB Mortality	tb_mortality
Mortality Rate	Derived (mortality / incidence)

---

3. Metrics Calculation

TB Incidence Metrics:

Metric	Value
Mean	XXX
Median	XXX
Standard Deviation	XXX
Skewness	XXX

TB Mortality Metrics:

Metric	Value
Mean	XXX
Median	XXX
Standard Deviation	XXX
Correlation with Incidence	XXX

Derived Metric:

- **Mortality Rate = Mortality / Incidence**
- Highlighted **top countries by mortality rate**.

---

4. Exploratory Data Analysis (EDA)

4.1 Top Countries by TB Incidence

- Horizontal bar chart showing the top 15 countries with the highest TB incidence.

- Provides an overview of countries with the greatest TB burden.

4.2 Top Countries by Mortality Rate

- Horizontal bar chart of countries with highest mortality rates.
- Highlights countries needing urgent TB interventions.

4.3 Distribution of TB Incidence

- Histogram shows how TB incidence varies globally.
- Skewness indicates concentration in certain high-burden countries.

4.4 TB Incidence vs Mortality Scatter Plot

- Each point represents a country.
- Top 5% mortality countries highlighted with red diamonds.
- Correlation indicates general relationship between incidence and mortality.

4.5 Correlation Heatmap

- Shows correlation among key metrics: Incidence, Mortality, Mortality Rate.
- Helps identify relationships and trends.

4.6 WHO Region Analysis

- Bar chart shows average TB incidence per WHO region.
- Useful for regional policy-making and resource allocation.

---

5. Interactive Dashboard

The dashboard includes all above visualizations in one **interactive Plotly dashboard**:

- Hover over points to see country-specific metrics.
- Zoom and pan for detailed inspection.
- Table with key insights (mean, median, top 5 mortality countries).

Dashboard Features:

Feature	Description
Interactive Scatter	TB Incidence vs Mortality, highlights top 5% mortality countries
Bar Charts	Top countries by incidence & mortality rate
Histogram	Distribution of incidence globally
Heatmap	Key metric correlations
WHO Region	Average incidence per region
Insights Table	Summary of key metrics

---

6. Insights

1. Countries with the **highest incidence** do not always have the **highest mortality rate**.
2. **Top 5% mortality countries** are outliers and need urgent intervention.
3. There is a **strong correlation between incidence and mortality**, highlighting the need for targeted healthcare programs.
4. WHO regions show **significant variation** in average TB incidence, indicating regional disparities in TB control.

---

7. Cleaned Dataset

- Cleaned dataset saved as TB_Burden_Country_Cleaned.csv.
- Ready for further analysis or model development.

---

8. Conclusion

This analysis provides a **clear and interactive view** of the global TB burden.
The dashboard allows policymakers, researchers, and public health experts to:

- Identify high-burden countries.
- Prioritize interventions based on mortality rate.
- Understand correlations between incidence, mortality, and region.