Analysis of Tuberculosis Affects on Different Countries

**Sashank Addanki Venkata Naga || Rithik Rathinavel Ragupathi ||
Irfan Saleemudeen || Kalidindi Saketh Varma**

Abstract

This project undertakes an in-depth analysis of the global tuberculosis (TB) burden using a dataset that spans multiple years and countries. This report presents key findings and statistical analysis on tuberculosis (TB) prevalence, mortality, and incident cases across various countries, based on historical data and future forecasts. The insights derived from these findings are critical for shaping public health strategies and interventions. The findings from this analysis underscore the persistent and evolving challenge of tuberculosis in India and Nigeria which happened to have the highest incident cases/deaths among other countries.

Introduction

Tuberculosis (TB) remains a significant public health challenge globally. Understanding the burden of TB is critical for developing effective interventions and policies. This project aims to analyze the "TB_Burden_Country" dataset to uncover patterns and insights that can aid in the fight against TB.

The research questions guiding this analysis is:

- Understand TB trends across regions and years.
- Highlight countries with the highest TB burden.
- Predict future trends to guide healthcare strategies.

Methods

Data Preprocessing: The dataset was first cleaned and preprocessed using SQL to handle missing values and ensure consistency.

Exploratory Data Analysis (EDA) & Statistical Analysis: by accessing the data frame from SQL we visualized data trends, distributions, and relationships to understand patterns of cases/deaths among the countries by employing various data analysis and visualization libraries such as Pandas, NumPy, Matplotlib, Scikit-learn, and Seaborn. Also, descriptive statistics were calculated to summarize the data, and statistical methods, including correlation analysis and regression modeling, were employed to explore relationships between different variables.

LangChain and OpenAI: Used GPT models (e.g., GPT-4, GPT-3.5-turbo) for generating insights

Results

- The bar plot(Figure 1) shows that India has the highest estimated prevalence of TB among the listed countries. The other countries following India in terms of high TB prevalence are China, Indonesia, and Pakistan.
- The bar plot(Figure 2) shows that India stands out with the highest TB mortality rate, highlighting a critical area for TB control and prevention efforts. Nigeria, Indonesia, and

China also show significant numbers of TB-related deaths, indicating these countries require substantial public health interventions. The countries listed are crucial targets for global health initiatives aiming to reduce TB mortality

- The bar plot(Figure 3) shows that Nigeria has the highest number of deaths from TB among HIV-positive individuals, followed by South Africa and India. Uganda, Mozambique, and Zimbabwe also have significant TB mortality rates. These countries highlight critical regions where TB control and HIV integration efforts are urgently needed.
- The correlation matrix(Figure 4) helps identify which TB metrics are closely related. For example, if the prevalence of TB increases, the incidence and mortality rates are likely to increase as well.
- The graph (Figure 5) indicates that TB prevalence initially increased slightly in the early 1990s, reached a peak in the early 2000s, and then gradually decreased.
- Geographical Insights of TB mortality rates(Figure 6):
    - High Mortality Regions: The dark purple areas highlight countries where TB mortality rates are the highest. These are primarily located in Sub-Saharan Africa and parts of Southeast Asia.
    - Moderate Mortality Regions: The blue areas show countries with moderate TB mortality rates, which include regions in South Asia, parts of Africa, and Eastern Europe.
    - Low Mortality Regions: The yellow area indicates a country with a relatively low TB mortality rate.
- Geographical Insights of TB prevalence(Figure 7):
    - High Prevalence Areas (Yellow): Parts of Sub-Saharan Africa and South-East Asia are highlighted in yellow, indicating these regions have the highest TB prevalence.
    - Moderate Prevalence Areas (Blue/Purple Gradient): Several countries in Africa, Asia, and parts of Eastern Europe show a moderate prevalence of TB, with varying shades between purple and blue.
    - Low Prevalence Areas (Dark Purple):North America, Western Europe, and some parts of Oceania are in darker purple, indicating these regions have the lowest prevalence of TB.
- Model fit/accuracy(Figure 8(a))
    - Model Accuracy: The high R-squared value (0.98) indicates that the model fits the data very well.

    - Error Margin: The MSE value (8287670.07) helps to understand the average squared error between the predicted and actual values.
    - Coefficient Impact: Each coefficient shows how each predictor variable affects the outcome. Positive values increase the outcome, while negative values decrease it.
    - Intercept: 5023.467134163877. This is the predicted value of the dependent variable when all independent variables are zero.
    - Scatter plot(Figure 8(b): the dots are clustered close to the red dashed line, indicating that the predictions are generally accurate.
- Future Forecast
    - Death cases:
        - Predict a slight decrease in TB deaths in India(Figure 9) from 2025 to 2030.
        - There is neither increase nor decrease in Nigeria(Figure 10) from 2025 to 2030.
    - Incident Cases:
        - In India(Figure 11)the number of cases from 2025 to 2030 seems to be at par with the number of cases registered in 2013.
        - In Nigeria(Figure 12)steady rise in incident cases over the years, and the forecasted continued increase, underscore the need for enhanced TB control and prevention measures to address this ongoing public health issue effectively.

Conclusion

Tuberculosis (TB) remains a significant public health and economic challenge, demanding targeted interventions and strategic resource allocation, particularly in high-burden regions. Economic development is crucial in reducing TB incidence and mortality, as broader socio-economic improvements positively impact public health. Future efforts should focus on:

- **Healthcare Investments:** Strengthening healthcare infrastructure and enhancing access to diagnostic tools and treatment facilities.
- **Policy Initiatives:** Implementing national TB control programs for early detection and ensuring free or subsidized TB medications in vulnerable regions.
- **Research and Development:** Investing in vaccine development and promoting studies on drug-resistant TB strains.
- **Collaborative Efforts:** Encouraging collaboration among governments, NGOs, and the private sector to combat TB effectively.

Addressing these aspects will help develop sustainable global strategies for TB control, improve health outcomes, and reduce the TB burden globally.

Appendix

Figure 1:

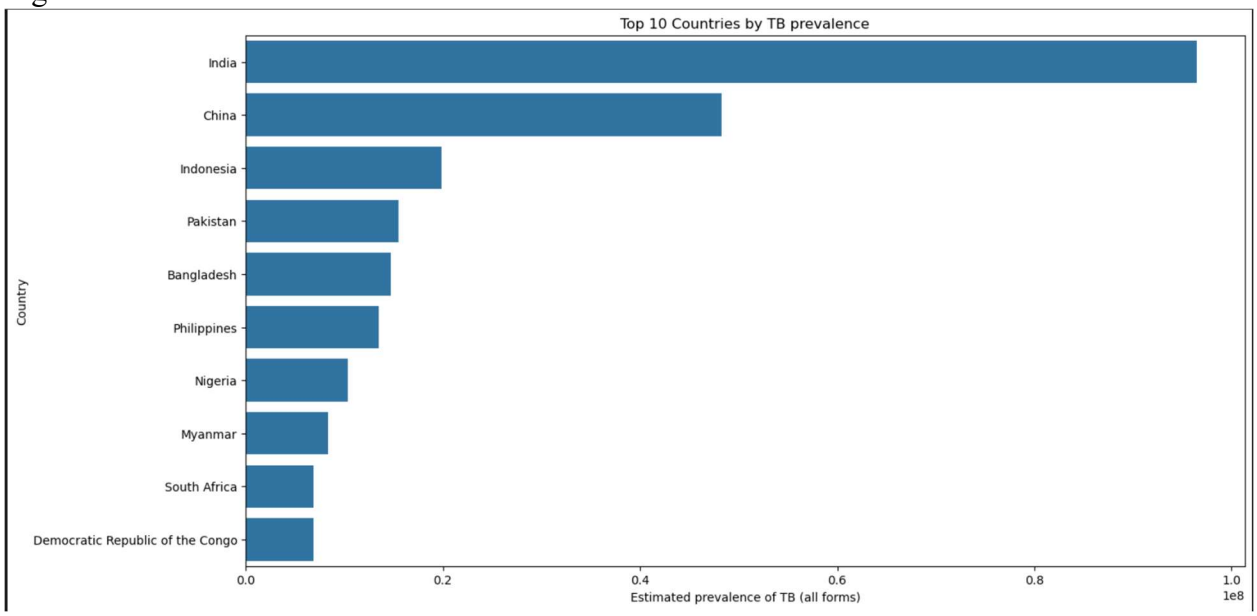


Figure 2:

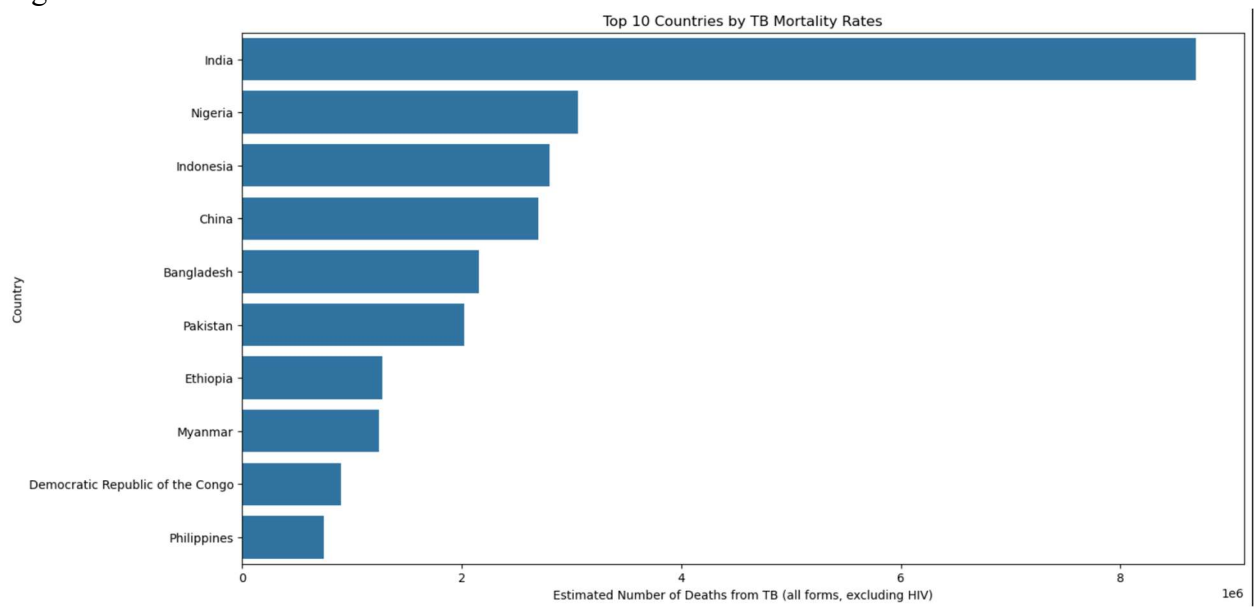


Figure 3:

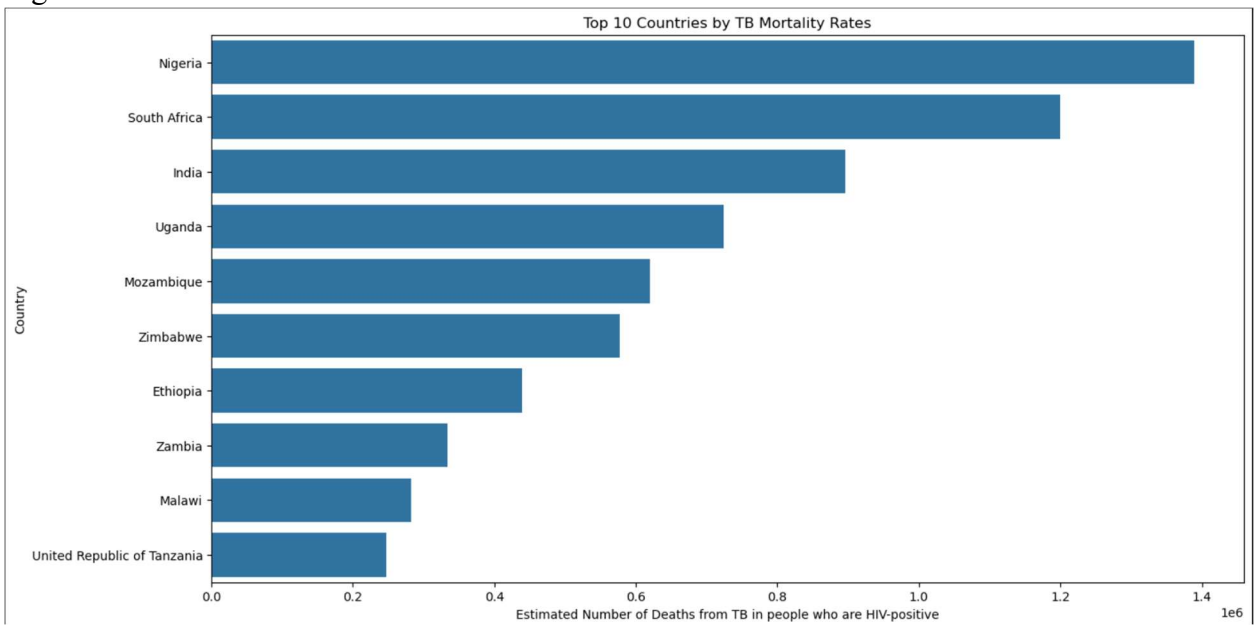


Figure 4:

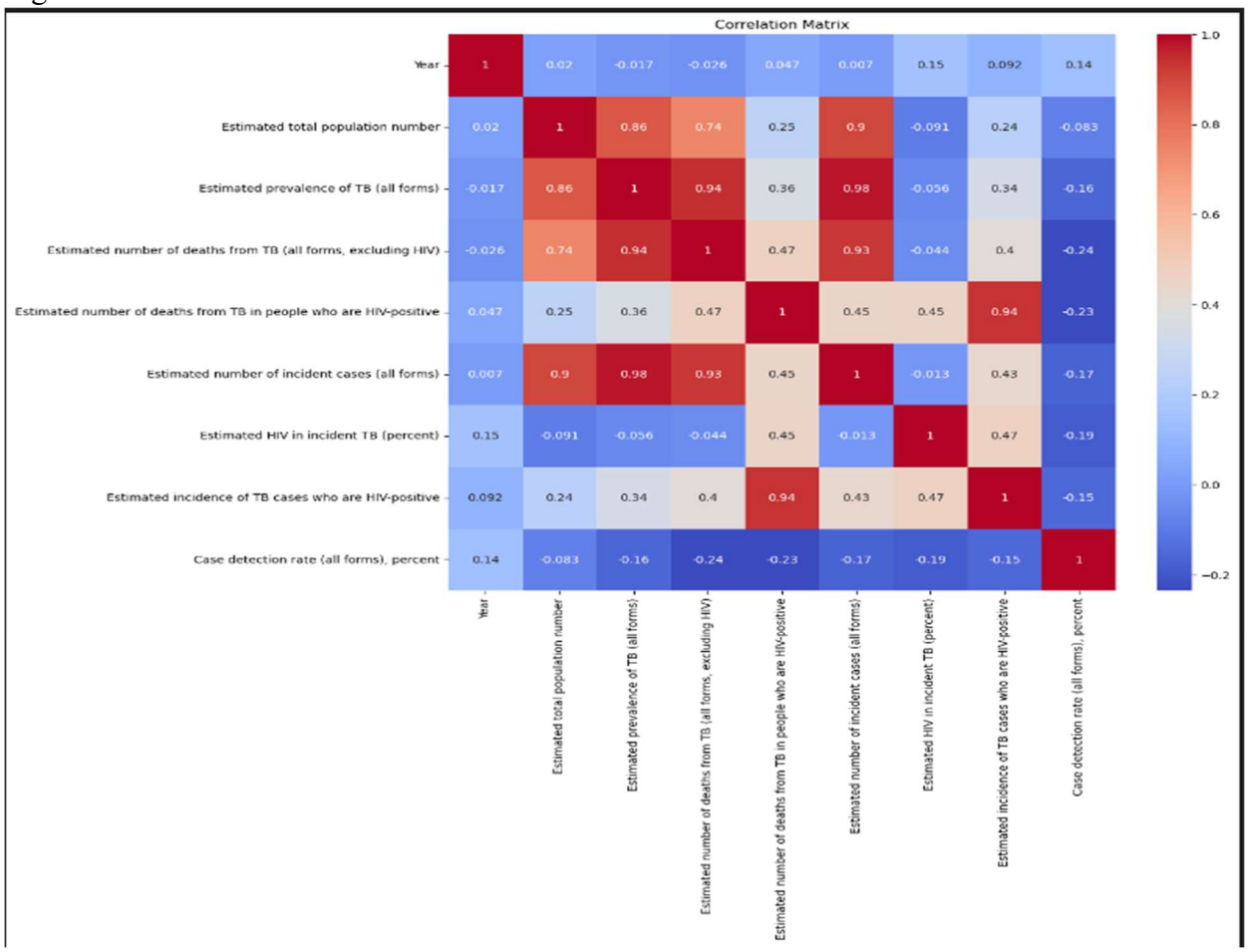


Figure 5:

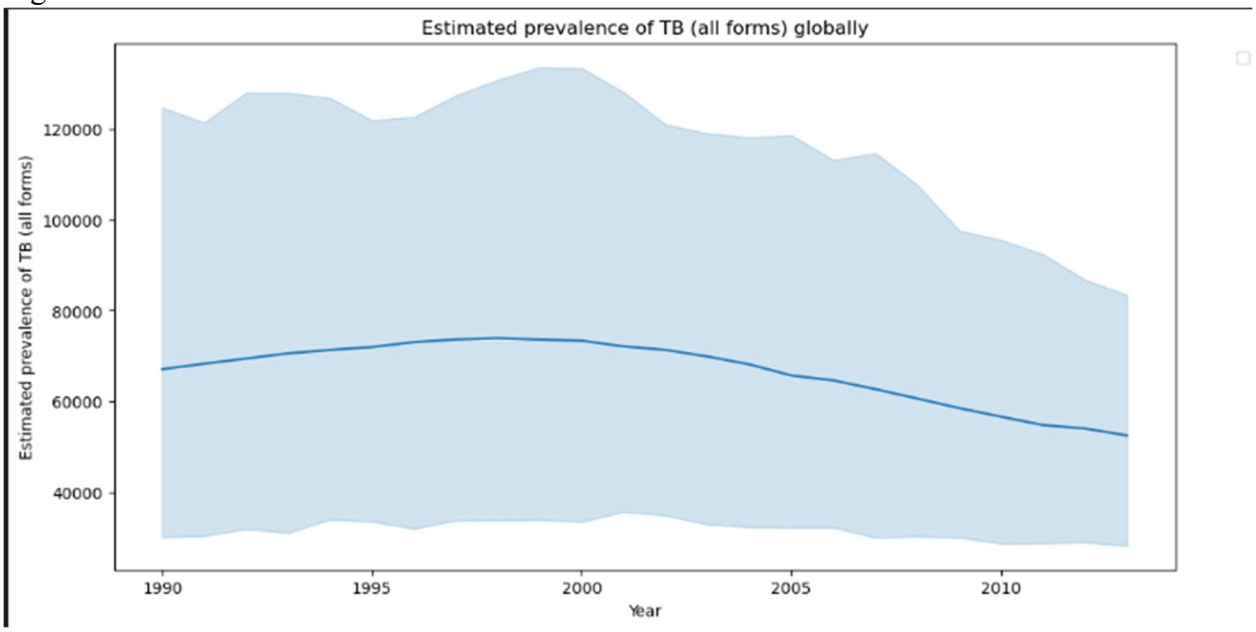


Figure 6:

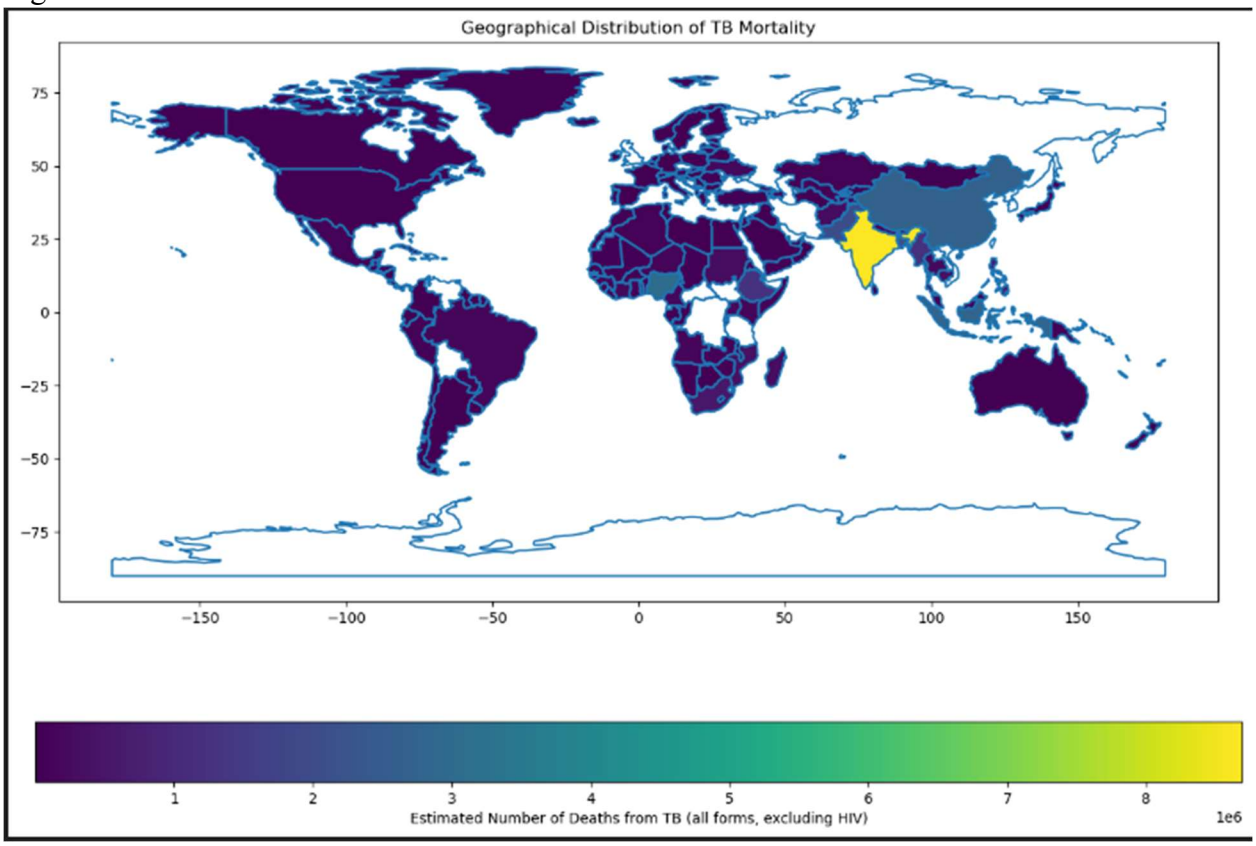


Figure 7:

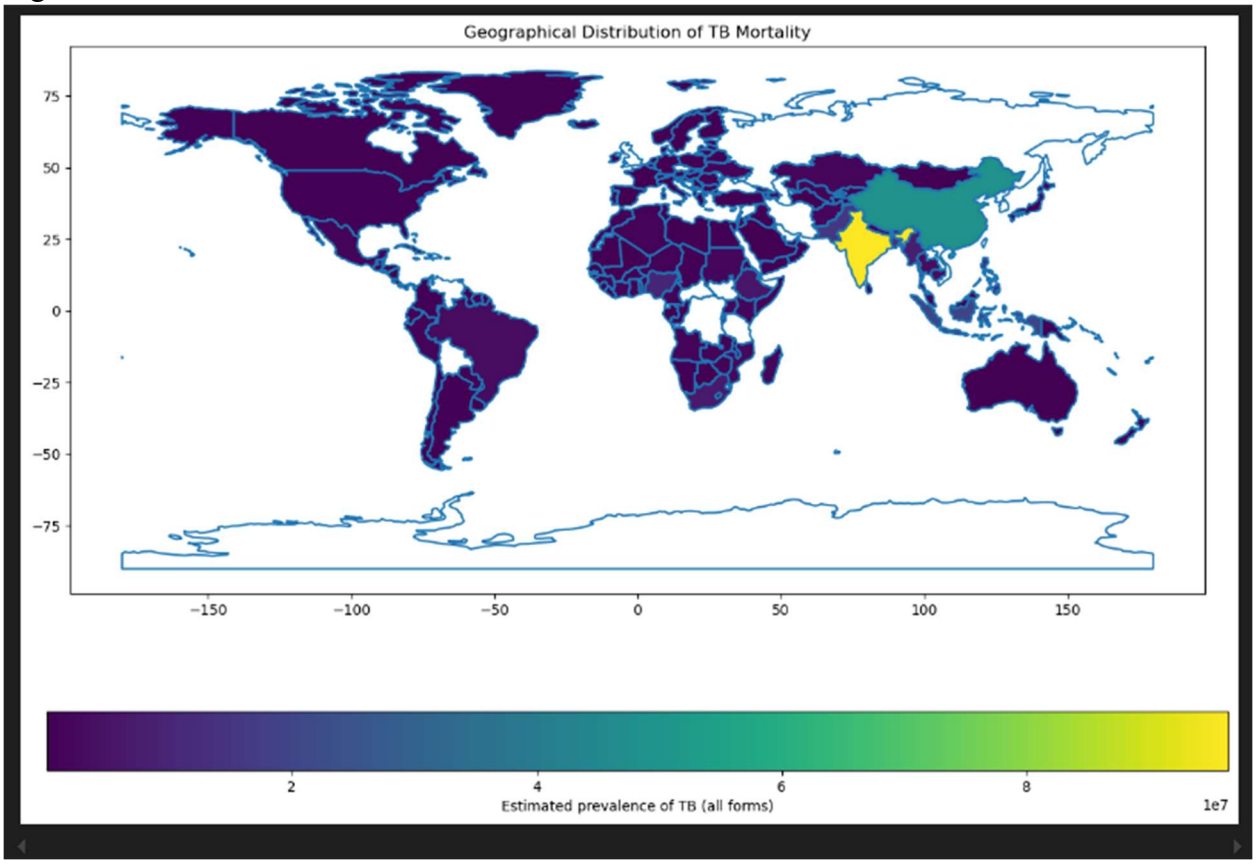


Figure 8 (a):

```
Mean Squared Error: 8287670.066046737
R-squared: 0.980715232598726 3
Regression Coefficients: [-1.78941698e-05  7.32458590e-02  7.42339953e-02 -5.31345696e+01
 -7.39475656e+02 -1.59863711e+03 -1.11644888e+03 -4.00354303e+02
 -2.95138284e+03 -4.74685960e+03 -3.99265823e+02 -1.86617310e+03
 -8.64520353e+02 -4.00130652e+02 -2.37344223e+02 -3.54424774e+02
 -1.04537087e+04 -4.87468878e+02 -4.07195697e+02  2.09086106e+04
 -1.03149187e+03 -1.21677460e+03 -3.36502689e+02 -3.59800279e+02
 -2.23559665e+03 -4.00098252e+02 -2.00509036e+03 -3.91324413e+03
 -4.00465381e+02 -1.77624339e+03 -2.57711468e+03 -7.82021857e+03
 -3.99752055e+02 -4.23134221e+02 -1.13693877e+03 -2.62718577e+03
 -2.13108101e+03 -2.73269444e+03 -2.57919433e+03 -2.24438241e+03
 -5.81009049e+01 -4.00384464e+02 -2.94792423e+03 -2.49570693e+03
 -3.71445820e+02 -1.12336358e+05 -1.15424430e+03 -4.51990420e+02
 -2.68227292e+03 -2.14329562e+03 -2.54664102e+03 -4.00474018e+02
 -1.34791820e+03 -4.30553137e+02 -1.53794022e+03 -3.98274637e+02
 -4.56231530e+02 -3.98561471e+02 -1.92558817e+03  5.30143289e+03
  2.80205434e+03 -3.82854292e+02 -1.27529517e+03 -3.98334132e+02
 -2.53002359e+03 -3.31007237e+03 -3.10268234e+03 -5.41927268e+02
 -4.20484785e+02 -3.36785735e+02 -3.94712684e+02  1.85216013e+04
 -3.00315749e+03 -3.12893468e+02 -3.41068396e+02 -4.02790685e+02
 -2.62196010e+03 -1.10425383e+03 -4.42200652e+03  7.41109148e+00
 -1.99395159e+03 -1.89477133e+02 -4.08645199e+02 -2.11234277e+03
 -4.03619142e+02 -3.82872621e+03 -2.12354236e+03 -2.86963362e+03
 -2.04064104e+03 -1.11199502e+03 -1.89431525e+03 -6.31612657e+02
 -3.87977602e+02 -7.65291526e+04  2.33077415e+04 -6.44811217e+02
  8.03347173e+02 -3.95183374e+02 -2.79529384e+02  1.60920654e+02
 -1.42294852e+03 -2.30773261e+03 -2.68880872e+02 -5.57323109e+03
 -7.47821378e+03 -9.52986362e+02 -4.47979947e+02 -2.50847828e+03
 -2.80678947e+03 -1.58101885e+03 -4.47083053e+01 -3.15192009e+03
 -2.80678947e+03 -1.58101885e+03 -4.47083053e+01 -3.15192009e+03
 -2.65246881e+03 -1.14395700e+03 -9.46727161e+02 -3.88680712e+02
  1.41320073e+02 -4.53545456e+03 -1.95939916e+03 -1.17227777e+03
 -1.94584573e+03 -2.75818052e+02 -7.77475161e+02 -2.20067809e+03
 -2.64654277e+03 -2.27379364e+03 -2.09663504e+03 -4.00165064e+02
 -3.17757349e+03 -4.26880823e+02 -3.99905899e+02 -2.93384137e+03
 -2.88898896e+03  1.12315045e+04 -2.97570429e+03 -4.00935275e+02
 -2.43593844e+03 -2.97173550e+02 -3.98116440e+02 -4.03282230e+02
 -3.94384003e+02 -4.56464443e+02 -1.38032326e+03  6.22558765e+04
 -4.00705097e+02 -4.04484962e+02 -3.82178946e+02 -3.81483809e+02
  7.32045270e+03 -4.01210695e+02 -7.90450189e+02 -1.99211672e+03
 -1.04255306e+03 -4.62591044e+03 -3.15816830e+04 -8.43835276e+02
 -7.56200990e+02 -3.30301475e+02 -4.44563775e+02 -5.54964956e+03
 -1.81401270e+03 -5.17545130e+03 -2.09057015e+03 -1.41676610e+03
 -3.99240126e+02 -3.96685660e+02 -3.11941892e+03 -1.28297868e+03
 -4.00450845e+02 -1.69371550e+03  1.09744628e+02 -1.84771017e+03
 -4.53837831e+02 -2.32319896e+03 -4.01611051e+02 -2.31496032e+03
 -5.29443943e+02 -4.00172821e+02 -4.41380257e+02 -3.97533960e+02
 -2.39439856e+03 -1.42746143e+03 -2.21438912e+04 -1.85286784e+03
 -7.81757049e+02 -2.31542262e+03 -6.76803463e+02 -3.34177551e+03
 -2.45135094e+03 -3.53421376e+02 -4.05438969e+02 -6.75257490e+02
 -4.11836208e+03 -6.73706329e+03  6.85819137e+01 -1.40901571e+03
 -2.02627352e+03 -4.00728660e+02 -1.23636200e+03 -4.47970392e+02
 -4.64759782e+02 -1.23210265e+03 -2.81466101e+03 -4.00711647e+02
 -1.54285346e+03 -3.98801281e+02 -5.51964922e+03 -2.85232123e+02
 -1.46551935e+03 -9.88219230e+01 -5.54342452e+03  2.41295573e+03
 -4.22855338e+02 -2.32658674e+03 -1.06515821e+03 -1.51733256e+03
 -6.54844824e+03 -2.62781060e+02 -3.80567192e+03 -1.57311708e+03
 -6.23409239e+03 -7.12245251e+03]
Intercept: 5023.467134163877
```

Figure 8 (b):

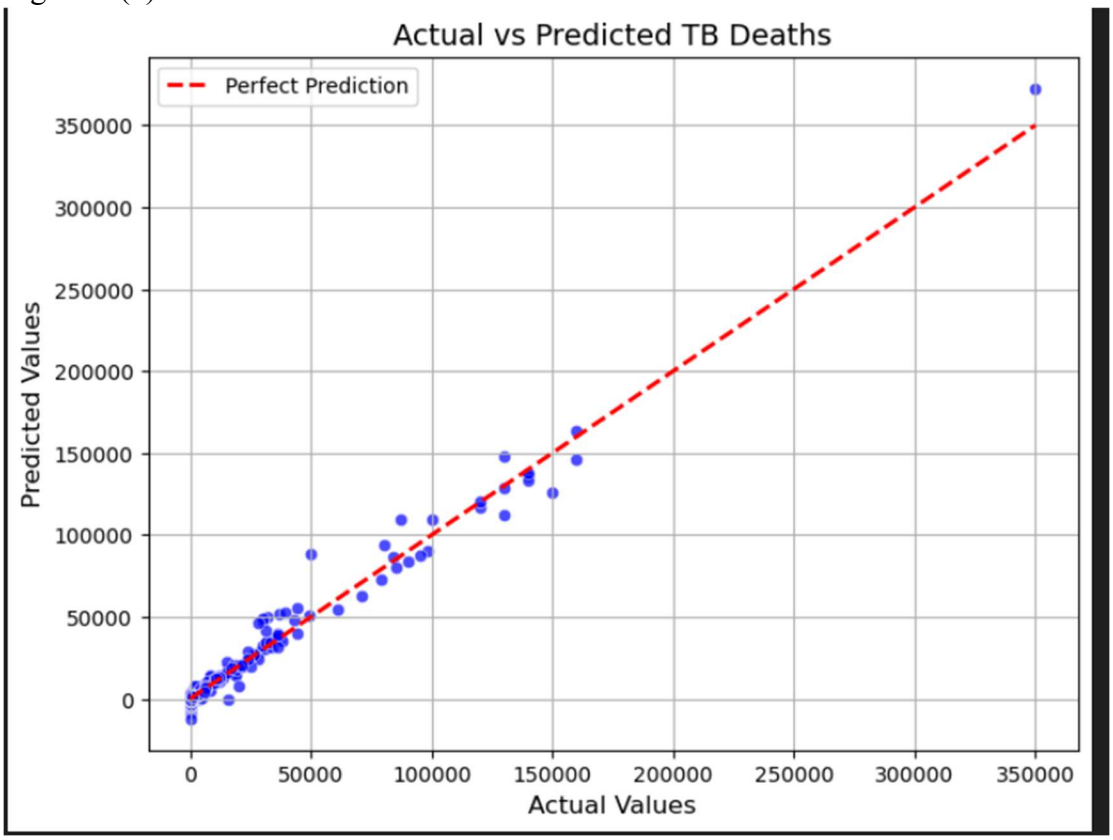


Figure 9:

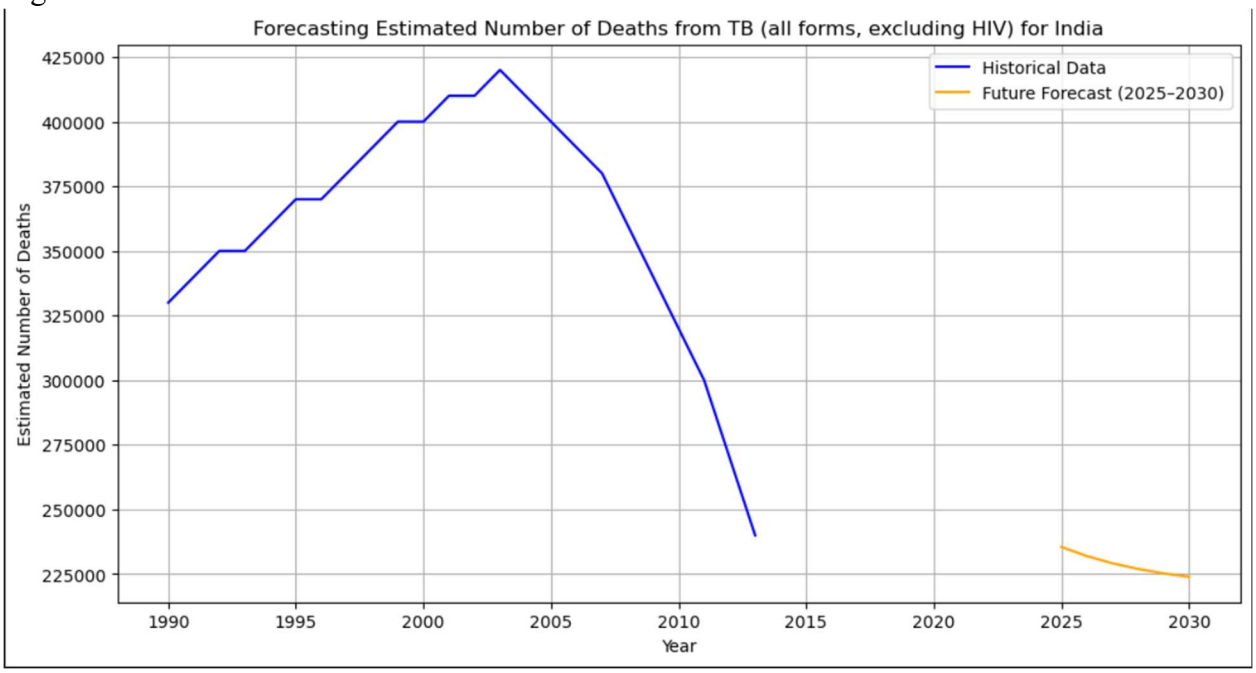


Figure 10:

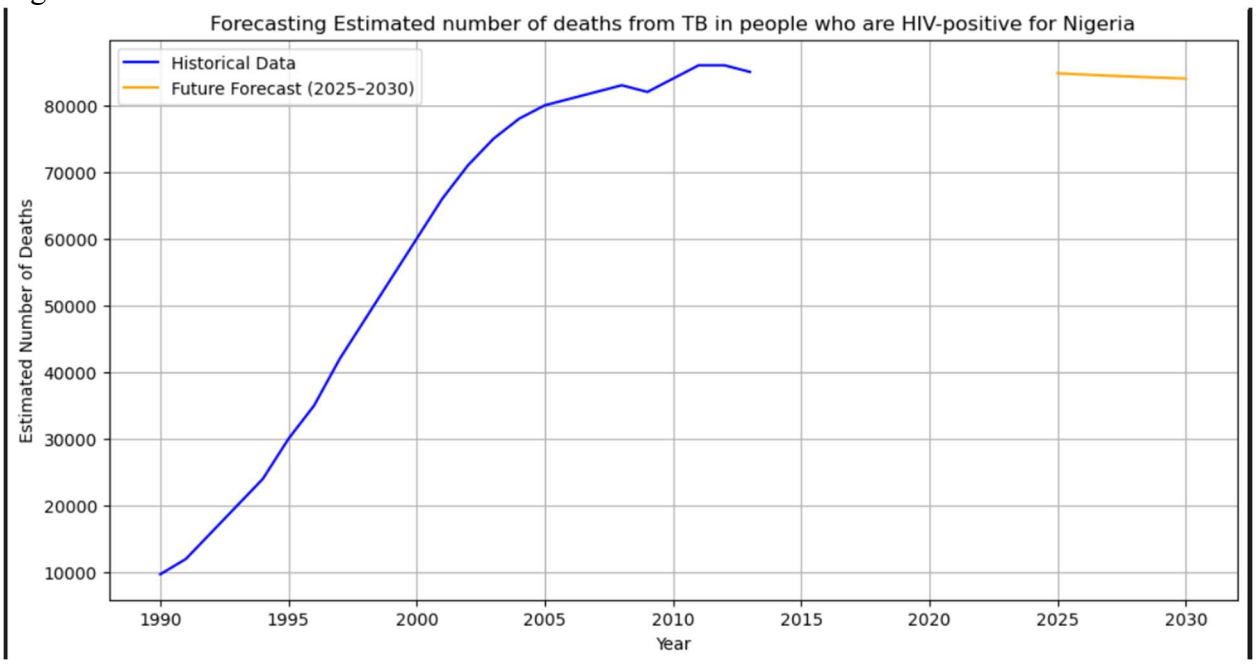


Figure 11:

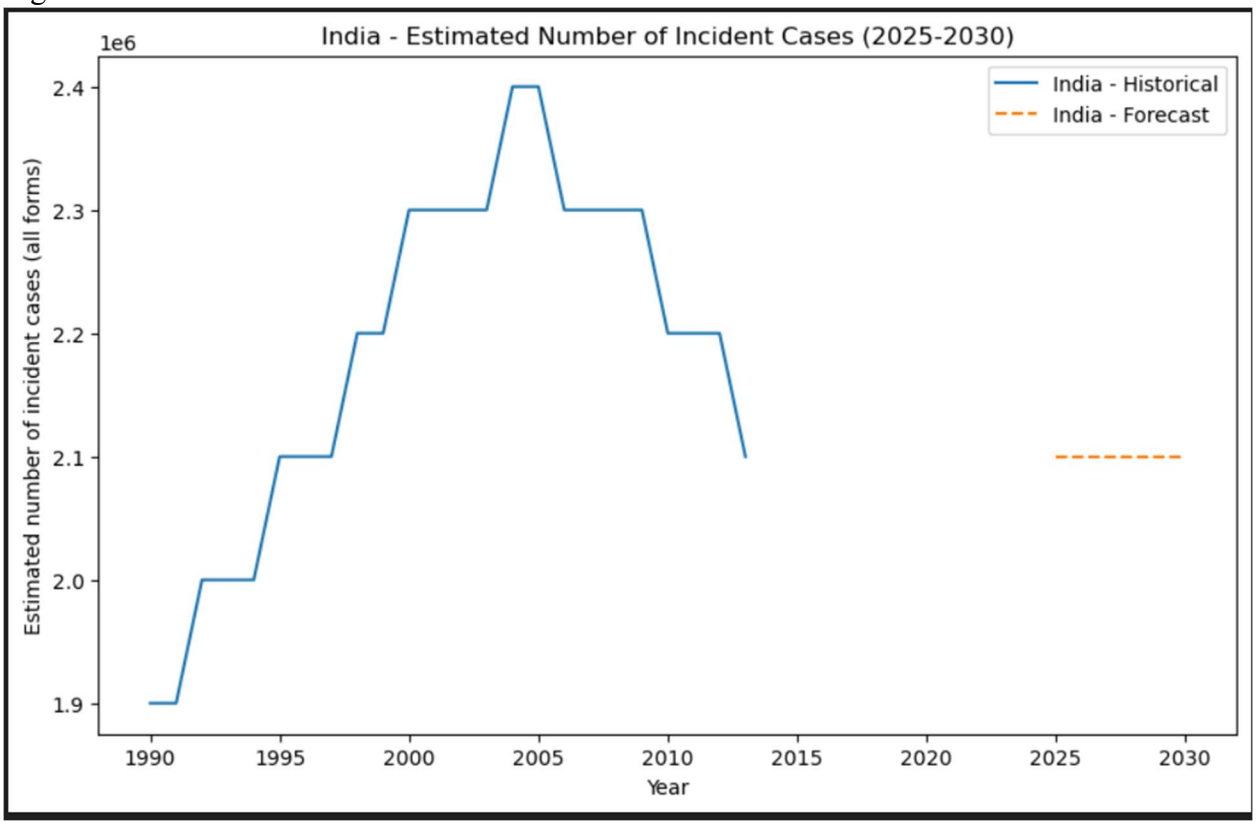
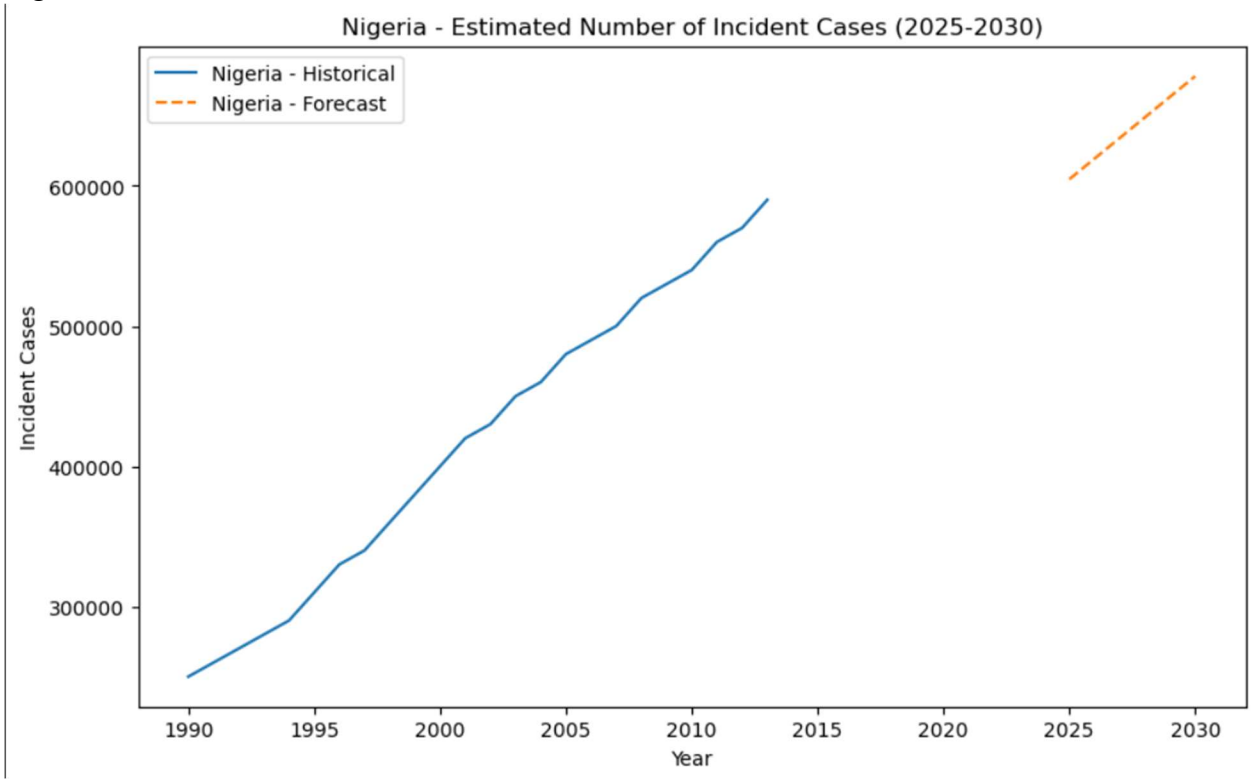


Figure 12:



References :

Dataset Source: https://public.tableau.com/app/learn/sample-data