

# MTH208

Group 24 Course Project

2025-2026

**TOPIC: GLOBAL TUBERCULOSIS  
ANALYSIS**

# INTRODUCTION

- Tuberculosis (TB) is caused by bacteria (*Mycobacterium tuberculosis*). It majorly affects the lungs.
- **India has the highest burden of TB** with two deaths occurring every three minutes from it.
- **Social factors:** poverty, undernutrition, literacy, and rapid urbanization influence TB prevalence.
- We aim to find correlations and draw analysis from the visualization of these factors with mortality, reporting of cases, etc.

# OBJECTIVES

- Track global and country specific trends in TB.
- Explore correlations of TB burden with various socioeconomic indicators
- Analyze the scale and trend of multidrug-resistant TB (MDR-TB) and rifampicin-resistant TB (RR-TB).
- Analyze the impact of COVID-19 on TB notifications and outcomes.
- To provide a user-friendly, data-driven, dashboard-app (Shiny app) that can be used and extended for further research purposes.

# **LIBRARIES USED :**

<b>LIBRARIES</b>	<b>DESCRIPTION</b>
<b>shiny</b>	<b>Creates app with interactive interfaces</b>
<b>shinydashboard</b>	<b>Used for building sidebar in shiny app</b>
<b>ggplot2</b>	<b>Mainly used for plotting of various types of data</b>
<b>dplyr</b>	<b>Used for grouping and summarizing data</b>
<b>plotly</b>	<b>Allows the creation of interactive web graphics from ggplot2 graphs</b>
<b>shinyWidgets</b>	<b>Collection of custom input controls and user interface components for 'Shiny'</b>

# **KEYWORDS**

## **1. TB Incidence**

Number of new active tuberculosis cases per 100,000 population in a year.

## **2. TB Mortality**

Number of deaths due to TB per 100,000 population in a year.

## **3. TB Notifications**

Number of cases officially reported to health authorities.

## **4. MDR-TB**

TB resistant to both isoniazid and rifampicin.

## **5. RR-TB**

TB resistant to rifampicin. All MDR-TB cases are RR-TB, but not all RR-TB cases are MDR-TB.

## **6. Latent TB Infection**

Person carries TB bacteria but shows no symptoms and is not contagious.

## **7. Treatment Success Rate**

Percentage of TB patients who successfully complete treatment.

## **8. Underreporting / Missed Cases**

Difference between estimated incidence and reported notifications.

## **9. TB Risk Factors**

HIV, diabetes, malnutrition, smoking, urbanization, and crowding.

## **10. COVID-19 Impact**

Healthcare disruptions reduced TB notifications during lockdowns.

# DATA DESCRIPTION

- The original data is collected from WHO's database. It consisted of multiple excel sheets with their TB data: budget, notifications, incidences, etc.
- Each of the excel files had multiple indicators, and from those, handful were picked for our analysis. The data was understood using their data dictionary
- The specific variables were then extracted from the excel sheets and combined with extra WDI indicators such as GDP and %urbanization. The final dataset was made into a csv file.
- All the data was combined on the basis of country codes.

country_who	year	g_whoregi	e_pop_num	e_inc_100k	e_inc_num	e_mort_100k	e_mort_num	e_inc_tbhiv_100	e_inc_tbhi	e_mort_tb	e_mort_tbcfr	cfr_pct	c_newinc_c_cdr	
Afghanistan	2000	EMR	20130323	190	38000	68	14000	0.03	6	0.17	34	0.37	37	35 19
Afghanistan	2000	EMR	20130323	190	38000	68	14000	0.03	6	0.17	34	0.37	37	35 19
Afghanistan	2001	EMR	20284311	189	38000	63	13000	0.03	6	0.3	61	0.35	35	50 26
Afghanistan	2001	EMR	20284311	189	38000	63	13000	0.03	6	0.3	61	0.35	35	50 26
Afghanistan	2002	EMR	21378110	189	40000	57	12000	0.03	6	0.27	58	0.31	31	65 34
Afghanistan	2002	EMR	21378110	189	40000	57	12000	0.03	6	0.27	58	0.31	31	65 34
Afghanistan	2003	EMR	22733047	189	43000	58	13000	0.03	6	0.29	66	0.32	32	61 32
Afghanistan	2003	EMR	22733047	189	43000	58	13000	0.03	6	0.29	66	0.32	32	61 32
Afghanistan	2004	EMR	23560660	189	44000	52	12000	0.03	6	0.29	67	0.28	28	78 41
Afghanistan	2004	EMR	23560660	189	44000	52	12000	0.03	6	0.29	67	0.28	28	78 41
Afghanistan	2005	EMR	24404569	189	46000	47	12000	0.03	7	0.31	75	0.26	26	90 47
Afghanistan	2005	EMR	24404569	189	46000	47	12000	0.03	7	0.31	75	0.26	26	90 47
Afghanistan	2006	EMR	25424099	189	48000	43	11000	0.03	9	0.32	82	0.24	24	100 53
Afghanistan	2006	EMR	25424099	189	48000	43	11000	0.03	9	0.32	82	0.24	24	100 53
Afghanistan	2007	EMR	25909852	189	49000	39	10000	0.04	10	0.33	85	0.21	21	111 59

# ANALYSIS AND RESEARCH

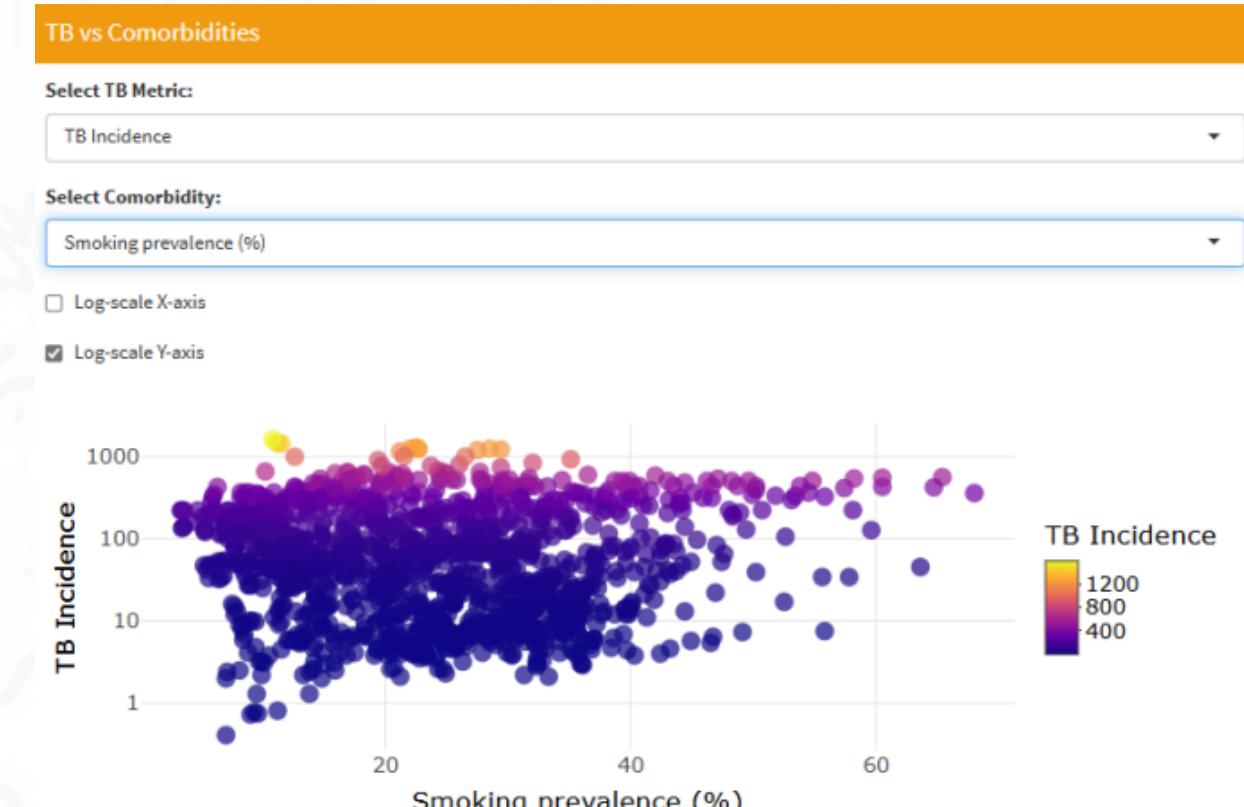
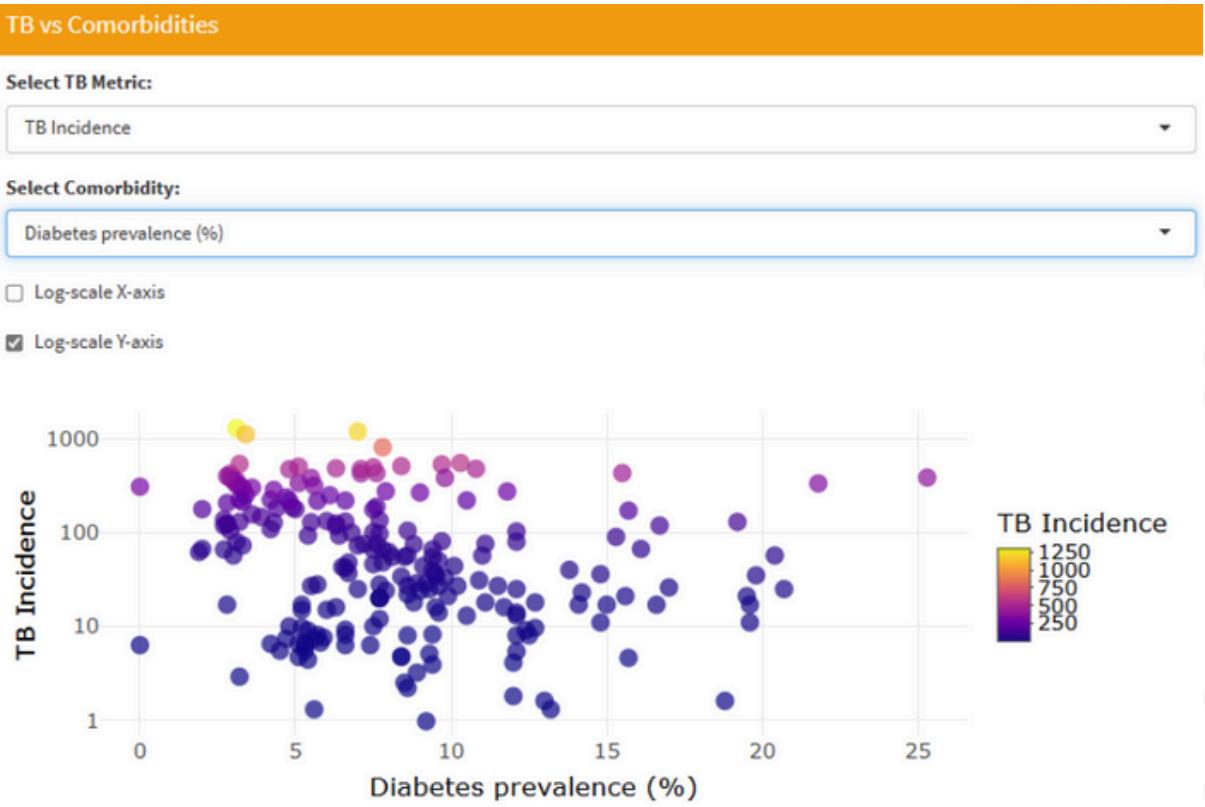
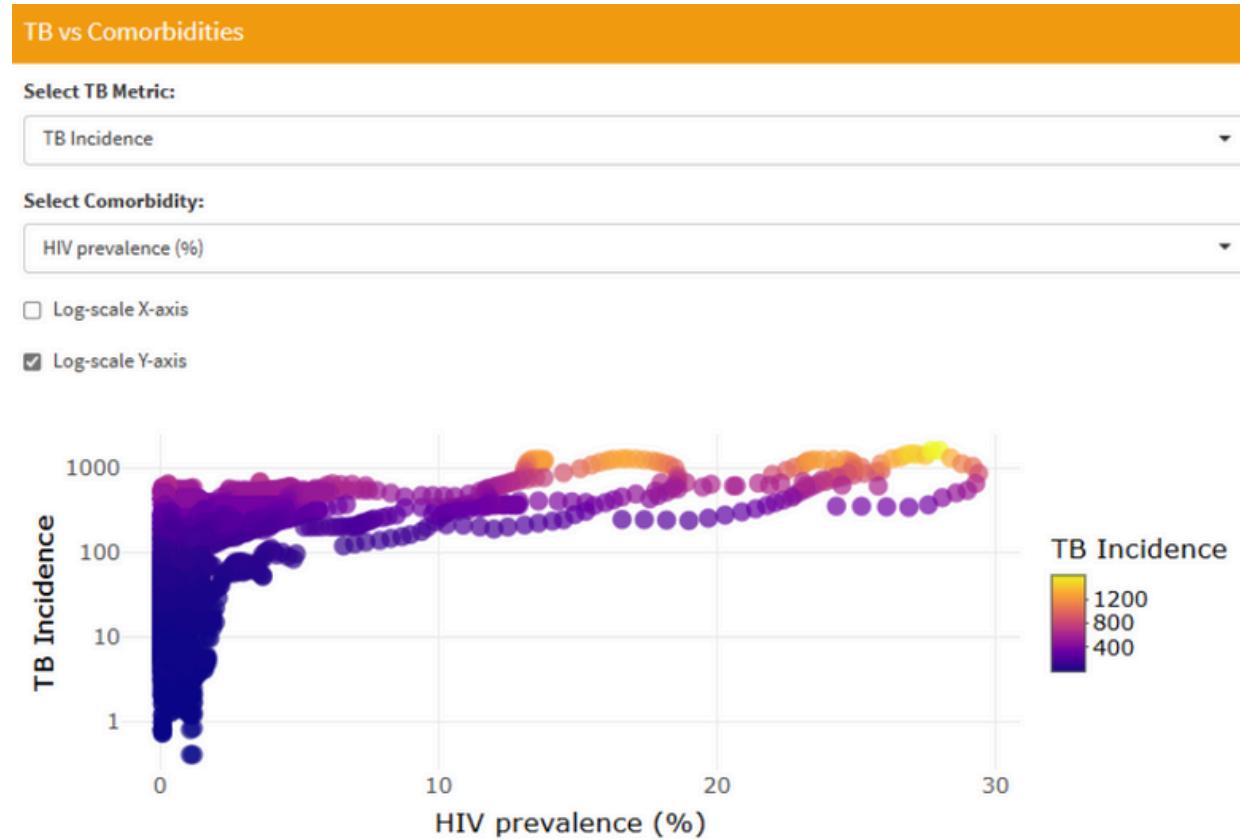
## 1. GLOBAL LEVEL FINDINGS



- TB incidence and mortality patterns show that African countries remain the most affected in recent years.
- Animated bubble plots reveal an inverse relationship between GDP per capita and TB incidence—countries with lower income levels experience higher TB rates.

# ANALYSIS AND RESEARCH

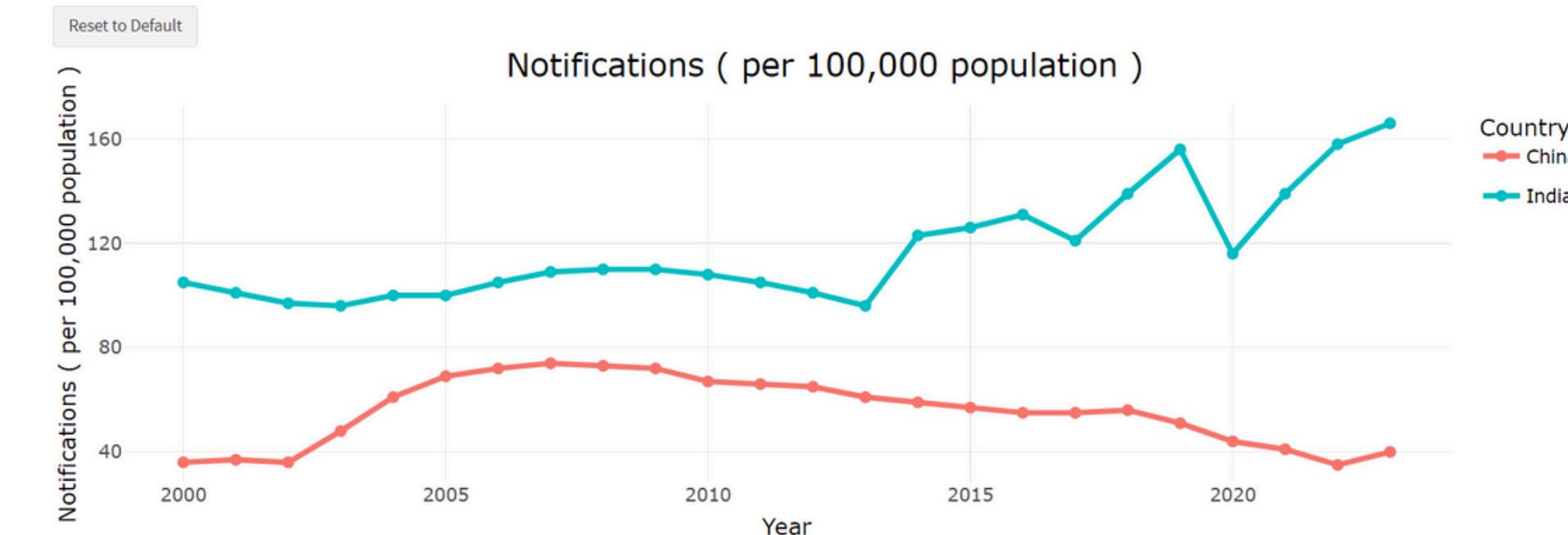
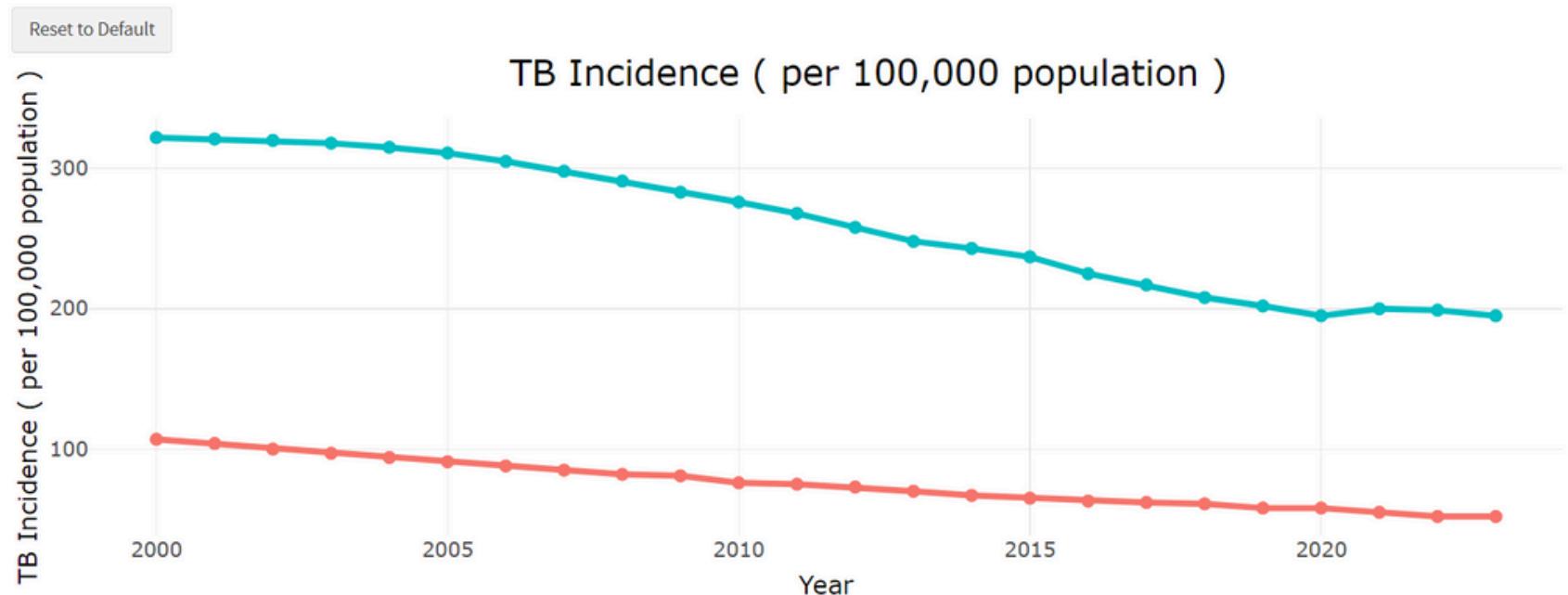
## 2. RELATIONS WITH COMORBIDIIES



- TB incidence rises sharply with higher HIV prevalence, confirming HIV as a key risk factor.
- Diabetes shows no clear pattern — data points are widely scattered.
- Smoking demonstrates a weak to moderate relationship with TB incidence.

# ANALYSIS AND RESEARCH

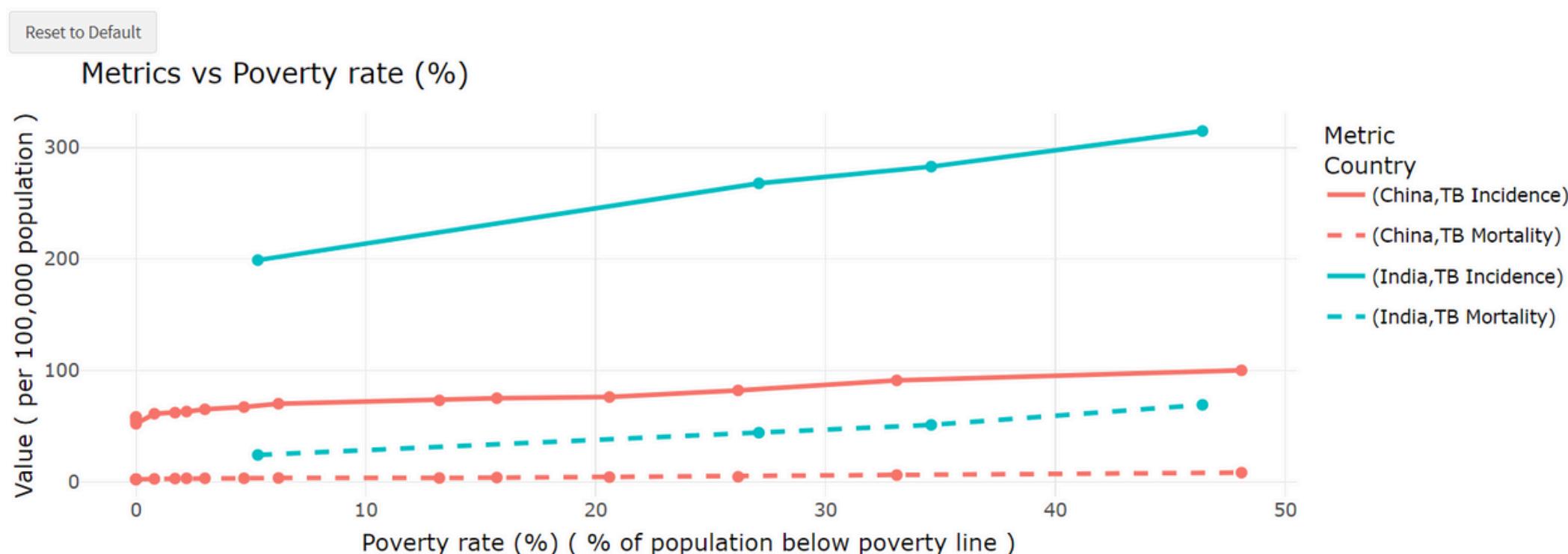
## 3. TB TRENDS: COMPARISON BETWEEN COUNTRIES



- India and China both show a consistent fall in TB incidence, indicating effective prevention strategies.
- China shows a gradual decline in notifications, while India displays a gentle upward trend with small ups and downs over the years.
- Overall trend shows progress, but with differing patterns of case reporting between the two nations.

# ANALYSIS AND RESEARCH

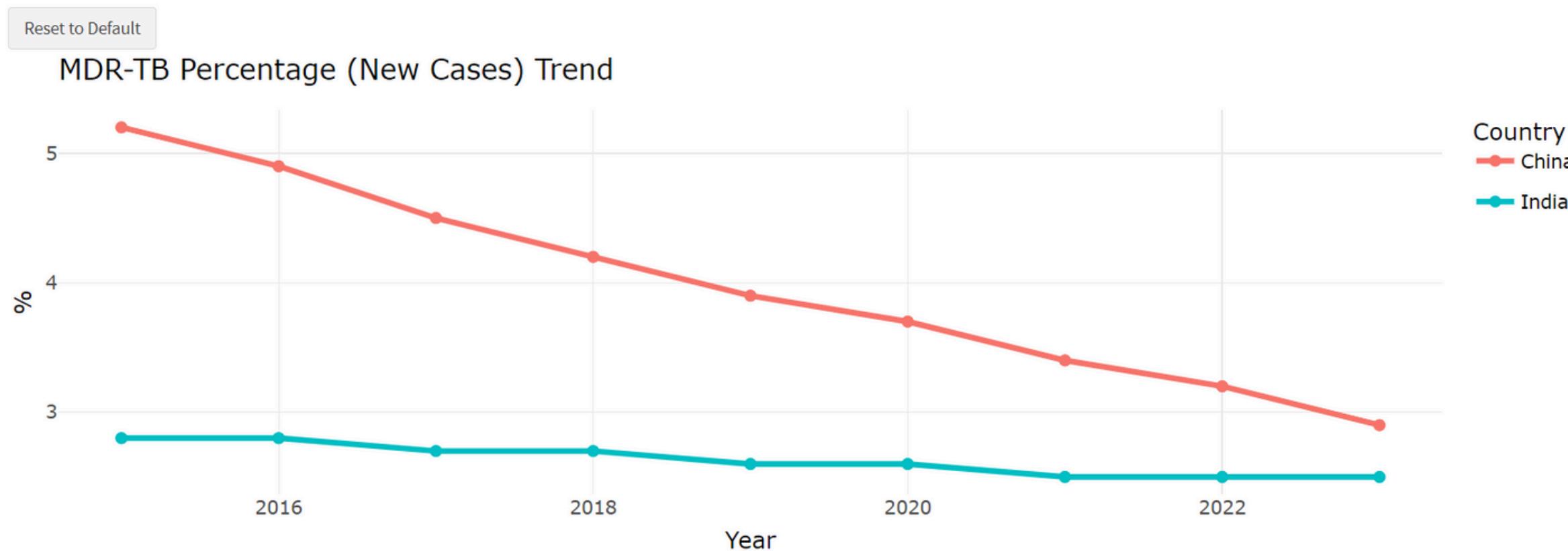
## 4. CORRELATIONS WITH SOCIOECONOMIC FACTORS



- At a comparable GDP, the incidence and mortality of India are higher than China.
- Countries with stronger economies show sharper declines in TB incidence and mortality.
- As poverty increases, TB burden rises noticeably, reflecting socioeconomic inequality.
- Improving income levels and living standards directly supports TB control.

# ANALYSIS AND RESEARCH

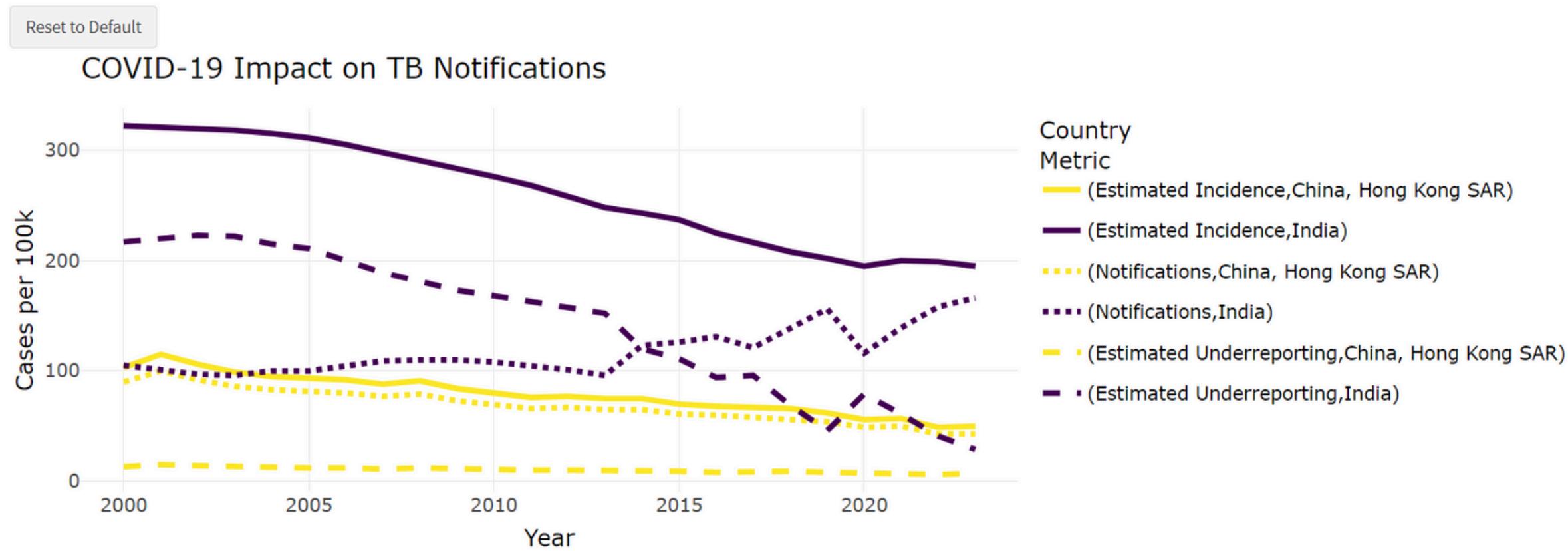
## 5. TRENDS IN MULTI-DRUG RESISTANT TB



- China shows a strong downward shift, indicating major success in reducing its initially high MDR-TB burden.
- India maintains a consistently low level, suggesting better baseline control but slower further improvement.
- Together, the trends reflect different stages of progress — rapid decline in China versus steady stability in India.

# ANALYSIS AND RESEARCH

## 6. ANALYSIS OF DISRUPTION CAUSED DUE TO COVID-19



- China's TB reporting and control systems were less disrupted by COVID-19.
- India, on the other hand, experienced a significant interruption in TB detection and notification due to pandemic-related healthcare strain.
- The contrast highlights how China maintained stability in TB control while India faced major pandemic setbacks.

# **CONCLUSION**

The project provides a statistical and visual analysis of global and country-specific TB patterns. The plots of T.B. incidence with variables like GDP, poverty levels, health expenditure, etc analyzes the dependencies of T.B. incidence with these factors. Additionally, the assessment of the COVID-19 pandemic's impact on TB notifications and treatment outcomes demonstrated how global health crises can disrupt essential disease control efforts. The Shiny dashboard provides a powerful tool for data visualization and policy-oriented research. Overall, the project contributes meaningfully to the understanding and monitoring of TB epidemiology and promotes data-driven strategies for global TB control.

# **LIMITATIONS**

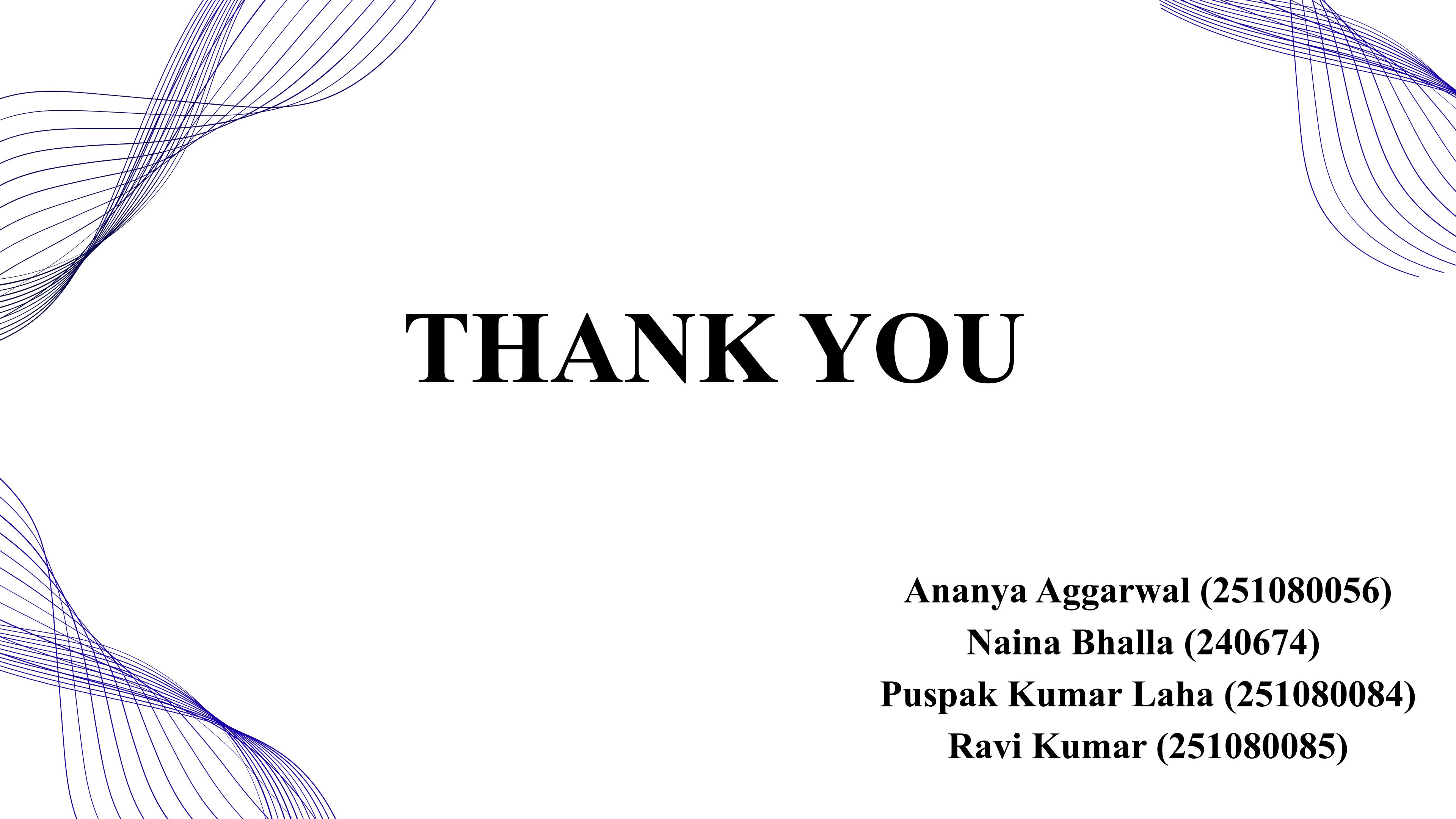
- Current analysis is descriptive, lacks predictive modeling.
- Missing or incomplete WHO or WDI data introduce uncertainty in trend estimation.
- Extremely handpicked indicators, cannot be directly used on another WHO dataset since the variables are named differently

# **FUTURE IMPROVEMENTS**

- Apply ARIMA or exponential smoothing for TB incidence forecasting.
- Incorporate machine learning (e.g., Random Forest, XGBoost) for risk prediction.
- Add data imputation using statistical or ML techniques (e.g. KNN)

# REFERENCES

- **WHO Global TB Programme Data:** <https://www.who.int/teams/global-tuberculosis-programme/data>
- **World Bank World Development Indicators and WDI library:**  
<https://data.worldbank.org/indicator>
- **Global Tuberculosis Report 2024:** <https://iris.who.int/server/api/core/bitstreams/7292c91e-ffb0-4cef-ac39-0200f06961ea/content>
- **The Social Determinants of Tuberculosis: From Evidence to Action:**  
<https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2010.199505>
- **Sociodemographic factors affecting knowledge levels of tuberculosis patients in New Delhi:**  
[https://journals.lww.com/jfmpc/fulltext/2024/13110/sociodemographic\\_factors\\_affecting\\_knowledg\\_e.60.aspx](https://journals.lww.com/jfmpc/fulltext/2024/13110/sociodemographic_factors_affecting_knowledg_e.60.aspx)



# **THANK YOU**

**Ananya Aggarwal (251080056)**

**Naina Bhalla (240674)**

**Puspak Kumar Laha (251080084)**

**Ravi Kumar (251080085)**