📄 **Project Report – Tobacco Use and Mortality (2004–2015)**

Project Title:
🚬 📉 *Tobacco Use and Mortality Analysis (2004–2015)*

Student Name:
Saurabh Badhani
BCA Student – DSB Campus, Nainital

GitHub Link:
https://github.com/saurabh-badhani/Tobacco-Use-and-Mortality

📌 Project Overview

This project analyzes trends in smoking prevalence and its relationship with various health and mortality indicators over time. The objective is to explore correlations between tobacco use and its impact on public health, with visual insights drawn from multiple government and healthcare datasets.

📊 Datasets Used

The analysis combines five related datasets:

- **smokers.csv** – Smoking prevalence by age, sex, and year

- **prescriptions.csv** – Prescription records for smoking cessation over time

- **metrics.csv** – Health and economic indicators (expenditure, affordability)

- **fatalities.csv** – Smoking-related fatality statistics

- **admissions.csv** – Hospital admissions for smoking-related illnesses

---

🧹 Data Cleaning & Preprocessing

- Converted Year to integer or datetime type for consistency

- Missing values in smokers.csv (especially in age and prevalence) were filled with column means

- Unnecessary characters (e.g., newlines) removed from column names

- Replaced '.' and invalid entries with NaN, then filled or handled appropriately

- No duplicate rows were found in the datasets

🔎 Exploratory Data Analysis (EDA)

1. **Smoking Prevalence**:

   o   Males showed consistently higher prevalence than females

   o   A downward trend was observed from 2004 to 2015

   o   Middle-aged adults had the highest prevalence

2. **Fatalities**:

   o   Fatalities caused by smoking showed a slow decline over the years

   o   Correlated closely with smoking prevalence levels

3. **Hospital Admissions**:

   o   Admissions due to smoking-related diseases remained steady

   o   Peaks often aligned with higher smoking rates in certain years

4. **Prescription Trends**:

   o   Prescriptions for cessation aids (like NRT, Bupropion) gradually increased

   o   Spending on Varenicline was higher than other aids

5. **Tobacco Expenditure**:

   o   Household expenditure on tobacco decreased over time

   o   The affordability index declined, indicating tobacco is becoming less affordable

📈 Correlation Analysis

A heatmap was generated to analyze relationships between smoking prevalence, fatality rates, healthcare costs, and prescriptions.

Key Findings:

Relationship	Correlation Strength
Smoking Prevalence vs Fatalities	Moderate Positive
Smoking Prevalence vs Prescriptions	Weak Positive
Smoking Prevalence vs Admissions	Very Weak

✅ Conclusion

- **Smoking prevalence** is declining, especially among younger populations

- There is a **clear positive correlation** between smoking and **fatalities**

- Public health initiatives may be **effectively reducing** tobacco usage

- **Men still smoke more than women**, requiring gender-specific interventions

- Tobacco's **decreasing affordability** also contributes to lower consumption

---

💡 **Future Recommendations**

- Use machine learning to predict future smoking-related health outcomes

- Apply classification models to identify high-risk populations

- Create a dashboard for public health stakeholders

- Include socioeconomic factors in future analysis for deeper insights