How India Is Evolving
MTH208 Project Report

Anshika Rai, Manas Jain Kuniya, Shreya saha, Khushi

Introduction

We are **DataToDuniya**, a team of passionate data scientists who believe data is transforming lives and will continue to do so. We help companies identify their weaknesses and strengths and support them in achieving their goals. Over the last few years, as leaders in this industry, we have seen a growing need for everyone to have access to real, analyzed data. With the rise of social media, it has become increasingly difficult for people to discern what is true and what is false. In response, we are launching a new branch in our company dedicated to providing reliable, analyzed data to enable anyone worldwide to make informed, judicious decisions.

We are starting this series of analyses with '**How India Is Evolving**' covering topics such as crime rates, crimes against women, development metrics like state rankings across India, and the representation of women in Indian elections. We will also be exploring public reactions and trends within these areas.

After our successful funding round with CEO Dootika Vats, we are very happy and grateful to share our first report.

How our data analysis is helping the user?

We hope this report will help government bodies, policymakers, and the public to drive data-driven decisions that target areas in need of improvement, and focusing women's welfare, security.

Data

Our major data set are:

- Rankings of Indian States based on various socio-economic and geographic factors, (saved in file *allStateIndicatorsData*) a few of them are listed as below:

Parameters
Growth rate in GSDP-Total
Population Density (Per Sq. Km.)
Total Literacy Rate (%)
Unemployment Rate (Usual Status (adjusted)) (%)
Dropout Rate: Primary Level (%)
Household availing banking services
Safe Drinking Water (Tap/Handpump/Tubewell) (%age)

- We have collected crime rates, conviction rates, chargesheeting rates, investigation rate, crime rates of all states in recent years, to better understand the current scenario (saved in file *allStateCrimeData*)

- Crime composition dataset provides a comprehensive overview of reported crimes against women in India, encompassing various offenses such as harassment, domestic violence, rape, and human trafficking (saved in file *crimeAgainstWomenDistribution*)

- We have Crime rate (per 100,000 population) head-wise 2021 of diferent crimes, along with female population vs female police ratio, to understand the relation, (saved in file *combinedCrime* after cleaning from factorsAffectingCrime and policePopulation)

- This dataset tracks the percentage of women representatives in India's Lok Sabha and Rajya Sabha over the years, highlighting trends in female political representation. (saved in file *rsPercentage*, *stateWisePoliticalData*)

Obtaining Data

1. **Extracting Data Using APIs**

   - We have used free government APIs to download state rankings, crime data, police population statistics, and other relevant datasets.
   - The API requests for fetching this data are written in the file *differentIndicatorsData*.

2. **Web Scraping**

   - We used Web Scraping techniques as discussed in class. Crime rates, state-wise statistics, and electoral data were scraped from different websites over the internet.
   - Relevant files for scraping include *politicalRatioData* and *CrimeDataExtractor*.

3. **Data Cleaning**

   - After extracting the data, we performed data cleaning to ensure that the datasets are structured correctly for analysis.
   - The major issue was the numerics were char, so used different functions to convert them into numeric, also we deleted the irrelevant rows and columns.

4. **Saving Data**

   - The cleaned data is saved locally in `.Rdata` format for future analysis or reporting. in Data/ directory.

Is there any biases in the data?

The data used in this analysis is sourced directly from trusted and official sources, including government APIs and Wikipedia, which in turn draws from the national census data. Since these sources rely on comprehensive, population-wide surveys rather than selective sampling or other errors, they avoid common biases such as sampling or survivorship biases.

Questions we imposed

Here are a few questions laid down by us looking at the dataset:

- Are there states with high rates of specific crimes, so we can focus more on awareness campaigns or public safety programs there?

- Can we list the major socio-economic problems that the states are facing, so that government bodies and activists can be better informed?

- Which crime is most prevalent against women?

- Which states have a mismatch between high crime rates and low women police representation?

- In which states more women-helping bodies , NGOs, and government body should focus?

- Are these specific states consistently showing high rates and need new policies or interventions?

- Are there states with rapidly increasing crime rates that need urgent attention?

Visualizing the dataset

Using the first tab in the shiny app, we have figured out some key things about different states of india as listed below:

- **Andhra Pradesh**
  Good Aspects: Have a good gender ratio and enrolment of children in schools, showing a healthy mindset of parents of not forcing them to child labor.
  Challenges: Lagging in the industry GDP, which can also be seen by its highway length ranking despite its large size.

- **Arunachal Pradesh**
  Good Aspects: Good sex ratio and a strong GDP in industry.
  Challenges: High population growth, possibly due to migration from Bangladesh (data from 2011). This could be a concern for both the state and central government.

- **Assam**
  Good Aspects: Despite being somewhat unreachable from the central government, Assam has shown very good performance in reaching door-to-door services. Every household can avail banking services. It has managed to reduce per capita consumption of electricity despite the growing global demand for energy.
  Challenges: Still behind in certain areas of development, which may reflect slower industrialization and modernization.

- **Bihar**
  Good Aspects: High growth rate of Net State Domestic Product (NSDP).
  Challenges: One of the lowest male literacy rates.

- **Delhi**
  Good Aspects: Despite having a very bad gender ratio and high population density, Delhi has successfully ensured the development of its nearby rural areas. This can be seen by the number of rural telephone connections.
  Challenges: Very bad gender ratio and extremely high population density.

- **Gujarat**
  Good Aspects: Has one of the least upper primary dropout rates. Highest per capita electricity consumption, a sign of increasing income and industrial growth.
  Challenges: Some inequalities persist, but overall, Gujarat has shown positive indicators for growth.

- **Haryana**
  Good Aspects: Has one of the highest rail routes.
  Challenges: Very bad gender ratio.

- **Himachal Pradesh**
  Good Aspects: Known for its tourism economy.
  Challenges: Has the lowest ranking in household access to banking services, (which can affect the most in the upcomming world of digial payments.) Also struggles with the availability of safe drinking water.

- **Kerala**
  Good Aspects: Excellent gender ratio and the highest literacy rates.
  Challenges: None reported, Kerala performs well on many socio-economic indicators.

- **Madhya Pradesh**
  Good Aspects: Has one of the lowest infant mortality rates.
  Challenges: Challenges persist in other areas, though they are not specified in the data.

- **Maharashtra**
  Good Aspects: Highest length of state highways per square kilometer.
  Challenges: Second-highest maternal mortality rate, indicating a need for improvement in healthcare systems.

- **Manipur**
  Good Aspects: Has the highest ranking for households availing banking services.
  Challenges: One of the lowest primary-level enrolment rates, suggesting that children are possibly being pushed into work instead of going to school. Low growth in per capita NSDP further points to economic struggles.

- **Meghalaya**
  Good Aspects: High growth rate of population.
  Challenges: The high population growth, possibly due to migration from Bangladesh, is a significant concern for the state and central governments. Has the highest gender gap in literacy.

- **Nagaland**
  Good Aspects: Unique in that it had a negative population growth rate in the 2010s, unlike many other states.
  Challenges: Nagaland had the highest death rates in 2011.

- **Punjab**
  Challenges: Has the highest fertility rate (which is a concern for women). Lowest rank in the availability of safe drinking water.

- **Rajasthan**
  Challenges: Highest dropout rates in upper primary education. Low female literacy rates and high gender gap in literacy. This may point toward forced early marriage of girls in Rajasthan, with parents marrying off girls as soon as they complete the minimum schooling requirement.

- **Uttar Pradesh**
  Good Aspects: Highest population (which can be both an asset and a liability). Has low birth rates, which could look favorable for the future.
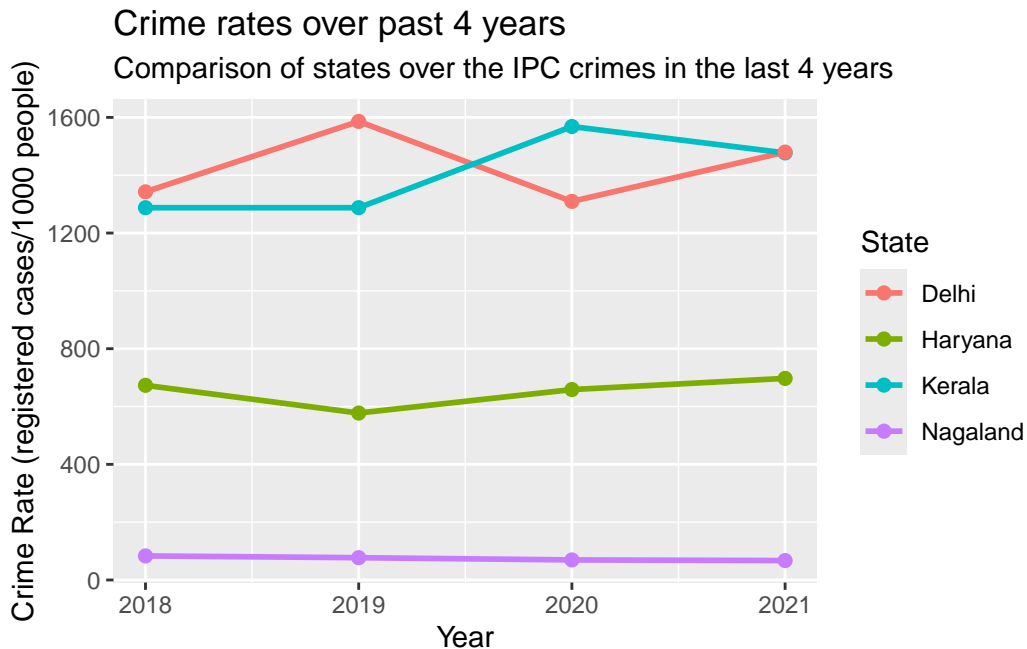 **Challenges:** Ranks low in safe drinking water facilities. Low per capita NSDP, which indicates inefficient use of resources for the large population.

- **West Bengal**
  Challenges: Highest fertility rates and very high population density. Low rankings in safe drinking water availability.

Trends in Crime Rates Across States

- The crime rates increases over years more or less in every states.
- In **Kerala and Delhi, crime rate increases drastically** over the years 2020 - 2021(on an average 1500 registered crime cases per 1000 people).
- All other states show less crime cases in the years 2018-2021. (can be taken in account of covid lockdown)
- Moreover, in **Nagaland, the crime rate is almost near 0**,i.e, the crime rate is much less than any other country in India.

Crime rates over past 4 years

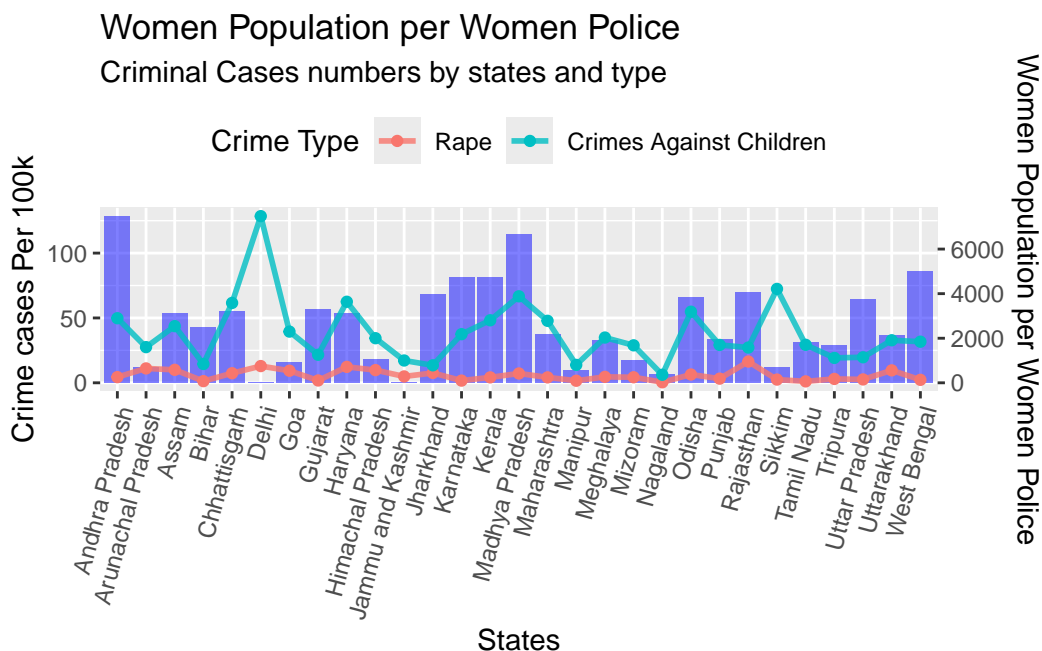Comparison of states over the IPC crimes in the last 4 years

- We can see the Women population / Women Police ratio is very high which can cause many problems like:

A shortage of women police officers can make it difficult for victims of crimes such as sexual assault, domestic violence, and harassment to report incidents. Victims may feel uncomfortable or intimidated when interacting with male officers, especially for sensitive issues.

With fewer women in law enforcement, there may be delays in investigating and addressing crimes, particularly those involving women and children. This can lead to increased vulnerability for women in society, as crimes may go unreported or unsolved.

Overburdened Workforce: A high women-to-police ratio leads to overburdening the few women officers available.

Women Population per Women Police

Criminal Cases numbers by states and type

- The relationship between women's participation in police forces and crime rates in Indian states is a complex one, shaped by a variety of social, cultural, and institutional factors. We were not a direct causal relationship is difficult

- Some states show that even in states with a high female police presence, there is still a higher level of criminal activity. This could be due to the stereotype that women can't handle both home and work, which may lead criminals to feel that no one can effectively stop them, resulting in a very difficult situation to manage.
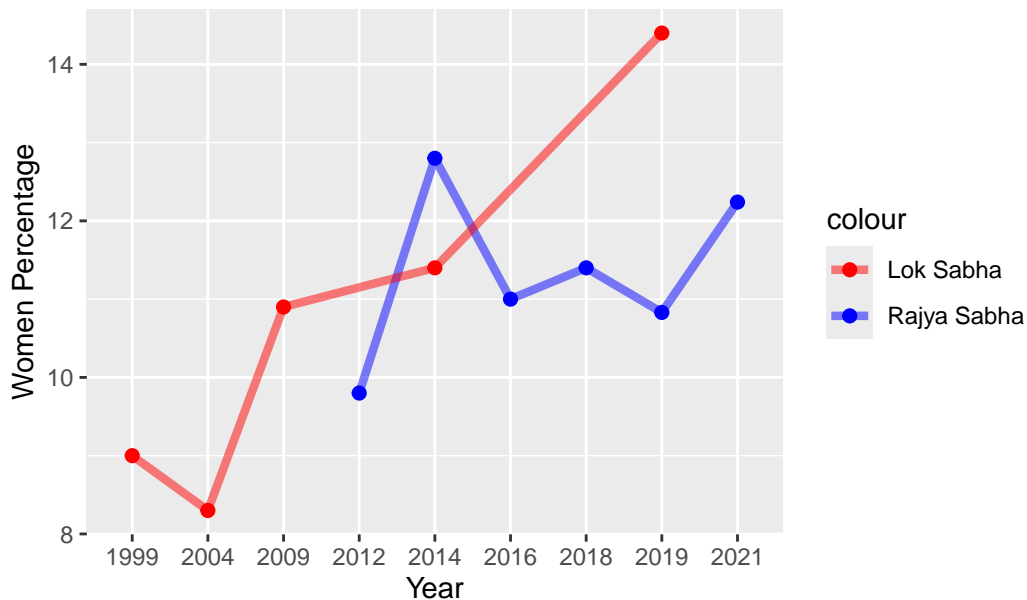
Crime against women

- Most Crime committed against women - **cruelty of family**,

- **Rajasthan has one of the highest occurrences of rape cases**, and *urgent action* is needed to address this issue.

Women and Elections

- The slope of the line graph for women's participation in the Lok Sabha elections is highest during the 2014-2019 period. From the data, it is evident that there was little to no participation of women in the Rajya Sabha elections between 1999 and 2012. Overall, the women's participation rate in the Lok Sabha elections is higher than in the Rajya Sabha.

- This trend suggests that **women are increasingly gaining public support**, as reflected in the higher participation in the Lok Sabha. However, **in the Rajya Sabha, where candidates are selected by state legislatures—often dominated by men**—it is more difficult for women to win seats.

- Overall, women's participation in both the Lok Sabha and Rajya Sabha remains very low, at around 15%, which is significantly less compared to the overall female population.

Percentage Women Elected (Lok Sabha and Rajya Sabh)

- Women Representation and Crime:

  We were unable to find a direct correlation between women's representation in government and the prevalence of criminal activities. For instance, in Rajasthan, despite a higher representation of women in governance, the rate of rape cases remains alarmingly

high. In Delhi, child kidnapping cases are notably high, while Punjab and Haryana face significant challenges related to drug trafficking and illegal drug production. These crimes appear to be more closely tied to socio-economic and regional factors than to the level of women's political participation.

In Rajasthan, the high incidence of rape cases suggests deep-rooted issues regarding the treatment and status of women in society, despite their increasing political involvement. In Punjab and Haryana, drug-related crimes are exacerbated by cross-border trafficking and local illegal production, creating a different set of challenges.

However, the growing participation of women in governance, especially in local and state politics, indicates a positive shift. Women are becoming key players in addressing these issues and working toward solutions,

Conclusion

Concluding this project looks closely at crimes against women in different states of India, showing important patterns and regional differences. By studying crime rates, police representation, and other factors, we provide useful information for making better policies. The findings emphasize the need for targeted interventions and improved resource allocation to address high-risk areas, support women's safety, and foster a more equitable society. The app serves as a decision-support tool for policymakers, empowering them to create effective, evidence-based strategies.

References

The websites we use to collect and scrape the data are following -

1. https://www.data.gov.in
2. https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_crime_rate
3. https://pib.gov.in/PressReleasePage.aspx?PRID=1809217
4. https://pib.gov.in/PressReleasePage.aspx?PRID=1805783