

Exploratory Data Analysis on Population, Sex-Ratio, Literacy of Indian Cities

Rohan Dey

CSE(Cybersecurity)

Meghnad Saha Institute of Technology

(Aff. To MAKAUT)

ABSTRACT:

The main goal of this paper is to study and analyze the demographic data of provided Indian states to find a pattern in the data regarding the population dynamics, literacy and sex ratio of the cities. The paper dives into 3 main sections, namely, sorting of data and calculation of sex ratio of the given data, calculation of literacy rate, effective literacy rate and literacy rate for every state alongside their graphical representation and finally our paper includes calculation and heatmap representation of the sex ratio of the top ten most populated states from the dataset.

This paper dives into the in depth analysis of literacy in every state and visualizes the performance of each state to create easily understandable data.

Keywords:

demographics, visualization, heatmap, literacy, EDA.

BODY:

The attached notebook does some Exploratory Data Analysis and visualization on the provided dataset. This analytical approach helps in understanding state-wise demographics and can inform further research or decision making. Some of the major patterns that can be observed from the above EDA and visualization are the following:

- The calculation of sex ratio highlights cases of gender disparity in some states. Lower sex ratios averaging around 75 can be seen in cities like Bhiwandi, Delhi Cantonment, Surat, Pithampur to name a few. This indicates varying demographic profiles across these cities.
- From the EDA done on the data we also get insights on the education from the literacy rates of the cities and states. The EDA gives us Maharashtra as the state having the maximum number of literate people. The scatter plot gives us a direct correlation between the number of graduates corresponding to effective literacy rate which is expected and the KDE plot gives us the density of literates in each state that directly gives us a visual to the literacy rate of each state.
- A dashboard is put together by merging all the plots done throughout the EDA process to give a synopsis of the trends in the data for population and literacy across the states.
- Finally, we filter out the top 10 states by their total population, and represent their sex ratio through a heatmap representation, with Maharashtra and Uttar Pradesh showing the maximum difference in male and female population while they are also the two states with the highest populations, followed by Andhra Pradesh, West Bengal and Gujrat but the difference from Arndhra and later states are quite negligible giving us the idea that there is a balance in the number of males and females.

CONCLUSION:

The EDA of demographic data across Indian cities reveals significant insights into sex ratios, literacy rates, and population dynamics. The analysis highlights gender disparity in several cities, such as Bhiwandi and Delhi Cantonment, where the sex ratio is notably low, suggesting a need for policies to address these imbalances. Literacy data shows Maharashtra leading in the number of literate individuals, with strong correlations between educational attainment and effective literacy rates. This is visualized through scatter and KDE plots, providing a clear view of educational achievements across states. The dashboard created consolidates these findings, offering a comprehensive view of population and literacy trends, making it easier to identify patterns and disparities. The heatmap analysis of the top ten

most populated states further illustrates differences in gender ratios, with Maharashtra and Uttar Pradesh exhibiting the largest discrepancies. These insights are crucial for policymakers and researchers to address gender imbalances and enhance educational initiatives. Overall, the EDA provides a valuable foundation for understanding demographic patterns and informing future research and policy development.