Analysis of Crime Against Women in India

By Tanaya Kavathekar & Madhuri Yadav

Data Science Department Columbian College of Arts and Science
George Washington University, Washington, DC, 20052

Objective

Defining the relationship between region, population, age, gender, illiteracy rate and increasing number of crimes against women in India

Introduction

In the era of big data and emerging machine learning algorithms, there is a growing interest in studying crime records data to take proactive instead of reactive measures. Crime is an integral part of society and there is a need to leverage advanced technologies to gain in-depth understanding about various factors. This project gives an overview of the crimes for 10 years (2001-2010). Hypothesis testing techniques such as ANOVA and Pearson correlation are performed on the data, in order to validate a relationship between the factors.

Over the years reported crimes have been increasing. There was a plateau observed in 3 consecutive years starting from 2001.But there was a **13% increase** 2006 from previous year.

Figure 1: Trend in crime cases reported, for a decade

Methods

Data collection: 1. Crime data: For this project, data is used from Kaggle website. Originally, data is gathered from National Crime Records Bureau(NCRB). 2. Census data: This data is present at the Registrar General and Census Commissioner of India. We have used demographic information from this source..

Data cleaning: We performed missing value and outlier detection of the data set. As the data set was already processed, we did not detect any major inconsistencies and inaccuracies were found.

Exploratory Data Analysis

We started with exploring the patterns and key trends in the data and found some interesting insights such as:

1. **Types of crimes:** After passing Immoral Traffic Prevention Act and Sati Prevention Act the crime rates has reduced in that decade. Even after spreading of awareness regarding dowry and making it illegal, no reduction is observed this category.

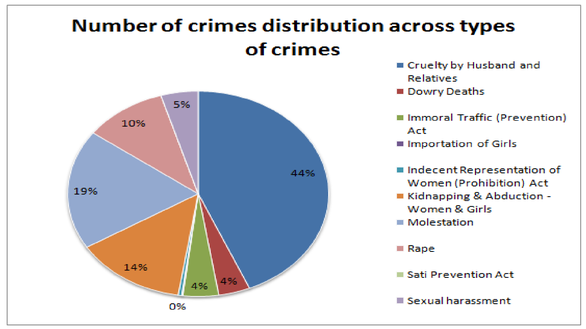

Figure 2: Crime distribution across types of crimes

2. **State wise distribution:** West Bengal and Andhra Pradesh has highest number of crimes whereas Lakshadweep has 0 crimes reported.
3. **Population:** Number of crimes is linearly increasing with population
4. **Illiteracy rate:** The average illiteracy rate in India for the year 2010 across states is 25%. Where Bihar (44%) and Rajasthan(43%) are the highest, and Lakshadweep(9%) and Kerala(10%) are the lowest
5. **Age and Gender :** Maximum crimes are observed between young adult group (18-30). And in terms of gender, as expected, men contribute to most of the crimes against women

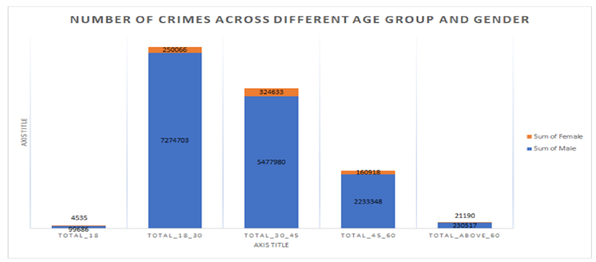

Figure 3: Crime distribution across age and gender

Results

Pearson Correlation: The Pearson product moment correlation coefficient is a measure of the strength of the linear relationship between two variables.
Figure 4 shows Pearson correlation for socio-economic factors.
Population (82%) and illiteracy rate (44%) is strongly positively correlated whereas employment rate is mildly correlated (10%).

Figure 4: Pearson Correlation heat map

Analysis Of Variance (ANOVA): This method is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. Anova is performed on types of crimes and location of crime with a significance level of 5%

Null hypothesis : Means between different types and region of crime are same.

Alternate hypothesis : Means between different types and region of crime are not same.

	Df	Sum	Sq Mean	Sq F value	Pr(>F)
Area_Name	31	576.4	18.593	6.152	<2e-16 ***
Residuals	288	870.4	3.022		
Group_Name	10	2.93E+11	2.93E+10	29.73	<2e-16 ***
Residuals	87	8.58E+10	9.86E+08		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Table 1: Anova results for the Region and the Type of the Crime

For both, types of crime and region we rejected null hypothesis with confidence interval of 95%.

Conclusion

We have successfully verified the factors such as region and different types of crime play a significant role in the increase of crime. Also socio-economic factors such as age, population, and illiteracy rate highly correlated with the increase in crime against women.

Recommendation

1. **Social awareness campaigns:** As we have identified age and gender affect crime rate, there is a need to drive social awareness campaigns to increase knowledge and empower and lift women rights.
2. **Strong laws and punishment:** Severe punishment for the heinous crimes and strong law should be passed to combat crimes and make country a safe place for women

Limitations

1. **Unreported crime:** Most of the crime goes unreported against women. Specially crimes such as sexual harassment and domestic violence. So it can be urged that the number is definitely more than what is reported.
2. **Unavailability of latest data:** Similar research needs to be performed on latest data to get more reliable results.

Future Work

The identified factors can be used to build a model to predict the number of crimes. This will be helpful to devise strategies for avoiding crimes.

References

Mukherjee, C., Rustagi, P., & Krishnaji, N. (2001). Crimes against Women in India. Economic and Political Weekly, 36(43), 1–3. Retrieved from https://www.jstor.org/stable/pdf/4411293.pdf?refreqid=excelsior:bc2f24f35e34baa66060f3a205e5977 ; Rai, D. (2019).Offences Against Women. Retrieved from https://blog.ipleaders.in/ offences-against-women/

Contact Details: madhuriyadav@gwu.edu, tanaya_10@gwu.edu