Media depiction of crime in UP

Kartik Gupta, Manali Patil, Deepankur Kansal

Problem definition

Background: The media motivates people to believe in false conspiracy theories and resort to violence. It sensationalises news and gives partisan coverage in its reports. In this way, it greatly aggravates the public attitude towards crime in this post-truth age. Along with this, a bias towards certain categories of crimes, the proportion of their depiction, and a correlation with the victim's background are observed.
Thus, there is a need to study and report these tendencies of the media.

Problem Statement: To perform a detailed quantitative and qualitative analysis of crime depiction in UP by media based on -
   a) Times of India news articles of Ghaziabad, Lucknow and Kanpur for the year 2019.
   b) Coverage of two crime events that had a nationwide reach and response.

Secondary Data Analysis: We will be using the **National Crime Records Bureau** data available on the NCRB website under Govt. Open Data License. It contains information regarding Metropolitan City - wise crimes under different categories. We will be analysing the four categories - Murder, Kidnapping, Crime against Women, Crimes against SC/ST. We shall study the distribution across subcategories and motives, age-wise and gender-wise statistics.

Following aspect was brought up after discussions with **Prof. Arvind Verma** and **Prof. Shankar Prawesh**

Qualitative Research: We will be examining the following two issues for this:
   a) **Vikas Dubey's Encounter Case** is a classic example of the media sensationalising a crime to manipulate people's minds. Most newspapers justified the encounter of the gangster without even questioning the police on the basic facts of the case. Times of India, the newspaper used for study, tries to keep its tone neutral, contrary to the Hindi newspapers like Danik Jagran.
   b) **Unnao Rape Case**, which despite its outrageous details didn't come into the public attention even during the critical 2019 elections. It highlights the bedrock of current public discourse - the male-centric approach. This agenda, coupled with political bias, deflected attention from the case and led the media to focus on other "greater" issues during this period. Women's safety was sidelined by other stories of "national interest".

Approach to the problem

We sampled 506 unique articles available in The Times Of India archive maintained online, representing the crime stories from Ghaziabad, Lucknow and Kanpur published from Jan 1 2019 to Dec 31 2019. We systematically annotated each story for crime categories, motive/intent tags or sub-category tags, number of children, adult male and adult female victims involved in each article. Instances containing multiple assignments within one field were also present. A sentiment analyser model is implemented to check for the outlook of the article, victim bias, accused bias, political bias, gender bias, religious bias and caste bias of these articles - which tests our hypothesis regarding the selective representation of cases and introduction of manipulative sentiment. This was then compared with crime statistics reported by the National Crime Records Bureau to find any possible patterns/ discrepancies. For the two case studies, multiple articles pertaining to the issues at hand were analysed, and a qualitative analysis was performed.
The findings are posited to be representatives of the topic in focus for the entire state.

Research Design

The topic of our study is Media depiction of crimes in UP, and to fulfil this purpose, we chose three major cities of UP (according to the availability of NCRB data): Ghaziabad, Kanpur and Lucknow.

Primary Data: For our Qualitative research,
   a) We selected a reputed national newspaper, 'The Times of India', to capture the idea of how the big names in the Indian media report crimes.
   b) We also collected 10+ articles referring to the case studies of our interest to gauge the sentiments and the tone depicted by them.

Secondary Data: For our Quantitative research, we explored the various types of crimes listed on the NCRB website, and selected four - murder, kidnapping, crime against women and crime against SC/ST.

Data collection:
   a) (Primary) The month-wise data of year 2019 was scraped from the ToI archives using a web-scraper. The collected articles formed ALL existing reports from the cities under consideration.
   b) (Primary) Case Study articles were searched on the internet by looking up different websites.
   c) (Secondary) NCRB data, available in the PDF format, was downloaded as per need.

Hypothesis:
   a) The media tends to prefer one case over the other based on the accused or victim's gender/ age/ caste/ religion, category of crime, political leaders' involvement, etc.
   b) The media tries to manipulate and sensationalise some specific cases over others.

We used both the Quantitative (Comparative Analysis) and Qualitative (Case Studies + Sentiment Analysis) methods to test this hypothesis.

Quantitative Research: The NCRB records (secondary data) were compared with reported statistics to capture the level of representation across categories and selectiveness across gender to visualise whether there is any misrepresentation of news articles and the difference in the distribution.

Qualitative Research:
   a) **Case Studies** - We analysed multiple articles to find how the entire story coverage by several media houses affected and manipulated readers' understanding of the event and how the content of the article influences or induces certain biases that are not inherent directly. Thus, we look at both "what" and "how" these articles have been written, which help us in finding how the media sensationalises specific cases.

   b) **Sentiment Analysis** - We rated the title as well as the article content from highly negative to highly positive that is -1 to +1 value based on word combinations, to evaluate the following traits:
   1. Tendency to grab reader attention by attaching a negative sentiment to the title as compared to body.
   2. Presence of positive and negative biases towards victims and accused, respectively.
   3. Presence of amplified victim/ accused bias in case of the Murder, Crime Against Women categories.
   4. Presence of caste bias through words like "dalit" and "SC/ST", and a negative sentiment.
   5. Presence of gender bias based on differentiating words like "boy"/"man" vs "girl"/ "woman"
   6. Presence of political bias based on words like "BJP","BSP","Samajwadi", and a positive sentiment.
   7. Presence of religion bias based on words like "Hindu" and "Muslim".
      The dataset and model is attached here.

Data Analysis and Results

Cleaning and Categorization: From 3600+ articles spanning over the 12 months, we first filtered out those about the required four crime categories. Then, for the 2 categories: Murder and Crime Against Women, we identified keywords associated with the motives and crime-type, respectively. Based on the following word cloud depictions and other relevant terms - we classified these 2 categories into subcategories.


1) Murder: Reasons - Property/Land Disputes/ Family Dispute/ Petty Quarrels/ Money Disputes/ Personal Vendetta/ Love Affairs/ Casteism/ Unknown

2) Crime Against Women: Types - Murder/ Dowry Deaths/ Suicide/ Kidnapping/ Acid Attack/ Cruelty by Husband/in-laws/ Rape only/ Assault on Women with Intent to Outrage her Modesty/ Cyber Crimes against Women/ Protection of Children from Sexual Offences Act.

For the remaining 2 categories, this sort of analysis was not possible due to the very low number of data points. For all 4 categories, the number and gender of victims were manually identified from the text.

**Quantitative Analysis**: **Comparative analysis as described in the Research Design section**

1. Gender-wise reporting number
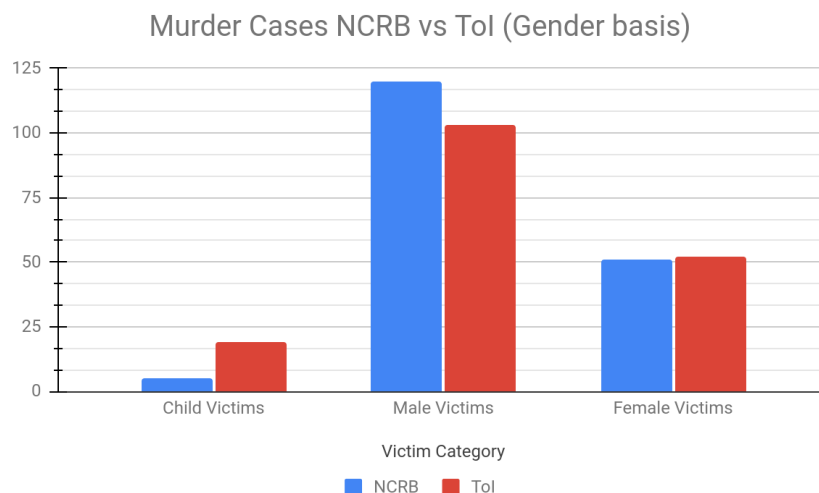
The higher number of reports than cases for child victims is due to follow-up/multiple reports of the same incident

Result - Near 100% reporting rate in case of female victims is observed, unlike male and child victims.

2. Subcategory-wise reporting rates

Crime against women (NCRB and ToI)
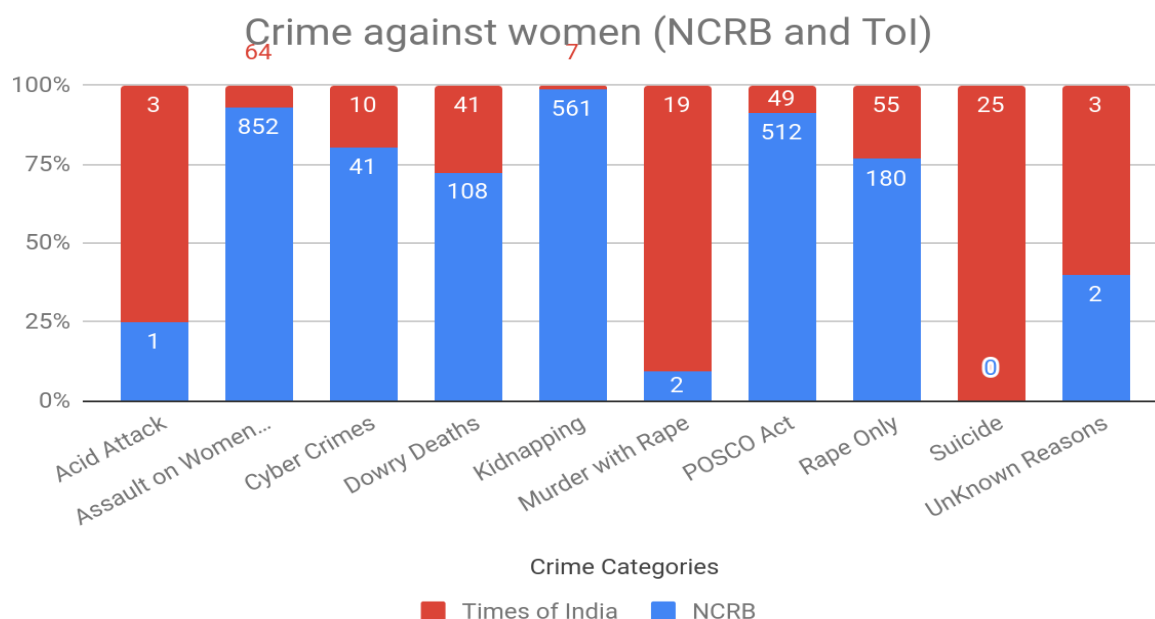
Result - Lack of reporting uniformity across types of crimes against women is revealed.
- Subcategories with less potential for sensationalising/ creativity in description such as Kidnapping, Cyber Crimes show less reporting rate.
- Subcategories worthy of serious attention such as Dowry Deaths and Rape don't show appropriate reporting rates, perhaps because they carry low receptivity from the audience due to social stigma.
- Subcategories with scope for extensive dramatic narration and higher appeal such as Acid Attack, Murder with Rape and suicide confirm a very high reporting rate.

3. Age-wise reporting number for kidnappings

Kidnapping (NCRB vs ToI)
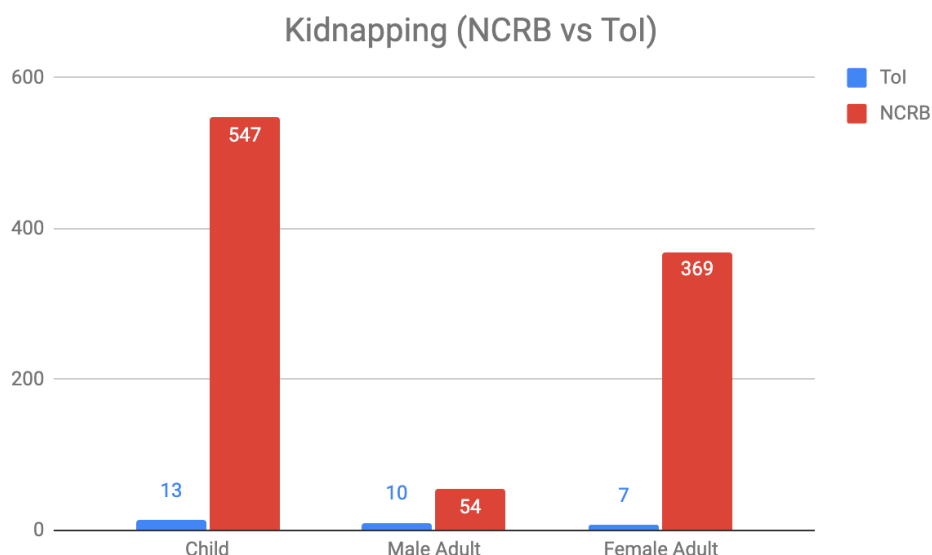
Kidnapping records the lowest proportion of reporting among all crime categories under consideration. Since there is little room for manipulative or critical writing here, this result is expected.

It is apparent that the highest reporting percentage is that of male adults (close to 18.5%) whereas that for child victims and female adults is nearly 2.37% - nearly 1/10 times that of male adults.
Result - This skewness highlights the extreme selectiveness in age/gender (20% vs 2%) is observed.

Qualitative Analysis: Case Studies

1) **Vikas Dubey Encounter Case** [*Citations:* 1 2 3 4]
The hunt for the "UP gangster", Mr Vikas Dubey, began after he, along with his accomplices, assaulted and killed eight policemen when they were coming to arrest him in response to a complaint against him at a nearby police station.

The entire incident shook the country, and people keenly followed up with the case and various media houses across the country exploited their sentiments. The local gangster was made into one of the most wanted gangsters of Uttar Pradesh due to the hype created by the media. Distortion and conspiracy theories were floating across newspapers, news channels creating a buildup of the gangster and portraying him to be extremely dangerous for the nation. Several media houses planted fake articles establishing his political connections with the current government and the negative implications of his arrest. The drama went on for eight days ending with his arrest and encounter by the UP Police and Special Task Force on the same day. Animations and short movies were displayed by Big Media Houses like Aaj Tak to gauge the audience, leaving aside other significant developments.

Most Hindi newspapers and media Houses in Delhi justified the encounter of the gangster without even questioning the police on the basic facts of the case. The Times of India, the newspaper used for our study, tries to keep its tone neutral by reporting the incidents keeping aside their inferences of the events. This was contrary to how Danik Jagran and several other prominent media houses reported the case.

There have been multiple instances in the past which depicts the media playing an instrumental role in sensationalising a particular case leaving behind other "not-so-important" issues. A few notable examples include Sushant Singh Rajput's Suicide Case and the Hathras Rape Case, where the media gave its undivided attention to prove someone or the other guilty of the crime.

2) **Unnao Rape** [*Citations*: 1 2 3]
In June 2017, a minor Dalit girl (17 years old), was kidnapped from her village of residence in Uttar Pradesh and raped by BJP MLA Kuldeep Singh Sengar, his brother Atul Singh, and others. She was found 17 days later, in a village nearly 100 km away from her hometown. This incident had shocked the entire nation.

The victim had accused the Sengar of rape in June 2017 itself. CM Yogi Adityanath's government, however, did not take action against the BJP MLA even after as long as a year had gone by. In the 2019 election season, the media entirely sidelined issues like this and focused on foreign policy, domestic militancy, aggressive nationalism and migration. No pressure was put on the ruling government by media houses for mismanagement of the case. Even during BJP's campaign, Sengar was still an MLA and a part of the BJP. The party continued to support him and he used his political power to repeatedly intimidate the victim.

When it comes to treating women and other (minority) groups as equal participants of democracy and politics, the media appears to skirt the issue. Associating politicism to grave societal issues like rape and murder highlights an unfortunate characteristic of Indian journalism - the absence of a 'gender lens'. The Unnao case eluded public attention until the news of the victim's father's death (in police custody) broke. It was only after this and her attempt at immolation, that an outrage happened due to reporting. The media's failure to portray the girl's perspective, her struggles encouraged the MLA to continue his corrupt actions.

This indifference towards women's issues and representation is facilitated by the media diverting attention within the community to political matters and a patriarchal viewpoint. This is greatly influenced by the gender imbalance in the workforce, which is especially pronounced in Hindi newspapers. In a 6-month survey conducted for 7 Hindi papers, of the 6,806 articles surveyed, only 11% were authored by women. And only 17% of the writers of 2084 articles studied were women - reference.

Qualitative Analysis: Sentiment Analysis

1. Tendency to grab reader attention by attaching a negative sentiment to the title as compared to body.
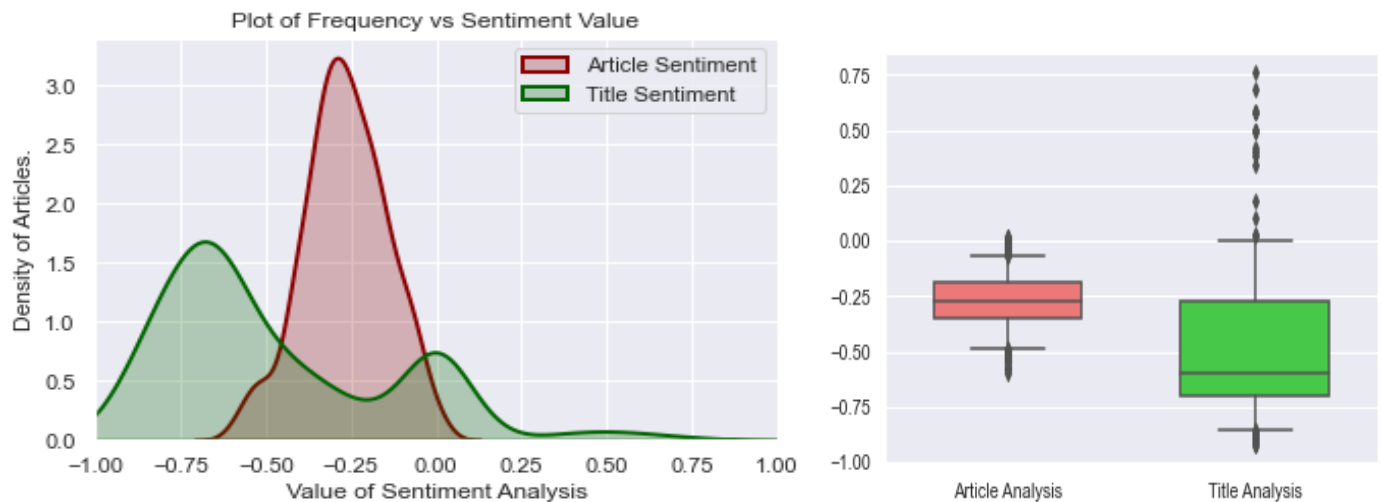
"Frequency Density of Sentiment" (left) and box plot for "Variation in sentiment" (right)

The title of news articles show a much larger negative sentiment on average than the article content. Also, most of the articles and the associated titles lie in the negative range. Considering the data points between 5th percentile to 95th percentile, all of them lie in the negative range of sentiment (see boxplot).
Result - Higher negative sentiment in the title than the body to attract attention is observed.

2. Presence of positive and negative biases towards victims and accused, respectively.
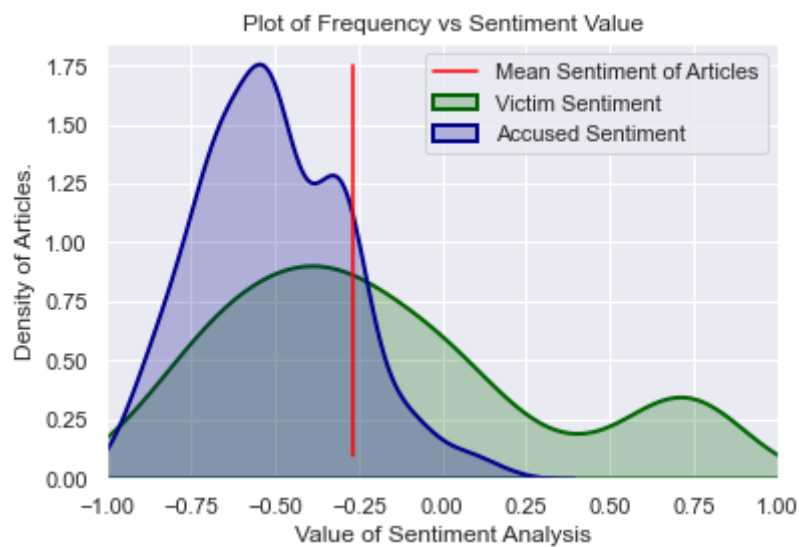

Distribution of the victim sentiment is more spread out and has a lot of values above the average value (denoted by the red line) with a mean of -0.18 (significantly higher than the article average). Hence, the articles are inducing a slight positive bias towards victims, and many induce a significant positive bias (+0.5 to +1) as shown by the distribution.

Distribution of the accused sentiment is more focused on the negative spectrum (peaks are present) and has a lot of values below the average value (denoted by the red line) with a mean of -0.50 (significantly lower than the article average). Hence, the articles are inducing a highly negative bias towards the accused, and induce a neutral or positive sentiment in a small minority of articles.

Result - A significant positive bias towards the victim and highly negative bias towards the accused.

3. Presence of amplified victim/ accused bias in case of the Murder, Crime Against Women categories.
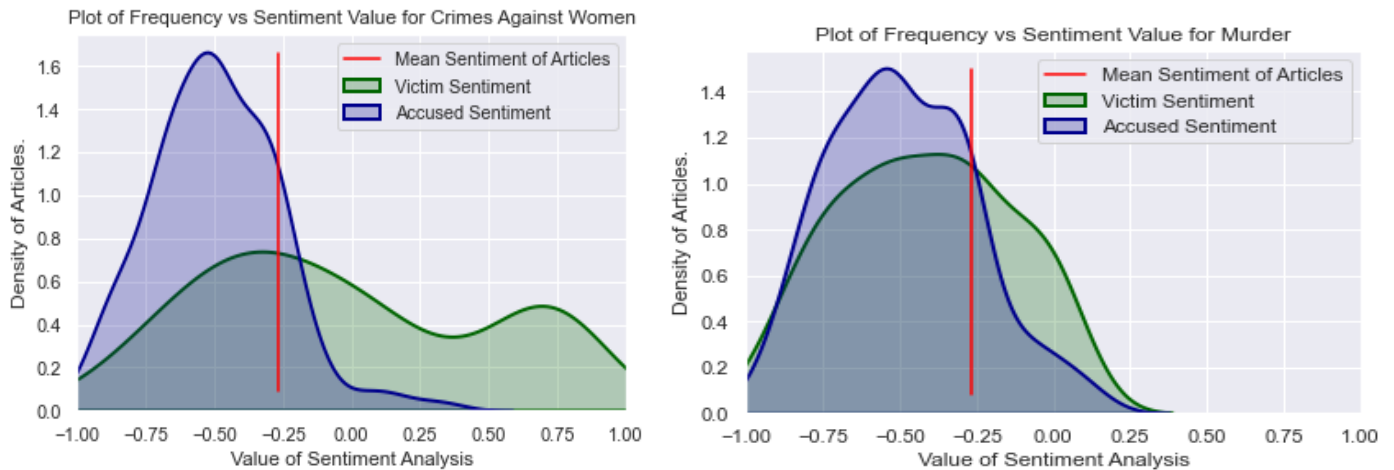

Crimes against Women category has a larger number of data points above the mean sentiment of the articles with a mean of -0.03 (victim) and -0.49 (accused).
Result - A high positive bias for victims as compared to the general scenario is observed as hypothesized.

Murder category has a larger number of data points below the mean sentiment of the articles with a mean of -0.41 (victim) and -0.49 (accused). We expect that this happened as murder when associated with victim conveys the emotion of "fear" and "sadness" much more prominently.
Result - A high negative bias for victims as compared to the general scenario is observed.

4, 5, 6, 7:   Presence of Caste, Gender, Political, and Religion bias.
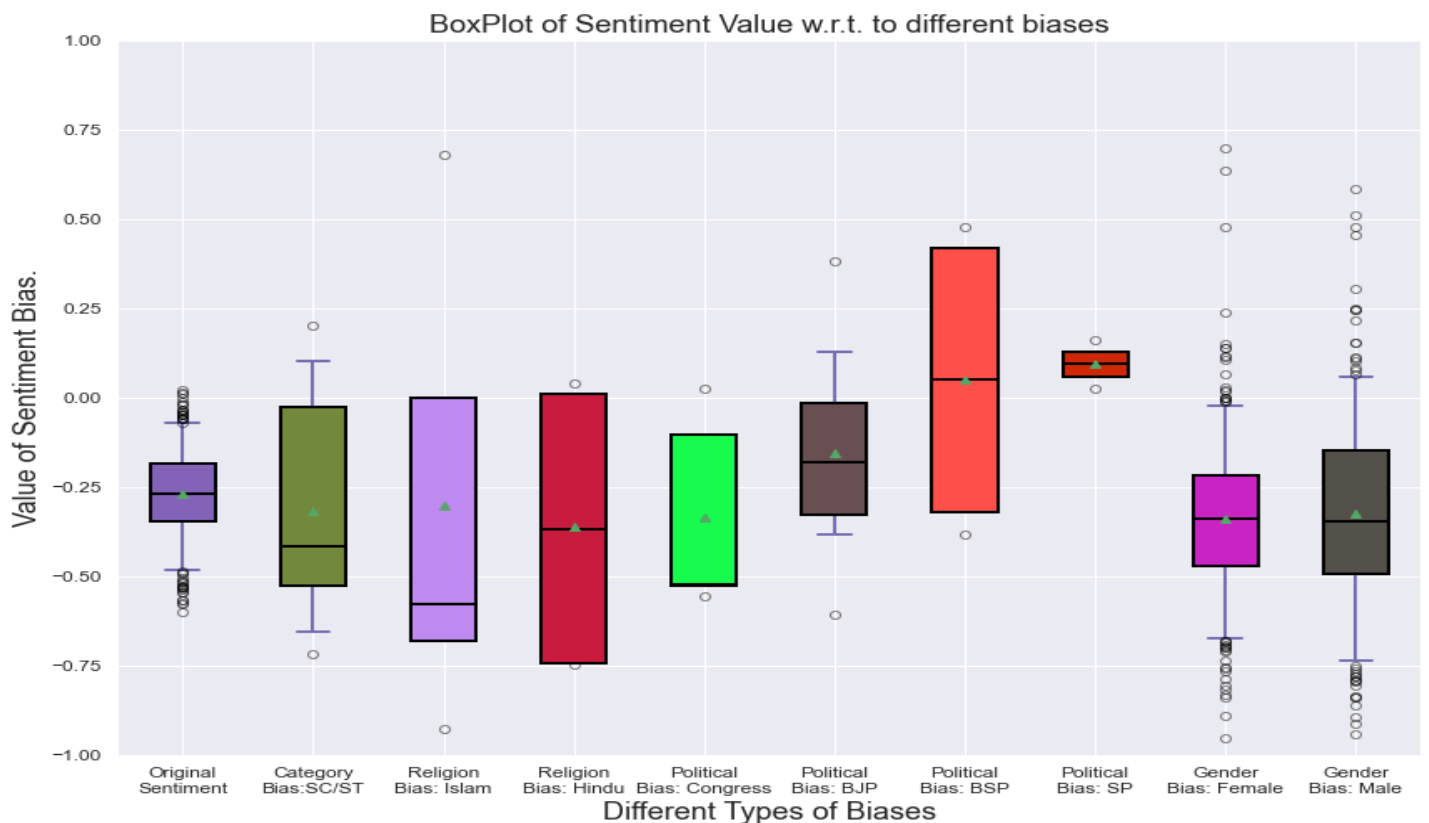
Caste Bias: The median sentiment is below the article average and and all the distribution from 0 to 75% percentile is in the negative spectrum. [Keywords: sc\st, dalit]
Result - A slightly negative bias is observed.

Gender Bias: The median sentiment of both the categories is approximately equal and the spectrum follows a very similar distribution. [Keywords:'girl', 'woman', 'female', 'women', 'wife', 'daughter', 'sister', 'mother'] [Keywords:''boy', 'man', 'male', 'men', 'husband', 'son', 'brother', 'father']
Result - No gender bias is observed.

Political Bias: Leaving Congress, the distribution of BJP, BSP and SP all induce a positive bias, varying slightly in case of BJP and most prominent in case of SP. For reasons, this has to be tested on more data points. Also, in Congress the sentiment is approximately the same as the original distribution.
[Keywords:bjp,bsp, congress, bahujan, samajwadi, bartia janta]
Result - A slight positive bias is observed.

Religion Bias: The percentile distribution shows a slight bias towards Islam as compared to Hindu religion. The median is significantly lower but the upper end of distribution follows a similar pattern.
[Keywords:Hindu, Islam,Muslim]
Result - On a general level it can be said that religion is inducing a negative effect in the articles.

Limitations and Caveats

1. In this endeavour, technological constraints limited the languages under study to only English.
2. Due to time constraints, analysis had to be performed only for reports from 3 cities.
3. Lack of clarity regarding definition and exclusiveness of categorisation parameters by NCRB.
4. (Possibly) Incomplete collection of total reported stories due to unavailability on ToI website.
5. Manually tagging and classifying data-points may have led to unavoidable human errors.

Conclusions and Recommendations

The media highlights specific categories of crimes more than others, as seen from the comparative analysis. Selectiveness in reporting is observed across ages, gender and type of crime. Case studies stress on the extent of sensationalization observed, and how political agenda influences reporting priorities. Sentiment analysis shows that the media induces a bias inherently which leads to swaying of opinions among readers.

Recommendation: Automating the category classification process can be explored, since the manual method is unfeasible for larger data sets. The model would be based on Machine Learning Classification Algorithms (Ex. LightGBM) where a test data set would check for its accuracy.

Team members and their contribution

Deepankur Kansal(10):
1. Writing Data Scraper for scraping data from TOI archives.
2. Data Cleaning: Category Classification
3. Qualitative Analysis:- Sentiment Analysis and Biases.

Kartik Gupta(10):
1. Data Cleaning - Filtering required articles from 3600+ stories, category classification
2. Qualitative Research - Vikas Dubey Case Study
3. Quantitative Analysis - Comparative analysis between the ToI and the NCRB records.

Manali Patil(10):
1. Data Cleaning - Filtering required articles from 3600+ stories, category classification
2. Preparatory Analysis - Word Cloud generation for detecting prominent sub-categories
3. Qualitative Research - Unnao Rape Case Study

Hypothesis Formulation and Report Compilation: All 3 members.