A Report on Visual Exploration of Sexual Violence in Nepal

A MINI-PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

COMP 482(Data Mining)

Submitted By:

Bishesh Kafle(09)

Anisha Lamsal(11)

Submitted To:

Dr. Rajani Chulyadyo

Department of Computer Science and Engineering

KATHMANDU UNIVERSITY

DHULIKHEL, NEPAL

23rd March, 2024

Acknowledgement

Abstract

This report offers a comprehensive visual analysis of sexual-related cases in Nepal, shedding light on the nature, location, timing, and sources of these incidents. Through meticulously crafted pie charts and scatter plots, we examine various aspects of these cases, aiming to provide a deeper understanding of this critical issue. The report delves into the types of sexual attacks that have occurred, categorizing them to reveal patterns and trends. By mapping the geographical distribution of these events, we uncover insights into where they have taken place, offering valuable information for policymakers, law enforcement, and advocacy groups. In addition, we dissect the times at which they occurred. This temporal analysis can help stakeholders allocate resources more effectively and implement preventive measures during high-risk periods. Furthermore, the report investigates the sources of data on sexual-related cases, including news outlets and other organizations that collect and store this information.

Overall, this report serves as a visual resource to facilitate a data-driven approach to addressing sexual-related cases in Nepal.

Table of Content

INTRODUCTION

Landscape of Sexual Violence in Nepal

Sexual violence, a horrific spectrum of acts including rape, assault, and harassment, casts a long shadow across Nepal. Survivors of these crimes face not only the immediate trauma of the assault but also the long-term physical, psychological, and social consequences. In a nation striving to build a more just and equitable society, understanding the true nature of sexual violence is necessary.

This project tackles this critical issue by delving into a dataset provided by Fightback Nepal, a prominent organization at the forefront of the fight against sexual violence. By continuously examining this data, our motive is to uncover the true extent of sexual violence in Nepal.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a technique which allows us to analyze and investigate datasets, summarizing their main characteristics using data visualization methods. EDA typically involves:

- Summarizing the Data: This involves calculating basic statistics like means, medians, and frequencies for each variable. It gives you a general sense of the data's central tendencies and spread.
- Data Visualization: EDA relies heavily on creating charts and graphs like bubble plot, scatterplots, and boxplots. These visuals help identify patterns, trends, and potential outliers within the data.
- Identifying Relationships: EDA explores how different variables in the dataset might be related to each other. This can involve techniques like correlation analysis to see if changes in one variable are associated with changes in another.
- Understanding the Data's Structure: EDA involves getting a sense of the data's overall structure, including missing values, data types, and potential inconsistencies. This ensures the data is prepared for further analysis.

Through techniques like descriptive statistics and visualizations, the EDA will uncover patterns and trends within the data.

Streamlit Dashboard

Streamlit is a Python library specifically designed to streamline the creation of web applications. It excels at transforming Python code, often used for data analysis and visualization with libraries like pandas and matplotlib, into interactive web dashboards. The core functionality lies in its collection of functions for building dashboards. These dashboards can display text, charts, and even incorporate interactive elements like sliders and buttons. This creates a powerful tool for sharing data insights with a broader audience, even those without programming experience. Users can engage with the data visually, filter information based on specific criteria, and gain a comprehensive understanding of the information presented. Streamlit effectively bridges the gap between data analysis and clear communication of the findings.

To translate the insights from the data analysis into an accessible format, a user-friendly dashboard was built using Streamlit. This dashboard will allow users to explore the data visually, filter information based on specific criteria, and gain a comprehensive understanding of the data.

Methodology

1. Data Collection:

Dataset for this project was retrieved from multiple sources, including Nepal Police, INSECOnline, Nepal Monitor, The Himalayan Times, Onlinekhabar, etc. The dataset covered a 10-year period from 2013 to 2022. Talking about the dataset, the dataset consisted of 2380 rows and the following attributes:

- Source of data
- Perpetrator's Age
- Victim's Age
- Date of Incident
- Location of Incident
- Time of Incident
- Attack Category
- Consequences of Attack
- Ethnicity of Perpetrator
- Types of Attack

2. Data Cleaning and Preparation:

The following steps were conducted to clean and prepare the data for exploratory data analysis:

a. Handled missing values using removal or imputation techniques.
b. Corrected inaccurate data entries where possible.
c. Consolidated and integrated data from multiple sources.
d. Normalized age data into 10-year age groups for visual analysis.
e. Converted 'Date of Incident' column to a datetime format for temporal analysis.
f. Created separate csv files for each particular analysis for better file handling.

3. Data Analysis

In order to gain insights from the collected data, we utilized Plotly Express, a powerful data visualization library, to create informative charts and graphs. Pie charts were employed to explore the distribution of attack categories. We further delved into the age demographics of both perpetrators

and victims using pie charts. To understand the temporal patterns of these incidents, a pie chart depicting the time of day attacks occurred was generated. Additionally, the ethnicity of perpetrators was visualized through a pie chart. Bar charts were used to analyze the distribution of reported incidents across various news sources, providing a perspective on media coverage. Another bar chart served to illustrate the occurrence of consequences, such as arrests or convictions, within different districts. For a temporal analysis of news source coverage, a line plot was constructed. Finally, to gain a spatial understanding of the incidents, a scatter plot map was created, pinpointing the geographical distribution of reported sexual violence. This comprehensive set of visualizations effectively communicated the key trends and patterns within the data.

4. Statistical Analysis

Firstly, we calculated descriptive statistics to summarize the numerical data. This could include measures like average age, total number of incidents reported by each news source, or the frequency of attacks occurring in specific time slots. Descriptive statistics provide a concise snapshot of the central tendencies and spread of the numerical values. Secondly, we conducted frequency analysis to examine the categorical variables. Categorical variables classify data into groups, such as perpetrator ethnicity, attack category, or district where the incident occurred. Frequency analysis calculates how often each category appears within the data set. This helps us understand the prevalence of different characteristics and identify any dominant patterns within the categorical data.

By combining descriptive statistics for numerical data and frequency analysis for categorical data, we were able to paint a more complete picture of the sexual violence related incidents, providing a solid foundation for further analysis and exploration of the underlying factors.

5. Interactive Visualizations

Plotly is employed extensively to generate various charts and graphs. These charts are then embedded into the Streamlit app. This allows users to interact with the visualizations and gain a deeper understanding of the data. For instance, users can hover over pie chart slices to view specific values or zoom in on scatter plots to focus on particular regions.

6. Textual Explanations

Streamlit's text formatting capabilities are utilized to incorporate clear and concise textual explanations alongside the visualizations. These explanations provide context, highlight key findings, and guide users through the data exploration process.

7. Dashboard Development and Deployment

The code structures the dashboard with informative titles and headers, and strategically positions various elements to create a visually appealing and user-friendly interface. Expanders are incorporated to house additional details or in-depth analysis, allowing users to control the amount of information displayed at once. Since Streamlit uses a static page, the dashboard is deployed using Github.

8. Ethical Considerations

This process was also guided by a strong commitment to ethical data practices. We strictly adhered to data privacy guidelines to safeguard the confidentiality of individuals involved. Sensitive data points were anonymized to prevent identification. Throughout the analysis, we ensured responsible and unbiased interpretation of the findings. This means we avoided making generalizations or drawing conclusions that could be discriminatory or misleading. Our focus remained on presenting the data accurately and objectively, allowing the insights to speak for themselves. Similarly, a login system was implemented for confidentiality purposes.
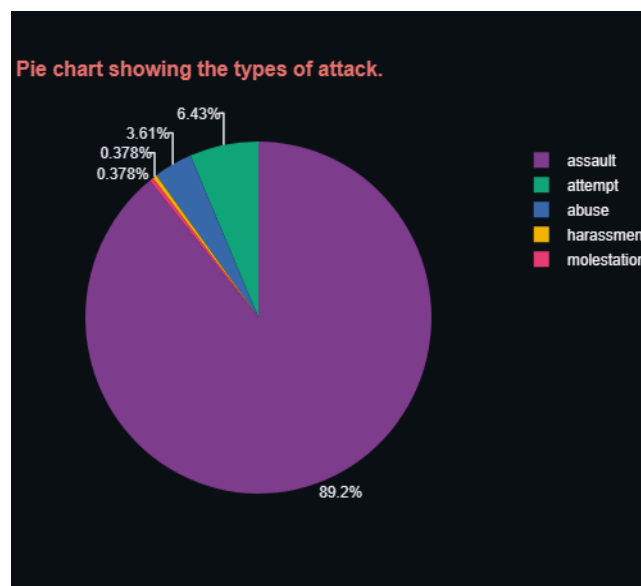
FINDINGS

Pie-charts

i. Pie-chart showing attack categories:
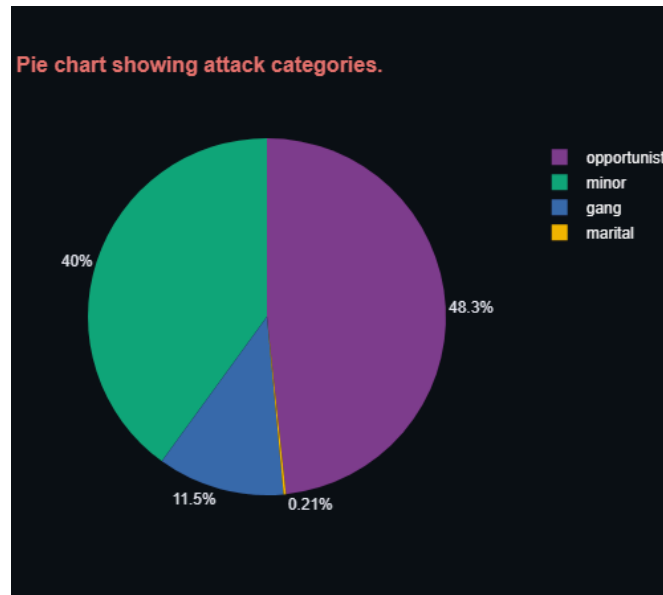
The pie chart shows the frequency of reported attacks by category. Opportunist attacks are the most common, making up nearly half (48.3%) of all attacks. Minor attacks are the second most common (40%), followed by gang attacks (11.5%) and marital attacks (0.21%).

ii. Pie-chart showing the types of attack:

Analyzing the above pie-chart,we observe that 'Assault' is the most common type of attack, representing 89.2% of the total. Other attack types are much less frequent. The remaining percentage represents all other reported types of attacks. These include abuse, harassment, molestation, and attempted attack, each constituting a very small percentage (between 0.38% and 6.43%) of the total attacks.

iii. Pie-chart showing perpetrator age groups:
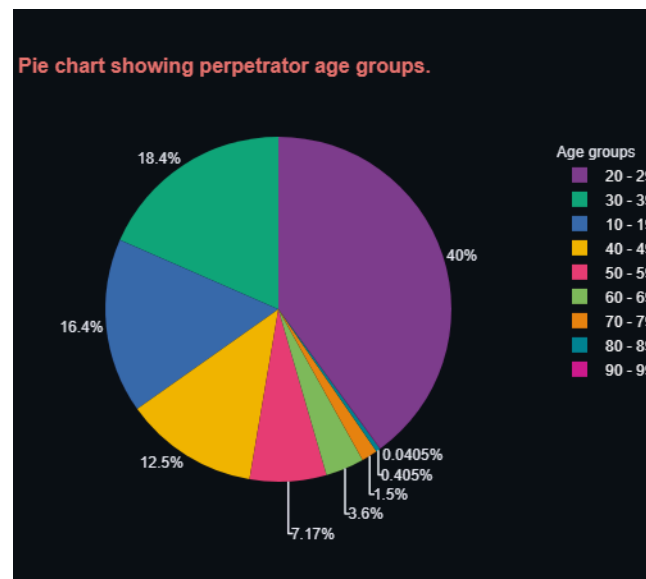

The largest percentage of perpetrators fall in the 20-29 age group, accounting for 40% of total perpetrators.The second largest group is the 30-39 age group, followed by the 10-19 age group at 18.4% and 16.4% respectively. Perpetrators aged 60 or over are much less frequent, constituting a very small percentage of the total perpetrators.

iv. Pie-chart showing victim age group:
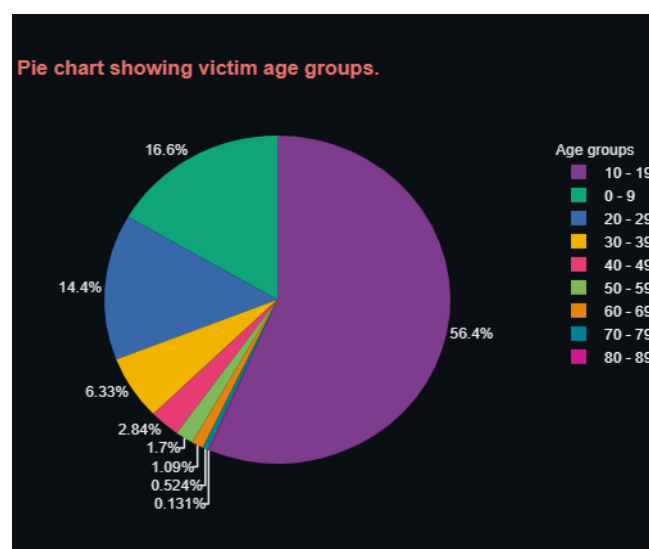
The majority of victims of social harassment are teenagers(10-19) age group (56.4%). About 16.6% of victims were children below the age 9. Then the percentage of victims followed by a steady decline with increase in age.

v. Pie-chart showing the time of attack

Pie chart showing the time of attack.
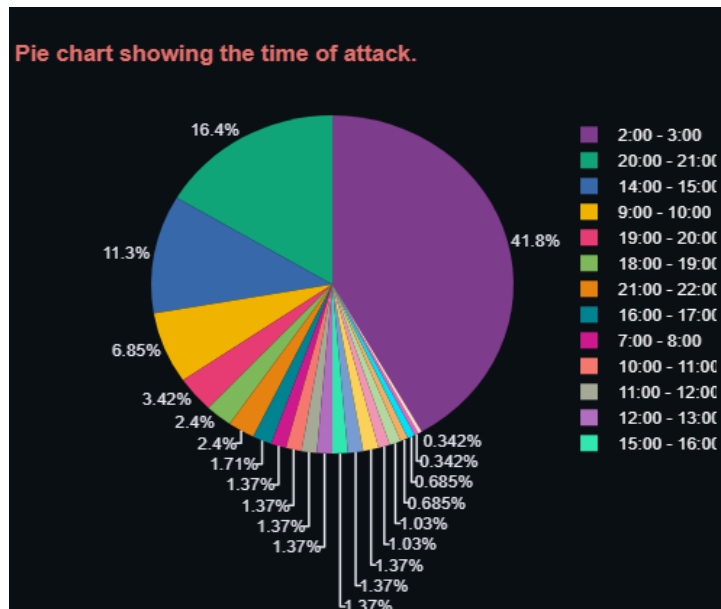
This data reveals a concerning trend: early morning seems the highest concentration of reported social harassment and rape incidents, with 41.8% occurring between 2:00 AM and 3:00 AM. Conversely, reports are least frequent during the evening and day time.

vi. Pie-chart showing ethnicity of perpetrator

Pie chart showing ethnicity of perpetrator.

The largest percentage of perpetrators identified by ethnicity are AVS, making up 25.1% of the total. MDS perpetrators are the second-most identified racial group, at 16.8%.

Sources

i. Bar plot of Stories by Sources

The plot above clearly illustrates the primary data sources for reports on sexual violence in Nepal over time. Unquestionably, Nepal Police stands out as the most prominent contributor, providing a significantly higher number of reports compared to other sources.

Interestingly, INSECOnline, a human rights-focused news portal in Nepal, also holds a notable number of records related to rape incidents. On the other hand, while Nepal Monitor has a respectable collection of records in general, it falls short of its potential in addressing this critical issue. Given its role as a monitoring website dedicated to incidents in Nepal, one would expect it to lead the way in reporting such sensitive incidents.Among all the top news sources in Nepal, only three—'The Himalayan Times,' 'Ratopati,' and 'Onlinekhabar'—demonstrate a strong commitment to covering stories of sexual incidents in the country.

In summary, Nepal Police rightfully takes the lead in covering stories related to sexual violence in Nepal, while specialized sources like 'INSECOnline' and 'Nepal Monitor' are significantly ahead in reporting various incidents. This divergence might be attributed to the fact that general news sources have a broader focus on all aspects of national events, whereas targeted news sources concentrate solely on human rights and incidents in Nepal.

ii. Line plot of Sources over time
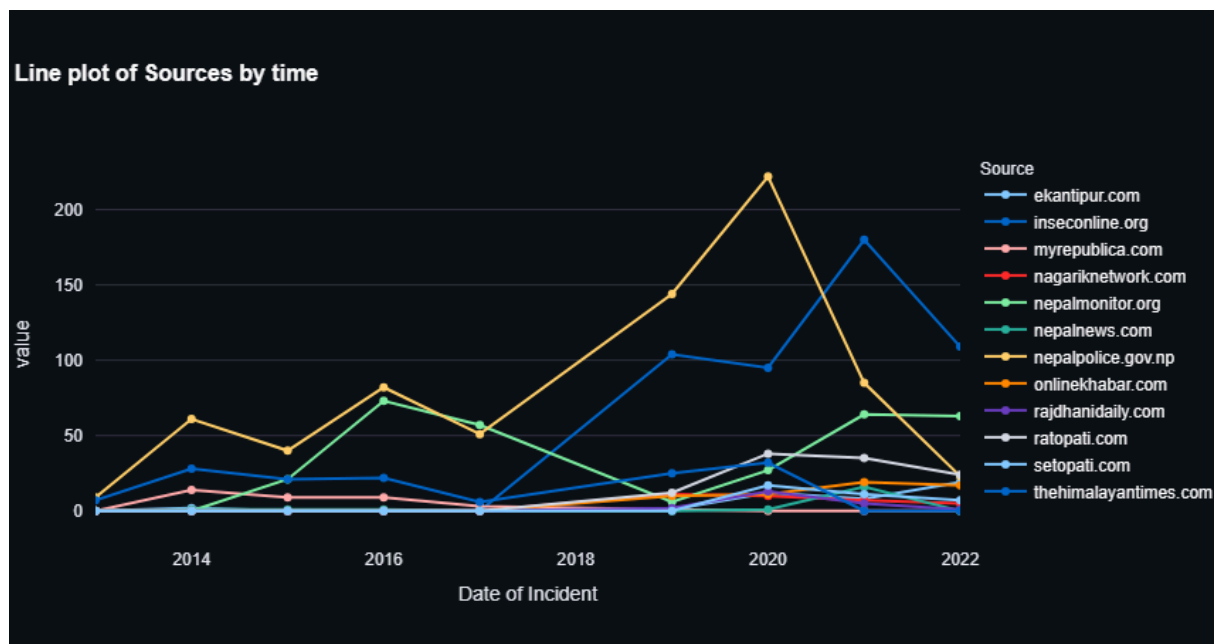
The line plot represents the coverage of news sources related to sexual assaults over a span of ten years, from 2013 to 2022, with time (years) on the x-axis and the number of records covered by various news sources on the y-axis. The plot provides valuable insights into the trends and patterns in media reporting on sexual assault incidents in Nepal.

At the outset, it is evident that the data reveals distinct trends for different news sources. The Nepal Police exhibits a notable peak in coverage in the year 2020, suggesting that this was a year when sexual assault incidents were prominently covered and reported by this source. However, this peak is followed by a consistent decrease in coverage in the subsequent years, indicating a potential decline in the frequency of reported incidents or a shift in the focus of the Nepal Police's reporting.

Another significant contributor to the coverage of sexual assault incidents is INSECOnline. The data highlights a rising trend in coverage from the base year of 2013, reaching its peak around 2021. This indicates that INSECOnline's reporting gained momentum over the years, culminating in a substantial amount of coverage in 2021. However, a slight decrease in coverage is observed in the following year (2022), suggesting a potential change in reporting strategies or a fluctuation in the incidence of sexual assault cases.

In contrast, Nepal Monitor starts off with relatively passive coverage until 2019. However, the data illustrates an upward trend in coverage from 2019 onwards. This upward trajectory indicates an increased emphasis on reporting sexual assault incidents by Nepal Monitor in recent years, possibly due to shifting societal awareness or changes in their reporting methodologies.

The analysis also identifies other news sources that appear to maintain a more consistent or passive level of coverage throughout the ten-year period. These sources might indicate a stable approach to reporting sexual assault incidents or might reflect the lack of significant changes in their reporting strategies over the years.

Overall, the line plot demonstrates the dynamic nature of media reporting on sexual assault incidents in Nepal. Peaks and troughs in coverage from different news sources reflect shifts in priorities, societal awareness, and reporting methodologies. The analysis highlights the importance of considering multiple news sources to gain a comprehensive understanding of trends in sexual assault reporting and to identify potential changes in the prevalence of such incidents over time. Further exploration of contextual factors and correlations with real-world events could provide deeper insights into the patterns observed in the line plot.

Consequences of the attacks

Among all the districts, we can analyze the trend and pattern which districts were highly prone to which consequences.

i. Bar plot of Consequences of attack
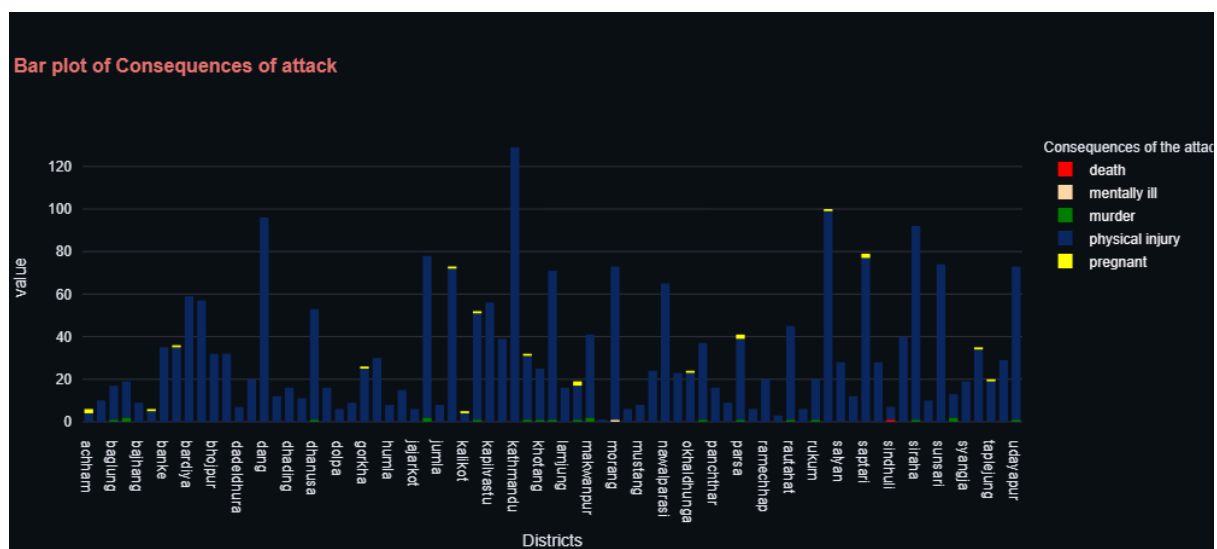
From the bar graph above representing the consequences,we observe that the highest rate of occurrence is attributed to "physical injury". This indicates that survivors most commonly experience physical harm as a consequence of the crime,underlining the immediate and tangible impact on their well-being.In Kathmandu,the prevalence of consequences resulting from the data set is highest, with a notable emphasis on "Physical Injury"(almost no other consequences).

As of the current data, the districts of Manang and Rukum exhibit the lowest incidence of consequences arising from rape compared to the others.

Survivors often experience significant psychological and emotional trauma. Common psychological consequences include mental illness,anxiety, depression, post-traumatic stress disorder (PTSD), feelings of shame, guilt, fear, and self-blame. The emotional impact can be long-lasting and may require therapy and counseling for healing.And for this,Morang stands out as a district warranting attention. Similarly , we can observe other consequences in districts from the graph.

ii. Line plot of district over time
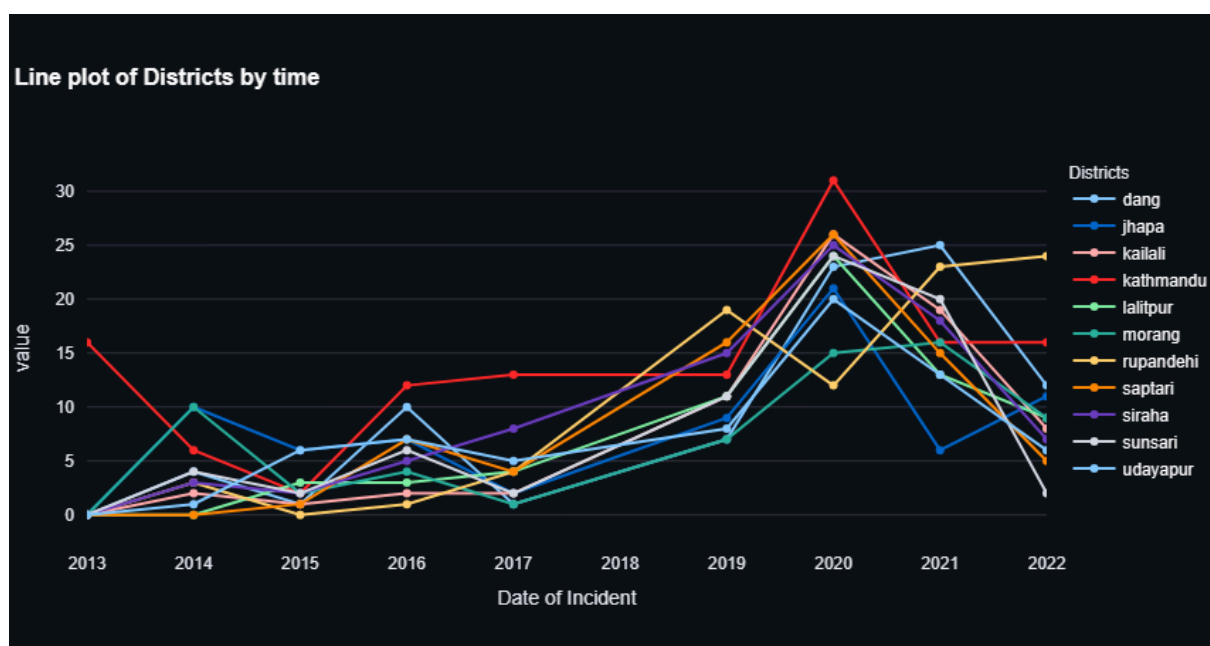
The line plot represents the distribution of incidents across different districts over time.It shows how the frequency of incidents has evolved over time.It allows us to identify periods of increased or decreased incident rates,highlighting potential trends and patterns.

In the vast tapestry of incident data,Kathmandu emerges as an undeniable focal point,adorned with the highest number of incidents,Kathmandu's data commands attention with its significant and imposing presence.Kathmandu's prominence serves as a potential call to action,urging us to channel our energy towards creating a safer,healthier and more secure environment for all.

Before the year 2021, Dang's incident data was unpredictable,with the number of incidents going up and down without a clear pattern.The number of incidents in Dang increased significantly,reaching the highest point ever recorded..

The number of incidents in Rupandehi was going up rapidlyBut in 2020, there was a decrease in incidents,providing some relief.However after that the incidents started rising again.

The data from Silyan creates an intriguing pattern-despite having fewer cases compared to districts like Kathmandu, the incident counts in Silyan remain steady or show an upward trend.

In conclusion,the data analysis reveals that Kathmandu has consistently high cases of incidents,which is a cause for concern.The steady trend of incidents indicates a persistent issue that demands urgent attention and effective interventions to ensure the safety.

Time of Attack

Among all the time durations, we can observe the time patterns at which the events are likely occurring.

i. Time by District

Here, for e.g. time = 3 denotes the time duration from 3 to 4
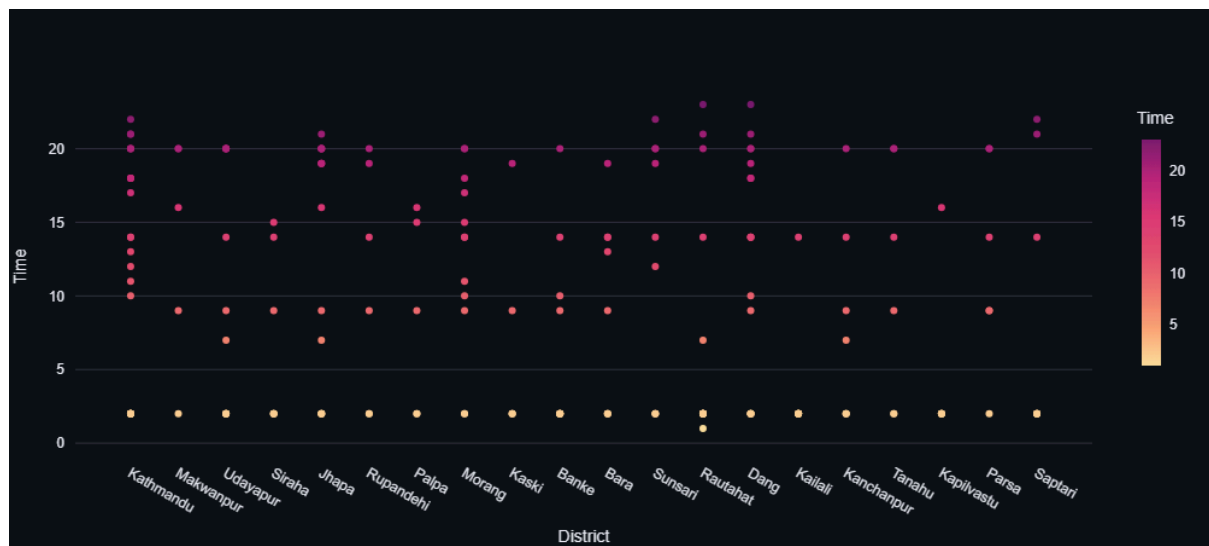
In our examination of incident data across various districts, a conspicuous pattern emerges, shedding light on the times when most incidents tend to occur. Two key timeframes consistently stand out as being particularly noteworthy in terms of incident frequency.

2 AM - 3 AM: In almost every district under scrutiny, the period between 2 AM and 3 AM emerges as the peak time for incidents. This late-night to early-morning hour appears to be associated with a heightened risk of various events. The reasons behind this phenomenon could be multifaceted, including reduced visibility, lower presence of individuals, and potentially altered social dynamics.

1 PM - 2 PM:Another time of increased incident occurrence across the surveyed districts is between 1 PM and 2 PM, notably in contrast to the early morning lull around 5 AM - 6 AM. This suggests that, for some reason, the early afternoon presents its own set of challenges or risk factors that contribute to incidents. Perhaps factors related to daily routines or social interactions play a role during this timeframe.

However, it is important to note that there is significant variation when we delve into district-specific data. In the case of Kathmandu, for instance, a different pattern emerges. Incidents in Kathmandu appear to peak between 10 AM and 3 PM, with the highest concentration of incidents occurring from 10 PM to 11 PM.

Conclusion

In summation, our analysis reveals compelling insights into the landscape of sexual-related cases in Nepal. Terai emerges as the epicenter of these incidents, registering the highest number of cases, while sexual assault remains the predominant form of abuse reported. Delving into the temporal dimension, patterns emerge, with 02:00 am appearing as a peak time for perpetrators, whereas particularly in the Kathmandu valley it is during the afternoon. The repercussions of these incidents are substantial, with physical injuries and a notable number of pregnancy cases being prevalent consequences.

To combat this troubling trend, we propose several proactive measures. Firstly, the allocation of security personnel during these active time periods could enhance public safety. Secondly, organizing self-defense training programs in areas with high incidents of physical injuries can empower potential victims to protect themselves. Finally, recognizing the significance of safe abortion services, especially in regions where pregnancy is a pronounced consequence, can provide essential support and care to those affected.

In conclusion, our findings underscore the urgency of addressing sexual-related cases in Nepal through a multifaceted approach that combines preventative measures, community support, and timely interventions to protect the well-being of individuals and the broader society.