**INTRODUCTION**

The issue of sexual violence and its impact on individuals and society has gained significant attention worldwide. As part of ongoing efforts to address this pressing societal concern, the collection and analysis of data related to rape incidents play a crucial role in informing policies, interventions, and advocacy efforts. This technical report aims to provide a comprehensive overview of a rape dataset with a case study of city of Los Angeles, California, United State of America between the years 2020 to 2024, exempting the year 2024 due to its incompleteness, focusing on its aims, objectives, description, and stakeholders involved.

Aim

The primary aim of this technical report is to provide a comprehensive analysis and interpret the rape dataset. This analysis is intended to shed light on the patterns, characteristics and trends of rape incidents that may exist within the data. It identifies factors contributing to rape incidents, such as time of day, relationship with the perpetrator, and reporting delays. The analysis explores the implications of the findings in a manner that is useful for stakeholders. It also assesses the effectiveness of law enforcement responses and support services in addressing rape incidents and supporting survivors.

By doing so, we hope to contribute to a better understanding of the factors surrounding these incidents, which could potentially aid in the development of more effective prevention strategies. To achieve this aim, the report sets forth the following objectives:

Objectives

* What is the relationship between most victims and their perpetrators, and how does this relationship influence the time taken to report the crime?
* Does this crime occur more frequently on weekends or weekdays, and is it more common during the day or at night?
* What is the yearly percentage of rape crimes being solved? is this percentage higher than the number of pending cases?
* What factors influence the resolution of the case?

Description of the Rape Dataset

The dataset under review contains information regarding reported incidents of rape within the city of Los Angeles, California, United State of America, spanning from 2020 to 2024. The dataset contains 5660 rows and 23 columns which includes data on various aspects of each incident, such as the date, time, and location of the incident, details about the victim and perpetrator, the relationship between the victim and perpetrator, and any subsequent legal proceedings or outcomes. The dataset may also include additional contextual information, such as reporting delays. The dataset represents a valuable resource for understanding the dynamics of rape incidents. It is important to note that the dataset only includes reported incidents, and the actual number of incidents may be higher due to underreporting.

1.4 Stakeholders

The insights derived from this analysis can be used by various stakeholders in the following ways:

* **Law Enforcement Agencies**: The patterns and trends identified in the data can help law enforcement agencies to understand the prevalence and distribution of rape incidents across different regions and times. This can aid in resource allocation, strategic planning, and crime prevention initiatives.
* **Policy Makers**: The analysis can provide valuable insights for policy makers in creating effective policies and laws to prevent such incidents. For instance, understanding the relationship between the victim’s relationship with the perpetrator and the crime status could inform policies aimed at protecting vulnerable groups.
* **Social Workers and Non-Profit Organizations**: These stakeholders can use the insights to develop targeted awareness campaigns, support services, and prevention programs. For example, the analysis of the delay in reporting the crime can help in understanding the barriers victims face in reporting such incidents, leading to initiatives aimed at encouraging timely reporting.
* **Researchers**: Academics and researchers in the fields of criminology, sociology, and psychology can use this analysis to further their research on the subject. The dataset and the findings can serve as a basis for conducting more detailed studies, developing theories, or building predictive models.
* **General Public**: While the data is sensitive, anonymized, and aggregated findings can be shared with the public to raise awareness about the prevalence of such crimes, encourage community vigilance, and inform individuals about the resources available for victims.

Key Findings and Result Interpretation

Our analytics focus on analysing the rape data, to answer the following questions and make recommendations to relevant policy makers and government officials. - What is the relationship between most victims and their perpetrators, and how does this relationship influence the time taken to report the crime? - Does this crime occur more frequently on weekends or weekdays, and is it more common during the day or at night? - What is the yearly percentage of rape crimes being solved? Is this percentage higher than the number of pending cases? - What factors influence the resolution of the case?

1. What is the relationship between most victims and their perpetrators, and how does this relationship influence the time taken to report the crime?

A screenshot of a graph

Description automatically generated

Figure Histogram of relationship with victim vs status description.

The image in Figure 3 shows most perpetrators are associates of the victims, meaning family or friends of these victims. It also shows that most crimes committed by associates of the victim end up as unsolved cases (investigation ongoing).

A graph of a report

Description automatically generated

Figure Report Delay vs. Status description.

Figure 4 shows that the delay in reporting this crime could affect the status of the crime, hence rape crimes that are reported after 10 days are less likely to be solved. For this reason, victims are encouraged to save any evidence in case they choose to press charges against the offender. Could the case's outcome be impacted by a delay in reporting a rape crime? Most open and unresolved cases have significant report delay days, as seen in Fig. 4. Why do victims wait to disclose that they were sexually assaulted? 1. Fear of not being accepted 2. Fear brought on by verbal threats. It is recommended that you call 911 as soon as possible after the crime is committed, preserve all tangible evidence including the clothes you were wearing at the time in a plastic bag, and refrain from eating, drinking, taking a shower or bath, brushing your teeth, washing your hands, or touching your privates until you have had a medical examination. If there is a delayed response, the inquiry will end if there is insufficient evidence to pursue the case, thus victims are encouraged to save any evidence in case they choose to press charges against the offender.

1. Does this crime occur more frequently on weekends or weekdays, and is it more common during the day or at night?

A diagram of a distribution of crime

Description automatically generated

Figure Distribution of crimes by day of the week.

A blue and white graph

Description automatically generated

Figure Distribution of crimes by time of the day.

The image in Figure 5 demonstrates that rape happens more often on the weekends, Saturday, Sunday, and Friday respectively. The frequency of this crime on weekends may be linked to recreational activities since victims may be at home with the associates that are perpetrators or out having fun. Figure 6 demonstrates that rape crimes tend to happen at night. This may be mostly since the offender and victim are typically associates, or persons with whom they are comfortable or were in the past.

1. What is the yearly percentage of rape crimes being solved? Is this percentage higher than the number of pending cases?

A screenshot of a graph

Description automatically generated

Figure Layout of Status Description by year.

Figure 7 shows the cases solved are consistently 90% less than the investigations ongoing yearly.

1. What factors influence the resolution of the case?

According to all the above findings, it is concluded that the two factors that affect the resolution of the case is the relationship with the victim and the delay in report time.

1. Time series analysis

A screenshot of a data

Description automatically generated

Figure Time series analysis for rape data.

To track the trend of rape events over time, a time series analysis was done. The date of occurrence was used to sort the data, and a time series was made to compile the total number of incidences that occurred each day.

**REFERENCES**

1. Los Angeles Police Department (LAPD) (2023) 'Rape Incident Data for Los Angeles (2020-2023)', LAPD Crime Data [Online]. Available at: <https://data.lacity.org/> (Accessed: o5, May 2024).
2. Kantardzic, M. (2011) Data Mining: Concepts, Models, Methods, and Algorithms. 2nd edition. Hoboken, NJ: John Wiley & Sons.
3. Han, J., Kamber, M. and Pei, J. (2011) Data Mining: Concepts and Techniques. 3rd edition. San Francisco, CA: Morgan Kaufmann.
4. Object Management Group (OMG) (2017) *Unified Modeling Language (UML) Specification*. Available at: https://www.omg.org/spec/UML/2.5.1/ (Accessed: 11, May 2024).
5. Miro (2020) ‘What is a UML Diagram? Different Types and Benefits’, *Miro*. Available at: https://miro.com/uml-diagram/ (Accessed: 11, May 2024).
6. Bretheim, D. and Watson, W. (n.d.). *Comparing SAS ® and Python -A Coder’s Perspective*. [online] Available at: <https://support.sas.com/resources/papers/proceedings19/3884-2019.pdf>.\
7. Statswork (2024). *A Comparative Analysis: Python, R, and SAS*. [online] Medium. Available at: https://statswork.medium.com/a-comparative-analysis-python-r-and-sas-fca47e23a6d3 [Accessed 23 May 2024].