Road accident analysis

Sai Yaaminie Ganda, Ann Alex, Preeti Aladakatti, Pawan Aditya Man
Department of Computer Engineering
San Jose State University
saiyaaminie.ganda@sjsu.edu, annalex@sjsu.edu, preetiprakash.aladakatti@sjsu.edu, pawanaditya.man@sjsu.edu

***Abstract*—Traffic accidents are a major issue for public health, causing a high number of fatalities and destruction of public property. Given these challenges, it is critical to implement efficient strategies to reduce accident risks and enhance road safety results. This study aims to discover and examine different kinds of accidents from a range of viewpoints using exploratory data analysis. Analyzing accident data patterns and trends provides valuable insights that can inform evidence-based interventions and policy development. The results show that most accidents happen after work hours, emphasizing the need for specific traffic control tactics during these times. By taking a comprehensive approach to studying road accidents, this study aims to help create effective strategies to lower accident rates and improve road safety for everyone using the roads.**

***Index Terms*—Road accidents, exploratory data analysis and patterns**

I. Introduction

Road traffic accidents pose a significant threat to public safety and economic development worldwide. In Ethiopia, the impact of road accidents is particularly severe, with the country facing one of the highest rates of road traffic accidents in the world [1]. According to the World Health Organization (WHO), road traffic accidents are the second leading cause of death in Ethiopia, resulting in significant human suffering and economic losses.

Road traffic accidents are a major problem in Addis Ababa, Ethiopia's capital and largest city. Rapid urbanization and increased motorization have caused a surge in traffic-related incidents in the city. Despite efforts by the Ethiopian government and various stakeholders to improve road safety, the number of accidents and casualties remains alarmingly high [2].

Understanding the factors contributing to road traffic accidents and their severity is crucial for developing effective road safety strategies. Data-driven approaches, such as data mining algorithms, can provide valuable insights into the underlying causes of accidents and help identify areas for intervention. By analyzing accident data from the Addis Ababa subcity, we aim to identify patterns and trends that can inform targeted road safety measures.

The dataset was collected from police departments for research purposes, covering the years 2017–20. The major causes of accidents were analyzed after careful preprocessing to exclude sensitive information. With 32 features and 12316 rows, the dataset provides a comprehensive view of road traffic accidents in the region.

In this study, we developed a data mining algorithm-based approach to analyze road traffic accident severity in the Addis Ababa sub city. The goal of our analysis was to identify the most frequent factors contributing to high-severity accidents. Policymakers and stakeholders in Ethiopia can gain actionable insights that will help improve road safety through an analysis of the major factors that contribute to road traffic accidents.

II. Motivation

Ethiopia faces a considerable issue with road traffic accidents, leading to a high number of injuries and deaths. Making sure the security of citizens is a top priority for every government, requiring a comprehensive analysis of the factors that lead to serious accidents and loss of life. In this study, we developed a data mining algorithm-based approach to analyze road traffic accident severity in the Addis Ababa subcity [1]. Our data collection includes a wide range of characteristics linked to these events, enabling us to perform an in-depth examination focusing on particular aspects. One surprising observation from our analysis is that individuals with 5-10 years of driving experience caused the majority of casualties. This unexpected finding underscores the importance of analyzing features that are often neglected but can provide valuable insights for spreading awareness and implementing targeted interventions to reduce road traffic accidents in Ethiopia.

Reports indicate that traffic accidents in Ethiopia lead to an annual loss exceeding 12,000,000 Ethiopian Birr [2]. The frequency of traffic accidents and pollution in Addis Ababa has gone up due to the increase in both motor vehicles and population. Even though Addis Ababa has a low number of vehicles, it has a large proportion of accidents nationwide, with an annual average traffic accident growth rate of 31.4% [2]. Due to the seriousness of this issue, it is imperative to implement data-driven strategies to tackle road traffic accidents in Ethiopia promptly. Our goal is to offer valuable insights through a thorough data-driven analysis to guide policy-making and interventions for enhancing road safety and saving lives.

III. Literature Review

Many studies have been conducted to detect skin cancer from dermoscopic images using machine learning models. The recent federated learning paradigm opened a new dimension of study focused on medical data privacy-preserving machine learning. Federated learning being a new concept, the number of studies applying it in healthcare has been limited.

The authors in [3] address the global issue of road accidents with the use of data mining techniques to understand, analyze and find patterns in road accidents and thereafter plan the preventive measures to reduce road accidents. The paper is divided into 2 sections - descriptive analysis and predictive analysis, the first section focuses on the use of descriptive analyses to find various factors responsible for accidents, in the second section explores three machine learning models to predict the hotspots of the road accidents.

Road traffic data is complex as it covers variables ranging from details about the driver to casualty information, accident type, spatial data, and temporal information. Due to this heterogeneity in the dataset, it is challenging to analyze it and find patterns, relations, and trends for road accident analysis. [4] proposes a data mining framework with a K-modes clustering algorithm to segment the heterogeneous road accident dataset and apply association rule mining to identify the relationships between various factors. The results showed that the proposed architecture is better at capturing hidden patterns than conventional methods of analyzing or modeling the entire dataset at once.

Environmental conditions and temporal factors influence the occurrence and severity of road accidents. Analyzing and deducing the patterns between them and road accidents will help road safety authorities to implement necessary preventive measures effectively, especially during specific environmental conditions or times. [5] analyzes the impact of environmental conditions on road accidents using statistical analysis to identify correlations between various attributes with accident severity and fatalities. The findings indicate that accidents are more likely to occur under normal weather conditions, with a higher risk of injury or death during night time.

Road accidents also have an indirect effect on medical costs and productivity costs irrespective of the severity of the accident. The study [6] conducted on Netherlands data highlights the importance of analyzing non-fatal accidents in addition to fatal incidents because non-fatal accidents are more frequent and result in higher medical and productivity costs. These costs show different trends with the gender and the age of the causality. The analysis showed that road user type is also an important attribute when analyzing accident severity. This study shows high total cost incurred due to bicyclists is more due to the high number of bicyclists in Netherlands.

The severity of the accidents majorly depend on the speed of the vehicle. This led governments to enforce road speed limits in narrow roads or accident prone areas. The study [7] shows the effect of lifting or relaxing the speed limit rules in urban and rural areas.

If not properly regulated, junctions prove to be a major area of traffic accidents. This makes it important to analyze the type of junction and type of accidents occuring in that region to prepare a better infrastructure or traffic regulation. The study [8] analyzes the junction structure of roads in Japan by employing machine learning techniques to understand the effect of junction type, intervals and bending angles on the accident severity. After through analysis and complex modeling considering of various factors like age of victim, type of collision, etc, the study finds that shorter intervals between junctions will better regulate traffic and minimize traffic accidents.

IV. Project Description

The aim of our project is to analyze the road traffic accidents dataset and finding the patterns or similarities among accidents which would further help us and also road safety authorities to make informed decision about necessary measures to mitigate accidents.

The objectives of this research include -

- Data Analysis 1 - Conduct exploratory data analysis to identify and comprehend general trends and patterns within the road accident dataset.
- Data Analysis 2 - Investigate the underlying structures and factors contributing to anomalies in road accidents

V. Data Analysis

A. Dataset

The dataset under consideration comprises records of road accidents, encompassing various features pertaining to the circumstances, individuals involved, and outcomes of these incidents. Comprising 32 distinct features and a total of 12,316 instances, the dataset offers a comprehensive insight into the factors contributing to road accidents. Each instance provides information regarding the time and day of the accident, demographic details of the drivers and casualties, as well as details regarding the vehicles involved, road conditions, and severity of the accidents. Notably, all sensitive information has been excluded through data encoding processes, ensuring the privacy and confidentiality of individuals involved. Through detailed analysis of this dataset, valuable insights can be gleaned to inform road safety measures, policy formulation, and accident prevention strategies.

B. Data Processing

Effective data preprocessing techniques are crucial for ensuring the accuracy and reliability of study findings. One common preprocessing step involves addressing missing values, which can significantly impact the quality of analyses and interpretations. In this study, we employed a robust approach to handle missing values, incorporating sub grouping and mode imputation techniques.

- Identification and Handling of Missing Values:
  Initial data inspection revealed the presence of various missing value representations, including 'unknown', 'Unknown', and 'na'. To ensure consistency and compatibility with subsequent analyses, these representations were uniformly converted to NaN (Not a Number), signifying missing values across the dataset.
- Total Number of Missing Values Assessment:
  Following the conversion of missing value representations to NaN, the total number of missing values across the dataset was calculated. This assessment provided critical insights into the extent of missing data, guiding

subsequent preprocessing steps and informing decisions regarding imputation strategies.

- Imputation of Missing Values using Mode with Subgrouping Technique:
  To address missing values effectively while preserving data integrity, a mode imputation strategy with subgrouping technique was employed. This approach involved: Identification of relevant subgroups within the dataset based on categorical variables or logical groupings. Calculation of the mode (most frequent value) within each subgroup to serve as the imputation value. Imputation of missing values within each subgroup using the respective mode value.

C. Identifying general trends

1) Peak Hours: Analysis of the dataset reveals significant patterns in the timing of accidents, closely tied to daily and weekly commuting patterns. Specifically, during weekdays, the hours of 7 to 9 AM and 5 to 8 PM witness a notable surge in traffic volume, aligning with the typical commute to and from workplaces. Correspondingly, this surge in traffic correlates with a higher frequency of accidents during these peak hours. This observed trend highlights risk of accidents during periods of increased traffic congestion.

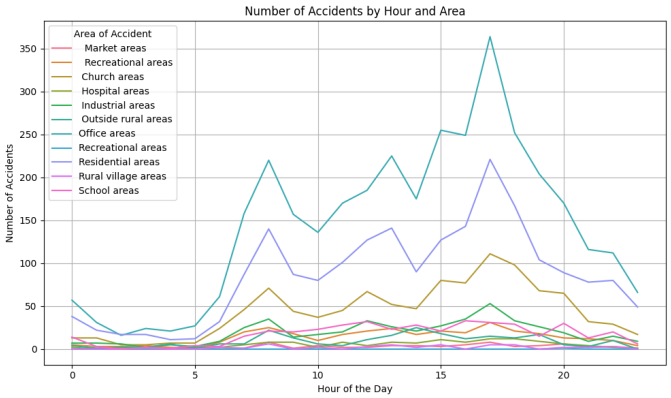

Fig. 1: Peak Hours of road accidents in Addis Ababa

Interestingly, weekends exhibit a different pattern, particularly during late-night hours. On Saturdays and Sundays, there is a notable spike in accidents occurring between 12 AM and 3 AM, contrasting with the relatively lower accident rates during early morning hours (6 to 10 AM). Subsequently, accident counts follow a similar path to weekdays, gradually increasing as the day progresses.

The discrepancy between morning and evening peak hours in terms of accident frequency may be attributed to workplace pressures and time constraints. During morning rush hours, individuals may prioritize punctuality and adhere to safer driving practices, whereas the rush to reach home after work during evening hours could lead to more aggressive driving behaviors and increased risk-taking on the road.

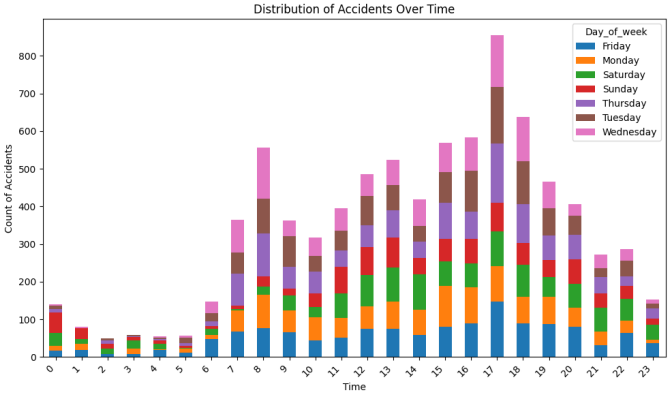

Fig. 2: Peak Hours of road accidents in Addis Ababa

Overall, these findings underscore the intricate relationship between commuting patterns, traffic volume, and accident occurrence. Understanding these dynamics can inform targeted interventions aimed at promoting safer driving behaviors during peak traffic hours and enhancing road safety measures during both weekdays and weekends.

2) Vehicle information: Vehicle owners may exhibit more varied driving behaviors due to their familiarity with their vehicles and frequent usage patterns, increasing the likelihood of accidents. Neglecting maintenance tasks and distractions while driving further contribute to higher accident rates among vehicle owners. Additionally, perceived lower risk levels and less stringent enforcement compared to professional drivers may exacerbate the problem. Addressing these factors through interventions like education campaigns and enhanced enforcement can help mitigate accident risks and improve overall road safety.

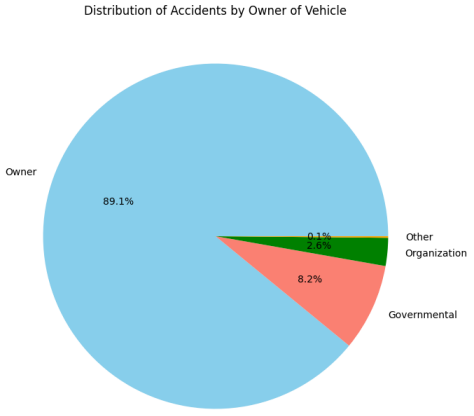

Fig. 3: Distribution of accidents by owner of the vehicle

Automobiles are responsible for a higher number of accidents primarily due to their sheer prevalence on the city's roads. As the most common mode of transportation, automobiles outnumber other types of vehicles, resulting in a greater

frequency of accidents involving them. The high volume of automobiles contributes to increased traffic congestion, leading to elevated accident risks and incidences.

TABLE I: Contribution of Various Vehicle Types to Total Accident Incidents

Type of accidents	Count of accidents
Automobile	2749
Lorry (large)	1480
Lorry (small)	378
Pickup upto 10Q	544
Public (12Q)	448
Stationwagen	438

3) Environment and road conditions: We have found a significant connection between higher accident rates and the existence of daylight and well-lit roads at night in our study. This occurrence is often linked to increased traffic levels in the daytime and people's tendency to steer clear of dark roads at night, resulting in greater chances of accidents. Additionally, our findings indicate that people are more careful while driving on poorly lit or unlit roads at night, which may help prevent accidents by increasing their alertness and practicing defensive driving techniques. These results emphasize the complex relationship between environmental factors, human actions, and road safety results.

Due to its tropical climate, Addis Ababa experiences minimal instances of rain or snowfall, which consequently results in fewer accidents during adverse weather conditions such as rain or snow. The infrequency of precipitation events in the city reduces the likelihood of slippery road surfaces and impaired visibility, factors commonly associated with weather-related accidents.

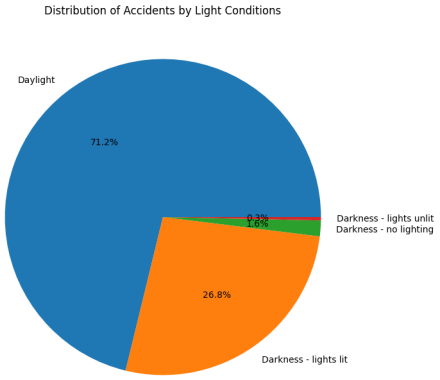

Fig. 4: Distribution of accidents by owner of the vehicle

Y-shaped intersections frequently have several traffic flows merging and diverging, which can make navigating the crossing more complicated. Drivers might have difficulty predicting the actions of cars approaching from various directions, resulting in confusion and possible accidents. In this city, we also notice that the majority of accidents happen at this particular type of intersection.

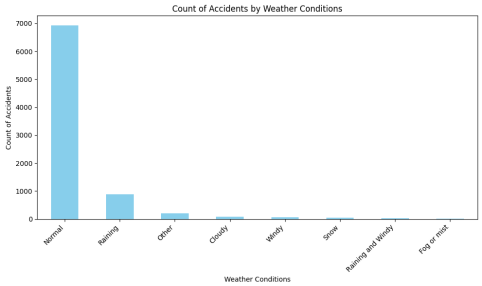

Fig. 5: Distribution of accidents by owner of the vehicle

4) Accident details: The higher occurrence of accidents when vehicles are going straight may be attributed to several factors. Firstly, on straight roads, drivers may perceive a reduced need for vigilance and may become complacent, leading to lapses in attention and slower reaction times to potential hazards. Additionally, straight roads often facilitate higher speeds, increasing the severity of collisions when accidents do occur. Moreover, the absence of curves or bends may lead to monotony, causing drivers to lose focus and engage in distracted behaviors such as texting or adjusting the radio. Lastly, straight roads may lack visual cues or landmarks, making it challenging for drivers to accurately judge distances and anticipate potential hazards, thereby increasing the likelihood of accidents.

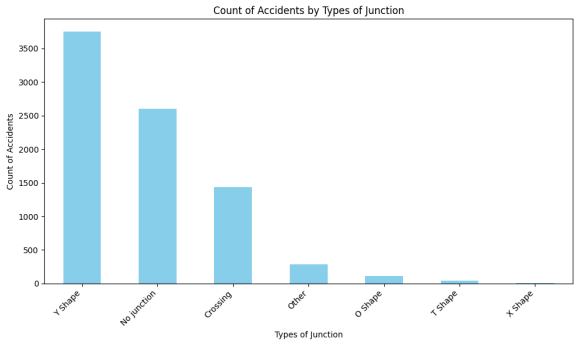

Fig. 6: Distribution of accidents by junction type

Our analysis reveals a prevailing trend in which the majority of accidents typically involve only one or two vehicles. Intriguingly, we observed a distinct pattern regarding accident severity corresponding to the number of vehicles involved. Specifically, accidents involving two vehicles exhibited heightened severity, with a notable increase in the occurrence of serious and fatal injuries compared to incidents involving only one vehicle. Surprisingly, our findings indicate that accidents involving three or more vehicles tend to be less severe, with fewer instances of serious injuries.

As a general conclusion, the casualty severity depends on accident severity.

D. Deviations from the usual trend

1) Vehicle Information: Service year, defined as the year when a vehicle was last serviced, serves as a crucial indicator

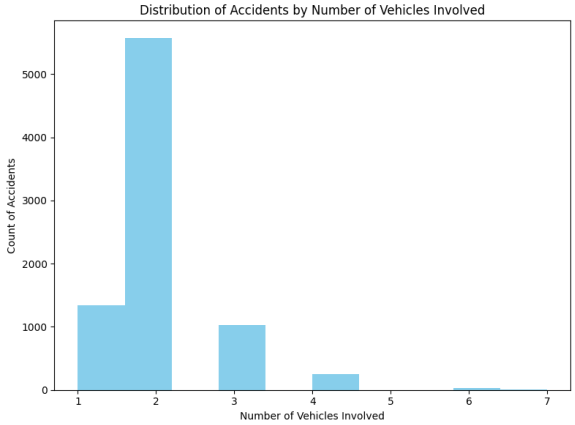

Fig. 7: Distribution of accidents by number of vehicles involved

of maintenance history and overall vehicle condition. In our study, we anticipated that vehicles with longer service periods would be associated with a higher frequency of serious accidents. However, our analysis yielded unexpected findings, revealing that vehicles with service periods ranging from 2 to 5 years were responsible for a disproportionate number of fatal and serious accidents.

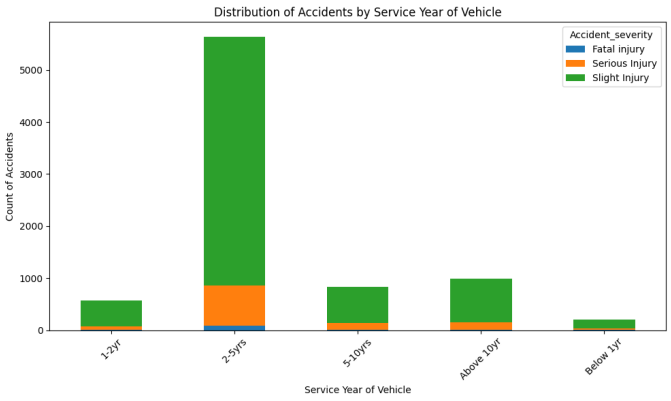

Fig. 8: Service year of vehicles

Our study revealed interesting patterns in the seriousness and rate of incidents with large vehicles like trucks. It was unexpected to discover that larger vehicles were linked to a lower number of serious accidents, indicating possible variations in crash dynamics or safety characteristics when compared to smaller vehicles. On the other hand, public transportation vehicles, although they transport fewer passengers, were found to be involved in more accidents, including many serious ones. Additionally, station wagons, although having a smaller coverage area, showed accident rates similar to public transport vehicles. These results highlight the complex connection among the type of vehicle, the seriousness of accidents, and the factors influencing road safety results.

2) Environment and Road conditions: A surprising anomaly appeared in our examination, disrupting traditional

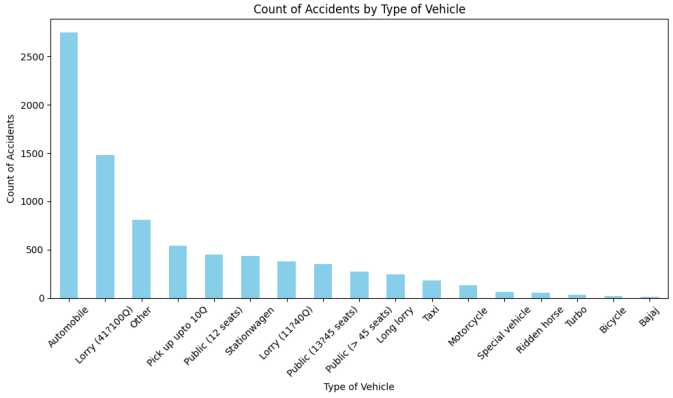

Fig. 9: Different types of vehicles involved in the accident

views on road safety. In contrast to what was expected, our research showed that two-way lanes with solid lines had fewer accidents than two-way lanes divided by broken lines. This unforeseen difference challenges the usual beliefs about the safety advantages of defined lanes, indicating possible complications in the relationship between road layout, driver actions, and accident frequency.

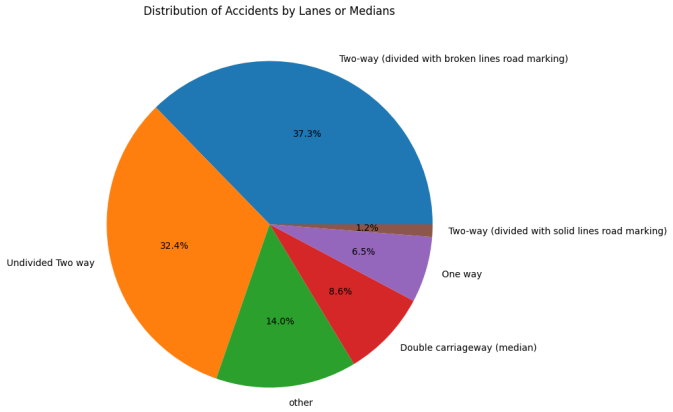

Fig. 10: Distribution of accidents by lanes or medians

VI. Community Contribution

Identifying Patterns and Trends: By analyzing the provided dataset, the community can identify patterns and trends related to road accidents. For instance, in our analysis, Y-shaped junctions are more prone to accidents. Sharing these conclusions of the analysis with the community makes our community more informed about the trends and they could avoid possible accidents by staying informed.

Education and Awareness: Launch educational campaigns and outreach initiatives to raise awareness about road safety issues and promote responsible behavior among road users. Collaborate with schools, universities, and local organizations to organize workshops, seminars, and training programs on safe driving practices, pedestrian safety, and accident prevention strategies. from the analysis of the data we also saw that

the office area that is downtown has recorded more accidents so we educated more people in the urban area with campaigns and road safety events.

Informing Policy Decisions: Insights from the analysis can be used to inform policy decisions aimed at improving road safety. For example, from the data analysis we did in the report we have seen certain types of roads have a higher number of accidents, policymakers can prioritize improvements or implement stricter regulations in those road types.

Data-driven Interventions: Finally, the insights gained from the analysis can be used to develop data-driven interventions aimed at reducing the number and severity of road accidents. These interventions can range from targeted enforcement of traffic laws to investments in infrastructure and education programs. We can also provide the results of our data analysis to authorities and people by putting up the findings on the website and sharing them with some influencer persons so that with their help we can reach to public and make them aware of the possible situations.

Emergency Response Planning: Understanding the types and severity of accidents can aid in emergency response planning. For instance, in our analysis the head-on-head type of collision is associated with more severe casualties, emergency services can be better prepared to handle such situations and provide appropriate medical care promptly. Making the persons involved in the accident get better care with prior information. Also, we can give insights from past accidents that have similar conditions to the emergency responders on what they can expect and make them prepared fully.

VII. Conclusions

Road accidents can be attributed to a lot of factors. The type of road accidents might differ from city to city. In summary, our thorough examination of traffic accidents in Addis Ababa reveals numerous important discoveries and perspectives that enhance our knowledge of road safety patterns and guide possible actions. Initially, we found specific times when accidents peak, with weekday rush hours having a higher number of incidents during commuting times and weekends having a different pattern, especially late at night. These results highlight the significance of implementing specific traffic control techniques during busy times. Additionally, when examining vehicle-related factors, we observed surprising patterns, including the overrepresentation of vehicles aged 2-5 years in fatal and serious accidents. Moreover, trucks had lower rates of serious accidents compared to public transportation vehicles, which had higher accident rates, questioning traditional beliefs about the correlation between vehicle size and safety. Next, our analysis of the environment and road conditions revealed the importance of well-lit roads in reducing accident rates, as well as the unexpected finding of fewer accidents on continuous two-way lanes in comparison to separated lanes. Finally, our analysis can inform the development of data-driven interventions that reduce the severity and number of road accidents in Addis Ababa and beyond.

References

[1] "Global status report on road safety," *World Health Organization*. [Online]. Available: https://www.mdpi.com/2071-1050/15/3/2722https://www.who.int/news/item/13-12-2023-despite-notable-progress-road-safety-remains-urgent-global-issue#:~:text=The%20latest%20WHO%20global%20status,and%20rising%20risk%20of%20death

[2] "Ethiopian ministry of transport." *World Health Organization*. [Online]. Available: https://unece.org/sites/default/files/2021-06/UNRSF_Annual_Report_2020_made_in_2021_2.pdf

[3] A. Comi, A. Polimeni, and C. Balsamo, "Road accident analysis with data mining approach: evidence from rome," *Transportation Research Procedia*, vol. 62, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352146522002265

[4] S. Kumar and D. Toshniwal, "A data mining framework to analyze road accident data," *Journal of Big Data*, vol. 2, p. 26, 11 2015. [Online]. Available: https://doi.org/10.1186/s40537-015-0035-y

[5] M. M. Jalilian, H. Safarpour, J. Bazyar, M. S. Keykaleh, L. Malekyan, and A. Khorshidi, "Environmental related risk factors to road traffic accidents in ilam, iran," *Medical archives (Sarajevo, Bosnia and Herzegovina)*, vol. 73, pp. 169–172, 06 2019.

[6] M. van der Vlegel, J. A. Haagsma, L. de Munter, M. A. C. de Jongh, and S. Polinder, "Health care and productivity costs of non-fatal traffic injuries: A comparison of road user types," *International journal of environmental research and public health*, vol. 17, 03 2020.

[7] J. I. Castillo-Manzano, M. Castro-Nuño, L. López-Valpuesta, and F. V. Vassallo, "The complex relationship between increases to speed limits and traffic fatalities: Evidence from a meta-analysis," *Safety Science*, vol. 111, pp. 287–297, 03 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753518306350

[8] Y. Wada, Y. Asami, K. Hino, H. Nishi, S. Shiode, and N. Shiode, "Road junction configurations and the severity of traffic accidents in japan," *Sustainability*, vol. 15, 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/3/2722