Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ttra21

Uncovering pedestrian midblock crash severity patterns using association rules mining

Rohit Chakraborty, Subasish Das & Md Nasim Khan

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2024.2379948

Uncovering pedestrian midblock crash severity patterns using association rules mining

Rohit Chakraborty [a], Subasish Das [b] and Md Nasim Khan [c]

[a]Graduate Student, Texas State University, 601 University Drive, San Marcos, TX, USA; [b]Assistant Professor, Civil Engineering, Texas State University, 601 University Drive, San Marcos, TX, USA; [c]Sr Engineer I, Transportation Planning West, Austin, TX, USA

ABSTRACT

The current study investigated the contributing factors and temporal variation in pedestrian crashes at midblock, with a particular focus on the severity levels: fatal/severe, moderate injury, and minor/no injury. It used association rules mining to uncover patterns between crash-contributing factors. By generating, evaluating, and visualising association rules for each severity level within each cluster, significant findings were discovered. Significant associations are observed between fatal crashes on weekdays and factors such as alcohol or drug impairment and nighttime. Similarly, factors including one-way roadway type, summer, and 25 MPH posted speed limit have a strong association with moderate injury crashes during weekdays. On weekends, nighttime crashes with non-motorised vehicles have the strongest association with fatal/severe injury crashes. Moreover, the generated rules for nighttime pedestrian fatal/severe injury crashes highlighted physical or drug-impaired pedestrians as the predominant attribute. The findings can enhance pedestrian safety at midblocks through targeted interventions.

ARTICLE HISTORY
Received 17 March 2024
Accepted 9 July 2024

KEYWORDS
Pedestrian safety; association rules mining; crash severity; midblock crash

1. Introduction

Walking, being the most basic and common form of transportation in daily life, provides health benefits as long as injuries caused by traffic crashes are avoided. The automobile transportation network is continuously increasing around the world. This also indicated the inclusion of more vehicles in the network. As vehicular traffic continues its upward trajectory, vulnerable road users, such as pedestrians, find themselves increasingly susceptible to traffic crashes (Zegeer and Bushell 2012). Over the past five years, pedestrian fatalities only in Texas have risen by 29.6%, representing one-fifth of all deaths on roadways. In the year 2022, the state witnessed 5,764 pedestrian-involved crashes, leading to 829 fatalities and causing severe injuries to an additional 1,526 individuals (TXDOT 2023). The Governors Highway Safety Association (GHSA) report, which was prepared utilising preliminary data from State Highway Safety Offices, indicates that 3,434 pedestrians lost their lives on U.S. roadways during the first half of 2022. A subsequent report examining state-reported

CONTACT Rohit Chakraborty xuw12@txstate.edu Graduate Student, Texas State University, 601 University Drive, San Marcos, Texas 78666, USA

data for the entire year highlighted the persistent high fatality rates for pedestrians. In 2022, there were 2.37 pedestrian deaths per billion vehicle miles traveled (VMT), marking an increase and continuing a concerning trend of elevated rates that began in 2020. The analysis also revealed a disconcerting safety imbalance for pedestrians, with a 77% increase in pedestrian deaths between 2010 and 2021, in stark contrast to the 25% rise observed in all other traffic fatalities (GHSA 2022).

Moreover, at least 7,508 pedestrians were killed by drivers on the roads of the U.S. during the year 2022. It was recorded as the most pedestrian deaths in the U.S. since 1981 (GHSA 2023). Pedestrian safety is paramount; however, it is challenging for several reasons. Pedestrians face greater susceptibility to traffic crashes compared to individuals using alternative modes of transportation, given the direct exposure of the human body to the forces at play. Compared to vehicle occupants, pedestrians lack physical barriers that protect them during crashes, thereby escalating the risk of serious injuries or fatalities. Therefore, in-depth investigations are essential to comprehend the complex factors that influence the likelihood of pedestrian-involved crashes.

Analysing pedestrian crashes at midblock locations is crucial due to the high vulnerability and fatality rates of pedestrians, particularly in areas lacking controlled crossing points. Traditional safety approaches often react to locations with a history of crashes, but a systemic analysis allows for a proactive strategy, identifying and addressing high-risk locations before crashes occur. A study focused on identifying and addressing severe midblock pedestrian crashes by integrating crash data with roadway, transit, census, and equity information to pinpoint high-risk segments revealed that principal arterials, minor arterials, and major collectors are facility types with a higher probability of severe pedestrian crashes at Massachusetts. Using binomial logit regression models, the study identified consistent risk factors across these facility types, including the number of lanes, traffic volume, population density, employment density, commute behaviours, and equity measures. The research emphasised a proactive systemic safety approach, contrasting it with reactive 'hot-spot' projects by broadly prioritising sites based on crash probability and applying low-cost safety countermeasures (Gooch et al. 2022). Another study investigated pedestrian crash risk factors in Seattle. The study employed multilevel mixed-effects Poisson models to account for correlations within and between locations over time. The study found that midblock locations accounted for 35% of pedestrian crashes from 2007 to 2013. Key risk factors for midblock crashes included wider streets and higher traffic volumes. Locations with marked crosswalks at midblock points had increased crash rates. Conversely, midblock locations on one-way streets and those in areas with higher intersection density had lower crash rates. The study emphasised the need for targeted safety improvements at midblock locations to reduce pedestrian crashes and promote safer walking environments (Quistberg et al. 2015). Kumfer et al. (2019) conducted a study addressing the rising concern of pedestrian fatalities in the United States, mainly focusing on midblock locations. The study used eight years of segment data from the Seattle Road network to develop safety performance functions (SPFs) for predicting two types of severe pedestrian crashes: crashes between motor vehicles traveling straight through and pedestrians and nighttime crashes involving pedestrians. Key findings included identifying risk factors such as the presence of four or more lanes, two-way left-turn lanes, on-street parking, speed limits of 30 or 35 MPH, right-turn lanes at adjacent intersections, and marked midblock crosswalks. These factors, along with pedestrian volumes and transit activity, were associated with higher crash likelihood. The study emphasised the importance of proactive, systemic safety analysis over traditional hotspot approaches, suggesting that addressing these risk factors with targeted countermeasures can enhance pedestrian safety across a broader network.

Researchers in the field of transportation have consistently shown a deep-seated interest in examining pedestrian safety. A significant volume of literature has been dedicated to this subject, employing traditional statistical methods as well as machine-learning approaches to forecast the severity of pedestrian crashes. Conventional and widely adopted discrete choice modeling approaches, such as mixed logit models, multinomial logit models, ordered logit/probit models, and partial proportional odds logit models, have been employed. While these statistical approaches can provide valuable insights into how different risk factors impact the levels of crash severity, they come with predefined assumptions. As a result, deviations from these assumptions may result in notable inaccuracies in predicting crash severity (Mondal, Bhuiyan, and Yang 2020; Tao et al. 2022). Regression models typically look at the average impact of risk factors and do not pay attention to specific categories. This leads interventions to target the average of the categorical variable without considering the unique needs of different groups. On the other hand, rule-based analysis looks at patterns of factors in specific categories of variables by considering how potentially influencing factors interact. Both methods can find similar influencing factors, but rule-based analysis is better at spotting multiple interactions between factors that define categorical variables (Haghighi et al. 2016). Moreover, regular statistical models usually make assumptions about how data is spread out or how independent and dependent variables relate (Rahman et al. 2021). Because of this, non-parametric data mining methods are gaining popularity in traffic safety research. These methods can uncover hidden connections among different crash factors without imposing strict rules on the type of variables used (Hossain et al. 2022b; Hossain et al. 2024). However, only a small number of studies have used unsupervised learning algorithms to study pedestrian crashes and investigate how crash severity depends on the underlying factors. To overcome the limitations of conventional statistical methods, we aimed to employ a robust data mining approach called association rules mining. Previous studies on pedestrian crash severity at midblock did not emphasise the risk factors related to pedestrian crashes during the day of the week and different lighting conditions. Therefore, the objective of this article is to investigate the factors contributing to pedestrian crashes and unveil patterns associated with pedestrian crash severity for weekday, weekend, daytime, and nighttime. Association rules mining presents a distinct advantage over traditional statistical modeling techniques as it facilitates the discovering and characterising relationships between variables (Das, Ahmed, and Ghasemzadeh 2019; Montella et al. 2011). Through a thorough analysis of traffic crash data, association rules mining offers valuable insights that may not be evident through conventional statistical approaches. This approach enables a more comprehensive understanding of the intricate interplay among various factors and the associated risk patterns in pedestrian-involved traffic crashes.

We adopted association rules mining for effectively analysing the pedestrian crash severity data at midblock. Through a systematic analysis, rules were generated, comprehensively evaluated, and subsequently visualised for three distinct levels of crash severity: fatal/severe injury, moderate injury, and minor/no injury. This research focused on cluster-based analysis to develop association rules. The rest of the paper is structured in the following manner: Section 2 explores an in-depth examination of literature related to pedestrian-involved crashes at midblock. Section 3 offers a comprehensive explanation of the association rules mining methodology employed in this research. Section 4 outlines the procedures undertaken to prepare the data for analysis, as well as results and discussions concerning association rules, supplemented by relevant graphical representations. Finally, Section 5 concludes the paper by summarising the investigative findings.

2. Literature review

2.1. Influencing factors on pedestrian injury severity

Several studies have investigated pedestrian behaviour at mid-block intersections. Abdullah, Oguchi, and Dias (2019) assessed pedestrian mid-block crossing behaviours, revealing four factors: risk-taking, wrong perception, walking for pleasure, and walking pattern. Demographic analysis showed that men and younger individuals exhibited higher risk-taking behaviour, aiding in the understanding and planning pedestrian crossing facilities. Avinash et al. (2019) aimed to assess pedestrian safety at urban midblock crossings in India through a video-graphic survey and Multiple Linear Regression (MLR) analysis. The regression model revealed the impact of pedestrian behaviour, vehicular gaps, and vehicle speed on pedestrian safety margins (PSM). Higher pedestrian and vehicular speeds and larger gaps increased safety margins while rolling behavior decreased safety.

In another study, Avinash et al. (2019b) explored pedestrian behavior at urban mid-block crossings amid mixed traffic. The findings indicated faster walking speeds among young pedestrians and the influence of city traits and gender on crossing behaviors. Safety recommendations included a minimum gap of 6.2 s, rising to 8 s in areas with more female or elderly pedestrians. Increased traffic lanes prolonged crossing times, leading to stage-wise crossing, while larger platoon sizes affected gap acceptance. Xue and Wen (2024) examined how drivers' route familiarity affects the severity of pedestrian injuries in vehicle crashes. The study found that for familiar drivers, factors like early morning, rainy weather, high-speed limits, young and intoxicated drivers, and elderly pedestrians increase injury severity. For unfamiliar drivers, foggy and rainy weather, dark conditions without road lights, elderly and female drivers, and continuous long downhill sections are associated with higher injury severity. The findings suggested the need for targeted safety measures, such as improved road lighting.

Some previous studies have explored the association of pedestrian age and gender with crashes. Toran Pour et al. (2018) examined how age and gender affect the spatial and temporal distribution of pedestrian crashes in Melbourne from 2004 to 2013. Key findings revealed that age and gender significantly influenced crash distributions. It was found that crashes involving school-aged pedestrians peak at 8:00 am and 3:00 pm during school commuting times; males aged 18-34 experience more nighttime crashes, likely due to higher alcohol consumption; and pedestrians aged 35-64 and those over 65 face higher risks during off-peak traffic periods, reflecting their walking patterns for fitness and daily activities. Spatial analysis showed that crash hotspots vary by age and gender, with younger pedestrians facing risks around schools, while older pedestrians are more vulnerable near parks and social centers. Toran Pour et al. (2017) found that elderly pedestrians (65+) faced higher risks, and crashes near points of interest (POIs) like bars and restaurants are more frequent and severe, particularly at night. Temporal analysis showed most crashes during off-peak times, with peaks around school and work commutes. Spatially, crashes are concentrated in the central business district during the day and shift to nightlife areas at night. Another study investigated the factors contributing to the severity of vehicle-pedestrian crashes involving school-aged children. The most significant factors include the proximity of crash sites to schools, with closer crashes being less severe, and socio-economic indicators, with lower-income areas experiencing more severe crashes. Additionally, higher traffic volumes and the distance of crashes from public transport stops are associated with increased severity (Toran Pour et al. 2017b). Similarly, Park, Abdel-Aty, and Lee (2019) evaluated the safety of school-aged pedestrians by assessing various roadway features and their effects on crash frequency. Key findings indicated that increasing shoulder and lane widths, installing flashing beacons at school zone speed limit signs, and decreasing the number of driveways are effective measures for reducing crash frequency. The study offered valuable insights for transportation practitioners and policymakers to enhance safety in school zones through empirical evidence and targeted safety measures. Another study investigates the impact of socioeconomic factors on the severity of vehicle-pedestrian crashes in a similar study area. The study focused on the neighborhoods where the crashes occur and where the road users live. The findings revealed that socioeconomic factors, particularly those related to the neighborhoods where pedestrians and drivers reside, significantly influence crash severity. Key factors include public transport usage, median age, and country of birth in the pedestrians' neighborhoods, and education level and ethnicity in the drivers' neighborhoods. Roadway and traffic-related factors such as distance from public transport stops, light conditions, and road gradients also played a crucial role (Toran Pour et al. 2017c).

2.2. Study methods

Several studies have employed various methods to investigate pedestrian crash risk factors. Angioi and Bassani (2022) investigated driver behavior in varying familiarity situations while examining driver-pedestrian interactions at uncontrolled mid-block crosswalks. Through a factorial experiment with crosswalk designs, driver familiarity, and pedestrian time gaps, it was found that route familiarity increased speed, while situation familiarity improved driver response, leading to speed reduction near crosswalks. Moreover, curb extensions enhanced pedestrian safety. Angulo and Smith (2021) implemented a pedestrian-to-vehicle communication system using C-V2X technology at a mid-block crosswalk. It was found that in-vehicle warnings significantly increased driver likelihood to stop for pedestrians are 2.44 times during the day and 1.79 times at night. Driver age, time of day, and warning presence significantly influenced stopping rates, highlighting the positive impact of C-V2X warnings on driver behavior and pedestrian safety.

Furthermore, Puterski et al. (1999) analyzed mid-block pedestrian crashes in Las Vegas and aimed to identify crash patterns, pedestrian characteristics, high-concentration roadway segments, and behaviors at these locations. GIS and cluster analysis pinpointed high-concentration areas, exploring connections to bus stops and marked mid-block crosswalks. The study demonstrated the feasibility of establishing crash rates based on traffic volume and mid-block crossings. Some studies investigated the association of temporal distribution of pedestrian crashes under different circumstances. Another study explored the effectiveness of time proximity-based (time-to-collision, TTC) and evasive action-based (maximum slope of step frequency, MSSF) indicators in assessing pedestrian-vehicle conflict severity. The study found that TTC is more effective in organized traffic environments, where interactions are predictable and compliant with traffic rules. Conversely, MSSF is more suitable for less-organized environments where sudden evasive actions are common (Tageldin and Sayed 2019). In a pedestrian safety-related study, Nasernejad, Sayed, and Alsaleh (2023) developed a multiagent simulation model to understand crash avoidance behaviors in pedestrian-vehicle interactions. Utilizing video data from a busy intersection, the study employed a Markov Game (MG) framework to model these interactions, recovering pedestrian and vehicle reward functions via Multi-agent Adversarial Inverse Reinforcement Learning (MAAIRL). The results showed that the multiagent approach outperforms traditional single-agent models in predicting road user behaviors and crash avoidance mechanisms, offering a more accurate reflection of real-world interactions.

Several safety studies on pedestrian crossings at mid-block have been conducted internationally. A study explored the factors contributing to the severity of vehicle-pedestrian crashes at mid-block in the Melbourne metropolitan area (Toran Pour et al. 2017d). The study developed three decision tree (DT) models using CART, bagging, and boosting techniques to identify these factors. The findings indicated that socio-economic characteristics, traffic volume, and neighborhood social characteristics significantly influence pedestrian crash severity. Higher traffic volumes, especially beyond 20,000 vehicles per day, and areas with certain demographic characteristics were associated with increased crash severity. The study concluded that the boosted DT model is more effective for analyzing and predicting pedestrian crash severity at mid-blocks compared to simple and bagged DT models. Zhang et al. (2017) analyzed pedestrian safety at multi-lane mid-block crosswalks in China. Video data was used to identify pedestrian-vehicle conflicts and assess their spatial distribution based on lane-specific post-encroachment time (LPET). Statistical analysis revealed conflict concentration in two lanes with varying rates. An OP model highlighted the positive effects of pedestrian refuges and the negative impacts of high vehicle speeds, traffic volume, rolling gap crossing patterns, and larger pedestrian groups on safety. Jain and Rastogi (2018) aimed to develop risk assessment criteria based on Psychological Gap Size (PGS) for safer pedestrian crossings in urban India's mixed traffic areas. This study established risk assessment criteria using 15th and 85th-percentile PGS values, offering a proactive tool to evaluate pedestrian-vehicle interaction severity and recommend appropriate crossing facilities for enhanced pedestrian safety.

Zhu et al. (2021) proposed a novel agent-based framework to assess pedestrian safety at un-signalized crosswalks, focusing on un-signalized mid-block crosswalks with refuge islands (OPR). Sensitivity analysis in this study highlighted factors increasing the probability of severe conflicts, such as longer driver reaction times, smaller safety margins, and visual obstructions near crosswalks. Additionally, the framework successfully simulated traffic crashes under extreme road conditions. Moreover, Zhu et al. (2021b) examined how autonomous vehicles interacted with pedestrians approaching mid-block crosswalks, considering visibility affected by built environments and surrounding vehicles. They created a model to assess visibility, embedded it in an agent-based framework, and tested different scenarios.

2.3. Analytical methods

A crash occurs due to a series of critical events. As crashes involve many factors like pedestrians, vehicles, roads, and the environment, they interact in different ways to cause crashes. Understanding these interactions is crucial for keeping pedestrians safe. Using methods like Association Rules Mining (ARM), the crash data can be analyzed to find patterns. Previous research has demonstrated the value of ARM in traffic safety analysis (Das and Sun 2014; Hossain et al. 2022a). Only a few studies examined association rules mining in traffic crash analysis in the past. In one study, ARM was applied to real traffic crash data collected from local police stations that offered significant insight into the phenomena of safety improvement (Marukatat 2007). Another study conducted in China, adopted a multidimensional association rules model of traffic crashes for the freeways. This study also presented preventive measures to reduce crashes (He et al. 2008). Pande and Abdel-Aty (2009) used ARM in safety analysis by generating closely associated crash characteristics in the form of rules. This study is based on this foundation by employing ARM to explore the factors associated with pedestrian-involved crashes. Unlike conventional statistical models, which may overlook complex interactions between variables, ARM provides a flexible and robust approach to uncovering hidden patterns in large datasets. By focusing on association rules rather than average impacts, ARM can identify multiple interactions between factors that define categorical variables, offering a more nuanced understanding of crash severity.

2.4. Research gaps and proposed method

Despite the insights gained from previous studies, significant gaps remain in the understanding of pedestrian crash severity at midblock locations, particularly concerning temporal and lighting conditions. Many studies have focused on specific factors such as pedestrian speed and behavior and driver behavior, but few have integrated these factors into a comprehensive analysis that considers the day of the week and lighting conditions. The literature search also revealed that temporal and lighting conditions have been explored in overall pedestrian safety studies, however, it was notable that the pedestrian safety studies particularly at midblock did not explore these factors extensively. As pedestrian safety is a concern at midblock crossing, it is necessary to explore these factors. Moreover, traditional statistical methods often fail to capture the complex, non-linear relationships between variables contributing to crash severity. The current study addresses these gaps by applying ARM to explore the risk factors based on the day of the week and lighting conditions. This allows for the identification of intricate patterns and associations without relying on predetermined assumptions. Additionally, the previous studies mostly focused on pedestrian road-crossing behavior at midblock, such as pedestrian speed and activities while crossing streets, driver behavior, driver distraction, and pedestrian-driver interaction. A few studies analyzed demographic factors of pedestrians and drivers, diverse traffic conditions, and CAV involvement. Moreover, previous studies did not analyze the possible influence of roadway characteristics, traffic control devices, and crash type based on different temporal clusters. The current scenario of pedestrian crashes at midblock necessitates an examination of the potential influence of these factors. This study addressed these research gaps by incorporating crash types, roadway characteristics and alignment, pedestrian and driver demographics, existing traffic control devices, season, and other factors as variables.

3. Methodology

3.1. Association rules mining

In this study, the utilization of association rules mining was applied to explore the underlying factors associated with pedestrian-involved crashes. In comparison to conventional statistical modeling techniques typically employed in crash data analysis, association rules mining presents several advantages. Statistical approaches, while capable of providing insights into the individual effects of different risk factors on crash risks, are constrained by their predefined assumptions. Violations of these assumptions can lead to biased or inaccurate results (Mondal, Bhuiyan, and Yang 2020). In contrast, association rules mining overcomes these limitations and provides distinct advantages. It exhibits superior performance, offers flexibility in handling various data distributions, efficiently manages large datasets, and reveals hidden relationships among multiple variables. A noteworthy advantage of association rules mining is its capacity to uncover concealed relationships within extensive databases. Unlike statistical modeling techniques relying on predetermined assumptions, association rules mining enables discovering and characterizing relationships between variables through rule formulation (Das, Ahmed, and Ghasemzadeh 2019; Montella et al. 2011). This approach effectively analyzes traffic crash data, offering valuable insights that are not readily apparent using traditional statistical modeling techniques. It allows for the identification of complex and non-linear relationships among various factors, contributing to a more comprehensive understanding of the factors and risk patterns associated with traffic crashes. Additionally, the efficient handling of big data by association rules mining makes it particularly suitable for analyzing large, intricate crash datasets common in traffic safety research.

Association rules mining is a robust descriptive data mining approach and a commonly used rule-based machine learning method to uncover meaningful relationships between variables in extensive databases. This technique aims to reveal hidden patterns within an itemset, encompassing various factors like environmental conditions, traffic characteristics, roadway geometry, and driver demographics in the context of this study. These factors concern a specific event, i.e. the severity of pedestrian-related crashes. Multiple algorithms exist for mining association rules, with the Apriori algorithm being extensively employed, initially introduced in a previous study (Agrawal, Imielinski, and Swami 1993). The Apriori algorithm utilizes a level-wise search strategy to extract frequent item sets, if subsets of frequently occurring item sets are themselves frequent. By employing this algorithm, subsequences or item groups recurrently appearing within a large dataset are extracted. In this study, the set of items $I = \{i_1, i_2, \ldots, i_m\}$ comprises crash classifications specific to pedestrian-involved incidents, and the collection of crash data $C = \{c_1, c_2, \ldots, c_n\}$ represents transactions in the database, where each crash record, c_i, consists of a subset of items selected from I. A k -itemset refers to an itemset containing k items, and an association rule is expressed as $Antecedent(A) \rightarrow Consequent(B)$, signifying that the occurrence of A implies a likelihood of B occurring. The selection of interesting rules and determination of association strength in this study is guided by four measures: support, confidence, lift, and coverage. Subsequent sections provide a comprehensive description of these measures.

3.1.1. Support

Support is a metric that quantifies how often an item set appears in a dataset. It indicates the ratio of transactions covered by a rule compared to the entire dataset (Hahsler, Grun, and Hornik 2005). Equation 1 defines the mathematical representation of support.

$$\text{Support, } S(A \to B) = \frac{A \cap B}{N} \tag{1}$$

Where,

$S(A \to B)$ = support of the association rule ($A \to B$),

$A \cap B$ = frequency of occurrences with both antecedent and consequent, and N = total frequency of occurrences

3.1.2. Confidence

The measure of confidence evaluates the dependability of a rule by determining its frequency of being accurate. It determines the extent to which item set B is observed when item set A is present. A high confidence value in an $A \to B$ relationship indicates that B consistently occurs in transactions where A is present as the antecedent (Hahsler, Grun, and Hornik 2005). The mathematical representation of confidence can be defined using Equation 2.

$$Confidence, C(A \to B) = \frac{S(A \to B)}{S(A)} \tag{2}$$

Where,

$C(A \to B)$ = confidence of the association rule ($A \to B$),

$S(A \to B)$ = support of the association rule ($A \to B$), and $S(A)$ support of antecedent A

3.1.3. Lift

The lift metric measures the observed occurrence of both the antecedent and consequence in relation to their expected co-occurrence. It calculates the ratio between the observed frequency and the anticipated frequency. When the lift value surpasses 1, it signifies positive independence, indicating that the antecedent and consequent appear together more often than expected. Conversely, a lift value below 1 indicates negative independence, suggesting that the antecedent and consequent appear together less frequently than expected (Montella et al. 2011). The mathematical representation of lift can be defined using Equation 3.

$$Lift, L(A \to B) = \frac{C(A \to B)}{S(B)} = \frac{S(A \to B)}{S(A).S(B)} \tag{3}$$

Where,

$S(A \to B)$ = support of the association rule ($A \to B$),

$S(A)$ = support of antecedent, and $S(B)$ = support of consequent

3.1.4. Coverage

The coverage is the support of the left-hand-side of the rule ($A \to B$), i.e. $S(A)$. It represents a measure of to how often the rule can be applied. Coverage can be quickly calculated from the rule's quality measures (support) stored in the quality slot. If this value is not present,

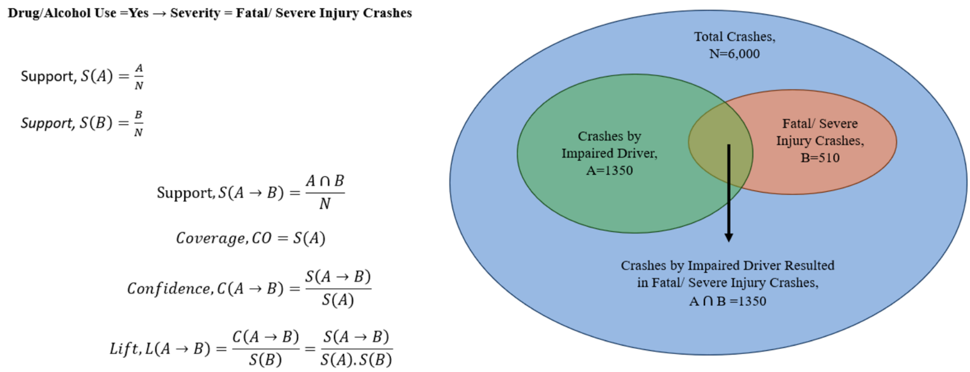


Figure 1. A Hypothetical Example of Estimating the Parameters of Association Rules.

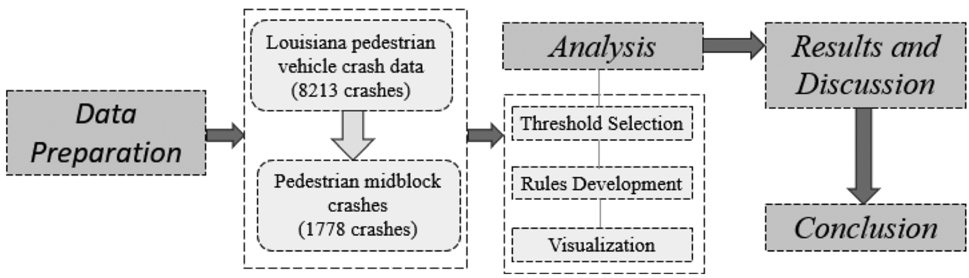


Figure 2. Flow-chart explaining research approach.

then the support of the LHS is counted using the data supplied in transactions.

$$Coverage,\ CO = S(A) \tag{4}$$

Figure 1 provides a hypothetical example of how support, confidence, coverage, and lift of an association rule can be calculated.

3.2. Data preparation

This section of the paper includes a brief explanation of the adopted research approach, collected crash data, and the crash locations. Figure 2 illustrates the research approach. We have followed four steps, from preparing the dataset to concluding the paper with policy implications. The first step includes the data preparation stage and variable selection. In the second stage, the analysis generates the association rules for all crash severity types. The results are discussed in the third step. The paper concluded by summarizing the research findings and providing several policy-making recommendations for the practitioners. The dataset considered in this research represents five years (2017-2021) of pedestrian-vehicle crash severity data from Louisiana. Initially, the data included 8213 pedestrian-vehicle crashes. From the dataset, 1778 crash data were finalized, relating only to the pedestrian crashes that occurred at midblock. The descriptives of the crash data are also discussed in this section.

Table 1. Descriptive of Pedestrian Crash Severity.

	Fatal/Severe Injury (KA)	Moderate Injury (B)	Minor/No Injury (CO)	Total
Count	450	739	589	1778
Percentage	25.31	41.56	33.13	100

The crash severity types are also denoted in the KABCO scale. The dataset includes 219 fatal (K) crashes, 231 severe injury crashes (A), 739 moderate injury crashes (B), 471 possible injury crashes (C), and 118 no injury crashes (O). Due to the low frequency of some crash severity levels, some categories are combined. To address data imbalance in data mining techniques, previous research has also merged different injury severity levels (Osman et al. 2016; Zhang et al. 2021). In the database, moderate injury crashes are dominant. Therefore, in this analysis, K and A combine the fatal/severe injury crashes, B is defined as moderate injury crashes and C and O are defined as minor/no injury crashes. Table 1 shows the descriptives of the 1778 pedestrian midblock crashes based on the severity. The descriptives show that most crashes caused moderate injury to the pedestrian, and fatal/severe injury crashes covered 25.31% of the dataset.

Table 2 shows descriptives of pedestrian crashes at midblock by day of week and lighting condition. Based on the day of the week, the descriptives show that more crashes occurred during the weekdays. Based on lighting conditions, more crashes were identified during the nighttime. Similar trends were followed for the fatal/severe injury crashes. A higher number of these crashes occurred during weekdays and nighttime. These statistics can emphasize the fact that weekdays and nighttime are crucial for pedestrian fatal/severe injury crashes. Pedestrian impairment is more significant on weekends (52.6% no impairment) compared to weekdays (61.5% no impairment) and at nighttime (47% no impairment at nighttime compared to 73.4% in daylight). The severity of crashes changes with the time of day; during daylight, minor injuries (MI) constitute 44.7%, but this decreases to 38.9% at nighttime, whereas fatal/severe injuries (FSI) increase from 14.1% in daylight to 34.8% at nighttime. The description also shows that the age of pedestrians involved in crashes varies significantly with time; for instance, 54.2% of middle-aged pedestrians are involved on weekdays compared to 64.3% during daylight. Driver conditions also vary, with drug/alcohol impairment reported in 8.09% of crashes at nighttime, which is higher than the 1.23% during daylight. Certain factors, such as pedestrian race and vehicle type, do not show substantial change between the compared times. The condition of pedestrians varies significantly between day and night; during daylight, 48.3% are distracted/inattentive, versus 32.7% at nighttime. Also, a larger proportion of pedestrians are in 'other' conditions at night (38.4%) compared to the day (25.3%). The *p*-values indicate the statistical significance of these differences, with several factors showing *p*-values less than 0.05.

Figure 3 shows the locations of crashes based on the crash severity types in the state of Louisiana. The maps indicate that most of the pedestrian crashes occurred at urban, suburban, and urban core locations.

4. Results and discussion

This section includes a detailed analysis and discussion of the findings of the analyses. Temporal cluster analyses have been discussed based on days of week and time of day. Figure 4

Table 2. Descriptive Statistics of Pedestrian Crashes by Day of Week and Lighting Condition.

	Weekday (N = 1308)	Weekend (N = 470)	*p*.overall		Daylight (N = 815)	Nighttime (N = 963)	*p*.overall
Severity			0.514	Severity			< 0.001
Fatal/Severe Injury (FSI)	325 (24.8%)	125 (26.6%)		Fatal/Severe Injury (FSI)	115 (14.1%)	335 (34.8%)	
Moderate Injury (MI)	540 (41.3%)	199 (42.3%)		Moderate Injury (MI)	364 (44.7%)	375 (38.9%)	
Minor/No Injury (MNI)	443 (33.9%)	146 (31.1%)		Minor/No injury (MNI)	336 (41.2%)	253 (26.3%)	
Pedestrian impairment (Ped_Impair)			0.003	Pedestrian impairment (Ped_Impair)			< 0.001
No	804 (61.5%)	247 (52.6%)		No	598 (73.4%)	453 (47.0%)	
Unknown	388 (29.7%)	173 (36.8%)		Unknown	195 (23.9%)	366 (38.0%)	
Yes	116 (8.87%)	50 (10.6%)		Yes	22 (2.70%)	144 (15.0%)	
Pedestrian Age (Ped_Age)			0.492	Pedestrian Age (Ped_Age)			< 0.001
Middle	709 (54.2%)	254 (54.0%)		Middle	344 (42.2%)	619 (64.3%)	
Old	98 (7.49%)	27 (5.74%)		Old	79 (9.69%)	46 (4.78%)	
Unknown	51 (3.90%)	23 (4.89%)		Unknown	32 (3.93%)	42 (4.36%)	
Young	450 (34.4%)	166 (35.3%)		Young	360 (44.2%)	256 (26.6%)	
Pedestrian Condition (Ped_cond)			0.013	Pedestrian Condition (Ped_cond)			< 0.001
Distracted/Inattentive	541 (41.4%)	168 (35.7%)		Distracted/Inattentive	394 (48.3%)	315 (32.7%)	
Impaired	85 (6.50%)	41 (8.72%)		Impaired	17 (2.09%)	109 (11.3%)	
Normal	280 (21.4%)	87 (18.5%)		Normal	198 (24.3%)	169 (17.5%)	
Other	402 (30.7%)	174 (37.0%)		Other	206 (25.3%)	370 (38.4%)	
Pedestrian State (Ped_State)			0.032	Pedestrian State (Ped_State)			0.453
Louisiana (LA)	1198 (91.6%)	414 (88.1%)		Louisiana (LA)	744 (91.3%)	868 (90.1%)	
Non-Louisiana (Non-LA)	110 (8.41%)	56 (11.9%)		Non-Louisiana (Non-LA)	71 (8.71%)	95 (9.87%)	
Pedestrian Gender (Ped_Gen)			0.944	Pedestrian Gender (Ped_Gen)			0.307
Female (F)	458 (35.0%)	161 (34.3%)		Female (F)	294 (36.1%)	325 (33.7%)	
Male (M)	818 (62.5%)	298 (63.4%)		Male (M)	498 (61.1%)	618 (64.2%)	
Unknown	32 (2.45%)	11 (2.34%)		Unknown	23 (2.82%)	20 (2.08%)	
Pedestrian Race (Ped_Rac)			0.899	Pedestrian Race (Ped_Rac)			0.013
Afri-Amcn	698 (53.4%)	245 (52.1%)		Afri-Amcn	443 (54.4%)	500 (51.9%)	
Caucasian	529 (40.4%)	195 (41.5%)		Caucasian	309 (37.9%)	415 (43.1%)	

(*continued*).

Table 2. Continued.

	Weekday (N = 1308)	Weekend (N = 470)	*p*.overall		Daylight (N = 815)	Nighttime (N = 963)	*p*.overall
Other	81 (6.19%)	30 (6.38%)		Other	63 (7.73%)	48 (4.98%)	
Season			0.338	Season			< 0.001
Autumn	349 (26.7%)	107 (22.8%)		Autumn	191 (23.4%)	265 (27.5%)	
Spring	359 (27.4%)	138 (29.4%)		Spring	262 (32.1%)	235 (24.4%)	
Summer	247 (18.9%)	99 (21.1%)		Summer	169 (20.7%)	177 (18.4%)	
Winter	353 (27.0%)	126 (26.8%)		Winter	193 (23.7%)	286 (29.7%)	
Access			0.454	Access			0.134
Full Control	42 (3.21%)	22 (4.68%)		Full Control	23 (2.82%)	41 (4.26%)	
No Control	1145 (87.5%)	408 (86.8%)		No Control	725 (89.0%)	828 (86.0%)	
Other	10 (0.76%)	2 (0.43%)		Other	3 (0.37%)	9 (0.93%)	
Partial Control	111 (8.49%)	38 (8.09%)		Partial Control	64 (7.85%)	85 (8.83%)	
Align			0.873	Align			0.037
Curve	27 (2.06%)	11 (2.34%)		Curve	20 (2.45%)	18 (1.87%)	
Other	29 (2.22%)	9 (1.91%)		Other	10 (1.23%)	28 (2.91%)	
Straight	1252 (95.7%)	450 (95.7%)		Straight	785 (96.3%)	917 (95.2%)	
Heat and Run (HAR)			<0.001	Heat and Run (HAR)			0.006
No	1108 (84.7%)	364 (77.4%)		No	697 (85.5%)	775 (80.5%)	
Yes	200 (15.3%)	106 (22.6%)		Yes	118 (14.5%)	188 (19.5%)	
Lighting			<0.001	Day of Week (DOW)			< 0.001
Daylight	633 (48.4%)	182 (38.7%)		Weekday	633 (77.7%)	675 (70.1%)	
Nighttime	675 (51.6%)	288 (61.3%)		Weekend	182 (22.3%)	288 (29.9%)	
Location (Loc)			0.034	Location (Loc)			< 0.001
Business	949 (72.6%)	310 (66.0%)		Business	539 (66.1%)	720 (74.8%)	
Industrial	23 (1.76%)	14 (2.98%)		Industrial	14 (1.72%)	23 (2.39%)	
Other	82 (6.27%)	32 (6.81%)		Other	49 (6.01%)	65 (6.75%)	
Residential	254 (19.4%)	114 (24.3%)		Residential	213 (26.1%)	155 (16.1%)	
Crash Type (CrashType1)			0.974	Crash Type (CrashType1)			< 0.001
No Collision with Motor Vehicle (NCWMV)	1028 (78.6%)	374 (79.6%)		No Collision with Motor Vehicle (NCWMV)	591 (72.5%)	811 (84.2%)	

(*continued*).

Table 2. Continued.

	Weekday (N = 1308)	Weekend (N = 470)	p.overall		Daylight (N = 815)	Nighttime (N = 963)	p.overall
Other	252 (19.3%)	86 (18.3%)		Other	197 (24.2%)	141 (14.6%)	
RearEnd	15 (1.15%)	5 (1.06%)		RearEnd	13 (1.60%)	7 (0.73%)	
Sideswipe	13 (0.99%)	5 (1.06%)		Sideswipe	14 (1.72%)	4 (0.42%)	
Road Realignment (RoadRel)			0.249	Road Realignment (RoadRel)			0.799
OnRoadway	1273 (97.3%)	461 (98.1%)		OnRoadway	797 (97.8%)	937 (97.3%)	
Other	10 (0.76%)	5 (1.06%)		Other	6 (0.74%)	9 (0.93%)	
Shoulder	25 (1.91%)	4 (0.85%)		Shoulder	12 (1.47%)	17 (1.77%)	
Road Type (RoadType)			0.010	Road Type (RoadType)			0.579
OneWay	104 (7.95%)	59 (12.6%)		OneWay	81 (9.94%)	82 (8.52%)	
Other	11 (0.84%)	5 (1.06%)		Other	7 (0.86%)	9 (0.93%)	
TwoWay	1193 (91.2%)	406 (86.4%)		TwoWay	727 (89.2%)	872 (90.6%)	
Weather			0.155	Weather			0.906
Adverse	267 (20.4%)	81 (17.2%)		Adverse	161 (19.8%)	187 (19.4%)	
Clear	1041 (79.6%)	389 (82.8%)		Clear	654 (80.2%)	776 (80.6%)	
Pre-occupancy (PreOcc)			0.204	Pre-occupancy (PreOcc)			0.492
No	1019 (77.9%)	352 (74.9%)		No	635 (77.9%)	736 (76.4%)	
Yes	289 (22.1%)	118 (25.1%)		Yes	180 (22.1%)	227 (23.6%)	
Driver Alcohol/Drug Impairment (DrAlcDrug)			<0.001	Driver Alcohol/Drug Impairment (DrAlcDrug)			< 0.001
No	1041 (79.6%)	328 (69.8%)		No	678 (83.2%)	691 (71.8%)	
Unknown	228 (17.4%)	104 (22.1%)		Unknown	127 (15.6%)	205 (21.3%)	
Yes	39 (2.98%)	38 (8.09%)		Yes	10 (1.23%)	67 (6.96%)	
Driver Age (DrAge)			0.642	Driver Age (DrAge)			0.084
Middle	781 (59.7%)	269 (57.2%)		Middle	475 (58.3%)	575 (59.7%)	
Old	144 (11.0%)	56 (11.9%)		Old	108 (13.3%)	92 (9.55%)	
Unknown	153 (11.7%)	64 (13.6%)		Unknown	92 (11.3%)	125 (13.0%)	
Young	230 (17.6%)	81 (17.2%)		Young	140 (17.2%)	171 (17.8%)	
Driver Condition (DrCond)			<0.001	Driver Condition (DrCond)			< 0.001
Distracted/Inattentive	130 (9.94%)	38 (8.09%)		Distracted/Inattentive	86 (10.6%)	82 (8.52%)	
Impaired	8 (0.61%)	14 (2.98%)		Impaired	3 (0.37%)	19 (1.97%)	
Normal	973 (74.4%)	323 (68.7%)		Normal	615 (75.5%)	681 (70.7%)	
Other	197 (15.1%)	95 (20.2%)		Other	111 (13.6%)	181 (18.8%)	
Diver Distraction (DrDistract)			0.005	Diver Distraction (DrDistract)			0.012
No	972 (74.3%)	317 (67.4%)		No	615 (75.5%)	674 (70.0%)	

(*continued*).

Table 2. Continued.

	Weekday (N = 1308)	Weekend (N = 470)	*p*.overall		Daylight (N = 815)	Nighttime (N = 963)	*p*.overall
Yes	336 (25.7%)	153 (32.6%)		Yes	200 (24.5%)	289 (30.0%)	
Driver Race (DrRace)			0.011	Driver Race (DrRace)			0.164
Afri-Amcn	534 (40.8%)	222 (47.2%)		Afri-Amcn	334 (41.0%)	422 (43.8%)	
Caucasian	585 (44.7%)	173 (36.8%)		Caucasian	367 (45.0%)	391 (40.6%)	
Other	189 (14.4%)	75 (16.0%)		Other	114 (14.0%)	150 (15.6%)	
Driver Gender (DrGen)			0.115	Driver Gender (DrGen)			0.008
Female (F)	517 (39.5%)	163 (34.7%)		Female (F)	342 (42.0%)	338 (35.1%)	
Male (M)	648 (49.5%)	244 (51.9%)		Male (M)	390 (47.9%)	502 (52.1%)	
Unknown	143 (10.9%)	63 (13.4%)		Unknown	83 (10.2%)	123 (12.8%)	
Posted Speed Limit (PSL)			0.060	Posted Speed Limit (PSL)			< 0.001
25 MPH or Lower	358 (27.4%)	152 (32.3%)		25 MPH or Lower	312 (38.3%)	198 (20.6%)	
30-45 MPH	796 (60.9%)	255 (54.3%)		30-45 MPH	435 (53.4%)	616 (64.0%)	
50-60 MPH'	131 (10.0%)	50 (10.6%)		50-60 MPH'	58 (7.12%)	123 (12.8%)	
65 MPH and Above	23 (1.76%)	13 (2.77%)		65 MPH and Above	10 (1.23%)	26 (2.70%)	
Prior Movement (PriorMov)			0.154	Prior Movement (PriorMov)			< 0.001
LeftTurn	68 (5.20%)	18 (3.83%)		LeftTurn	55 (6.75%)	31 (3.22%)	
Other	257 (19.6%)	79 (16.8%)		Other	189 (23.2%)	147 (15.3%)	
Stopped	17 (1.30%)	3 (0.64%)		Stopped	12 (1.47%)	8 (0.83%)	
StraightAhead	966 (73.9%)	370 (78.7%)		StraightAhead	559 (68.6%)	777 (80.7%)	
Traffic Control Device (TCD)			0.111	Traffic Control Device (TCD)			< 0.001
DashedLine	589 (45.0%)	186 (39.6%)		DashedLine	310 (38.0%)	465 (48.3%)	
NoControl	299 (22.9%)	127 (27.0%)		NoControl	240 (29.4%)	186 (19.3%)	
Other	296 (22.6%)	104 (22.1%)		Other	193 (23.7%)	207 (21.5%)	
SolidLine	124 (9.48%)	53 (11.3%)		SolidLine	72 (8.83%)	105 (10.9%)	
Vehicle Condition (VehCond)			0.223	Vehicle Condition (VehCond)			0.005
Defective	9 (0.69%)	2 (0.43%)		Defective	1 (0.12%)	10 (1.04%)	
NoDefects	1094 (83.6%)	379 (80.6%)		NoDefects	695 (85.3%)	778 (80.8%)	
Unknown	205 (15.7%)	89 (18.9%)		Unknown	119 (14.6%)	175 (18.2%)	
Vehicle Type (VehType)			0.046	Vehicle Type (VehType)			0.785
Bus/Truck	332 (25.4%)	106 (22.6%)		Bus/Truck	204 (25.0%)	234 (24.3%)	
Car	606 (46.3%)	236 (50.2%)		Car	381 (46.7%)	461 (47.9%)	
Other	60 (4.59%)	33 (7.02%)		Other	39 (4.79%)	54 (5.61%)	
SUV	310 (23.7%)	95 (20.2%)		SUV	191 (23.4%)	214 (22.2%)	
Visual Obstruction (VisObs)			0.020	Visual Obstruction (VisObs)			0.574
No	1002 (76.6%)	334 (71.1%)		No	618 (75.8%)	718 (74.6%)	
Yes	306 (23.4%)	136 (28.9%)		Yes	197 (24.2%)	245 (25.4%)	

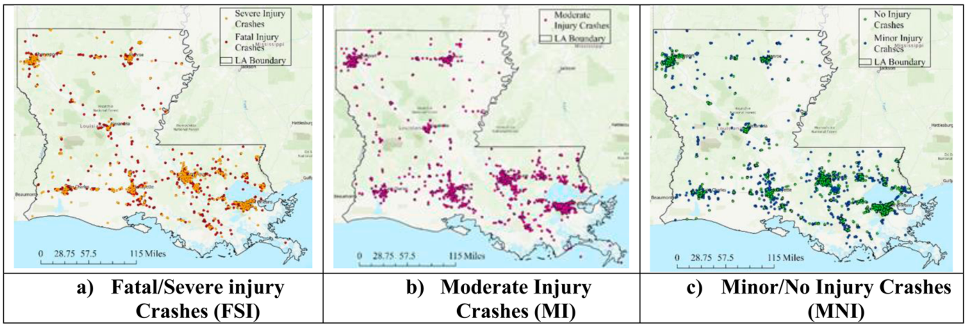


Figure 3. Crashes Locations based on Crash Severity: (a) Fatal/Severe injury Crashes (FSI) (b) Moderate Injury Crashes (MI) (c) Minor/No Injury Crashes (MNI).

provides an overview of the temporal cluster analysis output. As mentioned, the temporal clusters are based on the days of the week and the time of day. The days of the week have been categorized into weekdays and weekends, and the time of day has been categorized into daytime and nighttime. The output of the analyses is listed in the tables as shown in this figure. It should be noted that the top 15 rules among the generated association rules are portrayed in the tables for explanation. Finally, all association rules for each category were visualized using scattered plots, network graphs, and ballon plots.

4.1. Selection of support and confidence threshold

To derive meaningful rules accurately representing factors influencing each crash severity level, it's crucial to thoughtfully choose suitable thresholds for minimum support and confidence. Opting for excessively low thresholds may lead to an overwhelming number of redundant rules, making interpretation challenging. On the other hand, setting thresholds too high might significantly reduce the number of rules, potentially overlooking important ones and obscuring inherent relationships among different item sets. Hence, determining optimal thresholds is paramount, acknowledging the absence of a universally applicable method for this.

To tackle this challenge, our study employed a trial-and-error approach, adopting a grid search method as recommended by a previous study (Das, Ahmed, and Ghasemzadeh 2019). This method, widely used in data mining and optimization, systematically explores various combinations of parameter values to identify the most optimal configuration for a given model. This is comparatively a new method in pedestrian crash data analysis based on rules mining technique. This strategical advancement in data mining techniques can assist to get the most significant rules. Following this strategy, we defined a grid encompassing a range of minimum support and confidence values. Rules generated from each combination were then evaluated based on multiple criteria, including lift and coverage values, engineering judgment, and the number and meaningfulness of generated rules. To ensure a manageable number of rules, we deemed at least 150 rules for each severity level appropriate.

A lift-increase threshold was applied to select rules demonstrating significant contributions from all antecedents, thereby reinforcing overall rule strength. When sorting the rules,

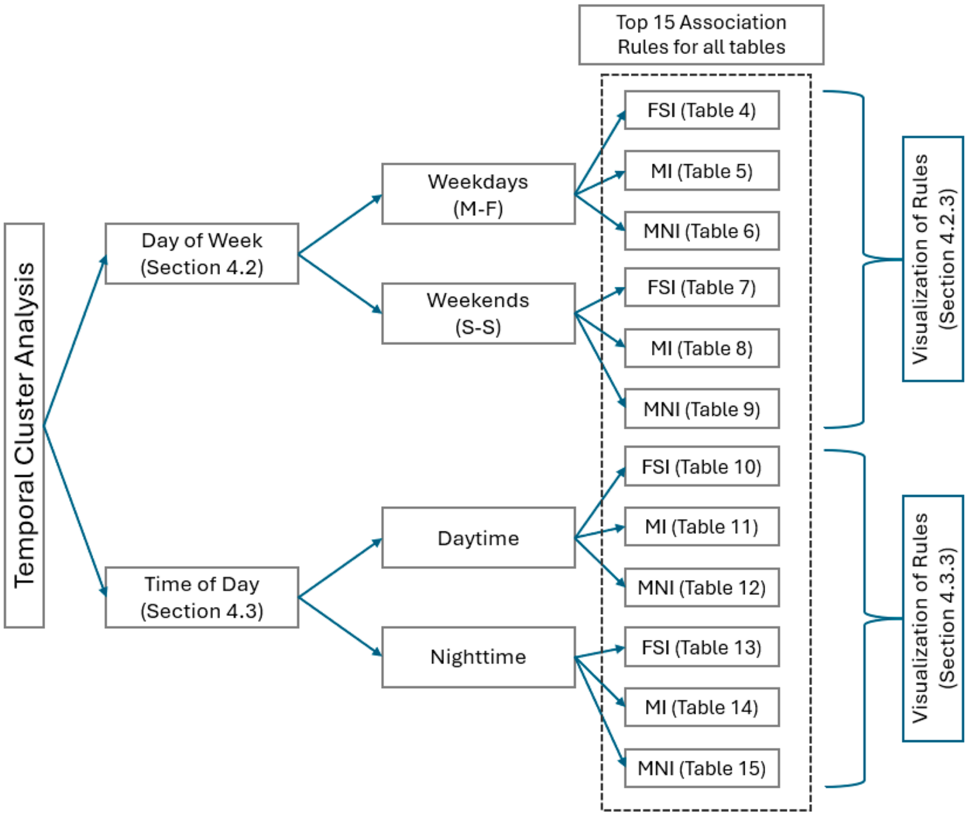


Figure 4. Temporal Cluster Analysis Overview.

preference was given to those with a greater number of antecedents over simpler rules, based on the lift-increase criteria as described in the following equation,

$$\frac{l(A_{n+1} \to B)}{l(A_n \to B)} \geq 1.01$$

Where,

A_{n+1} = antecedent of the rule with $n + 1$ items,

A_{n+1} = antecedent of the rule with n items, and B = consequent of each rule.

The outcomes of the sensitivity analysis are based on the crash data and are presented in Table 3. The sensitivity analysis was conducted to ascertain the optimal support and confidence thresholds based on the criteria mentioned earlier. The grid search method was employed to explore various combinations of support and confidence values for each crash severity category. In fatal/severe crashes on weekdays, initial consideration of support and confidence values 0.001 and 0.01 respectively resulted in the generation of 11,009 rules. As the confidence value increased, the number of rules gradually decreased. Finally, a support value of 0.01 and a confidence value of 0.7 were chosen, yielding a total of 184 rules. Further increment in support and confidence values produced a lower number of rules and eventually resulted in showing null values for association rules. Subsequently, other combinations with the grid were evaluated and the closest number of rules to 150 was found with a combination of support and confidence values 0.0. and 0.7 respectively. A similar process was adopted for finding the optimal support and confidence values for all crash severity levels based on weekends, daytime, and nighttime.

4.2. Temporal clusters by day of week

4.2.1. Weekdays

Table 4 presents the top 15 important rules (sorted by lift values) that contribute to pedestrian-involved Fatal/Severe Injury crashes at midblock during weekdays. The association rules were extracted and organized based on the decreasing lift values.

The rule with the highest lift value FSIWD01 {*DrAlcDrug* = *Yes* + *DrRace* = *Caucasian* → *Severity* = *FSI*}, indicates that the alcohol or drug impairment of the driver is strongly associated with pedestrian fatal/severe injury at midblock during weekdays. This signifies the strong association of impaired drivers with fatal/severe injury crashes. Previous studies have shown that alcohol or drug-impaired drivers tend to violate traffic rules such as speeding, no use of seatbelts, invalid driving licenses, and so on. These violations often result in fatal crashes (Valen et al. 2019). Additionally, this rule had a support value of 0.015, a confidence value of 1, a coverage value of 0.015, and a lift value of 4.025. The support value indicates that 1.5% of pedestrian fatal/severe crashes at midblock during the weekdays are attributed to these factors. The confidence value indicates that 100% of all pedestrian-involved crashes at midblock during weekdays involving Caucasian impaired drivers were fatal/severe injury crashes. The coverage value indicates that 1.5% of the crashes at midblock during weekdays in the dataset are associated with Caucasian-impaired drivers. The lift value indicates that the percentage of pedestrian-involved fatal/severe injury crashes at midblock during weekdays containing this combination is 4.025 times higher than the percentage of such crashes at midblock during the weekdays in the overall dataset. The second most important rule was found to be FSIWD02 {*DrAlcDrug* = *Yes* + *Lighting* = *Nighttime* + *Ped_Race* = *Caucasian* → *Severity* = *FSI*}. This rule signifies the influence of lighting conditions on severe crashes, justifying the research approach of considering the time of the day as one of the temporal clusters. Both support and coverage values for this rule were 0.014. The most common attribute in the top 15 rules was *DrAlcDrug* = *Yes*, which appeared in all rules and indicates that driver alcohol or drug impairment has a high correlation with fatal or severe pedestrian crashes at midblock. The second most common attribute was *TCD* = *DashedLine*, which appeared seven times out of 15 rules. The other frequent attributes were *Lighting* = *Nighttime* and *RoadType* = *TwoWay*. The rule with the highest count of 27 was FSIWD05 {DrAlcDrug = Yes + Lighting = Nighttime + RoadType = TwoWay → Severity = FSI}.

Table 5 lists the top 15 important rules that contribute to pedestrian-involved moderate injury crashes at midblock during weekdays. Based on the lift value, the topmost important rule is MIWD01 {*RoadType* = *OneWay* + *Season* = *Summer* → *Severity* = *MI*}. This rule indicates that the one-way road type and summer season are strongly associated with pedestrian moderate injury crashes at midblock during weekdays. This rule can be justified by the findings of a previous study that showed that no adverse weather conditions or dry conditions of roads often stimulate pedestrian crashes (Sherony and Zhang 2015). Moreover,

Table 3. Sensitivity Analysis to Determine the Optimum Support and Confidence Thresholds .

		Weekdays			Weekends			Daytime			Nighttime		
Support	Confidence	FSI	MI	MNI	FSI	MI	MNI	FSI	MI	MNI	FSI	MI	MNI
0.001	0.01	11,009	15,562	13,159	13,190	15,711	14,116	12,585	17,214	16,360	14,089	18,663	15,540
0.001	0.02	11,009	15,562	13,159	13,190	15,711	14,116	12,585	17,214	16,360	14,089	18,663	15,540
0.001	0.04	11,009	15,562	13,159	13,190	15,711	14,116	12,585	17,214	16,360	14,089	18,663	15,540
0.001	0.06	11,009	15,562	13,159	13,190	15,711	14,116	12,585	17,214	16,360	14,089	18,663	15,540
0.001	0.08	11,009	15,562	13,159	13,190	15,711	14,116	12,585	17,214	16,360	14,089	18,663	15,540
0.001	0.1	11,009	15,562	13,159	13,190	15,711	14,116	12,585	17,214	16,360	14,089	18,663	15,540
0.001	0.2	9,630	15,562	13,159	13,190	15,711	14,116	9,653	17,214	16,360	14,089	18,663	15,540
0.001	0.4	5,640	15,562	12,102	9,436	15,711	11,502	3,790	17,214	16,360	13,080	18,444	10,383
0.001	0.6	2,673	4,217	4,814	4,524	8,945	5,074	1,633	7,536	8,045	6,763	6,679	4,587
0.001	0.8	1,258	1,022	1,622	2,848	4,123	3,143	1,121	3,518	4,247	3,939	3,463	2,603
0.01	0.02	4,679	7,677	5,232	6,537	8,776	6,788	3,179	7,926	6,967	5,572	8,254	5,319
0.01	0.04	4,679	7,677	5,232	6,537	8,776	6,788	3,179	7,926	6,967	5,572	8,254	5,319
0.01	0.06	4,679	7,677	5,232	6,537	8,776	6,788	3,179	7,926	6,967	5,572	8,254	5,319
0.01	0.08	4,679	7,677	5,232	6,537	8,776	6,788	3,179	7,926	6,967	5,572	8,254	5,319
0.01	0.1	4,679	7,677	5,232	6,537	8,776	6,788	3,179	7,926	6,967	5,572	8,254	5,319
0.01	0.2	3,826	7,677	5,232	6,537	8,776	6,788	1,569	7,926	6,967	5,572	8,254	5,319
0.01	**0.4**	1,419	7,677	4,444	3,818	8,776	4,794	**192**	7,926	6,967	4,798	8,037	2,171
0.01	0.6	443	217	624	1,147	3,323	1,063	25	1,069	1,377	1,475	530	422
0.01	**0.61**	412	**153**	542	1,041	2,913	983	19	828	1,180	1,371	412	393
0.01	**0.65**	307	72	362	819	1903	774	13	394	740	1023	**178**	260
0.01	**0.7**	**184**	18	195	693	1,243	623	6	**180**	430	726	85	**153**
0.01	0.75	123	8	102	554	731	439	5	87	334	529	31	76
0.01	0.8	163	1	39	449	420	301	2	31	237	413	19	23
0.1	**0.2**	791	1,573	**191**	580	733	473	0	1,275	535	875	972	68
0.1	**0.4**	129	1,573	75	**174**	733	**172**	0	1,275	535	630	870	0
0.15	**0.42**	0	854	13	7	321	37	0	835	**174**	**167**	627	0
0.15	**0.45**	0	354	3	5	**191**	17	0	820	49	85	275	0
0.1	0.6	0	0	3	0	1	1	0	1	14	49	0	0
0.1	0.8	0	0	0	0	0	0	0	0	0	0	0	0
0.5	0.6	0	0	0	0	0	0	0	0	0	0	0	0
0.5	0.8	0	0	0	0	0	0	0	0	0	0	0	0

Notes: *Green cells with bold font indicate the support and confidence values, and the number of association rules that are developed based on the optimal support and confidence values*

Table 4. Top 15 Association Rules for Pedestrian Fatal/ Severe Injury Crashes at Midblock During Weekdays Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
FSIWD01	DrAlcDrug = Yes + DrRace = Caucasian	Severity = FSI	0.015	1.000	0.015	4.025	20
FSIWD02	DrAlcDrug = Yes + Lighting = Nighttime + Ped_Race = Caucasian	Severity = FSI	0.014	1.000	0.014	4.025	18
FSIWD03	DrAlcDrug = Yes + Ped_Age = Middle + Ped_Race = Caucasian	Severity = FSI	0.011	1.000	0.011	4.025	15
FSIWD04	DrAlcDrug = Yes + PSL = 30-45MPH + TCD = DashedLine	Severity = FSI	0.011	1.000	0.011	4.025	15
FSIWD05	DrAlcDrug = Yes + Lighting = Nighttime + RoadType = TwoWay	Severity = FSI	0.021	0.964	0.021	3.881	27
FSIWD06	DrAlcDrug = Yes + DrDistract = No + Lighting = Nighttime	Severity = FSI	0.020	0.963	0.021	3.876	26
FSIWD07	CrashType1 = No crash with motor vehicle (NCWMV) + DrAlcDrug = Yes + TCD = DashedLine	Severity = FSI	0.018	0.958	0.018	3.857	23
FSIWD08	DrAlcDrug = Yes + Ped_Race = Caucasian	Severity = FSI	0.016	0.955	0.017	3.842	21
FSIWD09	DrAlcDrug = Yes + Lighting = Nighttime + TCD = DashedLine	Severity = FSI	0.016	0.955	0.017	3.842	21
FSIWD10	DrAlcDrug = Yes + DrDistract = No + TCD = DashedLine	Severity = FSI	0.016	0.955	0.017	3.842	21
FSIWD11	DrAlcDrug = Yes + RoadType = TwoWay + TCD = DashedLine	Severity = FSI	0.016	0.955	0.017	3.842	21
FSIWD12	DrAlcDrug = Yes + DrCond = Normal + TCD = DashedLine	Severity = FSI	0.014	0.947	0.015	3.813	18
FSIWD13	DrAlcDrug = Yes + DrDistract = No + Season = Spring	Severity = FSI	0.013	0.944	0.014	3.801	17
FSIWD14	DrAlcDrug = Yes + RoadType = TwoWay + Season = Spring	Severity = FSI	0.013	0.944	0.014	3.801	17
FSIWD15	DrAlcDrug = Yes + Ped_Gen = M + TCD = DashedLine	Severity = FSI	0.013	0.944	0.014	3.801	17

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

one-way roadway provides the drivers more relaxation which may cause careless driving. This rule has a support value of 0.012, which means that 1.2% of pedestrian moderate injury crashes at midblock during weekdays are attributed to the combination of the factors one away roadway type and summer. The lift value 1.938 indicates that the occurrence of pedestrian moderate injury during weekdays with this combination of factors is 1.9 times higher compared to the overall moderate injury crashes on weekdays. The second topmost rule here is {*Ped_Race = Caucasian + PSL = 25MPHorLower + RoadType = OneWay → Severity = MI*}.

Among the top 15 rules, the attribute *RoadType = OneWay* appears most frequently on the antecedent, occurring four times. This portrays that one-way road type significantly contributes to moderate injury crashes during weekdays. The second topmost attribute that appears two times in the antecedent is *Season = Summer*. This means that pedestrian moderate injury crashes at midblock are more likely to occur more frequently during the summer. There are several reasons why one-way roadway type and summer have a strong association with moderate injury crashes at midblock. The rule with the highest count among all the combinations of factors for moderate injury during weekdays is {*PSL = 25MPHorLower + Season = Autumn + TCD = NoControl → Severity = MI*}.

Table 5. Top 15 Association Rules for Pedestrian Moderate Injury Crashes at Midblock During Weekdays Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MIWD01	RoadType = OneWay + Season = Summer	Severity = MI	0.012	0.800	0.015	1.938	16
MIWD02	Ped_Race = Caucasian + PSL = 25MPHorLower + RoadType = OneWay	Severity = MI	0.011	0.778	0.014	1.884	14
MIWD03	DrAge = Young + Ped_Cond = Distracted/Inattentive + Season = Summer	Severity = MI	0.011	0.778	0.014	1.884	14
MIWD04	Ped_Cond = Normal + Season = Autumn + VehType = SUV	Severity = MI	0.011	0.778	0.014	1.884	14
MIWD05	DrAge = Middle + DrGen = M + PriorMov = LeftTurn	Severity = MI	0.012	0.762	0.016	1.846	16
MIWD06	CrashType1 = NCWMV + RoadType = OneWay + TCD = Other	Severity = MI	0.012	0.762	0.016	1.846	16
MIWD07	DrGen = M + Ped_Impair = No + PriorMov = LeftTurn	Severity = MI	0.014	0.750	0.018	1.817	18
MIWD08	DrRace = Afri-Amcn + PreOcc = Yes + TCD = Other	Severity = MI	0.014	0.750	0.018	1.817	18
MIWD09	Lighting = Nighttime + Ped_Impair = No + PriorMov = LeftTurn	Severity = MI	0.011	0.737	0.015	1.785	14
MIWD10	PSL = 25MPHorLower + Season = Autumn + TCD = NoControl	Severity = MI	0.025	0.717	0.035	1.738	33
MIWD11	Access = NoControl + RoadType = OneWay + TCD = Other	Severity = MI	0.011	0.714	0.016	1.730	15
MIWD12	Ped_Gen = F + Ped_State = LA + PriorMov = LeftTurn	Severity = MI	0.013	0.708	0.018	1.716	17
MIWD13	DrCond = Normal + DrGen = M + PriorMov = LeftTurn	Severity = MI	0.013	0.708	0.018	1.716	17
MIWD14	Ped_Cond = Normal + Season = Autumn + TCD = NoControl	Severity = MI	0.013	0.708	0.018	1.716	17
MIWD15	Season = Autumn + TCD = NoControl + VehType = SUV	Severity = MI	0.013	0.708	0.018	1.716	17

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

Table 6 shows the top 15 rules that include the combination of factors responsible for pedestrian minor/no injury crashes at midblock during weekdays. The most important rule with the highest lift value is {*Align = Straight + CrashType1 = Other + RoadType = TwoWay → Severity = MNI*}. This rule illustrates that the combination of attributes *Align = Straight*, *CrashType1 = Other*, and *RoadType = TwoWay* has the strongest association with pedestrian minor/no injury crashes at midblock during the weekdays. The support value 0.102 indicates that 10.2% of pedestrian minor/no injury crashes at midblock during weekdays are attributed to the combination of factors that include straight road alignment, other crash types, and two-way roadways. In these attributes, the other crash types include head-on crashes, left turns, right turns, right angles, and other crashes. The confidence value for this rule is 0.605. This illustrates that among all the pedestrian minor/no injury crashes at midblock during weekdays, 60.5% of the crashes exhibit the attributes in this rule. The coverage value 0.168 indicates that 16.8% of minor/no injury crashes during weekdays are associated with straight road alignment, left/right turn, right angle, and other crash types. The lift value 1.785 indicates that the occurrence of the factors with this combination in the pedestrian minor/no injury crashes at midblock during the weekend is 1.785 times higher among others. The second topmost rule is {*Align = Straight + CrashType1 = Other → Severity = MNI*}.

Table 6. Top 15 Association Rules for Pedestrian Minor/ No Injury at Midblock During Weekdays Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MNIWD01	Align = Straight + CrashType1 = Other + RoadType = TwoWay	Severity = MNI	0.102	0.605	0.168	1.785	133
MNIWD02	Align = Straight + CrashType1 = Other	Severity = MNI	0.111	0.602	0.184	1.776	145
MNIWD03	CrashType1 = Other + RoadType = TwoWay	Severity = MNI	0.105	0.601	0.174	1.774	137
MNIWD04	CrashType1 = Other	Severity = MNI	0.115	0.599	0.193	1.769	151
MNIWD05	Ped_Impair = No + PSL = 25MPHorLower	Severity = MNI	0.102	0.473	0.216	1.398	134
MNIWD06	PreOcc = No + PSL = 25MPHorLower	Severity = MNI	0.103	0.472	0.219	1.394	135
MNIWD07	Lighting = Daylight + Loc = Business + Ped_Impair = No	Severity = MNI	0.114	0.470	0.242	1.388	149
MNIWD08	Access = NoControl + PSL = 25MPHorLower	Severity = MNI	0.114	0.469	0.243	1.383	149
MNIWD09	PSL = 25MPHorLower	Severity = MNI	0.128	0.466	0.274	1.377	167
MNIWD10	Access = NoControl + Lighting = Daylight + Ped_Impair = No	Severity = MNI	0.150	0.461	0.325	1.362	196
MNIWD11	Lighting = Daylight + Ped_Impair = No + PreOcc = No	Severity = MNI	0.131	0.461	0.285	1.362	172
MNIWD12	Lighting = Daylight + Ped_Impair = No + RoadType = TwoWay	Severity = MNI	0.151	0.460	0.327	1.359	197
MNIWD13	Lighting = Daylight + Ped_Impair = No + VisObs = No	Severity = MNI	0.129	0.457	0.283	1.349	169
MNIWD14	Lighting = Daylight + Ped_Impair = No	Severity = MNI	0.164	0.456	0.359	1.347	214
MNIWD15	Access = NoControl + Lighting = Daylight + PreOcc = No	Severity = MNI	0.154	0.451	0.341	1.331	201

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

The attribute that occurs the most in the antecedent is *Lighting = Daylight*, which *is* repeated seven times representing a high association of daylight with minor/no injury crashes. The frequent occurrence of these attributes indicates that daylight and head-on, left or right turns, and right-angle crash types have a strong association with pedestrian minor/no injury at midblock during weekdays. The rule with the highest count of 214 is {*Lighting = Daylight + Ped_Impair = No → Severity = MNI*}.

4.2.2. Weekends

Table 7 shows the top 15 association rules for pedestrian fatal/severe injury crashes at midblock during the weekend based on the lift. The top rule {*CrashType1 = No crash with motor vehicle or NCWMV + Lighting = Nighttime + VisObs = No → Severity = FSI*} signifies that if the crash is a no crash with motorized vehicles (*CrashType1 = NCWMV*) and it occurs during nighttime (*Lighting = Nighttime*) having no visual obstructions (*VisObs = No*) then it is highly likely that the crash severity would result in a pedestrian fatal/severe injury. Even if the crash doesn't involve a motorized vehicle, a pedestrian could be at risk of severe injury or fatality if, for example, they collide with a bicycle, skateboard, scooter, or other non-motorized mode of transportation. During the weekend, some walkways can be less crowded, allowing non-motorized mode users to travel at higher speed. The impact forces, particularly if the crash occurs at higher speeds, can still cause harm. Some studies showed an association between pedestrian fatality and non-motorized vehicles (Lee, Abdel-Aty,

Table 7. Top 15 Association Rules for Pedestrian Fatal/ Severe Injury Crashes at Midblock During Weekend Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
FSIWE01	CrashType1 = NCWMV + Lighting = Nighttime + VisObs = No	Severity = FSI	0.101	0.313	0.323	1.628	30
FSIWE02	Lighting = Nighttime + RoadType = TwoWay + VisObs = No	Severity = FSI	0.101	0.306	0.330	1.595	30
FSIWE03	Ped_Age = Middle + RoadType = TwoWay + VehCond = NoDefects	Severity = FSI	0.101	0.303	0.333	1.579	30
FSIWE04	CrashType1 = NCWMV + Ped_Age = Middle + RoadType = TwoWay	Severity = FSI	0.101	0.300	0.337	1.563	30
FSIWE05	CrashType1 = NCWMV + Lighting = Nighttime + RoadType = TwoWay	Severity = FSI	0.114	0.298	0.384	1.554	34
FSIWE06	Align = Straight + Ped_Age = Middle + VehCond = NoDefects	Severity = FSI	0.108	0.294	0.367	1.530	32
FSIWE07	DrGen = M + PriorMov = StraightAhead + RoadType = TwoWay	Severity = FSI	0.101	0.291	0.347	1.518	30
FSIWE08	Align = Straight + CrashType1 = NCWMV + Ped_Age = Middle	Severity = FSI	0.104	0.284	0.367	1.482	31
FSIWE09	CrashType1 = NCWMV + Ped_Age = Middle + RoadRel = OnRoadway	Severity = FSI	0.108	0.283	0.380	1.476	32
FSIWE10	Lighting = Nighttime + RoadType = TwoWay + Weather = Clear	Severity = FSI	0.108	0.283	0.380	1.476	32
FSIWE11	Ped_Age = Middle + RoadRel = OnRoadway + RoadType = TwoWay	Severity = FSI	0.114	0.281	0.407	1.464	34
FSIWE12	CrashType1 = NCWMV + Ped_Age = Middle	Severity = FSI	0.108	0.281	0.384	1.463	32
FSIWE13	Ped_Age = Middle + RoadRel = OnRoadway + VehCond = NoDefects	Severity = FSI	0.108	0.281	0.384	1.463	32
FSIWE14	Ped_Age = Middle + RoadType = TwoWay	Severity = FSI	0.114	0.279	0.411	1.452	34
FSIWE15	Access = NoControl + Lighting = Nighttime + RoadType = TwoWay	Severity = FSI	0.114	0.279	0.411	1.452	34

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

and Jiang 2015). Additionally, poor lighting conditions can also contribute to an increased pedestrian fatality rate (Chen, Ma, and Chen 2019).

The lift value 1.628 means that with a combination of these attributes, the chances of fatal/severe injury crashes at midblock during the weekend is 1.628 times higher than other combinations in the dataset for similar types of crashes. The support value indicates that 10.1% of fatal/severe injury crashes at midblock during the weekend include this combination of attributes. The second topmost rule {*Lighting = Nighttime + RoadType = TwoWay + VisObs = No → Severity = FSI*} illustrates that nighttime is an influencing factor for fatal or severe injury crashes. The most occurred attribute is *RoadType = TwoWay* which occurred 7 times in the antecedent. The rules with the highest count are FSIWE05, FSIWE11, FSIWE14, and FSIWE15. All these rules have 34 counts.

Table 8 illustrates the top 15 rules for pedestrian moderate injury crashes at midblock during the weekend. The association rules are enlisted based on the decreasing lift values. The topmost rule is {*CrashType1 = NCWMV + Ped_Impair = No + PreOcc = No → Severity = MI*}. This rule indicates that if the crash is not with a motorized vehicle, if the pedestrian is not impaired, and the roadway is not previously occupied, then it is highly likely that the crash severity would result in pedestrian moderate injury crashes at midblock during

Table 8. Top 15 Association Rules for Pedestrian Moderate Injury Crashes at Midblock During Weekends Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MIWE01	CrashType1 = NCWMV + Ped_Impair = No + PreOcc = No	Severity = MI	0.166	0.553	0.300	1.307	78
MIWE02	Align = Straight + CrashType1 = NCWMV + Ped_Impair = No	Severity = MI	0.198	0.531	0.372	1.255	93
MIWE03	CrashType1 = NCWMV + Ped_Impair = No	Severity = MI	0.206	0.530	0.389	1.252	97
MIWE04	CrashType1 = NCWMV + Ped_Race = Afri-Amcn + PreOcc = No	Severity = MI	0.164	0.524	0.313	1.237	77
MIWE05	CrashType1 = NCWMV + Ped_Cond = Distracted/Inattentive	Severity = MI	0.151	0.522	0.289	1.233	71
MIWE06	Access = NoControl + Ped_Age = Young	Severity = MI	0.164	0.510	0.321	1.204	77
MIWE07	DrAlcDrug = No + Ped_Race = Afri-Amcn + RoadType = TwoWay	Severity = MI	0.157	0.507	0.311	1.197	74
MIWE08	Ped_Race = Afri-Amcn + PreOcc = No + RoadType = TwoWay	Severity = MI	0.170	0.506	0.336	1.196	80
MIWE09	Ped_Gen = M + RoadRel = OnRoadway + VehType = Car	Severity = MI	0.157	0.503	0.313	1.189	74
MIWE10	CrashType1 = NCWMV + HAR = No + Ped_Race = Afri-Amcn	Severity = MI	0.157	0.503	0.313	1.189	74
MIWE11	CrashType1 = NCWMV + RoadRel = OnRoadway + VehType = Car	Severity = MI	0.183	0.503	0.364	1.188	86
MIWE12	CrashType1 = NCWMV + VehType = Car	Severity = MI	0.187	0.500	0.374	1.181	88
MIWE13	CrashType1 = NCWMV + DrAge = Middle + DrAlcDrug = No	Severity = MI	0.179	0.500	0.357	1.181	84
MIWE14	DrDistract = No + Ped_Race = Afri-Amcn + RoadType = TwoWay	Severity = MI	0.155	0.497	0.313	1.173	73
MIWE15	Ped_Gen = M + Ped_Impair = No + RoadRel = OnRoadway	Severity = MI	0.153	0.497	0.309	1.173	72

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

the weekend. This rule has a support value of 0.166, a confidence value of 0.553, a coverage value of 0.3, and a lift value of 1.307.

The support value shows that 16.6% of the pedestrian moderate injury crashes that occur at midblock during the weekend are attributed to these factors. The confidence value means that 55.3% of all the pedestrian moderate injury crashes that occur at midblock during the weekend involve pedestrians with no impairment. The coverage value shows that 30% of crashes at midblock during the weekend in the dataset are associated with non-impaired pedestrians. The lift value indicates that the percentage of pedestrian-involved minor injury crashes at midblock during the weekend containing this combination is 1.3 times higher than the percentage of such crashes at midblock during the weekend in the overall dataset. The count for this rule is 78. The second topmost important rule is {*Align = Straight + CrashType1 = NCWMV + Ped_Impair = No → Severity = MI*}. The support, confidence, coverage, and lift values for this rule are 0.198, 0.531, 0.372, and 1.255, respectively. The most common attribute among the top 15 rules is *CrashType1 = NCWMV*, which appeared 8 times in the antecedent. The rule MIWE03 has the highest count, and it includes the combination of attributes *CrashType1 = NCWMV + Ped_Impair = No*.

Table 9 lists the 15 rules for pedestrian minor/no-injury crashes at midblock during the weekend. The top rule is {*CrashType1 = Other → Severity = MNI*}. The support, confidence, coverage, and lift values for this combination are 0.113, 0.616, 0.183, and 1.984,

Table 9. Top 15 Association Rules for Pedestrian Minor/ No Injury Crashes at Midblock During Weekend Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MNIWE01	CrashType1 = Other	Severity = MNI	0.113	0.616	0.183	1.984	53
MNIWE02	DrDistract = No + Loc = Business + Ped_Impair = No	Severity = MNI	0.113	0.495	0.228	1.595	53
MNIWE03	DrCond = Normal + Ped_Cond = Distracted/Inattentive + Ped_Impair = No	Severity = MNI	0.111	0.491	0.226	1.579	52
MNIWE04	Loc = Business + Ped_Impair = No + VisObs = No	Severity = MNI	0.111	0.491	0.226	1.579	52
MNIWE05	DrCond = Normal + Loc = Business + Ped_Impair = No	Severity = MNI	0.117	0.487	0.240	1.567	55
MNIWE06	Ped_Cond = Distracted/Inattentive + Ped_Impair = No + Weather = Clear	Severity = MNI	0.102	0.485	0.211	1.561	48
MNIWE07	Ped_Cond = Distracted/Inattentive + Ped_Impair = No + VehCond = NoDefects	Severity = MNI	0.117	0.482	0.243	1.553	55
MNIWE08	DrAlcDrug = No + Ped_Cond = Distracted/Inattentive + Ped_Impair = No	Severity = MNI	0.115	0.482	0.238	1.552	54
MNIWE09	DrAlcDrug = No + Loc = Business + Ped_Impair = No	Severity = MNI	0.128	0.480	0.266	1.545	60
MNIWE10	HAR = No + Loc = Business + Ped_Impair = No	Severity = MNI	0.128	0.476	0.268	1.533	60
MNIWE11	HAR = No + Ped_Cond = Distracted/Inattentive + Ped_Impair = No	Severity = MNI	0.115	0.474	0.243	1.525	54
MNIWE12	Access = NoControl + Ped_Cond = Distracted/Inattentive + Ped_Impair = No	Severity = MNI	0.113	0.473	0.238	1.523	53
MNIWE13	DrDistract = No + Ped_Cond = Distracted/Inattentive + Ped_Impair = No	Severity = MNI	0.111	0.473	0.234	1.522	52
MNIWE14	DrAlcDrug = No + Ped_Impair = No + Weather = Clear	Severity = MNI	0.164	0.472	0.347	1.521	77
MNIWE15	Ped_Cond = Distracted/Inattentive + Ped_Impair = No + VisObs = No	Severity = MNI	0.102	0.466	0.219	1.500	48

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

respectively. This rule stands top among all the association mining rules for minor/no injury crashes at midblock during the weekend based on the highest lift value. The lift value 1.984 illustrates that the percentage of pedestrian-involved minor/no-injury crashes at midblock during the weekend containing this combination is 1.984 times higher than that of such crashes at midblock during the weekend in the overall dataset. The second topmost rule is {DrDistract = No + Loc = Business + Ped_Impair = No → *Severity = MNI*}. This rule implies that the roadways in commercial areas are susceptible to pedestrian minor/no-injury crashes when both the driver and pedestrian are in fully conscious condition. The count and support values of the first and second most important rules are similar. The rule with the highest count is {*DrAlcDrug = No + Ped_Impair = No + Weather = Clear → Severity = MNI*}. The most common attribute among these top 15 rules is *Ped_Impair = No* appearing 14 times in the antecedent. The second most common attribute is *Ped_Cond = Distracted/Inattentive,* which appeared eight times in the antecedent.

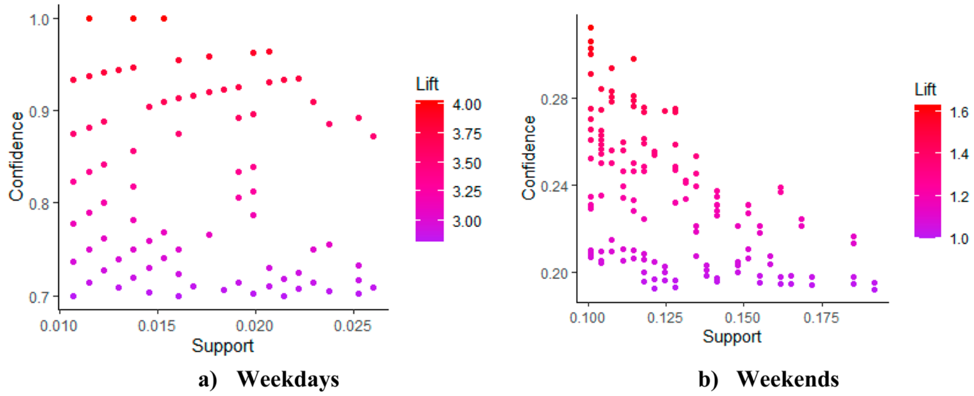


Figure 5. Scatterplot Plots of Association Rules for Fatal/ Severe Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

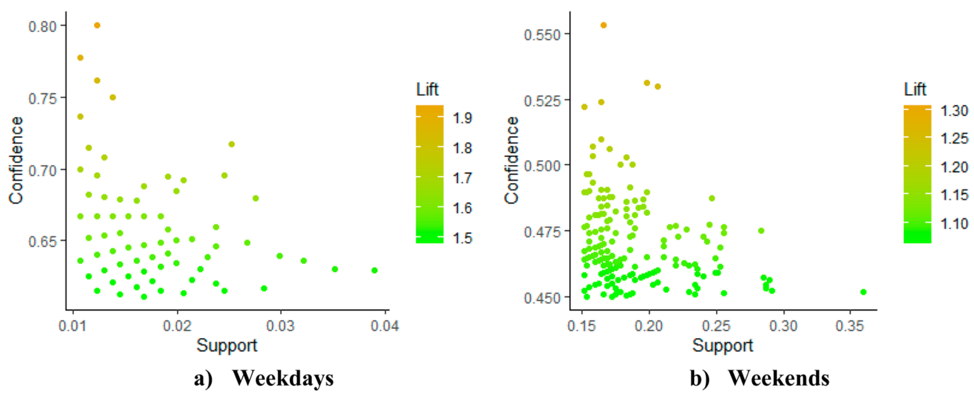


Figure 6. Scatterplot Balloon Plots of Association Rules for Moderate Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

4.2.3. Visualization of the rules

To gain a deeper understanding of the association rules for pedestrian crashes at midblock, we employed the 'arulesViz' package in the R software to visualize the generated rules pertaining to different severity levels resulting from pedestrian crashes. Figure 5, Figure 6, and Figure 7 demonstrate the scatter diagrams that show the association rules for different crash severity levels based on the day of the week. The X-axis defines the support values, and the Y-axis defines the confidence values. The color intensity of the points in the scatter plots defines the lift values. The points in red indicate higher lift values, and the point in purple indicates lower lift values. Figure 5(a,b) represent the scatter plots for pedestrian fatal/severe injury crashes on weekdays and weekends respectively. The weekday scatter plot in Figure 5(a) shows the significant rules are near the Y-axis and also have higher support values. In contrast, Figure 5(b) shows that the important rules appear to congregate near the border of support or confidence values, which indicates that the points situated near either the X-axis or Y-axis are crucial rules for fatal/severe injury crashes.

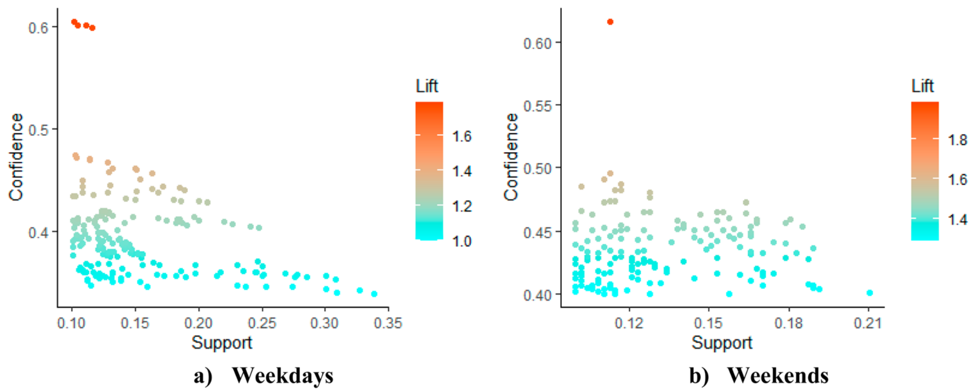


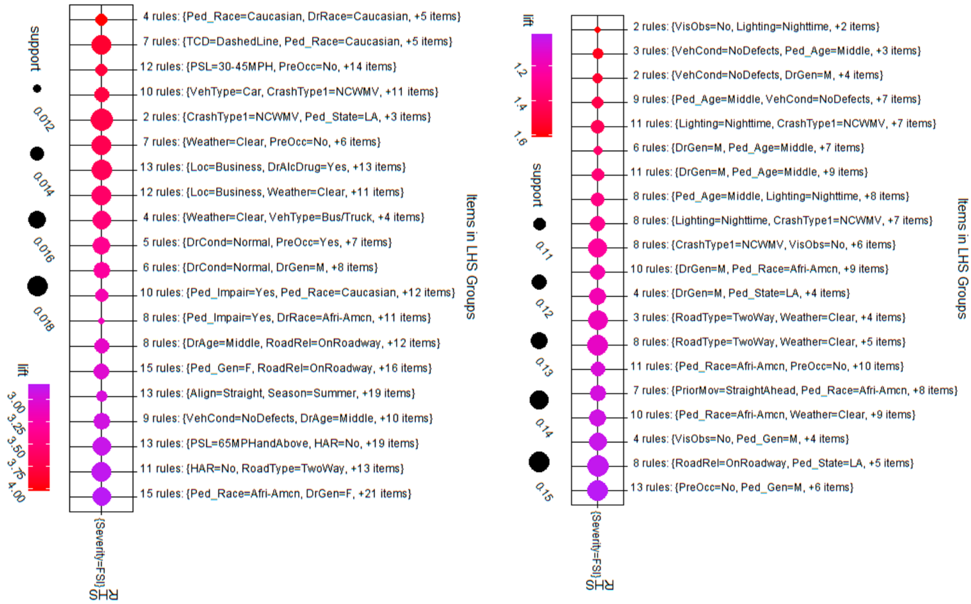
Figure 7. Scatterplot Plots of Association Rules for Minor/ No Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

Similarly, in Figure 6(a,b), it is noticed that the positions of the points are nearer to either the X-axis or Y-axis. The plot illustrates the association between the confidence, support, and lift values of all the association rules for pedestrian moderate injury crashes at midblock during weekdays and weekends. The points in golden color denote higher lift values, and the points in green denote lower lift values. The points with higher lift values usually have a high confidence and lower support values. The lowest confidence value among these two graphs is around 0.450 and the lowest support value is around 0.01. The points in the scatter plot for weekend pedestrian fatal/severe injury crashes remain closer to the vertical and horizontal axis compared to the scatter plot for weekdays.

Figure 7 illustrates the scatter plots of association rules for minor/no injury crashes at midblock based on days of the week. The higher lift values are shown in orange-red color, and the points in cyan denote lower lift values. The two plots in Figure 7(a,b) show that the points are nearer to the support values and have a lower confidence value among all scatter plots. However, based on lift values, the most important rules have a higher confidence value compared to the support value. The most important rules for weekday and weekend pedestrian minor/no injury crashes have higher confidence values and lower support values.

To get a better visualization of the top 15 rules, balloon plots were developed to illustrate the association between the antecedent and the consequent. The severity levels of the crashes, namely fatal/severe injury, moderate injury, and minor/no injury crashes, were considered as consequent (i.e. RHS) and arranged in the columns of the balloon plot. Conversely, the associated factors with each crash severity level were considered antecedents (i.e. LHS) and arranged in the rows of the balloon plots. In the balloon plots, the size of the balloon represents the support values, with larger balloons indicating a higher support value. Additionally, the shades of the balloons indicate lift values.

In Figure 8, the association between the antecedent and consequent of the top rules for pedestrian fatal/severe injury crashes at the midblock during weekdays and weekends is illustrated. The orange shades represent higher lift values, whereas the purple color of the balloons represents a lower lift value. Figure 8(a) shows that the most important group associated with fatal/severe injury crashes based on the lift has four rules. The rules include



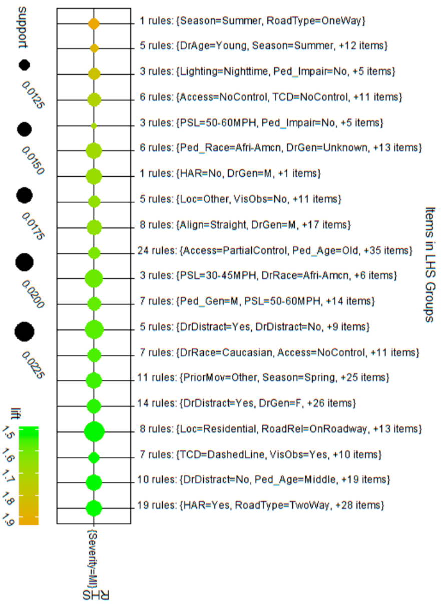
a) Weekdays **b) Weekends**

Figure 8. Grouped Balloon Plots of Association Rules for Fatal/ Severe Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

Caucasian pedestrian race and driver race and five other items in the antecedent. Combining higher lift and support values, the group that appears to be significant has seven rules, including the usage of the dashed line as a traffic control device, the pedestrian race is Caucasian, and five other items. For the top rules of pedestrian fatal/severe injury crashes at midblock during weekends, the most important group based on lift value has two rules. The items include no visual obstruction, nighttime lighting conditions, and five other items. This rule means that during the weekend at nighttime and under no visual obstructions, the likelihood of fatal/severe pedestrian injury crashes is highest.

Figure 9(a,b) represent the balloon plots for the pedestrian moderate injury crashes at midblock during weekdays and weekends, respectively. The most important group of rules for weekday pedestrian moderate injury crashes at midblock has one rule containing the attributes one-way road type and summer. This rule implies that it is highly likely that pedestrian crashes in one-way roadway type during summer will result in a moderate injury crash. The topmost point identifies the most important group that portrays the highest association between the antecedent and the consequent, whereas the bottommost point identifies the lowest association. The most important group of rules for the weekend pedestrian moderate injury balloon plot has one rule. The antecedents in this group are no impaired pedestrian, no preoccupancy, and no crash with motorized vehicles. This implies that if the pedestrian is not impaired and there is no crash with a motorized vehicle, and the roadway is not previously occupied, then it is highly likely that the crash will result in a pedestrian minor/no injury crash at midblock during the weekend.

Figure 10 represents the association between the antecedents and the consequents of the top rules for pedestrian minor/no injury crashes by the day of the week at midblock



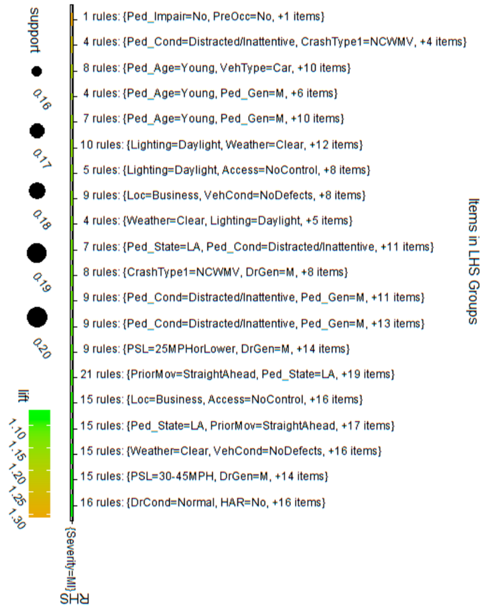
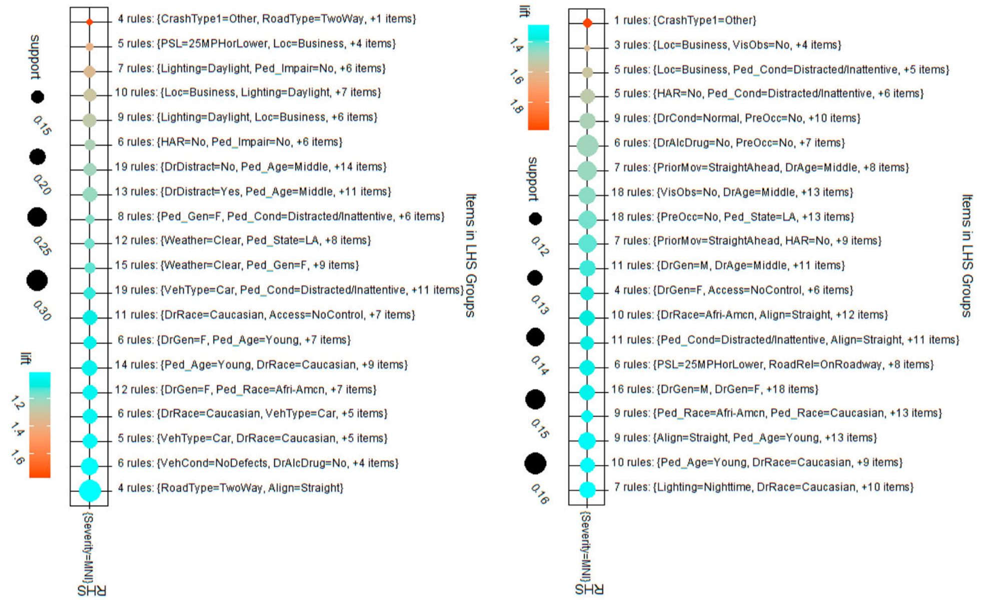


Figure 9. Grouped Balloon Plots of Association Rules for Moderate Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

through balloon plots. Similar to the previous balloon plots, the larger balloons indicate higher support values. The orange-red color indicates a higher lift value, and the cyan color indicates a lower lift value. Figure 10(a) illustrates the important groups for weekday minor/no-injury crashes based on lift value. The most important group has four rules, and the attributes are other types of crashes, two-way roadways, and one other item. This rule signifies that on two-way roadways, left turn, right turn, right angle, and other types of crashes are more likely to result in a pedestrian minor/no injury crash.

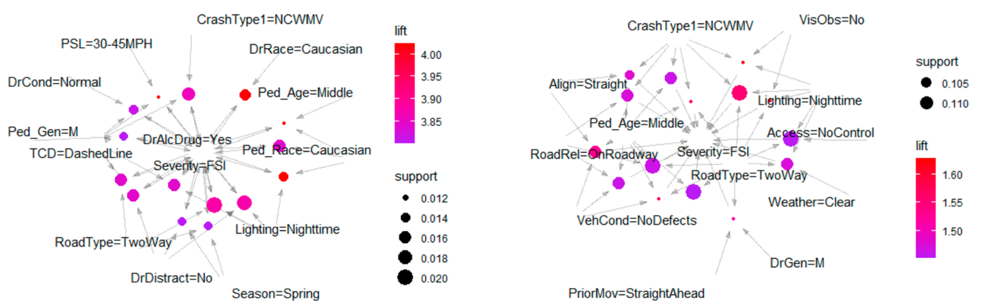
Furthermore, to enhance the understanding of the rules associated with each crash severity level of pedestrian crashes, network graphs for the rules were generated. These visual graphs contain the top 15 rules based on lift values. These plots highlighted significant patterns and identified trends. Figure 11 shows the network graphs for pedestrian fatal/severe injury crashes by day of the week. The plot for weekday pedestrian fatal/severe injury signifies that alcohol or drug-impaired drivers are more susceptible to causing pedestrian fatal/severe injury. The network graph in Figure 11(a) shows a strong relationship between dark lighting conditions such as nighttime. This implies that the likelihood that a drug or alcohol-impaired driver driving at nighttime will cause a pedestrian fatal/severe injury crash at midblock during weekdays is higher compared to others. Figure 11(b) shows that the attributes of no crash with motorized vehicles, no visual obstruction, and nighttime have the strongest association with fatal/severe injury crashes. It signifies that during nighttime, non-motorized vehicles are also responsible for causing fatal pedestrian crashes. Non-motorized vehicles moving at higher speeds can cause deadly injuries to pedestrians.

Figure 12 illustrates the network graphs containing the important rules for pedestrian moderate injury crashes by day of the week. For the weekday network graph in Figure 12(a),



a) Weekdays b) Weekends

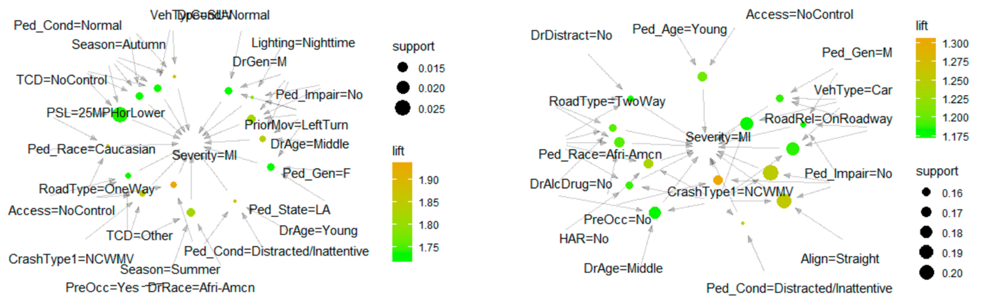
Figure 10. Grouped Balloon Plots of Association Rules for Minor/ No Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.



a) Weekdays b) Weekends

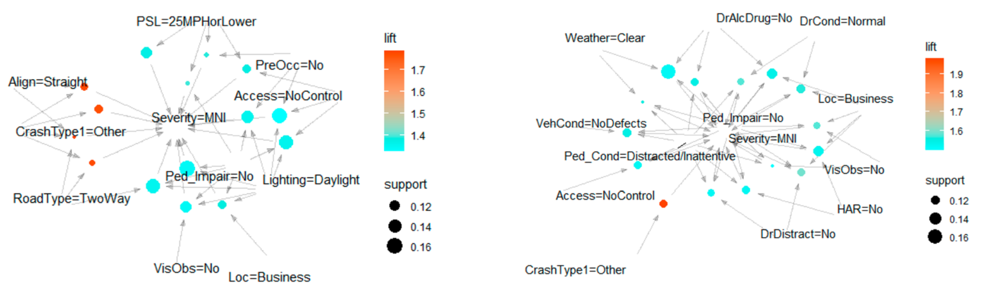
Figure 11. Network Graphs of Association Rules for Fatal/ Severe Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

it can be noted that the attributes one-way roadway, summer, and posted speed limit 25 MPH or lower have higher association compared to others. This implies that during weekdays in summer, it is highly likely that on the one-way roadway where the speed limit is 25 MPH or lower, the crash will end up in a moderate injury pedestrian crash. On roadways that have a lower speed limit, the vehicles will move at a slower speed and if a crash occurs then the likelihood of resulting in a moderate injury pedestrian crash is higher during weekdays. In Figure 12(b), the network graph for the weekend shows a strong association of the attributes 'Ped_Impair = No', 'CrashType1 = NCWMV', and 'Align = Straight'. This signifies that if the roadway has a straight alignment, if the pedestrian is not physically impaired



c) Weekdays **d) Weekends**

Figure 12. Network Graphs Plots of Association Rules for Moderate Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.



a) Weekdays **b) Weekends**

Figure 13. Network Graphs of Association Rules for Minor/ No Injury Crashes by Day of Week: (a) Weekdays (b) Weekends.

and the crash is not with a motorized vehicle then the likelihood is higher that the crash will result in a moderate injury crash. In both graphs, the golden color shades define higher lift values, and the green-colored points define lower lift values.

The network graphs in Figure 13 represent the association of the attributes that are reasonable for causing pedestrian Minor/No injury by day of the week. The orange-red colored plots denote higher lift values. In the weekday network graph in Figure 13(a), other crash types such as head-on, left turn, right turn crashes, straight roadway alignment, and two-way road types show the strongest association. These attributes significantly influence the likelihood of pedestrian minor/no-injury crashes during weekdays. Similarly, in the weekend network graph, right turn, left turn, head on, and other crashes play a significant role in causing minor/no injury pedestrian crashes at midblock during the weekend, along with other attributes that indicate no pedestrian impairment and no distraction by the driver.

Table 10. Top 15 Association Rules for Pedestrian Fatal/ Severe Injury Crashes at Midblock During Daylight Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
FSIDL01	Access = FullControl + DrDistract = No + Ped_Cond = Other	Severity = FSI	0.011	0.900	0.012	6.378	9
FSIDL02	Access = FullControl + Ped_Cond = Other	Severity = FSI	0.011	0.818	0.013	5.798	9
FSIDL03	DrDistract = No + Ped_Cond = Other + PSL = 50-60MPH	Severity = FSI	0.012	0.769	0.016	5.452	10
FSIDL04	Ped_Cond = Other + PSL = 50-60MPH + VehCond = NoDefects	Severity = FSI	0.012	0.769	0.016	5.452	10
FSIDL05	HAR = No + Ped_Cond = Other + PSL = 50-60MPH	Severity = FSI	0.012	0.769	0.016	5.452	10
FSIDL06	DrCond = Normal + Ped_Cond = Other + PSL = 50-60MPH	Severity = FSI	0.012	0.714	0.017	5.062	10
FSIDL07	Access = FullControl + DrCond = Normal + TCD = DashedLine	Severity = FSI	0.011	0.692	0.016	4.906	9
FSIDL08	Access = FullControl + RoadType = TwoWay + TCD = DashedLine	Severity = FSI	0.011	0.692	0.016	4.906	9
FSIDL09	Access = FullControl + CrashType1 = NCWMV + RoadType = TwoWay	Severity = FSI	0.011	0.692	0.016	4.906	9
FSIDL10	Access = FullControl + DrCond = Normal + RoadType = TwoWay	Severity = FSI	0.012	0.667	0.018	4.725	10
FSIDL11	Access = FullControl + DrDistract = No + RoadType = TwoWay	Severity = FSI	0.012	0.667	0.018	4.725	10
FSIDL12	Access = FullControl + HAR = No + RoadType = TwoWay	Severity = FSI	0.012	0.667	0.018	4.725	10
FSIDL13	CrashType1 = NCWMV + Ped_Cond = Other + PSL = 50-60MPH	Severity = FSI	0.012	0.667	0.018	4.725	10
FSIDL14	Access = FullControl + TCD = DashedLine	Severity = FSI	0.011	0.643	0.017	4.556	9
FSIDL15	Access = FullControl + Ped_Gen = M + RoadType = TwoWay	Severity = FSI	0.011	0.643	0.017	4.556	9

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

4.3. Temporal clusters by time of day

4.3.1. Daytime

Table 10 displays the top 15 significant rules contributing to pedestrian fatal/severe injury crashes at midblock during daytime. The highest lift value 6.378 indicates the topmost important rule {*Access = FullControl + DrDistract = No + Ped_Cond = Other → Severity = FSI*} for pedestrian fatal/severe injury crashes at midblock during daytime. This rule indicates that when there is a crash on a fully controlled access roadway (*Access = FullControl*), the driver is not distracted (*DrDistract = No*), and the pedestrian is physically/drug impaired or has illness/fatigue or experienced blackout (*Ped_Cond = Other*), it is likely to result in a pedestrian fatal/severe injury crash at midblock during daytime. The unhindered traffic flow on a full-control highway may increase the average speed of vehicles. The unawareness of pedestrians at midblock on such roadway might result in fatal/severe injury crashes. The lift value indicates that the occurrence of fatal/severe injury crashes at midblock during daytime with this combination is 6.378 times higher than the overall occurrence of similar crashes in the dataset.

The support, confidence, and coverage values for this rule are 0.011, 0.9, and 0.012, respectively. The support value indicates that 1.1% of the pedestrian fatal/severe injury crashes that occur at midblock during daytime are attributed to these factors. The confidence value means that 1.2% of all pedestrian fatal/severe injury crashes in this dataset involve drivers with no distractions and physically/drug-impaired pedestrians. The coverage value shows that 1.2% of the fatal/severe injury crashes at midblock during daylight are associated with this combination of factors. The second topmost important rule is $\{Access = FullControl + Ped\_Cond = Other \rightarrow Severity = FSI\}$.

The rules FSIDL03, FSIDL04, FSIDL05, FSIDL06, FSIDL10, FSIDL11, FSIDL12, and FSIDL13 have the highest count of 10 among the top 15 rules. The topmost attribute that occurred 10 times in the antecedent was $Access = FullControl$.

Table 11 shows the top 15 significant rules for pedestrian moderate injury cashes at midblock during daytime. Based on the lift values, the topmost important rule is $\{Loc = Business + Ped\_Impair = No + Ped\_State = LA \rightarrow Severity = MI\}$. This rule means that if a roadway is situated in a commercial area, and the pedestrian is not impaired then it is likely that the crash severity would be moderate injury crash. The lift value 2.239 means that the percentage of moderate injury crashes at midblock, including the combination of these factors, is 2.239 times higher than the percentage of such crashes during daylight in the overall dataset. The support, confidence, and coverage values are 0.011, 1.00, and 0.011. This rule has 9 counts. The rules MIDL12 and MIDL13 have the most counts in this dataset. The attribute $RoadType = OneWay$ appeared 9 times in the antecedent among the top 15 rules. The second most important rule $\{Loc = Residential + Season = Autumn + Weather = Adverse \rightarrow Severity = MI\}$ means that adverse weather conditions can lead to pedestrian moderate injury crashes on one-way roadways in residential areas.

Table 12 lists the top 15 rules for pedestrian minor/no injury crashes at midblock during daylight. The rules are extracted and organized based on lift values.

The top rule for pedestrian minor/no injury crashes is $\{Align = Straight + Ped\_Gen = F + RoadType = TwoWay \rightarrow Severity = MNI\}$. This rule signifies that if the road alignment is straight ($Align = Straight$), a pedestrian is female ($Ped\_Gen = F$), and if the roadway type is a two-way roadway, then it is highly likely that the crash severity type would be pedestrian minor/no injury. The support, confidence, coverage, and lift values for this rule are 0.155, 0.483, 0.320, and 1.171. The lift value for this rule signifies that the occurrence of pedestrian minor/no injury crashes in daylight with this combination of factors is 1.171 times higher compared to the overall dataset for pedestrian minor/no injury crashes in daylight. The support value suggests that 15.5% of pedestrian minor/no-injury crashes at midblock during daylight are attributed to this combination of factors. The second most important rule $\{Ped\_Gen = F + RoadType = TwoWay \rightarrow Severity = MNI\}$ means that if the pedestrian is female and the roadway is two-way, then it is also highly likely that the crash would result in a minor/no injury crash at midblock during daylight. The topmost attribute here is $Ped\_Gen = F$ which occurred 8 times in the antecedent. The second most frequent attribute in the antecedent is $Ped\_State = LA$ which appeared 5 times among the top 15 rules.

4.3.2. Nighttime

The association rules in Table 13 are for pedestrian fatal/severe injury crashes at midblock during nighttime. The topmost rule based on the highest lift value is $\{Ped\_Cond = Other + PriorMov = StraightAhead + VisObs = No \rightarrow Severity = FSI\}$. This

Table 11. Top 15 Association Rules for Pedestrian Moderate Injury Crashes at Midblock During Daylight Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MIDL01	Ped_Cond = Normal + RoadType = OneWay + Season = Summer	Severity = MI	0.011	1.000	0.011	2.239	9
MIDL02	Loc = Residential + Season = Autumn + Weather = Adverse	Severity = MI	0.011	1.000	0.011	2.239	9
MIDL03	Access = NoControl + RoadType = OneWay + Season = Summer	Severity = MI	0.015	0.923	0.016	2.067	12
MIDL04	DOW = Weekday + RoadType = OneWay + Season = Summer	Severity = MI	0.012	0.909	0.013	2.035	10
MIDL05	PreOcc = No + RoadType = OneWay + Season = Summer	Severity = MI	0.012	0.909	0.013	2.035	10
MIDL06	Access = PartialControl + PSL = 30-45MPH + Season = Winter	Severity = MI	0.011	0.900	0.012	2.015	9
MIDL07	PSL = 25MPHorLower + RoadType = OneWay + Season = Summer	Severity = MI	0.011	0.900	0.012	2.015	9
MIDL08	Ped_Age = Middle + RoadType = OneWay + Season = Summer	Severity = MI	0.011	0.900	0.012	2.015	9
MIDL09	CrashType1 = NCWMV + RoadType = OneWay + Season = Summer	Severity = MI	0.011	0.900	0.012	2.015	9
MIDL10	Ped_Impair = No + RoadType = OneWay + Season = Summer	Severity = MI	0.011	0.900	0.012	2.015	9
MIDL11	PreOcc = Yes + Season = Spring + TCD = Other	Severity = MI	0.011	0.900	0.012	2.015	9
MIDL12	RoadType = OneWay + Season = Summer	Severity = MI	0.017	0.875	0.020	1.959	14
MIDL13	Ped_Race = Caucasian + PriorMov = Other + Season = Summer	Severity = MI	0.017	0.875	0.020	1.959	14
MIDL14	DrAge = Middle + DrGen = M + PriorMov = LeftTurn	Severity = MI	0.015	0.857	0.017	1.919	12
MIDL15	DrRace = Afri-Amcn + PreOcc = Yes + TCD = Other	Severity = MI	0.015	0.857	0.017	1.919	12

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

rule indicates that if the pedestrian is physically/drug impaired or has illness/fatigues or experienced blackout (Ped_Cond = Other), if the vehicle has a prior movement of moving straight under no visual obstruction, then the crash is mostly likely to result in a fatal/severe injury crash. Even if there is no visual obstruction for both driver and pedestrian, sudden unauthorized movement by an impaired pedestrian may cause a fatal/severe injury crash as the driver would have very little time to react properly. The top rule has a lift value of 1.797, indicating that the occurrence of pedestrian fatal/severe injury crashes at nighttime with this combination of factors is 1.797 times higher compared to the overall dataset of similar types of crashes at nighttime. The support value refers that 15.1% of fatal/severe injury crashes that occurred at midblock in nighttime include this combination of factors. The second topmost rule {*Ped_Cond = Other + PriorMov = StraightAhead + VehCond = NoDefects → Severity = FSI*} also include the attributes *Ped_Cond = Other* and *PriorMov = StraightAhead*. The rule with the most count was {*CrashType1 = NCWMV + Ped_Cond = Other + RoadRel = OnRoadway → Severity = FSI*}. The attribute that occurred in all top 15 rules is *Ped_Cond = Other*. The second most common attribute that occurred in the antecedent is *PriorMov = StraightAhead*.

In Table 14, the top 15 association rules for pedestrian moderate injury type crashes at midblock during nighttime are displayed. The most important rule is

Table 12. Top 15 Association Rules for Pedestrian Minor/ No Injury Crashes at Midblock During Daylight Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MNIDL01	Align = Straight + Ped_Gen = F + RoadType = TwoWay	Severity = MNI	0.155	0.483	0.320	1.171	126
MNIDL02	Ped_Gen = F + RoadType = TwoWay	Severity = MNI	0.155	0.477	0.324	1.158	126
MNIDL03	Access = NoControl + Ped_Gen = F + Ped_State = LA	Severity = MNI	0.151	0.477	0.317	1.156	123
MNIDL04	Align = Straight + Ped_Gen = F + Ped_State = LA	Severity = MNI	0.160	0.476	0.335	1.155	130
MNIDL05	Access = NoControl + Ped_Cond = Distracted/Inattentive + PreOcc = No	Severity = MNI	0.163	0.473	0.345	1.148	133
MNIDL06	Ped_Gen = F + Ped_State = LA	Severity = MNI	0.161	0.473	0.340	1.147	131
MNIDL07	Access = NoControl + Align = Straight + Ped_Gen = F	Severity = MNI	0.156	0.472	0.330	1.145	127
MNIDL08	Ped_Cond = Distracted/Inattentive + Ped_State = LA + PreOcc = No	Severity = MNI	0.162	0.471	0.344	1.143	132
MNIDL09	Loc = Business + Ped_Impair = No + Ped_State = LA	Severity = MNI	0.209	0.471	0.443	1.142	170
MNIDL10	Access = NoControl + Ped_Gen = F	Severity = MNI	0.157	0.471	0.334	1.141	128
MNIDL11	Ped_Cond = Distracted/Inattentive + Ped_Impair = No + VisObs = No	Severity = MNI	0.156	0.470	0.331	1.141	127
MNIDL12	DOW = Weekday + Loc = Business + Ped_Impair = No	Severity = MNI	0.183	0.470	0.389	1.140	149
MNIDL13	Loc = Business + Ped_Impair = No + RoadType = TwoWay	Severity = MNI	0.201	0.470	0.428	1.140	164
MNIDL14	Align = Straight + Ped_Gen = F	Severity = MNI	0.167	0.469	0.356	1.138	136
MNIDL15	Loc = Business + Ped_Impair = No + VisObs = No	Severity = MNI	0.172	0.468	0.367	1.136	140

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

{*CrashType1 = NCWMV + TCD = NoControl + VehType = Other → Severity = MI*}. It means that the combination of attributes including no crash with a motor vehicle, absence of traffic control device on the roadway, and bicycle, bus, school bus, and other types of vehicles, is strongly associated with pedestrian moderate injury crashes at midblock during nighttime. All mentioned vehicles move at comparatively slower speeds, making it easier for the driver to react within a short period. This increases the likelihood of decreasing pedestrian injury severity. The lift value 2.225 signifies that the percentage of pedestrian moderate injury crashes at midblock during nighttime with this combination of factors is 2.335 times higher compared to the overall dataset of pedestrian moderate injury midblock crashes at nighttime. The most repeated attributes in the antecedent are *TCD = NoControl* and *VehType = Other*. The rules with the highest counts of 16 are MINT10 and MINT11.

Table 15 lists the top 15 association rules for pedestrian minor/no-injury crashes at midblock during nighttime that are based on lift values. The most important rule is {*CrashType1 = Other + PSL = 25 MPHorLower + Season = Autumn → Severity = MI*}. The rule signifies that if the crash type is a head-on or right turn crash (*CrashType1 = Other*) on a 25 MPH or lower posted speed limit roadway (*PSL = 25MPHorLower*) during autumn (*Season = Autumn*), then it is highly likely that the crash severity would be minor/no injury pedestrian crash at midblock during midnight. A posted speed limit can significantly impact pedestrian injury severity in a vehicle-pedestrian crash (Islam 2023). The support, confidence, coverage, and lift values for this rule are 0.012, 0.923, 0.013, and

Table 13. Top 15 Association Rules for Pedestrian Fatal/ Severe Injury Crashes at Midblock During Nighttime Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
FSINT01	Ped_Cond = Other + PriorMov = StraightAhead + VisObs = No	Severity = FSI	0.151	0.625	0.241	1.797	145
FSINT02	Ped_Cond = Other + PriorMov = StraightAhead + VehCond = NoDefects	Severity = FSI	0.161	0.620	0.260	1.782	155
FSINT03	CrashType1 = NCWMV + Ped_Cond = Other + PriorMov = StraightAhead	Severity = FSI	0.168	0.618	0.272	1.777	162
FSINT04	CrashType1 = NCWMV + Ped_Cond = Other + VisObs = No	Severity = FSI	0.151	0.609	0.247	1.751	145
FSINT05	HAR = No + Ped_Cond = Other + PriorMov = StraightAhead	Severity = FSI	0.157	0.609	0.258	1.750	151
FSINT06	Ped_Cond = Other + PriorMov = StraightAhead + RoadType = TwoWay	Severity = FSI	0.168	0.604	0.278	1.738	162
FSINT07	CrashType1 = NCWMV + Ped_Cond = Other + VehCond = NoDefects	Severity = FSI	0.163	0.604	0.270	1.736	157
FSINT08	DrDistract = No + Ped_Cond = Other + RoadRel = OnRoadway	Severity = FSI	0.154	0.602	0.255	1.729	148
FSINT09	DrDistract = No + Ped_Cond = Other	Severity = FSI	0.154	0.597	0.258	1.716	148
FSINT10	Ped_Cond = Other + PriorMov = StraightAhead + RoadRel = OnRoadway	Severity = FSI	0.184	0.594	0.309	1.707	177
FSINT11	CrashType1 = NCWMV + HAR = No + Ped_Cond = Other	Severity = FSI	0.156	0.593	0.263	1.704	150
FSINT12	CrashType1 = NCWMV + Ped_Cond = Other + RoadType = TwoWay	Severity = FSI	0.170	0.592	0.288	1.702	164
FSINT13	Ped_Cond = Other + PriorMov = StraightAhead	Severity = FSI	0.185	0.589	0.314	1.694	178
FSINT14	Ped_Age = Middle + Ped_Cond = Other + RoadRel = OnRoadway	Severity = FSI	0.153	0.586	0.261	1.684	147
FSINT15	CrashType1 = NCWMV + Ped_Cond = Other + RoadRel = OnRoadway	Severity = FSI	0.187	0.584	0.320	1.680	180

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

3.514. The lift value indicates that the percentage of pedestrian minor/no injury crashes at midblock during nighttime with this combination of factors is 3.514 times higher compared to all the other combinations of factors in this dataset. The support value indicates that among the dataset of pedestrian minor/no injury at midblock during nighttime, 1.2% of crashes occurred with this combination. The second most important rule includes the attributes *CrashType1 = Other*, *DrRace = Caucasian*, and *Ped_Age = Young*. The most occurred attribute is *CrashType1 = Other* which occurred in all top 15 rules. The rule MNINT08 has the highest count among these top 15 association rules.

4.3.3. Visualization of the rules

Like temporal clusters by day of the week, visualization of the association rules of the temporal clusters by lighting condition was also made in this study. Figure 14(a,b) illustrate the scatter plots of the association rules for pedestrian fatal/severe injury crashes under daylight and nighttime, respectively. The scatter plot for daylight shows that the points are situated along positions nearer to the X-axis and the Y-axis. The most important rule with

Table 14. Top 15 Association Rules for Pedestrian Moderate Injury Crashes at Midblock During Nighttime Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MINT01	CrashType1 = NCWMV + TCD = NoControl + VehType = Other	Severity = MI	0.010	0.909	0.011	2.335	10
MINT02	DrAge = Unknown + PriorMov = StraightAhead + Season = Winter	Severity = MI	0.015	0.875	0.017	2.247	14
MINT03	HAR = No + Ped_Gen = F + PriorMov = LeftTurn	Severity = MI	0.012	0.857	0.015	2.201	12
MINT04	PriorMov = StraightAhead + TCD = NoControl + VehType = Other	Severity = MI	0.012	0.857	0.015	2.201	12
MINT05	Ped_Gen = F + Ped_Impair = No + PriorMov = LeftTurn	Severity = MI	0.011	0.846	0.013	2.173	11
MINT06	DrAge = Middle + Ped_Gen = F + PriorMov = LeftTurn	Severity = MI	0.011	0.846	0.013	2.173	11
MINT07	DrDistract = No + Ped_Gen = F + PriorMov = LeftTurn	Severity = MI	0.011	0.846	0.013	2.173	11
MINT08	DrAlcDrug = No + Ped_Gen = F + PriorMov = LeftTurn	Severity = MI	0.011	0.846	0.013	2.173	11
MINT09	Ped_Gen = F + PriorMov = LeftTurn + VehCond = NoDefects	Severity = MI	0.011	0.846	0.013	2.173	11
MINT10	DrGen = Unknown + Ped_Impair = No + Season = Winter	Severity = MI	0.017	0.842	0.020	2.163	16
MINT11	DrAge = Unknown + Ped_Impair = No + Season = Winter	Severity = MI	0.017	0.842	0.020	2.163	16
MINT12	Ped_Impair = No + PriorMov = LeftTurn + PSL = 30-45MPH	Severity = MI	0.010	0.833	0.012	2.140	10
MINT13	Ped_Age = Middle + TCD = NoControl + VehType = Other	Severity = MI	0.010	0.833	0.012	2.140	10
MINT14	DrCond = Distracted/Inattentive + HAR = No + TCD = NoControl	Severity = MI	0.010	0.833	0.012	2.140	10
MINT15	Access = PartialControl + PSL = 25MPHorLower + Weather = Clear	Severity = MI	0.010	0.833	0.012	2.140	10

Notes: *Sup = Support, Con = Confidence, Cov = Coverage*

the highest lift value has a higher confidence value compared to others. Also, the confidence value for the same point is lower. In the plot for nighttime pedestrian fatal/severe injury association rules, the most important rules have higher confidence values. In both plots, the red colored points are denoting higher lift values.

In the scatterplots of the association rules for pedestrian moderate injury crashes at midblock by lighting conditions, it is notable that the points that represent the important rules tend to remain closer to both axes. Figure 15(a) illustrates the scatter plot for such type of pedestrian crashes under daylight, where the most important rule based on lift value has a higher confidence value. A similar crash is noticed in the scatter plot for nighttime, which is portrayed in Figure 15(a). The golden color in both figures defines higher lift values.

The cyan color in the plots of Figure 16 indicates lower lift values and orange-red color in the plots is indicating higher lift values. The points in Figure 16(b), show a tendency to remain nearer to the horizontal and vertical axes compared to the points in Figure 16(a) which denotes the scatter plot for daylight.

The visualization of the association rules for different crash severity types based on lighting conditions is further presented in balloon plots. These plots better illustrate the association between the antecedent and the consequent. Figure 17(a) shows the balloon plot for daytime crashes. The topmost important group has one association rule that includes the attributes '*DrDistract = No*', '*Ped_Cond = Other*', and another item that

Table 15. Top 15 Association Rules for Pedestrian Minor/ No Injury Crashes at Midblock During Nighttime Based on Lift.

Rules	Antecedent	Consequent	Sup	Conf	Cov	Lift	Count
MNINT01	CrashType1 = Other + PSL = 25MPHorLower + Season = Autumn	Severity = MNI	0.012	0.923	0.013	3.514	12
MNINT02	CrashType1 = Other + DrRace = Caucasian + Ped_Age = Young	Severity = MNI	0.012	0.923	0.013	3.514	12
MNINT03	CrashType1 = Other + DrAge = Young + DrDistract = Yes	Severity = MNI	0.011	0.917	0.012	3.489	11
MNINT04	CrashType1 = Other + DrRace = Caucasian + PriorMov = Other	Severity = MNI	0.010	0.909	0.011	3.460	10
MNINT05	CrashType1 = Other + Ped_Cond = Distracted/Inattentive + PSL = 25MPHorLower	Severity = MNI	0.010	0.909	0.011	3.460	10
MNINT06	CrashType1 = Other + DrRace = Caucasian + PSL = 25MPHorLower	Severity = MNI	0.016	0.882	0.018	3.359	15
MNINT07	CrashType1 = Other + DOW = Weekend + DrRace = Caucasian	Severity = MNI	0.015	0.875	0.017	3.331	14
MNINT08	CrashType1 = Other + DrRace = Caucasian + Ped_Race = Afri-Amcn	Severity = MNI	0.022	0.875	0.025	3.331	21
MNINT09	CrashType1 = Other + DrRace = Caucasian + Ped_Cond = Distracted/Inattentive	Severity = MNI	0.013	0.867	0.016	3.299	13
MNINT10	CrashType1 = Other + DrRace = Caucasian + TCD = Other	Severity = MNI	0.012	0.857	0.015	3.263	12
MNINT11	CrashType1 = Other + DrGen = F + Ped_Cond = Distracted/Inattentive	Severity = MNI	0.012	0.857	0.015	3.263	12
MNINT12	CrashType1 = Other + PriorMov = Other + Season = Winter	Severity = MNI	0.011	0.846	0.013	3.221	11
MNINT13	CrashType1 = Other + DrDistract = No + PriorMov = Other	Severity = MNI	0.011	0.846	0.013	3.221	11
MNINT14	CrashType1 = Other + DrGen = F + PreOcc = Yes	Severity = MNI	0.011	0.846	0.013	3.221	11
MNINT15	CrashType1 = Other + Loc = Business + RoadType = OneWay	Severity = MNI	0.010	0.833	0.012	3.172	10

Notes: *Sup* = *Support, Con* = *Confidence, Cov* = *Coverage*

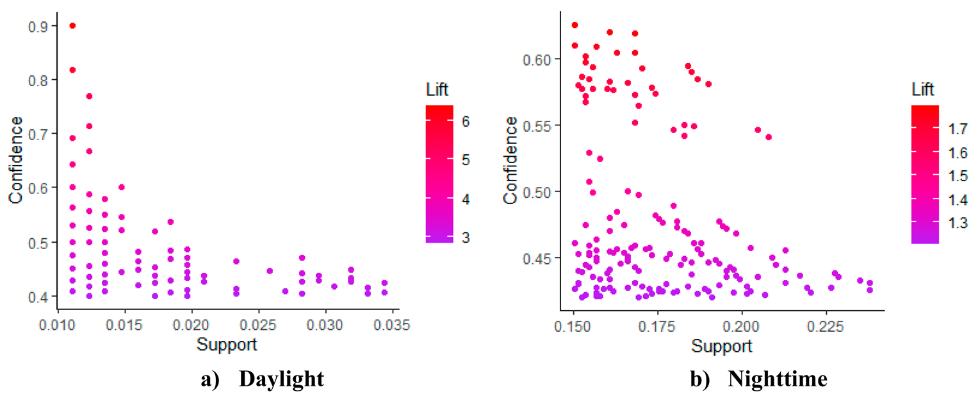


Figure 14. Scatterplot Plots of Association Rules for Fatal/ Severe Injury Crashes by Lighting Condition: (a) Daylight (b) Nighttime.

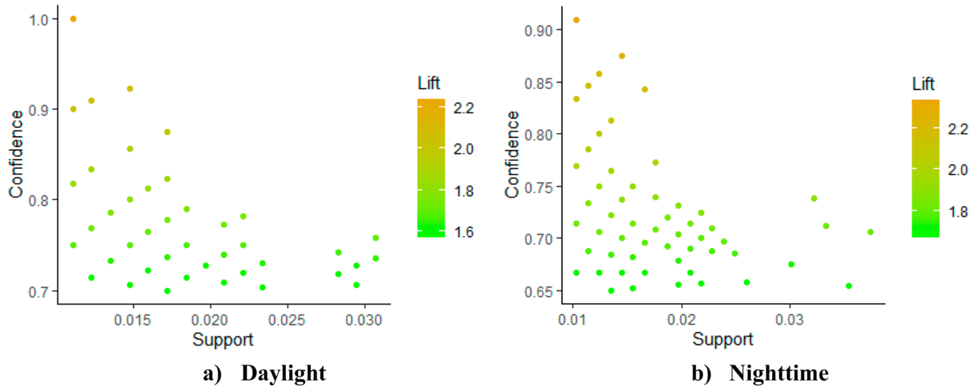


Figure 15. Scatterplot Plots of Association Rules for Moderate Injury Crashes by Lighting Condition: (a) Daylight (b) Nighttime.

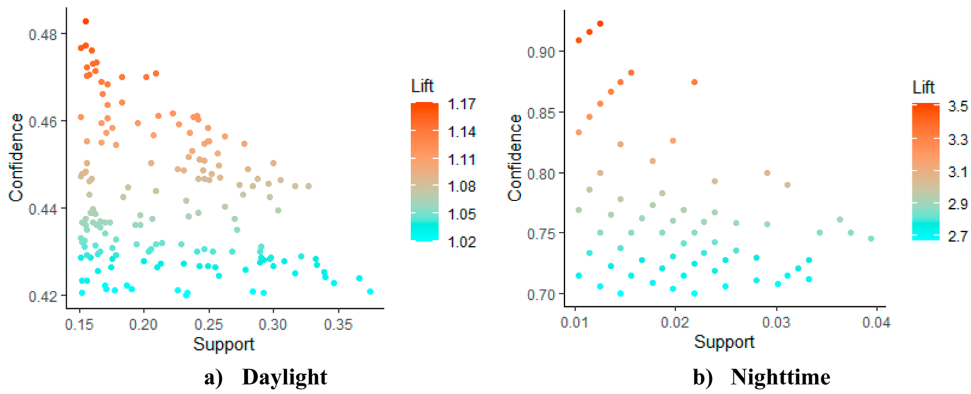
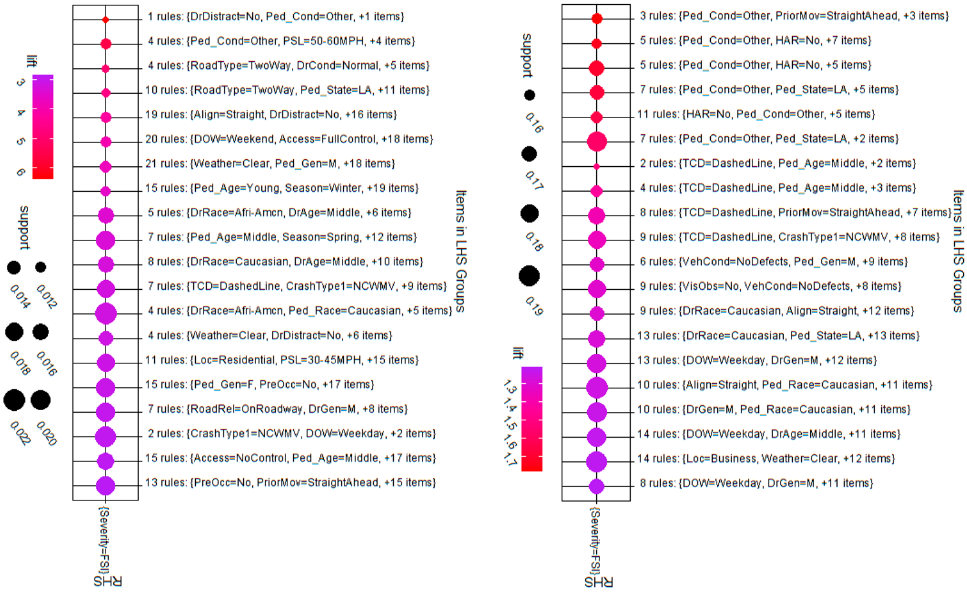


Figure 16. Scatterplot Plots of Association Rules for Minor/ NoInjury Crashes by Lighting Condition: (a) Daylight (b) Nighttime.

is '*Access* = *FullControl*'. This rule denotes that if the pedestrian is physically/drug impaired and the driver is not distracted while driving in a roadway that has complete control, then the likelihood is high that the crash would end up in a fatal/severe injury crash. Again, the most important group for the nighttime in Figure 17(a) has three rules. The attributes here include '*Ped_Cond* = *Other*', 'PriorMov = StraightAhead' and three other items. This group of rules indicates that under similar pedestrian conditions, the likelihood that a crash during nighttime would result in fatal/severe injury is higher. The purple color denotes lower lift values whereas the red colored point denotes higher lift value.

Figure 18(a,b) illustrate the balloon plots for pedestrian moderate injury crashes at mid-block during daytime and nighttime respectively. The size of the balloon plots indicates the support values. The larger points mean a higher support value. The golden color plots denote higher lift values. The most important group of rules in the balloon plot for daytime crashes has two rules. The attributes here are '*Loc* = *Residential*', '*Ped_Cond* = *Normal*' and four other items. In roadways that belong to residential areas, a likelihood that a crash would result in a moderate injury crash if the pedestrian were in normal condition. The



a) Daytime **b) Nighttime**

Figure 17. Grouped Balloon Plots of Association Rules for Fatal/ Severe Injury Crashes by Lighting Condition: (a) Daytime (b) Nighttime.

most important group for nighttime association rules has one rule and it denotes that if the pedestrian collides with any vehicles other than motorized vehicles, then the crash severity would result in a moderate injury.

Figure 19 shows two balloon plots of the association rules for daytime and nighttime pedestrian minor/no-injury crashes. The topmost important group for daytime in Figure 19(a) contains one rule, this group is associated with female pedestrians and two-way roadway type with one more item that is straight roadway alignment with pedestrian minor/no injury crashes at daytime. The plot for nighttime in Figure 19(b) shows that the topmost important group of association rules has 5 rules. This group signifies the association between the young pedestrian and distracted driver with a minor/no injury crash at nighttime. This means that if the driver is distracted and the crash involves a young pedestrian, then it is highly likely that the crash severity would be a minor/no injury crash. The orange-red colored points of the balloon plot define higher lift values. Moreover, the size of the points denotes support values. The larger size signifies a larger support value.

Network graphs were also constructed for better visualization of factors that are associated with pedestrian crash severity. The outcomes of the graphs depicted that the association rules plots align with the findings of Table 10 to Table 15. Figure 20(a,b) show the network graphs of association rules for fatal/severe injury crashes in daytime and nighttime, respectively. The network graph for daytime shows that physically/drug impaired pedestrians ('*Ped_Cond* = *Other*') are one of the most important factors that cause fatal/severe injury crashes. Moreover, fully controlled roadways are also a stimulating factor for pedestrian fatal/severe crashes. Similarly, the graph for nighttime shows that physically/drug-impaired pedestrians have strong associations with fatal/severe crash injury at midblock.

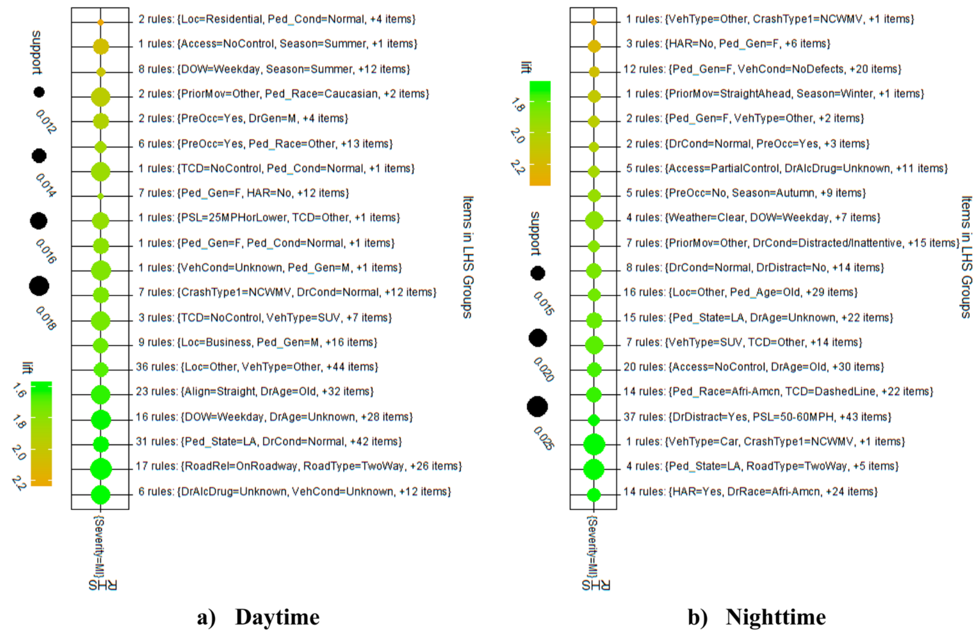


Figure 18. Grouped Balloon Plots of Association Rules for Moderate Injury Crashes by Lighting Condition: (a) Daytime (b) Nighttime.

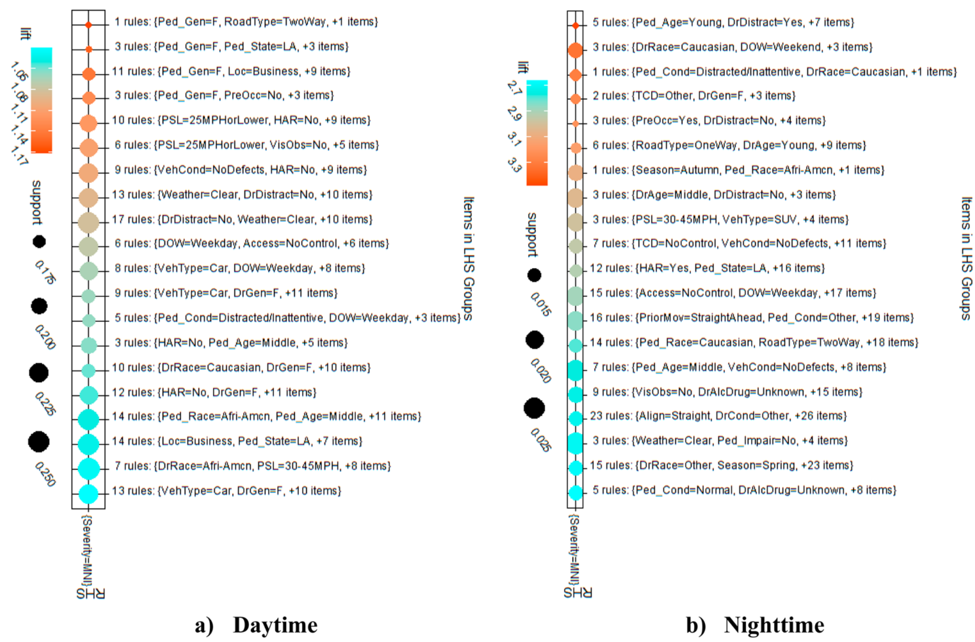
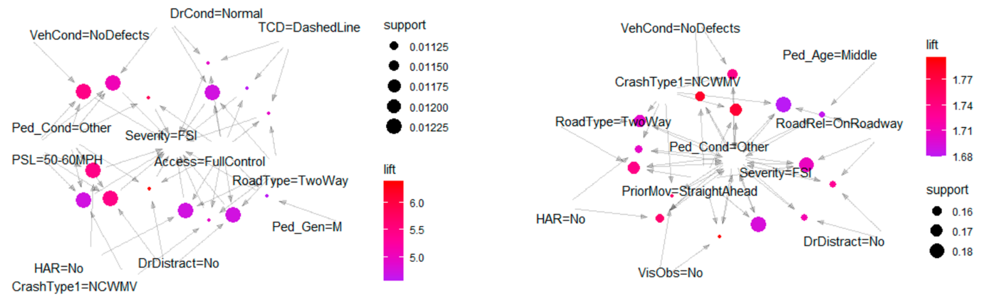
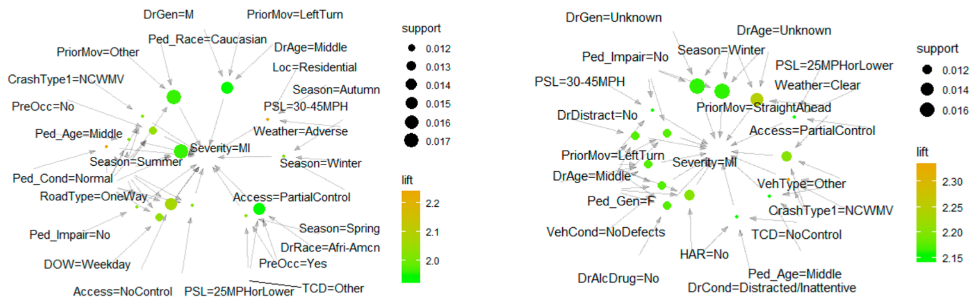


Figure 19. Grouped Balloon Plots of Association Rules for Minor/ No Injury Crashes by Lighting Condition: (a) Daytime (b) Nighttime.



a) Daytime b) Nighttime

Figure 20. Network Graphs of Association Rules for Fatal/ Severe Injury Crashes by Lighting Condition: (a) Daytime (b) Nighttime.



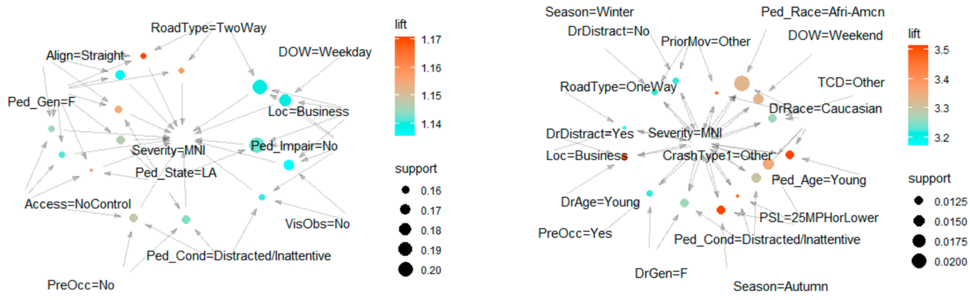
a) Daytime b) Nighttime

Figure 21. Network Graphs of Association Rules for Moderate Injury Crashes by Lighting Condition: (a) Daytime (b) Nighttime.

In addition, vehicles moving straight ahead with no visual obstructions are also significant factors in causing fatal/severe injury crashes.

The network graphs in Figure 21(a,b) illustrate valuable insights on the association rules for pedestrian moderate injury crashes at midblock during daytime and nighttime, respectively. During the daytime, the significant factors causing moderate injury at midblock are normal pedestrian conditions, one-way roadways, and the summer season. It implies that on one-way roadways during summer, a pedestrian in a normal condition is more likely to be involved in less severe injury crashes. The graph for nighttime shows that a pedestrian crash with non-motorized vehicles such as bicycles will result in a crash severity that will cause moderate injury.

The visualization analyses conclude with network graphs of association rules for pedestrian minor/no injury crashes at midblock by lighting condition. In Figure 22(a,b), the factors of association rules for such types of crashes are portrayed in network graphs. The network graph for daytime shows that straight roadway alignment of a two-way roadway plays a significant role in decreasing the crash severity level and resulting in pedestrian minor/no injury crashes. The network graph for nighttime illustrates the significant factor for such types of crashes. These factors include roadways with a posted speed limit of 25 MPH, other



a) Daytime **b) Nighttime**

Figure 22. Network Graphs of Association Rules for Minor/ No Injury Crashes by Lighting Condition: (a) Daytime (b) Nighttime.

crash types such as head-on left turns, right turn crashes, and young pedestrians. These signify that on a roadway with a lower speed limit, there is a likelihood that young pedestrians will end up in minor/no-injury crashes. This can also be explained by the lower reaction time capability of pedestrians at a younger age.

5. Major findings

The study highlighted a few major findings. These are described below:

1. Impaired drivers and crash severity: Alcohol or drug-impaired drivers are strongly associated with fatal/severe pedestrian crashes at midblock locations during weekdays. This finding highlighted the critical impact of impaired driving on pedestrian safety.
2. Nighttime crashes: Nighttime conditions significantly increased the likelihood of fatal/severe pedestrian crashes. Impaired drivers and impaired pedestrians are predominant attributes associated with these crashes during nighttime.
3. One-way roadways and seasonal effects: One-way roadways and the summer season are strongly associated with moderate pedestrian injury crashes at midblock locations during weekdays. This suggested that specific roadway designs and seasonal factors influence crash severity.
4. Speed limits and road types: Roadways with 25 MPH posted speed limits showed a strong association with moderate injury crashes during weekdays, indicating that even lower speed limits may not fully mitigate the risk of injury in certain road environments. Furthermore, two-way roadways, especially during nighttime, are associated with higher crash severity, indicating the need for targeted interventions on these types of roads.
5. Demographic factors: Specific demographic factors, such as driver race and pedestrian age, played a role in crash severity. For instance, Caucasian drivers under the influence of alcohol or drugs had a higher association with severe crashes.

6. Distractions: Non-distracted drivers and controlled access roadways are associated with a lower likelihood of fatal/severe pedestrian crashes during the daytime, indicating the importance of reducing driver distractions and maintaining controlled access.

6. Conclusion

We examined the impact of time of the day and day of the week in understanding the patterns of association factors in pedestrian crashes by using association rules mining. The study of temporal clusters based on days of the week reveals insightful patterns in pedestrian crashes. Notably, weekends exhibit increased vulnerability, with Sundays standing out as a day marked by elevated fatal/severe injury crashes. This trend emphasizes the need for targeted safety measures during weekends, potentially addressing factors like safety enhancement for increased leisure activities, varied road usage patterns, and potential tolerance in adherence to safety protocols. Moreover, the weekday association rules emphasize the significance of diverse attributes, including driver distraction, pedestrian impairment, and specific roadway types. These insights can initiate effective interventions, such as targeted awareness campaigns, improved signage, and enhanced weekday law enforcement, ultimately contributing to an overarching reduction in pedestrian crash severity.

Shifting the focus to the examination centered on lighting conditions, this study revealed the association between factors influencing pedestrian crashes in daylight and nighttime. Exploring different severity levels highlights essential distinctions, delineating specific scenarios where interventions can have the most significant impact. In the daytime, fully controlled access roadways, non-distracted drivers, and impaired pedestrians are critical factors in fatal/severe injury crashes. Commercial areas and adverse weather conditions in residential zones correlate with moderate injury crashes, while straight road alignments and two-way roadways are associated with minor/no injury crashes. These findings emphasize the complex nature of daytime crashes, urging for dynamic and effective strategies addressing both infrastructure and human factors.

Conversely, nighttime shows different patterns of risk factors, with impaired pedestrians and specific vehicle movements taking precedence in fatal/severe injury crashes. The absence of crashes with motor vehicles and the involvement of non-motorised vehicles emerges as pivotal factors in moderate injury crashes. Lower speed limits, specific crash types, and the presence of young pedestrians characterise minor/no-injury crashes during nighttime. These insights show the need for targeted interventions designed for the unique challenges posed by reduced visibility and altered traffic dynamics after dark.

In merging the conclusions drawn from analyses based on the day of the week and lighting conditions, readers can understand pedestrian crash-related factor association based on the time of the day and day of the week. The findings in the temporal analyses highlighted the significant association of lighting conditions with pedestrian crash severity type. These findings call for comprehensive strategies considering the dynamic interplay of temporal patterns, lighting conditions, and the diverse array of influencing factors. Some of the potential solutions include targeted awareness campaigns, infrastructural enhancements such as installation of proper lighting and audio or visual warning signs, and stringent enforcement protocols to create a safer environment for pedestrians, regardless of the day of the week or lighting conditions. Furthermore, increasing driving under the influence (DUI) checkpoints and public education campaigns can address the critical impact of impaired driving. Improving street lighting and installing reflective signage can reduce nighttime crashes. Implementing traffic calming measures on one-way streets and launching seasonal safety campaigns can mitigate moderate injury crashes. Stricter enforcement of 25 MPH speed limits and specific interventions on two-way roadways can reduce crash severity. Targeted education for high-risk groups and stricter distracted driving laws can further improve safety. Regular safety audits and community involvement are essential for identifying hazards and developing effective solutions. Utilising data-driven policymaking will prioritise interventions in high-risk areas, leading to safer midblock pedestrian crossings. These findings can also help transportation planners to understand better the key factors responsible for pedestrian crashes at midblock. Finally, implementing the mentioned policies might reduce certain types of pedestrian crashes at midblock.

Our study has several limitations. Firstly, the reliance on association rules mining implies correlation rather than causation, making it challenging to establish precise cause-and-effect relationships between identified factors and pedestrian crashes. Secondly, the generalizability of the findings may be constrained by the specific geographical focus of the study. Thirdly, the current study was based on police-reported crash data. This data may be subject to variations in reporting practices across jurisdictions and may not capture all pedestrian crashes, especially those that go unreported or are less severe. Furthermore, the effectiveness of the recommended interventions may vary based on local context, community behavior, and other contextual factors that are not explicitly considered in the analysis. Addressing these limitations would require strategies to enhance data completeness and accuracy, such as improved reporting mechanisms or data integration from multiple sources and multiple geographical locations with advanced modeling techniques, to ensure a more robust and representative analysis of pedestrian crash patterns.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Rohit Chakraborty http://orcid.org/0000-0002-1660-9764
Subasish Das http://orcid.org/0000-0002-1671-2753
Md Nasim Khan http://orcid.org/0000-0001-5996-091X

References

Abdullah, M., T. Oguchi, and C. Dias. 2019. Self-Reported Pedestrian Mid-Block Crossing Behavior: Effects of Gender, Age and Region. p. 15p.

Agrawal, R., T. Imielinski, and A. Swami. 1993. "Mining Association in Large Databases." *Proceedings of the ACM SIGMOD International Conference on Management of Data-SIGMOD* 93:207–216. https://doi.org/10.1145/170036.170072.

Angioi, F., and M. Bassani. 2022. "The Implications of Situation and Route Familiarity for Driver-Pedestrian Interaction at Uncontrolled Mid-Block Crosswalks." *Transportation Research Part F: Traffic Psychology and Behaviour* 90:287–299. https://doi.org/10.1016/j.trf.2022.09.003.

Angulo, A. V., and B. L. Smith. 2021. "Evaluation of Driver Performance with a Prototype Cyber Physical Mid-Block Crossing Advanced Warning System." *Journal of Safety Research* 79:237–245. https://doi.org/10.1016/j.jsr.2021.09.004.

Avinash, C., S. Jiten, A. Shriniwas, J. Gaurang, and P. Manoranjan. 2019a. "Evaluation of Pedestrian Safety Margin at Mid-Block Crosswalks in India." *Safety Science* 119:188–198. https://doi.org/10.1016/j.ssci.2018.12.009.

Avinash, C., S. Jiten, A. Shriniwas, J. Gaurang, and P. Manoranjan. 2019b. "Investigating Effect of Surrounding Factors on Human Behaviour at Un-Controlled Mid-Block Crosswalks in Indian Cities." *Safety Science* 119:174–187. https://doi.org/10.1016/j.ssci.2018.11.006.

Chen, Y., J. Ma, and N. Chen. 2019. "Analyzing Pedestrian Fatality Risk in Accidents at Mid-Blocks." *Journal of Transportation Technologies* 9 (02): 171–192. https://doi.org/10.4236/jtts.2019.92011.

Das, A., M. M. Ahmed, and A. Ghasemzadeh. 2019. "Using Trajectory-Level SHRP2 Naturalistic Driving Data for Investigating Driver Lane-Keeping Ability in fog: An Association Rules Mining Approach." *Accident Analysis and Prevention* 129:250–262. https://doi.org/10.1016/j.aap.2019.05.024.

Das, S., and X. Sun. 2014. Investigating the Pattern of Traffic Crashes Under Rainy Weather by Association Rules in Data Mining. Presented at the Transportation Research Board 93rd Annual MeetingTransportation Research Board.

GHSA. 2022. Pedestrian Traffic Fatalities by State: 2022 Preliminary Data | GHSA [WWW Document]. URL https://www.ghsa.org/resources/Pedestrians23 (accessed 12.18.23).

GHSA. 2023. Drivers Hit and Killed More Than 7,500 Pedestrians Last Year, Most Since 1981, New Projection Shows | GHSA [WWW Document]. URL https://www.ghsa.org/resources/news-releases/GHSA/Pedestrian-Spotlight-Full-Report23 (accessed 2.19.24).

Gooch, J., I. Hamilton, B. Polin, R. Tanzen, and T. Cohen. 2022. "Systemic Safety Analysis of Mid-block Pedestrian Crashes in Massachusetts." *Transportation Research Record* 2676 (11): 722–730. https://doi.org/10.1177/03611981221094566.

Haghighi, M., S. B. Johnson, X. Qian, K. F. Lynch, K. Vehik, and S. Huang. 2016. "A Comparison of Rule-Based Analysis with Regression Methods in Understanding the Risk Factors for Study Withdrawal in a Pediatric Study." *Scientific Reports* 6:30828. https://doi.org/10.1038/srep30828.

Hahsler, M., B. Grun, and K. Hornik. 2005. ". Arules-A Computational Environment for Mining Association Rules and Frequent Item Sets." *Journal of Statistical Software* 14), https://doi.org/10.18637/jss.v014.i15.

He, S.-B., Y.-J. Wang, Y. Sun, W.-W. Gao, Q. Chen, and Y.-Q. An. 2008. The Research of Multidimensional Association Rule in Traffic Accidents, in: 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing. Presented at the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1–4. https://doi.org/10.1109/WiCom.2008.2844.

Hossain, A., X. Sun, R. Thapa, and J. Codjoe. 2022a. "Applying Association Rules Mining to Investigate Pedestrian Fatal and Injury Crash Patterns Under Different Lighting Conditions." *Transportation Research Record* 2676 (6): 659–672. https://doi.org/10.1177/03611981221076120.

Hossain, M. M., H. Zhou, M. A. Rahman, S. Das, and X. Sun. 2022b. "Cellphone-distracted Crashes of Novice Teen Drivers: Understanding Associations of Contributing Factors for Crash Severity Levels and Cellphone Usage Types." *Traffic Injury Prevention* 23 (7): 390–397. https://doi.org/10.1080/15389588.2022.2097667.

Hossain, M. M., H. Zhou, X. Sun, A. Hossain, and S. Das. 2024. "Crashes Involving Distracted Pedestrians: Identifying Risk Factors and Their Relationships to Pedestrian Severity Levels and Distraction Modes." *Accident Analysis & Prevention* 194:107359. https://doi.org/10.1016/j.aap.2023.107359.

Islam, M. 2023. "An Exploratory Analysis of the Effects of Speed Limits on Pedestrian Injury Severities in Vehicle-Pedestrian Crashes." *Journal of Transport & Health* 28:101561. https://doi.org/10.1016/j.jth.2022.101561.

Jain, U., and R. Rastogi. 2018. "Risk Assessment for Pedestrian Safety Evaluation at Mid-Block Crossings Under Mixed Traffic Conditions." *ITE Journal* 88:44–49.

Kumfer, W., L. Thomas, L. Sandt, and B. Lan. 2019. "Midblock Pedestrian Crash Predictions in a Systemic, Risk-Based Pedestrian Safety Process." *Transportation Research Record* 2673 (11): 420–432. https://doi.org/10.1177/0361198119847976.

Lee, J., M. Abdel-Aty, and X. Jiang. 2015. "Multivariate Crash Modeling for Motor Vehicle and non-Motorized Modes at the Macroscopic Level." *Accident Analysis & Prevention* 78:146–154. https://doi.org/10.1016/j.aap.2015.03.003.

Marukatat, R. 2007. "Structure-Based Rule Selection Framework for Association Rule Mining of Traffic Accident Data." In *Computational Intelligence and Security, Lecture Notes in Computer Science*, edited by Y. Wang, Y. Cheung, and H. Liu, 231–239. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-74377-4_25

Mondal, A. R., M. A. E. Bhuiyan, and F. Yang. 2020. "Advancement of Weather-Related Crash Prediction Model Using Nonparametric Machine Learning Algorithms." *SN Applied Sciences* 2 (8): 1372. https://doi.org/10.1007/s42452-020-03196-x.

Montella, A., M. Aria, A. D'Ambrosio, and F. Mauriello. 2011. "Data-Mining Techniques for Exploratory Analysis of Pedestrian Crashes." *Transportation Research Record* 2237 (1): 107–116. https://doi.org/10.3141/2237-12.

Nasernejad, P., T. Sayed, and R. Alsaleh. 2023. "Multiagent Modeling of Pedestrian-Vehicle Conflicts Using Adversarial Inverse Reinforcement Learning." *Transportmetrica A: Transport Science* 19 (3): 2061081. https://doi.org/10.1080/23249935.2022.2061081.

Osman, M., R. Paleti, S. Mishra, and M. M. Golias. 2016. "Analysis of Injury Severity of Large Truck Crashes in Work Zones." *Accident Analysis & Prevention* 97:261–273. https://doi.org/10.1016/j.aap.2016.10.020.

Pande, A., and M. Abdel-Aty. 2009. "Market Basket Analysis of Crash Data from Large Jurisdictions and its Potential as a Decision Support Tool." *Safety Science* 47 (1): 145–154. https://doi.org/10.1016/j.ssci.2007.12.001.

Park, J., M. Abdel-Aty, and J. Lee. 2019. "School Zone Safety Modeling in Countermeasure Evaluation and Decision." *Transportmetrica A: Transport Science* 15 (2): 586–601. https://doi.org/10.1080/23249935.2018.1519646.

Puterski, R., E. S. Neumann, Z. Cui, and W. Voddraska. 1999. *Evaluation of Mid-Block Pedestrian Crash Patterns Within Las Vegas (Digital/other*.

Quistberg, D. A., E. J. Howard, B. E. Ebel, A. V. Moudon, B. E. Saelens, P. M. Hurvitz, J. E. Curtin, and F. P. Rivara. 2015. "Multilevel Models for Evaluating the Risk of Pedestrian–Motor Vehicle Collisions at Intersections and mid-Blocks." *Accident Analysis & Prevention* 84:99–111. https://doi.org/10.1016/j.aap.2015.08.013.

Rahman, M. A., Md.M. Hossain, E. Mitran, and X. Sun. 2021. "Understanding the Contributing Factors to Young Driver Crashes: A Comparison of Crash Profiles of Three age Groups." *Transportation Engineering* 5:100076. https://doi.org/10.1016/j.treng.2021.100076.

Sherony, R., and C. Zhang. 2015. Pedestrian and Bicyclist Crash Scenarios in the U.S, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. Presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems - (ITSC 2015), IEEE, Gran Canaria, Spain, pp. 1533–1538. https://doi.org/10.1109/ITSC.2015.250.

Tageldin, A., and T. Sayed. 2019. "Models to Evaluate the Severity of Pedestrian-Vehicle Conflicts in Five Cities." *Transportmetrica A: Transport Science* 15 (2): 354–375. https://doi.org/10.1080/23249935.2018.1477853.

Tao, W., M. Aghaabbasi, M. Ali, A. H. Almaliki, R. Zainol, A. A. Almaliki, and E. E. Hussein. 2022. "An Advanced Machine Learning Approach to Predicting Pedestrian Fatality Caused by Road Crashes: A Step Toward Sustainable Pedestrian Safety." *Sustainability* 14:2436. https://doi.org/10.3390/su14042436.

Toran Pour, A., S. Moridpour, A. Rajabifard, and R. Tay. 2017a. "Spatial and Temporal Distribution of Pedestrian Crashes in Melbourne Metropolitan Area." *Road & Transport Research* 26.

Toran Pour, A., S. Moridpour, R. Tay, and A. Rajabifard. 2017b. Influencing factors on vehicle-pedestrian crash severity of school-aged pedestrians. Presented at the URBAN TRANSPORT 2017, Rome, Italy, pp. 465–475. https://doi.org/10.2495/UT170401.

Toran Pour, A., S. Moridpour, R. Tay, and A. Rajabifard. 2017c. "Neighborhood Influences on Vehicle-Pedestrian Crash Severity." *Journal of Urban Health* 94:855–868. https://doi.org/10.1007/s11524-017-0200-z.

Toran Pour, A., S. Moridpour, R. Tay, and A. Rajabifard. 2017d. "Modelling Pedestrian Crash Severity at mid-Blocks." *Transportmetrica A: Transport Science* 13:273–297. https://doi.org/10.1080/23249935.2016.1256355.

Toran Pour, A., S. Moridpour, R. Tay, and A. Rajabifard. 2018. "Influence of Pedestrian age and Gender on Spatial and Temporal Distribution of Pedestrian Crashes." *Traffic Injury Prevention* 19 (1): 81–87. https://doi.org/10.1080/15389588.2017.1341630.

TXDOT. 2023. Pedestrian Safety Campaign [WWW Document]. URL http://txdot.gov/en/home/safety/traffic-safety-campaigns/pedestrian-safety.html (accessed 12.18.23).

Valen, A., Stig Tore Bogstrand, Vigdis Vindenes, Joachim Frost, Magnus Larsson, Anders Holtan, and Hallvard Gjerde. 2019. "Driver-related Risk Factors of Fatal Road Traffic Crashes Associated with Alcohol or Drug Impairment." *Accident Analysis & Prevention* 131:191–199. https://doi.org/10.1016/j.aap.2019.06.014.

Xue, G., and H. Wen. 2024. "Pedestrian-injury Severity Analysis in Pedestrian-Vehicle Crashes with Familiar and Unfamiliar Drivers." *Transportmetrica A: Transport Science* 20 (1): 2120784. https://doi.org/10.1080/23249935.2022.2120784.

Zegeer, C. V., and M. Bushell. 2012. "Pedestrian Crash Trends and Potential Countermeasures from Around the World. Accident Analysis & Prevention, Safety and Mobility of Vulnerable Road Usears: Pedestrians." *Bicyclists, and Motorcyclists* 44:3–11. https://doi.org/10.1016/j.aap.2010.12.007.

Zhang, X., H. Wen, T. Yamamoto, and Q. Zeng. 2021. "Investigating Hazardous Factors Affecting Freeway Crash Injury Severity Incorporating Real-Time Weather Data: Using a Bayesian Multinomial Logit Model with Conditional Autoregressive Priors." *Journal of Safety Research* 76:248–255. https://doi.org/10.1016/j.jsr.2020.12.014.

Zhang, C., B. Zhou, G. Chen, and F. Chen. 2017. "Quantitative Analysis of Pedestrian Safety at Uncontrolled Multi-Lane Mid-Block Crosswalks in China." *Accident Analysis & Prevention* 108:19–26. https://doi.org/10.1016/j.aap.2017.08.018.

Zhu, H., A. Almukdad, M. Iryo-Asano, W. K. M. Alhajyaseen, H. Nakamura, and X. Zhang. 2021a. "A Novel Agent-Based Framework for Evaluating Pedestrian Safety at Unsignalized Mid-Block Crosswalks." *Accident Analysis & Prevention* 159), https://doi.org/10.1016/j.aap.2021.106288.

Zhu, H., M. Iryo-Asano, W. K. M. Alhajyaseen, H. Nakamura, and C. Dias. 2021b. "Interactions Between Autonomous Vehicles and Pedestrians at Unsignalized Mid-Block Crosswalks Considering Occlusions by Opposing Vehicles." *Accident Analysis & Prevention* 163:106468. https://doi.org/10.1016/j.aap.2021.106468.