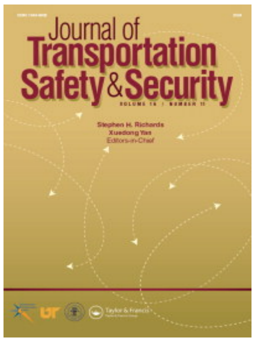

Journal of Transportation Safety & Security

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/utss20

Identifying attribute associations in fatal speeding crashes using latent class clustering and association rule mining

M. Ashifur Rahman, Rohit Chakraborty, Subasish Das, Nurul-Haq Mohammed, Md. Mahmud Hossain & Siam Junaed

To cite this article: M. Ashifur Rahman, Rohit Chakraborty, Subasish Das, Nurul-Haq Mohammed, Md. Mahmud Hossain & Siam Junaed (17 Nov 2024): Identifying attribute associations in fatal speeding crashes using latent class clustering and association rule mining, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2024.2429095

To link to this article: https://doi.org/10.1080/19439962.2024.2429095

Published online: 17 Nov 2024.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2024.2429095

Identifying attribute associations in fatal speeding crashes using latent class clustering and association rule mining

M. Ashifur Rahman[a,b], Rohit Chakraborty[c], Subasish Das[c], Nurul-Haq Mohammed[b], Md. Mahmud Hossain[d], and Siam Junaed[a]

[a]Department of Civil Engineering, University of Louisiana, Lafayette, Louisiana, USA; [b]Special Studies, Louisiana Transportation Research Center, Baton Rouge, Louisiana, USA; [c]Department of Civil Engineering, Texas State University, San Marcos, Texas, USA; [d]Department of Civil Engineering, Auburn University, Auburn, Alabama, USA

ABSTRACT

Speeding has been distinguished as one of the most frequent and persistent contributing factors and is a critical contributing factor to the degree of injury severity. In the United States, at least a quarter of nationwide annual fatal crashes during the last decade involved speeding. There is still a need for an overarching look at crashes involving speeding by considering a wider set of crashes, roadway, driver, and vehicle characteristics. Despite extensive research on speeding-related crashes, there is limited understanding of how collective variables in homogeneous crash clusters contribute to fatal speeding crashes. This paper addresses this gap by investigating these collective impacts using fatal crash data from the Fatality Analysis Reporting System (FARS). Using crash data from the 2015–2019 FARS repository, this study applies latent class clustering (LCC) to obtain homogeneous clusters of fatal speeding crashes, addressing the unobserved heterogeneity. Association rule mining (ARM) has been applied to homogeneous clusters to find hidden patterns. The finding of association rules, such as motorcycle speeding, single vehicle crashes during weekends, in dark, unlit conditions, etc. The results of this research and interpretative findings are expected to improve the knowledge of speeding-related crash mechanisms and to provide important insights on countermeasure development.

KEYWORDS

Speeding; latent class clustering; association rule mining; speeding countermeasures

1. Introduction

Speeding is one of the most frequent and serious contributors to fatal crashes in the United States. In the latest national roadway safety strategy, speeding has been distinguished as one of the three most frequent and

CONTACT M. Ashifur Rahman ashifur@louisiana.edu; ashifur@outlook.com; ashifur.rahman@la.gov Department of Civil Engineering, University of Louisiana, Lafayette, Louisiana, USA

persistent contributing factors alongside alcohol impairment and non-usage of seatbelts (USDOT, 2022). The U.S. Department of Transportation has recently integrated the Safe System approach as a fundamental element of its strategic plan, with Safe Speed designated as a central pillar (USDOT, 2022). In alignment with this approach, maintaining safe speeds continues to be a primary objective of the National Roadway Safety Strategy, aimed at reducing speeding-related crashes. Despite extensive research on speeding-related crashes, the interactions between multiple factors in homogeneous crash clusters remain underexplored. Understanding the collective impact of various factors—such as roadway conditions, driver characteristics, and vehicle types—on speeding-related fatalities is crucial for reducing crash rates.

Although speeding-related crashes could encompass a vast number of speeding scenarios associating different driver and vehicle characteristics, speeding behaviors indeed could be affected by location type (urban/rural), geometric configurations of roadway (curve/straight, intersection type), and/or other driving behaviors. For instance, it has been found to be linked to other aggressive behavior such as alcohol impaired driving (Høye, 2020). Identification of the associative attributes that depict speeding crash scenarios and result in the highest injury severity is critical to the development of the building blocks of the safe speed strategy under the Safe Systems approach.

The NHTSA's Fatality Analysis Reporting System (FARS) repository contains rich quality data encapsulating information of crashes that may not be found in typical crash repositories. Managed by a network of analysts from all US states who undergo formal training, FARS ensures that each data point—from police reports to medical examiner records—is consistently recorded and subject to rigorous quality checks. This involves automatic online verifications for data range and consistency, coupled with ongoing reviews to ensure accuracy, completeness, and timeliness. Such meticulous data handling enhances the reliability of FARS data, making it an invaluable resource for understanding traffic safety at a granular level. To gain insight into speeding-related crashes by exploring associative factors, the research team aimed to utilize the dataset.

Fatal crashes are diverse in nature, and it is helpful to identify subgroups with shared characteristics. By doing so, interventions can be tailored to these subgroups, making them more effective in preventing and managing fatal crashes. Whereas several studies have examined the relationship between speeding and crash outcomes, the majority of these have focused on individual factors, such as speed variance, or have explored limited crash scenarios. Few studies have investigated the collective impact of multiple variables within homogenous crash clusters, particularly with a focus on fatal speeding crashes. The lack of a comprehensive, data-driven approach to understanding how various crash, driver, and environmental factors interact within specific crash clusters has left a gap in the literature. Addressing this gap is critical for developing more targeted and effective countermeasures for speeding-related fatalities.

Using crash data from the FARS repository, this study applies latent class clustering (LCC) to obtain homogeneous clusters of fatal speeding crashes addressing the unobserved heterogeneity. In addition to addressing unobserved heterogeneity through LCC, this research combines Association Rule Mining (ARM) with LCC, an innovative approach for uncovering hidden relationships in crash data. By applying these methods to the FARS data from multiple years, this research introduces a novel way to examine the collective impact of crash, roadway, and driver characteristics on fatal speeding crashes. Moreover, the insights gained from this research can inform the development of more targeted interventions and safety policies, providing a practical framework for reducing speeding-related fatalities. This combined use of advanced data analysis techniques and real-world applications represents a key contribution to the field.

2. Literature review

A plethora of studies exist that involve speed and crashes, a selected review of which shows the breadth of research that has been conducted. Many studies examined the relationship between speed measures and crash outcomes such as crash frequencies and crash severity types. It is generally concluded that higher operating speed is associated with higher severity. However, several studies investigating the association between higher speed and crash occurrence, by using probe vehicle data, produce inconclusive results (Dutta & Fontaine, 2019; Das et al., 2022; Hutton et al., 2020; Park et al., 2021). It is mostly due to the data aggregation process and roadway facility types. Some early studies explored how to define the speed-crash relationship (Elvik, 2014; Hauer, 2009). Hauer (2009) examined the effect of speed on crash likelihood, severity, and highway safety. Elvik (2014) developed power and exponential models to associate traffic crashes with operating speed. Another factor that contributes to crash occurrence is speed distribution. Lee et al. (2002) applied an aggregated log-linear model to predict crash count and found that speed variation and traffic density are strong predictors of crash count. Wang et al. (2018) investigated the relationship between average speed, speed variation, and crash frequency on urban arterials and found that higher average speeds and higher speed variations lead to higher crash frequencies. Garber and Gadiraju (1989) showed that whereas a higher mean speed does not necessarily increase the crash rate, a higher speed variance can. Pei et al. (2012) used disaggregated speed and crash data to investigate the effect of mean operating speed on crash likelihood. They found that this association is the opposite for distance and time exposure (speed-crash association positive for distance and negative for time).

Several studies examined the causal patterns of speeding-related traffic crashes and crash-related injury types. Fitzpatrick et al. (2017) investigated speeding-related crash designation through the development of a series of logistic regression models derived from established speeding-related crash typologies and validated using a blind review of 604 crash narratives by multiple researchers. Only 53.4% of speeding-related crashes had narratives that mentioned speeding as a contributing factor. Using data from Ethiopia, Abegaz et al. (2014) developed a generalized ordered logit/partial proportional odds model to examine the influence of key contributing factors. The identified key factors are impairment, nighttime, and weather. Speeding was found to have varying coefficients for different injury levels (highest in fatal and incapacitating injury crashes). Watson et al. (2015) examined the behavioral patterns of 84,456 speeding offenses. The major classes of speeding offenders (once-only low-range offenders, repeat high-range offenders, and other offenders) were investigated based on their offenses, crash history, and criminal history. Results revealed different distinct patterns for each class of these offenders. Job and Brodie (2022) examined the role of safe speed and speeding behavior on the trauma associated with serious injuries by using data from New Zealand. Table 1 presents the findings from various selected studies that are relevant to research-specific observations on speeding-related crashes.

The cited literature emphasizes the consistent research effort of speeding-related crash studies. Whereas quantifying the speed-crash relationship had been a focal point in earlier studies, understanding the factors contributing to these crashes has gained importance, particularly as relevant crashes increased. Recent methodologies have advanced our understanding of factors affecting crash severity levels. However, a comprehensive examination of speeding-related fatal crashes, incorporating a broader spectrum of crash-, roadway-, driver-, and vehicle-related characteristics, is still necessary. Analyzing the entire dataset without segregating specific crash subgroups overlooks existing data heterogeneity. This lack of segmentation makes it difficult to capture the subtle differences in crash patterns, which are essential for developing targeted safety countermeasures. This paper addresses this research gap by analyzing the collective impacts of variables within homogeneous crash clusters, specifically focusing on fatal crashes using FARS data.

Table 1. Key selected studies and associated findings.

Author	Year	Methodology	Objective	Key Findings
(Yuan et al., 2023)	2023	Random parameter order models (with heterogeneity in means)	To examine global patterns and identifies severe crash clusters to systematically address the spatial distribution of speeding-related crashes.	Speeding-related crash severity varies by district, influenced by factors such as multi-vehicle head-on collisions, intersection types, work zones, dark lighting conditions, rural settings, older drivers, and risky driver behaviors like drunk driving and not wearing seatbelts.
(Wang et al., 2018)	2018	Hierarchical Poisson Log-Normal model (with random effects)	To comprehensively establish the relationship between mean speed, speed variation, and traffic crashes using an urban dataset to formulate effective speed management measures.	On urban arterials, crash frequency increases with both mean speed and speed variation, with a 1% increase in each leading to a 0.7% and 0.74% rise in crashes, respectively.
(Quddus, 2013)	2014	Poisson regression models	To explore the relationship between average speed, speed variations and accident rates.	Speed variation is associated with accident rate, with a 1% increase in speed variation linked to a 0.3% rise in accident rates.
(Park et al., 2021)	2021	Path analysis approach	To examines the relationship between speed and crashes on city streets by jointly modeling speed, roadway characteristics, and crashes.	Medians, curbs, and gutters decrease KABC crash rates, whereas signalized intersections, traffic volume, segment length, and posted speed limits increase them.
(Pei et al., 2012)	2012	Bayesian method (Markov Chain Monte Carlo approach)	To evaluates the relationship between speed and crash risk with respect to distance and time exposure.	• Rainfall, increased lane-changing opportunities, and bus stops heighten crash risk during speeding, particularly on weekends and public holidays. • Central dividers and more diverging ramps mitigate severe crashes during speeding.
(Yu et al., 2018)	2018	Random effect negative binomial model, Random effect logistic regression model	To examine the effect of different data aggregation methods on the relationships between operating speed and traffic safety.	• Operating speed negatively influences crash occurrence risk under low and moderate speed conditions, exhibits ambiguous effects at high speeds, and has a positive impact under free-flow conditions. • Segments with longer auxiliary lengths and more lanes experience fewer crashes and higher traffic volume per lane heightens crash occurrence exposure.
(Høye, 2020)	2020	Crash investigation data analysis	To characterize common risk factors among speeding and impaired (DUI) drivers involved in fatal crashes and propose corresponding mitigation strategies.	Excessively speeding drivers are often DUI, young males, frequently unbelted and unlicensed, driving older vehicles, and prone to single-vehicle crashes during low-volume conditions like nighttime, weekends, and on quiet roads.
(Se et al., 2024)	2023	Random parameter hierarchical ordered probit model	To examines the factors impacting the severity of driver injuries in single-vehicle speeding-	• Speeding in passenger cars or pickup trucks increases fatality risk for younger drivers, whereas rainy weather,

(*continued*)

Table 1. Continued.

Author	Year	Methodology	Objective	Key Findings
			related crashes across various age groups of drivers.	flush median roads, and evening peak hours elevate risk for older drivers. • Both age groups face higher fatal crash risk from speeding on roads without guardrails during adverse weather or on barrier median roads.
(Aarts & Van Schagen, 2006)	2005	Review of prior literature	To explore the impact of driving speed on road safety, emphasizing recent empirical research that investigates the relationship between speed and crash rates.	Lane width, junction density, traffic flow, and speed variability significantly influence the relationship between speed and crash rates.
(Gargoum & El-Basyouny, 2016)	2016	Structural equation modeling	To examines the correlation between average speed and collision frequency, whereas addressing confounding factors influencing this relationship.	Average speed, traffic volume, segment length, shoulder lanes, horizontal alignments and medians significantly affect crash frequency.

3. Methodology

3.1. Research approach

This study seeks to identify and analyze key crash scenarios associated with speeding that result in fatalities, focusing on the factors and unique characteristics of these incidents. Our research strategy is designed to detect patterns in unusual crash cases, which often involve complex interactions among road conditions, driver behaviors, and pedestrian actions leading to fatal results. We began by identifying homogenous data clusters, enabling us to leverage association rule mining to delve deeper into these patterns. The Latent Class Clustering (LCC) method was employed to examine the diverse nature of pedestrian crash data in Louisiana. Further analysis using Association Rule Mining (ARM) on each cluster aims to yield more intricate insights into the dynamics of fatal speeding incidents. Figure 1 illustrates the study's framework, detailing the important steps in our research methodology.

3.2. Data preparation

The FARS data was disaggregated across a wide variety of data tables. Selected important tables—ACCIDENT, DISTRACTION, PERSON, VEHICLE, AND VIOLATION—were merged. The National Highway Traffic Safety Administration (NHTSA) considers a crash to be speeding-related if any driver in the crash was charged with a speeding-related offense or if a police officer indicated that racing, driving too fast for conditions, or exceeding the posted speed limit was a contributing factor in the crash. As mentioned earlier, the annual datasets were filtered for all speeding types—"racing," "exceeded speed limit," "too fast for conditions," as well as "specifics for speeding unknown"—to capture all speeding-related crashes. In the "specifics for speeding unknown" category, the only known

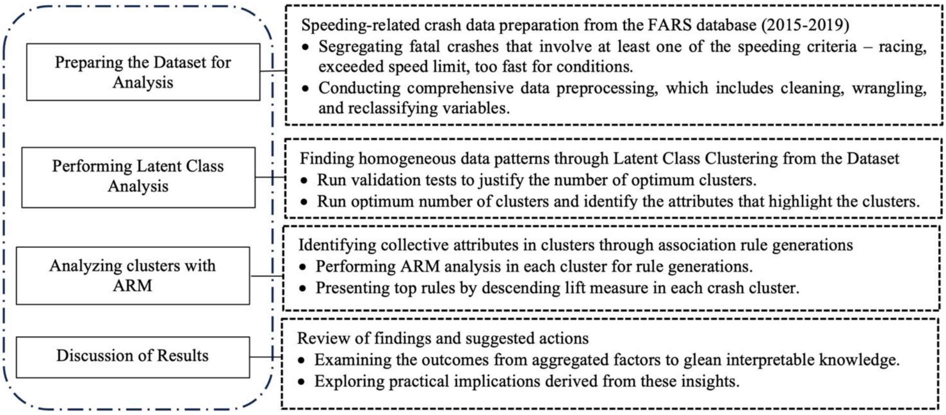
Preparing the Dataset for Analysis
Speeding-related crash data preparation from the FARS database (2015-2019)
• Segregating fatal crashes that involve at least one of the speeding criteria – racing, exceeded speed limit, too fast for conditions.
• Conducting comprehensive data preprocessing, which includes cleaning, wrangling, and reclassifying variables.
Performing Latent Class Analysis
Finding homogeneous data patterns through Latent Class Clustering from the Dataset
• Run validation tests to justify the number of optimum clusters.
• Run optimum number of clusters and identify the attributes that highlight the clusters.
Analyzing clusters with ARM
Identifying collective attributes in clusters through association rule generations
• Performing ARM analysis in each cluster for rule generations.
• Presenting top rules by descending lift measure in each crash cluster.
Discussion of Results
Review of findings and suggested actions
• Examining the outcomes from aggregated factors to glean interpretable knowledge.
• Exploring practical implications derived from these insights.

Figure 1. Study framework.

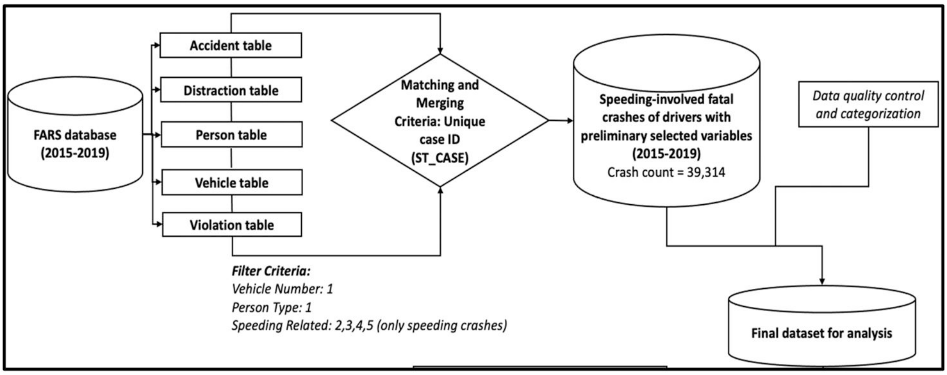
Accident table
Distraction table
Person table
Vehicle table
Violation table
FARS database (2015-2019)
Matching and Merging Criteria: Unique case ID (ST_CASE)
Speeding-involved fatal crashes of drivers with preliminary selected variables (2015-2019)
Crash count = 39,314
Data quality control and categorization
Filter Criteria:
Vehicle Number: 1
Person Type: 1
Speeding Related: 2,3,4,5 (only speeding crashes)
Final dataset for analysis

Figure 2. Crash data preparation.

information is "traveling at a high rate of speed." Collecting five years of data helps to avoid any annual anomalies that may exist in one specific year. The selected variables in the merged dataset from the two databases were categorized based on their frequency, potential research relevance, availability, and engineering judgment, as well as prior literature on speeding. Figure 2 exhibits the analytical framework of this study that also includes data pre-processing techniques.

Substantial research effort to perform data quality control prior to categorizing the variables. For instance, a driver is classified as alcohol-impaired when the Blood Alcohol Concentration (BAC) level is equal to or above 0.08 g/dL per NHTSA guidelines. Several U.S. states evaluate BAC driving at 0.15 g/dL or above as a "high BAC DUI" (driving under the influence) offense. In this study, we considered fixing five categories for the "alcohol" variable: no, BAC $<$ 0.08, BAC $=$ 0.08–0.15, BAC $>$ 0.15, yes_ BAC_unknown, and unknown. The final dataset had a total of 39,314 total crashes. The research team performed cluster validation tests on the final dataset to determine the optimum number of clusters. After clusters had been identified, we ran ARM on each cluster separately to identify frequent crash patterns within clusters.

3.3. Latent class clustering (LCC)

Cluster analysis is a meaningful allocation of observations to groups that are similar with respect to a set of observed variables. Similarity between observations may be defined in various ways depending on data specificities (e.g. measurement scales) and corresponding distance/similarity measures. LCC classifies heterogeneous data into useful homogenous groups, where parameters and properties of these classes (i.e. clusters) are typically unknown (Everitt et al., 2011; Kaufman & Rousseeuw, 1990). LCC addresses observation-specific variation in potentially influential crash characteristics, often referred to as "unobserved heterogeneity" (Mannering et al., 2016). LCC analysis uses a probabilistic approach to obtain meaningful segments associated with categorical variables that can be utilized to discover relationships in large-scale heterogenous crash data. LCC has recently been popular in transportation safety analysis (Depaire et al., 2008; Sasidharan et al., 2015; Sun et al., 2019).

LCC analysis is different from the conventional clustering algorithms that segregate data points by distance such as K-means clustering, hierarchical clustering etc. The LCC is rather a model-based probabilistic clustering approach that segmentalizes observed (i.e. endogenous) variables by an unobserved (i.e. exogenous) nominal categorical variable known as a latent variable. The LCC acts as a type of finite mixture model, as the latent variable involves a membership of a class. Endogenous variables are assumed to be statistically independent given latent variable values, a concept known as "local independence." The cluster parameters are estimated from the class membership probabilities of each crash that were estimated by the maximum likelihood method.

3.4. Theoretical background of LCC

Let us define the index for latent classes as c (where, $c = 1, 2, \ldots, C$), and let α_c represent the probability of a crash belonging to latent class cluster c. For the purpose of identifying clusters, suppose that crashes can be categorized based on M (where $M = 1, 2, \ldots, m$) attributes. The vector $Z_i = (Z_{i1}, Z_{i2}, \ldots Z_{iM})$ represents the attributes of crash i, with each attribute Z_{im} corresponding to one of the possible $\mathrm{r_m}$ categories (from 1 to $\mathrm{r_m}$). The probability $\rho_{\mathrm{m, r_m/c}}^{\mathrm{I(Z_m=r_m)}}$ denotes the likelihood that a crash exhibits the attribute $\mathrm{r_m}$ of characteristic $\mathrm{m_m}$, given its membership in latent class c. Therefore, the probability of observing a specific vector of outcomes can be formulated as follows:

$$\mathrm{P(Z_i = z)} = \sum_{\mathrm{c=1}}^{\mathrm{C}} \alpha_{\mathrm{c}} \prod_{\mathrm{m=1}}^{\mathrm{M}} \prod_{\mathrm{r_m=1}}^{\mathrm{R_m}} \rho_{\mathrm{m, r_m/c}}^{\mathrm{I(Z_m=r_m)}} \tag{1}$$

Where,

$$\mathrm{I(Z_m = r_m)} = \begin{cases} 1 \ \ \mathit{if} \ \ \mathrm{Z_m = r_m} \\ 0 \ \ \mathit{otherwise} \end{cases} \tag{2}$$

The estimated posterior probability of class membership is:

$$\mathrm{P(L = c|Z = z)} = \frac{\left(\prod_{\mathrm{m=1}}^{\mathrm{M}} \prod_{\mathrm{r_m=1}}^{\mathrm{R_m}} \rho_{\mathrm{m}, \frac{\mathrm{r_m}}{\mathrm{c}}}^{\mathrm{I(Z_m=r_m)}}\right) \gamma_{\mathrm{c}}}{\sum_{\mathrm{c=1}}^{\mathrm{C}} \alpha_{\mathrm{c}} \prod_{\mathrm{m=1}}^{\mathrm{M}} \prod_{\mathrm{r_m=1}}^{\mathrm{R_m}} \rho_{\mathrm{m, r_m/c}}^{\mathrm{I(Z_m=r_m)}}} \tag{3}$$

3.5. Bayesian criteria for cluster selection

The researcher determines the optimal number K of latent classes by selecting the configuration that yields the lowest values of several statistical information criteria (Akaike, 1974). These criteria include the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) (Schwarz, 1978), Bozdogan's Consistent AIC (CAIC) (Bozdogan, 1987), and the Sample Size Adjusted BIC (SABIC) (Sclove, 1987). These metrics assess model fit and complexity by modifying the log-likelihood to reflect the number of parameters used. Among these, the BIC and SABIC are often considered more precise than the AIC (Dziak et al., 2020). Nonetheless, in this instance, none of the information criteria indicated a minimized value (Figure 3).

3.6. Entropy criteria for cluster selection

Studies have also recommended using Entropy R^2 to estimate the accuracy of clustering classification, which is a measure of a weighted average of individual's posterior probabilities of membership and can be calculated by

$$\text{Entropy } R^2 = 1 - \frac{\sum_{i=1}^{n}\sum_{c=1}^{K} p_{ik} \ln(p_{ik})}{n \ln\left(\frac{1}{k}\right)} \tag{4}$$

where, p_{ik} is the posterior probability that a crash case i is assigned class k. The entropy values closer to 1 indicates better separation. An 8-cluster-solution showed maximum separation from the Entropy R^2 value (Figure 4).

3.7. Association rule mining

ARM was used to generate rules comprising co-occurring associations of multiple item sets under variables of concern from each cluster. The

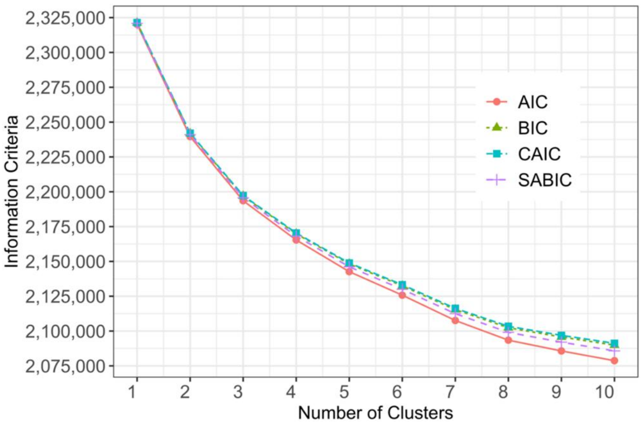

Figure 3. Information criteria by number of clusters.

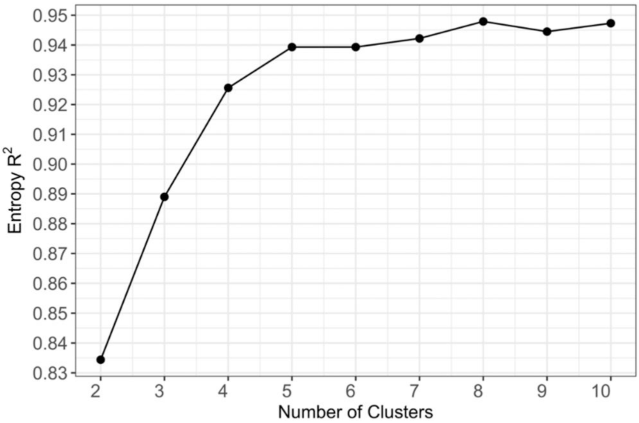

Figure 4. Entropy R^2 by number of clusters.

association rules were quantified through some measures of frequency. Three measurements are commonly used to quantify and rank the association rules:

3.7.1. Support

Support is an indication of the frequency of a combination of items in the dataset. If the antecedent A is a combination of items of variables and the consequent B is the targeted item $A \rightarrow \mathrm{B}$ an association rule and D is a complete observation in a dataset—the support of A denoted as $S(A)$, with regard to observation D is defined by the proportion of observations $'d'$ the dataset, which contains the combination of items A.

3.7.2. Confidence

Confidence is a measure of how often the rule, $A \rightarrow \mathrm{B}$, is true in the dataset i.e. how often each item in B appears in observations that contain S.

$$\mathrm{C(A \rightarrow B)} = \frac{\mathrm{S(AUB)}}{\mathrm{S(A)}} \tag{5}$$

3.7.3. Lift

The lift of a rule A → B is the confidence of the rule divided by the expected confidence, assuming A and B are independent of each other.

$$\mathrm{L(A \rightarrow B)} = \frac{\mathrm{S(AUB)}}{\mathrm{S(A)}*\mathrm{S(B)}} \tag{6}$$

A lift value greater than 1 is an indication that S and T appear more often together than expected. This can be restated as—the occurrence of S

has a positive effect on the occurrence of T or that S is positively correlated with T. A lift value smaller than 1 indicates that S and T appear less often together than expected, and therefore, S is negatively correlated with T. A lift value near 1 indicates that S and T appear almost as often together as expected. This means that the occurrence of S has almost no effect on the occurrence of T or that S and T have zero correlation.

Transportation safety research studies explored various algorithms to efficiently identify frequent itemsets tailored to the given dataset structure (Das et al., 2018; Das et al., 2021; Maimon & Rokach, 2010; Rahman, Sun, Sun, et al. 2021; Rahman, Sun, Das, et al. 2021; Rahman et al., 2023). One of the widely adopted approaches is the Apriori algorithm, originally introduced by Agrawal et al. (1993). This algorithm was specifically designed to find recurring itemsets using predefined threshold values, operating under the assumption that if a particular itemset is frequent, then all of its subsets must also be frequent.

Apriori uses a bottom-up strategy, starting with individual items and progressively expanding each subset to generate candidate itemsets. These candidate sets are evaluated iteratively against the dataset, and the algorithm continues building upon the successful expansions until no additional frequent itemsets can be identified. By employing a breadth-first search methodology, Apriori effectively navigates the search space, minimizing computational cost. For those interested in a deeper understanding, a comprehensive explanation can be found in Agrawal and Srikant (1994). To maintain consistency and enable a potential comparison of association rules among clusters, we applied a common threshold of $S \geq 0.1$, $C \geq 0.25$, and $L \geq 1.2$ for ARM across all clusters, based on the data distribution and after conducting several trial-and-error iterations. As the rules are generated within the clusters rather than across the entire dataset, they are expected to uncover hidden yet meaningful crash patterns by presenting the association of both frequent and infrequent factors.

4. Results

4.1. Cluster description

The current study utilized Latent Class Clustering (LCC) to determine the clusters and subsequently applied Association Rule Mining (ARM) to identify key rules within each cluster. Based on the cluster selection criteria, eight clusters were found to be the most optimal number for our dataset. Consequently, an analysis of eight clusters was conducted using LCC. The LCC analysis resulted in eight homogeneous clusters, labeled C1 through C8. Cluster 1 (C1) encompasses 23.82% of the dataset, representing the largest portion, whereas Cluster 8 (C8) is the smallest, comprising 2.34%.

Clusters 2 through 6 each account for between 10% and 15% of the dataset, ensuring a balanced distribution across these middle clusters. Table 1 below presents the distribution of attributes across clusters. It serves as a reference guide for informational purposes, offering a quick and concise overview rather than an in-depth description of cluster conditions. This distribution of variables highlights clusters with disproportionately high crash characteristics, enabling effective profiling and identification based on their distinct attributes. In profiling the clusters, we focused on the overrepresentation of characteristics that distinctly set each cluster apart from the others. The descriptions of the clusters are below:

- **Cluster 1 (C1: 23.82**%): Characterized by single-vehicle crashes (99.98%) on two-lane (97.03%), two-way undivided roads (95.32%), particularly on curves (100%) with no traffic control (90.68%), involving drivers with full licenses (90.29%) and no previous violations (75.04%), typically not at intersections (96.49%).
- **Cluster 2 (C2: 15.97%):** Characterized by single vehicle crashes (99.97%) mostly on straight two-lane, two-way undivided roads (96.73%), with the majority of drivers having a full license (87.14%) and no prior violations (76.76%), often with no traffic control present (92.30%).
- **Cluster 3 (C3: 14.63**%**):** Features single vehicle crashes (99.96%) primarily in urban areas (96.34%) on two-lane roads (47.16%), involving drivers going straight (67.34%) with no prior violations (70.99%), typically at non-intersections (69.01%) in dark-lighted condition (59.26%).
- **Cluster 4 (C4: 12.15%):** Comprised predominantly of single vehicle crashes (99.98%) on interstates, freeways, and expressways (82.16%), involving drivers with no other violations (73.53%) and typically not occurring at intersections (98.90%).
- **Cluster 5 (C5: 11.69**%**):** Largely consists of multivehicle crashes (0% single vehicle) in urban environments (91.15%) on two-lane roads (47.24%) at known intersection types (total of 70.23%), with a significant proportion of incidents involving male drivers (89.15%) with no previous violations (68.71%).
- **Cluster 6 (C6: 11.56**%**):** Characterized by muti-vehicle crashes (0% single vehicle crashes) on two-lane, two-way undivided roads (89.62%) mostly in daylight conditions (65.87%) in rural area (73.07%).
- **Cluster 7 (C7: 7.85%):** Predominantly occurring on interstates, freeways, and expressways (86.94%) at high speeds (60 mph or higher, 75.15%), multivehicle crashes mainly involving front-to-rear collisions (70.06%), often with drivers who are not unrestrained (76.28%).

- **Cluster 8 (C8: 2.34%):** Notably small in size, featuring drivers with unspecified previous violations (99.99% unknown), frequently not unrestrained (73.78%), and often involved in crashes at non-intersections (72.38%).

4.2. Description of rules by cluster

Association rules were generated using the "arules" package (Hahsler et al., 2023) within the R statistical software (R Development Core Team, 2024). The ARM algorithm was applied across the four identified clusters. The top 10 rules from the ARM analysis, sorted by descending lift values, are presented in Tables 2–9. Each table from Tables 3–10 presents the top 10 association rules for the respective clusters. Clusters are titled based on their key unique attributes derived from the cluster descriptions. The top 10 rules using a common threshold of support ($S \geq 0.1$), confidence ($C \geq 0.25$), and lift ($L \geq 1.2$) was used for applying ARM across all clusters are presented and highlighted below.

4.3. Cluster 1 – negotiating the curve whereas speeding on a curved road

Cluster 1 presenting multiple patterns of speeding-related crashes emerge from the top 10 rules, ranked by lift from highest to lowest. Table 3 shows the top 10 association rules from this Cluster 1. Motorcycles are notably the predominant vehicle type in fatal crashes involving speeding on curved roadways. The first pattern involves male motorcycle drivers primarily at fault for crashes occurring from Friday to Sunday, without any passengers. The speeding behavior associated with this pattern is categorized as "too fast for conditions". The rule [*Gender* = *male* + *Passenger_Presence* = *no* + *Speeding_Type* = *too_fast_for_conditions* → *Vehicle_Type* = *motored_cycle*] has a support, confidence, and lift value of 0.112, 0.361, and 1.609 respectively. The support value indicates that 11.2% of crashes in this dataset have this combination of factors. The confidence value indicates that if the antecedents involve these factors, 36.1% of crashes would result in fatalities that involve a motorcycle. The lift value highlights that the percentage of all fatalities containing this combination in cluster 1 is 1.609 times the percentage of all fatalities in overall cluster 1. This finding aligns with the fact that speeding by motorcyclists although navigating a curve and resulting in fatal crashes is often linked to diminished braking effectiveness when entering a curve with excessive speed. The reduced sight distance of a curve can also attribute to the factors and may further increase the risk of fatalities (Xin et al., 2017).

The second pattern involves an alcohol-intoxicated driver with a BAC greater than 0.15 g/dL. Other associated characteristics with this pattern are

Table 2. Attribute distribution among in-between clusters.

Variable	C1	C2	C3	C4	C5	C6	C7	C8	Overall
Cluster Percentage	23.82	15.97	14.63	12.15	11.69	11.56	7.85	2.34	100
Day_of_the_Week									
Fri-Sun	56.59	54.35	54.57	53.86	51.23	45.97	48.21	54.64	53.05
Mon-Thu	43.41	45.65	45.43	46.14	48.77	54.03	51.79	45.36	46.95
Area_Type									
rural	77.4	79.05	3.66	37.67	8.85	73.07	35.65	30.05	49.16
urban	22.6	20.95	96.34	62.33	91.15	26.93	64.35	69.95	50.84
Functional_System									
collector	39.37	32.52	11.67	0.03	9.2	25.87	0.09	14.47	20.69
interstate_freeway_expressway	1.04	0.5	1.82	82.16	1.82	2.25	86.94	16.28	18.25
local	25.92	33.5	18.86	0.09	13.17	8.63	0	19.58	17.29
minor_arterial	19.33	19.57	31.25	0.96	28.03	27.84	0.65	23.06	19.5
principal_arterial	14	13.58	36.14	16.76	47.7	35.27	12.28	26.51	24.06
other	0.34	0.33	0.27	0.02	0.09	0.13	0.03	0.11	0.2
Collision_Manner									
angle	0	0	0	0	62.11	36.05	9.97	13.83	12.53
front_to_front	0	0	0	0	8.7	38.5	5.27	3.91	5.97
front_to_rear	0	0	0	0	21.72	16.27	70.06	9.05	10.13
sideswipe_opposite_direction	0	0	0	0	1.19	4.69	0.33	0.87	0.73
sideswipe_same_direction	0	0	0	0	4.81	3.37	11.41	2.4	1.9
single vehicle	99.98	99.97	99.96	99.98	0	0	0	69.39	68.17
other	0.02	0.03	0.04	0.02	1.47	1.11	2.95	0.54	0.56
Intersection_Type									
4W (four-way intersection)	0.23	1.58	15.77	0.4	49.58	5.33	0.77	17.4	9.54
T (T intersection)	1.93	5.52	12.76	0.48	19.78	6.54	0.16	8.92	6.55
Y	1.08	0.43	1.15	0.12	0.78	0.45	0	0.33	0.66
not_intersection	96.49	92.2	69.01	98.9	29.15	87.66	99.07	72.38	82.83
traffic_circle_roundabout	0.05	0.05	0.57	0.07	0.09	0	0	0	0.12
other	0.23	0.21	0.74	0.03	0.63	0.03	0	0.98	0.3
Lighting_Condition									
dark-lighted	5.73	4.41	59.26	22.59	38.57	3.74	21.13	39.58	21.01
dark-unknown_lighting	0.75	0.87	0.44	0.32	0.42	0.31	0.27	1.2	0.55
dark-unlighted	46.04	44.45	9.99	29.63	7.27	24.48	28.69	24.09	29.62
dawn_dusk	4.47	4.88	3.4	3.81	4.28	5.53	3.64	3.26	4.31
daylight	42.23	44.73	26.68	43.37	49.3	65.87	46.18	31.87	44.12
other	0.77	0.66	0.23	0.28	0.16	0.07	0.08	0	0.39
Weather_Condition									
clear	66.19	64.76	73.22	60	73.8	53.25	70.54	72.05	66.11
cloudy	15.44	13.96	12.76	13.69	13.63	14.75	12.81	12.2	14.03
fog_smog_smoke	1.41	1.42	0.7	1.13	0.63	1.49	1.23	1.09	1.17
rain	7.27	9.41	7.15	14.43	4.44	15.05	7.53	8.81	9.09
snow_sleet_hail	1.36	2.53	0.46	4.52	0.57	8.2	2.68	1.63	2.61
other	8.32	7.93	5.7	6.24	6.93	7.27	5.2	4.23	7
Passenger_Presence									
no	70.07	66.18	67.8	65.25	74.9	70.78	76.23	48.05	69.15
one	19.49	20.26	20.25	19.98	15.95	19.84	16.09	18.23	19.11
more_than_one	10.44	13.56	11.95	14.77	9.15	9.39	7.69	33.72	11.74
Prev_Speed_Violations									
none	75.04	76.76	70.99	73.53	68.71	77.06	71.56	0.01	72
one	15.18	14.61	16.62	15.47	16.81	14.3	17.05	0	15.21
more_than_one	9.77	8.63	12.4	10.99	14.48	8.62	11.18	0	10.42
unknown	0.01	0	0	0	0	0.02	0.21	99.99	2.36
Other_Prev_Violations									
at_least_one	1.8	1.68	1.86	1.06	1.65	1.27	1.28	0	1.54
more_than_one	0.33	0.28	0.21	0.07	0.29	0.13	0.13	0	0.22
none	97.87	98.04	97.93	98.87	98.06	98.6	98.41	0.01	95.89
unknown	0	0	0	0	0	0	0.18	99.99	2.35
Trafficway									
2WD_median_barrier	0.68	0.1	6.33	51.25	5.74	0.5	64.86	12.72	13.44
2WD_unprotected_median	3.54	2.93	27.55	28.66	29.44	5.98	28.38	18.59	15.62
2WU	95.32	96.73	47.16	0.47	47.24	89.62	1.35	57.62	62.45
2WU_with_left-turn-lane	0.08	0	13.34	0.84	14.24	3.59	0.66	6.31	4.35
one-way	0.25	0.18	4.45	2.74	2.62	0.01	1.4	2.83	1.55
ramp	0.02	0	0.79	16.01	0.33	0.28	3.32	1.5	2.43

(continued)

Table 2. Continued.

Variable	C1	C2	C3	C4	C5	C6	C7	C8	Overall
other	0.12	0.06	0.38	0.04	0.39	0.02	0.04	0.44	0.16
Number_of_Lanes									
one	0.21	0.19	2.77	10.91	0.71	0.21	2.69	1.94	2.17
two	97.03	97.91	48.06	52.21	37.39	86.59	43.26	61.03	71.33
three	1.25	0.52	16.64	20.08	16.08	4.38	21.01	10.85	9.54
four	1.09	0.79	15.55	11.18	19.9	5.31	17.54	14.57	8.67
five	0.02	0.19	12.59	3.79	17.1	3.12	9.45	7.16	5.61
six	0	0	2.02	1.06	3.94	0.08	4.15	2.07	1.27
more_than_six	0	0	1.32	0.34	3.19	0.05	1.24	1.52	0.75
unknown	0.4	0.39	1.05	0.43	1.68	0.26	0.64	0.87	0.66
Speed_Limit									
25mph_or_less	6.4	7.17	16.36	1.23	9.56	2.11	0.06	13.05	6.88
30–35_mph	19.48	16.76	41.79	2.2	33.45	10.55	0.48	33.17	19.64
40–45_mph	25.19	21.53	34.82	5.87	42.7	21.35	2.62	19.13	23.36
50–55_mph	41.85	46	6.78	22.22	12.95	49.48	21.69	18.9	30.38
60mph_or_higher	7.08	8.54	0.26	68.48	1.33	16.52	75.15	15.75	19.74
Horizontal_Alignment									
curve	100	0.93	26.72	40.23	2.48	40.15	8.66	22.73	38.91
straight	0	98.58	68.98	58.72	93.67	59.37	88.9	74.44	59.5
other	0	0.49	4.3	1.05	3.85	0.48	2.43	2.83	1.6
Vertical_Alignment									
grade	40.02	23.19	17.82	29.36	12.38	31.89	19.34	18.54	26.49
hillcrest	3.29	4.29	1.46	1.87	2.12	4.77	2.17	2.36	2.94
level	52.36	67.95	66.94	61.51	75.23	58.42	68.96	66.27	63.09
other	4.33	4.57	13.79	7.26	10.26	4.92	9.53	12.83	7.48
Traffic_Control									
none	90.68	92.3	78.2	94	51.45	89.57	93.1	77	84.67
other	6.95	2.68	2.36	3.36	1.13	5.29	4.83	3.96	4.05
other_regulatory_signs	2.01	1	1.06	1.53	0.46	2.21	1.12	0.98	1.4
stop_sign	0.3	3.91	4.78	0.32	8.82	2.76	0.03	4.46	2.89
traffic_signal	0.06	0.12	13.6	0.79	38.13	0.17	0.91	13.59	6.98
Driver_Age									
19 y_or_less	10.94	14.1	9.79	7.12	10.03	12.97	3.5	12.07	10.38
20–29 y	30.09	30.38	42.7	34.04	42.32	31.94	32.67	20.27	34.08
30–39 y	19.21	20.42	22.46	22.07	22.64	19.07	23.72	11.79	20.79
40–49 y	14.69	13.42	10.72	14.87	12.09	12.79	16.32	7.18	13.36
50–59 y	13.84	10.98	7.68	10.91	6.4	11.64	12.31	4.58	10.67
60–69 y	7.02	6.91	3.58	7.01	3.65	6.1	7.75	2.81	5.96
70 y_or_older	4.17	3.78	3	3.98	2.82	5.45	3.67	1.5	3.8
unknown	0.05	0	0.08	0	0.04	0.04	0.06	39.8	0.97
Gender									
female	16.98	19.87	15.43	20.25	10.8	24.36	15.67	8.4	17.44
male	82.97	80.09	84.38	79.71	89.15	75.59	84.21	57.34	81.69
unknown	0.05	0.03	0.19	0.04	0.05	0.04	0.12	34.26	0.87
Alcohol									
BAC $<$ 0.08	4.04	3.98	4.07	3.29	4.1	3.17	3.01	1.72	3.71
BAC $=$ 0.08–0.15	8.06	7.32	6.98	4.87	5.12	3.79	4.88	3.73	6.21
BAC $>$ 0.15	19.98	16.56	15.97	11.52	10.9	8.61	8.52	6.63	14.23
no	32.57	38.23	29.18	41.02	37.51	48.8	47.46	25.47	37.46
unknown	29.92	28.33	37.67	34.81	37.22	32.25	31.7	57.7	33.31
yes_BAC_unknown	5.44	5.58	6.12	4.49	5.16	3.38	4.43	4.76	5.08
Drug									
no	42.95	46.08	36.71	42.81	37.5	45.41	50.52	25.68	42.36
unknown	44.21	40.61	49.58	44.84	46.65	40.88	40.13	67.47	44.62
yes	12.85	13.31	13.71	12.35	15.85	13.71	9.35	6.85	13.02
First_Harmful_Event									
bridge_culvert	4.52	5.13	1.36	4.12	0	0	0	1.63	2.64
curb_ditch_embankment	21.57	19.18	30.84	11.75	0	0	0	9.57	14.36
fence_wall	4.04	4.5	3.02	1.97	0	0	0	1.52	2.4
guardrail_face_end	5.39	2.93	1.5	16.87	0	0	0	2.61	4.08
mailbox	2.63	3.18	0.69	0.07	0	0	0	0.11	1.25
motor-vehicle_in_transport	0	0	0	0	100	100	100	30.61	31.81
other	4.3	4.6	5.32	3.63	0	0	0	2.72	3.04
other non-collision	0.67	1.39	1.29	1.22	0	0	0	0.44	0.73

(*continued*)

Table 2. Continued.

Variable	C1	C2	C3	C4	C5	C6	C7	C8	Overall
other non-fixed_objects	0.79	2.54	1.71	1.53	0	0	0	1.2	1.06
parked_vehicle	0.43	1.19	4.09	3.5	0	0	0	2.07	1.36
pedalcycle	0.13	1.37	3.21	0.13	0	0	0	3.92	0.83
pedestrian	0.36	2.61	15.5	2.81	0	0	0	21.53	3.61
post_pole_support	9.98	8.1	14.83	7.49	0	0	0	6.09	6.89
rollover	21.88	20.5	5.95	23.28	0	0	0	9.46	12.41
traffic_barrier	0.29	0.18	1.33	12.2	0	0	0	1.52	1.81
tree	23.02	22.59	9.36	9.43	0	0	0	5	11.72
Vehicle_Type									
bus	0.03	0	0.07	0.15	0.08	0.09	0.24	0.2	0.08
light_truck	34.9	42.02	25.6	36.06	23.29	34.45	29.03	30.65	32.85
motored_cycle	23.5	11.96	18.25	18.22	33.07	11.51	16.98	10.22	19.16
other	0.54	0.81	0.21	0.32	0.14	0.48	0.88	18.27	0.9
passenger_car	38.57	43.96	55.45	41.08	41.92	49.87	38.61	36.97	43.87
truck	2.47	1.25	0.42	4.17	1.51	3.6	14.27	3.68	3.15
Unrestraint_Driving									
no	41.08	41.38	61.62	57.47	64.38	68.43	76.28	73.78	55.53
yes	58.92	58.62	38.38	42.53	35.62	31.57	23.72	26.22	44.47
License_Type									
full_licensed	90.29	87.14	84.57	90.09	86.38	89.61	95	11.6	86.92
intermediate_license	3.32	4.51	1.81	1.34	1.75	4.42	0.73	0	2.71
learner_permit	1.08	1.21	1.19	0.83	1.14	0.77	0.38	0	0.98
not_licensed	5.09	7.08	12.16	7.57	10.45	5.07	3.78	23.26	7.69
others	0.22	0.06	0.27	0.17	0.28	0.13	0.1	65.13	1.7
Distraction									
cellphone_related	1.09	1.13	0.85	1.33	0.79	1.95	2.23	0.65	1.24
in-vehicle_source	0.53	0.94	0.81	0.78	0.57	0.99	1.97	0.11	0.83
no	37.19	36.16	37.11	39.79	37.66	37.99	42.51	31.4	37.76
other	4.37	5.39	5.96	5.57	5.84	6.47	11.12	6.38	5.9
unknown	56.82	56.37	55.27	52.54	55.14	52.6	42.15	61.46	54.28
Movement_Prior_Crash									
accelerating/decelerating	0.3	0.83	1.38	0.73	1.55	0.85	1.66	0.74	0.92
backing_up	0	0.02	0.03	0	0.02	0	0	0	0.01
changing_lanes/passing/overtaking/merging	1.36	4.48	5.24	11.19	8.37	12.58	17.65	7.22	7.15
disabled/parked/entering_or_leaving_parking_position	0	0.05	0.1	0.02	0	0.02	0.06	0.11	0.04
going_straight	4.36	91.57	67.34	52.11	85.23	48.84	73.26	67.57	54.78
making_a_u-turn	0.02	0.17	0.06	0	0.18	0.05	0.04	0	0.07
negotiating_a_curve	93.29	0	21.13	33.84	1.75	35.73	5.88	19.14	34.67
others	0.42	1.18	1.51	1.64	0.89	1.25	1.18	3.15	1.12
starting_in_roadway	0.02	0.02	0.06	0	0.13	0.02	0	0	0.04
stopped_in_roadway	0	0	0	0	0.02	0.02	0.2	0	0.02
turning_left	0.12	0.97	1.96	0.12	1.45	0.48	0.05	1.09	0.74
turning_right	0.11	0.7	1.18	0.34	0.41	0.14	0.03	0.98	0.44
Vehicle_Age									
0–3 y	12.37	9.34	16.6	15.36	19.78	14.16	27.7	14.67	15.2
4–7 y	11.94	10.8	14.42	12.71	16.02	13.83	20.02	9.23	13.48
8–11 y	20.57	20.03	21.81	23.15	22.05	21.07	20.08	15.41	21.05
12–14 y	18.79	21.12	19.4	20.88	16.74	19.09	14.18	15.23	18.86
15–17 y	15.39	17.4	13.25	14.11	12.39	14.85	9	10.06	14.2
18y_or_older	20.49	20.53	14.23	13.69	12.35	16.72	8.93	13.1	16.29
unknown	0.44	0.78	0.29	0.1	0.67	0.27	0.09	22.29	0.92
License_State									
in-state	91.5	93.17	95.18	82.77	94.78	90.64	80.94	27.11	89.19
out-of-state	8.48	6.8	4.82	17.18	5.17	9.36	19.06	23.3	9.63
unknown	0.02	0.03	0	0.04	0.04	0	0	49.59	1.18
Speeding_Type									
exceeded_speed_limit	38.46	43.38	44.09	31.89	53.32	36.73	25.46	37.49	39.76
racing	0.17	0.28	1.38	0.6	2.24	0.65	0.6	1.52	0.78
specifics_unknown	8.32	14.51	16.64	16.12	16.39	14.11	17.07	17.31	13.98
too_fast_for_conditions	53.06	41.82	37.89	51.39	28.06	48.5	56.87	43.68	45.47

Table 3. Top 10 association rules from cluster 1.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Day_of_the_Week = Fri-Sun + Gender = male + Passenger_Presence = no	Vehicle_Type = motored_cycle	0.125	0.379	1.609
2	Gender = male + Passenger_Presence = no + Speeding_Type = too_fast_for_conditions	Vehicle_Type = motored_cycle	0.112	0.361	1.532
3	Gender = male + Passenger_Presence = no	Vehicle_Type = motored_cycle	0.202	0.342	1.454
4	Day_of_the_Week = Fri-Sun + Passenger_Presence = no	Vehicle_Type = motored_cycle	0.128	0.335	1.425
5	Collision_Manner = single-vehicle + Gender = male + Lighting_Condition = dark-unlighted + Trafficway = 2WU	Alcohol = BAC > 0.15	0.104	0.276	1.384
6	Gender = male + Lighting_Condition = dark-unlighted + Trafficway = 2WU	Alcohol = BAC > 0.15	0.104	0.276	1.383
7	Collision_Manner = single-vehicle + Gender = male + Lighting_Condition = dark-unlighted	Alcohol = BAC > 0.15	0.108	0.276	1.382
8	Gender = male + Lighting_Condition = dark-unlighted	Alcohol = BAC > 0.15	0.108	0.276	1.381
9	Alcohol = BAC > 0.15 + Trafficway = 2WU	Lighting_Condition = dark-unlighted	0.120	0.630	1.366
10	Collision_Manner = single-vehicle + Lighting_Condition = dark-unlighted + Number_of_Lanes = two	Alcohol = BAC > 0.15	0.121	0.270	1.351

male drivers, single vehicle crashes, dark-unlighted conditions, and two-way undivided traffic. The rules suggest that alcohol significantly impairs driving ability, especially in poor visibility conditions, increasing the likelihood of severe crashes. The rule [*Collision_Manner = single-vehicle + Gender = male + Lighting_Condition = dark-unlighted + Trafficway = 2WU* → *Alcohol = BAC > 0.15*] shows a strong association with a high lift value of 1.384 indicating that these conditions are 1.384 times more likely to result in alcohol-related fatalities compared to crashes that include other combination of factors within this cluster. The impairment might cause loss-of-control of the vehicle which can also result in fatality (Shaheed & Gkritza, 2014). Also, the increased BAC level/alcohol impairment might intrigue aggressive driving behavior increasing the risk of fatalities (Rowden et al., 2016). It also emphasizes the danger of nighttime driving combined with alcohol consumption, reinforcing the need for targeted interventions such as improved street lighting and stricter DUI enforcement. Additionally, the rules reveal that specific trafficway configurations, for example, two-way undivided roads, combined with poor lighting, significantly correlate with alcohol-related crashes. The consistent association between dark, unlighted conditions and high BAC levels across multiple rules highlights the responsible risk factors that contribute to fatal crashes.

4.4. Cluster 2 [speeding whereas going straight on a straight road]

Cluster 2 identifies several patterns of speeding-related crashes on straight roads, revealing critical insights into the factors contributing to these incidents. The top 10 association rules from this Cluster 2 are presented in Table 4. The first pattern in this cluster is associated with "exceeded speed limit" as well as unrestrained driving, two-lane, two-way undivided roads, passenger cars, male drivers, and a rural area. The rule [*Gender = male + Number_of_Lanes = two + Trafficway = 2WU + Unrestraint_Driving = yes + Vehicle_Type = passenger_car* → *Speeding_Type = exceeded_speed_limit*] suggests that this demographic, when driving passenger cars on a two-way undivided highway, shows a significant tendency to speed and involve in fatalities. This consequence also has a strong association with unrestrained driving behavior. The lift value of 1.364 suggests that this combination of factors has 1.364 times higher likelihood of resulting in speeding related fatalities compared to the other combination of factors within cluster 2. Unrestrained driving behavior although speeding on two-way undivided highways might increase the risk of severe injuries or even ejection from the vehicle during the crash. Another rule [*Area_Type = rural + Unrestraint_Driving = yes + Vehicle_Type = passenger_car* → *Speeding_Type = exceeded_speed_limit*] includes rural areas as an

Table 4. Association rules from cluster 2.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Gender = male + Number_of_Lanes = two + Trafficway = 2WU + Unrestraint_Driving = yes + Vehicle_Type = passenger_car	Speeding_Type = exceeded_speed_limit	0.106	0.581	1.364
2	Gender = male + Number_of_Lanes = two + Unrestraint_Driving = yes + Vehicle_Type = passenger_car	Speeding_Type = exceeded_speed_limit	0.108	0.581	1.363
3	Gender = male + Unrestraint_Driving = yes + Vehicle_Type = passenger_car	Speeding_Type = exceeded_speed_limit	0.111	0.580	1.361
4	Unrestraint_Driving = yes + Vehicle_Type = passenger_car	Driver_Age = 20-29 y	0.100	0.411	1.353
5	Number_of_Lanes = two + Vehicle_Type = motored_cycle	Passenger_Presence = no	0.101	0.868	1.309
6	Trafficway = 2WU + Vehicle_Type = motored_cycle	Passenger_Presence = no	0.101	0.864	1.303
7	Vehicle_Type = motored_cycle	Passenger_Presence = no	0.103	0.864	1.303
8	Day_of_the_Week = Fri-Sun + Driver_Age = 20-29 y + Number_of_Lanes = two	Lighting_Condition = dark-unlighted	0.100	0.579	1.295
9	Area_Type = rural + Unrestraint_Driving = yes + Vehicle_Type = passenger_car	Speeding_Type = exceeded_speed_limit	0.105	0.550	1.290
10	Day_of_the_Week = Fri-Sun + Driver_Age = 20-29 y	Lighting_Condition = dark-unlighted	0.102	0.576	1.287

antecedent of similar types of fatalities. The support value of 0.105 in this rule indicates that 10.5% of all crashes in the dataset exhibit this combination of factors. The confidence value of 0.550 reveals that 55% of the crashes involving these antecedents (rural area, unrestrained driving, passenger car), demonstrate a high likelihood of speeding related fatalities. The lift value of 1.290 shows that the occurrence of these factors in such crashes is 1.29 times more likely than the combination of all other factors withing this cluster. Excessive speeding is a leading cause of road traffic crashes and, combined with unrestrained driving, it can significantly contribute to fatal outcomes (Bogstrand et al., 2015).

The second pattern is found in crashes with speeding motorcycle drivers (with no passengers) at fault on two-lane, two-way roads. Similar to the findings in the rules from cluster 1, the rules in cluster 2 indicate that solo motorcyclists are significantly more likely to exceed speed limits, suggesting that the absence of passengers might lead to riskier driving behavior. Solo riders were found to experience higher fatalities compared to those riding with passengers (Jou et al., 2012). Another pattern can be found with 20–29-year-old drivers speeding in dark-unlighted conditions between Friday and Sunday. This demographic in the rules shows a high tendency for engaging in risky driving behaviors during late hours, contributing to a higher likelihood of crashes. The lift values for these rules indicate strong associations between these specific factors and the occurrence of speeding-related crashes.

4.5. Cluster 3 [speeding in urban area in dark-lighted condition]

The top 10 rules in Cluster 3 (presented in Table 5) reveals significant patterns of crashes occurring in urban areas under dark-lighted conditions, particularly involving male drivers, passenger cars, drivers aged 20–29 years, absence of passengers, and incidents occurring from Friday to Sunday. The rules highlight specific scenarios associated with negotiating a curve. One prominent pattern involves crashes where male drivers in urban areas are negotiating a curve under dark-lighted conditions The rule [$Area\_Type = urban + Gender = male + Horizontal\_Alignment = curve + Lighting\_Condition = dark\text{-}lighted \rightarrow Movement\_Prior\_Crash = negotiating\_a\_curve$] has support, confidence, and lift values of 0.123, 0.813, and 3.943 respectively. This rule indicates that 12.3% of crashes in the dataset involve male drivers negotiating a curve in urban areas under dark-lighted conditions. The high confidence value of 81.3% means that when these antecedents are present, there is an 81.3% chance that the crash involved negotiating a curve. The lift value of 3.943 shows that the percentage of all crashes containing this combination in cluster 3 is almost 4 times the percentage of

Table 5. Association rules from cluster 3.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Area_Type = urban + Gender = male + Horizontal_Alignment = curve + Lighting_Condition = dark-lighted	Movement_Prior_Crash = negotiating_a_curve	0.123	0.813	3.943
2	Gender = male + Horizontal_Alignment = curve + Lighting_Condition = dark-lighted	Movement_Prior_Crash = negotiating_a_curve	0.125	0.811	3.932
3	Area_Type = urban + Horizontal_Alignment = curve + Vehicle_Type = passenger_car	Movement_Prior_Crash = negotiating_a_curve	0.109	0.810	3.928
4	Area_Type = urban + Horizontal_Alignment = curve + Lighting_Condition = dark-lighted	Movement_Prior_Crash = negotiating_a_curve	0.140	0.808	3.920
5	Horizontal_Alignment = curve + Lighting_Condition = dark-lighted	Movement_Prior_Crash = negotiating_a_curve	0.143	0.807	3.914
6	Horizontal_Alignment = curve + Vehicle_Type = passenger_car	Movement_Prior_Crash = negotiating_a_curve	0.112	0.805	3.905
7	Driver_Age = 20-29 y + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.102	0.805	3.903
8	Area_Type = urban + Gender = male + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.173	0.778	3.773
9	Gender = male + Horizontal_Alignment = curve + Passenger_Presence = no	Movement_Prior_Crash = negotiating_a_curve	0.122	0.778	3.772
10	Day_of_the_Week = Fri-Sun + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.118	0.777	3.770

such crashes with other combination of factors in the overall Cluster 3. The limited visibility and potential for glare from streetlights or headlights can impair the ability of the driver to judge the trajectory of a curve accurately. Visual distractions, combined with the inherent difficulty of maneuvering curves, especially at higher speeds, can lead to fatalities (Wang et al., 2017).

Another rule was found that supports this pattern. The rule [$Gender = male + Horizontal\_Alignment = curve + Passenger\_Presence = no \rightarrow Movement\_Prior\_Crash = negotiating\_a\_curve$] involves male drivers negotiating curves without passengers. This scenario appears in 12.2% of the crashes, with a confidence value of 77.8%, indicating a strong likelihood that these crashes involve negotiating a curve. The rule has a lift value of 3.772, which highlights the significant association, making this scenario nearly four times more likely than other scenarios including a combination of other factors within this cluster. These findings suggest the need for focused safety measures, such as promoting safer driving habits among solo male drivers and improving road conditions in areas prone to such crashes.

4.6. Cluster 4 [single vehicle speeding crashes on interstates]

The top 10 rules of Cluster 4 are represented in Table 6. This cluster associates speeding on curved segments on interstate highways with a speed limit of 60 mph or higher. It also shows associations with crashes on ramps. Negotiating a curve could be further complicated with vertical grades. This cluster also reveals significant patterns of single vehicle speeding crashes on interstate highways, particularly on curved segments with a speed limit of 60 mph or higher.

One notable pattern involves speeding on curves with speed limits of 60 mph or higher. The rule [Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions → Movement_Prior_Crash = negotiating_a_curve] indicates that this combination is strongly associated with fatal crashes although negotiating a curve. The support value of 0.185 shows that 18.5% of the crashes in the dataset involve these conditions. The confidence value of 88.3% indicates a very high likelihood that crashes involving these antecedents are associated with negotiating a curve. The lift value of 2.642 shows that the percentage of all crashes containing this combination in cluster 4 is 2.6 times the percentage of such crashes in overall Cluster 4. This means that high-speed on curves are particularly dangerous, and drivers are significantly more likely to crash when they speed on these segments (Wang et al., 2017). It can be also noticed on several functional classes of roadway that allow higher speed limit. Another rule, with

Table 6. Association rules from cluster 4.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Trafficway = ramp	Number_of_Lanes = one	0.101	0.625	5.695
2	Horizontal_Alignment = curve + Speed_Limit = 60 mph_or_higher + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.119	0.886	2.652
3	Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.185	0.883	2.642
4	Area_Type = rural + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.108	0.862	2.580
5	Horizontal_Alignment = curve + Number_of_Lanes = two + Speed_Limit = 60 mph_or_higher	Movement_Prior_Crash = negotiating_a_curve	0.112	0.858	2.567
6	Horizontal_Alignment = curve + Trafficway = ramp	Movement_Prior_Crash = negotiating_a_curve	0.111	0.857	2.565
7	Functional_System = interstate_freeway_expressway + Gender = male + Horizontal_Alignment = curve + Number_of_Lanes = two	Movement_Prior_Crash = negotiating_a_curve	0.112	0.849	2.541
8	Horizontal_Alignment = curve + Number_of_Lanes = two + Passenger_Presence = no	Movement_Prior_Crash = negotiating_a_curve	0.121	0.849	2.541
9	Horizontal_Alignment = curve + Vertical_Alignment = grade	Movement_Prior_Crash = negotiating_a_curve	0.151	0.848	2.539
10	Functional_System = interstate_freeway_expressway + Horizontal_Alignment = curve + Number_of_Lanes = two	Movement_Prior_Crash = negotiating_a_curve	0.137	0.848	2.538

antecedents "Functional_System = interstate, freeway, expressway", "Horizontal_Alignment = curve", "Gender = male", and " Number_of_Lanes = two", indicates a strong association with fatalities although negotiating a curve. The lift value of 2.541 highlights that the likelihood of occurrence of fatalities with this combination of factors is 2.5 times higher than the percentage of such crashes including other factors within this cluster. This increased the need for enhanced safety measures, such as improved signage and stricter speed enforcement on high-speed curves.

Crashes on ramps also reveal significant patterns. The rule "Trafficway = ramp" leading to "Number_of_Lanes = one" has a support value of 0.101, meaning 10.1% of crashes involve these conditions. The confidence value of 62.5% indicates that there is a 62.5% chance these crashes involve negotiating a curve. The lift value of 5.695 shows a very strong association, meaning crashes on ramps are almost 5.7 times more likely to involve negotiating a curve than random chance would suggest. This means that ramps, particularly those with one lane, are highly risky areas where drivers need to take extreme caution. Similarly, another rule suggests that drivers negotiating a ramp with curvature, also are highly likely to end up in fatal crashes with a lift value of 2.56.

4.7. Cluster 5 [speeding crashes at urban intersections]

Table 7 includes the top 10 association rules from Cluster 5. The top 10 high-lift rules in this cluster show an association with urban 4 W intersections controlled by traffic signals on principal arterials with a 40–45 mph speed limit. This pattern could also be associated with dark-lighted conditions and passenger cars. One notable pattern involves intersections in urban areas with principal arterials and traffic signals. The rule [$Area\_Type = urban + Functional\_System = principal\_arterial + Intersection\_Type = 4\,W \rightarrow Traffic\_Control = traffic\_signal$] has a support value of 0.167, meaning 16.7% of crashes in the dataset occur under these conditions. The confidence value of 78.8% indicates that when these conditions are present, there is a high likelihood of crashes. The lift value of 2.055 signifies that this scenario is just over twice as likely to occur compared to the other combination of factors in this cluster. Urban intersections with principal arterials and traffic signals are significant risk points for crashes primarily due to high traffic volumes and complex traffic patterns. Principal arterials are designed to handle large amounts of traffic, which can lead to congestion. This congestion increases the likelihood of crashes as drivers have to navigate through dense traffic, often with limited space and time to react to changes. Additionally, red-light violation under this scenario could result in more severe injuries that might cause fatalities (Datta et al., 2000).

Table 7. Association rules from cluster 5.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Area_Type = urban + Functional_System = principal_arterial + Intersection_Type = 4 W	Traffic_Control = traffic_signal	0.167	0.788	2.055
2	Area_Type = urban + Intersection_Type ntersectTrafficway = 2WD_unprotected_median	Traffic_Control = traffic_signal	0.108	0.783	2.043
3	Intersection_Type = 4 W + Trafficway = 2WD_unprotected_median	Traffic_Control = traffic_signal	0.116	0.778	2.029
4	Functional_System = principal_arterial + Intersection_Type = 4 W	Traffic_Control = traffic_signal	0.180	0.768	2.003
5	Area_Type = urban + Intersection_Type = 4 W + Speed_Limit = 40-45_mph	Traffic_Control = traffic_signal	0.132	0.740	1.931
6	Intersection_Type = 4 W + Speed_Limit = 40-45_mph	Traffic_Control = traffic_signal	0.139	0.718	1.872
7	Area_Type = urban + Intersection_Type = 4 W + Lighting_Condition = dark-lighted	Traffic_Control = traffic_signal	0.130	0.715	1.866
8	Intersection_Type = 4 W + Lighting_Condition = dark-lighted	Traffic_Control = traffic_signal	0.137	0.712	1.856
9	Area_Type = urban + Intersection_Type = 4 W + Vehicle_Type = passenger_car	Traffic_Control = traffic_signal	0.138	0.701	1.828
10	Intersection_Type = 4 W + Vehicle_Type = passenger_car	Traffic_Control = traffic_signal	0.146	0.679	1.770

Another significant pattern involves four-way intersections with unprotected medians. The rule $[Area\_Type = urban + Intersection\_Type = 4\,W + Trafficway = 2WD\_unprotected\_median \rightarrow Traffic\_Control = traffic\_signal]$ highlights the increased risk at urban four-way intersections with unprotected medians, suggesting a need for enhanced traffic control measures. The absence of medians allows for more complex and conflicting traffic movements, increasing the potential for driver errors and misjudgments (Kassu & Hasan, 2020).

The rule $[Intersection\_Type = 4\,W + Lighting\_Condition = dark\text{-}lighted \rightarrow Traffic\_Control = traffic\_signal]$ involves dark-lighted conditions at four-way intersections. It has a support value of 0.137, meaning 13.7% of crashes in the dataset occurred under these conditions. The confidence value of 71.2% indicates a high likelihood of crashes with this combination of factors. The lift value of 1.856 shows that the percentage of fatal crashes containing this combination in Cluster 5 is 1.856 times the percentage of such crashes in the overall Cluster 5. Reduced visibility during this condition makes it difficult for drivers to navigate and see the non-motorists on the road. This can lead to slower reaction times and misjudgments, increasing the likelihood of collisions.

4.8. Cluster 6 [multivehicle crashes in rural area]

The pattern in this cluster is associated with curve negotiation on rural, two-lane undivided roads, as per the top ten rules presented in Table 8. There is no passenger presence, and the speeding behavior is categorized as "too fast for conditions" in this scenario. One of the top rules $[Horizontal\_Alignment = curve + Speeding\_Type = too\_fast\_for\_conditions + Vehicle\_Type = passenger\_car \rightarrow Movement\_Prior\_Crash = negotiating\_a\_curve]$ indicates that speeding on curves significantly increases the likelihood of crashes, particularly for passenger cars. This rule has a support value of 0.105, meaning 10.5% of crashes in the dataset occur under these conditions. The confidence value of 93.4% indicates that when these conditions are present, there is a 93.4% chance that the crash involved negotiating a curve. The lift value of 2.631 signifies that fatal crashes are 2.631 times more likely to occur than by any other combination of factors within this cluster. In most cases, the crashes occurred when the vehicle was being operated at a higher speed on curves. Due to less pedestrian/non-motorist activities in rural areas, the drivers could drive with less care and at a higher speed, which have a high likelihood of resulting in fatality although negotiating a curve (Wang et al., 2017). The absence of passengers in vehicle can also increase the fatalities as the driver may involve in risky driving behavior (Orsi et al., 2013). This scenario is also supported by

Table 8. Association rules from cluster 6.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions + Vehicle_Type = passenger_car	Movement_Prior_Crash = negotiating_a_curve	0.105	0.934	2.631
2	Horizontal_Alignment = curve + Number_of_Lanes = two + Passenger_Presence = no + Speeding_Type = too_fast_for_conditions + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.106	0.923	2.602
3	Horizontal_Alignment = curve + Number_of_Lanes = two + Passenger_Presence = no + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.114	0.923	2.602
4	Horizontal_Alignment = curve + Number_of_Lanes = two + Speeding_Type = too_fast_for_conditions + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.153	0.923	2.601
5	Horizontal_Alignment = curve + Passenger_Presence = no + Speeding_Type = too_fast_for_conditions + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.118	0.923	2.601
6	Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.172	0.922	2.599
7	Horizontal_Alignment = curve + Passenger_Presence = no + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.134	0.921	2.597
8	Horizontal_Alignment = curve + Number_of_Lanes = two + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.165	0.921	2.596
9	Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.193	0.919	2.591
10	Collision_Manner = angle + Horizontal_Alignment = curve + Passenger_Presence = no	Movement_Prior_Crash = negotiating_a_curve	0.103	0.908	2.560

another rule [*Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions* → *Movement_Prior_Crash = negotiating_a_curve*] with a lift value of 2.591.

4.9. Cluster 7 [speeding crashes with front-to-rear collision on interstates]

The top 10 rules from Cluster 7 are represented in Table 9. This cluster reveals patterns of speeding-related crashes, specifically involving front-to-rear collisions on rural interstate highways with two-way divided unprotected medians. The speeding behavior is categorized as "too fast for conditions." These findings highlight critical scenarios that contribute to the risk of crashes on rural interstates. The rule [*Collision_Manner = front_to_rear + Speed_Limit = 60 mph or higher + Trafficway = 2WD_unprotected_median* → *Area_Type = rural*] highlights the increased risk of high-speed front-to-rear collisions on rural roads with unprotected medians. High speeds reduce reaction times, making it more difficult for drivers to stop in time if the vehicle in front slows down or stops unexpectedly (Arbabzadeh et al., 2019). On rural roads, where traffic flow can be unpredictable, and drivers might not anticipate sudden stops. Additionally, unprotected medians do not provide a physical barrier to prevent crossover incidents or help in mitigating the severity of crashes, allowing for more severe impacts (Chen & Chen, 2011; Kassu & Hasan, 2020). The rule shows a support value of 0.105, meaning 10.5% of crashes occur under these conditions. The confidence value of 70.0% indicates a high likelihood that these crashes occur in rural areas when the antecedents are present. The lift value of 1.966 suggests that this combination of factors is responsible for almost 2 times more fatal crashes compared to the other combination of factors in this cluster. This scenario in fatal crashes is also supported by the factors "Number_of_Lanes = two" and "Gender = male" in the subsequent rules.

4.10. Cluster 8 [speeding with unknown violation record]

The top 10 rules from Cluster 8 are highlighted in Table 10. This cluster shows associations with single-vehicle crashes, negotiating a curve, two-way undivided highways, and the speeding type of "too fast for conditions." These crashes occur from Friday to Sunday. The rule [*Collision_Manner = single-vehicle + Horizontal_Alignment = curve + Trafficway = 2WU* → *Movement_Prior_Crash = negotiating_a_curve*] has a support value of 0.116, meaning 11.6% of crashes in the dataset occur with this combination of factors. The confidence value of 0.915 indicates that when these factors are present, there is a 91.5% chance that the crash involved

Table 9. Association rules from cluster 7.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Area_Type = rural + Number_of_Lanes = two + Passenger_Presence = no	Trafficway = 2WD_unprotected_median	0.108	0.600	2.109
2	Area_Type = rural + Number_of_Lanes = two	Trafficway = 2WD_unprotected_median	0.148	0.590	2.075
3	Number_of_Lanes = two + Speed_Limit = 60 mph_or_higher + Speeding_Type = too_fast_for_conditions	Trafficway = 2WD_unprotected_median	0.107	0.564	1.984
4	Collision_Manner = front_to_rear + Speed_Limit = 60 mph_or_higher + Trafficway = 2WD_unprotected_median	Area_Type = rural	0.105	0.700	1.966
5	Collision_Manner = front_to_rear + Gender = male + Number_of_Lanes = two + Speed_Limit = 60 mph_or_higher	Area_Type = rural	0.136	0.689	1.935
6	Collision_Manner = front_to_rear + Number_of_Lanes = two + Speed_Limit = 60 mph_or_higher	Area_Type = rural	0.163	0.688	1.932
7	Number_of_Lanes = two + Speeding_Type = too_fast_for_conditions	Trafficway = 2WD_unprotected_median	0.128	0.545	1.917
8	Number_of_Lanes = two + Passenger_Presence = no + Speed_Limit = 60 mph_or_higher	Trafficway = 2WD_unprotected_median	0.135	0.545	1.916
9	Number_of_Lanes = two + Speed_Limit = 60 mph_or_higher	Trafficway = 2WD_unprotected_median	0.184	0.540	1.900
10	Gender = male + Speed_Limit = 60 mph_or_higher + Trafficway = 2WD_unprotected_median	Area_Type = rural	0.125	0.676	1.898

Table 10. Association rules from cluster 8.

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Collision_Manner = single-vehicle + Horizontal_Alignment = curve + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.116	0.915	4.780
2	Gender = male + Horizontal_Alignment = curve + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.107	0.883	4.615
3	Collision_Manner = single-vehicle + Gender = male + Horizontal_Alignment = curve + Number_of_Lanes = two	Movement_Prior_Crash = negotiating_a_curve	0.107	0.883	4.615
4	Collision_Manner = single-vehicle + Day_of_the_Week = Fri-Sun + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.104	0.881	4.604
5	Collision_Manner = single-vehicle + Horizontal_Alignment = curve + Number_of_Lanes = two	Movement_Prior_Crash = negotiating_a_curve	0.136	0.880	4.601
6	Collision_Manner = single-vehicle + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.165	0.879	4.593
7	Horizontal_Alignment = curve + Speeding_Type = too_fast_for_conditions	Movement_Prior_Crash = negotiating_a_curve	0.101	0.877	4.586
8	Horizontal_Alignment = curve + Trafficway = 2WU	Movement_Prior_Crash = negotiating_a_curve	0.136	0.874	4.569
9	Day_of_the_Week = Fri-Sun + Horizontal_Alignment = curve	Movement_Prior_Crash = negotiating_a_curve	0.120	0.846	4.423
10	Movement_Prior_Crash = negotiating_a_curve	Horizontal_Alignment = curve	0.191	1.000	4.402

negotiating a curve. The lift value of 4.780 signifies that crashes with this combination of factors are 4.7 times more likely to occur compared to the other combination of factors within this cluster. On two-way undivided highways, there are additional risks due to the lack of physical barriers separating opposing traffic (Chen & Chen, 2011; Kassu & Hasan, 2020). Additionally, the drivers could have the tendency to violate lane division rules and try to overtake the lead vehicle, which can result in head-on collisions with vehicles coming from the opposite direction. Furthermore, the likelihood of run-off crash fatality along with head-on collisions could also be higher (Al-Bdairi & Behnood, 2021).

The support values in most of the rules indicate how common these conditions are in the dataset, although the high confidence and lift values highlight the strong associations and increased likelihood of these scenarios leading to crashes. These findings highlight high-risk areas of crashes during weekend such as single-vehicle crashes on curves, two-way undivided highways, and speeding.

4.11. Summary findings and comparison of clusters

The most prominent and consistent finding across the eight clusters is the high prevalence of single-vehicle crashes, particularly on two-lane undivided roads and expressways. A comparison of association rules across Clusters 1 through 8 reveals several important factors contributing to fatal speeding crashes as the primary contributing factors—single-vehicle crashes on undivided roads, nighttime driving under dark-unlighted conditions, and motorcycle involvement on curved roads.

Clusters 1, 2, and 4, for instance, reveal that a significant proportion of fatal speeding crashes occur when drivers—often male—lose control of their vehicles on such road types. In Cluster 1, the data indicate that motorcyclists speeding on curved, rural roads are particularly vulnerable, although in Cluster 2, the crashes involve passenger cars on straight, undivided rural roads. Cluster 4 shows a similar pattern for expressways and high-speed interstates. Another condition observed across clusters is the role of nighttime driving with no lighting in fatal crashes.

Clusters 1 and 2, for example, show that crashes frequently occur in dark, unlighted environments, particularly in rural areas where infrastructure may be inadequate. This pattern is particularly evident in Cluster 1, where motorcyclists are involved in fatal crashes under dark-unlighted conditions, and in Cluster 2, where crashes occur in similar dark and unlighted environments. Motorcycle involvement in fatal crashes is another recurring observation among the clusters, especially in Clusters 1 and 2. Cluster 1 reveals that motorcyclists speeding on curves are at high risk of fatal

Table 11. Summary findings with interpretations.

Cluster	Key factors	Key associations and interpretation from rules
Cluster 1	Single-vehicle crashes on two-lane, two-way undivided roads, curves, no traffic control, drivers with full licenses, no previous violations	Predominantly male motorcyclists are at fault on weekends, often speeding "too fast for conditions." This is consistent with studies showing diminished braking effectiveness and increased risk on curves (Xin et al., 2017).
Cluster 2	Single-vehicle crashes mostly on straight two-lane, two-way undivided roads, drivers with full licenses, no prior violations, no traffic control	Key factors include exceeding speed limits, unrestrained driving on two-lane roads, and rural settings, highlighting the dangers of excessive speed and unrestrained driving (Bogstrand et al., 2015). Solo motorcycle riders on these roads are also prone to speeding, emphasizing riskier behavior without passengers (Jou et al., 2012).
Cluster 3	Single vehicle crashes primarily in urban areas on two-lane roads, drivers going straight, no prior violations, typically at non-intersections in dark-lighted condition	Male drivers aged 20–29, often in passenger cars, are involved in crashes although negotiating curves under poorly lit urban conditions. The lack of visibility and difficulty in maneuvering curves are significant risk factors (Wang et al., 2017).
Cluster 4	Single vehicle crashes on interstates, freeways, and expressways, drivers with no other violations, typically not at intersections	Speeding on curved interstate segments with higher speed limits is particularly dangerous, underscoring the need for enhanced signage and speed enforcement on such roads. Ramps are highlighted as high-risk areas, needing caution and better design to prevent crashes (Wang et al., 2017).
Cluster 5	Multivehicle crashes in urban environments on two-lane roads at known intersections, significant proportion of male drivers with no previous violations	High traffic volumes at urban intersections with traffic signals and principal arterials increase crash risks, particularly under dark-lighted conditions. The complex traffic patterns and high congestion are major concerns (Datta et al., 2000).
Cluster 6	Multi-vehicle crashes on two-lane, two-way undivided roads mostly in daylight conditions in rural areas	High speeds on rural, two-lane undivided roads without passenger presence are associated with higher fatalities, especially when negotiating curves. The limited activity in rural areas may encourage faster driving and increased risk-taking (Wang et al., 2017).
Cluster 7	Occurring on interstates, freeways, and expressways at high speeds, multivehicle crashes mainly involving front-to-rear collisions, drivers not unrestrained	High-speed collisions on rural interstates, particularly front-to-rear impacts with unprotected medians, are significant. These scenarios highlight the dangers of inadequate reaction times and the lack of physical barriers (Arbabzadeh et al., 2019; Kassu & Hasan, 2020).
Cluster 8	Drivers with unspecified previous violations, frequently not unrestrained, often involved in crashes at non-intersections	Single-vehicle crashes although negotiating curves on two-way undivided highways are prevalent, with additional risks on weekends. The potential for lane violations and head-on collisions is notably high (Chen & Chen, 2011; Kassu & Hasan, 2020).

crashes, particularly on weekends and under poor visibility conditions. In rural areas, as seen in Cluster 2, motorcycle crashes are similarly prominent.

Despite the above similarities, a few differences were also identified among the clusters. For instance, the type of roadway plays an important role in determining crash characteristics. Cluster 7 shows speeding crashes on high-speed expressways, where front-to-rear collisions are more common, in contrast to the single-vehicle crashes on rural roads in Clusters 1, 2 and 4. Driver demographics also vary among the clusters. In Cluster 1, young male motorcyclists are disproportionately involved in fatal speeding crashes, although in Cluster 7, male drivers in general are frequently associated with fatal crashes on high-speed roads. Considering the length of the discussion section with 8 clusters, Table 11 highlights the key findings and associative interpretations.

5. Conclusions

Understanding the patterns of speeding crashes and associated attributes is not only important from the perspective of speed limit enforcement but also in terms of speed-related crash reduction. This research attempt has been made to account for the heterogeneity of the crash data using an LCC approach. Subsequently, ARM has been applied to each of the identified latent clusters to further identify the collective presence of crash-contributing factors responsible for fatal crash injuries. The current study provides a fresh overview of speeding-related crashes by finding eight (8) homogenous clusters that were distributed in a heterogenous crash database for speeding.

The clustered segmentation and association rules provide additional insight into crashes that may not be discovered from statistical analysis. In addition to specifying a high-lift value, applying a threshold of supply and confidence measures often reveals crash patterns that may not always be highly frequent individually. However, they collectively could pose a high likelihood of crashes.

Countermeasures could encompass roadway improvement, traffic control measures, and educational and enforcement programs. Some of the key findings in this study are:

- Speeding in motorcycle crashes has been highlighted in a number of rules and therefore may require special attention. The association rules reveal that negotiating a curve although speeding a motorcycle on curves could be fatal, especially during dark-unlighted conditions and on weekends. Achieving safe motorcycle speed may require a multifaceted approach, and motorcycle safety programs addressing the issues through proper training prior to motorcycle licensure can be helpful. Such intervention programs have been proven to be successful (10). States in the United States have been undertaking such programs.

- Unrestrained driving issues continue to be an important factor in speeding-related fatal crashes. This trait has been found to be common among male passenger car drivers. Lowering the injury severity of speeding crashes largely depends on preventing unrestrained driving. Enhanced seat belt reminder systems in vehicles can increase seat belt usage and reduce fatalities (Kidd & O'Malley, 2023).
- Similar to unrestrained driving, crashes with highly intoxicated drivers (BAC $>$ 0.15 g/dL) can be prevented by educational campaigns and enforcement.
- In urban areas, traffic signals on 4-way intersections, especially on principal arterials, could be a fatal speeding crash concern. Peripheral transverse pavement marking, slow or speed limit pavement marking, and high-friction surface treatment have been proven to be beneficial countermeasures (11). Additionally, the installation of red-light cameras can effectively reduce red-light running and associated crashes at urban intersections (Jiang & Ouyang, 2017).
- Speed cameras can have safety benefits in terms of preventing high-speed collisions in urban areas (12). This can also be installed on roadways with straight/curved alignments as a countermeasure to control vehicles traveling with higher speed at specific road segments.
- Installation of rumble strips can assist in increasing driver attention and ensure lane assistance on roadways with straight or curved alignments. It can also help in reducing crashes. To mitigate fatalities on curves, countermeasures could also include installing curve warning signs and improving road markings.
- The use of rumble strips on unprotected medians can also reduce fatalities (Persaud et al., 2004). Additionally, median barriers can be installed to prevent vehicles from head-on collisions.
- Advanced driver assistance system (ADAS) in vehicles such as adaptive cruise control and collision avoidance systems can maintain safe following distances and reduce the likelihood of rear-end collisions (Tan et al., 2021).

By targeting a group of crash attributes from the rules generated in this study, strategic context-based countermeasures or prevention plans can be developed. The detailed findings and provided countermeasures in this study can be beneficial for developing strategies targeting speed-related crashes. Finally, the insights gained from this study emphasize the importance of a multifaceted approach to speed-related crash prevention, integrating infrastructure improvements, policy enforcement, and driver education. Future research must compare the factors with speeding crashes that occurred during and post-COVID era. Distinguishing between different injury severity levels is crucial for traffic engineers, policymakers, and planners as they devise the most effective interventions. These may include geometric modifications, traffic control enhancements, dedicated pedestrian zones, land use changes, educational initiatives, and enforcement measures.

Acknowledgments

The authors would like to thank the Louisiana DOTD for providing the data. The authors appreciate the comments of the discussants at the Transportation Research Board Annual Meeting, who provided interesting perspectives on the paper. This research did not receive any grants from funding agencies.

Authors contributions

The authors confirm their contribution to the paper as follows: study conception and design: M. Ashifur Rahman, Subasish Das; data collection: Md. Mahmud Hossain, and M. Ashifur Rahman; analysis and interpretation of results: M. Ashifur Rahman, Rohit Chakraborty, and Nurul-Haq Mohammed; draft manuscript preparation: M. Ashifur Rahman, Rohit Chakraborty, Nurul-Haq Mohammed, Subasish Das, Siam Junaed, and Md. Mahmud Hossain. All authors reviewed the results and approved the final version of the manuscript.

Disclosure statement

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

ORCID

M. Ashifur Rahman http://orcid.org/0000-0001-6940-1599
Md. Mahmud Hossain http://orcid.org/0000-0002-2737-6951

References

Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident, Analysis and Prevention*, *38*(2), 215–224. https://doi.org/10.1016/j.aap.2005.07.004

Abegaz, T., Berhane, Y., Worku, A., Assrat, A., & Assefa, A. (2014). Effects of excessive speeding and falling asleep while driving on crash injury severity in Ethiopia: A generalized ordered logit model analysis. *Accident, Analysis and Prevention*. 71, 15–21. https://doi.org/10.1016/j.aap.2014.05.003

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In 20th International Conference on Very Large Data Bases, VLDB (vol. 1215, pp. 487–499). https://doi.org/10.1345/aph.1K157

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, *22*, 207–216. https://doi.org/10.1145/170035.170072

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Al-Bdairi, N. S. S., & Behnood, A. (2021). Assessment of temporal stability in risk factors of crashes at horizontal curves on rural two-lane undivided highways. *Journal of Safety Research*, *76*, 205–217. https://doi.org/10.1016/j.jsr.2020.12.003

Arbabzadeh, N., Jafari, M., Jalayer, M., Jiang, S., & Kharbeche, M. (2019). A hybrid approach for identifying factors affecting driver reaction time using naturalistic driving data. *Transportation Research C*, *100*, 107–124. https://doi.org/10.1016/j.trc.2019.01.016

Bogstrand, S. T., Larsson, M., Holtan, A., Staff, T., Vindenes, V., & Gjerde, H. (2015). Associations between driving under the influence of alcohol or drugs, speeding and seatbelt use among fatally injured car drivers in Norway. *Accident; Analysis and Prevention*, *78*, 14–19. https://doi.org/10.1016/j.aap.2014.12.025

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370. https://doi.org/10.1007/BF02294361

Chen, F., & Chen, S. (2011). Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways. *Accident; Analysis and Prevention*, *43*(5), 1677–1688. https://doi.org/10.1016/j.aap.2011.03.026

Das, S., Dutta, A., Jalayer, M., Bibeka, A., & Wu, L. (2018). Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. *International Journal of Transportation Science and Technology*, *7*(2), 114–123. https://doi.org/10.1016/j.ijtst.2018.02.001

Das, S., Le, M., Fitzpatrick, K., & Wu, D. (2022). Did operating speeds during COVID-19 result in more fatal and injury crashes on urban freeways? *Transportation Research Record*, 036119812211095. https://doi.org/10.1177/03611981221109597

Das, S., Tamakloe, R., Zubaidi, H., Obaid, I., & Alnedawi, A. (2021). Fatal pedestrian crashes at intersections: Trend mining using association rules. *Accident; Analysis and Prevention*, *160*, 106306. https://doi.org/10.1016/j.aap.2021.106306

Datta, T. K., Schattler, K., & Datta, S. (2000). Red light violations and crashes at urban intersections. *Transportation Research Record*, *1734*(1), 52–58. https://doi.org/10.3141/1734-08

Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident; Analysis and Prevention*, *40*(4), 1257–1266. https://doi.org/10.1016/j.aap.2008.01.007

Dutta, N., & Fontaine, M. D. (2019). Improving Freeway segment crash prediction models by including disaggregate speed data from different sources. *Accident; Analysis and Prevention*, *132*, 105253. https://doi.org/10.1016/j.aap.2019.07.029

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Elvik, R. (2014). "Speed and road safety-new models." TØI Report, no. 1296/2014. https://trid.trb.org/View/1318116

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis: Fifth Edition. Wiley series in probability and statistics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470977811

Fitzpatrick, C. D., Rakasi, S., & Knodler, M. A. (2017). An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression. *Accident Analysis & Prevention*. 98, 57–63. https://doi.org/10.1016/j.aap.2016.09.017

Garber, N. J., & Gadiraju, R. (1989). Factors affecting speed variance and its influence on accidents. *Transportation Research Record*, 1213, 64–71.

Gargoum, S. A., & El-Basyouny, K. (2016). Exploring the association between speed and safety: a path analysis approach. *Accident; Analysis and Prevention*, *93*, 32–40. https://doi.org/10.1016/j.aap.2016.04.029

Hahsler, M., Grün, B., & Hornik, K. (2005). arules - A Computational environment for mining association rules and frequent item sets. *Journal of statistical software*, *14*(15), 1–25. https://doi.org/10.18637/jss.v014.i15

Hauer, E. (2009). Speed and safety. *Transportation Research Record*, *2103*(1), 10–17. https://doi.org/10.3141/2103-02

Høye, A. (2020). Speeding and impaired driving in fatal crashes—Results from in-depth investigations. *Traffic Injury Prevention*, *21*(7), 425–430. https://doi.org/10.1080/15389588.2020.1775822

Hutton, J. M., Cook, D. J., Grotheer, J. (2020). *Research utilizing SHRP2 data to improve highway safety: Development of speed—Safety relationships*. Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/44448

Jiang, Z., Ouyang, Y. (2017). Spillover effect and economic effect of red light cameras. https://apps.ict.illinois.edu/projects/getfile.asp?id=7427

Job, S., & Brodie, C. (2022). Understanding the role of speeding and speed in serious crash trauma: A case study of New Zealand. *Journal of Road Safety*, *33*(1), 5–25. https://doi.org/10.33492/JRS-D-21-00069

Jou, R.-C., Yeh, T.-H., & Chen, R.-S. (2012). Risk factors in motorcyclist fatalities in Taiwan. *Traffic Injury Prevention*, *13*(2), 155–162. https://doi.org/10.1080/15389588.2011.641166

Kassu, A., & Hasan, M. (2020). Factors associated with traffic crashes on urban freeways. *Transportation Engineering*, *2*, 100014. https://doi.org/10.1016/j.treng.2020.100014

Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data. *In* L. Kaufman & P. J. Rousseeuw. *Finding groups in data: An introduction to cluster analysis: Wiley series in probability and statistics* (Vol. 180). John Wiley & Sons, Inc.

Kidd, D. G., & O'Malley, S. (2023). Increasing seat belt use in the united states by promoting and requiring more effective seat belt reminder systems. *Traffic Injury Prevention*, *24*(suppl 1), S80–S87. https://doi.org/10.1080/15389588.2022.2134730

Lee, C., Saccomanno, F., & Hellinga, B. (2002). Analysis of crash precursors on instrumented freeways. *Transportation Research Record*, *1784*(1), 1–8. https://doi.org/10.3141/1784-01

Maimon, O., & Rokach, L. (2010). Introduction to knowledge discovery in databases. In *Data mining and knowledge discovery handbook*. Springer.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, *11*, 1–16. https://doi.org/10.1016/j.amar.2016.04.001

Orsi, C., Marchetti, P., Montomoli, C., & Morandi, A. (2013). Car crashes: The effect of passenger presence and other factors on driver outcome. *Safety Science*, *57*, 35–43. https://doi.org/10.1016/j.ssci.2013.01.017

Park, E. S., Fitzpatrick, K., Das, S., & Avelar, R. (2021). Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis. *Accident; Analysis and Prevention*, *150*, 105896. https://doi.org/10.1016/j.aap.2020.105896

Pei, X., Wong, S. C., & Sze, N.-N. (2012). The roles of exposure and speed in road safety analysis. *Accident; Analysis and Prevention*, *48*, 464–471. https://doi.org/10.1016/j.aap.2012.03.005

Persaud, B. N., Retting, R. A., & Lyon, C. A. (2004). Crash reduction following installation of centerline rumble strips on rural two-lane roads. *Accident; Analysis and Prevention*, *36*(6), 1073–1079. https://doi.org/10.1016/j.aap.2004.03.002

Quddus, M. (2013). Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. *Journal of Transportation Safety & Security*, *5*(1), 27–45. https://doi.org/10.1080/19439962.2012.705232

R Development Core Team. (2024). *R: A language and environment for statistical computing. no. 3.6.1*.

Rahman, M. A., Das, S., & Sun, X. (2023). Single-vehicle run-off road crashes because of cellphone distraction: Finding patterns with rule mining. *Transportation Research Record*, *2677*(3), 1261–1277. https://doi.org/10.1177/03611981221122781

Rahman, M. A., Sun, X., Das, S., & Khanal, S. (2021). Exploring the influential factors of roadway departure crashes on rural two-lane highways with logit model and association rules mining. *International Journal of Transportation Science and Technology*, *10*(2), 167–183. https://doi.org/10.1016/j.ijtst.2020.12.003

Rahman, M. A., Sun, X., Sun, M., & Shan, D. (2021). Investigating characteristics of cellphone distraction with significance tests and association rule mining. *IATSS Research*, *45*(2), 198–209. https://doi.org/10.1016/j.iatssr.2020.09.001

Rowden, P., Watson, B., Haworth, N., Lennon, A., Shaw, L., & Blackman, R. (2016). Motorcycle riders' self-reported aggression when riding compared with car driving. *Transportation Research F*, *36*, 92–103. https://doi.org/10.1016/j.trf.2015.11.006

Sasidharan, L., Wu, K. F., & Menendez, M. (2015). Exploring the application of latent class cluster analysis for investigating pedestrian crash injury severities in Switzerland. *Accident; Analysis and Prevention*, *85*, 219–228. https://doi.org/10.1016/j.aap.2015.09.020

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333–343. https://doi.org/10.1007/BF02294360

Se, C., Champahom, T., Jomnonkwao, S., & Ratanavaraha, V. (2024). Examining factors affecting driver injury severity in speeding-related crashes: A comparative study across driver age groups. *International Journal of Injury Control and Safety Promotion*, *31*(2), 234–255. https://doi.org/10.1080/17457300.2023.2300458

Shaheed, M. S., & Gkritza, K. (2014). A latent class analysis of single-vehicle motorcycle crash severity outcomes. *Analytic Methods in Accident Research*, *2*, 30–38. https://doi.org/10.1016/j.amar.2014.03.002

Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident; Analysis and Prevention*, *124*, 50–57. https://doi.org/10.1016/j.aap.2018.12.016

Tan, H., Zhao, F., & Liu, Z. (2021). Impact of adaptive cruise control (ACC) system on fatality and injury reduction in China. *Traffic Injury Prevention*, *22*(4), 307–312. https://doi.org/10.1080/15389588.2021.1896715

USDOT (2022). *National roadway safety strategy*. United States Department of Transportation.

Wang, B., Hallmark, S., Savolainen, P., & Dong, J. (2017). Crashes and near-crashes on horizontal curves along rural two-lane highways: analysis of naturalistic driving data. *Journal of Safety Research*, *63*, 163–169. https://doi.org/10.1016/j.jsr.2017.10.001

Wang, X., Zhou, Q., Quddus, M., & Fan, T. (2018). Speed, speed variation and crash relationships for urban arterials. *Accident Analysis & Prevention*, 113, 236–243. https://doi.org/10.1016/j.aap.2018.01.032

Watson, B., Watson, A., Siskind, V., Fleiter, J., & Soole, D. (2015). Profiling high-range speeding offenders: Investigating criminal history, personal characteristics, traffic offences, and crash history. *Accident; Analysis and Prevention*, *74*, 87–96. https://doi.org/10.1016/j.aap.2014.10.013

Xin, C., Wang, Z., Lin, P.-S., Lee, C., & Guo, R. (2017). Safety effects of horizontal curve design on motorcycle crash frequency on rural, two-lane, undivided highways in Florida. *Transportation Research Record*, *2637*(1), 1–8. https://doi.org/10.3141/2637-01

Yu, R., Quddus, M., Wang, X., & Yang, K. (2018). Impact of data aggregation approaches on the relationships between operating speed and traffic safety. *Accident Analysis & Prevention*, *120*, 304–310. https://doi.org/10.1016/j.aap.2018.06.007

Yuan, R., Ding, S., Fang, Z., Gu, X., & Xiang, Q. (2023). Investigating the spatial heterogeneity of factors influencing speeding-related crash severities using correlated random parameter order models with heterogeneity-in-means. *Transportation Letters*, *16*, 1–11. https://doi.org/10.1080/19427867.2023.2262201