Check for updates

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–20
© The Author (s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/03611981241242751
journals.sagepub.com/home/trr
Sage

Exploring Attribute Associations in Pedestrian-Involved Hit-and-Run Crashes through Cluster Correspondence Analysis

M. Ashifur Rahman[1], Subasish Das[2], Ahmed Hossain[1], Julius Codjoe[1], Elisabeta Mitran[1], and Xiaodaun Sun[1]

Abstract

Pedestrian-involved hit-and-run (PIHR) crashes represent a significant public health concern, and identifying patterns in these crashes can aid in developing effective countermeasures. Cluster correspondence analysis (CCA) is a multidimensional statistical technique that combines dimension reduction and clustering to identify patterns in categorical data. This method provides insights into underlying patterns and relationships among categories. The current study analyzed a Louisiana crash dataset of 2,201 PIHR crashes from 2015 to 2019 using CCA to identify underlying patterns. CCA identified six clusters, examined the top associative attributes, and assessed their cluster-to-dataset percentage ratio. The first two clusters, representing 66% of PIHR crashes, mainly involved crashes on city streets, occurring primarily during early night (7 to 11 p.m.) in Cluster 1 and the afternoon (12 noon to 4 p.m.) in Cluster 2. Clusters 3 and 4, accounting for 30% of PIHR crashes, predominantly exhibited crashes on U.S. and state highways. Cluster 4, which featured fatalities, primarily concentrated on state highways during the early morning hours (4 to 6 a.m.). Meanwhile, Clusters 5 and 6 focused on high-speed highways, specifically interstates involving pedestrian fatalities. A discussion on implementing strategic countermeasures tailored to the distinct characteristics of each cluster is presented. Alongside improvements in context-based countermeasures to ease pedestrian movement and enhance their visibility, strategies such as advocating for stringent hit-and-run laws, incentivizing the use of dashcams, and broadly publicizing resources for crash reporting are projected to be highly effective in curbing PIHR crashes.

Keywords

pedestrians, hit-and-run crashes, cluster correspondence analysis, correspondence analysis, pedestrian safety countermeasures

Pedestrians exhibit the highest susceptibility to hit-and-run incidents among all road users within the United States, as evidenced by nationwide comprehensive crash statistics. The fatal crash data obtained from the Fatality and Injury Reporting System Tool, or FIRST (*1*), strongly indicates the vulnerability of pedestrians to hit-and-run crashes, with 20.45% of the 30,228 pedestrian fatalities between 2015 and 2019 in the United States being hit-and-run crashes. Among all the hit-and-run crashes during the same period (9,665), more than two-thirds of them, 6,500 (67%), involved at least one pedestrian. A total of 6,181 (64%) of all hit-and-run crashes caused at least one pedestrian fatality. At the state level, more than a quarter of pedestrian crashes, 26.73% (i.e., 2,201 of 8,233), are identified as hit-and-run crashes in the crash repository system of Louisiana during the same period (*2*). In hit-and-run crashes, pedestrians are more likely to be involved than other road users (*3–5*). In fatal hit-and-run crashes, pedestrians are 10 times more likely to be involved compared with other road users (*6*). In addition to the disproportionate involvement of pedestrians in hit-and-run crashes a pattern indicating high

[1]Department of Civil Engineering, University of Louisiana, Lafayette, LA
[2]Ingram School of Engineering, Texas State University, San Marcos, TX

Corresponding Author:
M. Ashifur Rahman, ashifur@louisiana.edu; ashifur@outlook.com

severity levels in vehicle–pedestrian crashes (*7*) calls for an exclusive investigation of the contributing factors of pedestrian-only hit-and-run crashes.

As pedestrian safety remains a paramount concern in the development and implementation of comprehensive safety policies, the U.S. government has recently adopted the "Safe System" approach. This holistic strategy focuses on enhancing pedestrian safety as a key component within its "Safer People" initiative, aiming to address the multifaceted challenges associated with pedestrian protection, and promoting a secure transportation environment for all road users (*8*). Enhancing pedestrian safety countermeasures is a top priority when incorporating infrastructure improvements (*9*). Consistent with the national policies, the latest Louisiana strategic highway safety plan (SHSP) also prioritizes pedestrian safety infrastructure under the area emphasizing "Infrastructure and Operations" (*10*).

Hit-and-run crash scenarios associated with pedestrians possess a complex mix of behavioral and environmental factors. Understanding pedestrian and driver characteristics other than roadway and environmental factors are important, as pedestrian–vehicle interactions have a crucial role in preventing pedestrian crashes (*11*). Driver conditions such as distractions have been found to be influential in hit-and-run crashes (*12*). Intoxication can also contribute substantially to pedestrian hit-and-run crashes as an instigating factor (*13*). Apart from roadway environment conditions (*5*, *6*, *14*), pedestrian characteristics associated with their activities during crashes have been identified as important factors (*14*).

The aim of this study was to investigate pedestrian-involved hit-and-run (PIHR) crashes using a wide range of pedestrian characteristics in addition to other available crash characteristics. The use of clustering approaches has gained popularity among transportation safety researchers, as these methods facilitate the discovery of nonrandom structures in the analysis of crash datasets. A dimensionality reduction accounts for the limitations of categorical crash datasets, such as multidimensionality, multicollinearity. This study applied cluster correspondence analysis (CCA), a joint dimension reduction and clustering method, to find hidden patterns from the clustered datasets. CCA is specifically designed to analyze contingency tables, which represent the frequencies of the cooccurrence of two or more categorical variables. CCA combines dimension reduction and clustering to identify groups of rows and columns that show similar patterns of association. It allows for the simultaneous exploration of both the structure and relationships among categories in the data.

This research paper synthesizes key relevant research studies on PIHR crashes, organizing findings according to crash factors and summarizing the essential takeaways from these PIHR crash investigations. The "Data" section provides a concise overview of the data preparation process, addressing data limitations, and outlining the distribution of attribute characteristics. The "Methodology" section delves into a comprehensive explanation of the chosen methodologies and analytical contexts, with a particular focus on cluster selection. The "Results" section provides visual representations and thorough descriptions of the attribute associations in the identified clusters. The "Discussion" section mainly describes countermeasure implications and research limitations. Finally, the "Conclusions" section highlights research contributions, key findings, and directions for future research.

Literature Review

In the realm of transportation safety research, investigations into hit-and-run crashes hold significant importance. The repercussions of PIHR crashes are frequently grave, potentially resulting in fatalities or leaving critically injured individuals without aid. The human elements embedded within these crash studies extend across various disciplines, garnering substantial attention from scholars over the years. Efforts have been devoted to pinpointing the factors and trends linked to such behavior. This section is divided into two segments: a concise examination of the primary determinants and patterns associated with PIHR incidents, as uncovered by existing research, and an analysis of the methodological approaches employed in these investigations.

Factors Associated with Hit-and-Run Crashes

The majority of hit-and-run studies have sought to identify contributing factors by comparing hit-and-run crashes with non-hit-and-run crashes and estimating the likelihood of each factor's involvement. The subsequent discussion of factors primarily highlights those factors found to have a higher likelihood of being involved in hit-and-run crashes, based on these comparative analyses.

Temporal and Environmental Factors. Benson et al. found that drivers involved in crashes on weekends and during evening/night hours were more likely to be involved in hit-and-run incidents (*15*). MacLeod et al. (*14*) and Zhou et al. (*16*) noted that early mornings were associated with an increased risk of hit-and-run incidents, and crashes on weekends had a higher likelihood of drivers leaving the scene. Several studies emphasized the role of lighting conditions in hit-and-run incidents (*13*, *15*, *17*, *18*). Aidoo et al.

reported that nighttime conditions significantly increased the likelihood of such incidents (*3*). MacLeod et al. determined that poor light conditions were associated with an increased risk of hit-and-runs (*14*), and Sivasankaran and Balasubramanian found that pedestrian crashes in dark, unlit conditions were more likely to involve drivers leaving the scene (*18*). Unclear weather significantly increased the chances of hit-and-run incidents (*3*, *16*).

Pedestrian Factors. Pedestrian behavior can greatly affect pedestrian crashes, as factors like alcohol intoxication and the sociodemographic characteristics of pedestrians can influence the likelihood of hit-and-run incidents (*18*).

Driver Factors. Several studies have identified that younger, male drivers, and those without a valid license, who had rented their vehicles, were under the influence of drugs or alcohol, or had moving violations (i.e., any traffic law violation that occurs when the vehicle is in motion) had a higher likelihood of engaging in hit-and-run incidents (*5*, *13–15*, *19*). Benson et al. also mentioned that drivers of vehicles registered to a business or government entity were less likely to be involved in hit-and-runs (*15*). Kim et al. found that being male, a tourist, intoxicated, and driving a stolen vehicle increased the likelihood of a hit-and-run (*20*). Roshandeh et al. discovered that nondistracted drivers were 27% less likely to leave the scene after a crash compared with distracted drivers (*12*). Drivers without automobile insurance and lacking extensive driving experience were also both associated with hit-and-runs (*5*). Another study conducted by Tay et al. reported that male, minority, or drivers aged between 45 and 69 years were more likely to flee the crash location (*4*). From the analysis of available driver data, MacLeod et al. identified that drivers are more likely to be identified in crashes involving children aged 15 years and younger or women (*14*).

Collision Type and Vehicle Type. Single-vehicle collisions involving pedestrians were the most prevalent type of hit-and-run incidents (*17*). Benson et al. found that the majority of hit-and-runs involved the driver hitting nonmotorists (*15*). Tay et al. showed that when the crash involved two vehicles, two-wheeled vehicles (specifically motorcycles and bicycles), and vehicles from neighboring countries, the driver was more likely to engage in hit-and-run behavior (*4*).

Roadway Factors. Das et al. found that hit-and-run incidents involving segment-related crashes exhibited higher fatality rates compared with those occurring at intersections (*17*). The likelihood of a driver engaging in a hit-and-run incident increased on straight and flat road sections without medians and junctions (*3*). Zhou et al. reported varying contributing factors to hit-and-run crashes for different improper driving behaviors, with "following too closely" and "distraction by phone" models showing the most significant variables. Their study further identified that specific factors (e.g., locations with traffic signals, yield signs, shoulders, darkness, and less than three lanes) associated with these improper driving behaviors increased the likelihood of such crashes (*16*). MacLeod et al. noted that crashes occurring in locations other than roads or crosswalks had a higher probability of identifying the driver involved in the hit-and-run (*14*). Tay et al. discovered that hit-and-run collisions were more likely to occur on straight roads, whereas they were less frequent on undivided roads involving right-turn and U-turn maneuvers (*4*). In another study, Tay et al. explored the causal factors of PIHR crashes, suggesting that elements such as a lack of witnesses, low noise levels during the collision, and the inability of severely injured pedestrians to identify the driver, all played a significant role in influencing drivers to flee the scene (*6*).

Methods Adopted in Hit-and-Run Crash Analysis

Transportation safety researchers have also been undertaking different analytical approaches to address specific issues associated with hit-and-run crashes. In recent years, both unsupervised and supervised data mining methods have been applied for identifiable patterns and interpretive associations of attributes outside of the conventional logistic regression approach. For example, Das et al. utilized an association rule mining approach to explore rules presenting the simultaneous presence of multiple factors in statewide data (*17*). Das et al. applied a fast and frugal heuristics approach with a focus on extracting the significant factors that influence the occurrences of bicycle-involved hit-and-run crashes (*21*). Jha et al. applied several supervised learning classification models to predict the at-fault vehicle type (*22*).

Various methods have been used in previous hit-and-run crash studies. Table 1 shows the methods adopted in multiple selected PIHR crash studies and highlights their key findings. In addition to descriptive analysis (*23*, *24*), the most common method applied in these studies is logistic regression (*3*, *4*, *6*, *14*). Using this approach, segregated logit models have been developed in the by using various combinations of variable groups, such as crash environment, temporal characteristics, and characteristics of the fatal pedestrian victim and the hit-and-run driver (*14*). Models have mostly been developed by excluding and including available driver variables as driver-related data can be limited. The PIHR studies in the United States used only fatal crash data from the Fatality Analysis Reporting System (*13*, *14*, *23*, *25*).

Table 1. Findings from Selected Relevant Crash Studies

Reference	Hit-and-run study	Crash-year; location; analysis type	Key findings related to pedestrians
(*13*)	S J Solnick and Hemenway (1994)	1989 to 1990; USA; logit model	• Time of crash, age of car, driver gender, validity of the current license and previous DWI conviction are each associated with alcohol-related fatal pedestrian hit-and-run crashes. • Specifically, controlling for other variables, the results show that likelihood of a hit-and-run at night compared with during the day.
(*25*)	Sara J. Solnick and Hemenway (1995)	1989 to 1991; USA; logit model	• Nighttime remains a significant predictor of hit-and-run incidents in pedestrian crashes. • Drivers with previous "driving while intoxicated" arrests are twice as likely to run compared with those with no such arrests.
(*3*)	Aidoo, Amoh-Gyimah, and Ackaah (2013)	1998 to 2007; Ghana; logit model	• The findings revealed that inclement weather, nighttime, flat roads without medians and intersections played critical roles in hit-and-run crashes involving pedestrians. • Pedestrian hit-and-run crashes were shown to be more common in the early morning, during non-daylight hours, and on weekends. • Certain driver demographics (young, male), behavior (alcohol use), and history (e.g., suspended license or history of DWI/DUI convictions) were also associated with pedestrian hit-and-run crashes.
(*4*)	Tay, Rifaat, and Chin (2008)	1992 to 2002; Singapore; logit model	• A hit-and-run crash is more likely to occur on a high-speed road, at a nonintersection in an urban setting where pedestrians are more vulnerable. • Collision severity significantly influenced hit-and-run likelihood, with fatal and serious crashes having 1.52 and 1.27 times higher odds of a hit-and-run, respectively, compared with minor crashes.
(*19*)	Grembek and Griswold (2012)	1998 to 2007; USA; descriptive	• Drivers involved in hit-and-runs are more likely to be under the age of 25 and male. • Drivers are more likely to flee the scene of a collision if they believe that fault can be more clearly attributed to them.
(*14*)	MacLeod et al. (2012)	1998 to 2007; USA; logit model	• Hit-and-run was more common in lower-speed locations. • Alcohol use and early morning, the time frame when persons may be leaving bars and parties, were among the leading factors that increased the risk of hit-and-run.
(*23*)	A.J. Benson et al. (2017)	2016 to 2016; FARS; descriptive	• Drivers who have had their licenses suspended or revoked are more likely to flee than those who have legal licenses. • Hit-and-run crashes are more likely to occur at night, on weekends, on lower-speed local streets or roads.
(*18*)	Sivasankaran and Balasubramanian (2020)	2009 to 2017; Tamilnadu, India; logit model	• Unlit conditions increase the likelihood of pedestrian-involved hit-and-run (PIHR) crashes. • Pedestrians aged 18 + years and alcohol consumption of pedestrians are significantly associated with PIHR crashes.
(*24*)	Hitosugi et al. (2021)	2002 to 2016; Japan; descriptive	• The rate of hit-and-run cases was also significantly higher among pedestrians who were lying on the road. • Among fatally injured pedestrians not lying on the road, the rates with speeds of 30 km/h or greater did not differ significantly between hit-and-run and other cases.

Note: DWI = driving while impaired; DUI = driving under the influence; FARS = Fatality Analysis Reporting System.

Outside of the United States, studies that have undertaken PIHR crash analyses exclusively include Ghana (*3*), Singapore (*4*), India (*18*, *22*, *26*), Australia (*27*) and Japan (*24*, *28*).

Table 1 summarizes the key findings from select crash studies that delve into various facets of PIHR crashes. Although these studies have been conducted across diverse countries, they predominantly utilize fatal crash data for their PIHR crash analyses. They encompass a wide range of timeframes and geographic locations, aiming to pinpoint the factors that play a role in these incidents. The results indicate statistical association between PIHR crashes and numerous individual factors, including time of day, driver demographics, alcohol consumption, road conditions, and pedestrian actions, underscoring the intricate nature of comprehending and addressing these crashes.

The existing literature mainly relies on logit models and descriptive analyses to explore the relationships between individual factors and PIHR crashes. Moving beyond this approach, this study aims to uncover the collective dynamics of multiple potentially contributing elements in PIHR crashes. Specifically, we examine the combined associations of attributes in Louisiana using novel CCA methods, to offer a more comprehensive understanding of the intricate nature of PIHR crashes.

Data

Data Preparation

The annual traffic crash databases of the Louisiana Department of Transportation and Development (DOTD) consist of police-recorded information on the spatial and temporal environment of crashes, in addition to general roadway information, and include demographic and condition data related to the vehicle drivers and pedestrians involved. Each year, individual crash records noting driver, vehicle, pedestrian, and roadway details along with driving conditions are compiled in several disaggregated data tables, from which relevant attributes were extracted and merged based on the crash ID that is present in all of the tables. The "Crash Data" table containing temporal, road environment-related information of an individual crash event was filtered for mutually exclusive pedestrian crash and hit-and-run crash events. The "Vehicle Data" consisted of vehicle and driver information by presenting each row as a vehicle, from which information of only "Vehicle 1," the vehicle holding major responsibility for the crash, was extracted. Finally, the "Pedestrian Data" table had information about the pedestrian involved in the crash.

In the data preparation process of this research, the annual datasets for the 5 years from 2015 to 2019 were organized and compiled into a crash-level dataset in which each record represents one crash. On the compiled dataset, data wrangling was performed that involved removing duplicates, transforming data types, and rearranging data into a desired format. The final PIHR crash dataset had 2,201 crashes during the 5-year period of 2015 to 2019. A simplified overview of the data preparation process and the analysis steps are presented in

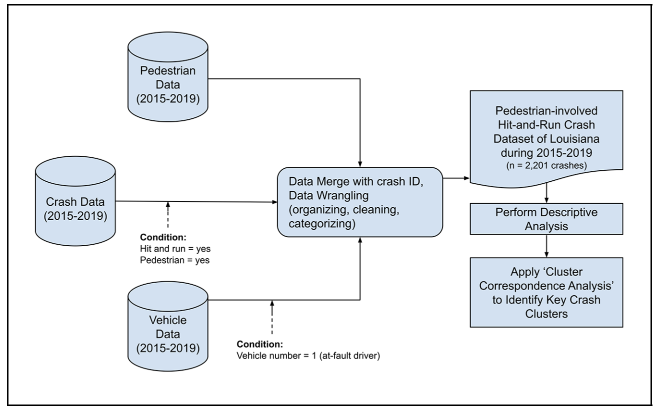

Figure 1. Outline for data preparation and analysis.

Figure 1. The analytical contexts and theoretical backgrounds are presented in the following subsections.

Limitations in the final dataset remain within the scope of this study. A combination of subjective judgment, lessons learned from past studies, and availability in the state-issued crash database was used for initial variable selection in this study. The unavailability of driver data in hit-and-run crashes owing to drivers leaving the scene poses a major challenge in recording accurate crash data for investigating police officers, who often need to rely on voluntary statement(s) from potential witness(es) in such crashes (*29*). Because of substantial underreporting of driver age and gender, 65% and 77% respectively, these two variables were removed from the dataset aiming to avoid uninterpretable results. Removing the crashes with unknown attributes would have largely reduced dataset sample size, which in turn emphasizes the issue of reporting difficulties of the PIHR crashes, broadly any hit-and-run crashes. Although this study is rooted in a specific dataset, for a relatively small amount of missing information, the gaps can be addressed by categorizing them as "unknown" or through imputation methods like random or model-based techniques, ensuring robust analysis even with incomplete data. Aiming not to reduce the sample size of the dataset, the unknown attributes of the remaining variables were retained as "unknown."

Conversely, the research team incorporated variables such as pedestrian condition and pedestrian action into the dataset, based on their inclusion in previous Louisiana studies on pedestrian crash data analysis (*30*, *31*), and the presumption that these factors could be influential in PIHR crashes as well. In addition, although the 2020 crash data became available at the time of submission of the research paper, they were not included in the study to prevent any potential misinterpretation of findings that could have arisen from the impact of the COVID-19 pandemic.

Variable	#	%	Barplot
Season			
Fall	606	27.53	
Winter	524	23.81	
Spring	608	27.62	
Summer	463	21.04	
Time			
Early-morning (4 a.m. to 6 a.m.)	92	4.18	
Morning-rush (6 a.m. to 9 a.m.)	124	5.63	
Late-morning (9 a.m. to 12 p.m.)	141	6.41	
Afternoon (12 p.m. to 4 p.m.)	317	14.40	
Evening-rush (4 p.m. to 7 p.m.)	390	17.72	
Early-night (7 p.m. to 11 p.m.)	646	29.35	
Late-night (11 p.m. to 4 a.m.)	113	5.13	
Unknown	378	17.17	
Day of the Week			
Fri-Sat-Sun	993	45.12	
Mon-Tue	598	27.17	
Wed-Thu	610	27.71	
Lighting Condition			
Daylight	943	42.84	
Dark-lighted	894	40.62	
Dark-unlighted	246	11.18	
Twilight	69	3.13	
Unspecified	49	2.23	
Weather			
Normal	1723	78.28	
Cloudy	277	12.59	
Rainy	137	6.22	
Unspecified	64	2.91	

Figure 2. Distribution of temporal and road environment variables.

In addition to the previously discussed data limitations, it is pertinent to address the potential application of our methodology beyond the Louisiana crash records database. This expansion in our data limitation discussion underscores our methodology's versatility and potential for broader application.

Descriptive Analysis

The descriptive analysis of the PIHR crash dataset offers insights into the data distribution, which is presented in percentages of attributes and barplots. These distributions have been separately presented in Figures 2 to 4. These categorical data distributions represent the initial exploration of the final crash dataset, consisting of 2,201 PIHR crashes, which also serve as a data reference for the final cluster analysis.

Temporal and Road Environment Characteristics. The times of day were categorized into unequal intervals to capture the potential effect of visibility as well as pedestrian volume in the crash clusters (Figure 2), recognizing that pedestrian activity is unevenly distributed throughout the day. Under the assumption that the direct link between the high frequency of PIHR crashes and the high volume of pedestrians exists, the prevalence of PIHR crashes was higher during early night and evening peak hours. When normalized per hour, the order of the three time intervals with the highest crash rates was as follows: early night at 161.5 crashes/hour, evening peak hour at 130 crashes/hour, and afternoon at 79 crashes/hour for the 5 year total. Prevalence is potentially the result of heightened pedestrian activity, diminished visibility, or a combination of the two. PIHR crashes during the weekend (during Friday to Sunday, 3 days) are understandably higher than during the two other weekday categories, each of which is spread across 2 days. However, crashes in the "Friday to Sunday" category are also higher (331 crashes per day) when normalized by the number of days, in comparison to Monday to Tuesday (299 crashes per day) and Wednesday to Thursday (305 crashes per day). PIHR crashes were predominantly observed during daylight hours and lighted dark conditions compared with unlighted dark conditions and twilight.

Variable	#	%	Barplot
Roadway Configuration			
One-Way	366	16.63	
Two-way with no separation	1269	57.66	
Two-way with separation	451	20.49	
Two-way with barrier	51	2.32	
Unspecified	64	2.91	
Highway Type			
Interstate	55	2.50	
U.S. highway	144	6.54	
State highway	349	15.86	
City street	1334	60.61	
Parish road	284	12.90	
Local road	34	1.54	
Off-road	1	0.05	
Posted Speed Limit			
25 mph or less	1049	47.66	
30 to 35 mph	654	29.71	
40 to 45 mph	329	14.95	
50 to 55 mph	118	5.36	
60 mph or more	51	2.32	
Area			
Rural	493	22.40	
Urban	1708	77.60	
Intersection			
Yes	1116	50.70	
No	1085	49.30	
Location Type			
Business + Residential	779	35.39	
Business continuous	510	23.17	
Manufacturing/Industrial	36	1.64	
Residential district	630	28.62	
Residential scattered	119	5.41	
School or Playground	29	1.32	
Open Country	54	2.45	
Unspecified	44	2.00	

Figure 3. Distribution of location and roadway variables.

Variable	#	%	Barplot
Pedestrian Sex			
Female	756	34.35	
Male	1353	61.47	
Unknown	92	4.18	
Pedestrian Age			
<15	170	7.72	
15-24	449	20.40	
25-44	804	36.53	
45-64	551	25.03	
65 and older	96	4.36	
Unknown	131	5.95	
Pedestrian Dark Cloth			
Yes	663	30.12	
Partial	783	35.57	
No	755	34.30	
Pedestrian Condition			
Normal	1130	51.34	
Distrated/Inattentive	202	9.18	
Intoxicated/Drowsy	155	7.04	
Unspecified	714	32.44	
Pedestrian Action			
Crossing at intersection	535	24.31	
Crossing not at intersection	298	13.54	
Vehicle entry, exit, or repair	32	1.45	
Walking against traffic	95	4.32	
Walking with traffic	266	12.09	
Working or playing in road	72	3.27	
Stationary on road	123	5.59	
Not in roadway	232	10.54	
Unspecified	548	24.90	
Pedestrian Injury			
Fatal (K)	126	5.72	
Severe (A)	199	9.04	
Moderate (B)	806	36.62	
Possible/Complaint (C)	795	36.12	
No injury (O)	158	7.18	
Unknown (U)	117	5.32	

Figure 4. Distribution of pedestrian variables.

Location and Roadway Characteristics. Figure 3 shows the distribution of location and roadway variables. Most PIHR crashes occurred on two-way roads without separation, followed by two-way roads with separation, and one-way roads typically found in urban areas and city centers, which often have a sizable pedestrian volume. The majority of the PIHR crashes happened on city streets, state highways, and parish roads. However, the PIHR crashes were more concentrated in areas where relatively lower speed limits were imposed possibly owing to known high pedestrian activity, since a substantial portion of crashes occurred in zones with a posted speed limit of 25 mph or less, as well as in areas with speed limits between 30 to 35 mph. Furthermore, more than three-quarters of these crashes took place in urban areas, with the location types associated with higher crash frequencies including mixed business and residential, residential districts, and continuous business areas. This observation further emphasizes the potential association between pedestrian activity and the prevalence of PIHR crashes. PIHR crashes were almost evenly distributed between intersections and nonintersection locations.

Pedestrian Characteristics. As presented in Figure 4, in the PIHR dataset, most of the pedestrians involved in crashes were male, whereas a smaller portion were female. The largest group of pedestrians fell into the 25 to 44 age range, followed by those aged 45 to 64 years, and 15 to 24 years. Pedestrians younger than 15 and those 65 and older represented smaller portions, whereas the age of a few pedestrians remains unknown. Considering dark clothing, the dataset showed a relatively even distribution between pedestrians wearing dark clothing, partially dark clothing, and nondark clothing. In relation to pedestrian condition, more than half appeared to be in normal condition, with smaller proportions being distracted or inattentive, intoxicated or drowsy, and unspecified conditions. Pedestrian actions covered a range of activities, with crossing at intersections being the most common, followed by crossing not at intersections and walking with traffic. Other actions included walking against traffic, stationary on the road, not in the roadway, and unspecified actions. The term "unspecified actions" refers to actions that are either unknown or represent a combination of low-frequency actions that have been merged for analysis purposes. A smaller proportion of pedestrians were working or playing in the road or involved in vehicle entry, exit, or repair. Finally, the severity of pedestrian injuries varied, with moderate injuries and possible injuries or complaints being the most prevalent. Severe and fatal injuries made up a smaller portion, whereas no injury or unknown injury status accounted for the remaining cases.

Methdology

Cluster Correspondence Analysis (CCA)

Correspondence analysis (CA) in general is popular with researchers who deal with high-dimensional datasets. This method reduces dimensionality by projecting the data to a lower-dimensional subspace by creating a simple structure from which important knowledge can be extracted. Transportation safety researchers utilize multiple correspondence analysis, a specific type of CA for categorical variables, to detect underlying structures in nominal categorical datasets (*32–38*). In these research studies, biplots, which are visual representations of two-dimensional projections of high-dimensional categorical datasets, were frequently used. The CCA approach, an extended version of CA, conjointly performs dimension reduction and cluster analysis for categorical data (*39*).

Unlike statistical techniques that often rely on predetermined assumptions, CCA does not necessitate prior assumptions for its execution. Additionally, effective analysis of interactive factors using logit models requires extensive domain knowledge and thorough, repeated examination of multiplicative factors to give dependable results of attribute association. CCA focuses on revealing associations between variables, making it less sensitive to the impact of missing data compared with traditional modeling techniques. Because of the challenges faced by police officers in retrieving quality driver demographic data, our analysis primarily focused on the available variables related to roadway characteristics, roadway conditions, and pedestrian factors. The CCA method offers flexibility, visualization, and robustness while providing interpretable representations of complex relationships from categorical datasets that may contain missing variables.

This characteristic of CCA facilitates a more adaptable examination of data, allowing the data itself to guide the discovery of relationships or patterns. Essentially, it is expected to help in generating the typologies of PIHR crashes based on their attributes, putting emphasis on patterns that might not be apparent through traditional cluster analysis alone. Through the dimension reduction integrated into the cluster analysis, CCA aims to provide insights into the characteristics of crashes assigned to each cluster, highlighting which characteristics are over- or underrepresented. The primary benefit of employing CCA alongside cluster analysis lies in its ability to handle datasets with numerous categorical variables more effectively. It helps uncover the relationships between categorical variables and clusters, providing a better understanding of the data. However, it is essential to remember that CCA is an exploratory technique, not a predictive one, and traditional modeling techniques may still be necessary for making predictions or establishing potentially causal relationships.

This novel method was recently proposed by van de Velden et al. (*39*). In relation to cluster convergence, this method has outperformed other approaches aiming to either sequentially or jointly apply dimension reduction and clustering (*40*). Recent transportation safety studies utilized the CCA approach to investigate issues such as rainy weather crashes (*41*), light delivery vehicle crashes (*42*), marijuana-impaired traffic crashes (*43*), and fatal crashes involving mopeds (*44*).

CCA Algorithm

Theoretically, CCA can be understood as applying correspondence analysis to the cross-tabulation of cluster membership and variable categories, which results in a cluster by categories contingency matrix. In mathematical terms,

$$F = \boldsymbol{Z}_K^\top \boldsymbol{Z} \tag{1}$$

where

F = A matrix cross-tabulating cluster memberships with categorical variables,

Z = A super indicator matrix representing data on n individuals across p categorical variables, and

Z_k = An indicator matrix for cluster membership, where n is the number of individuals and k is the number of clusters.

By applying CA to this matrix, the method derives optimal scaling values for rows (clusters) and columns (categories), ensuring the maximum between-cluster variance. Consequently, clusters are optimally separated based on the categorical variable distributions. Simultaneously, categories with differing distributions over clusters are also optimally separated. However, the method must determine unknown cluster memberships. It can be demonstrated that optimal category quantifications (i.e., column coordinates) and an optimal cluster allocation can be achieved by iterating between CA of the contingency matrix and applying *K*-means cluster analysis to the reduced space coordinates obtained using the CA category quantifications.

The CCA algorithm can be summarized in the following steps:

1. Generate an initial cluster allocation, $\boldsymbol{Z}_K$, (for example, by randomly assigning objects to clusters).
2. Obtain category quantifications, B, by applying CA to the contingency matrix, $\boldsymbol{Z}_K^\top \boldsymbol{Z}$.
3. Compute coordinates for the individuals (or objects) by averaging the centered scores using the category quantifications from Step 1: $\boldsymbol{Y} = \frac{1}{q}(\boldsymbol{I} - \mathbf{1}\mathbf{1}^\top/n)ZB$
4. Update $\boldsymbol{Z}_K$ by applying *K*-means clustering to $\mathbf{Y}$.
5. Repeat the procedure (i.e., return to Step 2) using $\boldsymbol{Z}_K$ for the cluster allocation matrix until convergence is reached, meaning $\boldsymbol{Z}_K$ (and therefore $\mathbf{Y}$ and $\mathbf{G}$) remain constant.

Convergence is ensured once the Calinski–Harabasz index (which means, the ratio of the sum of between-cluster variance to within-cluster variance) ceases to decrease in consecutive steps. For a more in-depth understanding of the theoretical foundation of CCA, readers are directed to van de Velden et al. (*39*).The researchers used the "clustrd" package (*45*) of R statistical software (*46*) to apply CCA on the PIHR crash dataset.

Selection of Number of Clusters

The CCA algorithm requires selection of the number of clusters before its application. Two approaches were

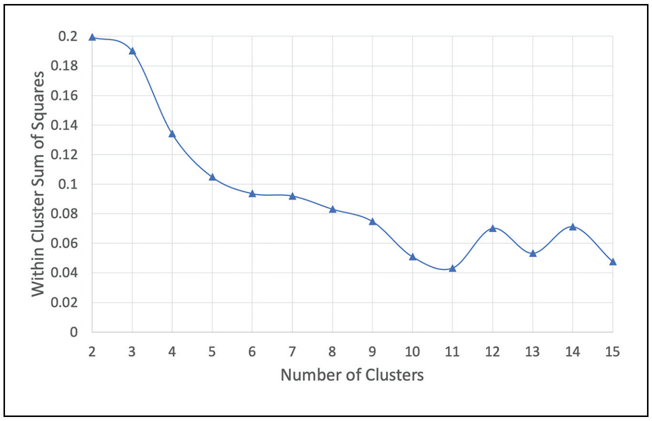

Figure 5. WCSS values by cluster quantity.
Note: WCSS = within-cluster sum of squares.

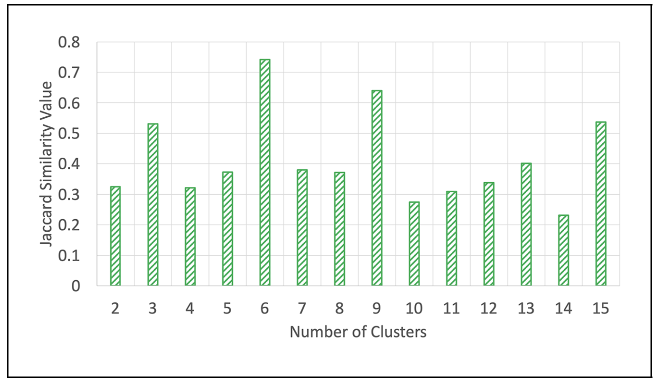

Figure 6. Jaccard similarity values by cluster quantity.

undertaken to identify the optimal number of clusters in the dataset, the first of which is a heuristic method referred to as "elbow method." The main idea behind the elbow method is to find the point at which adding more clusters does not lead to a substantial decrease in the within-cluster sum of squares (WCSS). This point is called the "elbow" because it often resembles an elbow shape on a plotted graph. Although the elbow method has its roots in the work of Thorndike (*47*), comprehensive details of elbow method in clustering can be found in Everitt et al. (*48*). Based on the WCSS values depicted in Figure 5, the 6-, 11-, and 13-cluster solutions seemed to be among the candidates of the optimal choices. Although we could have achieved a lower total within sum of squares by using more clusters, for example 11 or 13 clusters, the loss of WCSS tended to be inconsistent in this range. Furthermore, an increased number of clusters might lead to a less interpretable and more scattered combination of crash characteristics. This can complicate the analysis and hinder the identification of meaningful patterns within the data (*49*).

In addition to checking within-cluster compactness by WCSS, an evaluation of the clusterwise stability was conducted to confirm the selection of the number of clusters using the Jaccard similarity value. By running joint dimension and clustering algorithms repeatedly for different numbers of clusters on a bootstrap replica of the original data, and by examining the returns of corresponding cluster assignments and cluster agreement indices that compare pairs of partitions, number of optimal clusters was determined. The average of these similarities was used as a measure of the cluster's stability. This approach bears similarities to the method outlined by Hennig (*50*).

The Jaccard similarity index was calculated and found to be highest for a 6-cluster solution (Figure 6), demonstrating stability as it exceeded the threshold of 0.75 (*50*). On examination of the outcomes from both the elbow method and the Jaccard similarity index, the 6-cluster solution was deemed optimal. Consequently, the chosen number of clusters for the implementation of the CCA was six.

Results and Discussions

The two-dimensional solution obtained consists of six clusters with sizes of 770 (35%), 682 (31%), 484 (22%), 171 (7.8%), 49 (2.2%), and 45 (2%). The WCSS values for each of the six clusters in a cluster solution represent the amount of variation or dispersion within that cluster. The WCSS measure for each cluster was found to be 0.0187, 0.0188, 0.0172, 0.0156, 0.0133, and 0.0101. The WCSS values increased from Cluster 1 to Cluster 6, suggesting that the data points in each cluster become more closely grouped around the respective cluster centroids as the cluster number increases. In the biplot in Figure 7, a two-dimensional projection of the data points is presented, identifying six clusters: C1, C2, C3, C4, C5, and C6. In Figure 7, Dim. 1 and Dim. 2 refer to Dimension 1 and Dimension 2, respectively. Cluster 1 has the most compact and closely grouped cluster, whereas Cluster 6 has the largest dispersion of data points among all the clusters. This suggests that certain clusters, such as C1 and C2, exhibited more distinct and well-defined characteristics compared with other clusters like C5 and C6.

The ratio between the sum of squares and the total sum of squares for the 6-cluster solution was 89.88%, indicating that a large proportion of the variance in the data can be explained by differences between clusters. Table 2 presents the cluster size and dimension characteristics.

Description of Clusters

The CCA method facilitates the identification of attributes that exhibit significant deviations from the independence condition, thereby assisting in the

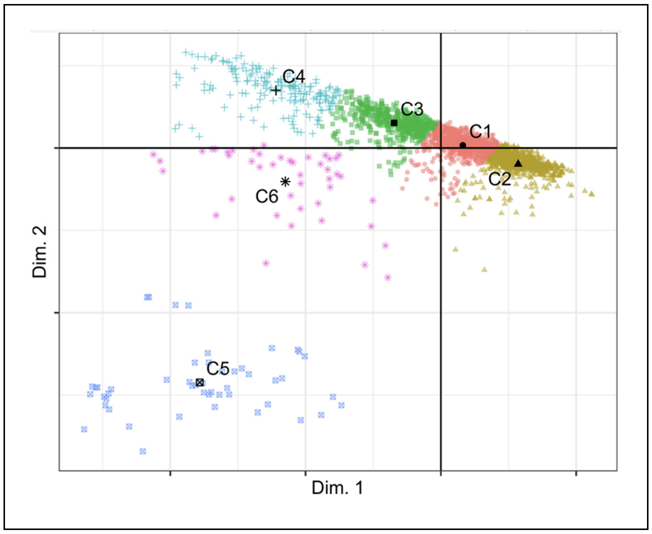

Figure 7. Identified clusters on a two-dimensional projection of data points.

interpretation of cluster characteristics. Figures 8 to 13 present six bar plots presenting the top 20 of the largest standardized residuals for Clusters 1 to 6. The bars in the resulting plot correspond to the highest (in absolute value) standardized residuals from the independence of the attribute distributions conditional on clusters. A positive residual means that the attribute has an above-average frequency within the cluster, whereas a negative residual means the attribute has a below-average frequency. For better interpretability, the attributes that are positively associated with the clusters are described. Clusters are primarily identified by the overrepresentation of certain attributes, which constitute at least 70% within each cluster, in relation to their percentages.

Cluster 1 (C1): City Streets, Urban Area. Cluster 1, as presented in Figure 8, is mainly associated with city streets, as the attribute 'highway type = city street' had the highest positive residual in the cluster. The proportion of city streets was the largest in the entire dataset (Figure 3), and it emerged as the most notable positively associated attribute in the largest cluster, thereby making it a crucial factor in PIHR crashes. Cluster 1 could further be positively associated with residential district, urban area, dark but lighted condition, early night (7 to 11 p.m.), and pedestrian aged below 15 years.

To gain insights into the distributional measures of Cluster 1, we examined the top positive attributes in relation to cluster memberships derived from the CCA analysis. Table 3 presents a comparison of the top positive attributes associated with Cluster 1, including their respective percentages within the cluster, and the cluster-to-dataset percentage ratio of each attribute. By calculating the cluster-to-dataset percentage ratio, we can effectively measure the strength of association between the attributes and Cluster 1. The estimated cluster-to-dataset percentage ratios presented in this paper are based on the precise values calculated using Excel, which may not match the values estimated from the rounded figures presented in the paper owing to rounding error.

City streets in general facilitate numerous formal and informal activities that pedestrians engage in—commuting, shopping, socializing, exercising, sightseeing, attending events, accessing public transportation, and so forth—and consequently allow for an increased exposure to pedestrian–vehicle crashes (*51*). Some 78% of crashes in Cluster 1 occur on city streets, whereas 60.61% of crashes occurred on city streets in the whole dataset (Figure 3). Proportionately (based on the proportion of the percentages within Cluster 1 in Table 3 and the percentages within the whole dataset in Figure 3), PIHR crashes on city streets in Cluster 1 were 1.28 times more likely compared with PIHR crashes in the entire dataset. In Cluster 1, 94% of crashes occurred in urban areas as opposed 77.6% in the whole dataset (Figure 3). As a result, PIHR crashes in Cluster 1 were 1.22 times more likely to take place in urban settings than in the whole dataset. A substantial amount of prior research has suggested that with this heightened crash exposure, the likelihood of hit-and-run incidents occurring in urban areas is also elevated (*3*, *4*, *18*, *23*). The finding of the association of urban areas with PIHR crashes is in line with the prior hit-and-run studies (*17*, *18*). The attribute

Table 2. Cluster dimensions and quality criterion values

Cluster	Cluster centroids		Distance from origin	Within sum of squares	Cluster size, *N* (%)
	Dimension 1	Dimension 2			
C1	0.0041	0.0008	0.0042	0.0187	770 (35)
C2	0.0143	−0.0049	0.0149	0.0188	682 (31)
C3	−0.0086	0.0076	0.0115	0.0172	484 (22)
C4	−0.0305	0.0175	0.0353	0.0156	171 (7.8)
C5	−0.0445	−0.0712	0.0839	0.0133	49 (2.2)
C6	−0.0287	−0.0102	0.0302	0.0101	45 (2)

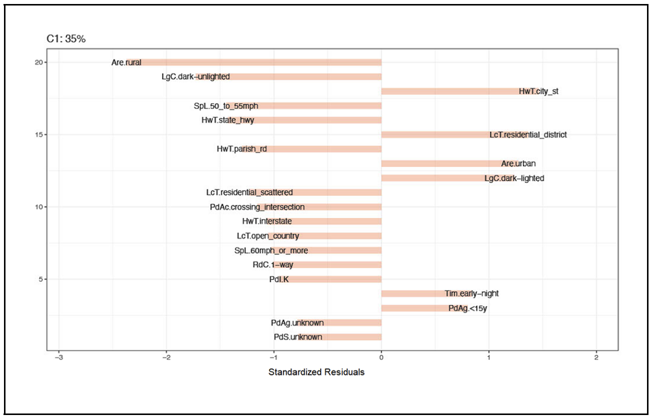

Figure 8. Top 20 of the largest standardized residuals of Cluster 1.

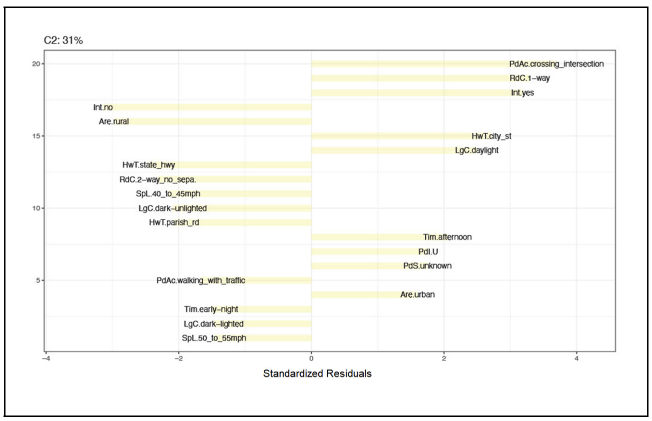

Figure 9. Top 20 of the largest standardized residuals of Cluster 2.

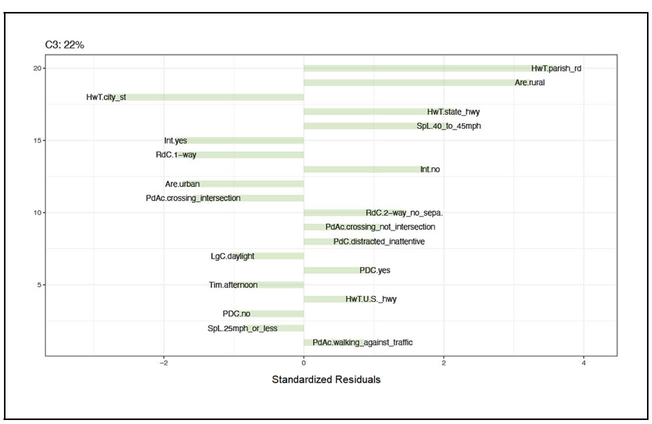

Figure 10. Top 20 of the largest standardized residuals of Cluster 3.

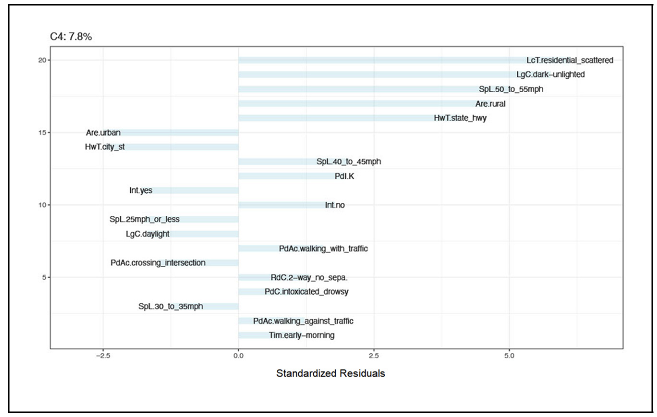

Figure 11. Top 20 of the largest standardized residuals of Cluster 4.

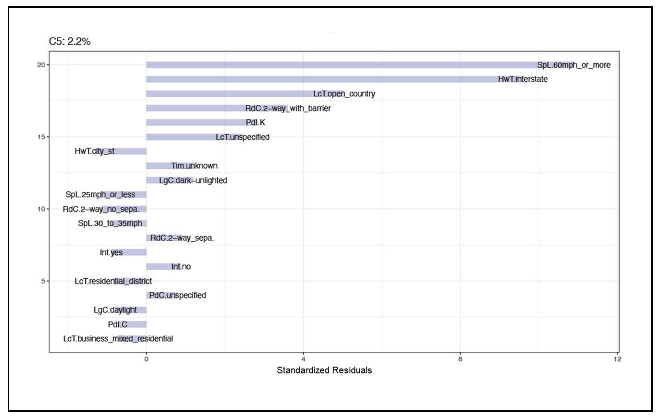

Figure 12. Top 20 of the largest standardized residuals of Cluster 5.

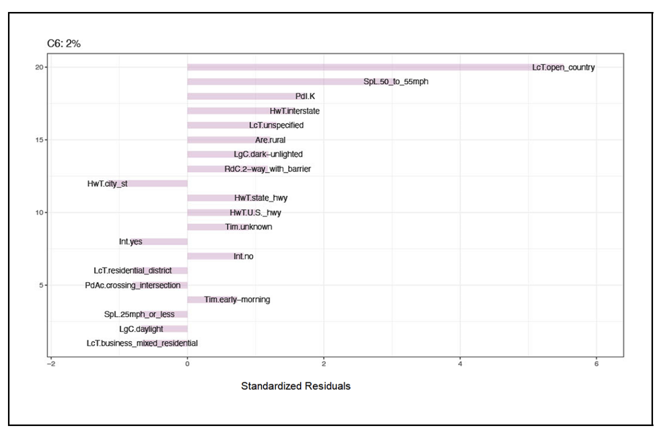

Figure 13. Top 20 of the largest standardized residuals of Cluster 6.

association of PIHR crashes with urban areas and city streets is apparently interrelated, with urban areas also covering highway types other than city streets.

The proportion of PIHR crashes under dark-lighted conditions was 52% in Cluster 1 to 40.6% in the whole dataset (Figure 2) resulted in a 1.29 times higher

Table 3. Comparison of Top Positive Attributes of Cluster 1

Top positive attributes of Cluster 1	Percentage within cluster	Cluster-to-dataset percentage ratio
Highway type = City street	78	1.28
Location type = Residential district	39	1.38
Area = Urban	94	1.22
Lighting condition = Dark-lighted	52	1.29
Time = Early night	36	1.23
Pedestrian age <15 years	11	1.43

likelihood of these conditions being associated with Cluster 1. Although poor visibility is expected to trigger PIHR crashes, Cluster 1 covers nighttime crashes in areas with streetlights. The cluster further pinpoints the crashes during 7 to 11 p.m. When examining the time of crashes, 36% of early night PIHR crashes were in Cluster 1, compared with 29.35% in the entire dataset, yielding a 1.23 times greater chance of Cluster 1 crashes occurring during this time period. In relation to the time of the day, pedestrian activity in low visibility conditions was highest during early night (7 to 11 p.m.) in city areas, not accounting for weather conditions, as supported by previous studies (*52*, *53*).

Comparing the 39% share of residential districts in Cluster 1 with their 28.62% representation in the entire PIHR dataset suggested a 1.38 times higher likelihood of association with Cluster 1. Notably, 11% of pedestrians involved in crashes within Cluster 1 were under 15 years of age in comparison to 7.72% in the whole dataset, demonstrating a 1.43 times stronger association between pedestrians aged less than 15 years and Cluster 1 crashes. It was interesting to discover that Cluster 1, which largely represents city streets and urban areas, was also associated with pedestrians aged less than 15 years. A study by Tay et al. identified that children are an infrequent yet more vulnerable pedestrian group, but they interpreted that drivers are less likely to flee if the pedestrian victims are children (*4*). The finding that the PIHR crash cluster associated with city streets in urban residential districts also captured pedestrians aged under 15 years is noteworthy.

Cluster 2 (C2): City Street, Urban Area, Intersection. Standardized residuals of Cluster 2 identified from the CCA are displayed in Figure 9. City streets and urban areas accounted for 60.61% and 77.6% (Figure 3), with an observed distribution pattern of 94% and 100%, respectively, in Cluster 1. Besides the positive association of these two attributes, Cluster 2 represents crashes associated with one-way road, intersection,

Table 4. Comparison of Top Positive Attributes of Cluster 2

Top positive attributes of Cluster 2	Percentages within cluster	Cluster-to-dataset percentage ratio
Pedestrian action = Crossing at intersection	53	2.18
Roadway configuration = One-way	38	2.27
Intersection = Yes	85	1.67
Highway type = City street	94	2.31
Lighting condition = Daylight	69	1.60
Time = Afternoon	27	1.85
Pedestrian injury = U	12	2.21
Pedestrian sex = Unknown	10	2.35
Area = Urban	100	1.29

Note: U = unknown injury.

Table 5. Comparison of Top Positive Attributes of Cluster 3

Top positive attributes of Cluster 3	Percentage within cluster	Cluster-to-dataset percentage ratio
Highway type = Parish road	37	2.88
Area = Rural	51	2.28
Highway type = State highway	32	2.01
Speed limit = 40 to 45 mph	30	2.00
Intersection = No	75	1.52
Roadway configuration = Two-way with no separation	78	1.35
Pedestrian action = Crossing not at intersection	21	1.56
Pedestrian condition = Distracted/ inattentive	15	1.67
Pedestrian dark clothing = Yes	45	1.51
Highway type = U.S. highway	11	1.71
Pedestrian action = Walking against traffic	8	1.77

daylight conditions. This cluster can further be associated with afternoon, unknown pedestrian injury, and unknown pedestrian sex.

Roadway configuration, such as one-way or two-way streets, may play an important role in hit-and-run crashes. One-way streets, which are typically undivided, demonstrate a higher tendency for crashes compared with divided highways without traffic barriers, and those featuring median strips and two-way continuous left-turn lanes, which offer added safety for pedestrians (*5*). This cluster partially represents some combinations of two predominant attributes: one-way roads, which are the most prevalent roadway configuration associated with PIHR crashes, and crossing at intersections, which is the most common pedestrian activity related to PIHR

Table 6. Comparison of Top Positive Attributes of Cluster 4

Top positive attributes of Cluster 4	Percentage within cluster	Cluster-to-dataset percentage ratio
Location type = Residential scattered	50	9.30
Lighting condition = Dark-unlighted	72	1.77
Speed limit = 50 to 55 mph	42	7.85
Area = Rural	92	4.10
Highway type = state highway	67	4.24
Speed Limit = 40 to 45 mph	40	2.66
Pedestrian injury = K	20	3.58
Intersection = No	91	1.84
Pedestrian action = Walking with traffic	29	2.42
Roadway configuration = Two-way with no separation	89	1.54
Pedestrian condition = Intoxicated/drowsy	18	2.49
Pedestrian action = Walking against traffic	12	2.85
Time = Early morning	12	2.80

Note: K = fatal.

incidents. Urban intersections have been identified as more likely to be linked with hit-and-run crashes than rural intersections (*54*).

Daylight conditions was 1.6 times more likely to be associated with Cluster 2 in comparison to the whole dataset, capturing 69% crashes in Cluster 2 (Table 4) out of a total 43% PIHR crashes (Figure 2). The likelihood of hit-and-run crashes are known to be higher during nighttime in reference to daytime (*4*); however, Cluster 2 encompasses PIHR crashes in daytime, implying PIHR crashes in city streets can occur during both the daytime and nighttime. The cluster represents 27% of the crashes to have occurred in afternoon (12 noon to 4 p.m.), showing afternoon crashes are likely to be linked to these daytime PIHR crashes in city streets.

This cluster revealed that approximately one-tenth of Cluster 2 primarily featured unknown pedestrian injury conditions and unknown pedestrian sex. Many crash narratives indicated that further investigation may not have been conducted to identify the drivers or pedestrians if they were unknown. In most instances, the victim may not have been available following the crash event, and the pedestrian's involvement in the crash was only reported by a witness.

Cluster 3 (C3): Undivided, Nonintersection. In Cluster 3 (Figure 10), the positively associated attributes were parish road, rural area, 40 to 45 mph speed limit, nonintersection, two-way roadways with no separation, pedestrians crossing at locations that do not have an intersection, distracted/inattentive pedestrian, pedestrians with dark clothing, both state highway and U.S. highway, and pedestrians walking against traffic. Table 5 shows parish roads and U.S. highways presented 37% and 32% of the PIHR crashes in Cluster 3. The attributes parish road, rural area, state highway, and 40 to 45 mph speed limit had at least a two times higher likelihood than the whole dataset. In contrast with Clusters 1 and 2, this cluster showed a higher likelihood of PIHR crashes in rural areas, 2.28 times that of the whole dataset.

One of the main highlights of this cluster was the pedestrians crossing at locations that do not have an intersection. From the clustered datasets, it was observed that these crashes can also be partially linked to distracted or inattentive drivers. This cluster also separately showed an association with pedestrians wearing dark clothes. An association with the 40 to 45 mph speed limit suggests that PIHR crashes, resulting from direct impacts with vehicles traveling at these speeds, can lead to a higher risk of severe injuries or fatalities compared with collisions at lower speeds (*55*, *56*). This also indicates that other collisions might have been indirect or occurred at lower speeds. A small portion of the PIHR crashes in this cluster also involved pedestrians walking against traffic, mainly in rural areas, as observed from the clustered dataset.

Cluster 4 (C4): Undivided, Rural, Nonintersection, Dark-Unlighted. Lighting conditions have been identified as a critical factor in a driver's decision to flee after a pedestrian crash (*5*, *17*). Because of the lower perceived likelihood of pedestrians being seen in unlit conditions, PIHR collisions occurring in the dark may encourage drivers to escape the scene of a crash in rural locations. Furthermore, there are unlikely to be witnesses in rural areas during such the dark-unlighted condition. The time interval that has been associated with this cluster is early morning (4 to 6 a.m.), which also represents the dark condition. In addition to dark-unlighted conditions, this cluster was associated with undivided roadways, scattered residential locations, rural areas, and nonintersection sites (Figure 11). Previous Louisiana studies indicate that scattered residential locations and rural areas are often closely related (*38*).

Unlike the prior clusters, 20% of this cluster showed an association with pedestrian fatalities (Table 6). This could be further linked to the speeds of vehicles in rural areas, as this cluster further demonstrated an association with both the 40 to 45 mph and 50 to 55 mph speed limit groups, as well as with state highways and roadways with no separation. These speed limits are representative of common noninterstate highways in rural areas.

Table 7. Comparison of Top Positive Attributes of Cluster 5

Top positive attributes of Cluster 5	Percentage within cluster	Cluster-to-dataset percentage ratio
Speed limit = 60 mph or more	100	43.16
Highway type = Interstate	92	36.75
Location type = Open country	49	19.96
Roadway configuration = Two-way with barrier	35	14.97
Pedestrian injury = K	45	7.84
Location type = Unspecified	22	11.23
Time = Unknown	47	2.73
Lighting condition = Dark-unlighted	35	3.10
Roadway configuration = Two-way with separation	45	2.19
Intersection = No	88	1.78
Pedestrian condition = Unspecified	59	1.82

Note: K = fatal.

Pedestrian conditions and actions that were positively highlighted in this cluster were intoxicated/drowsy pedestrians, and pedestrians who were walking both with and against the traffic. Although prior studies have extensively discussed the issue of a potential link between driver intoxication and hit-and-run crashes (*5*, *13*, *57*), the finding of intoxicated or drowsy pedestrians being associated with hit-and-run crashes also showed an unusual pattern within PIHR crashes.

Cluster 5 (C5): Interstate, Nonintersection, 60 mph or More. Cluster 5 was a relatively small cluster, representing only 45 (2.2%) of PIHR crashes from the entire dataset. All of these crashes occurred on highways with a speed limit of 60 mph or higher, primarily on interstates (Figure 12). These crashes were also associated with open country locations where a pedestrian presence is unexpected. Notably, this cluster was linked to pedestrian fatalities in 45% of its crashes (Table 7), which would be anticipated at such high speeds and in areas where pedestrians are not typically found. This cluster also exhibited a connection with nonintersection locations and the presence of barriers or another form of separation between opposing traffic, which are typical features of interstate highways. Moreover, dark-unlighted conditions are common on interstates during nighttime hours. Unknown crash times and unspecified pedestrian conditions might also be attributed to limited visibility and a lack of witnesses at interstate crash locations. Pedestrians involved in PIHR crashes might originate from disabled vehicles, unintentionally finding themselves on interstates or other freeways, with a portion of these crashes also potentially stemming from unauthorized pedestrian movements from surrounding areas (*58*).

Table 8. Comparison of Top Positive Attributes of Cluster 6

Top positive attributes of Cluster 6	Percentage within cluster	Cluster-to-dataset percentage ratio
Location type = Open country	56	22.64
Speed limit = 50 to 55 mph	49	9.12
Pedestrian injury = K	31	5.43
Highway type = Interstate	18	7.11
Location type = Unspecified	13	6.67
Area = Rural	58	2.58
Lighting condition = Dark-unlighted	36	3.18
Roadway configuration = Two-way with barrier	13	5.75
Highway type = state highway	42	2.66
Highway type = U.S. highway	22	3.40
Time = Unknown	40	2.33
Intersection = No	87	1.76
Time = Early morning	13	3.19

Note: K = fatal.

The small size of Cluster 5 led to noticeably high cluster-to-dataset percentage ratios when compared with the entire dataset. As a result of the low percentage of PIHR crashes on interstates in the overall dataset and Cluster 5's large representation of interstate PIHR crashes, attributes related to interstates, such as "Speed limit = 60 mph or more" and "Location type = Open country," exhibited large cluster-to-dataset percentage ratio estimates. Most notably, pedestrian fatalities within this cluster appeared to be eight times more likely compared with the entire dataset.

Cluster 6 (C6): Nonintersection. Cluster 6 was also a relatively small and dispersed cluster representing only 45 (2%) of the total PIHR crashes. This small cluster mainly highlights nonintersection crashes with further representation of a few attributes such as open country location, rural area, state highway, pedestrian fatality (Figure 13 and Table 8).

Across-Cluster Characteristics

The significant attributes' distribution across clusters, as shown in Table 9, indicated clear distinctions, emphasizing the clusters' heterogeneity. Across-cluster percentages in Table 9, summing to 100% for each attribute, delineated the distribution of these significant attributes across different clusters, as previously illustrated in barplots. This approach highlights the attributes' varying levels of prominence in each cluster, providing a focused insight into their distribution. For instance, whereas Clusters 1 and 2 were predominantly found across city streets. The prominence of city streets in the two largest

Table 9. Across-Cluster Percentages of Top Positive Attributes

Attribute	Cluster 1, %	Cluster 2, %	Cluster 3, %	Cluster 4, %	Cluster 5, %	Cluster 6, %
Time = Early morning	—	—	—	22	—	7
Time = Afternoon	—	57	—	—	—	—
Time = Early night	43	—	—	—	—	—
Time = Unknown	—	—	—	—	6	5
Lighting condition = Daylight	—	50	—	—	—	—
Lighting condition = Dark-lighted	45	—	—	50	—	—
Lighting condition = Dark-unlighted	—	—	—	—	7	7
Roadway configuration = One-way	—	70	—	—	—	—
Roadway configuration = Two-way with no separation	—	—	—	12	—	—
Roadway configuration = Two-way with separation	—	—	—	—	5	—
Roadway configuration = Two-way with barrier	—	—	—	—	33	12
Highway type = Interstate	—	—	—	—	82	15
Highway type = U.S. highway	—	—	38	—	—	7
Highway type = State highway	—	—	44	33	—	5
Highway type = City street	45	48	—	—	—	—
Speed limit = 40 to 45 mph	—	—	—	21	—	—
Speed limit = 50 to 55 mph	—	—	—	61	—	19
Speed limit = 60 mph or more	—	—	—	—	96	—
Area = Urban	43	—	—	—	—	—
Area = Rural	—	—	50	32	—	5
Intersection = Yes	—	53	—	—	—	—
Intersection = No	—	—	32	14	4	3
Location type = Residential district	48	—	—	—	—	—
Location type = Residential scattered	—	—	—	72	—	—
Location type = Open country	—	—	—	—	44	46
Location type = Unspecified	—	—	—	—	25	14
Pedestrian sex = Unknown	—	73	—	—	—	—
Pedestrian age < 15 years	50	—	—	—	—	—
Pedestrian dark clothing = Yes	—	—	29	—	—	—
Pedestrian condition = Distracted/ inattentive	—	—	37	—	—	—
Pedestrian condition = Intoxicated/drowsy	—	—	—	19	—	—
Pedestrian condition = Unspecified	—	—	—	—	4	—
Pedestrian action = Crossing at intersection	—	68	—	—	—	—
Pedestrian action = Crossing not at intersection	—	—	34	—	—	—
Pedestrian action = Walking with traffic	—	—	—	19	—	—
Pedestrian action = Walking against traffic	—	—	39	22	—	—
Pedestrian injury = K	—	—	—	28	17	11
Pedestrian injury = U	—	68	—	—	—	—

Note: K = fatal; U = unknown injury.
Long dashes indicate that the attribute is not among top attriputes within the cluster.

clusters (Clusters 1 and 2) suggests that urban dynamics heavily influence these clusters. Clusters 3 and 4 were more common on U.S. and state highways. Meanwhile, Clusters 5 and 6 were mainly characterized by interstates and open country areas.

Pedestrian actions and conditions offer insights into the circumstances surrounding crashes, such as a high percentage of pedestrians crossing at intersections in Cluster 2 or a significant number being intoxicated or drowsy in Cluster 4. Several shared attributes were noted, such as Clusters 1 and 4 both experiencing crashes in dark-lighted conditions, whereas rural area and nonintersection crashes were predominant in Clusters 3 and 4. When examining pedestrian actions, both Clusters 3 and 4 comprised a notable percentage of PIHR crashes involving pedestrians walking against the traffic. An association between the seasons and PIHR crashes may not, however, be influential in Louisiana, as they were not highlighted in any of the clusters.

Discussions

Practical Implications

Given that the highest number of PIHR crashes occurred on city streets as identified in Clusters 1 and 2, the findings underscore the importance of ensuring that infrastructure is designed and maintained to support safe pedestrian movement, particularly on city streets with a high volume of pedestrian traffic. Urban communities in the United States have acknowledged the need for infrastructural improvements to address the lack of adequate and well-connected sidewalk facilities for pedestrians, that is, insufficient sidewalks and unconnected paths to essential locations (*59*). Louisiana's SHSP has recognized the correlation between pedestrian crashes and intersections (*10*). Moreover, aiming to lower pedestrian exposure to hit-and-run crashes, emphasis should also be placed on enhancing existing facilities that often pose challenges, such as narrow or blocked pathways, poor surfaces, insufficient buffers, difficult street crossings, and poor connectivity, all of which make it difficult for pedestrians to navigate comfortably and safely. The association between city street-, dark-lighted condition-, and early night PIHR crashes in Cluster 1 hints at the need for improving street lighting and visibility during early night hours, especially in Louisiana's city streets. However, another Louisiana study found no specific factors relevant to lighting requirements and suggested that further investigation into the local context and factors may be necessary for more targeted recommendations (*60*).

With regard to protecting vulnerable age groups, the CCA analysis results of Cluster 1 on PIHR crashes indicated an association with pedestrians in the very young group, that is, under 15 years of age. Although this cluster did not find an association with a specific pedestrian injury type, a longer-term compilation of Louisiana's pedestrian fatality data (2012 to 2021) shows this under 15 age group in general constitutes 54% of pedestrian fatalities (*61*). In relation to PIHR crashes associated with this age group, observations from this study pointed to city streets, more specifically in residential areas. It is therefore essential to consider the unique factors contributing to PIHR crashes during these periods, for example, travel patterns to and from school, to develop tailored safety measures. Offering pedestrians safety-related education and outreach programs has proved to be beneficial (*62*); such an approach could provide greater emphasis on safety by targeting children under 15 years of age who are more vulnerable to PIHR crashes.

Prioritizing pedestrian crossing safety on city streets could be a key strategy in preventing PIHR crashes, as indicated by Cluster 2. In regions with high volumes of pedestrians, prioritizing their movement on city streets may be necessary. Zegeer et al. suggest several interventions for undivided roads, including high-visibility crosswalk markings, parking restrictions near crosswalks, adequate nighttime lighting at uncontrolled crossings, advance yield signs on roads with speed limits of 35 mph or higher, and pedestrian hybrid beacons for roads with speed limits of 40 mph or higher (*63*). Other measures such as "No Left Turn" and "Slow Down" signs, along with speed bumps in pedestrian-heavy areas, can help regulate driver behavior and enhance pedestrian safety (*64*). It is important to assess the safety benefits of these countermeasures before their implementation.

Pedestrian activities, including walking with or against traffic and crossing at nonintersection locations (as identified in Cluster 3), are notable concerns on rural nonintersection segments of U.S. and state highways, which typically prioritize mobility over pedestrian accessibility and often have lower pedestrian volumes. Previous Louisiana studies indicate that scattered residential locations and rural areas are often closely related (*38*). Zajac and Ivan argued that if rural roadways include village and downtown fringe areas including residential areas, their designs should be made compatible with those in compact residential areas by implementing changes such as reducing roadway width (*65*). Potential traffic safety countermeasures for Cluster 4 could include improved lighting and increased signage, particularly in areas with high speed limits and low pedestrian presence. Consideration should also be given to implementing pedestrian crossings or barriers on interstates to reduce pedestrian exposure to high-speed traffic.

Although the number of PIHR crashes in Clusters 4, 5, and 6 was lower, their association with pedestrian fatalities should not be overlooked and warrants emphasis. These clusters suggest that visibility continues to be an important factor in PIHR crashes on high-speed roadways. Considering that the Safe System approach prioritizes elimination of crashes that result in death, in the context of PIHR crashes, focus might be directed toward high-speed roads with limits of at least 50 mph. These highways typically have controlled accessibility varying from partial control to fully control. Considering Johnson's recommendations to expand assistance for pedestrians who unintentionally find themselves on interstates—increasing emergency call stations and roving roadside assistance vehicles equipped with emergency cellular telephone numbers to report disabled vehicles (*66*)—an expansion of Louisiana DOTD's current "Motorist Assistance Patrol" program (*67*) could potentially aid in preventing PIHR crashes. Inexpensive countermeasures such as pedestrian fencing could prove beneficial if applied in potentially high-frequented locations that experience illegal encroachment by pedestrians from outside. Further investigation of geospatial locations of PIHR crashes in these clusters could help to develop similar proactive measures in a cost-effective manner.

Although the police continuously work to identify drivers at fault in hit-and-run incidents, it is also important to minimize cases with unknown pedestrian information. Hopkins et al. proposed the provision of incentives for using dashcam recording technology to combat hit-and-run crashes—this could also facilitate the minimization of data loss in such crashes involving pedestrians (*68*). We excluded driver data from analysis owing to a lack of information; however, for Safe System countermeasures, it is reasonable to discuss contexts not related to drivers. Employing social campaigns to promote hit-and-run laws and the benefits of reporting a crash, along with publicizing the resources available for crash reporting represents another strategy that warrants exploration for reducing hit-and-run incidents. Florida's "Stay at the Scene" is one such hit-and-run awareness program (*69*). Promoting pedestrian use of light or reflective clothing during prolonged nighttime walks could also be an effective approach to enhancing their visibility to motorists, reducing the risk of nighttime PIHR crashes (*64*).

Limitations of the Study

The study is understandably not without limitations. Although several pedestrian-related attributes were explored, only variables that were present in the dataset were included, and several driver variables were omitted owing to substantial underreporting. Were a considerable quantity of data for driver-related variables available, comparing crashes with and without these variables within the context of the identified clusters could offer additional insights into the driver characteristics that might influence the decision to flee a crash scene. This comparison could reveal influential factors beyond those identified in the existing literature. Removal of the driver-related attributes, driver age and driver gender, from the dataset limited findings related to driver demographics in the clusters.

Conclusions

Preventing PIHR crashes is of paramount importance, therefore, discovering the unidentified patterns structuring PIHR crashes is critical for advancing our understanding of the salient features that will enable researchers to develop strategic countermeasures. Because of undue randomness, great uncertainty in data reporting, and the disproportionately severe outcomes, this study used a Louisiana crash dataset from 2015 to 2019 to identify underlying patterns using an unsupervised algorithm, CCA. Through a joint dimension reduction and clustering approach that overcomes the limitations of a multidimensional categorical dataset, six clusters were identified in the PIHR crash dataset. The top associative attributes within each cluster were further examined for their distribution within and across the clusters, as well as their cluster-to-dataset percentage ratio. Among the six clusters, the first two clusters, representing 66% of total PIHR crashes, primarily highlighted PIHR crashes on city streets, with most of them occurring in the early night (7 to 11 p.m.) in Cluster 1 and afternoon (12 noon to 4 p.m.) in Cluster 2. Clusters 3 and 4, accounting for 30% of PIHR crashes, predominantly exhibited PIHR crashes on U.S. and state highways. Clusters 5 and 6 focused on high-speed highways, specifically interstates involving pedestrian fatalities, in addition to the pedestrian fatalities presented in Cluster 4, which were mainly concentrated on state highways and associated with early morning (4 to 6 a.m.). Recognizing these patterns and associations in different clusters is crucial, as it points to the need for implementing targeted countermeasures to not only reduce pedestrian crashes in general, but specifically those involving hit-and-run incidents. Future studies in Louisiana could analyze direct links between individual attributes to PIHR crash severity by considering pedestrian volume. Although the study suggests several context-based countermeasures, further investigation into the safety benefits and feasibility of applying these countermeasures to local contexts is essential.

Acknowledgments

The authors express their gratitude to the Louisiana Department of Transportation and Development for providing the crash data. We acknowledge and appreciate the contributions of the two reviewers in enhancing the quality of our paper.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: M. Ashifur Rahman; data collection: M. Ashifur Rahman, A. Hossain; analysis and interpretation of results: M. Ashifur Rahman, S. Das, A. Hossain; draft manuscript preparation: M. Ashifur Rahman, S. Das, A. Hossain, J. Codjoe, E. Mitran, X. Sun. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

M. Ashifur Rahman https://orcid.org/0000-0001-6940-1599
Subasish Das https://orcid.org/0000-0002-1671-2753
Ahmed Hossain https://orcid.org/0000-0003-1566-3993
Julius Codjoe https://orcid.org/0000-0003-1958-8695
Elisabeta Mitran https://orcid.org/0000-0002-1778-9832
Xiaodaun Sun https://orcid.org/0000-0001-7282-1340

References

1. NHTSA. Fatality and Injury Reporting System Tool (FIRST). National Highway Traffic Safety Administration. https://cdan.dot.gov/query. Accessed September 30, 2021.
2. LADOTD. LADOTD Highway Crash List. Louisiana Department of Transportation and Development. http://www8.dotd.la.gov/crash1/home.aspx. Accessed October 1, 2021.
3. Aidoo, E. N., R. Amoh-Gyimah, and W. Ackaah. The Effect of Road and Environmental Characteristics on Pedestrian Hit-and-Run Accidents in Ghana. *Accident Analysis & Prevention*, Vol. 53, 2013, pp. 23–27. https://doi.org/10.1016/j.aap.2012.12.021.
4. Tay, R., S. M. Rifaat, and H. C. Chin. A Logistic Model of the Effects of Roadway, Environmental, Vehicle, Crash and Driver Characteristics on Hit-and-Run Crashes. *Accident Analysis & Prevention*, Vol. 40, No. 4, 2008, pp. 1330–1336. https://doi.org/10.1016/j.aap.2008.02.003.
5. Zhang, G., G. Li, T. Cai, D. M. Bishai, C. Wu, and Z. Chan. Factors Contributing to Hit-and-Run Crashes in China. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 23, 2014, pp. 113–124. https://doi.org/10.1016/j.trf.2013.12.009.
6. Tay, R., U. Barua, and L. Kattan. Factors Contributing to Hit-and-Run in Fatal Crashes. *Accident Analysis & Prevention*, Vol. 41, No. 2, 2009, pp. 227–233. https://doi.org/10.1016/j.aap.2008.11.002.
7. Das, S., S. Ashraf, A. Dutta, and L.-N. Tran. Pedestrians Under Influence (PUI) Crashes: Patterns from Correspondence Regression Analysis. *Journal of Safety Research*, Vol. 75, 2020, pp. 14–23. https://doi.org/10.1016/j.jsr.2020.07.001.
8. USODT. Safer People. Safe System Approach. National Roadway Safety Strategy. U.S. Department of Transporation. https://www.transportation.gov/NRSS/SaferPeople. Accessed March 18, 2023.
9. Belin, M.-Å., A. Hartmann, M. Svolsbru, B. Turner, and M. S. Griffith. Applying a Safe System Approach across the Globe. *Public Roads*, Vol. 85, No. 4, 2022, pp. 36–42.
10. DZD Louisiana DOTD. *Louisiana Strategic Highway Safety Plan*. Louisiana Department of Transportation and Development, Baton Rouge, 2022.
11. Kong, X., S. Das, Y. Zhang, and X. Xiao. Lessons Learned from Pedestrian-Driver Communication and Yielding Patterns. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 79, 2021, pp. 35–48. https://doi.org/10.1016/j.trf.2021.03.011.
12. Roshandeh, A. M., B. Zhou, and A. Behnood. Comparison of Contributing Factors in Hit-and-Run Crashes with Distracted and Non-Distracted Drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 38, 2016, pp. 22–28. https://doi.org/10.1016/j.trf.2015.12.016.
13. Solnick, S. J., and D. Hemenway. Hit the Bottle and Run: The Role of Alcohol in Hit-and-Run Pedestrian Fatalities. *Journal of Studies on Alcohol*, Vol. 55, No. 6, 1994, pp. 679–684. https://doi.org/10.15288/jsa.1994.55.679.
14. MacLeod, K. E., J. B. Griswold, L. S. Arnold, and D. R. Ragland. Factors Associated with Hit-and-Run Pedestrian Fatalities and Driver Identification. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 366–372. https://doi.org/10.1016/j.aap.2011.08.001.
15. Benson, A. J., B. C. Tefft, L. S. Arnold, and W. J. Horrey. Fatal Hit-and-Run Crashes: Factors Associated with Leaving the Scene. *Journal of Safety Research*, Vol. 79, 2021, pp. 76–82.
16. Zhou, B., A. M. Roshandeh, S. Zhang, and Z. Ma. Analysis of Factors Contributing to Hit-and-Run Crashes Involved with Improper Driving Behaviors. *Procedia Engineering*, Vol. 137, 2016, pp. 554–562.
17. Das, S., X. Kong, and I. Tsapakis. Hit and Run Crash Analysis Using Association Rules Mining. *Journal of Transportation Safety & Security*, Vol. 13, No. 2, 2021, pp. 123–142. https://doi.org/10.1080/19439962.2019.1611682.
18. Sivasankaran, S. K., and V. Balasubramanian. Investigation of Factors Contributing to Pedestrian Hit-and-Run Crashes in India. *Journal of Transportation Safety & Security*, Vol. 14, No. 3, 2022, pp. 382–403. https://doi.org/10.1080/19439962.2020.1781313.
19. Grembek, O., and J. B. Griswold. *On the Legal Deterrence of Pedestrian Hit-and-Run Collisions*. Working Paper. University of California, Berkley, Safe Transportation Research and Education Center, 2012.

20. Kim, K., P. Pant, and E. Y. Yamashita. Hit-and-Run Crashes: Use of Rough Set Analysis with Logistic Regression to Capture Critical Attributes and Determinants. *Transportation Research Record: Journal of the Transportation Research Board*, 2008. 2083: 114–121.
21. Das, S., A. Dutta, X. Kong, and X. Sun. Hit and Run Crashes: Knowledge Extraction from Bicycle Involved Crashes Using First and Frugal Tree. *International Journal of Transportation Science and Technology*, Vol. 8, No. 2, 2019, pp. 146–160. https://doi.org/10.1016/j.ijtst.2018.11.001.
22. Jha, A. N., N. Chatterjee, and G. Tiwari. A Performance Analysis of Prediction Techniques for Impacting Vehicles in Hit-and-Run Road Accidents. *Accident Analysis & Prevention*, Vol. 157, 2021, P. 106164. https://doi.org/10.1016/j.aap.2021.106164.
23. Benson, A. J., L. S. Arnold, B. C. Tefft, and A. J. Horrey. *Hit-and-Run Crashes: Prevalence, Contributing Factors, and Countermeasures (Research Brief)*. AAA Foundation for Traffic Safety, Washington, D.C., 2017.
24. Hitosugi, M., E. Kagesawa, T. Narikawa, M. Nakamura, M. Koh, and S. Hattori. Hit-and-Runs More Common with Pedestrians Lying on the Road: Analysis of a Nationwide Database in Japan. *Chinese Journal of Traumatology*, Vol. 24, No. 2, 2021, pp. 83–87. https://doi.org/10.1016/j.cjtee.2020.11.008.
25. Solnick, S. J., and D. Hemenway. The Hit-and-Run in Fatal Pedestrian Accidents: Victims, Circumstances and Drivers. *Accident Analysis & Prevention*, Vol. 27, No. 5, 1995, pp. 643–649. https://doi.org/10.1016/0001-4575(95)00012-O.
26. Jha, A. N., A. Kumar, G. Tiwari, and N. Chatterjee. Identification and Analysis of Offenders Causing Hit and Run Accidents Using Classification Algorithms. *International Journal of Injury Control and Safety Promotion*, Vol. 29, No. 3, 2022, pp. 360–371. https://doi.org/10.1080/17457300.2022.2040541.
27. Zhu, S., and J. Wan. Cost-Sensitive Learning for Semi-Supervised Hit-and-Run Analysis. *Accident Analysis & Prevention*, Vol. 158, 2021, p. 106199. https://doi.org/10.1016/j.aap.2021.106199.
28. Fujita, G., K. Okamura, M. Kihira, and R. Kosuge. Factors Contributing to Driver Choice After Hitting a Pedestrian in Japan. *Accident Analysis & Prevention*, Vol. 72, 2014, pp. 277–286. https://doi.org/10.1016/j.aap.2014.07.002.
29. LHSC. *Manual for Use of the Uniform Vehicle Traffic Crash Report*. Louisiana Highway Safety Commission, Baton Rouge, 2019.
30. Sun, M., X. Sun, and D. Shan. Pedestrian Crash Analysis with Latent Class Clustering Method. *Accident Analysis and Prevention*, Vol. 124, 2019, pp. 50–57. https://doi.org/10.1016/j.aap.2018.12.016.
31. Hossain, A., X. Sun, R. Thapa, and J. Codjoe. Applying Association Rules Mining to Investigate Pedestrian Fatal and Injury Crash Patterns Under Different Lighting Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, 2022. 2676: 659–672.
32. Das, S., and X. Sun. Association Knowledge for Fatal Run-off-Road Crashes by Multiple Correspondence Analysis. *IATSS Research*, Vol. 39, No. 2, 2016, pp. 146–155. https://doi.org/10.1016/j.iatssr.2015.07.001.
33. Jalayer, M., M. Pour-Rouholamin, and H. Zhou. Wrong-Way Driving Crashes: A Multiple Correspondence Approach to Identify Contributing Factors. *Traffic Injury Prevention*, Vol. 19, No. 1, 2018, pp. 35–41. https://doi.org/10.1080/15389588.2017.1347260.
34. Das, S., A. Dutta, M. Jalayer, A. Bibeka, and L. Wu. 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*, Vol. 7, No. 2, 2018, pp. 114–123. https://doi.org/10.1016/j.ijtst.2018.02.001.
35. Baireddy, R., H. Zhou, and M. Jalayer. Multiple Correspondence Analysis of Pedestrian Crashes in Rural Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 116–127.
36. Das, S., and X. Sun. Factor Association with Multiple Correspondence Analysis in Vehicle-Pedestrian Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2519: 95–103.
37. Hossain, M. M., M. A. Rahman, X. Sun, and E. Mitran. Investigating Underage Alcohol-Intoxicated Driver Crash Patterns in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 769–782.
38. Rahman, M. A., S. Das, J. Codjoe, E. Mitran, X. Sun, K. Abedi, and M. M. Hossain. Applying Data Mining Methods to Explore Animal-Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2023. 2677: 665–681.
39. van de Velden, M., A. I. D'Enza, and F. Palumbo. Cluster Correspondence Analysis. *Psychometrika*, Vol. 82, No. 1, 2017, pp. 158–185. https://doi.org/10.1007/s11336-016-9514-0.
40. Das, S., X. Sun, B. Dadashova, M. A. Rahman, and M. Sun. Sun Glare and Traffic Crashes: Identifying Patterns of Key Factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2676: 165–175.
41. Ashifur Rahman, M., S. Das, and X. Sun. Using Cluster Correspondence Analysis to Explore Rainy Weather Crashes in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board*, 2022. 2676: 159–173.
42. Das, S., A. Dutta, M. A. Rahman, and X. Sun. Pattern Recognition from Light Delivery Vehicle Crash Characteristics. *Journal of Transportation Safety & Security*, Vol. 14, No. 12, 2022, pp. 2055–2073. https://doi.org/10.1080/19439962.2021.1995800.
43. Das, S., L.-N. Tran, and M. Theel. Understanding Patterns in Marijuana Impaired Traffic Crashes. *Journal of Substance Use*, Vol. 26, No. 1, 2021, pp. 21–29. https://doi.org/10.1080/14659891.2020.1760381.
44. Das, S., M. M. Hossain, M. Ashifur Rahman, X. Kong, X. Sun, and G. M. Al Mamun. Understanding Patterns of Moped and Seated Motor Scooter (50 Cc or Less) Involved

Fatal Crashes Using Cluster Correspondence Analysis. *Transportmetrica A: Transport Science*, Vol. 19, No. 2, 2023, p. 2029613. https://doi.org/10.1080/23249935.2022.2029613.
45. Markos, A., A. I. D'Enza, and M. van de Velden. Beyond Tandem Analysis: Joint Dimension Reduction and Clustering in R. *Journal of Statistical Software*, Vol. 91, 2019, pp. 1–24. https://doi.org/10.18637/jss.v091.i10.
46. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.
47. Thorndike, R. Who Belongs in the Family? *Psychometrika*, Vol. 18, No. 4, 1953, pp. 267–276.
48. Everitt, B. S., S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. John Wiley & Sons, Chichester, 2011.
49. Jain, A. K. Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*, Vol. 31, No. 8, 2010, pp. 651–666. https://doi.org/10.1016/j.patrec.2009.09.011.
50. Hennig, C. Cluster-Wise Assessment of Cluster Stability. *Computational Statistics & Data Analysis*, Vol. 52, No. 1, 2007, pp. 258–271.
51. Park, Y., and M. Garcia. Pedestrian Safety Perception and Urban Street Settings. *International Journal of Sustainable Transportation*, Vol. 14, No. 11, 2020, pp. 860–871. https://doi.org/10.1080/15568318.2019.1641577.
52. Chen, L., I. Grimstead, D. Bell, J. Karanka, L. Dimond, P. James, L. Smith, and A. Edwardes. Estimating Vehicle and Pedestrian Activity from Town and City Traffic Cameras. *Sensors*, Vol. 21, No. 13, 2021, p. 4564. https://doi.org/10.3390/s21134564.
53. Tolford, T. M., M. Izadi, C. Ash, and J. Codjoe. *Pedestrians and Bicyclists Count: Developing a Statewide Multimodal Count Program*. Publication FHWA/LA.17/599. Louisiana Department of Transportation and Development, Baton Rouge, 2019, p. 132.
54. Tay, R. A Random Parameters Probit Model of Urban and Rural Intersection Crashes. *Accident Analysis & Prevention*, Vol. 84, 2015, pp. 38–40. https://doi.org/10.1016/j.aap.2015.07.013.
55. Tefft, B. C. Impact Speed and a Pedestrian's Risk of Severe Injury or Death. *Accident Analysis & Prevention*, Vol. 50, 2013, pp. 871–878. https://doi.org/10.1016/j.aap.2012.07.022.
56. Rosén, E., and U. Sander. Pedestrian Fatality Risk as a Function of Car Impact Speed. *Accident Analysis & Prevention*, Vol. 41, No. 3, 2009, pp. 536–542. https://doi.org/10.1016/j.aap.2009.02.002.
57. Liu, J., A. J. Khattak, C. Chen, D. Wan, J. Ma, and J. Hu. Revisiting Hit-and-Run Crashes: A Geo-Spatial Modeling Method. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 81–92.
58. Wang, J., and J. B. Cicchino. Fatal Pedestrian Crashes on Interstates and Other Freeways in the United States. *Journal of Safety Research*, Vol. 74, 2020, pp. 1–7. https://doi.org/10.1016/j.jsr.2020.04.009.
59. Sandt, L., L. Thomas, K. Langford, and D. Nabors. *A Resident's Guide for Creating Safer Communities for Walking and Biking*. Publication FHWA-SA-14-099. Federal Highway Administration, Washington, D.C., 2015.
60. Thapa, R., X. Sun, A. Hossain, and S. B. Mavra. *Reduce Pedestrian Fatal Crashes in Louisiana by Improving Lighting Conditions*. Publication FHWA/LA.17/674. Louisiana Department of Transportation and Development, Baton Rouge, 2022.
61. National Highway Traffic Safety Administration. Dashboard of State by State Pedestrian Fatalities: Louisiana - 2012-2021. National Roadway Safety Strategy (NRSS) Action Tracking Dashboard. https://explore.dot.gov/views/DV_FARS_PD/Home?%3Aembed=y&%3AisGuestRedirectFromVizportal=y%3AshowAppBanner&%3Adisplay_count=n&%3AshowVizHome=n&%3Aorigin=viz_share_link&%3Atoolbar=no&%3A%3Aembed=yes. Accessed April 30, 2023.
62. Venkatraman, V., C. M. Richard, K. Magee, and K. Johnson. *Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices, 10th Edition, 2020*. Publication DOT HS 813 097. National Highway Traffic Safety Administration, Washington, D.C., 2021.
63. Zegeer, C. V., J. R. Stewart, H. H. Huang, P. A. Lagerwey, J. Feaganes, and B. J. Campbell. *Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations: Final Report and Recommended Guidelines*. Publication FHWA-HRT-04-100. Federal Highway Administration, Washington, D.C., 2005.
64. Monsere, C., M. Figliozzi, S. Kothuri, A. Razmpa, and D. Hazel. *Safety Effectiveness of Pedestrian Crossing Enhancements*. Oregon Department of Transportation, Salem, 2017.
65. Zajac, S. S., and J. N. Ivan. Factors Influencing Injury Severity of Motor Vehicle–Crossing Pedestrian Crashes in Rural Connecticut. *Accident Analysis & Prevention*, Vol. 35, No. 3, 2003, pp. 369–379. https://doi.org/10.1016/S0001-4575(02)00013-1.
66. Johnson, C. D. Pedestrian Fatalities on Interstate Highways: Characteristics and Countermeasures. *Transportation Research Record: Journal of the Transportation Research Board*, 1997. 1578: 23–29.
67. Louisiana Department of Transportation and Development. Motorist Assistance Patrol. DOTD Operations: Motorist Assistance Patrol (MAP). http://wwwsp.dotd.la.gov/Inside_LaDOTD/Divisions/Operations/MAP/Pages/default.aspx. Accessed May 18, 2023.
68. Hopkins, M., S. Chivers, and G. Stevenson-Freer. *Hit-and-Run: Why Do Drivers Fail to Stop after an Accident? Contexts of Incidents, Driver Motivations and Preventative Strategies*. University of Leicester, 2017.
69. Florida Department of Highway Safety and Motor Vehicles. Stay at the Scene. Driving Safety: Hit-and-Run Awareness. https://www.flhsmv.gov/safety-center/driving-safety/hitrun/. Accessed May 18, 2023.