Identifying and Analyzing Pass-by Crashes for the Purpose of Designing Proper Intervention Measures to Mitigate Crashes Involving Rural Population

A Technical Report Submitted to the Rural Safe Efficient Advanced Transportation (R-SEAT) Center and United States Department of Transportation

FINAL REPORT

Principal Investigator:
O. Arda Vanli, Ph.D.
Professor
Department of Industrial & Manufacturing Engineering,
FAMU-FSU College of Engineering
2525 Pottsdamer Street
Tallahassee, FL 32310, USA
Phone: +1 (850) 410-6354
E-mail: oavanli@eng.famu.fsu.edu

Co-Principal Investigator:
Ren Moses, Ph.D., PE
Professor
Department of Civil and Environmental Engineering
Florida A&M University
2525 Pottsdamer Street, Suite 129
Tallahassee, FL 32310, USA
Phone: +1 (850) 410-6191
Email: ren.moses@famu.edu

Research Assistant:
Seyed Reza Abazari
Doctoral Student
Department of Industrial & Manufacturing Engineering,
FAMU-FSU College of Engineering
2525 Pottsdamer Street
Tallahassee, FL 32310, USA
E-mail: sabazari@fsu.edu

April 2026

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under the grant 69A3552348321 from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Identifying and Analyzing Pass-by Crashes for the Purpose of Designing Proper Intervention Measures to Mitigate Crashes Involving Rural Population</td><td>5. Report Date
04/01/2026</td></tr>
<tr><td>6. Performing Organization Code
59-1961248</td></tr>
<tr><td colspan="2">7. Author(s)
O. Arda Vanli, Ph.D., SeyedReza Abazari, Ren Moses, Ph.D.</td><td>8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Florida A&M University-Florida State University
2525 Pottsdamer Street
Tallahassee, FL 32310-6046, USA</td><td>10. Work Unit No. (TRAIS)</td></tr>
<tr><td>11. Contract or Grant No.
69A3552348321</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Rural Safe Efficient Advanced Transportation (R-SEAT)
2525 Pottsdamer Street
Tallahassee, FL 32310</td><td>13. Type of Report and Period Covered
Final Report
Period Covered: 01/01/2025-12/31/2025</td></tr>
<tr><td>14. Sponsoring Agency Code</td></tr>
<tr><td colspan="3">15. Supplementary Notes</td></tr>
</table>

16. Abstract

Rural areas in the United States are home to only 19% of the population, however, over 70% of the nation's 4 million miles of roadways are located in rural regions and the traffic fatality rates are significantly higher in these areas, about 1.5 times higher, than in urban areas nationwide. Pass-by crashes, incidents involving drivers far from their origin, have a notable share of rural crashes, particularly in the tourist destinations or during special events.
This research focuses on analyzing pass-by crashes in rural areas, particularly in FDOT District 3 (Northwest Florida) and and District 5 (Central Florida). The primary goal is to identify trends and factors contributing to the frequency crashes and propose interventions aimed at improving transportation safety for rural populations. By focusing on rural transportation, the study aligns with the broader objective of promoting equity and safety in regions that are often underserved in terms of infrastructure and access to transportation resources.
The project will explore innovative machine learning and statistical methods to analyze the complex interactions between drivers' social characteristics and roadway features that influence the frequency and severity of rural pass-by crashes. Specifically, the project examines the causal effects of special events on the likelihood of pass-by crashes in rural Florida, with particular attention to spatial and temporal variations across the two Florida Department of Transportation districts over 2021 and 2022. Unlike prior research that often overlooks rural contexts or treats special events as isolated factors, this work integrates crash records, roadway attributes, and socioeconomic characteristics to capture how event-driven travel behaviors interact with local conditions. Logistic regression models, combined with group lasso feature selection, were employed to isolate key factors influencing pass-by crash risks. A novel contribution of this study is the identification of county-level heterogeneity in how special events impact the likelihood of rural pass-by crashes, demonstrating that such effects are neither uniform across space nor consistent over time. The findings offer actionable insights for transportation planners by highlighting specific counties and roadway types where targeted interventions could mitigate the elevated risks associated with event-related traffic

<table>
<tr><td colspan="2">17. Key Words
Regularized Logistic Regression, Pass-by Crash, Rural Transportation Safety</td><td colspan="2">18. Distribution Statement
No restrictions</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
28</td><td>22. Price</td></tr>
</table>

ACKNOWLEDGEMENTS

This project was sponsored by the Rural Safe Efficient Advanced Transportation (R-SEAT) Center and United States Department of Transportation. The Principal Investigators would like to thank the representatives of the REAT Center for their valuable feedback throughout the project activities.

iv

EXECUTIVE SUMMARY

The United States Census Bureau reports that rural areas cover about 97% of the nation's land area and are a home to about 60 million people. About 19% of the American population lives in the rural area according to the Census Bureau. Although only 19% of the population lives in rural areas more than 70% of the 4 million miles of roadways in the United States are in rural areas. According to the NHTSA (2021) the fatality rate was 1.5 times higher in rural areas than in urban areas of the US. In Florida, the fatality rate per 100 million VMT (Vehicles Miles Travelled) in rural areas and urban areas were 2.06 and 1.64, respectively, giving a rural to urban fatality rate ratio of about 1.3. Similar to many other states of the country, on average, the road travel in rural areas in Florida is dangerous and riskier with respect to fatal accidents.

Pass-by crashes, defined as incidents involving drivers who are far from their point of origin, constitute a significant component of rural crashes, especially in touristic destinations and during special events. Unlike crashes involving residents or regular commuters, pass-by crashes occur among a transient population whose exposure to local road conditions is limited and often unpredictable. This research focuses on analyzing pass-by crashes in rural areas, particularly in FDOT District 3 (Northwest Florida) and District 5 (Central Florida). The primary goal is to identify trends and factors contributing to these crashes and propose interventions aimed at improving transportation safety for rural populations. By focusing on rural transportation, the study aligns with the broader objective of promoting equity and safety in regions that are often underserved in terms of infrastructure and access to transportation resources.

The project will explore innovative machine learning and statistical modeling methods to analyze the complex interactions between drivers' social characteristics and roadway features that influence the frequency and severity of rural pass-by crashes. Specifically, the project examines the causal effects of special events on the likelihood of pass-by crashes in rural Florida, with particular attention to spatial and temporal variations across the two Florida Department of Transportation districts over 2021 and 2022. Unlike prior research that often overlooks rural contexts or treats special events as isolated factors, this work integrates crash records, roadway attributes, and socioeconomic characteristics to capture how event-driven travel behaviors interact with local conditions. Logistic regression models, combined with group lasso feature selection, were employed to isolate key factors influencing pass-by crash risks. A novel contribution of this study is the identification of county-level heterogeneity in how special events impact the likelihood of rural pass-by crashes, demonstrating that such effects are neither uniform across space nor consistent over time. The findings offer actionable insights for transportation planners by highlighting specific counties and roadway types where targeted interventions could mitigate the elevated risks associated with event-related traffic

Data needed to train the models were sourced from the Florida Traffic Safety Dashboard, FDOT GIS Open Data Hub and US Census, focusing on crash events, roadway characteristics and driver demographics. To classify pass-by crashes, distances between crash locations and the drivers' home ZIP codes were calculated, with a threshold of 30 miles used to define a pass-by crash. Logistic regression and Random Forest models were used to analyze the factors influencing these crashes, with variables such as functional class, weather conditions, vision obstruction, and type of shoulder playing significant roles in predicting crash likelihood.

From the 2022-2026 USDOT RD&T strategic plan, the project predominantly addresses Safety as the strategic goal, and Data-Driven System Safety as the research priority. Specifically, within this priority, the tools and outcomes of this project will advance the knowledge and state of the art to help target the objectives to (1) identify and support strategies to increase user safety in rural roads (e.g., pedestrians, bicyclists, motorcyclists, and people with disabilities) and (2) develop and promote effective methods to assess and address traffic safety risks in rural communities.

Contents

List of Figures

List of Tables

1. INTRODUCTION

Traffic crashes continue to pose serious public health and road safety challenges, resulting in significant numbers of deaths, injuries, and financial burdens globally [1]. The scale of this problem is considerable, with many lives affected each year through direct injury, long-term disability, and the broader socioeconomic consequences borne by individuals, families, and healthcare systems. Even with ongoing improvements in automotive safety systems, road infrastructure, and regulatory measures, collision rates remain stubbornly elevated, especially in areas with difficult driving environments [2], [3]. This persistence suggests that existing countermeasures, while beneficial, have not been sufficient to fully address the underlying complexity of crash causation. Crashes arise from a complex interplay of elements, including driver behavior, road geometry, weather and environmental conditions, and vehicle-related factors [4], [5]. The interaction among these factors makes crash prediction and prevention particularly challenging, as no single intervention can adequately capture the full spectrum of contributing causes. Of these, human-related factors such as inattention, intoxication, and drowsiness are the most prominent contributors, collectively accounting for a substantial proportion of all recorded incidents. While crashes in urban areas are commonly linked to traffic congestion and high vehicle density, those occurring in rural areas, though comparatively less frequent, often produce more severe consequences due to faster travel speeds, underdeveloped infrastructure, and slower emergency response times. The disparity in crash outcomes between urban and rural environments underscores the need for context-sensitive approaches to safety analysis and intervention.

Rural road environments possess unique attributes that heighten crash severity relative to urban counterparts [6], [7]. Unlike urban corridors, which typically benefit from controlled intersections, street lighting, and dense surveillance infrastructure, rural roads are often characterized by extended stretches with minimal safety features and limited oversight. Characteristics such as unrestricted access points, inadequate lighting, narrow travel lanes, and tight horizontal curves add to the complexity of driving in these settings. These geometric and design features demand higher levels of driver attentiveness and skill, leaving less room for error. Beyond road geometry, unpredictable environmental factors including animal crossings and inconsistent surface conditions compound the associated risks. Seasonal variations, such as ice formation, fog, and flooding, can further degrade road conditions in ways that are difficult to anticipate, particularly for drivers who are not regularly exposed to such environments. While frequent local travelers may become accustomed to these conditions over time, developing adaptive driving behaviors and intuitive familiarity with known hazards, drivers passing through rural areas face added difficulties stemming from unfamiliarity with the road layout, appropriate speed levels, and location-specific hazards. This asymmetry in road familiarity between local and non-local drivers is a critical yet often overlooked dimension of rural crash risk.

Within this framework, pass-by crashes, defined as incidents involving drivers who are far from their point of origin, represent a notable yet insufficiently explored category of traffic incidents. Unlike crashes involving residents or regular commuters, pass-by crashes occur among a transient population whose exposure to local road conditions is limited and often unpredictable. These events are especially significant in rural settings, where long-distance travel and unfamiliar road conditions are prevalent. The combination of high-speed travel, infrequent rest opportunities, and limited access to real-time local information creates conditions that are particularly conducive to crash occurrence among passing drivers. Drivers who are not local to an area may depend excessively on GPS navigation, misread roadway indicators, or be unprepared for region-specific dangers. In many cases, navigation systems may fail to convey the nuanced characteristics of rural roads, such as sharp bends, unmarked intersections, or areas prone to wildlife activity, leaving drivers without adequate preparation for the conditions they encounter [8]. Furthermore, extended periods of driving can heighten fatigue and diminish situational awareness, thereby increasing the likelihood of a crash. The cumulative effect of prolonged travel, reduced alertness, and environmental unfamiliarity creates a compounding risk profile that distinguishes pass-by crashes from other crash types and warrants dedicated analytical attention.

1.1. Role of Special Events on Pass-by Crashes

One significant yet inadequately examined factor contributing to variability in pass-by crash risk is the presence of special events, including sporting competitions, public holidays, and community festivals. Such events are distinctive in their capacity to generate sudden, concentrated surges in travel demand that differ fundamentally from routine daily traffic patterns, often resulting in disruptions to normal traffic flow and increased congestion [9], [11]. Unlike predictable commuter flows or seasonal travel trends, event-driven traffic is often characterized by its temporal concentration, geographic directionality, and the high proportion of non-local travelers it introduces to unfamiliar road networks. Such events can temporarily modify both the volume and composition of traffic by drawing large numbers of out-of-area travelers to particular locations [9], [11]. This shift in traffic composition is particularly consequential in rural areas, where road networks are typically designed to accommodate local travel demands and may not be well-equipped to handle the sudden influx of unfamiliar drivers. Consequently, drivers unfamiliar with the area are more likely to encounter roadway environments that differ from what they are accustomed to or have previously experienced. The resulting mismatch between driver expectations and actual road conditions can cause delayed reaction times, inappropriate speed choices, and misjudgments at intersections or curves, all of which elevate the probability of a crash occurring.

These event-driven changes introduce multiple mechanisms that may increase crash risk. First, the influx of non-local drivers can lead to greater heterogeneity in driving behavior, as drivers differ in their familiarity with local traffic norms, roadway geometry, and speed expectations. This behavioral diversity can disrupt the flow of traffic and create unpredictable interactions between drivers operating under different assumptions about road use, right-of-way, and safe following distances. Second, increased reliance on navigation systems may cause abrupt maneuvers, late lane changes, or hesitation near intersections and decision points. In rural environments, where signage may be limited and decision points are often encountered at higher speeds, such navigation-induced behaviors can have particularly severe consequences. Third, special events often coincide with longer travel distances and irregular travel schedules, which can contribute to driver fatigue and reduced situational awareness. Drivers who have traveled extended distances to attend an event, or who are returning home late at night following its conclusion, may be operating under conditions of significant physical and cognitive impairment. Together, these factors can amplify the likelihood of crashes involving pass-by drivers, particularly in rural settings where roadway conditions are less forgiving and the margin for driver error is considerably narrower than in more controlled urban environments.

Prior empirical studies have also found that sporting events are associated with measurable increases in vehicular crashes, particularly in the periods surrounding event activity [10]. In addition, special events can place temporary stress on transportation infrastructure by increasing traffic demand beyond typical levels [9], [11]. Rural road networks, which are often designed with relatively low design-hour volumes in mind, may be particularly vulnerable to the operational disruptions caused by large-scale event traffic. This may lead to congestion in areas not designed to accommodate high volumes, limited availability of traffic control measures, and increased interaction between local and non-local drivers. In the absence of adequate traffic management resources, drivers may encounter unexpected queuing, spontaneous detours, or uncontrolled merging situations that further complicate navigation for those already unfamiliar with the area. Such conditions further elevate the complexity of the driving environment and may disproportionately affect drivers who lack familiarity with the area. The compounding effect of infrastructure stress and driver unfamiliarity creates a particularly hazardous set of conditions that warrants dedicated investigation, yet has received limited attention in the existing traffic safety literature, despite evidence that planned and large-scale special events significantly alter traffic demand, flow stability, and crash risk conditions [9]-[11].

1.2. Project Objectives

This project addresses a critical gap in the literature by examining how event-induced changes in traffic patterns influence the likelihood of crashes involving non-local drivers. Traffic safety research has long recognized the role of demand fluctuations in shaping crash risk, yet the specific mechanisms through which special events alter the driving environment for unfamiliar road users remain poorly understood. Special events can significantly alter both traffic demand and driver composition, potentially increasing the exposure of unfamiliar drivers to high-risk roadway environments, particularly in rural settings where limited infrastructure and high operating speeds leave little tolerance for navigational errors. However, the extent to which these changes causally affect pass-by crash occurrence remains insufficiently understood, and establishing a causal rather than merely associative relationship is essential for generating actionable evidence that can meaningfully inform safety planning decisions.

To address this gap, the primary objective of this study is to estimate the causal effect of special events on the likelihood of pass-by crash occurrence. This is achieved using a comprehensive dataset

that integrates crash records with information on event timing, weather conditions, roadway characteristics, and driver demographics. A logistic regression-based causal inference framework is employed to control for confounding factors and isolate the specific contribution of special events to crash risk, producing estimates that are both statistically rigorous and practically interpretable.
In addition to the primary objective, this study pursues the following secondary objectives:

- Identify key factors associated with pass-by crash occurrence, including roadway features, environmental conditions, and driver-related characteristics.
- Examine spatial and temporal heterogeneity in pass-by crash risk across counties and across different time periods.
- Evaluate how event-driven changes in traffic composition influence crash likelihood, particularly in rural regions with limited infrastructure.
- Provide evidence-based insights for targeted interventions, including event-specific traffic management strategies, improved signage, and safety awareness campaigns.

The findings of the project aim to support transportation agencies in developing more effective and context-specific safety policies. By identifying when and where pass-by crash risks are elevated, decision-makers can implement targeted interventions to reduce crash occurrence and improve roadway safety, particularly for rural populations most vulnerable to the consequences of event-driven traffic surges.

1.3. Report Structure

The remainder of this report is organized as follows. Section 2 presents a review of the relevant literature. Section 3 describes the study area and data sources used in the analysis. Section 4 outlines the methodological framework. Section 5 presents and discusses the results, and Section 6 provides concluding remarks and implications.

2. LITERATURE REVIEW

2.1. Crash Severity and Contributing Factors

Traffic crashes represent a leading cause of fatalities and injuries on a global scale. Existing research underscores the importance of pinpointing key risk factors to inform and guide prevention efforts. One investigation utilizing a hybrid ensemble machine-learning model identified crash type, injury severity, location, and time of occurrence as strong predictors of fatal outcomes [12]. Related work focusing on vulnerable road users, including pedestrians, cyclists, and motorcyclists, has demonstrated that environmental conditions, vehicle characteristics, and road user behavior are critical determinants of crash severity [13]. The complexities inherent in crashes involving pedestrians and cyclists further expose the limitations of existing crash databases in supporting accurate classification and pattern analysis [14].

Machine-learning techniques have also been broadly applied to injury severity prediction, with research demonstrating the effectiveness of feature selection methods in enhancing model performance [15]. Comparative evaluations of various machine-learning architectures, including deep learning approaches, indicate that model effectiveness varies as a function of data structure, computational demands, and prediction goals [16]. While these contributions advance understanding of crash severity and prediction, they are largely focused on general crash mechanisms and do not explicitly address differences attributable to driver familiarity or travel behavior patterns.

2.2. Driver Familiarity and Non-Local Driver Risk

Driver familiarity with road networks is a significant determinant of traffic safety, shaping risk perception, driving behavior, and decision-making processes. Research consistently reveals meaningful behavioral differences between local and non-local drivers, with route familiarity emerging as a central factor in crash risk [17]. For instance, non-local drivers have been found to be disproportionately involved in fatal wrong-way crashes under unlighted rural conditions, whereas local drivers are more commonly associated with alcohol impairment and nighttime urban driving incidents [18].

A study of freeway crashes in China found that while factors such as road gradient, temperature, and precipitation affected both local and non-local drivers in comparable ways, vehicle type and road curvature produced differential effects, with non-local drivers exhibiting greater difficulty navigating certain roadway geometries [19]. Conversely, research on urban vehicle-to-vehicle crash severity suggested that unfamiliar drivers may actually be associated with lower injury severity relative to their familiar counterparts [20]. A study of single-vehicle crashes on mountainous highways further identified distinct severity risk factors for unfamiliar versus familiar drivers, with unfamiliar drivers showing greater sensitivity to roadway lighting conditions, driver age, sharp curves, and rural terrain, while familiar drivers were more influenced by speed limits, impaired driving, and seatbelt use [21].

Research conducted along a U.S. highway corridor near national parks found that rising traffic volumes increase crash risk for non-local drivers while being associated with fewer crashes among local drivers, indicating that local drivers develop adaptive responses to higher traffic conditions over time [22]. Elevated crash risk among non-local drivers has similarly been observed in international settings, where long-term residents with sustained exposure to local road networks demonstrate lower crash risk compared to visiting travelers [23].

While unfamiliarity can elevate crash risk, excessive familiarity may also introduce safety concerns of its own. Drivers on well-known routes are more susceptible to mind wandering, tailgating, and slower responses to unexpected events, all of which can increase crash likelihood [24]. Naturalistic driving studies have further shown that distracted behaviors such as eating, drinking, and mobile phone use occur more frequently and persist for longer durations on familiar roads. Drivers on familiar routes are also more inclined to exceed speed limits and delay braking near intersections, suggesting that overconfidence and diminished attentiveness can independently elevate crash risk [25].

Despite these contributions, existing studies predominantly examine general familiarity effects and do not explicitly consider pass-by crashes or the safety implications of temporary disruptions such as special events.

2.3. Research Gap: Pass-by Crashes and Special Events

The body of literature reviewed demonstrates that road familiarity plays a meaningful role in crash risk, with non-local drivers facing distinct challenges related to unfamiliar road conditions, signage, and geometric features. Nevertheless, pass-by crashes as a distinct crash category have received limited dedicated attention, particularly in rural environments where roadway conditions are less accommodating and driver unfamiliarity may carry more serious safety consequences.

Furthermore, while prior studies have investigated crash risk factors and driver behavior in various contexts, comparatively little research has examined how temporary, event-driven shifts in traffic composition affect crash occurrence among non-local drivers. Special events have the potential to introduce abrupt increases in traffic demand and alter the proportion of unfamiliar drivers present on the road network, yet their causal impact on pass-by crash occurrence remains poorly characterized.

To address this gap, this study proposes a regularized logistic regression framework to investigate how special events influence pass-by crash likelihood while accounting for socioeconomic and roadway-related confounders. Drawing on crash data from two Florida transportation districts, the analysis offers insights into spatial and temporal variations in crash risk. The findings are intended to assist transportation planners in designing targeted intervention strategies aimed at enhancing safety along rural and high-speed roadway corridors.

3. STUDY AREA AND DATA SOURCES

3.1. Data Sources and Variable Construction

This study draws on data from multiple sources to investigate the relationship between special events and the probability of pass-by crash occurrence in rural areas of Florida during 2021 and 2022. Crash and driver information were sourced from the Signal 4 Analytics database [26], which provides detailed records of crash incidents and associated driver characteristics. Roadway attributes at crash locations were obtained from the Florida Department of Transportation's Roadway Characteristics Inventory (RCI) [27]. Socioeconomic and demographic data were gathered from the U.S. Census Bureau [28] and the American Community Survey (ACS) [29] at the census tract level and linked to drivers' residential ZIP codes to capture the characteristics of their home communities.

3.2. Study Area Description and Regional Characteristics

The study focuses on FDOT District 3 and District 5 [30], as illustrated in Figure 1. These two regions were selected to represent a range of environments, encompassing rural, semi-urban, and urban areas with varying population densities, traffic conditions, and special event activity. FDOT District 5 includes the Orlando metropolitan area, a major hub for tourism, entertainment, and higher education through the University of Central Florida, all of which regularly attract large numbers of out-of-region visitors. FDOT District 3, by contrast, is predominantly rural in character, encompassing smaller municipalities, major travel corridors such as Interstate 10, Gulf Coast beach destinations, and several universities including Florida State University, Florida A&M University, and the University of West Florida, each of which generates considerable non-local traffic during event periods.

It is important to note that while large urban centers such as Orlando inherently attract substantial visitor traffic, the primary focus of this research is on isolating the causal effect of special event periods on pass-by crashes in rural settings, with confounding influences addressed through the inclusion of appropriate control variables in the analytical framework. To further contextualize these two study regions, Table 1 presents key descriptive variables capturing demographic and economic differences between District 3 and District 5. Variables highlighted in bold, identified using the one standard deviation rule, appear to differ meaningfully between the two districts.

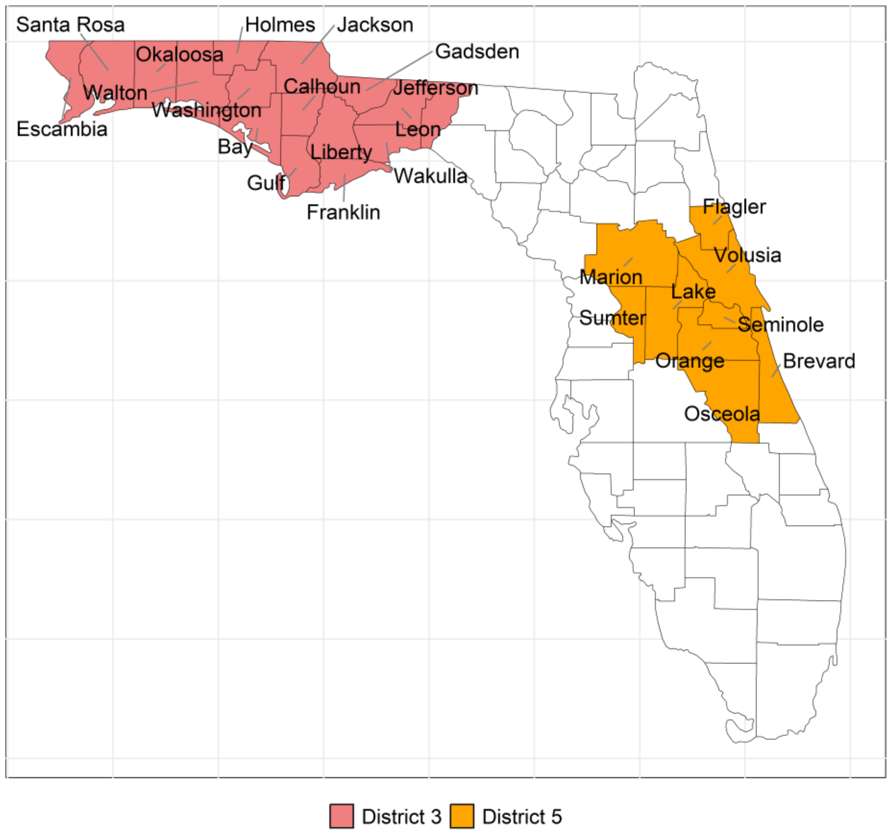


Figure 1: Study area

Table 1: Socioeconomic variables in FDOT Districts 3 and 5.

Variable	District 3	District 5
Percentage of College-Age Population	3.64 (0.49)	3.12(0.31)
Percentage Enrolled in College	23.03(1.07)	22.03(0.56)
Percentage of Households Without a Vehicle	**5.46(0.49)**	**4.44(0.24)**
Percentage of Renter-Occupied Housing	26.09(1.93)	27.06(2.98)
Average Commute Time (Minutes)	25.54(0.99)	24.88(0.91)
Percentage Employed in Service Occupations	**0.804(0.1)**	**0.972(0.11)**
Percentage Employed in Transportation Occupations	**1.95(0.12)**	**2.47(0.31)**
Median Household Income (USD)	**58,698(3181.0)**	**67,964(2277.6)**
Average Household Size	**2.44(0.03)**	**2.52(0.12)**
Poverty Rate	**14.7(1.22)**	**11.08(0.66)**
Employment Rate	**98.18(0.79)**	**99.74(0.07)**

The maps presented in Figure 2 highlight an important spatial distinction in the residential origins of drivers involved in pass-by crashes across FDOT Districts 3 and 5 over the two-year study period. The left panels, depicting crashes occurring in District 3, indicate that the majority of involved drivers originate from concentrated geographic areas, most notably the Orlando and Tampa metropolitan regions. This pattern suggests that a substantial share of pass-by crashes in District 3 involve drivers traveling from these urban centers into the more rural portions of the district. In contrast, the right panels, which focus on crashes occurring in District 5, reveal a closer correspondence between crash locations and drivers' places of residence. Most crashes in District 5 involve drivers from within or in close proximity to the district, as well as from the Miami metropolitan area, with relatively few originating from the Florida Panhandle. This spatial pattern remained consistent across both 2021 and 2022.

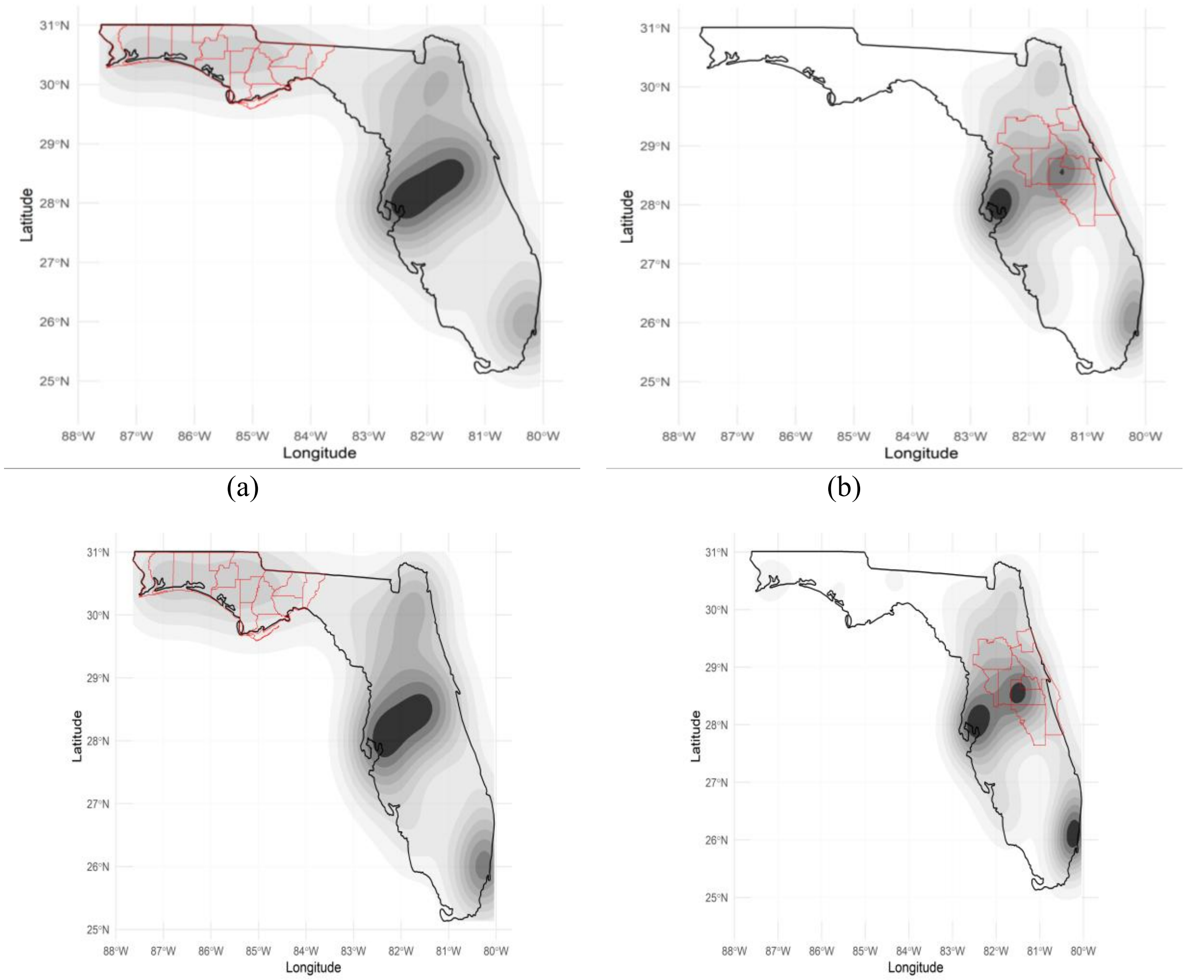


Figure 2: Drivers' residential ZIP code characteristics associated with rural pass-by crashes in (a) district 3, 2022; (b) district 5, 2022; (c) district 3, 2021; and (d) district 5 ,2021

4. METHODOLOGY

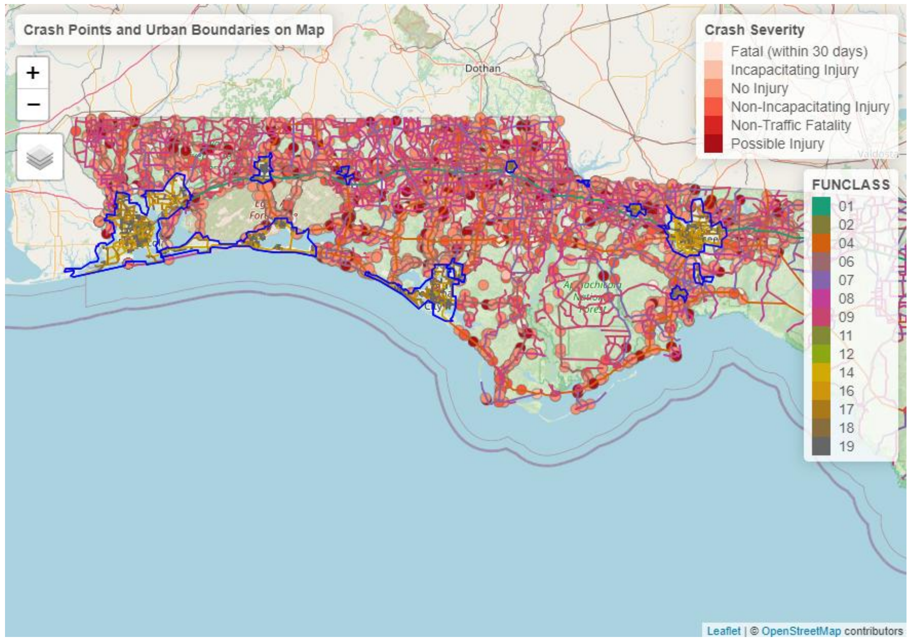
4.1. Data Preparation

Initial data preparation involved removing variables with excessive missing values and eliminating redundant variables that duplicated information already captured elsewhere in the dataset. For variables containing a large number of categorical levels, categories were consolidated to reduce dimensionality and enhance model interpretability. All data preparation, spatial joins, and statistical analyses, including logistic regression and correlation analysis procedures, were carried out using the R programming language [31].

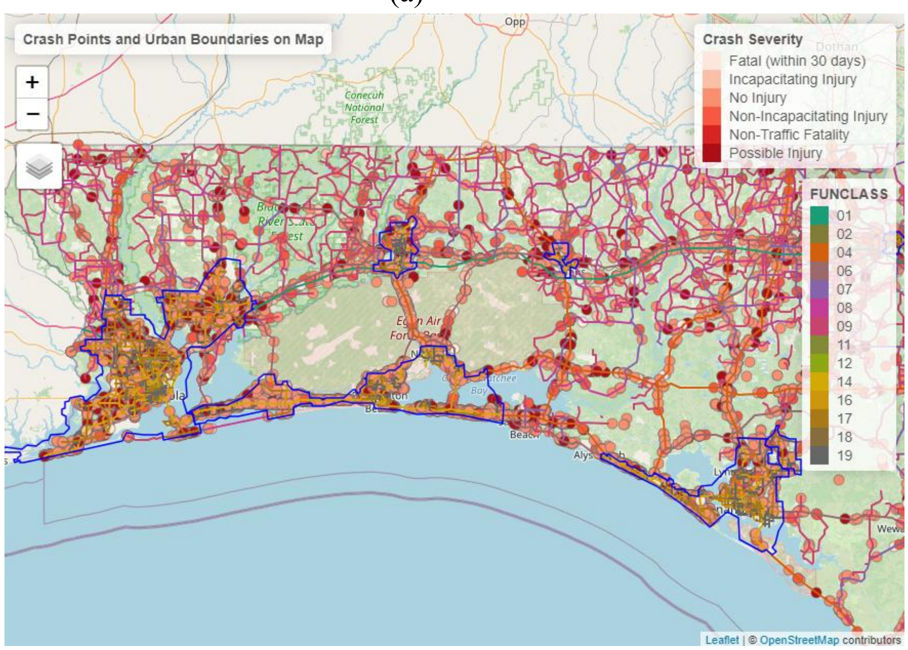
Following data cleaning, the dataset was refined to concentrate specifically on rural crash records. Urban crashes were identified and excluded through a spatial join performed using FDOT's defined urban boundary shapefiles [32], as illustrated in Figure 3, which displays both the full set of crashes and the resulting rural subset. After removing urban crashes, crash records were merged with driver information using the report number as a common linking key to produce a unified analytical dataset. Socioeconomic and demographic variables were subsequently joined to the driver data via ZIP codes, establishing a connection between each driver's place of residence and the corresponding census and ACS data. Roadway characteristics, including functional class, median type, and Annual Average Daily Traffic (AADT), were additionally incorporated through a spatial join to crash locations using geographic coordinates.

4.2. Identification of Special Events

To capture the potential influence of special events on pass-by crash likelihood, a binary special event indicator variable was constructed, assigned a value of 1 when the timing of a crash coincided with a special event period and 0 otherwise. U.S. federal holidays were identified, and a two-day buffer was applied around each holiday to account for extended travel periods associated with these occasions. Crashes falling within these buffered windows were coded as event period observations. In addition, dates of home football games hosted by major universities within the study region were incorporated as special event periods, though this designation was applied exclusively to crashes occurring within the counties in which those institutions are located.



(a)



(b)

Figure 3: Pass-by crashes (a) all crashes; (b) rural crashes

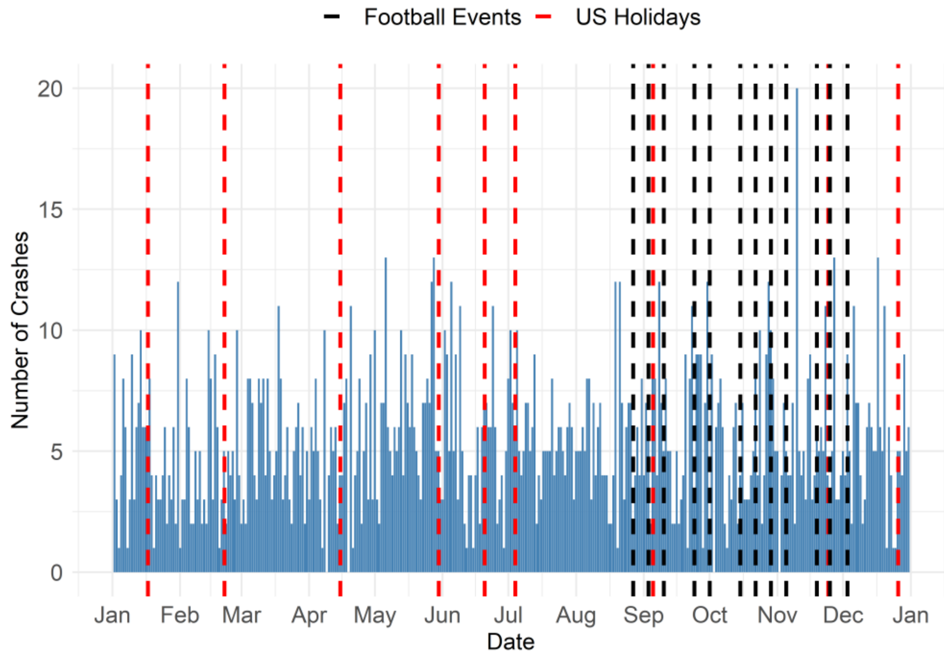


Figure 4: Aggregated daily pass-by crashes in District 3, 2022

4.3. Identification of Pass-by Crashes

A fundamental step in this analysis is the identification of pass-by crashes, defined as crashes involving drivers who were traveling considerably farther than what would be anticipated for routine local travel. A binary variable was constructed to distinguish pass-by crashes from non-pass-by crashes within the modeling framework. To generate this classification, the Haversine distance was calculated between each crash location and the centroid of the driver's residential ZIP code. Average commute times for each census tract were retrieved from the ACS [25] and converted into expected travel distances under the assumption of an average travel speed of 45 miles per hour. ZIP codes were linked to their corresponding census tracts using the HUD-USPS ZIP-tract crosswalk [33]. When the distance between a driver's residential ZIP code centroid and the crash location exceeded the expected travel distance derived from the area's average commute time, the crash was classified as a pass-by event and assigned a value of 1. Crashes that did not meet this threshold were classified as non-pass-by and assigned a value of 0.

4.4. Regularized Logistic Regression for Causal Effects

A regularized logistic regression model is proposed to estimate the causal effect of special events on the likelihood of pass-by crash occurrence:

$$\log \frac{\pi(D,X)}{1-\pi(D,X)} = \beta_0 + \beta_D D + \beta_1 X_1 + \cdots + \beta_K X_K$$

where $\pi(D,X) = \Pr(Y = 1|D,X)$ represents the probability that a given crash is a pass-by event, and $Y \in (0,1)$ is a binary variable indicating pass-by crash status. D denotes the treatment variable whose causal effect X is the primary object of estimation, represents the set of control variables, and β_D captures the causal effect of the treatment on crash likelihood. Within a causal inference framework, control variables are not themselves of direct interest; rather, they are included as regressors to hold constant factors that, if omitted, could introduce bias into the estimated causal effect of interest [34]. The special event indicator and its interaction with County are incorporated as treatment variables in D, while the control variables X encompass roadway characteristics, environmental conditions, crash-related factors, and demographic information. Figure 5 presents the correlation matrix of the 30 numerical variables contained in for District 5 in the year 2022. The discrete control variables include "Vision_Obstructed", "Alcohol_Involved", "Drug_Involved", "Driver_Distracted", "Speeding_Involved", "Road_Surface", "Junction_Flag", "Day_Night", "Lane_Departure", "Hit_and_Run", "Day_of_Week", "Month", "Crash_Condition", "Sex", "Shoulder_Type", "Light_Condition", "Time_of_Day", "Weather", "Crash_Severity", "County", "Functional_Class", "Median_Type", "Inside_Shoulder_Type", and "Outside_Shoulder_Type". The full dataset for this case comprised potential control variables.

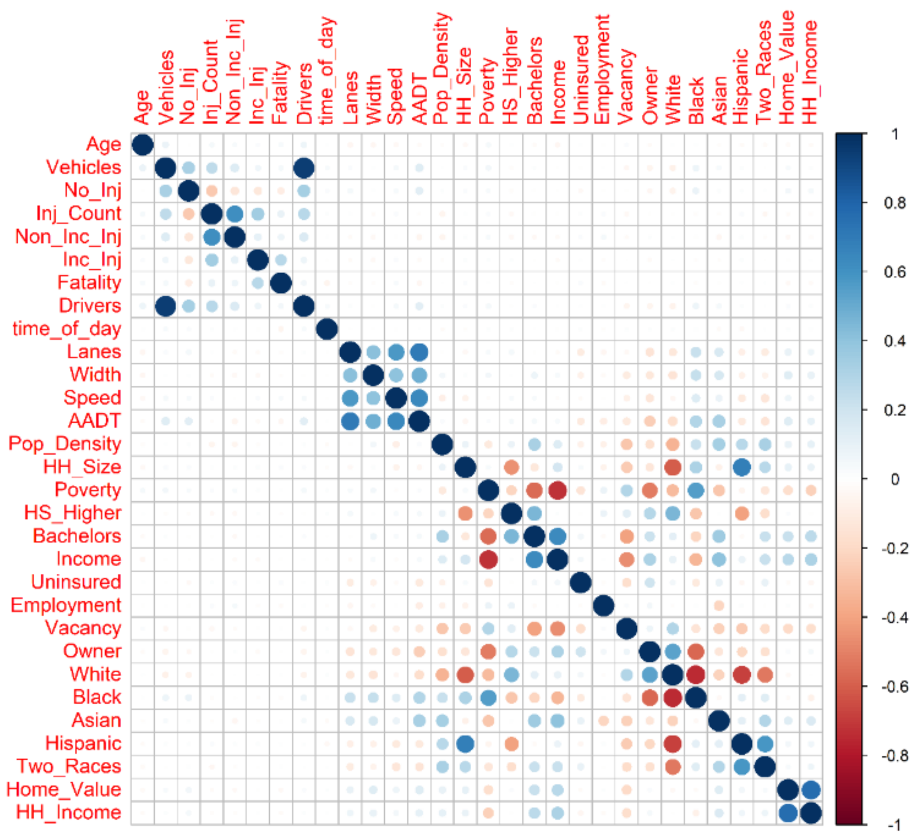


Figure 5: The correlation matrix for district 5 in 2022

Given the large number of candidate control variables, the analysis proceeded through a structured feature reduction process. First, a pairwise correlation matrix was constructed to identify and eliminate highly correlated variables, thereby reducing redundancy and minimizing the risk of multicollinearity. Group lasso regression was subsequently applied to perform further variable selection, enabling retention of the most influential predictors while improving model interpretability and generalizability. To illustrate this process, consider the correlation matrix presented in Figure 5 for District 5. Based on these results, two highly correlated variables—total number of vehicles involved and number of drivers involved—were excluded from the analysis to guard against multicollinearity, reducing the candidate variable set to 53. Group lasso regression was then applied to further refine the model [35]. Within the lasso framework, a penalized least squares fitting procedure estimates using values that minimize the penalized log-likelihood function:

$$LLB(Y) + \lambda \sum_{i}^{K} |\beta_i|$$

where $LLB(Y)$ represents the negative binomial log-likelihood of the observed data [35]. It is important to note that the treatment variable D is always included in the model and is therefore excluded from the penalty term.

Group lasso extends the conventional lasso framework by applying an L2 penalty to predefined groups of related variables, such as the dummy variables generated from categorical features, such that entire variable groups are retained or eliminated together rather than individual coefficients being penalized in isolation [35]. This approach is particularly well-suited for handling multilevel categorical predictors and helps maintain the interpretability of the resulting model. The group lasso penalty parameter (λ) was tuned using ten-fold cross-validation. As shown in the cross-validation error plot presented in Figure 6, the minimum cross-validation error was achieved at a particular value of $\lambda = 0.002$. However, to promote model parsimony and reduce the likelihood of overfitting, the one-standard-error rule was adopted as the parameter selection criterion. Under this rule, one standard error is added to the minimum cross-validation error, and the largest value of λ whose corresponding cross-validation error falls within this expanded threshold is selected as optimal. Application of the one-standard-error rule yielded $\lambda = 0.012$ as the preferred value of the penalty parameter, indicated by the vertical solid line in Figure 6, which resulted in the retention of 19 control variables. The following variables were consequently removed from the model: Vision_Obstructed, Alcohol_Involved, Drug_Involved, Driver_Distracted, Speeding_Involved, Road_Surface, Junction_Flag, Day_Night, Lane_Departure, Hit_and_Run, Day_of_Week, Month, Driver_Age, Minor_Injuries, Serious_Injuries, Fatalities, Time_of_Day, Road_Width, Household_Size, Education_Rate (HS+), Employment_Rate, Vacancy_Rate, Asian_Proportion, Hispanic_Proportion, and Multiracial_Population.

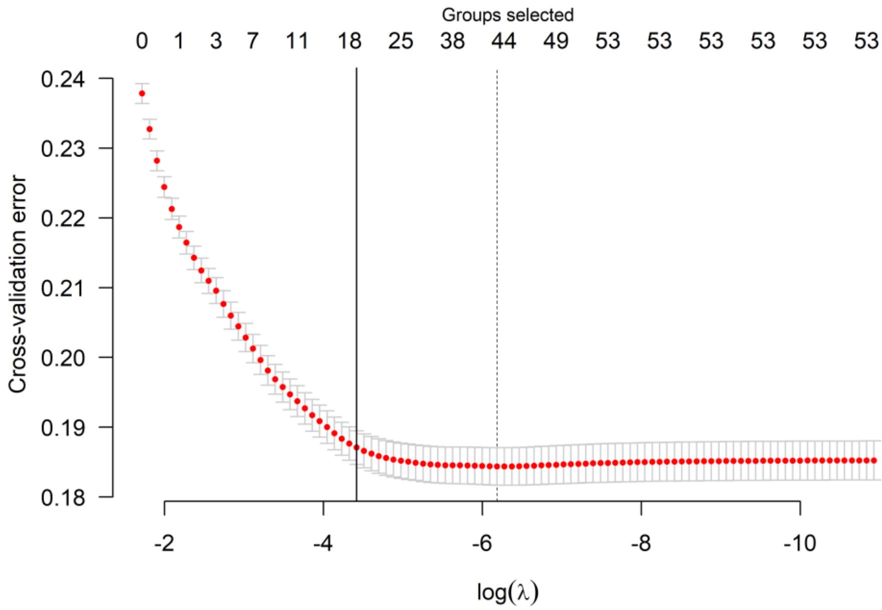


Figure 6: Ten-fold cross-validation error across different lambda values for group lasso

5. RESULTS AND DISCUSSION

5.1. Model Framework and Statistical Significance

This section presents the findings from logistic regression models examining the causal effect of special events on pass-by crash likelihood across rural counties in Florida for 2021 and 2022. For each district and year combination, two models were developed: a district-level model considering all crashes occurring across the counties within the district, and a county-level model focused exclusively on the counties identified through the district-level model as exhibiting the strongest causal effect of special events.

To assess whether county-specific effects of special events were statistically significant, Table 2 presents the Analysis of Variance (ANOVA) for the district-level logistic regression models across Districts 3 and 5 over the study period, showing the association of the model variables with pass-by crash frequencies. For conciseness, only the ANOVA tables are reported rather than full regression summaries. Notably, the interaction between county and special event was not statistically significant at the 0.1 level for District 3 in 2022 or District 5 in 2021. However, in the corresponding regression summaries, the county variable alone reached significance at the 0.1 level. These findings highlight how the significance of spatial effects can vary depending on the specific modeling output under consideration. A dashed line in the table denotes variables that were excluded from the model.

Table 2: Results of the district-level logistic regression models for Districts 3 and 5 over the study period.

Variable Type	Variable Name	District 3, 2021	District 3, 2022	District 5, 2021	District 5, 2022
		$\Pr(> Chi)$			
Treatment Variable	County×Near_Special_Event	0.073 ·	0.390	0.265	0.0762 ·
	Near_Special_Event	0.079 ·	0.266	0.015 *	0.68
	County	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***
Control Variable	Crash_Condition	-	-	-	<0.001 ***
	Sex	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***
	Alcohol_Involved	0.009 **	0.001 **	-	-
	Drug_Involved	-	0.056 ·	-	-
	Driver_Distracted	0.084 ·	-	-	-
	Shoulder_Type	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***
	Light_Condition	0.137	0.151	-	-
	Road_Surface	-	<0.001 ***	-	-
	Crash_Severity	-	0.019 *	-	-
	Lane_Departure	<0.001 ***	<0.001 ***	-	-
	Functional_Class	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***
	Day_of_Week	0.026 *	<0.001 ***	<0.001 ***	-
	Inside_Shoulder_Type	0.037 *	<0.001 ***	-	-
	Outside_Shoulder_Type	-	0.009 **	-	-
	Total_Persons_Involved	-	-	<0.001 ***	-
	No-Injury_Count	-	-	0.103	0.063 ·
	Fatalities	0.125	-	-	-
	Time_of_Day	-	0.179	0.113	-
	Lane_Count	-	-	0.046 *	0.724
	Speed	0.722	0.101	0.040 *	0.153
	AADT	-	-	0.473	0.676
	Population_Density	0.05 ·	0.007 **	<0.001 ***	<0.001 ***
	Household_Size	0.008 **	0.086 ·	-	-
	Poverty_Rate	-	-	-	<0.001 ***
	Education_Rate (HS+)	-	-	0.004 **	-
	Education_Rate (BA+)	-	0.259	<0.001 ***	<0.001 ***
	Median_Income	-	-	<0.001 ***	0.508
	Uninsured_Rate	-	0.648	<0.001 ***	<0.001 ***
	Employment_Rate	-	-	0.140	-
	Vacancy_Rate	0.002 **	<0.001 ***	0.315	-
	Owner_Occupancy_Rate	-	0.0015 **	-	0.002 **
	White_Proportion	0.936	0.977	<0.001 ***	0.014 *
	Black_Proportion	0.571	0.601	<0.001 ***	0.366
	Asian_Proportion	-	0.771	-	-
	Median_Home_Value	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***
	Median_Household_Income	<0.001 ***	-	-	0.076 ·
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '·' 0.1 ' ' 1					

5.2. Spatial and Temporal Effects of Special Events

The logistic regression models identified specific counties where the odds of a pass-by crash during special event periods differed meaningfully from those of the reference counties. To facilitate interpretation of these county-level differences and clearly illustrate the causal effects of special events on pass-by crash probability, Figure 7 presents the effects of the interaction between counties and special events on pass-by crash probability predicted by the model in the two districts depicting how predicted crash probabilities shift between event and non-event periods across all counties within each region, based on the fitted models. These plots represent the interaction effect between County and the special event indicator along with the confidence intervals of the probabilities which quantify the uncertainty or variability in model predictions for this region and time period.

In Figure 7 the counties with notably higher predicted probabilities are highlighted in bold letter fonts on the x-axis to draw attention to areas of elevated risk. In District 3 in 2021, Calhoun, Jefferson, and Bay counties exhibited elevated pass-by crash probabilities during special event periods and in 2022, Jefferson and Liberty counties continued to show this pattern. In District 5 during 2021, Orange, Flagler, and Osceola counties demonstrated higher predicted pass-by crash probabilities during special event periods and in 2022, Orange and Flagler counties continued to show elevated probabilities.

These effect plots underscore the spatial and temporal variability in the causal relationship between special events and pass-by crashes, confirming that certain counties face consistently heightened vulnerability during event periods.

5.3. County-Level Analysis and Model Refinement

Based on these findings, subsequent analysis concentrated on counties where special events were associated with elevated pass-by crash likelihood. County-level models were then employed to explore how roadway characteristics, driver demographics, and environmental conditions contribute to this heightened risk, providing insight into the factors that extend beyond the events themselves. To further investigate these patterns, the dataset was filtered to retain only the counties identified as having increased pass-by crash probabilities during special events within each district and year. Figure 8 presents a map highlighting these counties in distinct colors, illustrating those associated with higher crash counts across Districts 3 and 5 during the study period. This figure demonstrates the robustness of the modeling approach by revealing largely overlapping sets of counties across districts and years. While the precise counties identified are not identical across all combinations, the differences are minimal and the highlighted counties remain geographically adjacent, suggesting consistent regional patterns. Separate analyses were conducted for each subset—District 3 (2021, 2022) and District 5 (2021, 2022)—with group LASSO regression applied within each to select the most relevant predictors of pass-by crash likelihood. These predictors were subsequently used to re-estimate logistic regression models, yielding four refined models in total.

5.4. Effect of Special Events on Pass-by Crash Likelihood

Table 3 and Table 4 present the variables found to be statistically significant at the 0.1 significance level for the counties located in District 3 and District 5, respectively, across the two years of the study period. For categorical variables, all levels are reported regardless of whether only one level attains statistical significance.

Special events consistently demonstrated a positive and statistically significant effect on pass-by crash likelihood across both District 3 and District 5, with variation in the magnitude of this effect observed across years and geographic locations. In District 3 during 2021, the odds of a pass-by crash occurring during a special event period nearly doubled, reinforcing the heightened risk associated with such periods in rural settings. In District 3 for 2022, the effect remained substantial, with the odds of a pass-by crash increasing by approximately 89%. In District 5 during 2021, special events were associated with a 51% increase in pass-by crash likelihood. Although this effect size was comparatively smaller than that observed in District 3, it nonetheless underscores the temporal vulnerability to pass-by crashes introduced by special events in these areas. The most pronounced effect was recorded in District 5 during 2022, where the odds of a pass-by crash were nearly three times higher during special event periods. Across all four models, these results consistently support the conclusion that special events elevate pass-by crash risk, with the variation in effect magnitude likely attributable to district-level differences in road network characteristics, traffic management capacity, and the degree of exposure to non-local drivers.

Table 3: Coefficients of significant variables for counties in District 3 over the study period.

	Term	Estimate	P value
Bay, Calhoun and Jefferson (2021)	Near_Special_Event	0.68	<0.0001
	SexMale	0.31	0.043
	Functional_Class Major Collector	0.37	0.285
	Functional_Class Minor Arterial	0.31	0.335
	Functional_Class Minor Collector	0.37	0.285
	Functional_Class Principal Arterial—Interstate	2.79	<0.0001
	Functional_Class Principal Arterial—Other	0.92	0.001
	Lane_Count	-0.53	0.067
	Population_Density	-0.004	0.06
	Poverty_Rate	-4.61	0.002
Jefferson and Liberty (2022)	Near_Special_Event	0.654	0.023
	Road_Surface Not Dry	0.638	0.087
	Hit_and_RunY	1.189	0.034
	Functional_Class Major Collector	-0.239	0.69
	Functional_Class Minor Arterial	-0.013	0.984
	FUNCLASSMinor Collector	0.026	0.971
	Functional_Class Principal Arterial—Interstate	2.366	0.029
	Functional_Class Principal Arterial—Other	0.41	0.443
	Median_Type 02	-1.951	0.106
	Median_Type 08	-1.026	0.030
	Median_Type 17	-2.109	0.003
	Median_Type 35	-13.671	0.983
	Median_Type 44	-0.233	0.865
	Inside_Shoulder_Type 1	-0.484	0.648
	Inside_Shoulder_Type 2	-0.979	0.333
	Inside_Shoulder_Type 6	-0.395	0.67
	Inside_Shoulder_Type 8	-1.809	0.070
	Asian_Proportion	-285.443	0.094

Table 4: Coefficients of significant variables for counties in District 4 over the study period.

	Term	Estimate	P value
Orange, Flagler and Osceola (2021)	Near_Special_Event	0.41	0.058
	CountyOrange	1.97	0.006
	CountyOsceola	4.63	0.002
	Functional_Class Major Collector	1.41	0.051
	Functional_Class Minor Arterial	1.6	0.037
	Functional_Class Minor Collector	0.98	0.169
	Functional_Class Principal Arterial—Free& Expways	3.34	0.0001
	Functional_Class Principal Arterial—Interstate	6.63	<0.0001
	Functional_Class Principal Arterial—Other	1.87	0.007
	Inside_Shoulder_Type 1	-0.74	0.09
	Inside_Shoulder_Type 2	-0.04	0.932
	Inside_Shoulder_Type 6	-0.01	0.991
	Inside_Shoulder_Type 8	-0.54	0.592
	Population_Density	-0.0006	0.054
	Owner_Occupancy_Rate	-12.88	0.004
Orange, Flagler (2022)	Near_Special_Event	1.039	0.0004
	Junction_Flagon-Junction	-0.216	0.534
	Junction_FlagOther	-2.546	0.026
	CountyOrange	-0.943	0.075
	Functional_Class Major Collector	-0.085	0.908
	Functional_Class Minor Arterial	1.842	0.051
	Functional_Class Minor Collector	-0.491	0.499
	Functional_Class Principal Arterial—Interstate	15.763	0.981
	Functional_Class Principal Arterial—Other	0.568	0.389
	Minor_Injuries	0.664	0.014
	Lane_Count	1.064	0.04
	Employment_Rate	162.971	0.08
	White_Proportion	-5.039	0.072
	Median_Home_Value	5.6e-06	0.029

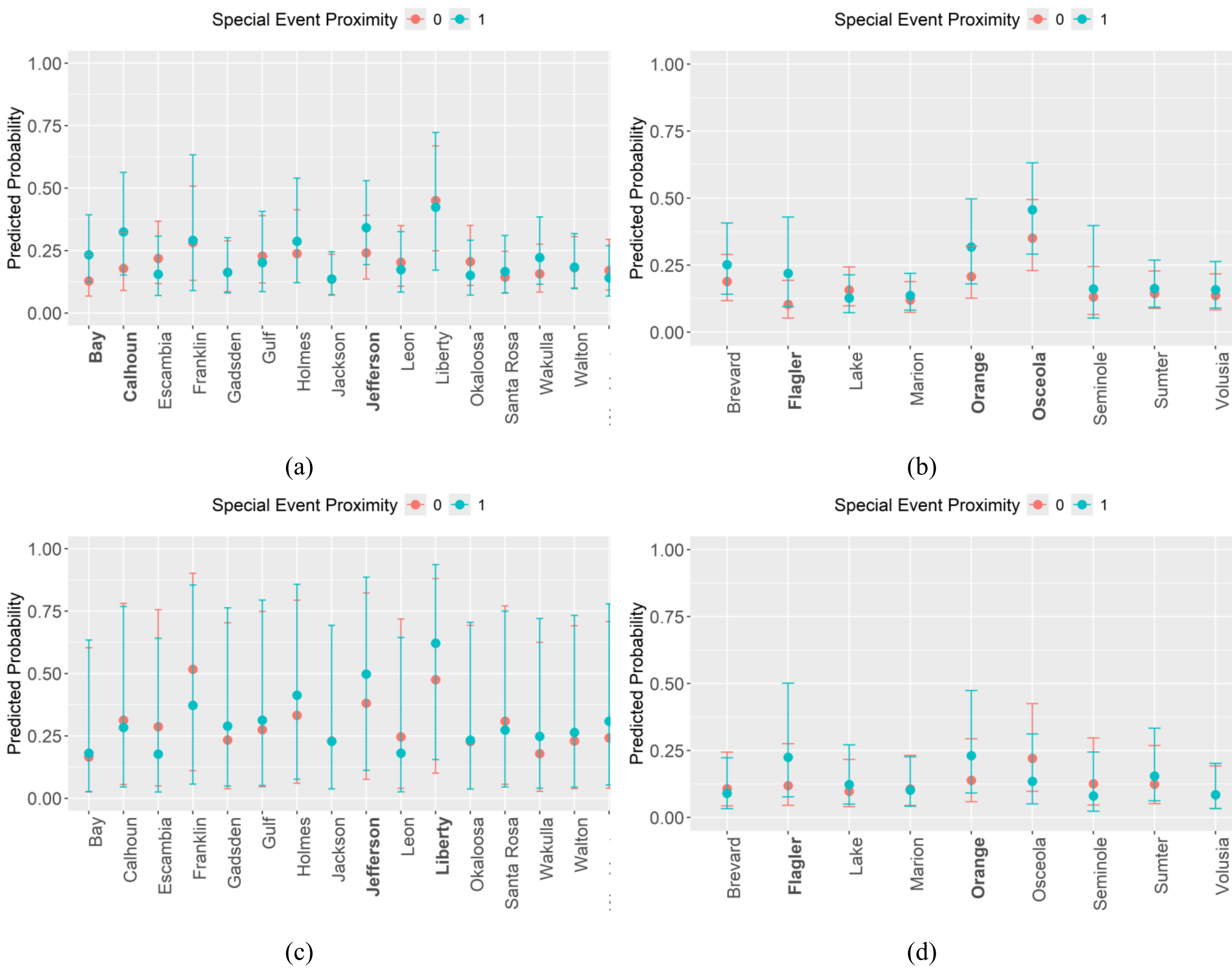


Figure 7: Effects of the interaction between counties and special events on pass-by crash probability (a) district 3, 2021; (b) district 5, 2021; (c) district 3, 2022; and (d) district 5, 2022. Counties with notably higher predicted probabilities are highlighted in bold letter fonts in x-axis labels.

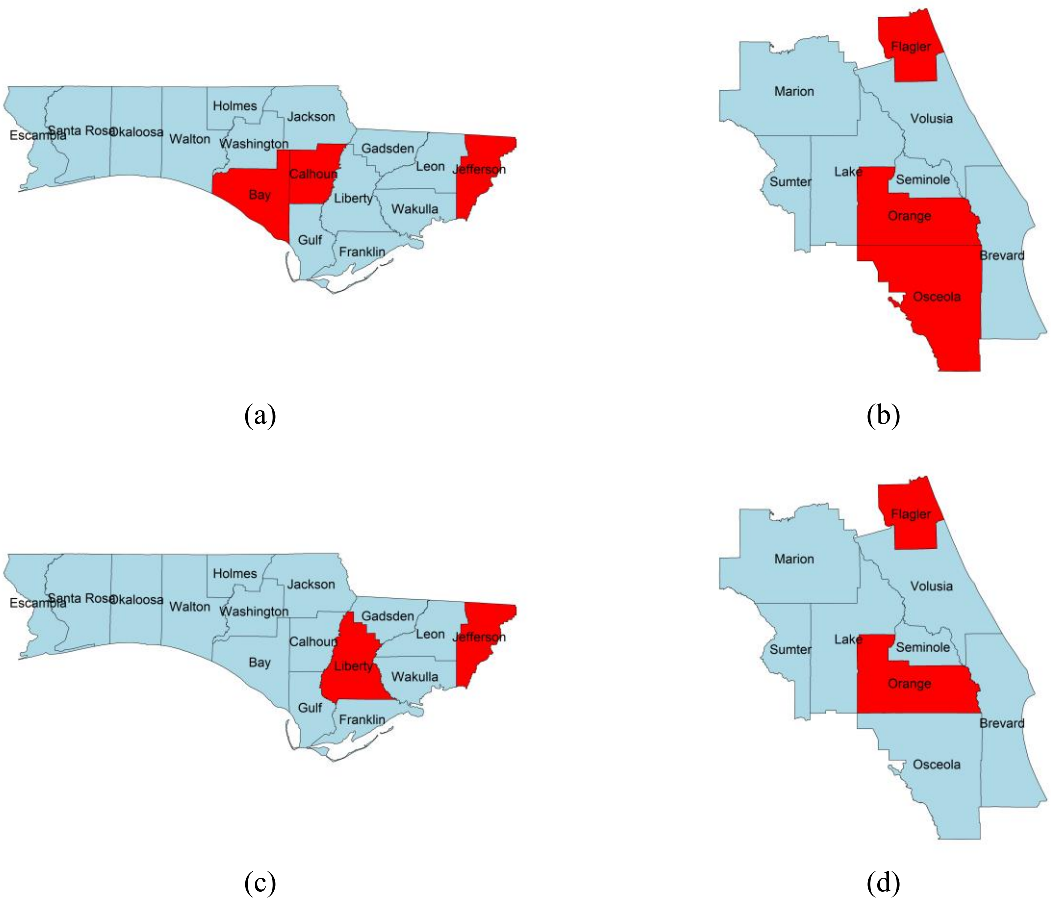


Figure 8: Counties associated with higher pass-by crash counts during special events, shown by district and year: (a) District 3, 2021; (b) District 3, 2022; (c) District 5, 2021; (d) District 5, 2022

5.5. Control Variable Effects

The results for the control variables were largely consistent with prior expectations. Roadway functional class consistently exhibited increased odds of pass-by crashes for higher-order facilities such as Interstates and Principal Arterials, a finding that was anticipated given that these road types are predominantly used by non-local travelers covering longer distances. In District 3, Interstates were associated with odds ratios of 16.3 in 2021 and 10.65 in 2022, reflecting persistent elevated risk along major travel corridors. In District 5, this effect was even more pronounced in 2021, where Interstates were associated with an increase in the odds of a pass-by crash by more than 750 times, underscoring the substantial influence of these facilities on regional travel patterns during special event periods. Conversely, certain median types, including vegetation and curb-and-vegetation configurations, were associated with reduced odds of pass-by crashes, suggesting a protective effect likely attributable to roadway design features and enhanced access control, a pattern that aligns with established expectations in the literature.

The number of lanes produced mixed results across the two districts. In District 3, a greater number of lanes was associated with reduced crash likelihood, whereas in District 5 the relationship was positive, potentially reflecting differences in traffic volumes or roadway design characteristics between the two regions. Socioeconomic factors linked to drivers' residential ZIP codes also exhibited consistent patterns. Higher poverty rates and lower homeownership rates were associated with decreased odds of pass-by crash involvement, possibly reflecting reduced propensity for long-distance travel among lower-income or more transient populations. In contrast, higher employment rates and greater median home values tended to increase crash odds, likely capturing the greater mobility and discretionary travel activity characteristic of residents in more economically active or affluent areas. The proportion of the white population showed a negative association with pass-by crash involvement, though the mechanisms underlying this relationship warrant careful and cautious interpretation. Overall, the control variables behaved largely in accordance with expectations, with roadway characteristics and socioeconomic factors shaping exposure to pass-by crash risk in ways that are broadly consistent with existing literature on non-local travel behavior and rural traffic safety.

6. CONCLUSIONS

This study examined the causal effects of special events on the likelihood of pass-by crashes in rural Florida, with particular emphasis on spatial and temporal variations across two FDOT districts over the 2021 and 2022 study period. Multiple regression models were developed to serve this analytical purpose. The first regression model was used to identify counties exhibiting significant differences in pass-by crash probabilities during special event periods. Effect plots were additionally employed to facilitate clearer interpretation of the causal effects of special events on pass-by crash probability at the county level. Counties with elevated predicted probabilities were subsequently selected for inclusion in a second logistic regression model to represent their respective districts for each study year. Jefferson County in District 3, and Orange and Flagler Counties in District 5, were consistently identified as having heightened pass-by crash probabilities during special event periods. A follow-up model applied to these high-risk counties further explored the localized factors contributing to elevated crash likelihood. Collectively, these findings underscore the importance of accounting for spatial and temporal context when assessing event-related crash risk, and point to the need for targeted traffic management strategies during special event periods to reduce pass-by crash occurrence. Interstates and principal arterials consistently exhibited increased odds of pass-by crashes, while certain median types were associated with a reduced risk. The analysis also yielded noteworthy insights related to the demographic characteristics of drivers involved in these crashes, which merit further investigation.

Several limitations of this study present opportunities for future research. First, the records available for identifying special event periods may not have fully captured unofficial or smaller-scale events that could nonetheless exert a meaningful influence on traffic patterns. Second, although multiple covariates and interaction effects were incorporated into the models, unobserved factors related to driver behavior, temporary changes to infrastructure, and the presence of law enforcement may still exert influence on the results. Finally, the application of advanced machine learning methods in future work could help identify nonlinear relationships and spatial dependencies in crash patterns, thereby offering a more nuanced and comprehensive understanding of event-related crash risks.

REFERENCES

[1] R. C. Wunderlich and E. M. Shipp, "Perspectives on road safety through the lens of traffic crashes in the United States," in *Advances in Transportation and Health*, Elsevier, 2020, pp. 35–58. doi: 10.1016/B978-0-12-819136-1.00002-4.

[2] G. Xue, Z. Zhu, Q. Liu, and H. Wen, "Modeling faults among mountainous highway fatal crash-involved drivers with different route familiarity levels," *Traffic Inj Prev*, vol. 24, no. 2, pp. 140–146, Feb. 2023, doi: 10.1080/15389588.2023.2167200.

[3] Z. Sun, S. Liu, D. Li, B. Tang, and S. Fang, "Crash analysis of mountainous freeways with high bridge and tunnel ratios using road scenario-based discretization," *PLoS One*, vol. 15, no. 8, p. e0237408, Aug. 2020, doi: 10.1371/journal.pone.0237408.

[4] A. J. Khattak, N. Ahmad, B. Wali, and E. Dumbaugh, "A taxonomy of driving errors and violations: Evidence from the naturalistic driving study," *Accid Anal Prev*, vol. 151, p. 105873, Mar. 2021, doi: 10.1016/j.aap.2020.105873.

[5] Y. Li, D. Ma, M. Zhu, Z. Zeng, and Y. Wang, "Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network," *Accid Anal Prev*, vol. 111, pp. 354–363, Feb. 2018, doi: 10.1016/j.aap.2017.11.028.

[6] R. Tay, "A random parameters probit model of urban and rural intersection crashes," *Accid Anal Prev*, vol. 84, pp. 38–40, Nov. 2015, doi: 10.1016/j.aap.2015.07.013.

[7] N. S. S. Al-Bdairi and S. Hernandez, "Comparison of contributing factors for injury severity of large truck drivers in run-off-road crashes on rural and urban roadways: Accounting for unobserved heterogeneity," *International Journal of Transportation Science and Technology*, vol. 9, no. 2, pp. 116–127, Jun. 2020, doi: 10.1016/j.ijtst.2020.01.004.

[8] I. Shrira and K. Noguchi, "Traffic fatalities of drivers who visit urban and rural areas: An exploratory study," *Transp Res Part F Traffic Psychol Behav*, vol. 41, pp. 74–79, Aug. 2016, doi: 10.1016/j.trf.2016.05.003.

[9] Raccagni, S., Ventura, R., & Barabino, B. (2025). Evaluating road crash risk for a mega-event: an application to 2026 Milano-Cortina Winter Olympic games. *Transportation Research Interdisciplinary Perspectives*, *34*, 101709.

[10] Jakar, G., Gordon, K., & He, Q. (2023). The temporal and spatial relationships between professional sport events and reported vehicular crashes: an analysis of Cleveland, Ohio. *Sport Management Review*, *26*(4), 628-648.

[11] Skrodenis, D. (2018, December). Road Traffic Management During Special Events.

In *Conference Vision Zero for Sustainable Road Safety in Baltic Sea Region* (pp. 104-109). Cham: Springer International Publishing.

[12] A. J. Ghandour, H. Hammoud, and S. Al-Hajj, “Analyzing Factors Associated with Fatal Road Crashes: A Machine Learning Approach,” *Int J Environ Res Public Health*, vol. 17, no. 11, p. 4111, Jun. 2020, doi: 10.3390/ijerph17114111.

[13] M. M. R. Komol, M. M. Hasan, M. Elhenawy, S. Yasmin, M. Masoud, and A. Rakotonirainy, “Crash severity analysis of vulnerable road users using machine learning,” *PLoS One*, vol. 16, no. 8, p. e0255828, Aug. 2021, doi: 10.1371/journal.pone.0255828.

[14] S. Das, M. Le, and B. Dai, “Application of machine learning tools in classifying pedestrian crash types: A case study,” *Transportation Safety and Environment*, vol. 2, no. 2, pp. 106–119, Aug. 2020, doi: 10.1093/tse/tdaa010.

[15] K. Assi, “Traffic Crash Severity Prediction—A Synergy by Hybrid Principal Component Analysis and Machine Learning Models,” *Int J Environ Res Public Health*, vol. 17, no. 20, p. 7598, Oct. 2020, doi: 10.3390/ijerph17207598.

[16] K. Sattar, F. Chikh Oughali, K. Assi, N. Ratrout, A. Jamal, and S. Masiur Rahman, “Transparent deep machine learning framework for predicting traffic crash severity,” *Neural Comput Appl*, vol. 35, no. 2, pp. 1535–1547, Jan. 2023, doi: 10.1007/s00521-022-07769-2.

[17] P. Intini, P. Colonna, and E. Olaussen Ryeng, “Route familiarity in road safety: A literature review and an identification proposal,” *Transp Res Part F Traffic Psychol Behav*, vol. 62, pp. 651–671, Apr. 2019, doi: 10.1016/j.trf.2018.12.020.

[18] M. R. Abbaszadeh Lima, M. M. Hossain, H. Zhou, and Y. Song, “Data Mining Approach to Explore the Contributing Factors to Fatal Wrong-Way Crashes by Local and Non-Local Drivers,” *Future Transportation*, vol. 4, no. 3, pp. 985–999, Sep. 2024, doi: 10.3390/futuretransp4030047.

[19] Q. Zeng, X. Wang, H. Wen, and Q. Yuan, “An empirical investigation of the factors contributing to local-vehicle and non-local-vehicle crashes on freeway,” *Journal of Transportation Safety & Security*, vol. 14, no. 2, pp. 323–337, Feb. 2022, doi: 10.1080/19439962.2020.1779422.

[20] M. Vahedi Saheli, “Understanding the Factors Affecting Urban Vehicle-to-Vehicle Crash Severity with Focus on Drivers’ Route Familiarity,” *COMPUTATIONAL RESEARCH PROGRESS IN APPLIED SCIENCE & ENGINEERING*, vol. 8, no. S1, pp. 1–7, 2022, doi: 10.52547/crpase.8.2227.

[21] H. Wen and G. Xue, “Injury severity analysis of familiar drivers and unfamiliar drivers in

single-vehicle crashes on the mountainous highways," *Accid Anal Prev*, vol. 144, Sep. 2020, doi: 10.1016/j.aap.2020.105667.

[22] Y. Wang, D. Veneziano, S. Russell, and A. Al-Kaisy, "Traffic safety along tourist routes in rural areas," *Transp Res Rec*, vol. 2568, pp. 55–63, 2016, doi: 10.3141/2568-09.

[23] G. Yannis, J. Golias, and E. Papadimitriou, "Accident risk of foreign drivers in various road environments," *J Safety Res*, vol. 38, no. 4, pp. 471–480, Jan. 2007, doi: 10.1016/j.jsr.2007.01.014.

[24] M. R. Yanko and T. M. Spalek, "Route familiarity breeds inattention: A driving simulator study," *Accid Anal Prev*, vol. 57, pp. 80–86, Aug. 2013, doi: 10.1016/j.aap.2013.04.003.

[25] J. Wu and H. Xu, "The influence of road familiarity on distracted driving activities and driving operation using naturalistic driving study data," *Transp Res Part F Traffic Psychol Behav*, vol. 52, pp. 75–85, Jan. 2018, doi: 10.1016/j.trf.2017.11.018.

[26] "Signal Four Analytics." Accessed: Feb. 26, 2025. [Online]. Available: htps://s4.geoplan.ufl.edu

[27] "Transportation Data and Analytics, Road Characteristics Inventory Data." Accessed: Feb. 26, 2025. [Online]. Available: htps://www.fdot.gov/statistics/gis/default.shtm#Roadway

[28] "U.S. Census Bureau (2021) American Community Survey." Accessed: Feb. 26, 2025. [Online]. Available: htps://www.census.gov/content/dam/Census/library/publications/2021/acs/acs_summary_file_handbook_2021. pdf

[29] "U.S. Census Bureau (2010) Census Summary File 1 – Technical Documentation." Accessed: Feb. 26, 2025. [Online]. Available: htps://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/complete-techdocs/summary-file/sf1.pdf?#

[30] "FDOT District Maps." Accessed: Jul. 15, 2025. [Online]. Available: https://www.fdot.gov/publications/distmap/index.shtm

[31] R Core Team, "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*, 2022.

[32] "U.S. Census Urban Area Boundaries 2020." Accessed: Jul. 15, 2025. [Online]. Available: https://gis-fdot.opendata.arcgis.com/datasets/fdot::u-s-census-urban-area-boundaries-2020/about

[33] "HUD USPS ZIP Code Crosswalk." Accessed: Jul. 15, 2025. [Online]. Available: https://www.huduser.gov/portal/datasets/usps_crosswalk.html

[34] Martin Huber, *Causal Analysis: Impact Evaluation and Causal Machine Learning with Applications in R*. The MIT Press, 2023.

[35] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity*. Chapman and Hall/CRC, 2015. doi: 10.1201/b18401.