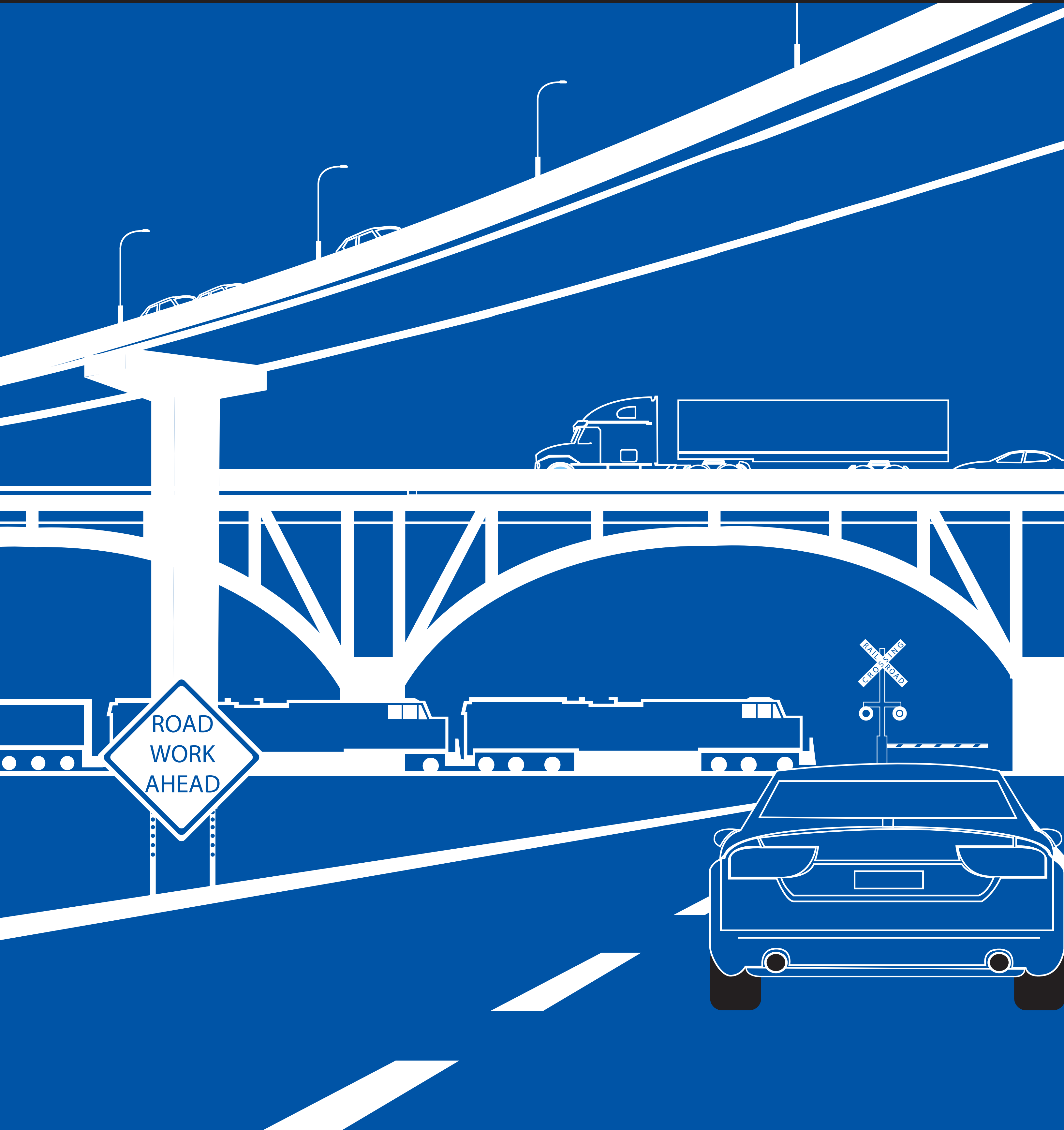
Vehicle on Shoulder and Crash — Correlation or Causation?

Report Number: KTC-20-25/SPR20-598-1F

DOI: https://doi.org/10.13023/ktc.rr.2020.25



Kentucky Transportation Center
College of Engineering, University of Kentucky, Lexington, Kentucky

in cooperation with
Kentucky Transportation Cabinet
Commonwealth of Kentucky

The Kentucky Transportation Center is committed to a policy of providing equal opportunities for al persons in recruitment, appointment, promotion, payment, training, and other employment and education practices without regard for economic, or social status and will not discriminate on the basis of race, color, ethnic origin, national origin, creed, religion, political belief, sex, sexual orientation, marital status or age.

Kentucky Transportation Center
College of Engineering, University of Kentucky, Lexington, Kentucky

in cooperation with
Kentucky Transportation Cabinet
Commonwealth of Kentucky

KENTUCKY
Transporation Center

Kentucky Transportation Center • University of Kentucky
176 Raymond Building • Lexington, KY 40506 • 859.257.6898 • www.ktc.uky.edu

University of Kentucky

Research Report

KTC-20-25/SPR20-598-1F

Vehicle on Shoulder and Crash — Correlation or Causation?

Mei Chen, Ph.D.
Professor

Xu Zhang, Ph.D.
Research Engineer

Benjamin Blandford, Ph.D
Research Scientist

and

Eugene Antwi Boasiako
Graduate Research Assistant

Kentucky Transportation Center
College of Engineering
University of Kentucky
Lexington, Kentucky

In Cooperation With
Kentucky Transportation Cabinet
Commonwealth of Kentucky

July 2020

<table>
<tr><td>1. Report No.
KTC-20-25/SPR20-598-1F</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No</td></tr>
<tr><td colspan="2">4. Title and Subtitle
Vehicle on Shoulder and Crash — Correlation or Causation?</td><td colspan="2">5. Report Date
July 2020
6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s):
Mei Chen, Xu Zhang, Benjamin Blandford, Eugene Antwi Boasiako</td><td colspan="2">8. Performing Organization Report No.
KTC-20-25/SPR20-598-1F</td></tr>
<tr><td colspan="2">9. Performing Organization Name and Address
Kentucky Transportation Center
College of Engineering
University of Kentucky
Lexington, KY 40506-0281</td><td colspan="2">10. Work Unit No. (TRAIS)
11. Contract or Grant No.
SPR 20-598</td></tr>
<tr><td colspan="2">12. Sponsoring Agency Name and Address
Kentucky Transportation Cabinet
State Office Building
Frankfort, KY 40622</td><td colspan="2">13. Type of Report and Period Covered
14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes
Prepared in cooperation with the Kentucky Transportation Cabinet</td></tr>
<tr><td colspan="4">16. Abstract
This study sought to evaluate the relationships between vehicles on shoulder, congestion, and crashes. Three pertinent datasets on interstates were obtained and then integrated using a spatiotemporal approach. The analysis showed that about 36% of the crashes had vehicle on shoulder present in their vicinity, defined as 0.25 miles upstream and downstream of a crash site and 30 minutes before crash occurrence. The percentage increased considerably as spatial/temporal window expanded. In addition, congestion was found to be associated with about 25% of the crashes. The presence of both vehicle on shoulder and congestion was found for 11.7% of the crashes, signifying a high correlation between them and crashes. Based on crash narrative review, 1.8% of all crashes directly involved vehicles on shoulder and 23% of the carshes cited congestion as a contributor. However, there's little indication in the crash narratives on how vehicles on shoulder contributed to crashes, beyond their direct involvement, or how they contributed to congestion which may led to crashes. Only 6 out of the 512 crashes flagged for review through the keyword search process specified a vehicle on shoulder as a contributor to congestion and subsequent crashes. While a small fraction of crashes were attributed to vehicles on shoulder, these crashes tended to be more severe than average interstate crashes.</td></tr>
<tr><td colspan="2">17. Key Words
vehicle on shoulder, crash, traffic congestion, crash narrative, Waze, crowdsourced data, crowdsourcing</td><td colspan="2">18. Distribution Statement
Unlimited with approval of the Kentucky Transportation Cabinet</td></tr>
<tr><td>19. Security Classification (report)
Unclassified</td><td>20. Security Classification (this page)
Unclassified</td><td>21. No. of Pages
30</td><td>19. Security Classification (report)</td></tr>
</table>

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Crashes result in fatalities, economic losses, and often significantly affect roadway capacity. Crashes which occur on road shoulders reduce roadway capacity by up to 19% (Transportation Research Board, 2016), and for every 20 minutes a roadway remains uncleared, the likelihood of a secondary crash occurring increases by up to 7% (Goodall, 2017). As such, vehicles on the shoulder of limited access highways are a safety hazard.

Conventionally, operators in a traffic management center learn of potential safety hazards and traffic incidents through sources such as traffic cameras and emergency services dispatch. Recently, however, road user-generated (crowdsourced) data have begun to provide a cost-effective alternative to traffic monitoring. Researchers have studied crowdsourced traffic incident data from Waze, a GPS navigation application. Waze incident alerts have been found to have moderate spatial accuracy with low false alarm rates (Amin-Naseri et al., 2018; Goodall & Lee, 2019; Liu et al., 2019). Waze data, however, contain multiple reports of the same incident — creating unnecessary redundancy. In response, various approaches leveraging spatiotemporal and semantic information have been proposed to mitigate redundancy and integrate Waze data with other data sources (Amin-Naseri et al., 2018; dos Santos et al., 2017a; Eriksson, 2019; Lenkei, 2018; Liu et al., 2019).

Researchers have described several applications that use Waze data to monitor traffic safety, including traffic crash estimation (Flynn et al., 2018), traffic crash monitoring (Young et al., 2019), and freeway traffic risk assessment (Turner et al., 2020). While such innovative applications of crowdsourced Waze data exist in literature, no previous researchers have attempted to use Waze road user alerts to evaluate how traffic safety and flows are impacted by vehicles on the shoulder of limited access highways.

1.1 Motivation

The Kentucky Transportation Cabinet's (KYTC) partnership with Waze through the Connected Citizens Program lets the agency access traffic alerts reported by Waze users, including accidents, traffic jams, potholes, objects on the road, vehicles stopped on shoulders, and other hazards. The Cabinet uses this information to provide situational awareness to the travelers and support real-time incident management.

By reviewing traffic data alongside crowdsourced reports, KYTC has frequently observed the concurrence of the following chain of events: *Vehicles On Shoulder*, speed drops to well below free-flow condition, and then reports of *Crash*. Developing a better understanding of the correlation and/or causational relationships among these events can help KYTC develop operational strategies and policies to prevent future crashes and fatalities.

1.2 Objective

The objective of this research is to explore the correlation and/or causation between vehicles on shoulder, traffic slowdowns, and crashes. The goal of the adopted approach is to establish a spatiotemporal linkage between each crash, speed, and Waze alert. Doing so offers better knowledge of the circumstances leading up to the crash. Crash reports are analyzed and their narratives are reviewed for additional insight into the factors contributing to crashes.

Chapter 2 Review of Literature

2.1 Crowdsourced Data

Crowdsourced traffic data offer a cost-effective alternative for traffic monitoring. With crowdsourced data, traffic managers can monitor roadway conditions and coordinate more effective incident management, thereby improving road safety. Unlike conventional traffic data sources — sensors, cameras and floating car studies — which are neither economically viable nor feasible at all locations and times (Yoon et al., 2007), crowdsourced traffic information is generated by road users. Thus, it can potentially supplement conventional sources of traffic data and provide wider spatial coverage.

Many researchers have studied the utility of traffic data crowdsourced through Waze, a smartphone application. Waze is a GPS navigation app with over 115 million users worldwide (https://www.waze.com/about). It relies on reports generated by its users to provide services to the Waze community. Through mutually beneficial partnerships with transportation agencies, Waze provides real-time crowdsourced data to traffic managers, comprising Waze incident alerts and jams. Major types of incident alerts are accidents, road closures, hazards, and user-reported traffic jams. These alerts also have subtypes. For example, road construction, potholes, heavy rains, floods, and cars stopped on shoulders are all subtypes of hazards. A Waze incident alert contains information about the incident type, publication time, and location — including street name, city, and county of an incident. Other data attributes are number of confirmations, confidence level, and reliability. The reliability measure is contingent on the rank of the user reporting the incident and how other Waze users react to an alert.

Waze generates traffic jam data for road segments by comparing passively collected Waze user speed data to the historical average speed of the corresponding time and free-flow speed of that road segment. User-reported jams are processed to identify and present jams (Lenkei, 2018). Jam reports present information on publication time, location, average speed, expected delay on the jammed road segment, length of the jam, its start and end node, the road type on which the jam occurred, and the amount of congestion caused by the jam. Regardless of the alert type, Waze assigns a unique identification number (UUID) to each incident alert and jam.

2.2 Characterizing Crowdsourced Reports

Understanding the characteristics of crowdsourced data helps in assessing their reliability and additional traffic monitoring coverage they may potentially offer. Lenkei (2018) in his study on Waze data found a higher proportion of Waze incident alerts were made along major roads during the daytime and on weekdays than on minor roads, during the night and on weekends. Amin-Naseri et al. (2018) corroborated this, finding a high correlation between road type and time and the probability that an incident is reported in Waze. Outside of urban areas, however, Waze had much broader coverage than the conventional traffic incident detection sources. Also, a linkage was observed between the spatial distribution of crowdsourced reports and report reliability. Waze evaluates user reliability by ranking them. A reliability score is computed for incident alerts based on the rank of the reporter and the number of confirmations the alert receives from other users. Regions with a higher number of Waze users generate a larger number of Waze alerts. Correspondingly, alerts from these regions receive a higher number of confirmations. The probability of a Waze incident alert receiving a high reliability score is proportional to the number of Waze users in the vicinity of the incident (Lenkei, 2018).

Waze users passing the location of a previously reported alert are prompted by the app to confirm or deny the alert. Typically, Waze alerts are removed after 30 minutes if no users confirm it. An alert may be terminated before the typical report duration of 30 minutes if denied by other users. Alert durations may be longer depending on number of confirmations received. No linkage was established between Waze alert validity and Waze alert reliability score and duration (Goodall and Lee, 2019).

2.3 Challenges Associated with Crowdsourced Data

2.3.1 False Alarms

An inherent challenge associated with using crowdsourced traffic incident reports is sorting through false reports and multiple reports of the same incident. To identify and quantify false alarms, Amin-Naseri et al. (2018) and Goodall & Lee (2019) compared Waze alerts to screenshots of traffic camera video feed captured at time intervals of 5 minutes and 1 minute, respectively. They found false alarm rates for Waze data to be quite low. Of 319 Waze reports in the month of October, Amin-Naseri et al. (2018) reported that only one (0.3%) was a false alarm. Goodall and Lee (2019) documented false alarm rates of 5% for crash reports and 23% for disabled vehicle reports. The variance in false alarm rates between the two studies could potentially be attributed to the frequency at which ground truth data were collected for Waze incident validation and differences in study locations.

2.3.2 Redundancy

Various approaches leveraging spatiotemporal as well as semantic information (e.g., incident type, road name, direction) have been proposed to mitigate redundancy in Waze data by aggregating multiple alerts referring to the same incident. Lenkei (2018) and Eriksson (2019) made a clear distinction between algorithms for matching related Waze incident alerts and matching related Waze jams and incident alerts. Amin-Naseri (2018) did not make this distinction in his study, however, but developed a tool for lessening redundancy.

To match related Waze incident alerts, Lenkei (2018) and Eriksson (2019) employed road names as well as the spatiotemporal proximity of the alerts. In their studies, multiple alerts were treated as referring to the same incident if their existence overlapped in the data feed or the alerts were published within a one-hour period. Also, multiple alerts were considered as related in space if the distance between them was less than 70 m. For incident alerts on the same road, they were considered related if they were separated by less than 500 m. The result was a cluster of related alerts referring to the same incident and independent alerts not related to any alerts. Consequently, a cluster of related alerts was represented as one alert. This approach has a limitation — it cannot differentiate distinct incidents close in space and time. According to Eriksson (2019), in order of precedence, the alert types that may be chosen to represent a cluster are crashes and weather hazards. In cases where a cluster contains at least three crash or weather hazard alerts, the one closest to the centroid of the cluster is used. For clusters without crash or weather hazard alerts, the first alert in the cluster list represents the cluster.

To match related Waze jams and incident alerts, Lenkei (2018) considered a jam as related to an incident in time if they were published within a time period of 15 minutes. If the distance between them was less than 70 m, or they were on the same road and separated by less than 1,000 m, then they were also considered related. Thus, a jam related to an alert in a cluster could be treated as related to that cluster. Eriksson (2019) applied the same spatiotemporal proximity criteria as Lenkei (2018). Waze alerts inherently may hold an attribute referring to the UUID of the potential Waze jam caused by it. Conversely, Waze jams may also have the UUID of the potential alert causing it. Thus, Eriksson (2019) employed the UUIDs as an additional criterion when matching related Waze jams and alerts. When dealing with real time data, however, the temporal proximity criterion was not required since the jam and alert in question overlapped in the Waze data feed (Eriksson, 2019). Eriksson (2019) found that spatial proximity constraints accounted for no less than 59% of the total matches between Waze jams and incident alerts. Specifying spatial proximity constraints was more effective in matching jams and alerts than relying solely on UUIDs for matching. Comparing Waze jams to Waze incident alerts in Stockholm, he found that 75% of all Waze jams were related to Waze incident alerts and 81% of all Waze incident alerts had jams associated with them.

Amin-Naseri (2018) investigated the selection of clustering methods for crowdsourced traffic incident reports. He developed a tool which uses density-based clustering methods to cluster Waze data and reduce redundancy. His clustering methods included density-based spatial analysis of clusters with noise (DBSCAN), Space-Time DBSCAN (STDBSCAN), and Hierarchical DBSCAN (HDBSCAN), which were considered as best fits for Waze data. The general idea is to specify 1) the minimum number of data points that form a cluster and 2) the spatiotemporal constraints so that the tool clusters Waze alerts based on user-specified constraints. The result is a cluster of related jams, a cluster of related incident alerts, or a cluster of related jams and incident alerts.

2.3.3 Data Integration

Integrating crowdsourced data with traffic incident data obtained through conventional data sources in real time is a necessary step to fully realize the benefits of crowdsourced data in traffic management. However, the inherent differences in the two datasets pose a challenge. Moreover, the reliability and validity of crowdsourced reports is of importance to traffic managers. To this end, various methods have been proposed for data integration. Eriksson (2019) proposed a tool for integrating Waze data and official traffic data in real time. To match crowdsourced data containing clusters of related alerts and independent alerts to the official traffic dataset in real time, he considered reports as related if the distance between them was less than 70 m, or 500 m if they were on the same road. Similarly, Goodall and Lee (2019) matched Waze incident alerts to an official dataset using spatiotemporal constraints. They treated a

Waze incident alert as related to a report in the official dataset if both incidents were published within a time period of 30 minutes and were 0.5 mile apart on the same road and in the same direction.

Since the output of different integration tools and methodologies hinges on spatiotemporal constraints, their efficiency is affected by the spatial and temporal accuracy of crowdsourced reports. Typically, Waze alerts filed earlier — between 2.2 to 9.8 minutes — than corresponding event reports in official datasets (Amin-Naseri et al., 2018; Lenkei, 2018; Liu et al., 2019; Young et al., 2019). On average, the spatial difference between Waze crash and stopped vehicle report locations and official traffic data locations was less than 0.001 mile and 0.0025 mile respectively, on a Tennessee interstate (Liu et al., 2019). Thus, integrating crowdsourced data into traffic management systems can be achieved with a reasonable level of accuracy. Generally, studies report more than 40% of official records are detected by Waze. However, only a small fraction of all Waze data are reported in official records (Amin-Naseri et al., 2018; Eriksson, 2019; Flynn et al., 2018). This highlights that Waze can potentially supply additional data to traffic management systems, particularly for low-severity crashes, which are often underrepresented in crash reports.

2.4 Crashes Involving Vehicles on Shoulder

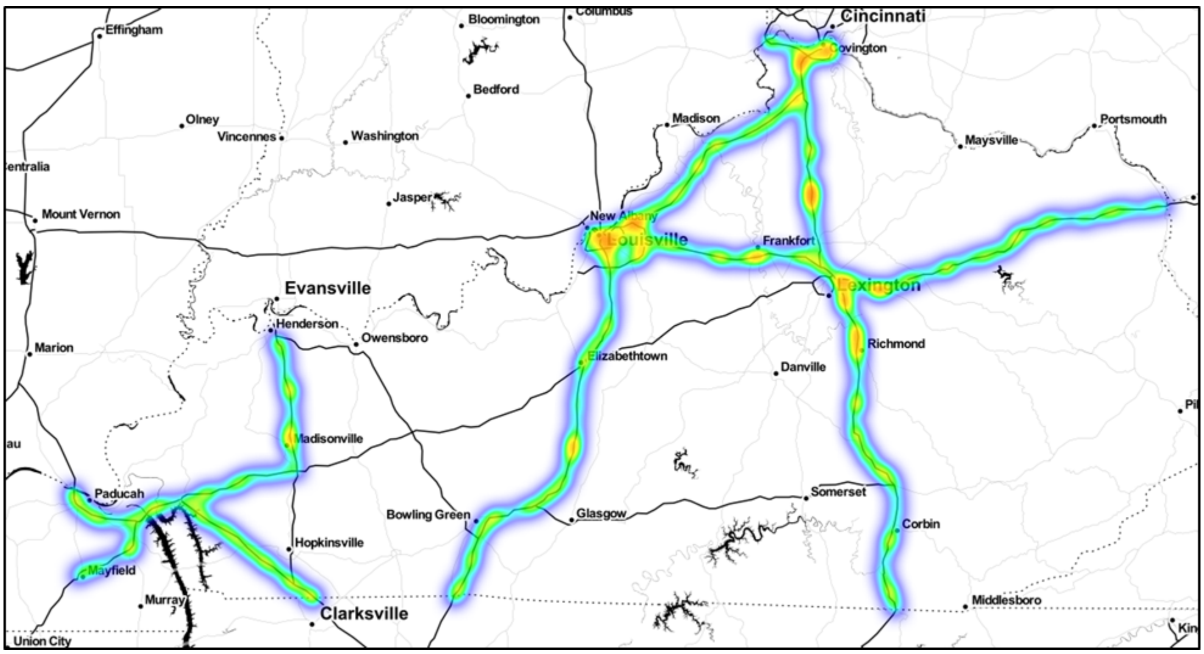
In a study to assess the influence of vehicles on the shoulders of limited access highways on crashes, Agent and Pigman (1989) conducted observational surveys and analyzed crash data covering a three-year period. On average, they found that 1.9 crashes per 100 million vehicle miles were caused by a vehicle on the shoulder of an interstate or parkway — with a higher rate on interstates than parkways. A smaller percentage of crashes was attributed to vehicles pulling out of the shoulder. Most crashes were, however, caused by vehicles that had stopped on the shoulder for several reasons, including mechanical failure and leisure. Similarly, Chimba and Kutela (2014) reported that 76% of secondary crashes were associated with abandoned or disabled vehicles parked on the shoulders of interstates and parkways in Tennessee. Crashes involving a vehicle on shoulder were very prevalent during the night hours and more severe than all other accident types observed (Agent and Pigman, 1989). The majority of the vehicle-on-shoulder crashes occurred between 12:00 am and 6 am, with tractor trailers overrepresented in shoulder crashes.

Chapter 3 Data Collection and Integration

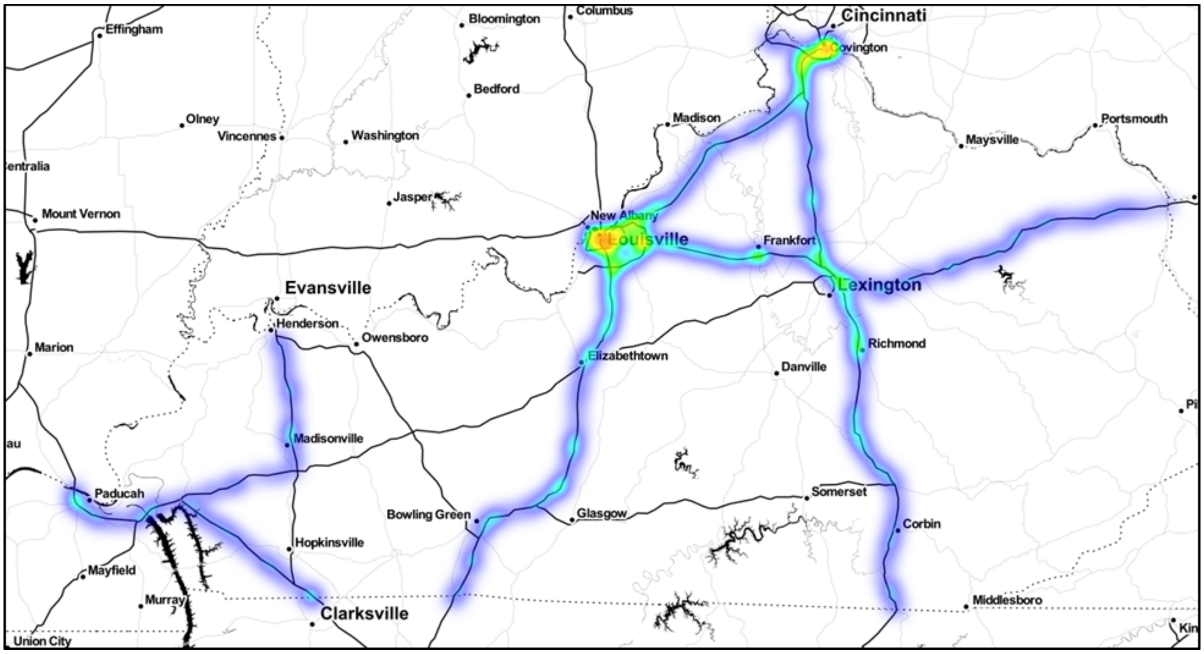
3.1 Data Collection

The study relied on three data sources: 1) Official crash reports from KyOPS (https://kyops.ky.gov), 2) archived speed data for Kentucky highways, 3) and user-generated traffic incident alerts from Waze. Data were obtained for the second half of 2018 — from July to December — for all mainline interstates in Kentucky. These data had been pre-conflated with KYTC's road network. Each data point, i.e., Waze alert or crash, thus had a unique route identifier field that defines the county, road name, and travel direction. Archived GPS-based speed data from HERE Technologies were also conflated with the highway network. These data items were available at two-minute intervals for each travel direction.

Figures 3.1 and 3.2 illustrate the spatial distribution of crashes and vehicle-on-shoulder alerts statewide and within the Louisville Metro area, respectively. Waze vehicle-on-shoulder hotspots coincide with crash hotspots. Additionally, freeway crashes are more concentrated in urban areas.

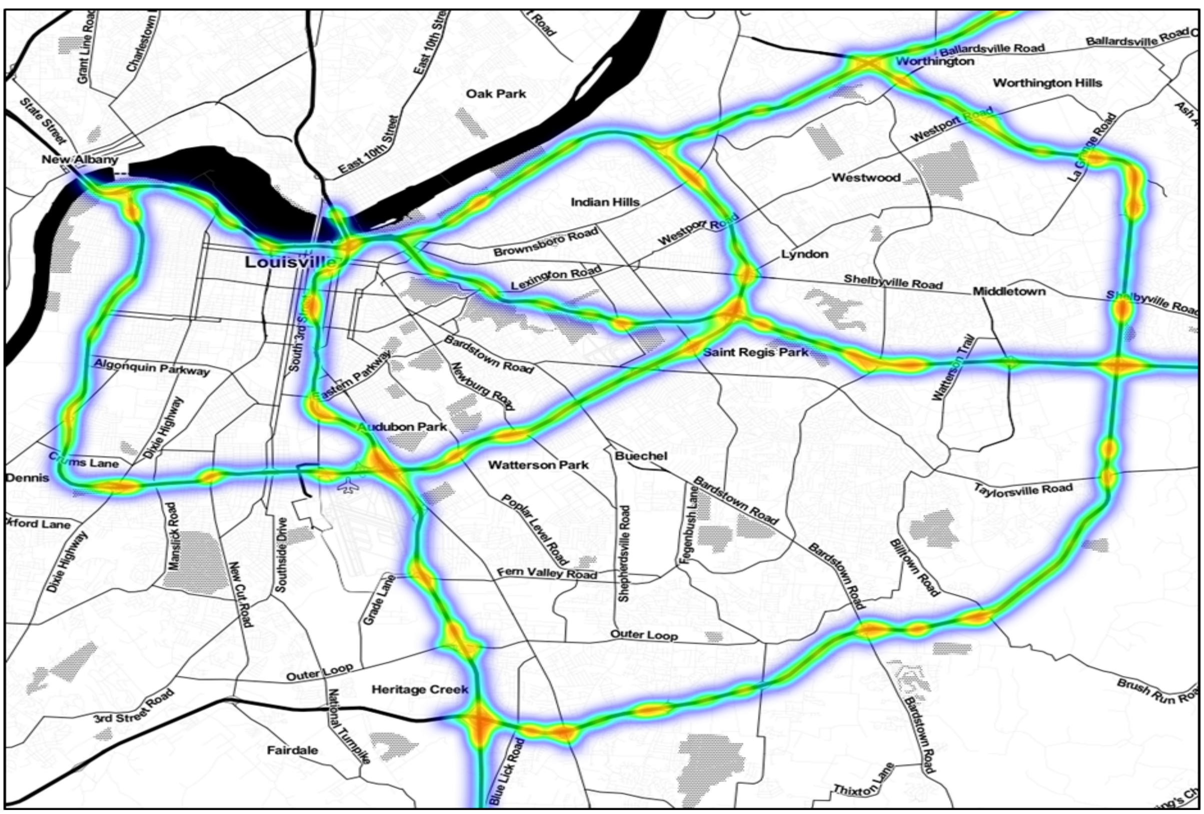


(a)

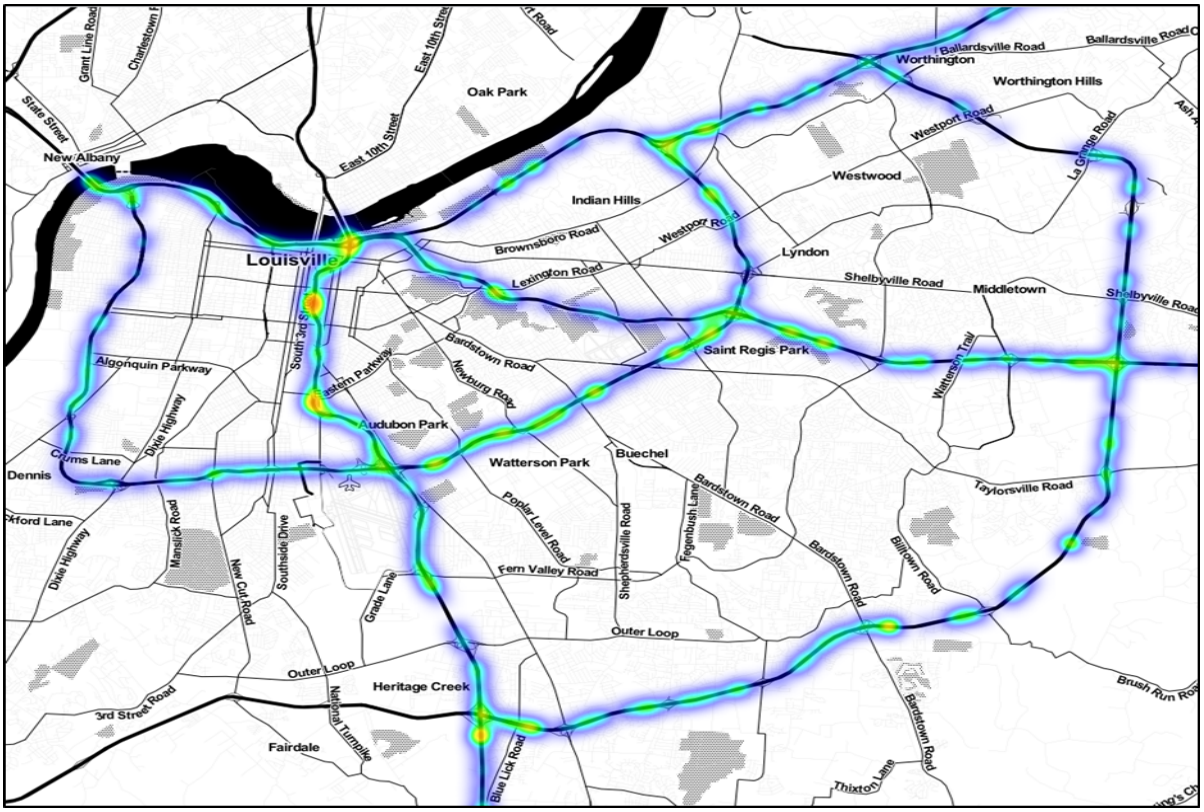


(b)

Figure 3.1 Statewide spatial distribution of (a) Waze vehicle on shoulder alerts and (b) Crashes



(a)



(b)

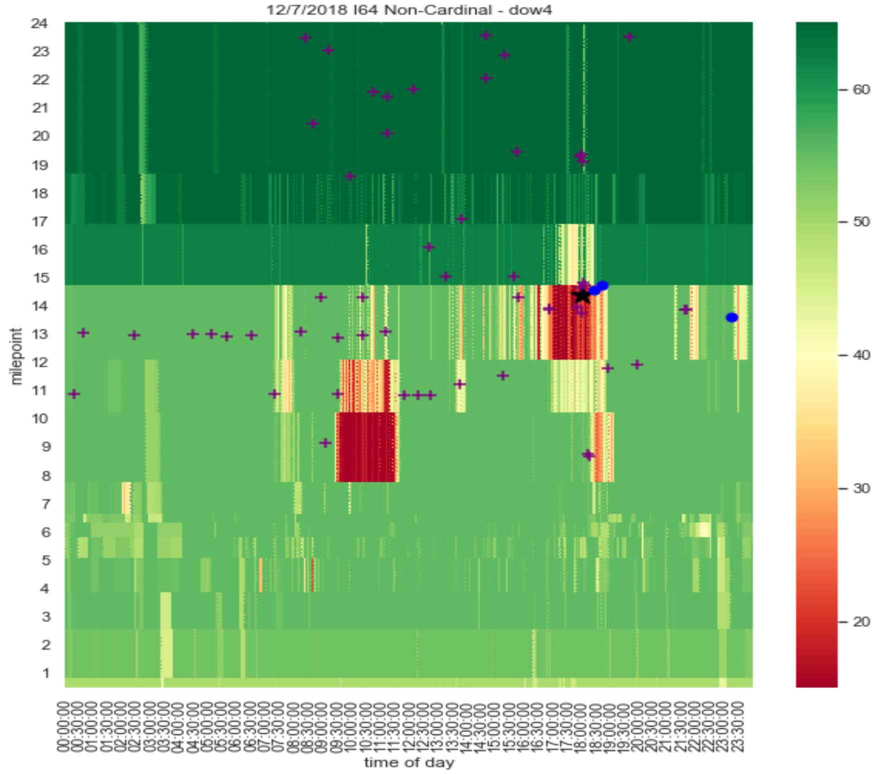
Figure 3.2 Distribution of (a) Waze vehicle on shoulder alerts (b) Crashes in Louisville Metro area

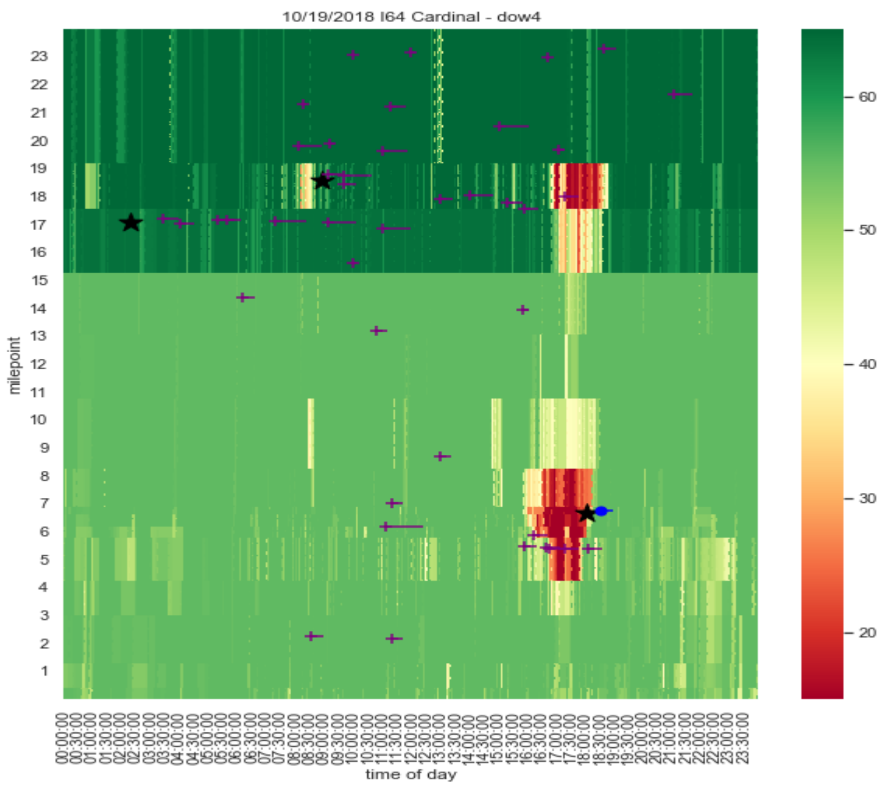
3.2 Data Integration

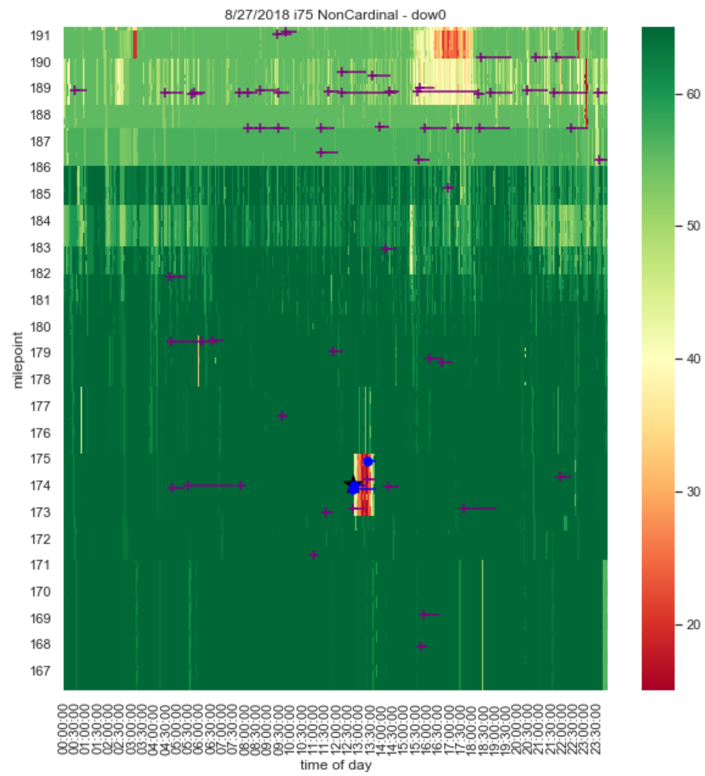
To determine whether congestion and a parked, disabled, or abandoned vehicle on the road shoulder were present before a crash occurred, the aforementioned data sources were integrated. For each day a reported crash occurred, speed data for the road corridor on which the crash occurred were queried and visualized using a heatmap that plotted speeds throughout the corridor by mile point against time of day. Then, the crashes, Waze crash alerts, and Waze vehicle-on-shoulder alerts were also plotted on the heatmap. All Waze incident alerts were charted based on their start times, with elongated symbology to signify their durations in the Waze data feed before they disappeared.

Figure 3.3 illustrates the heatmaps generated. Each plot in Figure 3.3 represents one travel direction only — either cardinal or non-cardinal. Mile points increase along the cardinal direction and decrease along the non-cardinal direction. As such, lower mile points signify upstream in the cardinal direction and downstream in the non-cardinal direction of flow. A descriptive legend of the symbology is provided below the plots.

Albeit not as frequent, there was a chain of events in which reports of parked, disabled, or abandoned vehicles on the shoulder were succeeded by congestion and crashes (Figure 3.3(a)). There were instances where Waze vehicle-on-shoulder alerts were received following a crash report (Figure 3.3(b)). These vehicle-on-shoulder alerts could potentially refer to the vehicles involved in a crash. To ensure the measured effects were for vehicles on the shoulder potentially leading to crashes, and not the other way around, only vehicle-on-shoulder alerts received prior to a crash report were considered for analysis. Additionally, only congestion prior to crash reports was considered. This was to ensure the measured effects were attributable to congestion leading to crashes and not crash-induced congestion as seen in Figure 3.3(c).


12/7/2018 I64 Non-Cardinal - dow4
milepoint
time of day
60
50
40
30
20
(a)


10/19/2018 I64 Cardinal - dow4
milepoint
time of day
60
50
40
30
20
(b)



(c)

Legend:
★ Crashes
✚ Waze vehicle on shoulder alerts
● Waze reported crashes

Figure 3.3 Heatmap Plots of speed with Waze Incidents and Crashes

A spatiotemporal approach was adopted to establish the presence of a vehicle on the road shoulder prior to crash occurrence. Different space-time thresholds were tested. The results of this step are presented in Table 3.1, showing the percentage crashes (from KyOPS) that had a vehicle-on-shoulder alert within its spatial and temporal vicinity. For example, in Jefferson County, 1,608 crashes took place between July and December 2018, and 66% of these crashes had active vehicle-on-shoulder alert(s) 30 minutes prior to the crash and within a distance of 0.5 miles upstream and downstream of the crash site.

Increasing the thresholds significantly increases the number of matches between Waze vehicle-on-shoulder alerts and crashes, as shown in **Table 3.1**. However, a spatiotemporal threshold of 0.25 miles upstream and downstream of a crash site and 30 minutes before crash occurrence was adopted for this study. This was viewed as a reasonable threshold for identifying crashes that may have resulted from the presence of a parked, disabled, or abandoned vehicle on the shoulder of a limited access highway. Based on the data shown in **Table 3.1**, about 36% freeway crashes statewide had vehicle-on-shoulder alert(s) in their vicinities. The percentages were 48% for Jefferson and Campbell Counties, much higher than the statewide rate.

Table 3.1 Percentage Crashes With a Vehicle-on-Shoulder Alert Within Its Spatiotemporal Vicinity

		Statewide	Jefferson	Fayette	Kenton	Boone	Campbell
Total Crashes (Jul-Dec 2018)		5768	1608	240	598	315	272
% Crashes with vehicle on shoulder alerts within the spatiotemporal vicinity							
30 min before	0.25 mi	36	48	28	32	35	48
	0.50 mi	54	66	46	48	52	64
	1.0 mi	72	83	65	66	70	82
30 min before and after	0.25 mi	47	59	41	44	45	59
	0.50 mi	64	76	60	59	62	76
	1.0 mi	80	89	77	75	79	89

To determine if congestion was present before a crash, GPS-based speed data at the crash location were queried two minutes prior to crash time. Congestion was considered present if the queried speed was below 45 mph.

3.3 Spatiotemporal Pattern

Aggregating crashes by hour of day, an exploratory analysis was carried out to ascertain the temporal pattern of both potential vehicle-on-shoulder–related crashes — based on the spatiotemporal criteria — and crashes not seemingly attributable to the presence of a vehicle on the shoulder. On Kentucky's interstate highways, more crashes occurred during the peak hour times of the day than any other time. The same is true for crashes potentially involving a vehicle on the shoulder (Figure 3.4). An evaluation of the spatial distribution of potential vehicle-on-shoulder and congestion-related crashes indicated more of these crashes occurred in urban areas, particularly in northern Kentucky (Cincinnati Metro Area) and the Louisville Metro Area. The spatial distribution is presented in Figures 3.5, 3.6 and 3.7.

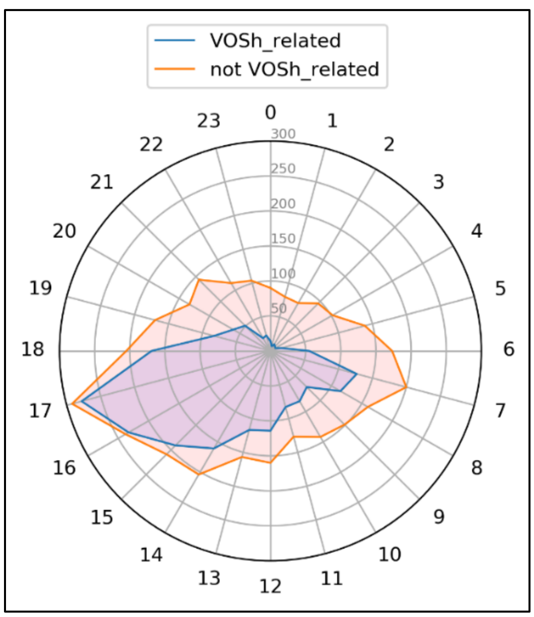
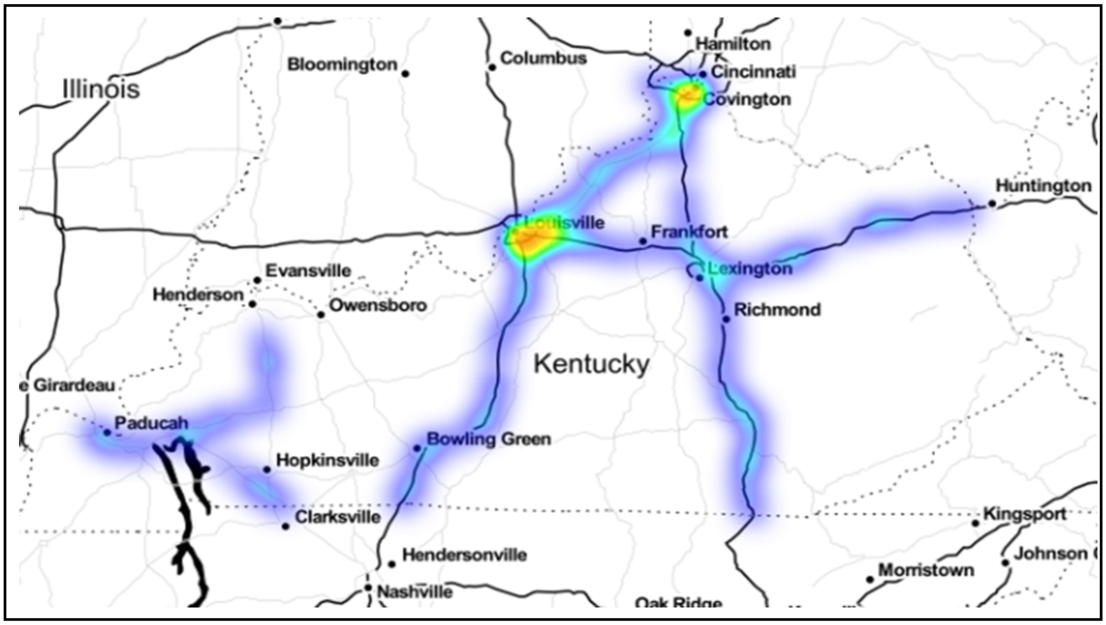
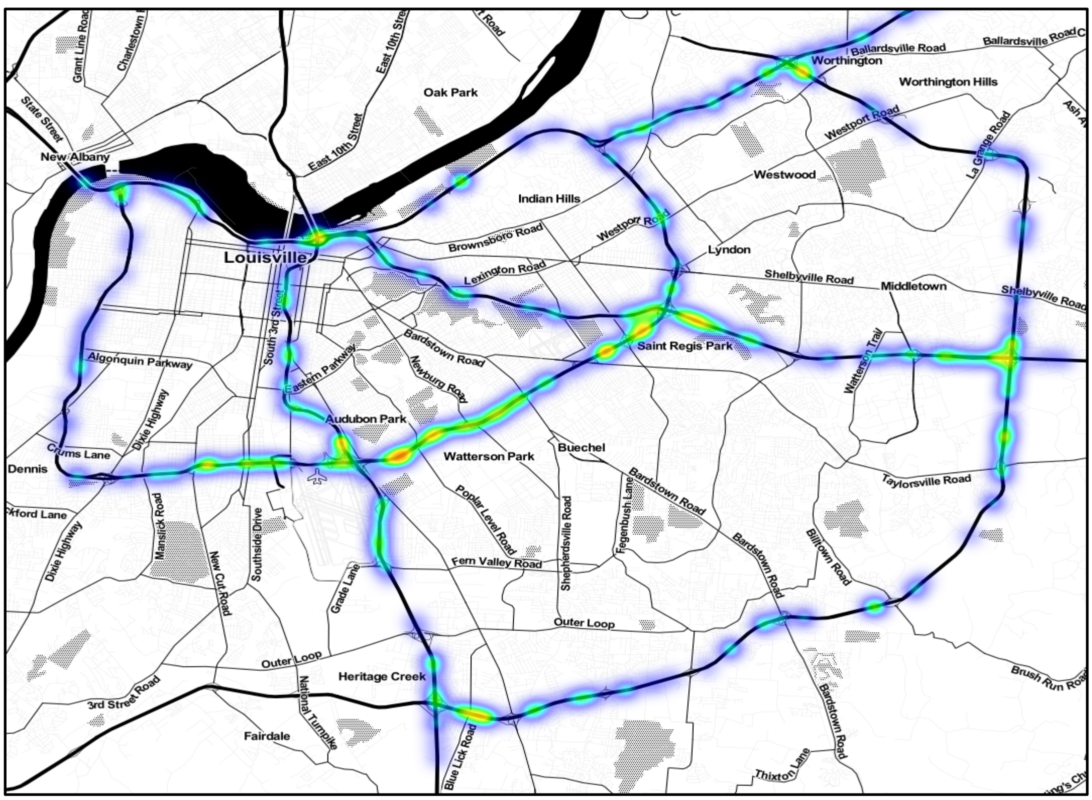


Figure 3.4 Temporal Analysis of Crashes by Hour of Day

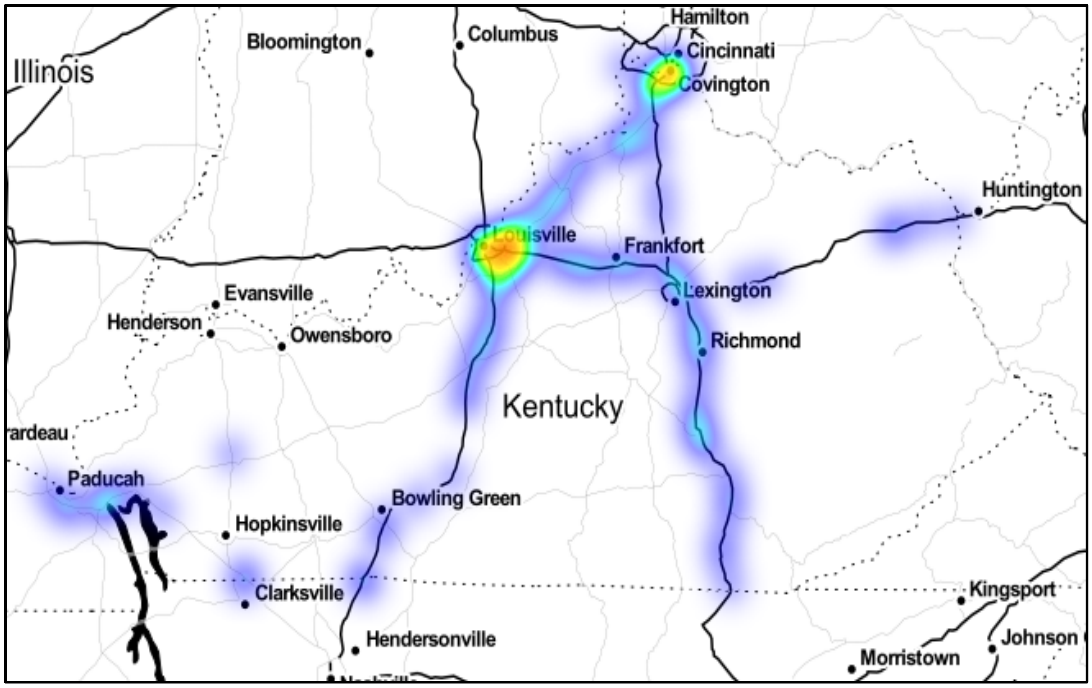


(a)

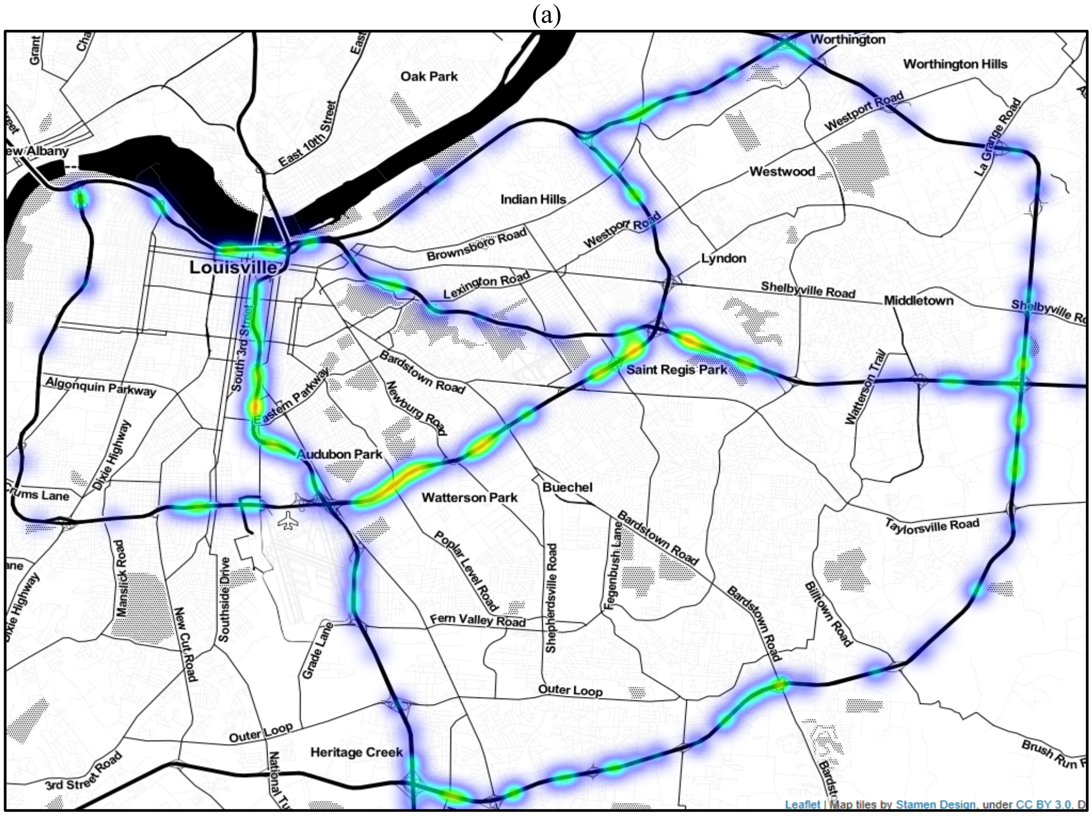


(b)

Figure 3.5 Spatial Distribution of Vehicle-on-Shoulder–Related Crashes in (a) Kentucky (b) Louisville Metro

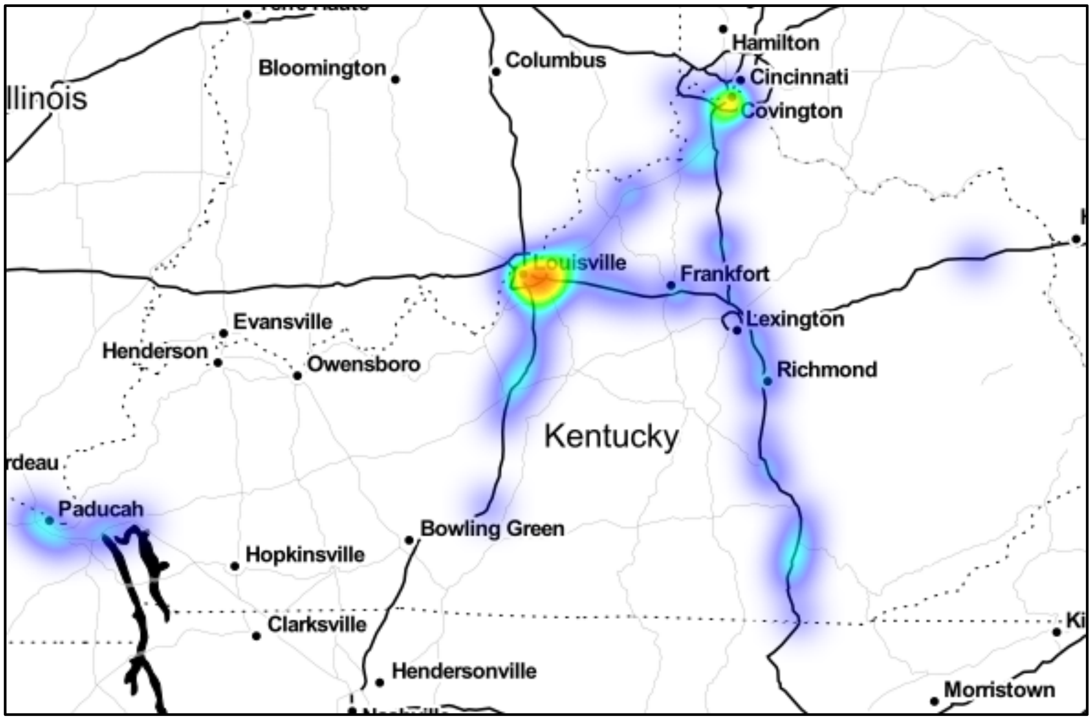


(a)

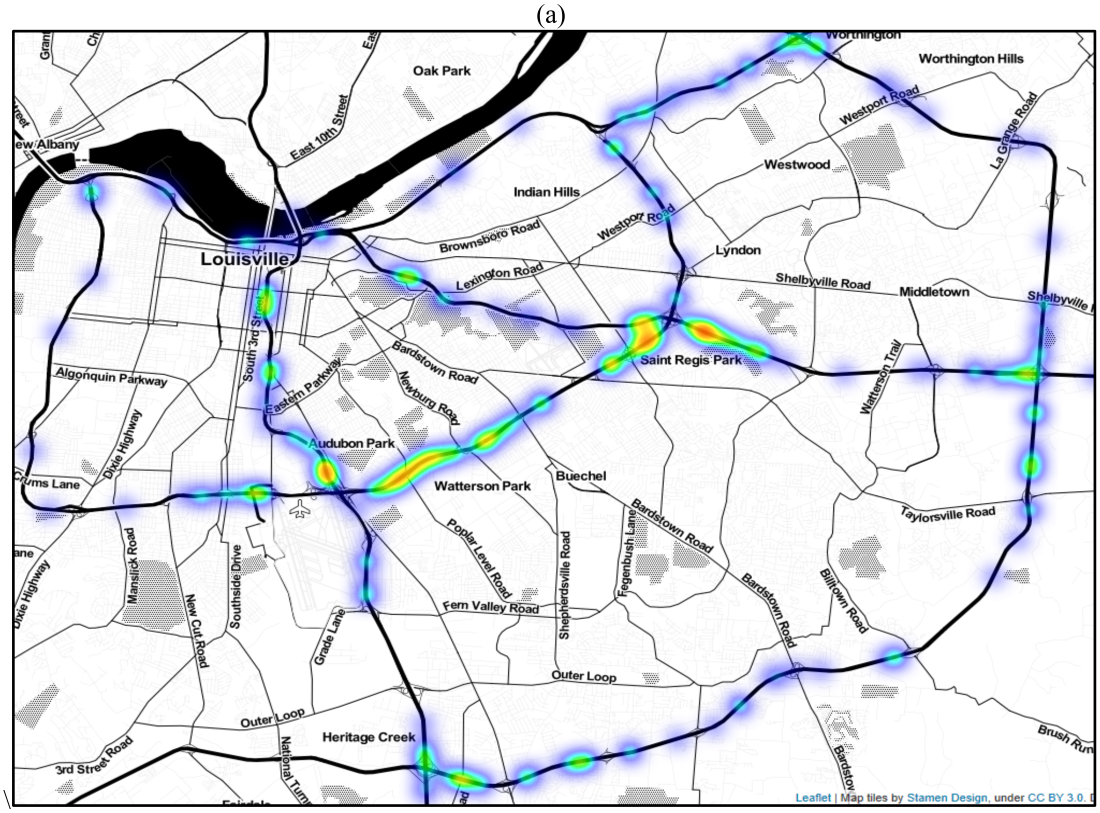


(b)

Figure 3.6 Spatial distribution of Congestion-Related Crashes in (a) Kentucky (b) Louisville Metro Area



(a)



(b)

Figure 3.7 Spatial Distribution of Congestion- and Vehicle-on-Shoulder–Related Crashes in (a) Kentucky (b) Louisville Metro Area

Chapter 4 Correlation/Causation Between Vehicle-on-Shoulder Alert and Crash

4.1 Exploring Correlation

Although crashes may have multiple causes, these causes may generally be classified as one of three major factors: human factors, environmental factors, or vehicle factors. To assess the potential contribution of factors, including vehicles on the shoulder and congestion, to crashes, and the degree of their correlation with limited access highway crashes, association rule mining was employed.

Association rule mining identifies correlations or associations among a set of items in a large data set based on a support-confidence framework. Based on Agrawal et al. (1993), it may be defined as follows.

Let I = $\{i_1, i_2, \ldots, i_m\}$ be a universal set of crash-related factors, including human, environmental, and vehicle factors. Let D = $\{c_i, c_{i+1}, \ldots, c_n\}$ be a set of the crashes from the crash data, where each crash has a unique crash ID (Cid) and an item set (C-itemset) consisting of the factors related to this specific crash. Let $X \subseteq I$, $Y \subseteq I$ each be a subset of the universal crash contributory factors. An association rule is the implication $X \rightarrow Y$ such that $X \cap Y = \emptyset$, $p(X) \neq 0$ and $p(Y) \neq 0$ where X is the antecedent and Y the consequent. The significance or effectiveness of a rule can be measured by indicators such as support, confidence, and lift.

Support of a rule here is computed using Equation 1, which refers to how frequently the antecedent and consequent of a rule occur together in the crash data.

(Eq. 1) $$\text{Support } (X \rightarrow Y) = \frac{P(X \cap Y)}{N}$$

Confidence (see Equation 2) refers to the strength of a rule's implication and is the proportion of crashes involving contributing factor X that also contain Y.

(Eq. 2) $$\text{Confidence } (X \rightarrow Y) = \frac{P(X \cap Y)}{N}$$

Although the support-confidence framework is a common model for mining association rules, it does not provide a test for identifying the correlation between two item sets (Zhang and Zhang, 2002). As such, the lift measure, which measures the dependence between factors, was employed as a third measure. With values ranging from 0 to ∞, a lift value of 1 indicates factors are independent, values greater than 1 denote positive dependence, and values less than 1 indicate negative dependence between factors. Mathematically, lift is computed as:

(Eq. 3) $$\text{Lift } (X \rightarrow Y) = \frac{P(X \cap Y)}{P(X) \times P(Y)}$$

Human factors extracted for this study include driver impairment, distraction, inattention, driving too fast for conditions, improper vehicle maneuvers, failure to yield right of way, and following too closely. Environmental factors included roadway character — presence of curves and grades, inclement weather, poor visibility based on the lighting condition field in crash reports, animal/debris, water pooling, slippery road surface, and construction work zone. Additionally, the presence of traffic congestion and vehicles on the shoulder before the crash were established from GPS speed data and Waze data, respectively. Vehicular factors from crash reports were classified as vehicle defects. Each of contributing factor was encoded as a binary variable, indicating its presence or absence for each crash.

Using the 'MLxtend' python package (Raschka, 2018), the Apriori algorithm was employed with a minimum support of 5% and a minimum lift of 1 so that only positive correlations between factors were reported.

Among the large set of crash contributing factors, the most frequently occurring crash factors in order of decreasing support were poor visibility, presence of a vehicle on the shoulder, inclement weather, improper driver maneuver, driver inattention, congestion, slippery road surfaces and water pooling, and the presence of grades and curves. The first five factors indicate that at night or during inclement weather conditions when a vehicle is present on the shoulder of the road that a crash is likely to occur if a driver is inattentive or makes an inappropriate steering maneuver. The top extracted association rules of interest are presented in Tables 4.1 and 4.2. They are sorted by lift value in descending order. These rules are split into two-item and three-item rules and reveal the underlying correlations between the crash contributing factors.

Antecedent support and consequent support in Tables 4.1 and 4.2 refer to the proportion of crashes involving the antecedent and consequent, respectively. The higher lift association rules from Table 4.1 indicate that human and environmental factors are highly correlated with crashes. For example, the first rule in Table 4.1 indicates that when a driver drives too fast on slippery road surfaces or locations where water has pooled on the road, a crash is very likely to occur. These two-item set of rules help us understand the relationships between individual contributing factors. Three-item rules, however, clarify the interactions between more than two factors, particularly if they have a higher lift value, which indicates a stronger correlation between the interaction of those factors and crashes. The two-item rules suggest that congestion coupled with the presence of a vehicle on the shoulder of the road were correlated with crashes, accounting for 11.7% of limited access highway crashes over the study period. When two contributing factors were present — vehicle on shoulder presence and congestion — and combined with driver inattention crashes were more likely, as seen in Table 4.2, with a higher lift value compared to the presence of vehicles on shoulder and congestion only. In 44.3% of crashes where a vehicle on shoulder and congestion may have contributed to the crash, driver inattention was also a contributing factor. Again, it should be noted that the rules only suggest correlations between a combination of factors and crashes.

Antecedents	Consequents	Antecedent Support	Consequent Support	Support	Confidence	Lift
{'DrivingTooFast'}	{'SlipperySurf/WaterPool'}	0.079	0.231	0.051	0.646	2.796
{'BadWeather'}	{'DrivingTooFast'}	0.323	0.079	0.060	0.185	2.338
{'Congestion'}	{'FollowTooClose'}	0.257	0.098	0.052	0.203	2.081
{'Animal/Debris'}	{'BadVisibility'}	0.085	0.375	0.058	0.679	1.811
{'Congestion'}	{'Inattention'}	0.257	0.268	0.112	0.435	1.624
{'SlipperySurf/WaterPool'}	{'ImproperManeuver'}	0.231	0.273	0.086	0.373	1.366
{'ImproperManeuver'}	{'BadWeather'}	0.273	0.323	0.114	0.418	1.295
{'Vehicle on Shouder'}	{'Congestion'}	0.367	0.257	0.117	0.318	1.237
{"Vehicle on Shouder '}	{'Inattention'}	0.367	0.268	0.111	0.303	1.130

Table 4.1 Three-Item Association Rules

Antecedents	Consequents	Antecedent Support	Consequent Support	Support	Confidence	Lift
{'BadWeather', 'CurvePresent'}	{'SlipperySurf/WaterPool'}	0.071	0.231	0.051	0.726	3.142
{'BadWeather'}	{"Vehicle on Shouder ', 'SlipperySurf/WaterPool'}	0.323	0.070	0.064	0.200	2.833
{'BadWeather'}	{'ImproperManeuver', 'SlipperySurf/WaterPool'}	0.323	0.086	0.078	0.242	2.812
{'Congestion'}	{"Vehicle on Shouder ', 'Inattention'}	0.257	0.111	0.052	0.201	1.809
{"Vehicle on Shouder ', 'Congestion'}	{'Inattention'}	0.117	0.268	0.052	0.443	1.652

4.2 Exploring Causation

Association rules imply correlation *but not* causation. To further explore the nature of vehicle-on-shoulder crashes and infer the causal relationship between the presence of a vehicle on the shoulder or congestion and crashes, crash narratives were reviewed. Reviewing crash narratives manually requires substantial human resources. Consequently, a combination of crash code analysis and keyword searches of crash narratives were conducted to identify potential vehicle-on-shoulder and congestion-related crash narratives for review.

4.2.1 Congestion Related Narrative Review

To identify the crashes potentially caused by congestion and flag their narratives for review, several steps were taken. First, the crash narratives that contained any one of the following words or phrases — *hydroplane*, *animal*, *deer*, *stop sign*, *standing water*, *pool of water*, and *patch of ice* — were excluded from further consideration, as it is unlikely these crashes were the result of traffic congestion. Second, before the keyword search word tenses were converted to present to improve search results. For example, *built*, *backed*, and *slowing* were converted to *build, back, and slow,* respectively. During the keyword search, crash narratives had to contain one of the keywords *congestion* or *congested;* or at least one sentence had to contain the word *traffic* and any of the following keywords: *slow, stop, build up, backup, heavy, halt, standstill, sitting.* In addition, narratives with at least one sentence containing the word *vehicle* and any of the keywords *slow, stop, build up, backup* were flagged.

The above process flagged 4,081 crashes as potential congestion-caused crashes. Given the large number of returns, a subset of 351 crashes was randomly selected and manually reviewed. After reviewing their narratives, 266 (75.8%) crashes were confirmed as having congestion as a contributing factor. Two narratives with personally identifiable information removed are provided as examples.

- *Unit 2 was traveling in slow moving EB traffic congestion on Interstate XX in the right lane just prior to the XX Freeway. Unit 1 was traveling EB in the right lane approaching stopped traffic. According to a statement obtained from the operator of Unit 1 he attempted to brake but did not stop in time. As a result; Unit 1 struck the rear of Unit 2 with the front of Unit 1. When officer arrived on-scene both vehicles had moved to the right shoulder of roadway.*
- *Units 1 and 2 were West on I-XX with traffic backing up. Unit 1 struck the rear of unit 2 with the front of his vehicle when he couldn't get stopped in time.*

Extrapolating from the sample suggested that across the entire set of 4,081 crashes, 3,102 were likely to have had congestion as a contributing factor. Thus in 23.1% of the 13,414 interstate and parkway crashes over the study period congestion likely was a contributing factor. This corroborates results from association rule mining analysis in which 25.7% of crashes were congestion related, demonstrating a causal relationship between congestion on roadways and crashes.

4.2.2 Vehicle on Shoulder Related Narrative Reviews

Using keyword searches of crash narratives to identify potentially relevant crashes, a crash was considered a likely vehicle-on-shoulder–related crash and its narrative flagged for review if its crash narrative contained at least one of the following keywords: *parked, abandoned, unoccupied*; or at least one sentence contained one keyword from each of these two lists: 1) *shoulder, emergency lane, emergency strip*, and 2) *disabled, stationary, pulled over*.

This search yielded 512 crashes. First, the narratives for these crashes were reviewed to identify any that included the chain of events where a vehicle on the shoulder caused congestion or slowing down of traffic which then led to a crash. Of these crashes, only six were found where vehicles on the shoulder were specified as contributing to congestion and subsequent crashes. A portion of the narratives from several of these crashes are provided as examples.

- *Both vehicles traveling in the right lane when ahead of unit 2 traffic slowed suddenly due to an abandoned vehicle that was on the fog lane traffic slowed to get by it. Unit 2 had slowed when he was struck in the rear; operator of unit 1 said that when he saw traffic coming to a stop he applied the brakes unable to stop slid into the rear of unit 2.*
- *Unit 2 was traveling Northbound on Interstate XX near the XXMM in the left lane. Unit 1 was behind unit 2 in the left lane traveling in the same direction. Unit 2 operator stated he was slowing with traffic as it passed an emergency vehicle parked on right shoulder. Unit 1 made contact with the rear of Unit 2. Damage to Unit 1 was in the front. Unit 2 came to final rest on left shoulder. Unit 1 came to final rest in median approx. 300 feet behind Unit 2.*
- *This officer was dispatched to a two vehicle injury accident on eastbound I-XX in the area of the XX mile marker. Unit 2 was parked in the right emergency shoulder due to experiencing a flat front right tire. Unit 2 had its hazard lights activated. Unit 1 was traveling east bound in the far right lane. Unit 1 encountered slowed traffic in its lane and in reaction; the driver veered to the right to avoid a collision. In doing so;*

unit 1 drove onto the right emergency shoulder where it collided with the rear of unit 2; and then the guardrail face of the concrete bridge wall. A previous occupant of unit 2 was standing outside of the vehicle at the time and was struck; and killed as a result of the collision.

It should be noted that most narratives associated with crashes and congestion did not generally specify the cause of the congestion. Therefore, we were only able to verify these six occurrences of the sequence of events: vehicle-on-shoulder alert – congestion – crash.

Additional crash narratives were reviewed to see if any crashes were indirectly caused by a vehicle-on-shoulder incident, i.e., the vehicle on shoulder was not an involved party in the crash but was cited as a potential factor in the crash report. No such case was identified through reviewing 31 randomly chosen crash reports from the flagged 512 crashes.

Though this review found only six crashes that met the ascribed criteria, it is likely that there were more than just the six crashes that followed the pattern of vehicle on shoulder leading to congestion leading to crash. However, absent the full chain of events being captured in the crash narrative, it is impossible to determine exactly how many there were. Based on this analysis, the high correlation between vehicles on the shoulder and crashes did not appear to suggest causation at the same rate.

Chapter 5 Crashes Involving Vehicles on Shoulder

This chapter provides more detailed analysis by classifying and examining factors involved in crashes where a vehicle on the shoulder was directly involved. The findings from the analysis are compared to those of Agent and Pigman (1989).

5.1 Procedure

For this analysis, only crashes that occurred along interstates and parkways were included. A total of 13,414 crashes occurred on interstates and parkways — 8.5% of crashes statewide for 2018. Of these, 12,623 (94%) occurred on the interstate and parkway through lanes, while 791 (6%) occurred on ramps or other lane types.

Several methods were used to identify interstate and parkway crashes involving a vehicle on the shoulder. First, crash codes were used to identify potentially relevant records. In the field *Location of First Event*, crashes coded (04) Outside Shoulder-Left, (05) Outside Shoulder-Right, and (06) Shoulder were selected. Additionally, in the *Manner of Collision* field, crashes coded (09) Single Vehicle were removed. This combination of codes yielded 225 results. Crash narratives were obtained and reviewed to confirm which crashes involved a vehicle on the shoulder. After removing non-relevant records, 175 crashes remained.

To identify additional potentially relevant crashes, results from the narrative keyword searches for vehicle-on-shoulder (described in Section 4.2.2) were considered. These searches had identified 512 crashes, some of which, though, had already been reviewed through the crash code identification. Of the new crashes, 65 were relevant.

In all, 240 crashes were identified as involving a vehicle on the shoulder — 1.8% of all crashes on interstates and parkways. This, remarkably, is the identical percentage of vehicle-on-shoulder crashes found by Agent and Pigman in a similar study covering the years 1985 to 1987.

Of the 240 vehicle-on-shoulder crashes, 208 occurred on interstates and 32 on parkways. Vehicle-on-shoulder crashes on interstates comprised 1.8% of all interstate crashes, which was slightly lower than the percentage of vehicle-on-shoulder crashes for parkways, which was 2.3% of all parkway crashes.

5.2 Reason and Explanation for Vehicle on Shoulder Crash

Agent and Pigman (1989) used six categories to classify vehicle-on-shoulder crashes: a vehicle stopped on the shoulder, a vehicle pulling from the shoulder, a vehicle pulling onto the shoulder, a vehicle moving on the shoulder, an occupant from a vehicle stopped on the shoulder, or a crash caused by a vehicle on the shoulder even though that vehicle was not involved. Agent and Pigman manually searched three years of crash records (1985 to 1987) to identify and analyze vehicle-on-shoulder crashes. Agent and Pigman's typology was applied to the 2018 crash data for comparison.

The most frequent type of crash was a vehicle stopped on the shoulder (78%). The next most frequent was a vehicle pulling onto the shoulder (9%). Many of these crashes occurred when a driver experienced sudden slowing/stopping of traffic ahead and then pulled onto the shoulder to avoid a collision only to then crash into another vehicle pulling onto the shoulder for the same reason. The third most frequent type was a vehicle moving on the shoulder (5%). Such collisions included work vehicles, emergency response vehicles, and aggressive drivers trying to avoid congestion. The fourth most frequent was a vehicle pulling from the shoulder back to the roadway (4%). The remaining types of crashes included motorist outside vehicle (3%) and secondary crashes (1%). Figure 5.1 summarizes vehicle-on-shoulder crash types.

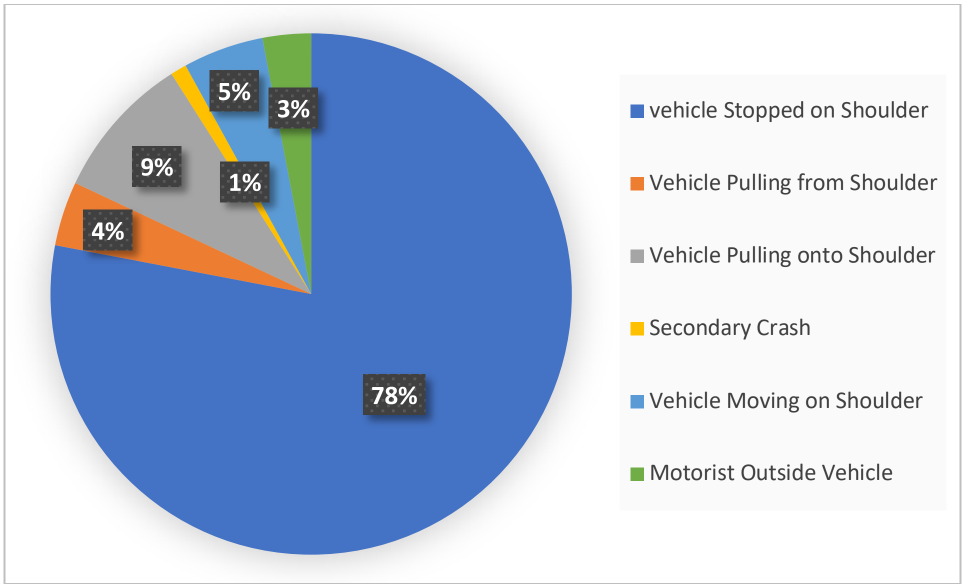


Figure 5.1 Types of Vehicle-on-Shoulder Crashes

For each of the 240 vehicle-on-shoulder crashes, crash narratives were reviewed to ascertain, where possible, the reason and explanation for why a vehicle was on the shoulder. This effort, again, mirrored the work of Agent and Pigman and used many of the same categories. Table 5.1 summarizes the reason and explanation derived for these crashes.

The most common reason for a vehicle on the shoulder was an emergency, which occurred in 119 crashes, representing 50% of vehicle-on-shoulder crashes. The most common explanations for the emergency were other crash (23), mechanical problem (18), swerved to avoid stopped traffic (18), and tire problem (14).

The next most common reason was work, which accounted for 18 crashes. In 14 of these, the vehicle on the shoulder was a work vehicle (e.g., a road construction vehicle).

The third most common reason was leisure, which accounted for 12 crashes. Of these, 9 involved sleeping, which generally entailed a commercial vehicle parked on the shoulder while the operator rested in the sleeper cab.

For 91 crashes, the reason why the vehicle was on the shoulder was unknown or unspecified. In 45 of these crashes, the vehicle was abandoned, likely due to mechanical problems. For 39 crashes, the narrative did not specify why the vehicle was on the shoulder.

Table 5.1 Reason and Explanation for Why Crash-Involved Vehicles Were on the Shoulder

REASON	EXPLANATION	Crashes
Emergency	Other Crash	23
	Mechanical Problem	18
	Swerved to Avoid Stopped Traffic	18
	Tire Problem	14
	Abandoned Vehicle	12
	Assist Other Driver	7
	Bad Weather	6
	Police/EMS Vehicle	5
	Out of Gas	5
	Check on Vehicle	4
	Passing in Emergency Lane	3
	Pickup Item that Fell from Vehicle	2
	Rest	1
	Unknown	1
	TOTAL	119
Leisure	Sleeping	9
	Rest	1
	Restroom	1
	Use phone	1
	TOTAL	12
Work	Work Vehicle	14
	Police/EMS Vehicle	4
	TOTAL	18
Unknown	Abandoned Vehicle	45
	Unknown	39
	Passing in Emergency Lane	7
	TOTAL	91

5.3 Crash Severity

Table 5.2 summarizes crash severity for vehicle-on-shoulder crashes. Of the 240 crashes, 9 resulted in a fatality, 67 resulted in an injury, 158 were property damage only, and 6 involved a hit and run where injury severity could not be determined. The 9 fatal crashes represented 10.7% of all fatal crashes on interstates/parkways in 2018. This was similar to the 11.1% of fatal crashes that Agent and Pigman found in their study.

Compared to all interstate/parkway crashes, the severity of vehicle-on-shoulder crashes was higher. For shoulder crashes, 3.75% resulted in a fatality, compared to 0.63% of all interstate/parkway crashes. Additionally, 27.9% of shoulder crashes resulted in an injury, compared to 16.3% of all interstate/parkway crashes.

Table 5.2 Crash Severity for Vehicle on Shoulder Crashes

CRASH SEVERITY	Crashes	% of all VOS crashes	% Statewide
Fatal	9	3.8%	0.6%
Injury	67	27.9%	16.3%
Property Damage Only	158	65.8%	82.6%
Unknown	6	2.5%	0.4%

5.4 Time of Crash

Table 5.3 summarizes the time of day when vehicle-on-shoulder crashes occurred. While vehicle-on-shoulder crashes occurred throughout the day, most happened during the high-volume traffic hours of 6:00 am to 9:00 am and 3:00 pm to 6:00 pm. A total of 95 vehicle-on-shoulder crashes (40%) occurred during morning and afternoon rush hours.

Comparing vehicle-on-shoulder crashes to all interstate/parkway crashes, a higher proportion of vehicle-on-shoulder crashes occurred at night. Between the hours of 12:00 am and 6:00 am, 22% of vehicle-on-shoulder crashes occurred, compared to 11% of all interstate/parkway crashes.

Table 5.3 Time of Day when Vehicle on Shoulder Crashes Occurred

TIME OF CRASH	Crashes	% of all VOS crashes	% Statewide
Midnight to 3:00 AM	22	9%	5%
3:00 AM to 6:00 AM	30	13%	6%
6:00 AM to 9:00 AM	45	19%	17%
9:00 AM to Noon	25	10%	12%
Noon to 3:00 PM	25	10%	15%
3:00 PM to 6:00 PM	50	21%	24%
6:00 PM to 9:00 PM	24	10%	13%
9:00 PM to Midnight	19	8%	8%

5.5 Crash Location

Vehicle-on-shoulder crashes were plotted to highway segments using GIS to further analyze their locational attributes. Of the 240 vehicle-on-shoulder crashes, 132 (55%) took place along rural highway segments (Table 5.4). Of those along urban highway segments, 62 (26%) were in Louisville Metro/Jefferson County and 29 (12%) were in northern Kentucky.

Table 5.4 Location of Vehicle on Shoulder Crashes by Urban Area

CRASHES BY URBAN AREA	Crashes	% of all VOS crashes	% Statewide
Rural	132	59.2%	44.3%
Louisville/Jefferson County	62	27.0%	31.3%
Northern KY	29	12.4%	16.5%
Lexington-Fayette County	7	3.0%	2.3%
Bowling Green	3	1.3%	1.0%
London-Corbin	3	1.3%	0.6%
Elizabethtown-Radcliff	1	0.4%	0.6%
Hopkinsville	1	0.4%	0.3%
Mount Sterling	1	0.4%	0.1%
Berea	1	0.4%	1.1%

By highway, I-75 had the highest number of vehicle-on-shoulder crashes, followed by I-65, I-71, and I-64 (Table 5.5).

Table 5.5 Location of Vehicle on Shoulder Crashes by Route

CRASHES BY ROUTE	Crashes	% of all VOS crashes	% Statewide
I-75	56	23.3%	24.3%
I-65	42	17.5%	16.0%
I-71	25	10.4%	6.9%
I-64	24	10.0%	14.8%
I-264	21	8.8%	8.9%
I-275	14	5.8%	5.1%
I-265	12	5.0%	5.3%
I-24	8	3.3%	4.2%
WK-9001	7	2.9%	2.2%
WN-9007	7	2.9%	1.3%
I-69	5	2.1%	2.1%
BG-9002	5	2.1%	1.7%
LN-9008	3	1.3%	1.1%
EB-9004	3	1.3%	1.0%
AU-9005	2	0.8%	0.5%
HR-9006	2	0.8%	0.8%
JC-9003	2	0.8%	1.1%
I-471	1	0.4%	2.2%
KY-9009	1	0.4%	0.3%

Highway segments defined by route and county were also examined for vehicle-on-shoulder crashes (Table 5.6). Using this segmentation, I-264 in Jefferson County had the most crashes — 21. It was followed by I-65 in Bullitt County (16), I-265 in Jefferson County (12), and I-71 in Jefferson County (10).

Table 5.6 Location of Vehicle on Shoulder Crashes by Route/County

CRASHES BY ROUTE/COUNTY	Crashes	% of all VOS crashes	% Statewide
I-264 Jefferson	21	8.8%	8.9%
I-65 Bullitt	16	6.7%	2.7%
I-265 Jefferson	12	5.0%	5.3%
I-71 Jefferson	10	4.2%	2.3%
I-65 Jefferson	9	3.8%	6.3%
I-75 Fayette	9	3.8%	3.4%
I-75 Kenton	9	3.8%	6.7%
I-75 Boone	8	3.3%	3.4%
I-275 Kenton	7	2.9%	2.2%
I-75 Rockcastle	7	2.9%	2.1%
I-75 Scott	7	2.9%	2.3%
I-64 Shelby	6	2.5%	2.4%
I-65 Warren	6	2.5%	1.8%
I-75 Grant	6	2.5%	1.8%
I-275 Boone	5	2.1%	0.9%
I-64 Jefferson	5	2.1%	6.4%
I-75 Whitley	5	2.1%	1.3%
I-71 Gallatin	4	1.7%	0.9%
I-71 Henry	4	1.7%	0.8%
I-75 Madison	4	1.7%	2.0%

5.6 Crash Conditions

The roadway character of locations where vehicle-on-shoulder crashes occurred was analyzed and compared to all interstate/parkway crashes (Table 5.7). Most vehicle-on-shoulder crashes (68%) occurred on straight-level sections of highway. For each roadway character type, the percentage of all vehicle-on-shoulder crashes was similar to the percentage of all interstate/parkway crashes statewide.

Table 5.7 Roadway Character of Vehicle-on-Shoulder Crashes

ROADWAY CHARACTER	Crashes	% of all VOS crashes	% Statewide
Straight-Level	162	68%	63%
Straight-Grade	35	15%	16%
Straight-Hillcrest	10	4%	3%
Curve-Level	19	8%	9%
Curve-Grade	12	5%	7%
Curve-Hillcrest	2	1%	2%

Roadway condition was also analyzed (Table 5.8). Most vehicle-on-shoulder crashes (67 percent) occurred on dry roadways; 22 percent occurred on wet roadways; and 11 percent took place on snow/icy road conditions. Compared to all interstate/parkway crashes statewide, vehicle-on-shoulder crashes were twice as likely to occur on snowy/icy roadways.

Table 5.8 Roadway Condition of Vehicle-on-Shoulder Crashes

ROADWAY CONDITION	Crashes	% of all VOS crashes	% Statewide
Dry	160	67%	62%
Wet	53	22%	31%
Snow/Ice	26	11%	5%

Table 5.9 summarizes the light conditions for vehicle on shoulder crashes. A slight majority of these crashes (52%) occurred during daylight, while 41% occurred in darkness, and 6% occurred during dusk/dawn. The 41% of crashes occurring at night was considerably higher than the 29% of all interstate/parkway crashes occurring at night.

Table 5.9 Light Condition of Vehicle-on-Shoulder Crashes

LIGHT CONDITION	Crashes	% of all VOS crashes	% Statewide
Daylight	125	52%	65%
Dawn	10	4%	3%
Dusk	5	2%	2%
Darkness-Lighted	42	18%	13%
Darkness-Not Lighted	56	23%	16%

5.7 Human Factors

Human factors from crash reports were examined. *Inattention* was the most common human factor, cited in 78 of the 240 vehicle-on-shoulder crashes. Other common factors included *Not Under Proper Control* (60 crashes), *Misjudged Clearance* (18 crashes), *Distraction* (16 crashes), *Too Fast for Conditions* (14 crashes), *Alcohol Involvement*, (9) and *Drug Involvement* (7). In 49 of the 240 shoulder crashes, no human factors were recorded.

Human factors can be grouped into *Aggressive Driving*, *Distracted Driving*, or *Impaired Driving*. Of the 240 vehicle-on-shoulder crashes, *Aggressive Driving* factored into 39 (16.3%) . For all interstate/parkway crashes, aggressive driving factored into 20.6%. Distracted Driving factored into 91 crashes (37.9%), while distracted driving factored into 29.6% of all interstate/parkway crashes. Impaired Driving factored in 14 of the vehicle-on- shoulder crashes (5.8% of all vehicle-on-shoulder crashes), while 2.6% of all interstate/parkway crashes involved impaired driving.

5.8 Vehicles

Trucks were involved in 72 of the 240 vehicle-on-shoulder crashes. In 52 crashes, a truck was the mainline vehicle, while in 44 of the crashes, a truck was on the shoulder. In 24 crashes, a truck was both the mainline vehicle and a vehicle on the shoulder. Trucks were involved in 30% of all vehicle on shoulder crashes, compared to 14.8% of all interstate/parkway crashes.

Chapter 6 Conclusion

6.1 Research Findings

This study evaluated relationships between vehicles on the shoulder, congestion, and crashes. Researchers obtained three relevant interstate datasets and integrated them using a spatiotemporal approach. Analysis found that in about 36% of crashes, a vehicle was present on the roadway shoulder in the vicinity of the crash (defined as 0.25 miles upstream and downstream of a crash site no more than 30 minutes before crash occurrence). Expanding the spatiotemporal window significantly increased this percentage. For example, 54% of crashes had vehicle on shoulder in the vicinity when the spatial threshold was increased to 0.5 miles. In addition, congestion was associated with roughly 25% of crashes. In 11.7% of crashes, both a vehicle on the shoulder and congestion were present, indicating a high correlation between these conditions and crashes. The subsequent association rule mining analysis confirmed the association between vehicles on shoulders, congestion, and crashes was statistically significant. The level of significance ranked this relationship behind combinations of several important human and environmental factors, such as bad weather, slippery surface, driving too fast, following too closely, and executing an improper maneuver. While these human and environmental factors are hard to remedy, vehicles on the shoulder and congestion could potentially be alleviated through incident management and operational strategies.

Because spatiotemporal analysis and association rule mining only demonstrated a correlation between vehicles on the shoulder, congestion, and crashes, researchers reviewed crash narratives to better understand what caused crashes. This review found that 1.8% of all crashes directly involved vehicles on the shoulder, while 23% of crash narratives cited congestion as a contributing factor. However, crash narratives did not shed much light on how the vehicles on shoulders contributed to crashes (beyond saying they were directly involved) or how they influenced the congestion which led to the crashes. Only 6 of 512 crash narratives flagged for review specified a vehicle on the shoulder as contributing to congestion and subsequent crashes. Hence, there was insufficient evidence to conclude this chain of events occurs at a frequency which is statistically significant. Nevertheless, while a small fraction of crashes were attributed to vehicles on the shoulder, these crashes tended to be more severe than the average interstate crash. For crashes involving vehicles on the shoulder, 3.75% and 27.9% resulted in a fatality and injury, respectively (compared to 0.63% and 16.3%, respectively, of all crashes). Additionally, crashes involving a vehicle on the shoulder accounted for 10.7% of all fatal crashes on interstates and parkways. These findings demonstrate that vehicles on shoulder constitute an operational hazard and may cause greater losses if directly involved in crashes.

6.2 Implementation Consideration

This study quantitatively analyzed the correlative and causal relationships between vehicles on the shoulder and crashes. In addition to the findings summarized above, data also indicate that some vehicle-on-shoulder alerts stay active for an extended period. Over 35% of the vehicle-on-shoulder alerts stayed active for at least 30 minutes, while 12% remained active for at least one hour. Vehicles remaining on shoulders for extended periods increase crash risk. KRS189.450 prohibits the parking of disabled vehicles on shoulders of limited access highways for more than 24 hours. KRS189.753 states that a vehicle left on the highway's right of way for three consecutive days may be considered abandoned — the Kentucky State Police can order the removal of such a vehicle. Adopting measures to respond to incidents and promptly remove vehicles located on the shoulder will help lower crash risk. Examples may include increase highway patrol and improve coordination with local law enforcement agencies on quick removal of abandoned vehicles from freeway shoulder.

Data used in this study, and the analytical methods proposed, offer much-needed insights into the challenges posed by vehicles on roadway shoulders. For instance, hotspot maps (i.e., Figure 3.7) showing locations where vehicles on shoulder, congestion, and crashes coincide can inform decisions about where to patrol freeways. The heatmap (i.e.. Figure 3.3) can help detect abandoned vehicles if the vehicle-on-shoulder alerts appear at the same location for an extended period of time, and inform patrol officers of these locations. While this study focused on vehicle-on-shoulder alerts, other incident alerts (e.g., jam, road object, road ice) available from Waze can provide additional context on crashes and therefore should be included in future analysis.

References

Agent, K. R., & Pigman, J. G. (1989). *Accident Involving Vehicles Parked on Shoulders of Limited Highways*. College of Engineering, University of Kentucky, Lexington.

Agrawal, R., Imieliński, T., & Swami, A. (1993). 170035.170072.pdf. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207–216.

Amin-Naseri, M., Chakraborty, P., Sharma, A., Gilbert, S. B., & Hong, M. (2018). Evaluating the Reliability, Coverage, and Added Value of Crowdsourced Traffic Incident Reports from Waze. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(43), 34–43. https://doi.org/10.1177/0361198118790619

dos Santos, S. R., Junior, C. A. D., & Smarzaro, R. (2017). Analyzing Traffic Accidents based on the Integration of Official and Crowdsourced Data. *Journal of Information and Data Management*, *8*(1), 67–82.

Eriksson, I. (2019). *Towards Integrating Crowdsourced and Official Traffic Data*. Uppsala Universitet.

Flynn, D. F. B., Gilmore, M. M., & Sudderth, E. A. (2018). *Estimating Traffic Crash Counts Using Crowdsourced Data* (DOT-VNTSC-BTS-19-01; p. 48).

Goodall, N. J. (2017). Probability of Secondary Crash Occurrence on Freeways with the Use of Private-Sector Speed Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2635*(1), 11–18. https://doi.org/10.3141/2635-02

Goodall, N., & Lee, E. (2019). Comparison of Waze crash and disabled vehicle records with video ground truth. *Transportation Research Interdisciplinary Perspectives*, *1*, 100019. https://doi.org/10.1016/j.trip.2019.100019

Lenkei, Z. (2018). *Crowdsourced traffic information in traffic management: Evaluation of traffic information from Waze*.

Liu, Y., Hoseinzadeh, N., Han, L. D., & Brakewood, C. (2019). *Evaluation of Crowdsourced Event Reports for Real Time IMplementation—Spatial and Temporal Accuracy Analyses*.

Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *The Journal of Open Source Software*, *3*(24). https://doi.org/10.21105/joss.00638

Transportation Research Board. (2016). *Highway Capacity Manual 6th Edition: A Guide for Multimodal Mobility Analysis*. The National Academies Press. https://doi.org/10.17226/24798

Turner, S., Martin, M., Griffin, G., Le, M., Das, S., Wang, R., Dadashova, B., & Li, X. (2020). *Exploring Crowdsourced Monitoring Data for Safety* (TTI-Student-05). Texas A&M Transportation Institute. https://www.vtti.vt.edu/utc/safe-d/wp-content/uploads/2020/04/TTI-Student-05_Final-Research-Report_Final.pdf

Yoon, J., Noble, B., & Liu, M. (2007). Surface street traffic estimation. *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services - MobiSys '07*, 220. https://doi.org/10.1145/1247660.1247686

Young, S. D., Wang, W., & Chakravarthy, B. (2019). Crowdsourced Traffic Data as an Emerging Tool to Monitor Car Crashes. *JAMA Surgery*, *154*(8), 777. https://doi.org/10.1001/jamasurg.2019.1167

Zhang, C., & Zhang, S. (2002). *Association rule mining: Models and algorithms*. Springer.