Journal of Transportation Safety & Security

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

Pattern recognition from light delivery vehicle crash characteristics

Subasish Das, Anandi Dutta & M. Ashifur Rahman

To cite this article: Subasish Das, Anandi Dutta & M. Ashifur Rahman (2021): Pattern recognition from light delivery vehicle crash characteristics, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2021.1995800

To link to this article: https://doi.org/10.1080/19439962.2021.1995800

Published online: 25 Oct 2021.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2021.1995800

Pattern recognition from light delivery vehicle crash characteristics

Subasish Das[a], Anandi Dutta[b], and M. Ashifur Rahman[c]

[a]Roadway Safety Division, Texas A&M Transportation Institute, Texas A&M University System, Bryan, Texas, USA; [b]Department of Computer Science, University of Texas at San Antonio, San Antonio, Texas, USA; [c]Department of Civil Engineering, University of Louisiana, Lafayette, Louisiana, USA

ABSTRACT

In the era of food delivery and grocery delivery startups, traffic crashes associated with light delivery vehicles have increased significantly. Since the number of these crashes is increasing, it is important to investigate light vehicle crashes to gain insights into potential contributing factors. This study collected seven years (2010-2016) of data from traffic crash narrative reports and structured traffic crash data from Louisiana. Using text search options and manual exploration, a database of 1,623 light delivery-related crashes was examined with a comparatively robust clustering method known as cluster correspondence analysis. The findings identified six clusters with specific traits. The key clusters are fatigue, alcohol impairment, young drivers on low to moderate speed roadways, open country and moderate speed state/U.S. highways, and interstate-related crashes due to inattention. Policymakers can use the findings of the current study to perform data-driven policy development and promote safety for delivery-related travels.

KEYWORDS

Light delivery vehicles; crashes; severities; correspondence analysis; contributing factors

1. Introduction

Light delivery vehicles are an important part of the urban fleet, which is a vital component in the overall transportation system. With the rise of food and grocery delivery startups and autonomous delivery vehicles, it is anticipated that the number of crash events associated with light delivery vehicles will increase in the near future. A recent report announced that the National Highway Traffic Safety Administration (NHTSA) granted 'Nuro,' one of the autonomous delivery vehicle (ADV) companies, a temporary exemption from certain low-speed vehicle standard requirements. The exemption will allow the company to deploy a low-speed, occupant-less electric delivery vehicle called R2 (NHTSA, 2020). There has been a limited number of studies focused on the safety issues associated with light delivery

CONTACT Subasish Das s-das@tti.tamu.edu Texas A&M Transportation Institute, Texas A&M University System, 1111 RELLIS Pkwy, Bryan, TX-77807, USA

vehicles. Thus, an investigation into light delivery vehicle-related crashes is needed.

This study is limited to light delivery vehicle-related crashes, so it is important to know that conventional crash databases do not provide sufficient information specific to light delivery vehicle-related crashes. As autonomous delivery vehicle companies are entering the market, it is important to gain insights from conventional light delivery vehicle crashes. The findings from this study can be applied to the intelligent design of autonomous delivery vehicles in the future by improving collision-avoiding abilities. This study aims to answer two research questions: 1) is there any cluster or sub-group effect in the light delivery vehicle-related crash databases, and 2) what are the patterns of key contributing factors in light delivery vehicle-related crashes? This study aims to answer these research questions by collecting seven years (2010-2016) of traffic crash data and detailed narratives from police reports in Louisiana. Both text search and manual efforts were conducted to develop the database. A crash is a complex juxtaposition of multiple factors and scenarios; this study attempted to account for this by collecting a wide range of variables to determine the association between key contributing factors. Cluster correspondence analysis is a less explored clustering technique in transportation safety engineering. The current study considered this method to be suitable due to its performance and robust interpretation capability.

The rest of the paper is as follows. The next section provides a brief overview of relevant studies. Next is the methodology section, which contains three subsections, including data preparation, exploratory data analysis, and a short overview of the theory of cluster correspondence analysis. The results and discussions are explained later. The final section is the conclusions.

2. Literature review

The number of studies on light delivery vehicle-related safety issues is limited. This section provides a brief overview of the relevant studies. Some of these studies are not directly related to safety issues associated with delivery vehicles. Delivery vehicle-related safety issues are not limited to driving only- they are also associated with delivery-related scenarios. The following discussions are related to the broader contexts of light delivery vehicles and associated safety concerns.

Crash data analysis on light delivery vehicles is very limited. Serre, Perrin, Dubois-Lounis, and Naude (2014) conducted a comprehensive study on light goods vehicle (LGV)-related traffic crashes by collecting data from France. The sample size of the analysis was only 88 traffic crashes involving LGVs. Some of the key features, such as safety concerns, maintenance and loading issues, LGV design, crash characteristics, vehicle features, and passive safety, were explored in this study. Kuehn et al. (2011) explored the safety issues associated with LGVs. This study provided hotspots of LGV crashes in Germany. The results show that the patterns of LGV crashes are similar to passenger crash crashes. Some key variations are associated with backing-related incidents, and occupant/passenger safety. Davidse and Duijvenvoorde (2012) conducted an in-depth study by examining crashes involving delivery vehicles to gain insight into the circumstances and factors that influence the occurrence and consequences of crashes in urban areas. To influence the safety-related risks of delivery riders, Zheng, Ma, Guo, Cheng, and Zhang (2019) conducted a study illuminating the role of working conditions. The findings show that fatigue, a huge workload, and risk-taking behaviors have a high likelihood of being related to a crash. Time pressure and several work-related traits indirectly affect crash involvement by influencing riders' feelings of fatigue and their riding behaviors. Zhang, Huang, Wang, and Casey (2020) conducted a study to identify the extent and motives behind cell phone use by deliverymen while driving. The findings indicate that efficient intervention efforts can effectively curb these behaviors. Shen, Zhang, Lv, Wei, and Sun (2020) applied the theory of planned behavior to examine the psychological characteristics of delivery riders' risky behavior of red-light running. Conformity tendency and the traffic environment were found to be the key contributing factors. Delivery vehicles usually extend to bicycles and motorcycles, but few studies have explored this area. Chung, Song, and Yoon (2014) conducted a study to explore crash severity patterns of delivery-related motorcyclist crashes. Some of the crucial predictors identified in this study are traffic violation behaviors, broadside crashes, nighttime, and speeding. Truong and Nguyen (2019) examined the mobile phone usage patterns in delivery-related motorcycle trips. The results show that ride-hailing motorcycle taxi services have a high likelihood of mobile phone-related crashes. Zhang, Ji, Lv, and Ma (2021) conducted a study to examine the risky behavior of delivery-related e-bike riders' red-light running.

Parking and routing issues are critical for light delivery vehicle-related safety issues. Several studies focused on this particular issue. Han, Chin, Franzese, and Hwang (2005) developed a geographically based combinatorial model to estimate the extent of capacity losses and subsequent traffic delays from illegal parking of delivery trucks used for pickup and delivery. The estimate of the national pickup and delivery effect is somewhat conservative because only daytime and weekday activities were studied. Amer and Chow (2017) presented an on-street parking model for downtown areas in urban centers to incorporate the often-neglected demand for delivery trucks. They examined the relationship between truck delivery behaviors and passenger vehicles' parking. They also provided tools for policymakers to optimize the tradeoffs in pricing, parking space allocation, and aggregate network congestion. Using data collected at eight large residential buildings from four boroughs in New York City, Salas, Chen, and Conway (2018) characterized the food delivery trips generated by in-store and online home-delivery services. The study investigated delivery frequencies, delivery times, parcel lot sizes, and modes of delivery. Based on the B2C e-commerce environment, Mao, Wu, Li, and Li (2020) conducted a study on the home delivery of fresh food. The findings showed a nonlinear negative correlation between temperature loss and home delivery distance for fresh food. Hess, Spinler, and Winkenbach (2021) explored predictive routing applications using data from a meal delivery company 'Uber Eats.' The results show the machine learning models perform better in forecasting compared to statistical models. Park and Saphores (2021) developed a graph theory-based solution to the optimized fleet size problem to serve all the delivery demands without any delay in an autonomous driving setting.

Few new studies have explored ADV-related issues. To investigate the users' acceptance of ADV in Germany, Kapser and Abdelrahman (2020) adapted an extended Unified Theory of Acceptance and Use of Technology (UTAUT2) to the context of ADVs in last-mile delivery. Figliozzi and Jennings (2020) performed an analysis by targeting two major research questions: (a) what are the existing capabilities of ADVs, and (b) what are the energy and emissions reductions that ADVs can bring about? Their results indicated that ADVs have significant potential in reducing CO_2 emissions and energy consumption in urban areas.

The existing literature on delivery vehicle-related studies is mainly focused on driver-related issues, parking-related issues, routing optimization, and automated delivery performances. However, few studies have addressed the safety issues associated with delivery vehicle crashes. The present study aims to bridge this gap by identifying the extent of and the key contributing factors of delivery vehicle-related crashes.

3. Methodology

3.1. Data integration

This study utilized seven years (2010-2016) of crash data from the Louisiana Department of Transportation and Development (LADOTD). The dataset has three major files: 1) crash file, 2) vehicle file, and 3) roadway inventory (known as DOTD table) file. The crash file contains general information about crash characteristics and circumstances. The roadway inventory file contains information about crash location, roadway type,

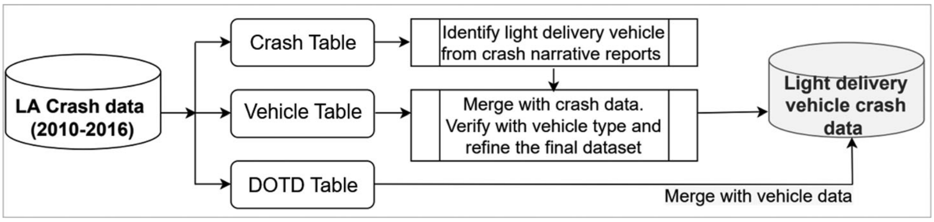


Figure 1. Data preparation flowchart.

traffic volume, segment length, and other relevant geometric information. As delivery vehicles are not classified as a specific vehicle type in Louisiana crash data, the researchers needed to explore the police-reported crash narrative data to identify which crashes involved delivery vehicles. A set of delivery-related keywords (e.g., delivery, pizza delivery, food delivery, grocery delivery) was used to identify the crash reports associated with delivery-related crashes. A manual effort was made to identify delivery-related crashes using broader vehicle categories (limited to passenger car, light truck, van, and pickups) and manually reading the crash narrative reports. After identifying these events in the 'vehicle' table, a data merging method was conducted to develop a database with crash, roadway, vehicle, and driver information (see Figure 1). The final dataset contained 1,623 unique crashes with 3,015 involved drivers.

3.2. Exploratory data analysis

It is important to perform a data-driven variable selection method when performing a robust analysis. As the crash dataset contains a wide range of variables (e.g., numerical, integer, nominal or categorical, and ordinal), it is important to determine which variables can provide intuitive knowledge about the crash occurrences. Variables examined in the related studies discussed in the literature review were explored for the preliminary selection of variables. After removing redundant variables (e.g., district name, agency name, driver registration) for this study, this study selected a list of 40 relevant variables for preliminary exploration. To determine the most relevant variables, several considerations, such as correlated variables, variables with significant missing values (20% or above), and variables with a highly skewed attribute (95% coverage), were conducted. For example, the hour of the day is correlated with the lighting condition, vehicle operating speed has 50% missing values, and pavement type has 96% paved (asphalt or concrete) roadways. After performing all quality checks, sixteen variables were selected for the final analysis. The final dataset was conducted at the person level (3,015 drivers), in which each row indicates the driver level information, and each column indicates a selected variable. With one hot encoding,

Table 1. Descriptive statistics.

Category	Perc.	Category	Perc.
Alignment (Align.)		Posted Speed Limit (PSL)	
Straight-Level	93.77	30 mph or less	58.79
Straight-Level-Elev.	0.90	35-45 mph	35.04
Curve-Level	2.78	50-60 mph	5.33
Curve-Level-Elev.	0.30	65-70 mph	0.84
Dip, Hump-Straight	0.13	Number of Vehicles Involved (NVeh)	
Hillcrest-Curve	0.07	Multi Vehicle	89.21
Hillcrest-Straight	0.57	Single Vehicle	10.75
On Grade-Curve	0.20	Not reported	0.03
On Grade-Straight	0.84	Intersection Type (Intersec.)	
Other	0.44	Intersection	70.82
Highway Type (Hwy.)		Segmentation	29.15
City Street	45.46	Not reported	0.03
Parish Road	27.71	Driver Gender (Gen.)	
State Hwy	16.28	Female	24.66
U.S. Hwy	7.87	Male	57.39
Interstate	1.74	Unknown	17.82
Toll Road	0.94	Driver Race (Race)	
Roadway Type (Road)		White	47.07
2-Way No Sep.	68.27	African American	32.40
2-Way with Sep.	16.08	Others	2.68
One-Way Road	12.76	Asian	0.03
2-Way with Barr.	1.64	Unknown	17.82
Other	1.24	Driver Age (Age)	
Locality Type (Locality)		15-24	12.8
Business Cont.	31.69	25-34	20.37
Mixed	33.97	35-44	16.31
Residential	25.76	45-54	15.61
Industrial	2.21	55-64	11.22
Residential Scatt.	3.45	> 65	5.93
Open Country	0.97	Driver Injury Type (Inj.)	
School/Playground	0.67	Not reported	17.72
Other	1.27	Incapacitating/Severe	0.50
Lighting Condition (Lighting)		Non-Incapacitating/Moderate	1.64
Daylight	89.15	Possible/Complaint	7.37
Dark - Cont. St. Light	6.16	No Injury	90.49
Dark - No St. Lights	1.68	Driver Alcohol Test (Alc.)	
Dark - Int. St.Light	1.04	No Test Given	77.86
Dawn	0.10	Test Given, Bac	0.37
Dusk	1.31	Test Given, Results Pending	0.03
Other	0.57	Not reported	21.61
Weather Type (Weather)		Driver Condition (Cond.)	
Clear	75.54	Normal	48.68
Cloudy	17.92	Inattentive	25.56
Rain	5.49	Distracted	3.52
Blowing Sand, Soil, Dirt, Snow	0.07	Drinking Alcohol - Impaired	0.27
Fog/Smoke	0.07	Drinking Alcohol - Not Impaired	0.10
Not reported	0.03	Drug Use - Impaired	0.10
Severe Crosswind	0.03	Apparently Asleep/Blackout	0.23
Sleet/Hail	0.07	Fatigued	0.13
Snow	0.07	Physical Impairment (Eyes, Ear, Limb)	0.07
Other	0.70	Other	21.34
Day of Week (DOW)			
MTWT	71.69		
FSS	28.31		

the final matrix was 3015 by 91. Table 1 lists the distribution of the key variable categories. There were several interesting findings shown in the table. One finding is that the majority of the crashes occurred on straight aligned roadways (for overall crashes, straight-level crashes are approximately 85% of all crashes). Another finding is that approximately 90% of crashes occurred on city streets, parish roads, and state highways, which is higher than the percentage of crashes on these roadways for overall crashes (around 70%). This is obvious as these crashes occurred more in the urban locations for loading or unloading. Two-way undivided roadways are associated with 68% of crashes, and nearly 92% of crashes occurred in business, residential, and mixed localities. Most of the crashes were multiple vehicle crashes. Most of the crashes occurred on low posted speed limit roadways, and approximately 70% of crashes occurred at an intersection. Around 80% of drivers were either white or African American, and most of the drivers involved in delivery-related crashes were in the age group between 25 to 64 years (above 80%; for overall crashes, this age group is associated with around 60% of crashes). There were no significant insights in lighting, weather, or day of the week variables. Injury crashes were lower in delivery-related crashes, and there were no fatal crashes during the study period. Alcohol involvement was also rarely found; the drivers were asked for alcohol tests in only about 0.40% of cases. Inattention and distraction were associated with approximately 29% of delivery vehicle-related crashes.

3.3. Cluster corresponding analysis

For categorical data analysis, correspondence analysis (CA) is one of the most popular data analysis techniques. The core idea is to perform dimension reduction from simple two-way and multi-way tables, which contain an association between the columns and rows from a multifaceted dataset. Different variants of CA have been used in transportation studies (Ali, Dissanayake, Bell, & Farrow, 2018; Baireddy, Zhou, & Jalayer, 2018; Das, 2021; Das, Ashraf, Tran, & Dutta, 2020; Das, Brimley, Lindheimer, & Zupancich, 2017; Das, Jha, Fitzpatrick, Brewer, & Shimu, 2019; Das, Sun, Dadashova, Rahman, & Sun, 2021, Hosseini, Jalayer, & Das, 2021; Das et al., 2020; Das & Dutta, 2020; Das & Sun, 2015; 2016; Jalayer, Zhou, & Das, 2018; Wen & Chen, 2011). Cluster correspondence analysis, a variant of CA, combines both cluster analysis and the dimension reduction method for categorical datasets. The cluster correspondence analysis algorithm assigns individuals to optimal scaling values and assigns variable attributes to clusters in order to achieve variance maximization objectives. A brief overview of this method is presented here, which is mostly based on the theory developed by Van de Velden, Enza, and Palumbo (2017).

First, consider the data is associated with n individuals (e.g., drivers involved in delivery vehicle-related crashes) for p categorical variables (e.g., roadway alignment). This information can be expressed by super indicator matrix $\mathbf{Z}$ with $n \times Q$ dimension, where $Q = \sum_{j=1}^{p} q_j$. The user can use an indicator matrix $\boldsymbol{Z}_K$ to develop a tabular format to cross-tabulate cluster memberships with the nominal or categorical variables such as $\boldsymbol{F} = \boldsymbol{Z}'_K\boldsymbol{Z}$, where $\boldsymbol{Z}_K$ is the $n \times K$ indicator matrix indicating cluster membership. The application of the CA framework to this matrix populates optimal scaling values for rows (as clusters) and columns (as categories). The clusters are optimally separated with respect to the distributions over the categorical variables in the two-dimensional plane. In the same way, the categories with differing distributions over the clusters are separated optimally, which can be expressed as:

$$max\emptyset_{clusca}(\boldsymbol{Z}_K, \boldsymbol{B}^*) = \frac{1}{p} trace\boldsymbol{B}^{*'}\boldsymbol{D}_z^{-1/2}\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}_K\boldsymbol{D}_Z^{-1}\boldsymbol{Z}'_K\boldsymbol{M}\boldsymbol{Z}\boldsymbol{D}_z^{-1/2}\boldsymbol{B}^* \quad (1)$$

Where,
$\boldsymbol{M} = \boldsymbol{I}_n - 1_n 1'_n/n$
$\boldsymbol{B} = \sqrt{np}\boldsymbol{D}_z^{-\frac{1}{2}}\boldsymbol{B}^*$
$\boldsymbol{D}_K = \boldsymbol{Z}'_K\boldsymbol{Z}_k$, a diagonal matrix with cluster sizes
$\boldsymbol{D}_z$ is a diagonal matrix so that $\boldsymbol{D}_z 1_Q = \boldsymbol{Z}' 1_n$

This study used an open-source R package 'clustrd' (Allard, 2019) to perform the analysis. Out of several validity measures, this study used the Calinski-Harabasz measure. This measure, also known as the valence ratio criterion, is the ratio of the sum of between-clusters dispersion and inter-cluster dispersion for all clusters. The higher the score, the better the performance. This measure is used for the application of k-means clustering to complete clustering for different k measures.

4. Results and discussions

The selection of an appropriate number of clusters was one of the critical issues. After performing several tests, the research team finally limited the number of clusters to 6. Due to the nature of the data and unknown information, all of the other trials using a different number of clusters had more than 50% of the information contained in cluster 1, which presents a general trend of the overall data. This study determined that a two-dimensional, six-cluster solution would be the most suitable for this analysis. The objective criterion value of the final model is 7.2714. In the framework of CA, the origin indicates the mean profile, and all other coordinates depict deviations from this average profile. The locations of the cluster centroids answer research question 1 by indicating that there are cluster or sub-group effects in the light delivery vehicle-related crash databases. Table 2 lists the

Table 2. Location of the cluster centroids and other measures.

Cluster	Size (percentage)	Sum of Squares	Dimension 1	Dimension 2
Cluster 1	1444 (47.4%)	0.0126	−0.0069	−0.0049
Cluster 2	737 (24.7%)	0.0149	−0.0108	0.0022
Cluster 3	524 (17.6%)	0.0217	0.0333	0.0012
Cluster 4	212 (7.1%)	0.0076	0.0051	−0.0021
Cluster 5	73 (2.4%)	0.0141	−0.0081	0.0364
Cluster 6	25 (0.8%)	0.0068	−0.0112	0.0956

cluster size, sum of squares, and coordinates of the cluster centroids on two axes or dimensions.

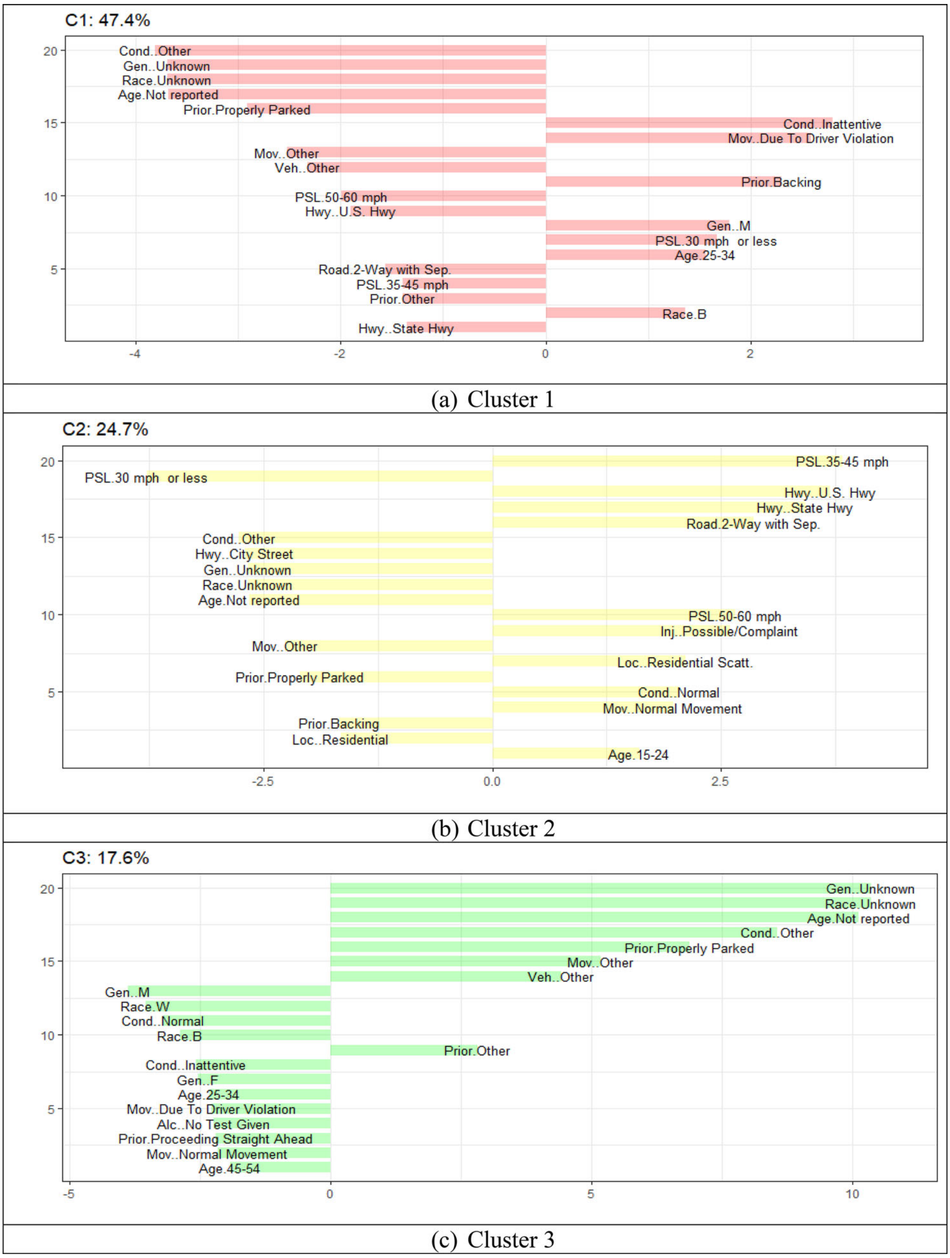
The six plots in Figure 2 and Figure 3 show the 20 attributes for each cluster with the highest standardized residuals (positive or negative). A positive (or negative) residual means that the attribute has an above (or below) average frequency within the cluster. The longer the bar is, the more dominance that attribute has (in terms of above or below average frequency) in that cluster. The purpose of this study was to explore the association among the contributing attributes in each cluster. Thus, only the bars with positive values on the right-hand side will be discussed. The following explanation considers attributes with positive residual means. Cluster-based analysis answers research question 2 by explaining the patterns of the risk factors that are associated with light delivery vehicle crashes.

4.1. Cluster 1

This cluster has seven attributes with positive residual means: inattentive driver condition, movement due to driver violation, prior movement backing up, male drivers, posted speed limit of 30 mph or less, drivers aged 25-34 years, and race of the driver as black. This indicates that there is an association between black male drivers aged 25-34 years and inattentive driving, driver violations, and a prior movement of backing up on roads with a posted speed limit of 30 mph or less.

4.2. Cluster 2

This cluster has ten attributes with positive residual means: posted speed limit of 35-45 mph, U.S. highway, state highway, 2-way road with separation, posted speed limit of 50-60 mph, possible injury or complaint, located in a residential scattered area, normal driving condition, normal movement condition, and drivers aged 15-24 years. This indicates that there is an association between possible injuries or complaints from drivers aged 15-24 years with normal driver and movement conditions on U.S. highways, state highways, or 2-way roads with separation with a posted speed limit of 35-45 mph or 50-60 mph in residential scattered areas.



(a) Cluster 1

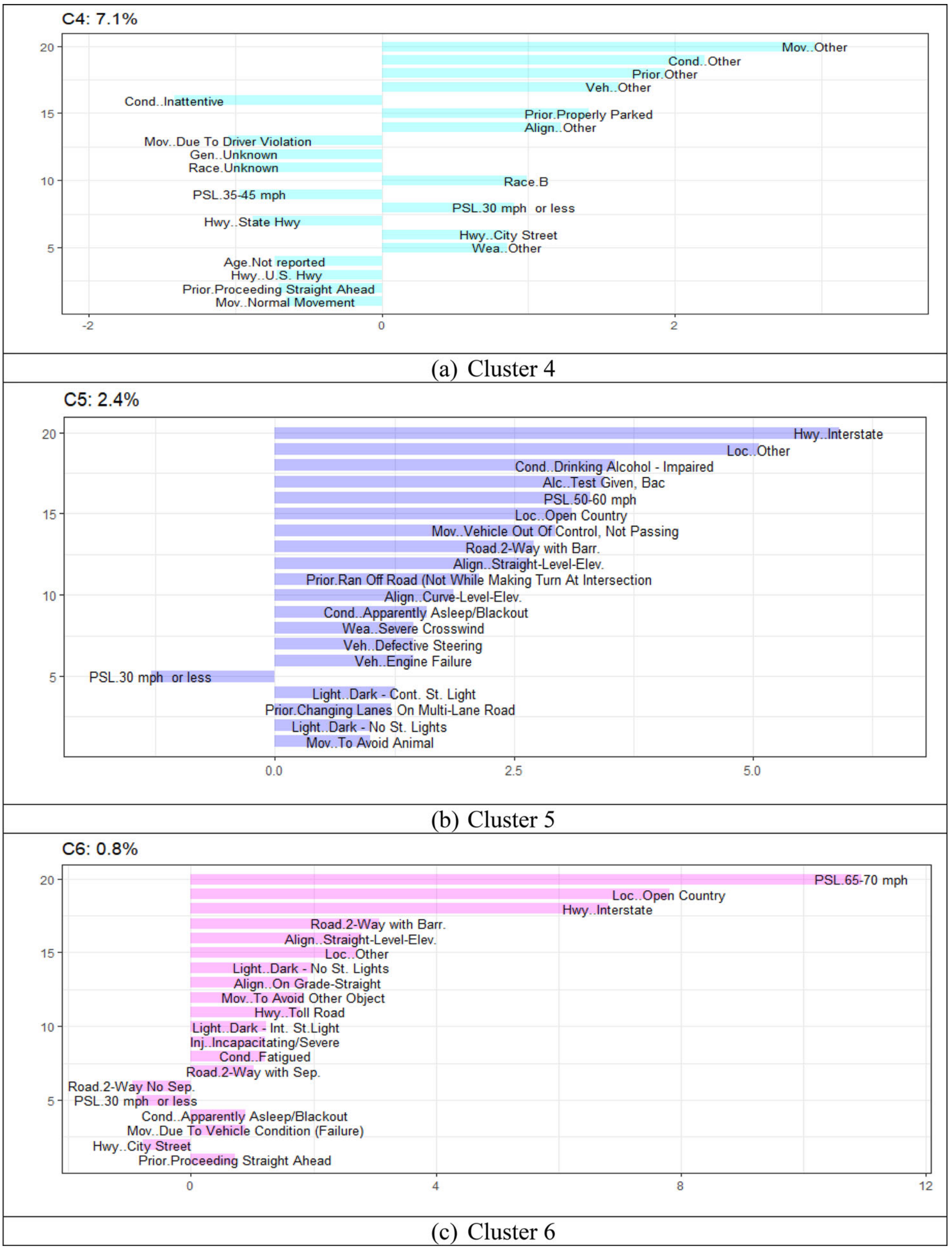
(b) Cluster 2

(c) Cluster 3

Figure 2. Top 20 largest standardized residuals per cluster (cluster 1-3).

4.3. Cluster 3

This cluster has eight attributes with positive residual means: unknown gender, unknown race, unreported age, other driving condition, properly parked prior condition, other movement condition, other vehicle type, and other prior condition. This indicates that there is an association between drivers with unknown or unreported gender, race, and age and conditions of ‘other’



(a) Cluster 4

(b) Cluster 5

(c) Cluster 6

Figure 3. Top 20 largest standardized residuals per cluster (cluster 4-6).

for driving condition, movement condition, vehicle type, and prior condition. These are also associated with properly parked as the prior condition.

4.4. Cluster 4

This cluster has ten attributes with positive residual means: other movement, other driving condition, other prior condition, other vehicle type, properly parked prior condition, other alignment, Black drivers, posted speed limit of 30 mph or less, highway city street, and other weather. This indicates that there is an association between a city street with a posted speed limit of 30 mph or less and 'other' for prior condition, driver condition, movement, weather, vehicle type, and roadway alignment. These are also associated with African American drivers and properly parked as the prior condition.

4.5. Cluster 5

This cluster has nineteen attributes with positive residual means: highway interstate, other location, driving condition impaired by drinking alcohol, blood alcohol content test given, posted speed limit of 50-60 mph, open country location, movement of vehicle out of control, 2-way road with a barrier, straight-level-elevated alignment, prior condition of running off the road, curve-level-elevated alignment, asleep or blackout driver condition, severe crosswind weather condition, defective steering in the vehicle, engine failure in the vehicle, continuous streetlights, prior condition changing lanes on a multi-lane road, no streetlights, and movement to avoid an animal. This indicates that there is an association between a very large number of variables. Drivers that were asleep or blackout, impaired by drinking alcohol, and given a blood alcohol content test were associated with out-of-control movements of the vehicle, running off the road, defective steering in the vehicle, engine failure in the vehicle, changing lanes on a multi-lane road, and movement to avoid an animal. These are also associated with highway interstates in open country or other locations with a speed limit of 50-60 mph. The road types associated with this were 2-way roads with a barrier and straight-level-elevated or curve-level-elevated alignment, severe crosswind weather conditions, and either continuous streetlights or no streetlights. One study (Das et al., 2017) showed that reduced visibility during inclement weather conditions is associated with higher crash occurrences.

4.6. Cluster 6

This cluster has seventeen attributes with positive residual means: a posted speed limit of 65-70 mph, open country location, highway interstate, 2-way roads with a barrier, straight-level-elevated alignment, other location, no streetlights, on grade-straight alignment, movement to avoid an object, toll road highway, intermittent streetlights, incapacitating or severe injury, fatigued driver condition, 2-way roads with separation, asleep or blackout driver condition, movement due to vehicle condition, prior movement proceeding straight ahead. This indicates that there is an association between crashes with incapacitating or severe injuries and highway interstates, toll roads, or 2-way roads with a barrier or separation in open country or other locations with intermittent or no streetlights. These were also associated with drivers that were asleep or blackout and fatigued, as well as a posted speed limit of 65-70 mph, straight-level-elevated or grade-straight alignment, movement to avoid an object, separation movement due to vehicle condition, and prior movement proceeding straight ahead.

One of the major advantages of this analysis is the ability to generate 'proportion odds' for the attributes in each cluster (see Table 3). One general observation is that cluster 5 and cluster 6 have very high odds for a few attributes compared to other clusters. These two clusters are associated with freeway-related crashes, which represent around 3% of the data.

The key findings from this table are stated below:

- Cluster 1 does not have drastically high odds measures for any other attributes. The highest odds measures for each of the variables are: alignment (hillcrest-curve = 1.43), highway type (Parish road = 1.27), roadway type (2-way undivided = 1.14), locality (industrial = 1.27), lighting (dawn = 2.00), weather (blowing sand/dirt = 1.43), day of the week (Monday to Thursday = 1.03), posted speed limit (30 mph or less = 1.26), intersection type (non-intersection = 1.02), gender (male = 1.28), race (Asian = 3.33), age (25-34 years = 1.41), injury type (no injury = 1.05), alcohol test (no test given = 1.13), and driver condition (inattentive = 1.65). It is important to note that this cluster contains around 47% of the overall information. This cluster also shows some odds measures are zero. Based on the values, this cluster is not associated with interstate, open country, and driver impairment.
- Cluster 2 shows higher odds for some key attributes: curve alignment, dark as lighting condition, snow/sleet weather, 35-60 mph posted speed limit, severe and moderate injury, drug-impaired, and distracted driver. This cluster represents around 25% of the data, and it indicates that drug-impaired drivers are associated with delivery-related crashes under certain conditions.
- Cluster 3, with 18% of the information, mainly presents that there is a significant number of crashes with inadequate information regarding some key variables. There is a need for additional efforts in the completion of the crash data characteristics, which is currently out of the scope of this study.
- Cluster 4 represents 7% of the data. For categorical variables, some of the categories and attributes are not well defined. These attributes are clustered together as 'others' to make the number of categories limited.

Table 3. Proportion odds of the attributes by clusters.

Variable	Category	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Align.	Straight-Level	1.05	0.93	1.02	1.00	0.70	0.55
Align.	Straight-Level-Elev.	0.00	1.33	0.44	0.00	15.22	26.67
Align.	Curve-Level	0.29	2.23	0.76	0.86	3.96	2.88
Align.	Curve-Level-Elev.	0.00	1.67	0.67	0.00	18.33	0.00
Align.	Dip, Hump-Straight	0.00	3.08	0.00	0.00	10.77	0.00
Align.	Hillcrest-Curve	1.43	0.00	2.86	0.00	0.00	0.00
Align.	Hillcrest-Straight	0.88	2.46	0.00	0.00	0.00	0.00
Align.	On Grade-Curve	0.50	2.50	0.00	0.00	0.00	0.00
Align.	On Grade-Straight	0.00	2.74	0.00	1.07	3.21	19.05
Align.	Other	0.23	0.00	2.50	6.36	0.00	0.00
Hwy.	City Street	1.16	0.37	1.44	1.38	0.30	0.00
Hwy.	Parish Road	1.27	0.67	0.87	0.99	0.44	0.00
Hwy.	State Hwy	0.62	2.36	0.42	0.35	0.93	0.00
Hwy.	U.S. Hwy	0.23	3.13	0.27	0.24	2.26	0.51
Hwy.	Interstate	0.00	0.06	0.11	0.00	23.62	45.98
Hwy.	Toll Road	0.11	1.60	1.17	2.55	0.00	17.02
Road	2-Way No Sep.	1.14	0.84	0.98	0.97	0.48	0.00
Road	2-Way with Sep.	0.56	2.15	0.53	0.67	1.45	3.23
Road	One-Way Road	0.94	0.39	1.69	1.48	1.94	0.94
Road	2-Way with Barr.	0.00	1.71	0.37	0.55	11.71	21.95
Road	Other	0.89	0.08	2.02	2.66	0.00	0.00
Loc.	Business Cont.	0.95	1.34	0.78	0.95	0.65	0.00
Loc.	Mixed	1.01	0.93	1.17	0.99	0.57	0.12
Loc.	Residential	1.20	0.49	1.20	1.30	0.32	0.00
Loc.	Industrial	1.27	1.04	0.68	0.23	0.63	0.00
Loc.	Residential Scatt.	0.35	2.84	0.38	0.55	1.19	0.00
Loc.	Open Country	0.00	0.00	0.00	0.00	16.91	70.10
Loc.	School/Playground	0.90	1.04	1.49	0.75	0.00	0.00
Loc.	Other	0.00	0.39	0.79	0.00	23.70	22.05
Light.	Daylight	1.05	0.96	0.95	1.03	0.78	0.72
Light.	Dark - Cont. St. Light	0.67	1.10	1.54	0.76	3.56	0.00
Light.	Dark - No St. Lights	0.00	2.44	0.77	0.30	4.88	14.29
Light.	Dark - Int. St. Light	0.38	2.12	1.06	0.48	0.00	11.54
Light.	Dawn	2.00	0.00	0.00	0.00	0.00	0.00
Light.	Dusk	1.15	0.92	0.84	1.07	0.00	0.00
Light.	Other	0.18	0.00	4.39	2.46	0.00	0.00
Wea.	Clear	0.97	1.05	0.99	1.05	0.91	1.11
Wea.	Cloudy	1.17	0.78	0.98	0.69	1.14	0.89
Wea.	Rain	0.95	1.06	0.97	1.04	1.75	0.00
Wea.	Blowing Sand, Soil, Dirt, Snow	1.43	1.43	0.00	0.00	0.00	0.00
Wea.	Fog/Smoke	0.00	0.00	5.71	0.00	0.00	0.00
Wea.	Not reported	0.00	0.00	6.67	0.00	0.00	0.00
Wea.	Severe Crosswind	0.00	0.00	0.00	0.00	46.67	0.00
Wea.	Sleet/Hail	0.00	1.43	2.86	0.00	0.00	0.00
Wea.	Snow	0.00	4.29	0.00	0.00	0.00	0.00
Wea.	Other	0.71	0.00	2.14	4.00	0.00	0.00
DOW	MTWT	1.03	0.97	0.98	0.96	0.82	1.17
DOW	FSS	0.91	1.08	1.04	1.10	1.45	0.57
PSL	30 mph or less	1.26	0.22	1.42	1.35	0.14	0.00
PSL	35-45 mph	0.74	2.04	0.45	0.51	1.25	0.00
PSL	50-60 mph	0.02	2.85	0.19	0.53	8.22	2.25
PSL	65-70 mph	0.00	0.00	0.00	0.00	4.88	104.76
NVeh	Multi Vehicle	0.95	1.05	1.04	1.03	0.97	1.08
NVeh	Single Vehicle	1.39	0.56	0.66	0.79	1.27	0.37
NVeh	Not reported	0.00	0.00	6.67	0.00	0.00	0.00
Intersec	Intersection	0.99	0.91	1.11	1.03	1.14	1.02
Intersec	No-intersection	1.02	1.21	0.72	0.92	0.66	0.96
Intersec	Not reported	0.00	0.00	6.67	0.00	0.00	0.00
Gen.	Female	1.07	1.48	0.02	1.15	1.39	0.81
Gen.	Male	1.28	1.10	0.01	1.15	1.05	1.25
Gen.	Unknown	0.00	0.00	5.52	0.29	0.31	0.45

(*continued*)

Table 3. Continued.

Variable	Category	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Race	White	1.18	1.37	0.01	0.87	1.22	1.19
Race	Afrin American (Black)	1.29	0.99	0.02	1.52	1.14	0.86
Race	Others	1.04	1.27	0.22	1.75	0.00	2.99
Race	Asian	3.33	0.00	0.00	0.00	0.00	0.00
Race	Unknown	0.00	0.00	5.52	0.29	0.31	0.45
Age	15-24	1.01	1.73	0.06	0.55	1.28	1.56
Age	25-34	1.41	0.90	0.01	1.20	0.74	0.59
Age	35-44	1.13	1.26	0.08	1.18	1.51	1.96
Age	45-54	1.15	1.24	0.06	1.39	1.14	1.02
Age	55-64	1.28	1.05	0.04	1.27	1.22	0.71
Age	>65	1.25	1.30	0.03	0.79	0.69	0.67
Age	Not reported	0.00	0.00	5.42	0.48	0.46	0.45
Inj.	Incapacitating/Severe	0.00	2.80	0.00	1.00	5.40	16.00
Inj.	Non-Incapacitating/Moderate	0.37	2.80	0.12	0.55	2.50	0.00
Inj.	Possible/Complaint	0.54	2.54	0.00	0.57	2.42	1.63
Inj.	No Injury	1.05	0.83	1.10	1.04	0.83	0.88
Alc.	No Test Given	1.13	1.15	0.49	0.96	0.69	1.18
Alc.	Test Given, Bac	0.00	0.81	0.00	0.00	29.73	10.81
Alc.	Test Given, Results Pending	0.00	3.33	0.00	0.00	0.00	0.00
Alc.	Test Refused	0.00	0.77	3.08	0.00	10.77	0.00
Alc.	Not reported	0.56	0.44	2.85	1.18	1.58	0.19
Cond.	Normal	1.09	1.48	0.11	0.89	1.10	1.23
Cond.	Inattentive	1.65	0.74	0.02	0.16	0.80	0.31
Cond.	Distracted	0.99	1.88	0.00	0.26	0.40	3.41
Cond.	Drinking Alcohol - Impaired	0.00	0.37	0.00	0.00	35.56	0.00
Cond.	Drinking Alcohol - Not Impaired	1.00	1.00	2.00	0.00	0.00	0.00
Cond.	Drug Use - Impaired	0.00	4.00	0.00	0.00	0.00	0.00
Cond.	Apparently Asleep/Blackout	0.00	1.74	0.00	0.00	17.83	17.39
Cond.	Fatigued	0.00	2.31	1.54	0.00	0.00	30.77
Cond.	Physical Impairment (Eyes, Ear, Limb)	1.43	0.00	0.00	0.00	0.00	0.00
Cond.	Other	0.06	0.05	4.40	2.41	0.52	0.56

The other attributes with higher odds in this cluster are city streets and roadways with a posted speed limit of 30 mph or less. The other attributes also indicate the trivial nature of the variable attributes.

- Cluster 5 represents alcohol-impaired crashes on interstate roadways in open country localities. Driver condition as apparently asleep/blackout also shows higher odds. Another interesting feature of this cluster is that severe crosswind shows higher odds for this cluster.
- Cluster 6 properties are also similar to Cluster 5. However, this cluster is not associated with driver impairment. Driver condition as apparently asleep/blackout and being fatigued also shows higher odds.

To meet the requirements of the FAST Act of 2015, the Louisiana Freight Mobility Plan aims to assist the unique needs of the LADOTD and its partners to enhance freight mobility by distinguishing demands, proposing policies, and developing implementation approaches. The 2015 Louisiana Statewide Transportation Plan (STP) illustrates the State's transportation system, including passenger and freight. The second major goal of this plan is safety, which refers to providing safe and secure travel conditions across all transportation modes through physical infrastructure improvements, operational controls, programs, and public education and awareness. The key risk factors identified in this study can help policy-makers in understanding patterns in order to implement the necessary strategies for the safe and secure mobility of light delivery vehicles.

5. Conclusions

The occurrence of light delivery-related crashes is a critical issue in the urban atmosphere. Due to the recent rise of delivery vehicle automation, it is important to understand the safety concerns of conventional light vehicle delivery-associated crashes. The findings from this study can enhance scenario planning and crash avoidance in the design of autonomous delivery vehicles. To answer the two research questions, this study applied a comparatively new categorical data analysis method that combines both cluster analysis and correspondence analysis to determine the key contributing clusters in light delivery vehicle-related crashes. This method functions as a correspondence analysis of a cluster by variable association tables. In addition to a low-dimensional approximation by presenting the centroids of clusters and variable categories, there is a cluster partitioning of individuals based on the profiles over the categorical variables. To determine the nature of the underlying cluster structures, this method outperforms CA and Multiple Correspondence Analysis (MCA). Using seven years (2010-2016) of crash data from Louisiana, this study empirically determined the relative contribution of key factors for different cluster groups. Given that a crash is typically the complex and interrelated result of several factors (for example, vehicle, human, roadway, and surroundings), this study contributes to the current safety literature by identifying high-risk scenarios where delivery vehicle-related crashes are more likely to occur through interactions with other related factors. This study developed six clusters with associated factors and provided odds ratio measures by the attributes. These findings will provide a deeper understanding of light delivery vehicle-related crashes.

This study is unique and has significant contributions in identifying association trends from light delivery vehicle crashes; however, it has some limitations. One major limitation is that the data selection is primarily based on police-reported crash narrative documents. Missing information related to ‘delivery vehicles’ in the crash narrative will prohibit the inclusion of light delivery vehicle-related crashes. The second limitation is the unavailability of some important variables such as driveway density, presence of business entities, and business types. These variables could provide additional insights. Future studies can use datasets from different states to identify light vehicle crashes using text mining and structured crash databases. Additionally, count or severity analysis can be performed to identify the locations with high numbers of these crashes. Advanced statistical and machine learning modeling can be applied to understand the risk factors and potential countermeasures in securing the safe mobility of light vehicles. The third limitation is the sample size. The sample size is limited compared to all vehicle crashes. However, the previous comprehensive study conducted by Serre et al. (2014) was based on 88 LGVs. The current study used 1,623 unique light delivery vehicle crashes (with the inclusion of 3,015 driver level information). Future studies can consider using a larger set of data to increase the sample size.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753

References

Ali, F., Dissanayake, D., Bell, M., & Farrow, M. (2018). Investigating car users' attitudes to climate change using multiple correspondence analysis. *Journal of Transport Geography*, *72*, 237–247. doi:10.1016/j.jtrangeo.2018.09.007

Allard, J. (2019). clustrd: Methods for joint dimension reduction and clustering. https://cran.r-project.org/web/packages/TaxicabCA/TaxicabCA.pdf.

Amer, A., & Chow, J. Y. J. (2017). A downtown on-street parking model with urban truck delivery behavior. *Transportation Research Part A: Policy and Practice*, *102*, 51–67. doi:10.1016/j.tra.2016.08.013

Baireddy, R., Zhou, H., & Jalayer, M. (2018). Multiple correspondence analysis of pedestrian crashes in rural Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(38), 116–127. doi:10.1177/0361198118777088

Chung, Y., Song, T.-J., & Yoon, B.-J. (2014). Injury severity in delivery-motorcycle to vehicle crashes in the Seoul metropolitan area. *Accident; Analysis and Prevention*, *62*, 79–86. doi:10.1016/j.aap.2013.08.024

Das, S. (2021). Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis. *Transportmetrica A: Transport Science*, *17*(4), 593–614. doi:10.1080/23249935.2020.1802362

Das, S., Ashraf, S., Tran, L., & Dutta, A. (2020). Pedestrians under influence (PUI) crashes: Patterns from correspondence regression analysis. *Journal of Safety Research*, *75*, 14–23. doi:10.1016/j.jsr.2020.07.001

Das, S., & Dutta, A. (2020). Extremely serious crashes on urban roadway networks: Patterns and trends. *IATSS Research*, *44*(3), 248–252. doi:10.1016/j.iatssr.2020.01.003

Das, S., Jha, K., Fitzpatrick, K., Brewer, M., & Shimu, T. H. (2019). Pattern Identification from Older Bicyclist Fatal Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(6), 638–649. doi:10.1177/0361198119841863

Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle–pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2519*(1), 95–103. doi:10.3141/2519-11

Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, *39*(2), 146–155. doi:10.1016/j.iatssr.2015.07.001

Das, S., Sun, X., Dadashova, B., Rahman, M. A., & Sun, M. (2021). Identifying patterns of key factors in sun glare-related traffic crashes. *Transportation Research Record: Journal of the Transportation Research Board*. doi:10.1177/03611981211037891

Das, S. Ridesharing services and car-seats: Technological perceptions and usage patterns, (2020). https://arxiv.org/ftp/arxiv/papers/2011/2011.02277.pdf

Das, S., Brimley, B., Lindheimer, T., & Zupancich, M. (2017Safety impacts of reduced visibility in inclement weather. Transportation Research Board 96th Annual Meeting, Washington DC.

Davidse, R., & Duijvenvoorde, K. Delivery vehicle crashes: Characteristics, crash scenarios and possible interventions: Results of an in-depth study into urban crashes involving delivery vehicles. (2012).

Figliozzi, M., & Jennings, D. (2020). Autonomous delivery robots and their potential impacts on urban freight energy consumption and emissions. *Transportation Research Procedia*, *46*, 21–28. doi:10.1016/j.trpro.2020.03.159

Han, L. D., Chin, S.-M., Franzese, O., & Hwang, H. (2005). Estimating the impact of pickup- and delivery-related illegal parking activities on traffic. *Transportation Research Record: Journal of the Transportation Research Board*, *1906*(1), 49–55.

Hess, A., Spinler, S., & Winkenbach, M. (2021). Real-time demand forecasting for an urban delivery platform. *Transportation Research Part E: Logistics and Transportation Review*, *145*, 102147. doi:10.1016/j.tre.2020.102147

Hosseini, P., Jalayer, M., & Das, S. (2021A multiple correspondence approach to identify contributing factors related to work zone crashes. Transportation Research Board 100th Annual Meeting

Jalayer, M., Zhou, H., & Das, S. (2018). Exploratory analysis of run-off-road crash patterns. In A. Alavi, W. G. Buttlar (Eds.), *Data Analytics for Smart Cities*. Boca Raton, FL: CRC Press.

Kapser, S., & Abdelrahman, M. (2020). Acceptance of autonomous delivery vehicles for last-mile delivery in Germany. *Transportation Research Part C: Emerging Technologies*, *111*, 210–225. doi:10.1016/j.trc.2019.12.016

Kuehn, M., Bende, J., Sferco, R., Schaefer, R., Georgi, A., Niewoehner, W., Schepers, A., Pastor, C.-H., & Scheerer, J. (2011). Safety of light goods vehicles – findings from the german joint project of BASt, DEKRA, UDV and VDA. Presented at the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.

Mao, M., Wu, Z., Li, L., Li, Z., & American Society of Civil Engineers. (2020). Home delivery distance and free shipping policy for fresh food considering the loss.

NHTSA. (2020). NHTSA grants nuro exemption petition for low-speed driverless vehicle. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/nuro_grant_notice_final-unofficial.pdf Accessed: July 31, 2020.

Park, B., & Saphores, J.-D. M. (2021). What is the optimal fleet size for online food delivery companies? An application to San Francisco. Presented at the Transportation Research Board 100th Annual Meeting, Transportation Research Board (virtual).

Salas, J., Chen, Q., & Conway, A. Food deliveries to residential buildings in New York City: An assessment of delivery characteristics. (2018).

Serre, T., Perrin, C., Dubois-Lounis, M., Naude, C. (2014). In-depth study of accidents involving light goods vehicles. *Presented at the ESAR 2014 - 6th International Conference: Expert Symposium on Accident Research*.

Shen, X., Zhang, F., Lv, H., Wei, S., & Sun, Z. (2020). The application and extension of the theory of planned behavior to an analysis of delivery riders' red-light running behavior in China. *Accident Analysis & Prevention*, *144*, 105640. doi:10.1016/j.aap.2020.105640

Truong, L. T., & Nguyen, H. T. T. (2019). Mobile phone related crashes among motorcycle taxi drivers. *Accident; Analysis and Prevention*, *132*, 105288. doi:10.1016/j.aap.2019.105288

Van de Velden, M., Enza, A. I. D., & Palumbo, F. (2017). Cluster correspondence analysis. *Psychometrika*, *82*(1), 158–185.

Wen, C.-H., & Chen, W.-Y. (2011). Using multiple correspondence cluster analysis to map the competitive position of airlines. *Journal of Air Transport Management*, *17*(5), 302–304. doi:10.1016/j.jairtraman.2011.03.006

Zhang, Y., Huang, Y., Wang, Y., & Casey, T. W. (2020). Who uses a mobile phone while driving for food delivery? The role of personality, risk perception, and driving self-efficacy. *Journal of Safety Research*, *73*, 69–80.

Zhang, F., Ji, Y., Lv, H., & Ma, X. (2021). Analysis of factors influencing delivery e-bikes' red-light running behavior: A correlated mixed binary logit approach. *Accident; Analysis and Prevention*, *152*, 105977.

Zheng, Y., Ma, Y., Guo, L., Cheng, J., & Zhang, Y. (2019). Crash involvement and risky riding behaviors among delivery riders in China: The role of working conditions. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(4), 1011–1022. doi:10.1177/0361198119841028