DO TRANSPORTATION NETWORK COMPANIES AFFECT ROAD SAFETY OUTCOMES: A SPATIALLY DETAILED ANALYSIS IN SAN FRANCISCO

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Engineering at the University of Kentucky

By
Vedant Shriniwas Goyal
Lexington, Kentucky
Director: Dr. Gregory Erhardt, Professor of Civil Engineering
Lexington, Kentucky
2023

https://orcid.org/0000-0001-9149-4330

ABSTRACT OF DISSERTATION

DO TRANSPORTATION NETWORK COMPANIES AFFECT ROAD SAFETY OUTCOMES: A SPATIALLY DETAILED ANALYSIS IN SAN FRANCISCO

US traffic fatal deaths have steadily risen since 2010, with the past few years witnessing an unusual trend increase. To reverse such a dangerous trend, one must understand how and why road crashes occur and which factors are causing them. Emerging transportation technologies have shown the potential to improve mobility and safety. However, such technologies are not inherently beneficial and could worsen road safety if not effectively implemented. One such transportation technology that warrants investigation is the rise of ridesharing services, also called Transportation Network Companies (TNCs).

The primary goal of the dissertation is to explore the statistical relationship between road safety outcomes and TNC service components like curbside pick-ups and drop-offs (PUDO) or through the TNC-involved vehicles miles traveled (VMT). It evaluates the relationship between TNC service components like PUDO and Tot TNC VMT with five main types of road crash frequency: the total number of road crashes, fatal and severe injury crashes, crashes involving pedestrians and bicyclists, and crashes involving drink-driving (DUI) and property-damage-only (PDO) crashes using San Francisco (SF) county data.

A fixed-effect Poisson Regression Model with a robust covariance matrix compares San Francisco (SF) county's 2010 safety outcomes when TNCs were negligible to safety outcomes for the exact locations in 2016 for which spatially detailed TNC data is available. Dependent variables like Total Crashes, Fatal and Injury Crashes, Pedestrian and Bicyclist Crashes, DUI Crashes, and PDO Crashes are evaluated using the model, controlling for vehicle speed, Total VMT, and TNC service components, namely TNC VMT and PUDO. We apply that model to 2010 and 2016 scenarios and counterfactual scenarios that estimate what would have occurred in 2016 without specific aspects of TNC operations.

The results show that TNCs indirectly increased total crashes by 4% due to higher exposure and 7% due to changes in vehicle speeds. The direct effect of TNCs on crashes offsets these increases, reducing crashes by 14%, but this effect depends upon the model specification and is insignificant in other specifications tested. The results for other types of crashes are similar in direction but lower in significance. Overall, the results suggest that TNCs are a minor factor in road safety outcomes, at least within the limits of what we can measure with the available data. This finding is broadly consistent with past research on the topic.

These results interest engineers, planners, and policymakers seeking to improve road safety. Those aiming to reduce traffic crashes would be well-advised to avoid getting distracted by TNCs in one direction or another and instead focus on known solutions, including road design, vehicle technology, and reducing exposure through reducing vehicle miles traveled.

KEYWORDS: Road Safety; Transportation Network Companies; Ride-Hailing; Fixed-Effects Poisson Regression.

Vedant Shriniwas Goyal

(Name of Student)

01/18/2023

Date

ॐ

भूर्भुवः स्वः

तत्सवितुर्वरेण्यम्

भर्गो देवस्य धीमहि

धियो यो नः प्रचोदयात् ॥

|| ॐ शांति शांति शांति ||

To

my parents, Shriniwas and Anita Goyal, who ushered continuous blessings,

my brother Mukul Goyal for his constant motivation,

my wife Parul, my life companion, for all the strength, inspiration,

my sons Nityaansh and Rayaansh for their unwavering love,

my friend Kartik Kumar and his ears for hearing me out,

my ex-bosses and now close friends Manfred Breithaupt and Grant Ennis,

and to all my friends, advisors, and future mentors,

I dedicate this work to you all.

In loving memory of my dear friend, mentor, and colleague Dr. Dominik Schmid who left for his heavenly abode on 17th Oct 2021.

I pray for you to attain moksha.

|| Om Shanti Om ||

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my dissertation Chair, Dr. Gregory Erhardt, who nurtured me with the proper tools and balance to undertake the research. His insights are invaluable and exemplify the high-quality scholarship I aspire to be. I am forever indebted to his valuable advice, constructive criticism, positive appreciation, and counsel in my research tenure. Thank you for believing in me, teaching me the art of patience, and restoring my enthusiasm and passion for continuing working toward the "*big picture*" questions. You are, in a true sense, the Guru, which everybody needs to move forward in their cosmic chaos.

Next, I thank Dr. Christopher Bollinger, who was always kind to answer all my statistical queries. It will not be an understatement to say that I achieved my statistical aha! Moments during my in-person and email chats with him. I appreciate his willingness to meet me at short notice every time and his guidance throughout my research journey.

I also like to thank my other Ph.D. committee advisors, Dr. Nikiforos Stamatiadis and Dr. Reginald Souleyrette, for their helpful advice and guidance in general. They are like an endless ocean of knowledge waiting for me to ask the right question.

A special thanks to Berry Bettie and Kimberly Kelley from the Department of Civil Engineering at the University of Kentucky. You were always there to help me right from the day I began planning my travel - from India to Lexington, USA.

A big shout to Mathews James Hunter and Nancy Christiano, whom I met when I stayed at their AirBNB property during my first visit to Lexington. You and your extended family have provided us with the warmness and comfort of a home away from home.

Thank you for teaching me the American Culture and experiencing the country's hospitality. You truly exemplify the *Vasudhaiva Kutumbakam – The World is One Family!* shlok from the Maha Upanishad mantra VI – 72.

To all those off-topic books, which I read, and re-read; to numerous digital forums, discussion pages, and posts from which I took inspiration; to the vast amount of YouTube videos I gobbled, and last but not the least to the UK Graduate Housing in providing a comfortable shelter to live and work during the pandemic times. I thank you all for your support.

Finally, I would like to thank my family. My children for understanding why I am always on my computer. My brother who took care of my parents in my absence. My parents who understood my desire to pursue this journey and move away from home. And, of course, my wife for taking all responsibility (*jimmevārī*) to ensure everything is in place and every step I take is in the direction of my goal.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ABBREVIATIONS AND ACRONYMS

Abbreviations/Acronyms	Description
AADT	annual average daily traffic
ABS	Anti-locking braking system
AM	6 to 9 am
APIs	application programming interface
BAC	blood alcohol concentration
CAGR	Compound Annual Growth Rate
CAR	Bayesian Poisson Mode
CBD	Central Business District
CHP	California Highway Patrol
CT	Census Tract
DiD	difference-in-difference
DWI	Driving while intoxicated
EA	3 to 6 am
EV	6-30 pm – 3.00 am
FARS	Fatality Analysis Reporting System
GCS	Geographic Coordinate System
GHSA	Governors Highway Safety Association
GIS	Geographic Information System
HVAC	Heating, ventilation, and air conditioning
ITS	Institute of Transportation Studies (ITS), University of California, Davis
ITS	Intelligent Transportation System
LBS	location-based technology
MD	9 am - 3.30 pm
NAIC	National Association of Insurance Commissioners
NAS	National Academy of Science, Engineering and Medicine's
NB	Negative Binomial

NHTSA	National Highway Traffic Safety Administration
NNJoin	Nearest neighbor join
OECD	Organization for Economic Co-operation and Development
PCS	Projected Coordinate system
PDO	Property Damage Only
PLZ	passenger loadings zone
PM	3.30 – 6.30 pm
PUDO	TNC pick-up and drop-off
SARAR	Spatial panel fixed-effects lag and Spatial Autoregressive with additional Autoregressive error structure
SF	San Francisco
SF-CHAMP	San Francisco Chained Activity Modeling Process
SFCTA	San Francisco County Transportation Authority
SFDPH-PHES	San Francisco Department of Public Health's Program on Health, Equity, and Sustainability
SFDPW	San Francisco Department of Public Works
SFMTA	San Francisco Municipal Transportation Agency
SWITRS	Statewide Integrated Traffic Records System
TNCs	Transportation Network Companies
TODs	Time of Day
VMT	vehicle miles traveled
WHO	World Health Organization
Year-over-Year	Y-o-Y

CHAPTER 1 INTRODUCTION

1.1 Overview of Road Safety Scenario

According to the World Health Organization (WHO) "Global Status Report on Road Safety 2018", around 1.30 million people die, while approximately 20-50 million are injured yearly on the world's roads. Road-related traffic crashes are now the eighth leading cause of death among other non-communicable diseases (WHO, 2020).

The situation in the US is not as problematic as in low- and middle-income countries. However, the Organization for Economic Co-operation and Development (OECD) data reveals that among all its member countries, the US led the traffic fatality rate per million inhabitants (117.4) in 2020 (OECD, 2020). The number is three times higher than other OECD member countries, most of which are high-income economies with a high Human-Development Index, see Figure 1-1

The National Highway Traffic Safety Administration (NHTSA), an agency in charge of reducing vehicle-related crashes and associated deaths and injuries in the US, reports a 10.5% year-over-year (YoY) increase in traffic fatalities from 2019 to 2020 compared to an average rise of 1.70% (YoY) observed between 2010-2019 (NHTSA, 2022d), see Figure 1-2. It contrasts the general expected trend during COVID19-pandemic which forced a large part of the county and its business to shut down, resulting in a decline of annual vehicle miles traveled (VMT) of 11% (Stewart, 2022). Given that VMT can be considered a measure of exposure to crashes, it is surprising to see crashes increase while VMT decreases.

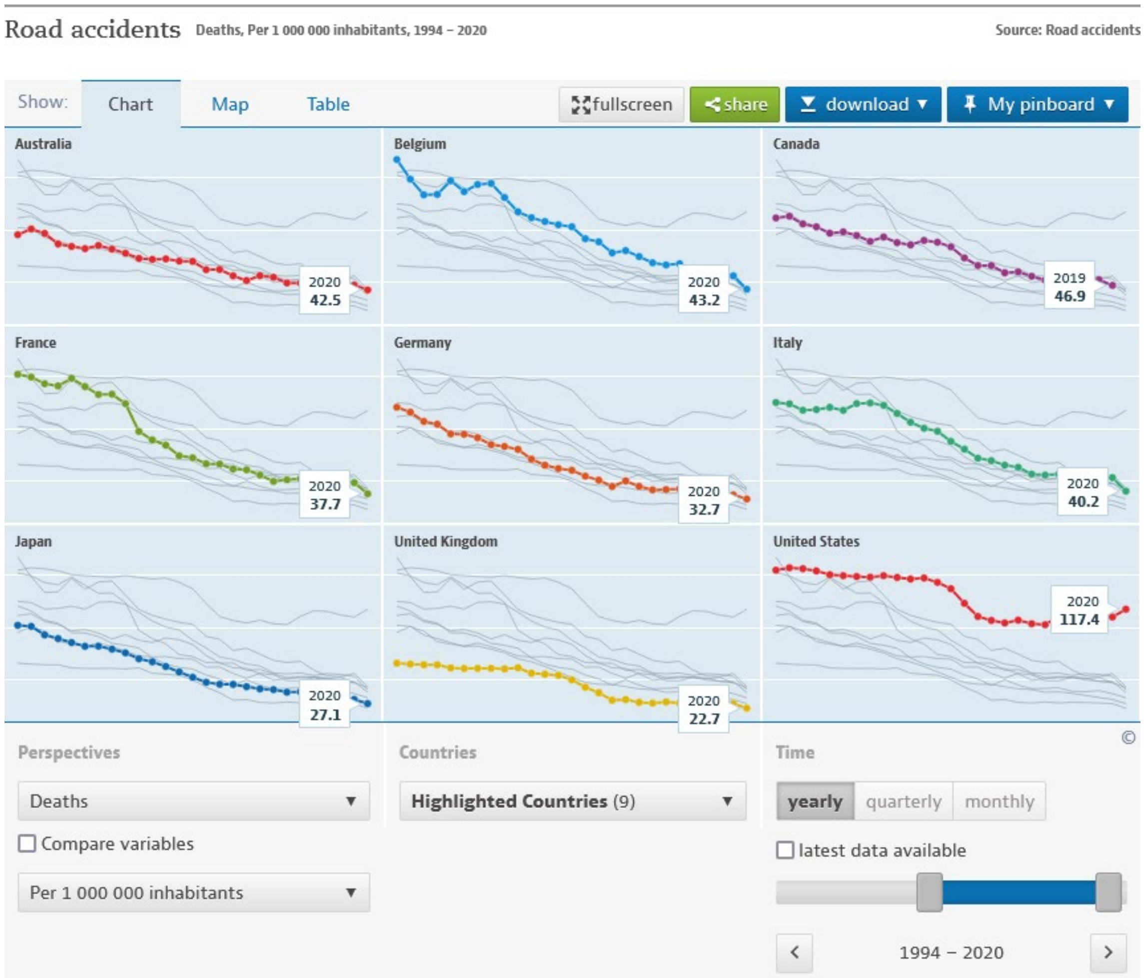

Figure 1-1: Fatal Road Crashes per million inhabitants (1994 - 2020)

These trends are not uniform across all areas or populations within the United States. The annual number of deadly deaths in urban settings jumped y 8.5% compared to 2.5% in rural areas in 2020 (Stewart, 2022). In addition, the number of fatal crash types since 2011 has sharply risen for pedestrians (61%), bicyclists (54%), and motorcyclists (42%) (Stewart, 2022).

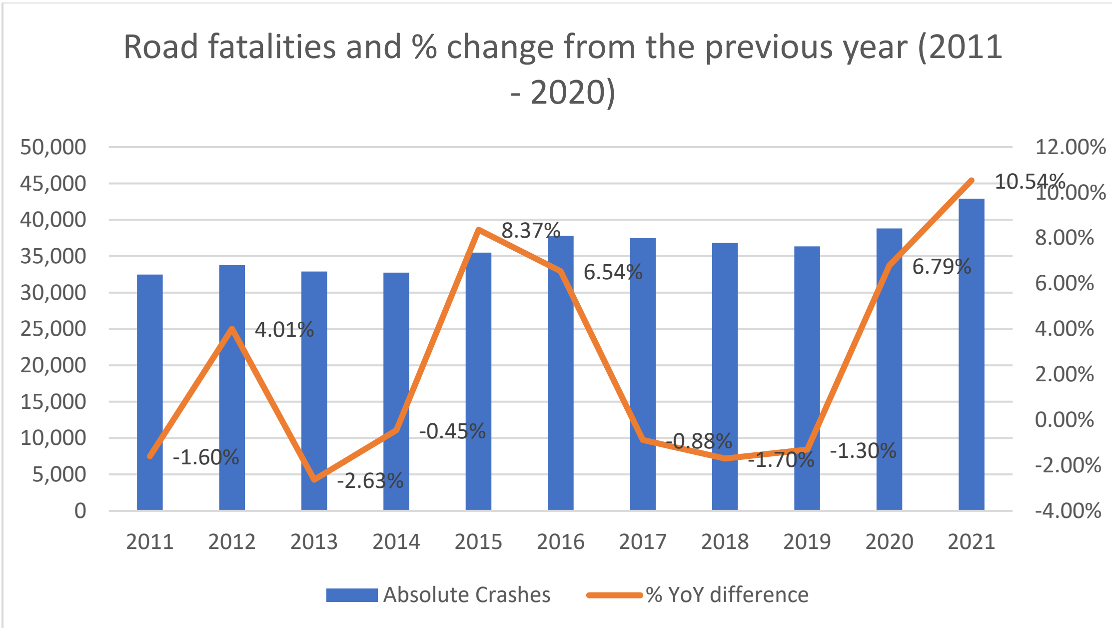

Figure 1-2: Road fatalities and percentage change in crashes from the previous year

Furthermore, these crashes have become deadlier and more frequent (Stewart, 2022). For example, pedestrian fatalities made up the increasing proportion of fatal road crashes between 2010 – 2018 (Macek, 2022). These crashes occur away from intersections, on busy main roads, or/and in the dark (IIHS, 2018).

To reverse such a dangerous trend, we need to understand how and why road crashes occur and which factors cause them. While Speed, impaired driving, helmet, seatbelt usage, unsafe vehicles, and not obeying road traffic safety rules remain the critical risk factors related to crash causation, other transportation systems and their elements need to be studied to understand the challenges and opportunities they present in achieving the road safety goals.

Recently, the concept and technology of the transportation sector have been rapidly evolving. From automated vehicles to self-driving trucks, from the fleet of electric scooters to rideshare services, emerging technologies have transformed how we move around. These technologies have changed how people and goods move, and their importance may continue to grow. While emerging transportation technologies offer the potential to improve mobility and safety, such technologies are not inherently beneficial and could worsen specific outcomes if not deployed effectively. One such transportation technology that warrants investigation is the rise of ridesharing services, also called Transportation Network Companies (TNCs).

1.2 Transportation Network Companies (TNCs)

Transportation Network Companies (TNCs) [1] have disrupted urban transportation across the USA and globally. Unlike traditional vehicle-for-hire services like taxis and limousines, TNCs do not own or operate a fleet of vehicles. Instead, TNC companies like Uber and Lyft leverage the geolocation capability of smartphones to connect potential passengers with private drivers. The estimated cost of travel is pre-calculated by the TNC app's algorithm, with both the driver and passenger agreeing to the arrangement before the trip is confirmed.

Initially, Uber and Lyft started their services to target young, affluent customers; however, triggered by the digital revolution, which reduced the transaction cost, and the rise of smartphone ownership, TNCs quickly succeeded in attracting other sections of the population. TNCs offer a direct, affordable, and comfortable door-to-door trip, so TNCs quickly became a viable alternative to public transportation.

By 2015, TNCs were present in about 75% of all US urban areas with a population of 100,000 and more (Andrew J Hawkins, 2015). Uber's service is available in over 70 countries, covering 900+ cities, with ~1.6 – 1.9 billion trips per quarter completed worldwide. Uber expects its market to grow by 19% Compound Annual Growth Rate (CAGR), reaching around USD 180 billion from $61.3 billion in 2018 (Arevalo, 2020). However, a pandemic in the form of COVID-19 may have delayed the company's growth plans (Ewoldsen, 2020).

Given their large footprint, it is essential to understand the impacts of TNC services on the transportation system.

[1] The app-based alternative urban transportation options are referred to by many names: ride-hailing, ride-sourcing, e-hailing, or transportation network companies (TNC). However, in this document, we would refer to such modes as TNCs as it is the legally preferred name among the service providers and city agencies.

1.3 Relevance of study

Research on TNCs' impact on travel behavior, general usage patterns of TNCs, and the economics of their operations are widely available (Alemi et al., 2018; Clewlow & Mishra, 2017; Tirachini, 2019). Similarly, the negative effect of TNCs on public transportation usage, car ownership, and traffic congestion are well documented (G. Erhardt et al., 2019; G. D. Erhardt et al., 2020; Henao & Marshall, 2019). The debate about the safety risk of accepting a ride from a stranger vs. access to means of transportation is being regularly framed and reframed, e.g., the regulatory decisions in London (Browne, 2019, 2020; Feikert-Ahalt, 2020). In addition to driving, research has shown that TNCs are attracting trips from non-motorized transport modes like walking and biking, thereby inducing vehicle trips that would not have taken place at all in the first place (Clewlow & Mishra, 2017; Feigon & Murphy, 2018; Gehrke et al., 2018; Rayle et al., 2016). The ongoing COVID-19 pandemic has forced companies like Uber to focus on food deliveries to compensate for their loss in passenger ridership. New York's Department of Transportation annual mobility report estimates that 50% of all TNC app users switched from using transit services (NYC Department of Transportation, 2018). Such induced trips have the potential to generate unwanted pressure on urban road networks and increase congestion and emissions from the urban transport sectors.

TNC providers claim their service has reduced motor vehicle fatalities and alcohol-involved crashes (Uber, 2017, 2019; Uber & MADD, 2015). However, critics argue their conclusion given the limitation in verifying the data accuracy, the absence of a transparent and scientific methodological approach, and preliminary or partial analysis (Jones, 2015). Independent studies on the potential linkages of TNC operations and their impact on road traffic safety outcomes also do not present precise, established trends (Barrios et al., 2019; Brazil & Kirk, 2016; Dills & Mulholland, 2018b; Greenwood & Wattal, 2017; Kirk et al., 2020; Morrison et al., 2018, 2022; Ward et al., 2021).

One of the primary reasons TNC-related research produced mixed results may be the non-availability of quality TNC trip-level data. The above studies measure either crash outcomes at the scale of a metropolitan area (Barrios et al., 2019; Brazil & Kirk, 2020, 2016; Dills & Mulholland, 2017; Greenwood & Wattal, 2017) or through the causal relationship between an increase in admissions rate at the health care facilities and presence of TNC services (Barreto et al., 2021; Huang et al., 2019). However, TNC trips are not evenly distributed throughout metropolitan areas and are highly concentrated at the center of large cities (Fehr & Peers, 2019; SFCTA, 2017). TNCs rarely share data related to their operations, limiting the opportunity to measure their impact on road traffic crashes at more spatially detailed levels (Seth, 2020a, 2020b, 2020c, 2020d). Researchers also failed to find a mechanism to understand the factors that can potentially influence TNC motor vehicle crashes or alcohol-involving (DUI) crash instances, which is the primary motivator for the current research.

1.4 Research Objectives and Approach

This study aims to measure the effect of TNC operations on road safety outcomes. We consider that TNC services could change crash outcomes in either direction. TNCs could improve road safety by reducing alcohol-involving (DUI) crashes or by replacing trips made by regular drivers with trips by more experienced professional drivers. Conversely, they could worsen road safety by increasing the exposure to crashes through more vehicular travel or by disrupting traffic flow with curbside pick-ups and drop-offs, mainly if those pick-ups and drop-offs conflict with bicycles or pedestrians, which are also negotiating the traffic on busy arterials. To test either of these is the case, we measure whether the emergence of TNCs is associated with a decrease or increase in:

- Total crashes,
- Fatal and severe injury crashes,
- Crashes involving pedestrians and bicyclists, and
- alcohol-involving (DUI) crashes
- property-damage-only (PDO) crashes

while controlling for other changes, including exposure and changes to vehicular speeds.

We evaluate these outcomes using San Francisco (SF) as a case study. TNC services had operated continuously in San Francisco since May 2010, when Uber first started offering such trips to the world, and it is home to the headquarters of the two dominant TNCs in the US, Uber and Lyft. More importantly, the San Francisco County

Transportation Authority (SFCTA) successfully collected data on TNC activity across the road network, thereby allowing to study of TNC's impact using spatially detailed data. Furthermore, the city maintains a high-quality geocoded crash database and estimates the changes in vehicle miles traveled and vehicle speeds.

We compare 2010 safety outcomes when TNCs were negligible to safety outcomes for the exact locations in 2016 when we have TNC data. We control for changes to vehicle miles traveled, vehicle speed, and road network over this period, allowing us to generate a counterfactual scenario estimating what would have occurred in 2016 if TNCs had not entered.

If our results show that TNCs improve road safety, it will enable planners to develop strategies that will leverage them to improve safety. Conversely, if our results show that TNCs worsen road safety, it will enable planners to develop strategies to mitigate those effects. If our results show that TNCs are not an essential factor in road safety outcomes in either direction, it will allow planners to focus on other safety strategies that are known to be effective.

1.5 Paper Organization

The study is organized into six chapters to answer the main research questions in section 1.4. An overview of each of the subsequent chapters is presented below.

CHAPTER 2 starts with defining the Transportation Network Companies (TNC) and how they differ from traditional vehicle-for-hire services like taxis and limousines. It briefly overviews the advantage of TNCs and how it provides an alternative to traditional car-based urban transportation. The study attempts to present a theoretical framework to explain how TNC services may, directly and indirectly, influence road safety outcomes. The chapter ends with a comprehensive review of the scientific literature focusing on TNCs and road safety crashes. The literature review is a synthesis of published and unpublished papers/literature on the impacts of TNC service trips on road traffic crashes and injuries

CHAPTER 3 explains why San Francisco County is chosen as the focus study area. It later provides a comprehensive overview of the primary data used for the current study, including the detailed traffic flows, TNC-related information on the road segment, the road network itself, and road crashes for 2010 and 2016, respectively. It also presents a detailed procedure to transform and merge all three mentioned data sources into one dataset. The produced singleton dataset is the base for all future statistical analyses and feeds into the statistical model. The later part of the chapter presents the statistical model's structure used to estimate the relationship between road crash frequency and TNC service components like TNC Pick-ups and Drop-offs (PUDO) and Vehicle Miles Travelled (VMT) and controlling for Total VMT and vehicular speeds on the broadly defined facility types.

CHAPTER 4 presents results from the explorative analysis of traffic crashes, traffic estimates, and TNC data to understand the spatial distribution of these activities. It provides countywide fatal and injury trends for the primary dependent variables: total crashes, fatal and injury crashes, pedestrian and bicyclist crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes. It provides an overview of the spatial diversity in the occurrence of road crashes and how it might impact our road crash predictions. It further decomposes the road crashes by road facility types and explains where the Total TNC activities are happening and areas witnessing variations in congested speed.

CHAPTER 5 provides an after-implementation summary of the statistical model presented in CHAPTER 3. The summary tables are provided for all dependent variables like Total Crashes, Fatal, and Injury Crashes, Pedestrian and Bicyclist involving crashes, alcohol-involving (DUI) crashes, and Property Damage Only (PDO) crashes. Further, it presents the model application results of the additive effect of introducing an independent variable and predicting its impact on overall predicted crash values. Such scenario generation and evaluation approach allowed the study to understand the potential impacts of the selected independent variable on road crashes. Further, it presents the results of several alternative model specifications estimated if including both TNC PUDO and TNC VMT makes the predictions better than the final model results.

Finally, CHAPTER 6 revisits the study's objectives and summarizes its findings. It outlines the findings related to TNC's direct and indirect effects on road crashes.

Limitations are identified while recommendations are provided for undertaking potential future research opportunities.

1.6 Chapter Summary

With 1.35 million people dying each year on the world's roads and another 20 million to 50 million seriously injured, road deaths and injuries are a silent epidemic on wheels that needs an immediate resolution. Back home, NHTSA predicts a 10.5% jump in fatal road deaths from 2019 – 2020 despite the COVID-19 pandemic forcing the annual VMT to decline by 11%. To reverse such a devastating trend, there is a need to enhance the understanding of how and why road crash occurs and which factors cause them. The relationship between traditional factors and road safety outcomes is exhaustively studied, and other emerging modes that rely on technology should be subject to scrutiny. One such modern transportation mode which has risen exponentially is ride-hailing services, also known as Transportation Network Companies (TNC). Few researchers' attempts to establish a relationship have produced mixed results, necessitating reinvestigation. Using data from San Francisco (SF), one of the few cities where TNC services had continued to operate uninterrupted since May 2010, when Uber first started offering such trips to the world, the study aims to explore the relationship between TNC services and road safety outcomes. Precisely, it measures whether TNC services increase or decrease total crashes, fatal and severe injury crashes, crashes involving pedestrians and bicyclists, alcohol-involving (DUI), and property-damage-only crashes. It compares 2010 and 2016 conditions in San Francisco before and after the introduction of TNCs.

CHAPTER 2 LITERATURE REVIEW

2.1 Background

The chapter starts with defining the Transportation Network Companies (TNC) and how they differ from traditional vehicle-for-hire services like taxis and limousines services. It briefly overviews the advantage of TNCs and how it provides an alternate mobility option to traditional car-based urban transportation. The study attempts to present a theoretical framework to explain how TNC services may, directly and indirectly, influence road safety outcomes. The chapter ends with a comprehensive review of the scientific literature focusing on TNCs and road safety crashes. The literature review is a synthesis of published and unpublished papers/literature on the impacts of TNC service trips on road traffic crashes and injuries

2.2 What are Transportation Network Companies (TNC)?

Transportation Network Companies (TNC), also known as rideshare, e-hail, or ride-hailing companies, provide on-demand transportation services. Unlike traditional vehicle-for-hire services like taxis and limousines, TNCs do not maintain their vehicle fleets and rely on an army of individual private drivers contracted by them to transport potential passengers from one place to another. Potential passengers request a ride through TNC online platform application, typically installed in a smartphone, and get matched to a private vehicle at the predetermined rate (trip cost). The TNC application uses an algorithm to take into account various real-time variables that could influence trip time and cost like driver availability near passenger's location, rider demand, road congestion, and many other factors to arrive at (estimated) dynamic prices which both the passenger and the driver agree before the start of the trip.

Relative to traditional taxis, TNC services offer customers several potential advantages:

1. Convenient booking and payment – booking a ride through an app may be more convenient than hailing a taxi on the street or calling dispatch. Many customers may prefer paying with a credit card through an app to paying a cash fare.

2. Greater supply – while reliable numbers on TNC supply are not widely available, the supply of TNCs appears to be higher than that of taxis in most US cities. It includes smaller markets where part-time drivers operating their personal vehicles increase the supply and large markets like San Francisco, where there are 12 times as many ride-hailing trips on a typical weekday as taxi trips. With a higher supply, waiting times can be shorter.
3. Cheaper travel cost – TNC trips have historically been less expensive than taxis, but recent reporting suggests this may no longer be true in 2021 and 2022. Comprehensive data on TNC fares is not widely available, but reporting suggests a post-pandemic price spike (Evans, 2021). It is reasonable to expect that TNCs can operate more efficiently than a taxi company because they do not own any vehicles or dispatch centers, the drivers are not employees, and they do not comply with the traditional regulatory framework that applies to taxis. However, some reporting suggests that TNCs may have higher overhead costs than taxi companies (Lee, 2022), and pricing may be more a function of business interests than operating expenses (i.e., offering low-cost trips to build a market versus generating profit/revenue for investors).
4. Transparency in trip cost and time - before the trip is confirmed, the driver and the passenger are provided with the estimated trip price and duration.
5. Location and traceability - when the trip is confirmed, the passenger's geolocation gets shared with the driver, which is also suggested with the most economical route (path) to reach the customer's start location, thereby reducing the delays. At the same time, passengers can track the driver's position, its route and communicate with the driver if required.

It is important to note that TNCs remain evolving businesses, so their operations may change further.

2.3 Framework by which TNC service might influence road safety outcomes

As mentioned in the previous chapter, road safety is not only an engineering problem but is a multi-faceted issue requiring political, public health, and economic support. Therefore, to identify appropriate solutions to the problem, we must understand the causes of road crashes and identify the factors that may lead to crash severity outcomes.

Bayam et al. were the first to present a meta-analysis framework to approach road crash systems (Bayam et al., 2005). According to Bayam et al., the model consists of different independent variables: drivers, vehicles, road environment, geographical conditions, road network, vehicle occupants, and other road users groups (pedestrians, bicyclists, motorcyclists, and other vehicle users from both the light and heavy commercial vehicle drivers and passengers) (Bayam et al., 2005). These independent variables interact with each other creating different road traffic scenarios. One of the traffic scenarios then becomes a crash scenario, leading to a motor vehicle road crash. The dependent variable is the outcome of the motor vehicle road crash, with the independent variable being the attributes of the participant factors at the moment of the crash.

Extending the framework explained in the previous paragraph, Figure 2-1 below presents the mechanisms by which TNC services could get involved in a road crash.

Broadly, the interaction between TNC vehicle and road crash system variables can be divided into five categories, namely

1. Road infrastructure consists of road attributes such as road conditions, surface type, traffic conditions, and traffic management systems installed on the roadways.
2. The road environment comprises information about the weather condition, lightning conditions, time of the day, and the area type (urban, suburban, rural). We also include enforcement-related traffic safety management laws and policies (seatbelt use, helmet usage, blood alcohol concentration (BAC) levels), including specifically to the TNC (driver verification, license requirements, pick-up and drop-offs pre-conditions, lane usage restrictions)
3. Vehicle characteristics comprise attributes like vehicle make, type, age, and road crash safety features available or pre-installed and include the information related to the condition of the vehicle, whether it is in good shape to operate
4. Personal characteristics revolve around the behavioral aspect of TNC driver

5. Other road users consist of information demographics of the road section or the subject area and represent their characteristics, including the total number of users, age, and gender.

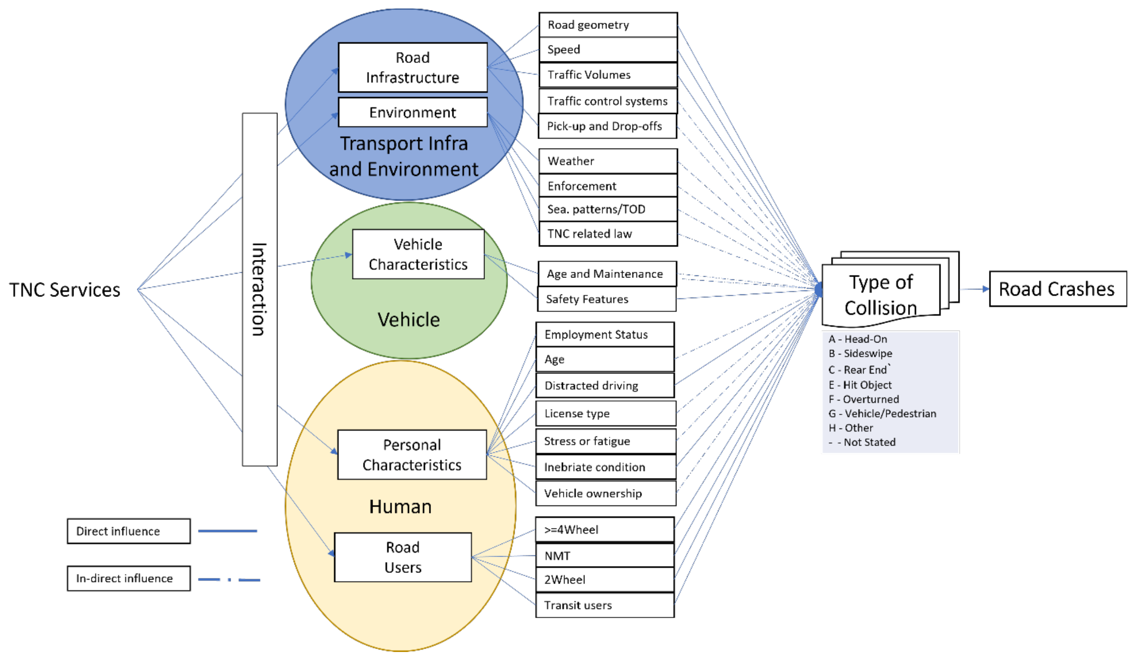

Figure 2-1: Potential ways by which TNC services could get involved in a motor vehicle road crash

The list is not exhaustive; numerous permutations and combinations between the factors can create a random crash scenario.

These factors may interact independently or in combination, often directly or indirectly influencing the road crash scenario, which later becomes the dependent variable. The associated information like road crash type, severity, number of fatalities and injuries, point of impact (contact) between two or more road user groups, and the causes leading up to the road crash becomes secondary information.

Depending upon the nature of the study, the road crash analysis could be categorized as a pre-crash (scenarios/variables leading to the crash), post-crash (impact of such a road crash on a particular road user group or causes of a road crash or crash itself).

Given the available data, the current study limits its investigation to post-crash scenarios to explore the statistical relationship between road infrastructure variables and TNC services.

2.4 The mechanism by which TNC trips can lead to more or fewer crashes

2.4.1 Unregulated markets

Unlike taxi and limousine services, TNCs drivers and vehicles are not part of the vehicle-for-hire industry and, as a result, are entirely exempted from the regulatory practice put in place by the local administration. It prevents them from undertaking numerous checks like mandatory vehicle inspection, criminal background checks, medical certification, and obtaining or renewing a driving license annually (Dills & Mulholland, 2018a). Though such an administratively long process is cumbersome, it helps filter out drivers more likely to create errors and get involved in road crash incidents (Dills & Mulholland, 2018a). Therefore, we can say that unregulated or the absence of standard-defined policies for TNC may lead to the penetration of (unskilled) drivers into the road crash system.

2.4.2 Quality of Vehicles

As mentioned in the previous bullet point, there is a difference in regulatory requirements desired for commercial vs. private vehicles. Inspection of commercial vehicles is much more intense. It involves frequent vehicle safety inspection standards involving batteries, seat belts, brake systems, doors and trunks, engine/transmission, fuel, heating, ventilation, and air conditioning system (HVAC), headlights, and many more compared to their private vehicles counterparts (Flor et al., 2022; Schaller, 2017, 2018). Fewer inspection instances mean more opportunity to wear and tear on the vehicles, primarily tires and brakes, increasing the likelihood of a system failure.

2.4.3 Increased exposure

One of the most significant and valuable perceptions of being a TNC driver is that the work hours are inherently flexible. It means drivers ostensibly have greater independence and flexibility in choosing when to log on and off from the TNC apps. While such perception is correct, a driver's flexibility in choosing when to log on and off is limited by an individual's desire to maximize profits w.r.t time spent on driving searching for passengers (Barrios et al., 2019). Given that the market demand and the chances of getting a ride as soon as they log in or finish a trip is out of their control, TNC drivers often drive for long hours and multiple TNC-app platforms. As a result, the number of vehicles on the roads, even after assuming that TNC trips are directly replacing private-vehicle trips, increases on the road. The increased number of vehicles on the road increased the probability of getting involved in road crash collisions.

2.4.4 Road congestion and speed

Theoretically, road congestion increases when VMT rises. Fehr and Peers, whom Uber and Lyft engaged to determine the VMTs in six metropolitan cities of the US, including San Francisco, found that the two TNCs contributed 13% of VMT (Fehr & Peers, 2019). It is twice that of what is estimated by the San Francisco County Transportation Authority (SFCTA) (SFCTA, 2017). Sneha et al. revisited such estimates by carrying out a series of (hypothetical) travel demand model run supplemented with observed TNC data. The study found that TNCs are associated with about half the increase in VMT between 2010 and 2016 (Roy et al., 2020). It also identified TNC volumes as the single biggest contributor to increased congestion over the same period (Roy et al., 2020). In addition, the pick-up and drop-off (PUDO) activity at the start and end of TNC trips disrupts traffic flow and deteriorates traffic congestion (Roy et al., 2020). The relationship between road congestion and traffic crashes is complex and has been under research for more than a century (Retallack & Ostendorf, 2019; Vey, 1939). A U-shaped function is generally accepted; however, a positive relationship between these variables is not uncommon (Retallack & Ostendorf, 2019).

Contrary to the congestion, the relationship between speed and road crashes is also complicated and non-linearly in nature. According to WHO Speed Manual, an increase in vehicular speed increase the likelihood of crashes due to a) less time to react to the incoming hazard and b) an increase in the probability of the other road users misjudging the real intentions of the speeding driver (WHO, 2018). The implication is that as the speed increases, the severity and the consequence of the road crash increase. On the contrary, lower average speeds result in fewer crashes relative to the number of vehicles on the road. According to WHO, a "*5% increase in average speed leads to approximately a 20% increase in fatal crashes.* (WHO, 2018)*"*

2.4.5 Driver's driving skills

Like taxi drivers, TNC drivers may be subject to fatigue and tiredness due to prolonged and non-traditional driving hours, making them more prone to errors and triggering their involvement in road crash incidents. Using 42 Sydney metropolitan taxi drivers across two years, Dalziel et al. explored the relationship between fatigue-related variables and road crash incidents (Dalziel & Job, 1997b). The study concluded that even professional drivers suffer from optimism bias. Most of the selected drivers of the study self-reported that they considered themselves more competent than their peers; however, with one caveat - the optimism is less pronounced when explicitly asked about their "*ability to drive safely when very tired*." According to Dalziel and Job, one possible explanation is that taxi "*drivers are frequently reminded of the effects of fatigue on their driving abilities due to the tiredness induced by the long hours they work, which leads to a somewhat more realistic assessment*." (Dalziel & Job, 1997b). On the other hand, the study also found that given those taxi drivers spent substantially significant time (59 hours or more per week) compared to the average driver significantly) on the road, the driving experience of a professional taxi driver far exceeds that of the average motorist.

According to Dalziel et al., the taxi driver's job improves the driver's driving capability. It explains that the job as a taxi driver involves "negotiating traffic, seeking out the next fare, responding to passenger conversation, keeping attention on the radio/computer job dispatching service, remembering how to get to the destination, and maintaining personal safety. So while the job may be tiring, in the context of attentional and cognitive resources, it is not fatiguing in the way that the minimal stimulation (and subsequent boredom) of a long straight country road at night can be" (Dalziel & Job 1997b).

Maag et al. found that the driver's age strongly influences the number and severity of crashes. Taxi drivers younger than 30 were much more prone to be involved and injured in crashes than the older group (Maag et al., 1997).

Theoretically, one can expect the role of the TNC driver to be somewhat parallel to the taxi driver's role. Therefore, one can conclude that as TNC drivers spend more time on the roads, their driving skills get superior to ordinary drivers.

2.4.6 Change in alcohol-involving crash instances

Jackson and Owens investigate the relationship between new transportation options (public transit) and alcohol consumption and found the relationship ambiguous at the aggregate level (Jackson & Owens, 2011). Using Washington D.C as the case study, the authors found that when the metro services are expanded by three hours from midnight to 3 am, the neighborhoods with at least one bar within 100 m of the metro station see DUI arrests dropping by 14%. In contrast, DUI rates remained unchanged for the neighborhood, which had no bar within a 100m distance. Uber tried to answer a similar question using data from Chicago, Illinois, in their in-house study. According to their report, late-night Uber requests within 50 meters of a bar, restaurant, or hotel that serves alcohol went up by 45.8% during peak drinking hours of 10 pm and 3 am, compared to 28.5% off-peak hours (Uber & MADD, 2015). Greenwood et al. argue that the probability of inebriated drivers opting for alternate transport services, including TNCs, is not limited to the ease and availability of the services but also depends upon the individual's ability to comprehend the actual opportunity cost. It includes the cost of getting caught, court costs, the cost of the trip, social stigma, and jail sentences (if any) (Greenwood & Wattal, 2017).

To conclude, the relationship between TNC services and alcohol-involving (DUI) crash instances is complex. TNC services may lower the DUI-related instances in the leisure areas or areas subject to more targeted enforcement, but at the aggregate level, they may not bring any substantial change. Therefore, the net effect of DUI-related trip generations is somewhat ambiguous.

2.4.7 Increased road crash risk at pick-up and drop-off locations (PUDO)

Morrison et al. explored the TNC service issue and its association with road traffic crashes using 372 million trips recorded between 2017-2018 in New York. Using a case-crossover design to overcome the absence of traffic volume data, the authors compared the activities (-1 week prior and +1 week after) in areas where a crash has occurred (Morrison et al., 2020). They compared the number of TNC trips, taxi pick-ups, and drop-offs during the duration. The team discovered that more crashes involved motorist and pedestrian injuries when a location or block saw an uptick in TNC services (Morrison et al., 2020). A similar trend did not exists between taxi rides and pedestrian/cyclist injuries (Morrison et al., 2020).

The key takeaway from this study is that TNC trips increase the number of crashes for motorists and pedestrians at pick-up and drop-off locations (PUDO).

2.4.8 An increased instance of distracted driving

National Highway Transport Safety Authority (NHTSA) defines distracted driving as "*any activity that diverts attention from driving, including talking or texting on your phone, eating and drinking, talking to people in your vehicle, fiddling with the stereo, entertainment or navigation system*" (NHTSA, 2022a). According to NHTSA's latest report, around 8% of the fatal crashes, 14% of total injury crashes, and 13% of all police-reported motor crashes were reported as distracted driving crashes, leading to 3142 people dead and 324,652 people injured (Stewart, 2022).

TNC trips are inherently distracting given that the drivers constantly carry out various activities like searching for passenger pick-up and drop-off locations and routes assigned to the ongoing trip by the TNC algorithm. Furthermore, when potential passengers place a ride request on the TNC app, the passenger's location is locally broadcasted to the nearest available drivers in the form of an alert sound. The driver has a particular time window to determine whether they will accept or decline the ride based on the distance and time to reach the passenger's location (NHTSA, 2022a; Richtel, 2014).

Executing multiple tasks in addition to providing quality services every time is challenging and therefore increases the probability of getting involved in a road crash.

2.4.9 Insurance and liability coverage for TNC

Before 2016, in order to enroll as a TNC driver, the individual needed to fulfill three basic requirements:

a. be of minimum age limit
b. hold a valid driver's license
c. maintain valid personal auto insurance policies (PAPs) to cover their driving activities

However, transporting personnel by undertaking commercial activities, often termed "*livery exclusion,*" is strictly prohibited under most of the PAPs insurance schemes. Insurance corporations include "*livery exclusions*" in the PAPs scheme to protect themselves from unwanted risks arising due to increased exposure resulting from a) additional miles of driving as a result of transporting a passenger from point A to point B., b) driving in unfamiliar areas, and in highly dense urban and traffic locations c) distracted driving instances.

After much deliberation with the public, on September 2013, the California Public Utilities Commission (CPUC) determined that TNCs are transportation charter-party (TCP) and required

- TNC drivers are to provide proof of their personal and commercial insurance in case of an accident
- TNCs are to hold a commercial liability insurance policy that is more stringent than the CPUC's current requirement for limousines[2], which shall include a minimum of one million dollars ($1,000,000) per-incident coverage for incidents involving

[2] CPUC General Order 115-G requires a carrier operating a vehicle that seats seven passengers or fewer to carry $750,000 of insurance coverage

TNC vehicles and drivers in transit to or during a TNC trip, regardless of whether personal insurance allows for coverage;

- Submit annual operational data to the CPUC, which includes reports regarding incidents/accidents and zero-tolerance complaints

However, the insurance requirements did not cover items related to comprehensive/collision, medical payments, uninsured/underinsured motorist (UM/UIM) for drivers, or any other optional coverages. According to the California Department of Insurance (CDI), this led to a coverage gap as it allowed "*TNC's operators and their liability policy not to provide coverage for 1) bodily injury to the TNC driver, 2) damages to the TNC driver's car or 3) bodily injury or physical damage caused by an uninsured or underinsured motorist.*"

The insurance coverage gap was further exposed in a judicial hearing for a road crash in January 2014 in San Francisco where a TNC driver hit a family of three, resulting in the death of a 6-year-old child and severe injuries. The victim's family sued Uber for her death because the driver was on the TNC operator app at the time of the crash. The TNC operator argued that there were no passengers when the car hit the family, so the company is not responsible for providing insurance protection to the victim and its driver.

It triggered the California Public Utilities Commission (CPUC), which regulates TNCs in the state and the state Department of Insurance, to develop models and recommendations to overcome insurance and coverage gaps (CPUC, 2016). Three significant changes were brought about as a result of the tragedy. The changes were passed into law in September 2014 and enacted on July 1, 2015, as Assembly Bill 2293 (Bonilla).

1. Explicitly define the TNC trip - CPUC splits the whole TNC operation into three periods and requires all three periods to be covered by commercial insurance.
    - Period 1 - When the driver opens the app and waits for a ride to match
    - Period 2 - When a ride gets matched, and the driver accepts the ride and traverses to the passenger's location

- Period 3 - The passengers board the TNC driver's car, travel to its destination, and safely exit it.

2. TNCs were mandated to provide a $1 million commercial liability policy when a TNC driver has a smartphone app turned on to accept rides.
3. It directed the insurance industry to develop new schemes tailored explicitly for TNC drivers.

In response to point 3, the National Association of Insurance Commissioners (NAIC), led by the then-California Insurance Commissioner, also adopted a white paper on insurance coverage for TNC services (NAIC, 2015, 2016). The white paper provides models and standards that state regulators can use to draft their policies related to insurance schemes for TNC services.

Overall, given the complications over the coverage gap, TNC drivers have little incentive to engage in rash driving, making them extra cautious of their choices to complete the TNC service rides.

2.4.10 Driver and passenger ratings

All the big TNC service operators provide both the driver and the passengers an option to rate each other and provide feedback at the end of the trip. The rating and feedback are anonymous, thereby ensuring the privacy of each other views. Uber and Lyft allow drivers and passengers to rate each other by providing 1 to 5 stars. If the rating is below 5, the customer and the driver can provide feedback as a comment (Lyft, 2023a; Uber, 2023b).

When the passenger gives five stars to the driver, it means that the ride conformed to TNC operators' expected level of service, which includes (but is not explicitly mentioned)

- Punctuality - arrives at the customer’s pick-up point address within the expected travel time
- Good navigation - follows the GPS-suggested route but does not follow blindly. The driver is aware of the routes, understands the traffic situation, and accepts the

passenger's route guidance if it believes they are safe and helps reach the destination on time.

- Safe driving - maintain speed within the posted speed limits or appropriate to the urban landscape and traffic conditions. The driver abides by traffic laws like obeying stop signs, wrongfully overtaking vehicles or bicycles, engaging in gradual lane changes, and negotiating speed-reducing measures that do not cause customers discomfort.
- Sociability - is always courteous and can adapt to the client's mood, thus making the rider feel safe and comfortable and making the ride fun. Accommodate passenger requests which are not illegal, unsafe, or unreasonable, like switching or completely shutting off the radio stations, requesting to drive further than the drop-off location, or taking a different route than what GPS indicates.
- Comfort - ensure the car is clean (exterior and interior), pleasant, and clutter-free. It means keeping the seats and trunk clean and free from crumbs, dust, hair, dirt, and other debris.

Companies understand that the ratings are perceptions and may differ with each customer. Uber averages these ratings once every 500 rides, while Lyft averages them once in 100. The companies take passenger ratings for their drivers far more seriously than the other way around. The ratings may have positive and negative consequences. The ratings get used to identifying patterns that may compromise the safety of the passengers (Brett Helling, 2018).

If the driver's rating is too low, it may affect its ability to match rides. If the driver has less than five stars, Uber allows the passenger to cancel the ride within 5 minutes after the initial request. If the driver's feedback is consistently below average, it may result in the deactivation of the driver from its platform. On the other hand, both Uber and Lyft incentivize good rating drivers by allotting them higher trip matches and premium service trips. In addition, Uber enrolls its higher-rating drivers in its driver's reward program, which rewards the drivers with cashback on gas and free tuition coverage (Brett Helling, 2018; Lenzo, 2016; Lyft, 2023b; Uber, 2023b, 2023a). Lyft provides additional perks like free tax services, public charging facilities, and roadside assistance (Lyft, 2023b).

2.5 TNC-focused studies

Given the general understanding of the framework and the mechanism by which TNC trips can potentially lead to a road crash, let us review the aspects covered by scientific literature. The following section synthesizes published and unpublished papers/literature on the impacts of TNC service trips on road traffic crashes and injuries.

Greenwood & Wattal were the first to investigate the impact of Uber's services, specifically UberX and Uber Black, on alcohol-related motor vehicle homicides in California counties (Greenwood & Wattal, 2017). Using DUI data available for these counties between 2009 - 2015, the paper suggested that the introduction of UberX contributed to reducing DUI rates by 3.6%. However, the paper found no association between Uber Black service DUI crashes. Their paper argued that one of the primary reasons Uber Black operations did not lead to reduced DUI instances and fatal crashes is the difference in trip costs. UberX offers trips at significant (~20-30%) price reductions compared to taxi rates, while Uber Black trips are typically ~20-30% more expensive than the taxi rates. The study's key takeaway is that the cost and availability of TNC services are crucial to attracting passenger demand (Greenwood & Wattal, 2017). Unless and until the cost of a TNC trip falls below the total perceived cost incurred using alternative modes of urban transportation (private vehicles or other public transportation model options), people are unwilling to make such a shift and continue to engage in illegal activity leading to DUI instances.

Dills & Mulholland's paper investigated the impact of Uber's entry in reducing road fatalities by each month and year for all US Counties from 2007 through 2015. The study found that road fatalities reduce by 1.6% (unweighted regression) and by 0.7%(weighted regression) for each passing quarter of the year Uber is available (Dills & Mulholland, 2018b). The proposed estimates are much smaller in percentage than those proposed by (Greenwood & Wattal, 2017). The paper also experiments with various forms of statistical model specification and finds such reduction rates robust. It also finds that once Uber has operated in a given county for more than a year, the fatal crashes declined by 17 - 30%

(Dills & Mulholland, 2018b). Contrary to the positive association with road fatalities, the paper finds no association between TNC services and DUI rates.

(Brazil & Kirk, 2016) also attempted to test Uber's claim that its presence within the city is positively helping the cities curb alcohol-involving (DUI) crashes and associated fatalities. The paper examines the relationship between Uber service (staggered) deployment and subsequent monthly traffic fatalities per year within the 100 largest metropolitan areas across the United States from 2005 to 2014. However, it fails to find any association between ridehailing services, total crashes, alcohol-involving (DUI) crashes, and weekend and holiday-specific traffic fatalities. Because Uber's service role out, market penetration, and service maturity at the city level may not be uniform across cities, the authors felt the need to revisit the concept. Their new study extended the period from 2009 - 2014 to 2009 – 2017. The revised research also evaluated the results for local variables (contexts to accommodate heterogeneity effects) like population size, density, urban centrality (the percent of the county's population that lives inside a census-designated urban area), counties demographic and socioeconomic characteristics, vehicle ownership, public transportation access and usage and alcohol access and consumption (Brazil & Kirk, 2020). The revised experiment also failed to find any association between Uber's availability and traffic fatalities at the county (aggregate) level.

One argument the supporters of TNC services present is that the combined annual vehicle miles traveled (VMTs) of all TNC companies operating in the United States is small and may not exceed 1-3% of the total 3.23 trillion miles clocked in 2019 (Federal Highway Administration, 2019). A study commissioned by Uber estimates the TNC trips to be in the range of 1-2.9% of the total VMTs for the six most significant metropolitan areas, namely Boston, MA; Chicago, IL; Los Angeles, CA; San Francisco, CA; Seattle, WA; and Washington, DC (Jackson & Owens, 2011). However, that same report shows that TNCs contribute a much larger share of VMT (2% to 13% of Total VMT) in the central counties of each of those metropolitan areas. Therefore, TNCs may have a significant effect where they are a substantial portion of VMT, but metropolitan areas are too large of an analysis area to detect an effect.

There is a legitimate concern about why TNC's share needs to be evaluated and is related to increased vehicle miles traveled (VMT). For example, conducting a quasi-natural experiment, Hennao and Marshall found that for 311 trips the study surveyed, the TNC services will increase trip VMT estimates by 83.5% compared to the system with TNC services not existing. (Henao & Marshall, 2019). For instance, a distance of 1 mile between two points A to B covered by a private car when replaced by a TNC service trip will end up becoming 1.83 miles long trip (VMT). Since the number of road crashes is directly proportional to the amount of travel, a rapid increase in the usage of TNC has the potential to amplify the risk of road crash exposure and reverse the road safety achievements of the last two decades. Furthermore, research carried out by the Institute of Transportation Studies (ITS), University of California, Davis, does not find any decrease in vehicle ownership among those who used TNCs and those who did not (Clewlow & Mishra, 2017), further reaffirming the concerns that such services are not single-handedly capable of reversing or pausing the increasing car ownership trends.

The study by (Barrios et al., 2019) critiques the premise of TNCs replacing private vehicle trips on two fronts. First, it argues that there is no rationale for the claim that TNC trips only replaces self-driven trips with the same mileage. On the contrary, the fact that a TNC driver needs to drive from their present location to the customer's location before transporting the customer to their desired location indicates that TNC trips are not just replacements for individual private vehicle trips. Second, it transforms the car (TNC service provider) into a productive asset to maximize its profits, i.e., offering the maximum number of trips at the shortest possible duration of time at the service. The desire to maximize individual profits leads TNC drivers to be nonstationary and drive more between places in the city in search of better fare prospects. Furthermore, TNC apps further incentivize its driver to stay on the road, even when the utilization is low. Covering the period between 2001 – 2016, for all areas with a population greater than or equal to 10,000 within the continental USA, (Barrios et al., 2019) study indicates that TNC services have increased VMT and led to a 3% rise in crash rates.

Given the wide diversity in the geography in the form of topology, public transportation, people's lifestyle, and socioeconomic trends, the impact of TNC usage at the regional scale may not be entirely forthcoming. Uber's in-house study to understand its service impact on metropolitan's total VMTs indicates a vast variation in their percentage contributions (Fehr & Peers, 2019). TNCs' contribution to total VMT rises from a marginal 1-2.9% at the metropolitan level to nearly 12.2 – 13.4% in denser (core counties) areas of San Francisco but remains stagnated at 1.7- 2.0% in Seattle (Fehr & Peers, 2019). Therefore, can one understand TNC's true impact by limiting the study to a given urban area?

Motivated by this question (Morrison et al., 2018) focused on studying the impact of Uber on road injury crashes and alcohol-involving (DUI) crash instances in 4 US cities - Las Vegas, Nevada; Reno, Nevada; Portland, Oregon; and San Antonio, Texas. Their selection of cities was simple: a) in all these cities, Uber must have launched itself between January 1, 2013 - December 31, 2015; b) abruptly ceased its operation either voluntarily or due to a regulatory ban after running its operation for at least three months or more, c) resumed its operations continuously between January 1, 2016, and June 30, 2016. According to the researchers, adopting such a methodology allowed them to measure the immediate impact of services on motor crashes. Except for Reno, the data analysis partially supported their hypothesis, meaning the resumption of TNC operations was associated with a decreased incidence of alcohol-involving crashes. It found a ~60% reduction in the relative accident rate related to alcohol and the implementation of ride-hailing platforms in Portland, Oregon. In San Antonio, Texas, such relationships stood at 58% (Morrison et al., 2018). However, the said model was a poorer fit than the null model. Further, no correlation between Uber's resumption and fewer overall road crashes in any of the cities. On the contrary, Uber's service resumption in San Antonio resulted in increased injury-related road crashes (Morrison et al., 2018).

(Peck, 2017) also studied the short-term impact of the Uber service launch on alcohol-related collisions in New York City at the borough level using data between January 2007 and July 2013. Overall, the New York City boroughs witnessed an average 17-35% (~40 collisions per month) reduction in alcohol-related collision rates. However, at the individual borough level, Uber services had the most significant (reducing) impact in Manhattan, middling effects for Brooklyn and the Bronx, and the most negligible impact in Queens borough. According to the author, one of the primary reasons these results are not equally spread over the New York borough and also not congruent to the results observed in other parts of the country has to do with the fact that other forms of public transit readily available in New York for over a century (Peck, 2017). The availability of alternative public transport modes has led many New York residents not to own a car and are therefore open to adopting a newer service mode as and when offered.

On the contrary, (Kontou & McDonald, 2020a) investigated the influence of Ride Austin, a local TNC service in Travis County, Texas, on vehicular crashes, injuries, fatalities, and c (DWI). By leveraging the real-world data available from RideAustin's open records, road crash data from Travis County in Texas along with monthly travel demand for the period between 2012 to Apr 2017, their analysis concludes that every 10% increase in TNC trips leads to ~0.12% decrease in road crashes, ~0.25% decrease in road injuries, and ~0.36% decrease in DWI incidence (Kontou & McDonald, 2020a). However, Austin is a predominantly car-oriented society, with less than 10% of the trips originating from other modes like walking, bicycling, and transit (Multimodal Community Advisory Committee, 2018). Therefore, there are not enough conflict points generated between different road user groups, and drivers are less likely to engage in aggressive driving. Furthermore, the origin-destination trips index is a weaker proxy for travel demand than VMT, so the results may not capture actual impacts.

(Anderson & Davis, 2021) reasons that using proxy variables to map the impact of TNC services in the urban area is inefficient and can only explain about 3% of variation in Census Tract (CT) levels. Therefore, utilizing proxy variables to explain the relationship between TNC and road crashes is a highly inefficient way to capture TNCs contribution. The paper utilizes Uber's propriety data (trip counts aggregated at the CT level) to explore its relationship with total crashes and alcohol-involving (DUI) crashes for all CT per month in the USA between July 1, 2017 - January 1, 2017 (excluding where either origin/destination = Airport or CTs (Census Tract) of Seattle, New York). The paper demonstrates that the statistical model is insignificant when a binary variable gets used to confirm the presence of TNC service in CTs. However, when proprietary data replaces the binary variable, the regression model is statistically robust, significant, and negative for alcohol-related traffic fatalities.

(Morrison et al., 2021) explored the issue of TNC service and its association with road traffic crashes using 372 million trips recorded between 2017-2018 in New York. Using a case-crossover design to overcome the absence of traffic volume data, the authors compared the activities (-1 week prior and +1 week after) in areas where a crash has occurred. They compared the number of TNC trips, taxi pick-ups, and drop-offs during the duration. The team discovered that more crashes involved motorist and pedestrian injuries when a location or block saw an uptick in TNC services. A similar trend did not exist between taxi rides and pedestrian/cyclist injuries, indicating that taxi and TNC services impact are entirely different.

In order to study the impact of the National Academy of Science, Engineering and Medicine's (NAS) recommendation to make alternate transportation (including ridehailing services) available at low cost, (Humphreys et al., 2021) examined the impact of philanthropic ridesharing service program ran by Evesham and Voorhees township under Evesham Saving Lives Program (between Sep 2015 - 2018). Using the difference-in-difference method and Bayesian Poisson Model (CAR), the paper attributes around 11% (morning), 20% (afternoon), and 38% (night) reduction in road crash due to the subsidized scheme (Humphreys et al., 2021). However, the magnitude of the contribution attributable to the scheme is dramatically higher than other studies mentioned above that struggled to find consistent results.

2.5.1 Is there something to learn from the taxi and limousine market sector?

Given no definite trend arising from the research, it is natural to broaden the scope and ask a generic question – do other transportation services similar to TNCs have the same effect on road crashes? Notably, the taxicab and limousine services and whether there is any overwhelming evidence to either support or argument against such ambiguous trends.

Dalziel and Job examined various aspects of taxi drivers and their relationship with road safety outcomes and submitted a detailed report to the Federal Office of Road Safety, the Government of Australia (Dalziel & Job, 1997a, 1997b). The study indicated that behavioral aspects such as the average length of shifts, vehicle types, increased anger expression, and economic pressure (need for more income) were much more important predictors than commonly assumed factors like age, time, kilometers traveled, employment, and shift types. The report also found that a) among all crash types, collisions with pedestrians were over-represented b) taxi crash rates were highest at the end of the weekend night shift. It also found that negative perception of taxi drivers, their safety, earnings, and respect by passengers often leads to increased anger expressions in the form of risk-taking behaviors and maneuvers, which increases taxicab's exposure to a road crash.

In a separate study, after observing a control group of 42 taxi drivers across two years, Dalziel and Job found that drivers with less break time in a day due to busy workloads were more likely to be involved in a road crash than others (Dalziel and Job 1997a). The paper also found older drivers to be safer than their younger counterparts. The study attributes the superiority of older drivers driving skills acquired over the years, refined to such an extent that it now far exceeds the average motorist and is better at handling fatigue and long hours of shift/workloads (Dalziel & Job, 1997a, 1997b).

Similarly, back home, after observing 130,000 taxies and livery crashes between 1990 and 1999, (Schaller, 2001) found that taxi and livery-cab drivers have lower crash rates than other drivers. The author suggests that strict licensing requirements and the driver's knowledge of streets and urban areas are a few possible reasons for lower crash rates among taxi drivers.

On the contrary, the (Schaller, 2001) study indicates that despite overall declines, injury rates remain high for taxi passengers. The authors find unrestrained taxi passengers as the prime reason for such high injury rates and estimate that such passengers are three times as likely to be injured compared to seatbelt-using passengers. Furthermore, the study concludes that bicycles are twice likely to collide with taxicabs than other vehicles because of "dooring," in which the passenger in a taxicab suddenly opens their doors without any inspection of oncoming bikes. Such incidents are also very prevalent in the TNCs service industry. Chicago City and Illinois have laws to prevent such crashes. The Chicago ordinance requires motorists to look before opening their car door and imposes fines of up to $500.00 on the motorist for obstructing a cyclist's path if such an event occurs.

Using the economic principle of "moral hazard" (Tay & Choi, 2016) presented an interesting perspective on the risk-taking capabilities of non-taxi owners vs. taxi owners. The study found that non-taxi owners with no incentives to drive carefully attempt to maximize their economic profitability by engaging in hazardous driving behaviors, thereby increasing their exposure to road crashes.

2.5.2 Mobile phone usage and distracted driving

Mobile phones, especially smartphones, are an inherent part of the success of TNC and its services. It is mainly because the TNC app is an intermediary platform that connects its contracted drivers and potential passengers. Furthermore, the smartphone is the primary medium for passengers to get trip quotes and assist in real-time vehicle tracking. From the driver's perspective, it helps identify the customer's pick-up and drop-off location and trace the optimal route based on real-time traffic conditions. However, there is an unwanted cost to it.

As explained by (Richtel 2014) - a driver receives a service call alert from Uber, Lyft, or any other TNC platform, in the form of a loud beeping alarm on the phone (through the TNC app). The driver typically has 15 seconds to respond to such a request. In these 15 seconds, the driver has to mentally visualize the area, calculate the distance between its present location and the customer address, and anticipate the time required to cover the distance before making a decision. Failure to respond within the stipulated time may mean the trip gets assigned to a different driver. If such instances get repeated several times in a row, it may also lead to the temporary suspension of the driver from the TNC app. It is like participating in a quiz contest show - to win, the participant is expected to hit the buzzer as soon as the question is displayed. Add the complexity of the urban landscape and the urban transport system, where many actors interact at varying spaces and times. The driver must circumnavigate such complexities while remaining vigilant on the road surface. Could this extensive engagement with the mobile device lead to distracted driving among TNC drivers?

Many past studies have shown that mobile phone usage harms driving performance (Barkana et al., 2004; Lesch & Hancock, 2004; McCartt et al., 2006; Mccartt et al., 2006; Foss et al., 2009). By reduction in driving performance, the studies meant vehicle control (Treffner & Barrett, 2004), attention (Beede & Kass, 2006; McCarley et al., 2016), workload (Patten et al., 2004), impaired eye scanning (Harbluk et al., 2007; Maples et al., 2008), and reaction time (Caird et al., 2008; Horrey & Wickens, 2006; Strayer & Drews, 2004; Troglauer et al., 2006). Past experimental studies indicate that one hour per month of cell phone usage while driving is directly associated with a 400-900% jump in the likelihood of the driver getting involved in a road crash (McEvoy et al., 2005; Violanti, 1998; Violanti & Marshall, 1996) and thus represents a serious traffic safety problem. Furthermore, studies have shown that switching from hand-held to hands-free devices does not help negate such risks (Caird et al., 2008; McCartt et al., 2006; Mccartt et al., 2006; McEvoy et al., 2005). However, it is worth noting that it is equally challenging to quantify the number or percentage of road crashes or deaths attributable to mobile phone usage, as drivers may not admit to using mobile phones when other physical causes are evident.

The National Association of Insurance Commissioners (NAIC), in their 2015 white paper titled "Transportation Network Company Insurance Principles for Legislators and Regulators," also acknowledges the risks of distracted driving (NAIC, 2015). Though the main objective of the white paper is to identify means to minimize the underwriting losses, the paper acknowledges the insurance coverage gaps that arise due to the activities undertaken by the TNC driver and its vehicles. Because most TNC drivers use personal cars to offer commercial services, the personal auto insurance lenders get exposed to additional risk than contemplated, thereby increasing their exposure to losses. The white paper identifies the following reasons for such heightened risks:

a) the driver drives additional miles against when the vehicle gets used for personal travel
b) because most of the TNC trips start and end in the urban area, which presents a geographic hazard which is high-traffic locations
c) TNC drivers often take routes that are non-familiar to them
d) TNC apps cause distractions while driving

e) risk generated due to mental stress from accepting a ride request and picking up and dropping off the passengers on time ("Transportation Network Company Insurance Principles for Legislators and Regulators," 2015).

2.5.3 Any lessons from abroad and those from developing countries?

Much of the discussion in the previous section of this chapter concentrated on the urban areas of the USA, where the traffic death rate per million inhabitants far exceeds those observed in countries like the UK, Germany, and Spain. So the question arises if TNCs impact other cities worldwide similarly to those found in US cities.

(Kirk et al., 2020) explored the question by exploiting the difference in the deployment of Uber across Britain to understand the implications of TNCs services on road traffic injuries. Their analysis found a marginal reduction in severe road crash injury cases but no change in serious crashes. Interestingly, they found contrasting results for London and the rest of Britain. Whereas severe injury cases declined outside of London, the number of cases increased within it. One of the explanations for such an observed trend is that TNC services are used as a substitute for rash or alcohol-involving (DUI) crash instances outside of London boroughs. At the same time, within London, it is a substitute for public transit (like subways and buses), which are already much safer than cars.

Using an annual number of victims (injured/death) between YR 2013 - 2019 (T= 7) in the 21 (N) districts of Madrid (NxT =147), (Flor et al., 2022) aimed to investigate whether the severity of traffic crashes in the city of Madrid has increased after the introduction of TNC services like Uber and Cabify. The paper results indicate that TNC services decrease severe injuries and deaths by 25% (Flor et al., 2022). However, the same is not valid for other road crash types who witnessed increased road crash instances, i.e., total victims (+3%) and minor injuries (+5.6%). These results are comparable to (Kirk et al., 2020), who found a 9% reduction in severe injuries in London after the arrival of Uber. The results are in-sync with those found by (Greenwood & Wattal, 2017), which showed that the introduction of TNC results in a 3.6%-5.6% decrease in the rate of motor vehicle homicides per quarter in California. However, the authors caution that such a trend may reverse in the coming years as TNC continues to attract users from other modes of transport

(Flor et al., 2022). If TNCs attract public transportation or nonmotorized users in a much larger capacity than their replaced car trips (riskier drivers), then the number of riskier drivers on the roads would not decline. As a result, the number of road crashes and associated injuries/deaths may subsequently increase.

Using road crash data corresponding between 2014 - 2018 for Madrid, (Flor et al., 2021) found that TNC services like Uber and Cabify have a positive (reducing) effect on seriously injured or fatal crashes during weekends and policy holidays. However, it has no significance on crashes related to alcohol and drugs. The paper's findings indicate that TNC service usage is prevalent in areas that are either denser or have an increased presence of leisure establishments. By definition, leisure establishments are dedicated areas or places offering catering, leisure, and entertainment activities. This finding contrasts with the (Brazil & Kirk, 2020) findings in which the author found an increased association between Uber service and total, weekend- and holiday-specific, and alcohol-involved fatalities in counties with high population density and urban centrality.

Blazquez et al. attempted to evaluate Uber's impact on alcohol-involving crash in the city of Santiago, Chile - one of the biggest markets for Uber in South America (Blazquez et al., 2021). The study takes a unique approach to the effect of TNC on road safety. According to the paper, most early Uber adopters were either highly educated or from well-off (higher household income) families, with easy access to or high penetration of Credit cards and smartphones. Since Uber transaction settlement primarily required a credit card and smartphone, its adoption during the initial period was concentrated mainly in the high socioeconomic community. Therefore, any analysis undertaken during earlier times of the Uber service launch must consider such disparity in use. The study further goes to classify 34 municipalities of Santiago into five economic groups High, High-middle, Middle, Middle-low, and Low and finds a mild positive (reducing) effect between Uber and high socioeconomic municipalities. A similar impact between Uber and vulnerable (low- socioeconomic) municipalities was not found.

2.6 Methodological traits used in understanding TNC impact on road safety

Due to the absence of available TNC-related trips, the primary independent variable of interest in most of the previous literature was limited to a dichotomous treatment indicator. i.e., 1: to indicate if the TNC service ran its operation; 0: to indicate if the TNC service did not run its operation.

Depending upon the geographical spread of the research and the type of road crashes examined, such binary variables get populated for areas (a_i), and measured over time t, usually a week, month, or quarter. Since the launch of TNC services in these areas is not always simultaneous, such a scenario turns into a classic post vs. pre-policy level experimental study. Difference-in-difference (DiD) estimation can then be used to exploit post and pre-intervention scenarios and thereby estimate the causal relationship between TNC and road crash variable. Indeed, it is the case for most studies (Barrios et al., 2019; Brazil & Kirk, 2020, 2016; Dills & Mulholland, 2018b; Peck, 2017) published in the intervening years since (Greenwood & Wattal, 2017) adopted such a technique to establish the effect of Uber's entry on alcohol-related motor vehicles.

Using a binary variable to represent TNC services presence, both (Morrison et al., 2018; Nazif-Munoz et al., 2022) exploited the limitation imposed by the interrupted time-series analysis method to explore the relationships between Uber operations and different motor vehicle road crashes. The interrupted time-series study is a time-series analysis conducted by viewing an observational unit repeatedly over time at equal intervals. The said time-series data is interrupted by an intervention at a known time. Hypothetically, had the interruption not taken place, the underlying trend of the time series (expected trend) may have continued unchanged. This counterfactual scenario provides a comparison for the evaluation of the impact of the intervention by examining any change occurring in the post-intervention period.

(Brazil & Kirk, 2020) adopted a more straightforward form of analysis in traditional count regression to model road fatality data. At the same time, Uber availability in the counties is still represented in binary form; the structure design allowed a test of the influence of other heterogeneity variables. Instead of presenting the coefficients of the list of parameters in the model, the authors categorize the association as high/mid/low. Categorization transforms the association between Uber service and road crashes from a change in percent increase/decrease to a marginal association, i.e., whether the association is a non-existent, weak, or strong category. Using a simplified regression model, (Anderson & Davis, 2021) demonstrated that replacing a binary variable acting as a proxy for TNC service with a continuous proprietary dataset may lead to a complete reversal of the results. Such that, having TNC services in cities may positively reduce alcohol involving fatalities.

Kontou et al. used Spatial panel fixed-effects lag and Spatial Autoregressive with additional Autoregressive error structure (SARAR) models to overcome the absence of a control group in their study to investigate the use of rideshare services influences the rate of vehicular crashes, injuries, fatalities, and DWI Offenses (Kontou & McDonald, 2020a). Unlike DiD, which allows no inter-temporal dependence of the events, the spatial panel data model assumes that the observations units are geographic and that there is some degree of dependence between these units (spatial) due to their location and distance between each other. By including the spatial and time fixed-effects, their paper reduced the bias from unobserved factors generated from the continuous interaction between human behavior, vehicle, road infrastructure, and environmental conditions over space and time.

2.7 Factors influencing the impact of TNC on road safety outcome

Brazil et al. hypothesize that local, i.e., county demographic and socioeconomic characteristics, have a more significant influence on the consumption of TNC services and, therefore, should be part of the statistical model structure (Brazil & Kirk, 2020). They evaluate their hypothesis at various levels. Initially, only county-level characteristics participate in the model as covariates. The parameters include total population, density, and urban centrality, where Urban centrality gets defined as the percentage of the county population living inside a census-designated urban area.

Later on, the paper goes on to explore the influence of transport accessibility and usage in influencing the TNC and road crash relationship by including variables like the number of available vehicles per housing unit, the percentage of total vehicle miles traveled (VMT) attributed to public transit vehicles, the percentage of total VMT attributed to rail specific public transit, and measures of public transit coverage and service frequency, where

- Public transit coverage is measured as the share of working-age residents living in block groups with access to at least one transit stop within ¾ mile of their population-weighted centroid.
- Public Transit Service frequency is measured as the average time in minutes the commuters must wait between bus or train stops.

In order to test the strength of the relationship between Uber availability and traffic fatalities, they include socioeconomic characteristics like log median household income, the percent of residents 25 years and over with a college degree, and the percent of residents between 20 to 39 years old. The assumption is that higher-income people, single-earning young professionals (aged between 20 -39 years), have more profanity in adopting new technology-based services than other community sections. At the same time, they also have more elasticity to absorb the price difference TNC service offers.

Lastly, to test Uber's claim that their service helps the city reduce drink-driving-related road instances, they include parameters like the number of drinking establishments per county area, the percentage of adults reporting drinking any alcohol in the past 30 days before being surveyed, and the percent of adults reporting binge or heavy drinking in the past 30 days before being surveyed. The assumption is that TNC service encourages people to spend more time socializing with their colleagues and friends, leading to increased binge drinking and exposing them to newer risks during transportation, where:

- drinking establishments are that property which has licenses to sell and consume alcohol on its premise;
- binge drinking is those instances involving where male indulges in 5 or more drinks or four or more drinks on occasion for women;
- heavy drinking is defined as drinking 15 or more drinks for men or eight or more drinks for women per week

2.8 Summary

Transportation Network Companies (TNC), also known as rideshare, e-hail, or ride-hailing companies, provide on-demand transportation services. Unlike traditional vehicle-for-hire services like taxis and limousines, TNCs do not maintain their vehicle fleets and rely on individually contracted private drivers to transport potential passengers from one place to another. Potential passengers request a ride through TNC online platform application, typically installed in a smartphone, and get matched to the nearest driver who transports the customer to the desired destination at a predetermined rate (trip cost). While such an arrangement presents numerous advantages in terms of cheaper travel costs, travel time, ease of booking and payment, and greater supply of transportation choices, their service is not short of disadvantages.

Road crash is one of the many traffic externalities of TNC Service operations. What separates road crashes from other externalities like vehicle miles traveled, road congestion, and increased energy consumption is that it is not apparent at first glance. Therefore it explains the framework and, later on, the mechanism by which TNC trips can lead to more or fewer crashes. Various means, namely the impact of the increased vehicle on roads (exposure), increased vehicle miles traveled (VMT), deteriorating road congestion and travel speed, increased interaction with other road users at pick-up and drop-off locations (PUDO), and distracted driving, are discussed.

A review of key published and unpublished papers/literature on the impacts of TNC service trips on road traffic crashes and injuries follows it. A few of the key messages emerging from the existing literature are:

- In the absence of trip level, most studies resorted to using a dichotomous variable to measure the presence or absence of TNC services in the urban area, with the majority of the studies opting for using difference-in-difference methods for evaluating TNC service relationship with road crash outcomes;
- Most of the peer-reviewed studies were set in the US, with a few set outside in the UK, Spain, Brazil, Chile, and South Africa.

- A few of the primary dependent variables whose association with TNC services are explored include total crashes, fatal crashes, alcohol-involving (DUI) crashes, and property-damage-only (PDO);
- Most of the studies opted for a 2-dimensional panel data structure with either county, metropolitan areas, or city representing the spatial unit, while week or month is the time dimension. Few smartly utilized disruptions which occurred in TNC service operations like citywide strikes or bans followed by resumption of services as natural breaks over time to analyze the relationship on a single spatial location.
- Overall, even after utilizing the same method and data setup, the results are found to be surprisingly mixed
    - (Greenwood & Wattal, 2017) found DUI rates to reduce by 3 – 4%; (Morrison et al., 2018) showed that only one out of the four selected city DUI crashes reduced due to TNC operations; (Peck, 2017) demonstrates a 17-35% drop in DUI rates in New York; using proprietary data obtained from Uber, (Anderson & Davis, 2021) found Uber's relationship with alcohol-related traffic fatalities negative; (Humphreys et al., 2021) paper attributes around 11% (morning), 20% (afternoon), and 38% (night) reduction in road crash due to the subsidized TNC scheme; (Kontou & McDonald, 2020a) concludes that every 10% increase in TNC trips leads to ~0.12% decrease in road crashes ~0.25% decrease in road injuries, and ~0.36% decrease in DWI incidence,
    - (Dills & Mulholland, 2018a) found a marginal negative association between total crashes and TNC operations but found no association between TNC services and DUI rates
    - (Brazil & Kirk, 2020, 2016) fails to find any association between ridehailing services, total crashes, drink-driving, and weekend and holiday-specific traffic fatalities,
    - (Flor et al., 2022) TNC services decreased severe injuries and deaths by 25% but contributed to increased total victims (+3%) and minor injuries (+5.6%) in Madrid, Spain.

    - (Barrios et al., 2019) study indicates that TNC services have increased VMT and led to a 3% rise in crash rates.
    - (Morrison et al., 2021) found both TNC trips, pick-ups, and drop-offs to be positively associated with the motorist and pedestrian injuries.
    - (Kirk et al., 2020) found severe injury cases declined outside of London, while the number of cases increased within it.
- TNC operation resembles a traditional taxi and limousine service quite a lot. The section reviews the taxi literature to understand road safety trends related to taxi crashes.
    - The research indicates that behavioral aspects are much more impactful than traffic characteristics
    - Older taxi drivers are safer than their younger counterparts.
    - The strict bureaucratic process acts like a natural barrier and allows taxi drivers to have far lower crash rates than other drivers
- To measure the association between TNC travel and road safety
    - most studies used dichotomous dummy variables to represent TNC's presence in the urban area, mainly because granular local-scale TNC trip data is hard to obtain.
    - Difference-in-difference (DiD) estimation is the preferred statistical method to exploit the naturally occurring post and pre-intervention scenarios (Barrios et al., 2019; Brazil & Kirk, 2020, 2016; Dills & Mulholland, 2018b; Greenwood & Wattal, 2017; Peck, 2017).
    - The interrupted time-series analysis method gets utilized to explore the relationship between Uber operations and motor vehicle crashes (Morrison et al., 2018; Nazif-Munoz et al., 2022).
    - A more traditional road safety model in the form of count regression models and spatial data models is also used to tap spatiotemporal units of analysis, yet findings remain inconsistent (Brazil & Kirk, 2016; Kontou & McDonald, 2020b).
- The relationship between TNC services and various road traffic crash types is closely related to the following:

- Vehicular traffic flow (volume, speed, pattern),
- Local spatial characteristics of the urban area play a vital role in evaluating the impact of TNC services on road crashes and, therefore, should be studied. That is to say that the relationship between TNCs and road safety outcomes is a function of residents' travel habits and travel patterns and how they use TNCs for local commutes.
- The local (urban) impact may differ widely from the overall (on average) impact at the metropolitan level. Therefore, it implies a need to undertake a spatially complex network-level analysis between TNC services and road safety outcomes.

CHAPTER 3 DATA AND METHODS

3.1 Background

Chapter 2 provided a detailed overview of the objective of the past scientific literature, the methods adopted, and their findings related to TNC service and road safety crashes. Although the literature demonstrates the varied adoption of methodological (statistical models) traits, it fails to provide a definitive trend in the outcomes. Findings are also not consistent across studies that follow similar methods. The findings suggest that heterogeneous effects at the geographical and road network levels significantly influence the actual impacts between TNC services and road crashes.

This chapter is divided into two major parts. The data section provides details of the data used to perform the analysis. The method section presents the framework adopted to analyze the relationship between road crashes (crash frequencies) and TNC-related parameters on a road network level.

3.2 Research methodology overview

Like past research, the study presents the association between TNC services and road safety outcomes in a before-and-after layout, where 2010 represents a scenario when TNC activities were negligible in numbers (assumed to be zero), while 2016 represents a condition when they were not. Figure 3-1 provides a graphical representation of the whole process. For each year, an estimation data file is compiled with one observation for each road facility type (FT). There are four FTs – Freeways, Arterials, Collectors, and Locals, with each length aggregated at the census tract (CT) level.

There are three parts to the data a) traffic flow estimates like vehicular speed, VMT, which represents average weekday conditions in the fall of each year b) TNC-related estimates in the form of TNC VMT, PUDO and also represents average weekday conditions in the fall of 2016, while zero for 2010 c) estimates of road crash frequency counts that occurred at each of the four facility types clustered again at the CT level. The main categories of road traffic crashes analyzed are total crashes, fatal and injury crashes, pedestrian and bicyclist crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes.

Fixed-effects count (panel) outcome models are estimated where the dependent variable is road crash frequency counts. The descriptive variables include the background traffic levels, vehicular speeds, TNC volumes, and TNC PUDO. Variables that do not change between the two years, such as roadway geometry (unless there is a construction project), are absorbed into the fixed effects. The result is that coefficients are estimated based on the change between 2010 and 2016 conditions, mitigating the risk that an excluded variable that does not change biases the coefficient estimates.

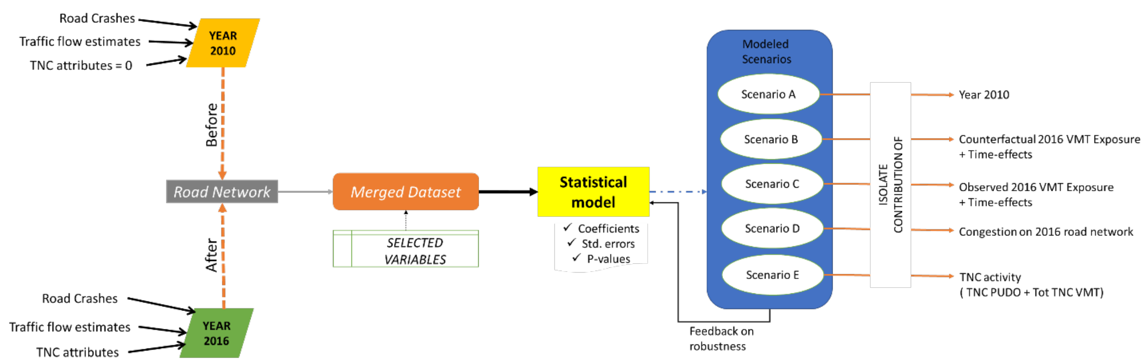

Figure 3-1: Research methodology used to evaluate the relationship between TNC operations and road crash outcomes

3.3 Data

This section provides details of the data used to perform the analysis. It includes the source of traffic attributes, namely TNC trip data, traffic flow characteristics, and road crashes. It also provides the detailed procedural steps followed to convert and merge different sources of these data into a singleton dataset used to conduct the statistical analysis.

3.3.1 Data Sources

3.3.1.1 Focus Study Area

The research uses data from San Francisco (SF) County, California. The county covers 47 square miles of land area at the end of the San Francisco Peninsula. It represents the densest residential and commercial location in the Bay Area, with an estimated population of 815,201 as of 2021 (US Census Bureau, 2022). It is also the first city where Uber rendered the inaugural TNC service trip in May 2010. Subsequently, Lyft hits San Francisco Streets in the year 2012. Together they expanded and now control the dominant share of TNC trips in the San Francisco Area. Fehr & Peers estimates that 12 to 14 percent of the total VMT generated in the San Francisco county limit is caused by TNC services like Lyft and Uber (Fehr & Peers, 2019). Studies by Roy et al., and Erhardt et al., demonstrate that between 2010 and 2016, almost half of the SF VMT increase is attributable to TNCs (G. Erhardt et al., 2019; Roy et al., 2020).

3.3.1.2 Traffic flow estimates

The study leverages local-scale granular traffic data made available by San Francisco's travel demand model, the SF-CHAMP model, for both the 2010 and 2016 periods. The traffic estimates produced for all road segments for both 2010 and 2016 are sensitive to San Francisco's observed travel patterns, the average vehicle ownership characteristics, population, employment, usage of the public transport system, non-motorized facilities, and temporal variation by time of day (*SF-CHAMP Modeling*, 2002). The SF-CHAMP model has been continuously tweaked to reflect changes and used for almost two decades to ensure that the local trade-off factors like demographic influence, availability, and quality of alternative mode choices are sufficiently captured (*SF-CHAMP Modeling*, 2002). It is successfully applied to analyze policy and infrastructure changes (Brisson et al., 2012; Castiglione et al., 2006).

The dataset is generated using SF-CHAMP 5.2.0. The data represent average weekday conditions and is available for each road segment for five different time-of-days (ToD) of each year. It is worth noting that the SF CHAMP model uses actual inputs, not forecasts, thereby avoiding discrepancies and errors (*SF-CHAMP Modeling*, 2002).

It also means that the sensitive travel demand like SF-CHAMP that includes link level attributes for every street and city like length, number of lanes, capacity, turn restrictions, and facility type and which takes into account the 2016 population, employment, and network inputs to produce a robust counterfactual case where TNCs do not exist. This dataset is termed Counterfactual (CF) 2016 and will be used to test and validate the impact of TNC services on road safety outcomes.

3.3.1.3 TNC data estimates

TNC data compliments the traffic flow estimates described in section 3.3.1.2. Between mid-November to mid-December of 2016, excluding dates around the Thanksgiving 2016 holiday period, the San Francisco County Transportation Authority (SFCTA), in collaboration with Northeastern University, generated a profile of local TNC usage in the SF county area by scrapping Uber and Lyft servers every five seconds via their application programming interface (APIs). Cooper et al. describe the procedure adopted to collect and process the data in considerable detail (Cooper et al., 2018). The data were further processed and cleaned to identify pick-up and drop-off (PUDO) locations, with vehicle trajectories tracked back between these locations to define a TNC trip. Erhardt et al. 2019 further processed the data to identify the total number of TNC volumes and pick-ups and drop-offs (PUDO) occurring on each road segment in San Francisco county by the time of day (TOD) (G. Erhardt et al., 2019).

The calibrated SF-CHAMP model for 2010 does not account for TNCs and is assumed not to exist during this period. We have two scenarios from SF-CHAMP for 2016: with and without TNCs. Both were developed by (Roy et al., 2020). The 2016 scenario with TNCs incorporates TNC volumes in three ways:

1. It pre-loads TNC deadheading travel (travel without a passenger), directly observed through data collection.
2. It includes a trip table of TNC vehicle trips serving passengers as a separate class in a multi-class traffic assignment. For these trips, we observe the origin and destination of the trips, with the route inferred through the assignment.
3. It assigns the pick-ups and drop-offs to specific network links and estimates their effect on traffic congestion with previously estimated coefficients that can indicate how long each pick-up or drop-off blocks the right-most traffic lane (G. Erhardt et al., 2019).

These travel estimates are for typical weekday conditions, segmented by the time of day. The 2016 scenario without TNCs excludes these three factors and estimates how population growth, employment growth, and network changes changed traffic volumes and speeds between 2010 and 2016.

3.3.1.4 Road Crash Data

Crash data representative of San Francisco County, California, were obtained from Statewide Integrated Traffic Records System (SWITRS) (*CHP-SWITRS*, 2019). By California law, all enforcement agencies must report fatal or injury collisions (classified under the KABCO injury scale) to the California Highway Patrol (CHP), which updates and pushes the data into the SWITRS database (*CHP-SWITRS*, 2019). The study extracts all road crashes in SF County between January 1, 2010 - December 31, 2010, and January 1, 2016 - December 31, 2016, to represent total crashes for YR 2010 and YR 2016, respectively. The crash data is re-examined to identify missing information, especially latitudes, and longitudes. For records missing such information, we retrieve their geographic coordinates using *geocoding* (*Overview | Geocoding API*, 2022) process, in which land-based information (e.g., street address, primary road, secondary road, the direction of the road segment) fetches geographic coordinates. Section 3.3.1.5 explains the process in detail.

There are five main crash types captured in the SWITRS database, namely:

1. Fatal Injury (K): Death due to injuries sustained in a collision or an injury resulting in death within 30 days of the collision.
2. Severe Injury (A): An injury other than a fatal injury that includes the following
    a. Broken or fractured bones,
    b. Dislocated or distorted limbs,
    c. Severe lacerations, Skull, spinal, chest, or abdominal injuries that go beyond “Other Visible Injuries,”
    d. Unconsciousness at or when taken from the collision scene
3. Other Visible Injuries (B): An injury other than a fatal or severe injury is evident to observers at the collision scene. Other visible injuries include

a. Bruises, discoloration, or swelling,
b. Minor lacerations or abrasions and
c. Minor Burns

4. Complaint of Pain Injuries (C) - This classification could contain authentic internal, other non-visible injuries and fraudulent injury claims. "Complaint of Pain" includes
    a. persons who seem dazed, confused, or incoherent (unless such behavior is attributed to intoxication, extreme age, illness, or mental infirmities)
    b. persons who are limping or complaining of pain or nausea but do not have visible injuries
    c. any person who may have been unconscious as a result of the collision, although it appears he/she has recovered
    d. persons who say they want to be listed as injured but do not appear to be so
5. Property Damage Only [PDO] (O) – Any road crash that neither results in fatal nor personal injury is classified as a property damage only (PDO) crash. Here the primary damage is to the real property instead of a person or persons. The primary type of property damage that may occur in a road crash is damage to the vehicles involved or the personal property inside the car. Any PDO crash amounting to 1000$ or more is required to be reported to the state agency by California law.

Classification of road crashes by crash types, as mentioned above, is also sometimes referred to as KABCO Injury Classification Scale.

The series of figures below presents the univariate distribution of road crashes (by types) for each year, i.e., 2010 and 2016, by counting the number of observations that fall within discrete bins having a width equal to 1.

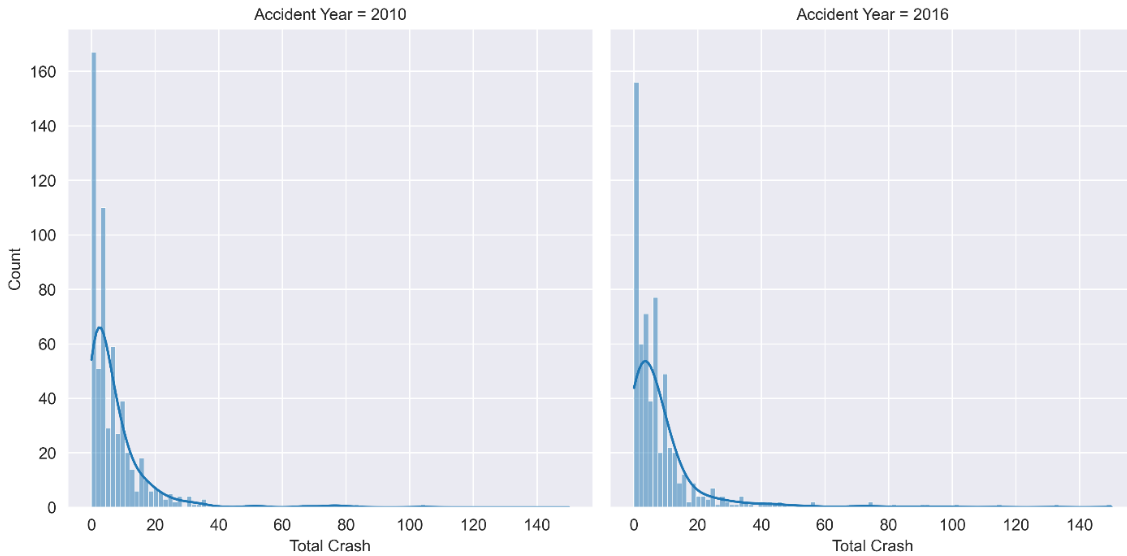

Figure 3-2: frequency distribution for total crashes by individual year

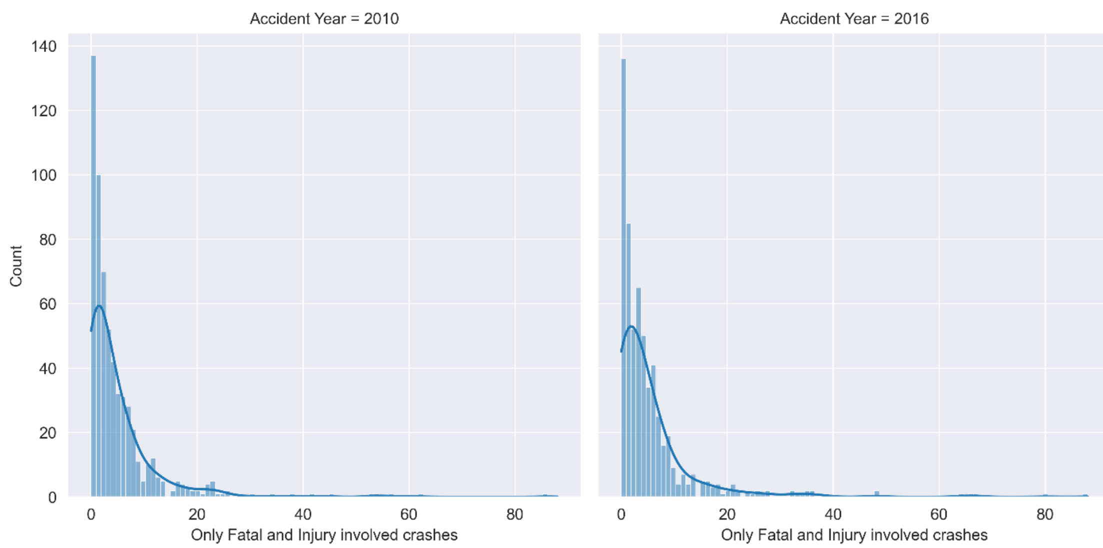

Figure 3-3: frequency distribution for all fatal and injury crashes by individual year

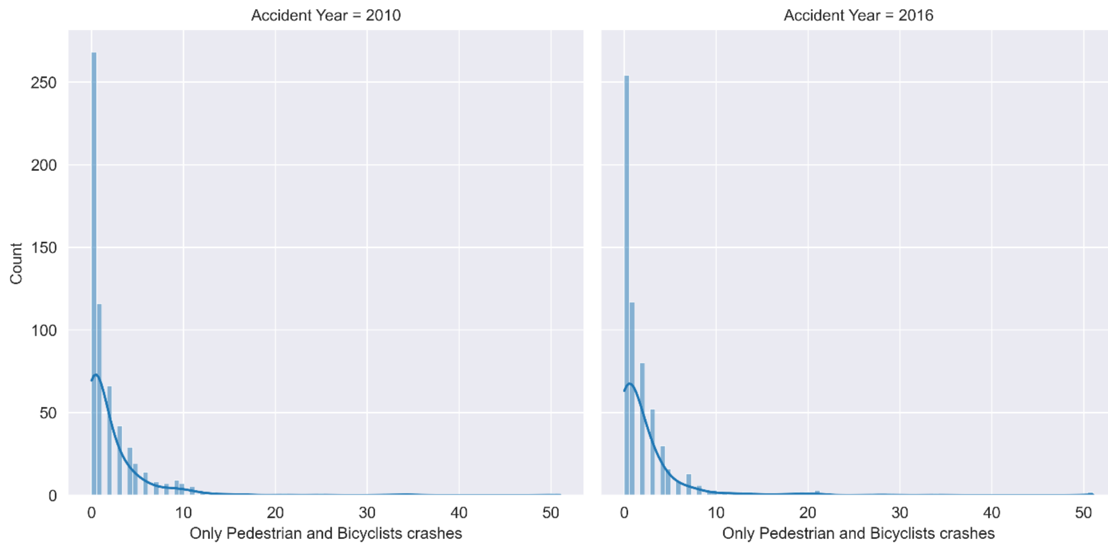

Figure 3-4: frequency distribution for crashes involving pedestrians and bicyclists by individual year

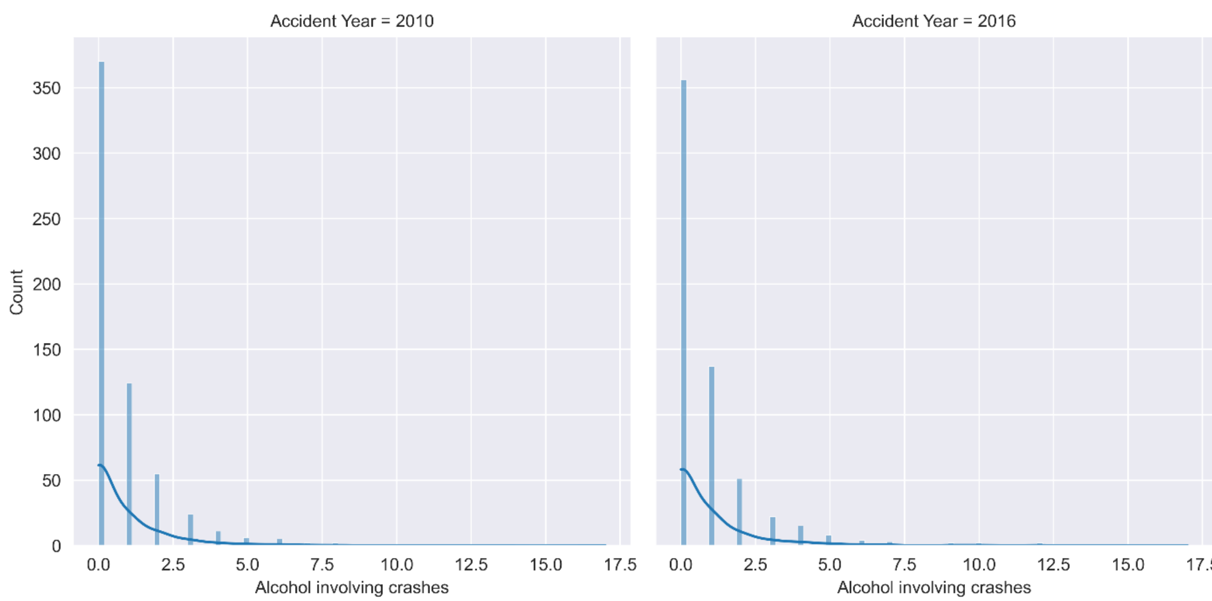

Figure 3-5: frequency distribution for alcohol-involving crashes

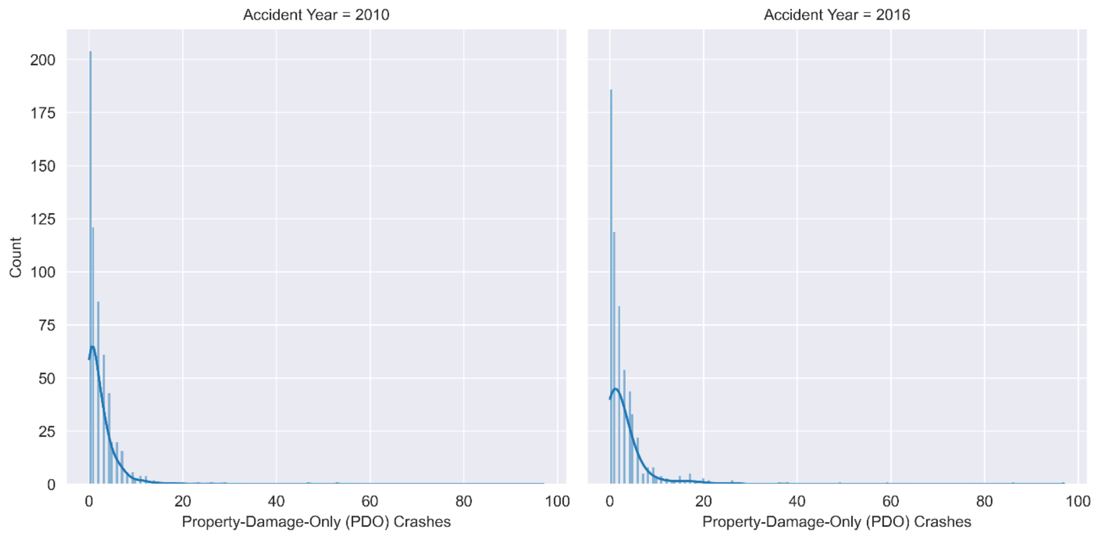

Figure 3-6: frequency distribution for property-damage-only crashes

3.3.1.5 Fetching missing geo-coordinates of a road crash

Road crashes for 2010 and 2016 are fetched from the SWITRS database using Query & Map functionality available online. A step-by-step method to access crash data is found here. The SWITRS database contains information about all types of crashes, including fatal, injury (severe), injury (other visible), injury (complaint of pain), and PDO (property damage only).

Approximately 90% of all the fatal and severe injury crashes maintained in the SWITRS database are geocoded through the process developed at SafeTREC University of California, Berkeley. However, crashes involving injury (complaint of pain) and PDOs do not have spatial locations because most data is not entered immediately. When a road crash gets reported to the police, the officer's first response is to clear the original crash site and ensure that traffic flow gets restored to normalcy. Later, a crash form is filled out if the officer identifies that the crash damages are >1,000$ or the vehicle needs to be towed. Suppose the crash form is filled out after the vehicle is moved. There is a high probability that geolocation information of the road crash is left blank or highly inaccurate. If the road crash is self-reported, the crash parties may not provide the crash location or do not fully recollect the crash scene resulting in the missing information.

Given that the research objective is to understand the impact of TNC service operation on road safety outcomes, there is a considerable expectation of a) mapping out all the road crashes, including PDOs b) geolocation of each road crash with reasonable accuracy. Therefore, a custom geocoding process is created to parse the text information in "primary road," "secondary road," "distance," "direction," and "intersection" attributes into geographic coordinates.

The process is like the method adopted by SafeTREC but instead utilizes open-source data and tools. The process can be divided into three parts.

A. Step I - Identify coordinates of the road intersections - For this step, we utilize Overpass Turbo API, a freely available web-based data mining tool for OpenStreetMap (OSM). We use the API query functionality to manipulate OSM data elements to extract coordinates of all road intersections within the SF county. The outcome is checked for various conditions to ensure that false positive road intersections are kept to a minimum. The resultant output contains the list of all road intersections with SF county, its coordinates, and the street names of all participating road segments.
B. Step II - Address locator – Road crash data records are passed over a custom address locator script, which reads the information in "primary road," "secondary road," and "intersection" and searches the information in the output from Step A. If the script finds a match, a (point geometry) marker object is created. The marker is assigned the coordinates of the matched road intersection.
C. Step III - Offsetting the coordinate to identify the actual position – The "direction" attribute is used to update the orientation of the marker position. In contrast, "distance" is used to move the marker from the road intersection by the distance along the primary road.

The position of the marker after undertaking steps A, B, and C is the projected coordinate values of the road crash. The script returns no such information if either of the "primary road," "secondary road," or "intersection" values are missing in a crash record.

3.3.2 Data Processing

3.3.2.1 Unit of analysis

Various styles of observational units have been used in the crash-count modeling literature. It includes counties, regions, districts, census tracts, road intersections, and roadway segments. Each has its merits and demerits, with each level of data aggregation leading to somewhat different results.

For instance, area-based crash prediction models allow area-level demographic and socioeconomic characteristics to be accounted for in the statistical models. These datasets are more readily available, enabling the analysis to infer explanatory variables' effects more accurately. They are used without or in combination with engineering-related variables like average speed, traffic volume, roadway geometry, and pavement characteristics. In this regard, spatial econometric models are gaining traction but are relatively complex to implement. Kontou et al. used spatial distributions of TNC usage, road safety outcomes, and other socio-economic characteristics at the Census Tract (CT) level to test associations of the launch of TNCs with road injuries (Kontou & McDonald, 2020b), while Wang and Kockelman and Li et al. demonstrated such methods as a viable alternative to the discrete/integer counts, at smaller geographical levels (Li et al., 2007; Wang & Kockelman, 2013).

On the contrary, road segment-based analyses are more focused and used to study the effect of traffic characteristics (like VMT, AADT, and speed limits), road features (like horizontal, vertical alignment, shoulders, curbs, road surface), and other variables like weather conditions on the temporal patterns of motor vehicle crashes (rate, frequency, fatality, injury, duration, severity) measured either at hourly, daily, monthly, and yearly scale (Balkin & Ord, 2001; Fridstrom et al., 1995; Johansson, 1996; J. Ma & Kockelman, 2006; Miaou, 1994, 1996, 2001; Miaou et al., 1993; Miaou & Lum, 1993; Pernia et al., 2004; Vogt, 1999; Vogt & J.G, Bared, 1998; Zegeer, C.V., et al., 2002)

So how do we choose which unit of analysis is appropriate for our analysis?

Miranda-Moreno et al. did an interesting study on the influence of build environment (BE) characteristics like land use types, road network, transit supply, and demographic factors on pedestrian activity or pedestrian-vehicle collision occurrence. Two-equation modeling framework was used to simultaneously model pedestrian activity (in log-linear form, as an exposure variable) and crash counts (using a standard negative binomial model structure) at signalized intersections in Montreal, Canada. Their analysis concludes that while BE is a strong predictor of pedestrian activity, after controlling for exposure, it has a minor effect on collision frequency (Miranda-Moreno et al., 2011).

Furthermore, there will always be some unobserved explanatory variables that influence crash frequency but could not be made part of the model. Such a situation is overcome by arranging them in a panel data structure form and introducing fixed effects which can eventually help deal with heterogeneity among missing variables.

Given that the objective of the dissertation is to study the impact of TNC services on road safety outcomes in the form of crash frequencies, it was decided to choose the road segment as the unit of analysis.

However, we quickly faced another hurdle in developing a sensible crash count model because of the sheer number of short-length (unequal) road segments within the road network. In addition, urban road segments vary in design and use levels. It also meant that crashes on short-length road segments were few and hard to model, which increased the preponderance of zero crash observations, i.e., the results at this scale appeared to be noisy. Shifting the overall sample mean to near zero makes the count data model extremely inefficient.

Therefore, to overcome these issues, we adopted another approach. We first classified roadway segment observations into relatively homogeneous facility types (FT), as in Table 3-1. The FT is so organized that it brought minimum variance to speed variables, one of the main drivers of road crashes. Later, the FT attributes - length and the number of crash counts are aggregated into a spatial unit. A variety of spatial units have been employed in the literature, such as regions (Washington et al., 1999), counties (Miaou et al., 2003), zip codes (Girasek & Taylor, 2010), census tracts (Wang & Kockelman, 2013), block groups (Levine et al., 1995), and traffic analysis zones (Xu et al., 2014).

We use Census Tract (CT) because:

a) it is the most common spatial unit to undertake macro-level model development

b) it allowed the cluster to be less dominated by local roads, which had the largest share of observations but a minority of crashes.

Table 3-1: Reclassified Facility Types (FT)

(reclassified) Facility Type (FT)	FT code	Description
Freeways	1,2,3,5	Roads classified as Fwy-Fwy Connector or Freeway or Expressway or Ramps
Arterials	7,12,15	Major Arterial, Minor Arterial, and Super Arterial roads
Collectors	4	Collector roads
Locals	11	All local roads

3.3.2.2 Road Network

The road network and its corresponding attributes are stored in a tabulated format called *shapefiles*. ESRI developed and regulated the shapefile format, the most widely used data format in geospatial sciences. Shapefile is part of open format data (interoperable) specifications, and most GIS suites, including [ESRI](), [QGIS](), or any third-party proprietary geographic information system (GIS), could easily read the information. Typically, a shapefile maintains the topological and nontopological attributes in a fixed tabular format. The shapefile can store simple (vector features) like points, lines, polylines, or a polygon to non-simple, complex geometry structures.

As mentioned earlier, SF-CHAMP divides the whole day into five times of day, i.e., 3 to 6 am (EA), 6 to 9 am (AM), 9 am to 3.30 pm (MD), 3.30 – 6.30 pm (PM), and 6-30 pm – 3.00 am (EV) and therefore the study had five different road network files available for each year, i.e., YR 2010 and YR 2016.

Therefore, the initial step is to merge all these road network files into a singular data. The study uses Python programming language and the QGIS platform to undertake desired manipulation and visualization.

3.3.2.3 Selecting the reference coordinate system

Explicitly defining the map projections before conducting any spatial exercise is usually a good practice. Having all spatial features (road network and road crashes) in the same coordinate system avoids any topographic errors while undertaking further complex geometry analyses. There are two types of a projection system

- Geographic Coordinate system (GCS) - Geographic coordinate system uses a three-dimensional spherical surface to define the location of an object (point, line, and polygon) on Earth. The object's location is measured from the center of the Earth using two relative planes: a plane parallel to the equator and an imaginary line connecting the North and South Pole and Greenwich, England. The distance is traditionally measured either in decimal degrees or in degrees, minutes, and seconds (DMS). The angular distance parallel to the equator is termed as latitude (+90 deg towards the North Pole, -90 deg towards the South Pole), while longitude range from -180 deg (traveling to the west) to +180deg (traveling to the east) with endpoint constantly merging with North & South poles. They are optimal when someone needs to locate places on the Earth or analysis is to be undertaken globally. GCS is not suitable for tasks involving "distance." Therefore the study adopts a projected coordinate system (PCS). (*Desktop Help 10.0 - Map Projections*, 2013; Gimond, 2022; Holdgraf & Wasser, 2020)
- Projected Coordinate system (PCS) - In a projected coordinated system (PCS), the 3-dimensional representation of an object on the Earth's surface gets transformed into a 2-dimensional object. It transforms an object from a spherical or curved surface (like our Earth's body) to a flat planar coordinate. The coordinate system consists of an x-axis and a y-axis. The X-axis and Y-axis intersect at right angles, and the point of intersection is also called the *origin* with coordinates (0,0). Transforming from GCS to PCS is termed *reprojection.* Reprojection is always

associated with a mathematical algorithm (and the reference point called *datum*) that controls the distortion that may arise from transforming a 3-D object into a 2-D object. Therefore, the measurement units become consistent and equally spaced across both axes. As a result, calculating the distance or an area is relatively easy to work with, provided the covered area is not too large (*Desktop Help 10.0 - Map Projections*, 2013; Gimond, 2022; Holdgraf & Wasser, 2020; Mieno, 2022).

Given that the study area is roughly 47 square miles, the study uses ESPG:3857 as its spatial projection system. ESPG:3857 is a popular projected coordinate system used by online mapping engines like Google Maps and OpenStreetMap.

3.3.2.4 Keeping only the relevant records by filtering

After the reprojection, the shapefile is filtered to keep only those features belonging to facility types as described in SF-Champ documentation (see the Table below). The shapefile is also checked for assigned free-flow speed and capacity values to remove any inconsistencies in the data records. If there is a discrepancy in the observed free-flow speed values, then the values in shapefiles are updated to the desired values.

Table 3-2: referenced free-flow speed as per SF-CHAMP guidelines

FREEFLOW SPEED

FT \ AT	Description	0	1	2	3	4	5
		Regional Core	CBD	Urban Biz	Urban	Suburban	Rural
1	Fwy-Fwy Connector	30mph	35mph	40mph	40mph	50mph	50mph
2	Freeway	45mph	50mph	55mph	60mph	65mph	65mph
3	Expressway	45mph	45mph	55mph	60mph	60mph	60mph
4	Collector	20mph	25mph	30mph	30mph	30mph	30mph
5	Ramp	25mph	25mph	30mph	30mph	35mph	35mph
7	Major Arterial	25mph	30mph	35mph	35mph	35mph	35mph
11	Local	20mph	25mph	30mph	30mph	30mph	30mph
12	Minor Arterial	25mph	25mph	35mph	35mph	35mph	35mph
15	Super Arterial	30mph	35mph	40mph	45mph	45mph	45mph

3.3.2.5 Clip the road network

The output feature class layer still contains road segments extending to other parts of the Bay Area (green lines in the figure) that are not within the SF County limits. Therefore, a spatial intersection between the polygon area (thick black polygon) and the road network is performed. The cropped road network (in brown) contains only those road segments within the SF County boundary limits. The road segments are updated to reflect the revised GIS length. At the end of this step, there remain 26,276 road segments inside the SF County Area.

Figure 3-7: Clipping the SF County road network from the Bay Area road network

3.3.2.6 Merge the road network available by Tod into one dataset

Each TOD's clipped road network gets merged into a singular data source at this stage, and the dataset looks like something below

Table 3-3: merging the ToD records into one record for each road network segment

A	B	peak	GIS_length_mil	CAP	FT	SPEED	DISTANCE	ONEWAY
6985	6980	AM	0.068958149	1400	5	35	0.2193	0
6985	6980	PM	0.068958149	1400	5	35	0.2193	0
6985	6980	EA	0.068958149	1400	5	35	0.2193	0
6985	6980	EV	0.068958149	1400	5	35	0.2193	0
6985	6980	MD	0.068958149	1400	5	35	0.2193	0

3.3.2.7 Creation of unique segment id's and aggregation of the road network attributes

A unique field is created by concatenating two existing columns, "A" and "B." A & B represent the start and end nodes of the road link. The newly created column A_B acts as a unique identifier of the road segments and permits to carry out of mathematical aggregation functions like mean, median, summation, maximum, and minimum on each (desired) attribute. For each attribute desired, the aggregation function is carefully assigned. For example, the Total Volume flowing on a given road segment per day is the sum of all traffic volume observed across each TOD. At the same time, given that variables like facility type, number of lanes, and GIS length are constant, only the first value of the record is kept. Variables like the congested Speed, congested travel time, and free flow travel time depend upon the traffic flowing through the road segment. Therefore, the weighted average method is adopted for their calculations, with the total volume being the predetermined weight parameter.

At the end of the process, two unique road network datasets for 2010 and 2016 emerge. The road network contains precisely one entry (a record) for each road segment within SF County, with attribute values corresponding to the expected average daily conditions.

3.3.2.8 The spatial intersection of the road network with the SF Census Tract

The spatial intersection is performed between the aggregated road network and the SF Census Tract (polygon shapefile) layer. During the intersection process and depending upon the number of census tracts through which a road segment traverses, the road segments get exploded into sub-segments. Each subsegment would always be within the given census tract.

For example, before the spatial intersection is performed, the (red) line 24870_24872 is partly inside Census Tract (CT) 010100, while a tiny fragment is in CT 010500. The spatial intersection splits the road segment into two sub-segments, each independent of the other (see other figures). It allows the research to have better control over the length of road segments in a given census tract. Variables like free flow time, congested time, and GIS length are recalculated for each of these subsegments

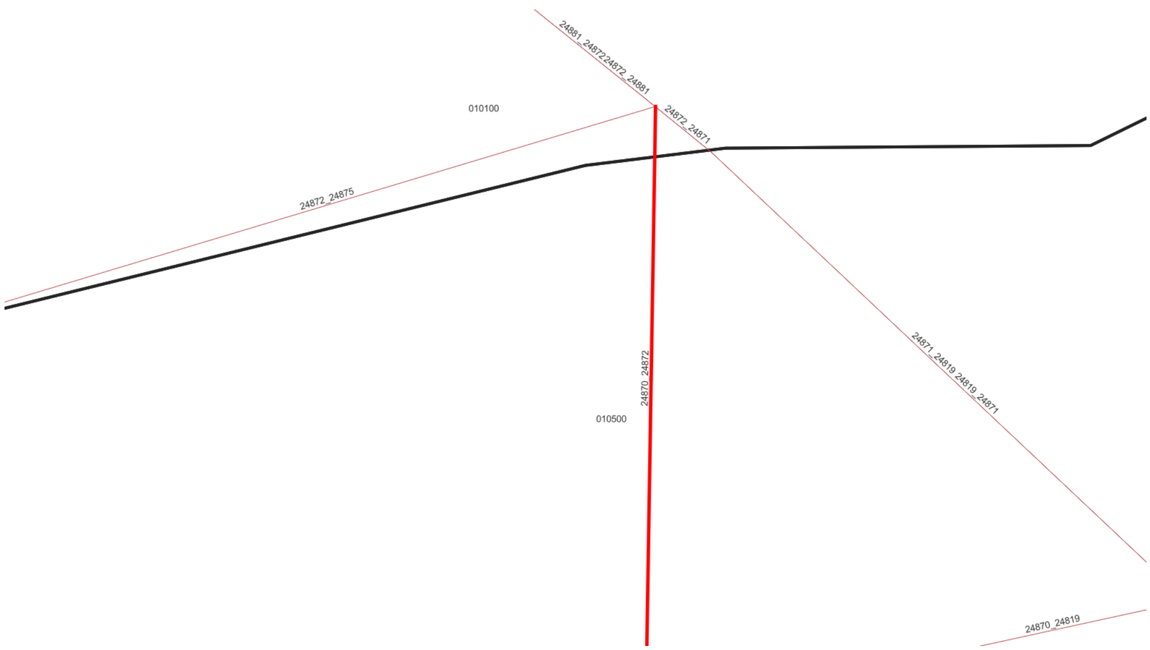

Figure 3-8: Road segment before spatial intersection between road network and census tract polygon shapefile

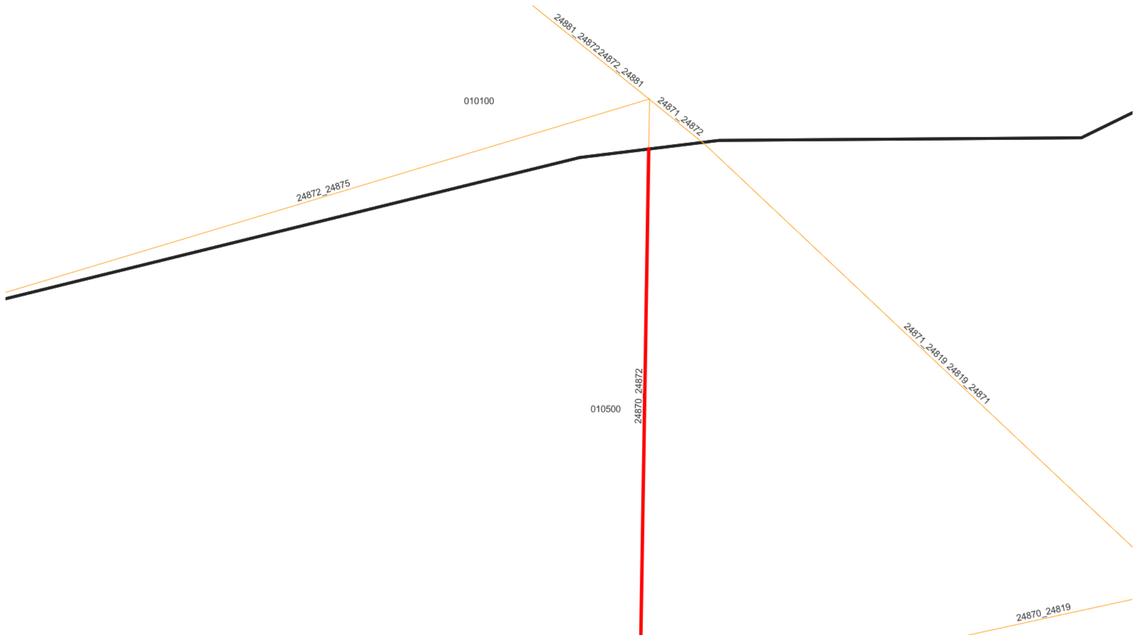

Figure 3-9: Road segment after spatial intersection between road networks

3.4 Methods

This section is divided into two major parts.

Section 3.4.1 details explanatory variables that will be used to conduct a thorough statistical analysis.

Section 3.4.2 provides a detailed procedure adopted to develop the Poisson fixed-effects model, which will be used to test out different explanatory variables identified in Section 3.4.1. The model is developed to derive additional crashes generated due to explanatory variables' involvement.

3.4.1 Selection of explanatory variables

The literature section briefly mentions many factors that can potentially contribute to the likelihood of road crashes. The contributing factors can be broadly classified into road environment, vehicle factors, driver characteristics and behavior, and road design. Ideally, all categories of factors should participate in road crash frequency analysis. However, not all variables can have equal attention, possibly due to the limitation of capturing the data. Therefore, the selection of a variable is very critical.

3.4.1.1 Quantity of Travel

The most intuitive and perhaps critical factor affecting road crash frequency and severity is the quantity of travel. The amount of travel can be hourly volume, annual average daily traffic (AADT), or vehicle miles traveled (VMT). VMT is widely used as a primary indicator of travel by policymakers and transport professionals due to how consistently and comprehensively it is monitored and documented over time by local or federal agencies.

Moreover, studies by (Fridstrom et al., 1995; Jovanis & Chang, 1986; Song et al., 2006) have demonstrated that vehicle miles traveled are an acceptable variable to measure road crash frequency. Similarly, while analyzing the urban transportation system in Honolulu, Hawaii, (Levine et al., 1995) found that more road mileage is associated with more road crashes. While examining the effects of VMT, Tarko found that higher VMT, especially on the urban road, is directly related to the number of crashes (Tarko et al., 1996). The results are consistent with the studies by (Karlaftis & Tarko, 1998; Hadayeghi et al., 2003; Ladrón de Guevara et al., 2004; Hadayeghi et al., 2006; Aguero-Valverde & Jovanis, 2006; M. A. Quddus, 2008) who found VMT to be positively associated with road crashes.

One can use urban sprawl phenomena to explain the relationship between VMT and road crashes (Ewing et al., 2002, 2003). As urban sprawl increases (city limits increase in size), people are forced to travel more frequently and for longer distances resulting in increased time spent on roads, thereby exposing them to more traffic-related crashes. Other factors like population density, job-housing locations, unemployment rate, the natural topology of the city, and availability and efficiency of public transportation also shape VMT. In other words, VMT mainly reflects the exposure measure for traffic volume; therefore, with increased VMT, the probability of getting involved in a crash is likely higher.

Therefore, this study selects VMT as one of the explanatory variables to test the statistical model. To differentiate the exposure of TNC vehicles from the rest of the vehicles on the road network, the study splits the total VMT into two parts a) non-TNC

VMT and b) TNC VMT. As the names indicate, non-TNC VMT corresponds to the VMT generated by all other vehicles except TNC vehicles. TNC VMT is the summation of the vehicle miles traveled by the TNC service-rendering vehicles.

3.4.1.2 Speed

Like VMT, abundant scientific literature links crash occurrence with traffic speed (Solomon, 1964; Elvik, 2001; Hauer, 2009). In simplistic terms, one can assume that the relationship between speed and road crash occurrence is strictly linearly positive. It is because drivers traveling at higher speeds have less time to react to potentially hazardous scenarios, given that the reaction time and braking distance are proportional to increased speed. At very high speeds, the driver may lose control of the vehicle or misjudge the road scene leading to a dangerous situation.

However, past studies have suggested otherwise (Solomon, 1964; Elvik, 2001; Hauer, 2009). Solomon found a U-shaped relationship between the number of crashes per distance and travel speed, i.e., road crashes decrease as the speed increases. After a certain speed, the road crashes started to rise again. Furthermore (Elvik et al., 2006) did an extensive study to explore the statistical association between speed and road safety and found such a relationship casual. According to (Elvik et al., 2006), a 10% reduction in average speed results in a 37.8% reduction in fatalities. Similarly, (M. Quddus, 2013) found that speed variations can positively influence road crash rates.

Alternatively, the widespread use of Intelligent Transportation System (ITS) techniques and location-based technology (LBS) services has made it feasible to measure speeds and travel time more dynamically and thereby use these variables as a proxy to explain traffic congestion (Albalate & Fageda, 2021; Shefer & Rietveld, 1997; Ivan et al., 2000; Lord et al., 2005). Given that the SF CHAMP network models include average speed (congested speed) as one of three congestion-centric measures besides vehicle hours of delay and VMT, this study utilizes congested speed to represent congestion in the statistical model.

(Shefer, 1994) established an inverse U-relationship between road congestion and safety mainly due to reduced speed. The author uses a simple graph, as shown below, to explain the relationship.

Stage I – initially, with few vehicles on the road, the possibility of an individual getting involved in a crash is rare. Therefore, the possibility of a vehicle leading to a fatal crash is rare. However, as the number of vehicles keeps increasing, the number of fatal crashes keeps rising. At the same time, as the number of vehicles keeps increasing, the link's (average) speed starts to decrease but is still more than the operational speed.

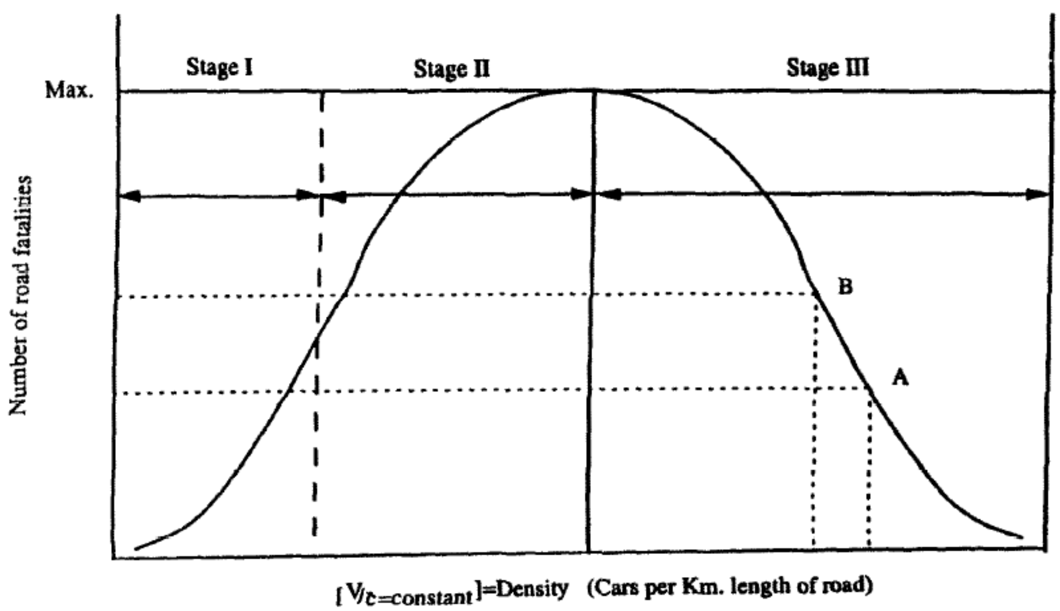

Figure 3-10: Hypothetical relationship between road fatalities and road density proposed (Shefer, 1994)

Stage II - At the beginning of this stage, the vehicular speed tends to decrease, but it is still possible to maintain the speed well above the design limit. The short steep slope line represents the stage. It indicates that the fatalities increase faster than the initial Stage I. However, as more and more vehicles get added, the speed starts to decrease non-linearly, resulting in congestion. The rise in congestion further flattens the slope of the curve and thus reducing the overall rate of fatality.

Stage III – Further, a point arises where the slope of the curve is parallel to the X-axis, starting in which the number of fatalities goes down and first traverses the graph like in Stage II and then Stage I, even when the V/C ratio increases.

Shefer and Reitvield further explored the relationship between congestion and road crashed using average speed, speed variance, and vehicle mix and concluded that congestion has an unwanted gain potentially in the form of lower road fatalities (Shefer & Rietveld, 1997). They found that the fatality rate is lower during peak hours than at non-peak hours.

Therefore, the study also includes congested speed in the list of explanatory variables. Furthermore, to represent a non-linear relationship between road crashes and congestion, the study adds the square of congested speed variable to the list of explanatory variables. Further, the study maintains two variables to acknowledge that congested speed on freeways differs from other city roads in terms of duration and variation.

3.4.1.3 TNC Pick-up and Drop-off (PUDO)

The primary source of the research crash data is the SWITRS database. Like NHTSA, it provides information about crashes that local and governmental agencies reported. However, both databases are restricted to those crash records in which motor vehicles are involved, i.e., where motor vehicles are in transport. Moreover, it does not have an attribute to capture crashes resulting from passengers' loading/unloading (PUDO) behavior.

The taxi industry has long acknowledged the issue somewhat limitedly in the form of *dooring* and its implication on two-wheeler motorcycles and bicycles but has no comprehensive solution. The Australian road safety community discusses its impact on vulnerable road user groups, especially bicyclists w.r.t parked vehicles but fails to acknowledge problems arising from the emerging transportation sector (Bolitho, 2013; Cumming, 2012) like TNC operations.

Compared to personal (vehicle) transportation which starts or ends at a parking place in either a home, office or on-street or off-street parking area, most TNC trips end at the nearest road curb leading to the entry/exit of the premise. This come-and-find curbside parking approach may have unintended road safety consequences, mainly when another road user tries to use the same area simultaneously.

Past research to understand the impact of PUDOs on road safety is rare. (Schimek, 2018) indicates that almost 12-27% of bicycle crashes involve parked vehicles on the curbside. Researchers from the University of Washington studied the behavior of TNC driver's PUDO in the South Lake Union (SLU), Seattle neighborhood, to develop a strategy to improve traffic flow. The SLU neighborhood is the site of the main Amazon campus and generates many TNC trips. The study found that during peak hours, the PUDO activity amounts to 29 - 39% of the total traffic volume, with many vehicles performing PUDO activities in the travel lane (Goodchild et al., 2019). The physical survey finds that more than half of the TNC trips in the neighborhood either start or end outside the dedicated curb space. Such a pattern is prevalent even when the curb space in the passenger loadings zone (PLZ) is empty and even when both Uber/Lyft enforces the geofencing technique, which directs drivers and passengers to designated PUDO locations on the street. The research observed high-volume pedestrian (30-40%) crossings at street locations where no crosswalks exist. While these erratic loading/unloading patterns and high pedestrian crossing volumes did not create traffic safety issues in the survey neighborhood, they may lead to hazardous scenarios in other areas, primarily where multi-modal transport activity occurs or in high-density areas.

More recently, (Kwayu et al., 2022) tried to summarize bicycling safety issues in nine Canadian cities by analyzing self-reported near-miss and collisions event obtained from the crowdsourcing platform BikeMaps.org. Her research found that the probability of cyclists getting involved in crashes increases as motor vehicles encroach a bike lane or on-street parking close to or into the bike lane resulting in dooring and driver violations at the crosswalk.

In a nutshell, no prior research predicts the impact of TNC-related PUDO on road safety outcomes. The current study is the first to attempt to model such behavior.

3.4.1.4 Exposure variable

The notion of exposure has widespread use in road safety crash analysis. Chapman and Carroll (Carroll 1973; Chapman 1973) have extensively studied and reviewed the concept of exposure. However, there is no definitive agreement on what it means. Many researchers attempted to use exposure for their statistical crash analysis without trying to define the concept. However, this study will use the definition proposed by Carroll, which states that "exposure is the frequency of traffic events which create a risk of road crashes." In other words, exposure helps study the potential crash risk and estimate the probability of crash occurrence at a particular site. Knowing such parameters allows for comparing a given location at different periods or two or more sites in the same period reliably or meaningfully. Without the exposure measure, the analysis results may mask the proper relationship between the factors and their relationship with road safety crash frequency.

There is a general disagreement in the scientific literature regarding which exposure is more desirable than the others. Many argue that given that different exposure variables produce different results, the proxy for exposure must be carefully identified and complement the analysis's overall objective (Kam, 2003).

Exposure variables can be classified in two ways a) aggregate or b) disaggregate. The latter is more complicated to obtain and measure. For example, examining individual choices based on individuals characteristics and preferences is much more complex than getting (aggregated) attributes like traffic volume (Miaou, 1994; Mountain et al., 1996; Qin et al., 2004, 2006; Wong et al., 2007; Van den Bossche et al., 2005), conflicts (Bie et al., 2005; Wong et al., 2006), travel distance (Li et al., 2003), travel time (Chipman et al., 1993), and population or fuel consumption (Fridstrøm et al., 1995) are much easier to obtain and widely used.

For road crash analysis like the current study or regional, national, and international comparison, it is adequate to compare the number of road safety crashes against some gross estimate of the amount of travel. Therefore Total VMT is chosen as an exposure variable for future analysis.

3.4.2 The model

3.4.2.1 Expectations from the linear regression model

The model must adhere to certain assumptions no matter the form of multiple linear regression. These assumptions are often termed as Gauss Markov Theorem or Conditions, and these are (Jeffrey Wooldridge, 2019):

- Linearity: the parameters the regression model estimates using the OLS method must be linear.
- Random: the data must have been randomly sampled from the population.
- No Perfect Collinearity: none of the independent variables is constant, and there is no exact linear relationship between them.
- Exogeneity: independent variables are not correlated with the error term u.
- Homoscedasticity: no matter the independent variables' values, the variance error is constant.

It is with this principle that we approach the following section.

3.4.2.2 Regression Structure

Road crashes are discrete or count data variables since they can only take values between $0 \rightarrow n$ where n = 0,1,2,3,4, 5, … ∞

While using count data as a dependent variable in the linear regression model (Ordinary Least Square (OLS) regression) setup is not prohibited, it cannot handle the truncated data at zero and therefore predicts non-integers or negative values. It makes the expected value from OLS regression inconsistent and non-sensical. Thus, a linear regression model is unsuitable for predicting count data variables.

An excellent alternative to OLS regression is to take the natural logarithm (log) of both the predicted and independent variables. The log-linear relationship for predicting crash frequency using explanatory variables presented in section 3.4.1 can be represented as:

$$\log(total\ crash) = \beta_0 + \beta_1 * \ln(Tot\ VMT) + \beta_2 * (CongSpeed\ Freeways) + \beta_3 * (CongSpeed\ Freeways)^2 + \beta_4 * (CongSpeed\ Otherroads) + \beta_5 * (CongSpeed\ Otherroads)^2 + \beta_6 * \ln(PUDO) + \beta_7 * \ln(Tot\ TNC\ VMT) + \varepsilon$$

Equation 3-1

In the equation above, variables on the equation's left-hand side (LHS) and right-hand side (RHS) get logged. As a result, the coefficients of the RHS variable, i.e., β_1,$\beta_2, \ldots, \beta_7$. The proportional change in total crash frequency resulting from proportional change due to Tot VMT, Congested Speed – Freeways, Congested Speed – Other Roads, PUDO, and Tot TNC VMT in a given area. Taking the exponent on both sides of the equation yields the following equation

$$total\ crash = e^{\beta_0} * e^{\beta_1 * \ln(Tot\ VMT)} * e^{\beta_2 * (CongSpeed\ Freeeways)} * e^{\beta_3 * (CongSpeed\ Freeways)^2} * e^{\beta_4 * (CongSpeed\ Otherroads)} * e^{\beta_5 * (CongSpeed\ Otherroads)^2} * e^{\beta_6 * \ln(PUDO)} * e^{\beta_7 * \ln(Tot\ TNC\ VMT)} * e^{\varepsilon}$$

Equation 3-2

Compared to Equation 3-1, Equation 3-2 indicates that the impact of variables on total crash frequency is not linear but rather multiplicative. The interpretation is as follows: 1 unit change in TNC VMT increases the road crash frequency by e^{β_7}, ceteris paribus.

If β_7 is negative, the multiplicative impact is less than one, and therefore the road crashes increased by a factor of less than one;

If β_7 is positive, then Tot TNC VMT will contribute towards increasing road crashes in a given road category per census tract by a factor more significant than one.

The log-linear OLS regression, as in Equation 3-1, gives us an estimate of $E(\log(total\ crash))$ and whereas the primary question of interest is $\log\big(E(total\ crash)\big)$. An alternative to OLS regression is to use count regression models, which can predict $\log\big(E(totalcrash)\big)$.

For example, the simplest is the Poisson Regression model, which can handle such skewed data. Unlike OLS, the Poisson model can easily handle heteroscedasticity and preserve the count nature observed in the datasets as observed in road crashes.

Alternatively, past literature suggests that in scenarios where the predicted values have a more significant frequency of zeros, zero-inflated models like Zero-inflated Poisson (ZIP) or negative binomial model (ZINB), or Hurdle model can be utilized.

However, both ZIP and ZINB assume that the dependent variable is a mixture of two individual groups (Hu et al., 2011).

- One group whose counts could be explained by Poisson or Negative Binomial regression Model and
- another group with zero probability of a count greater than zero. Typically logistic regression model is used to predict which group an individual belongs

The literature does not indicate when to use or not use such zero-inflated models. Even if utilized and the model fits much better than traditionally counted data models, the outputs are difficult to estimate and interpret (Allison, 2012).

Similarly, Hurdle Model divides the data into a two-part decision-making process (Hu et al., 2011):

- explaining whether the count is zero or positive
- determining whether the count is positive.

To model condition one, we can use Bernoulli probability to govern the binary outcome, i.e., whether the count variate has a zero or positive realization. Condition two triggers only when the count data is positive, i.e., when the hurdle is crossed (hence the name), and is governed by conditional distribution for a truncated-at-zero count data model. However, such an assumption is too strong given that there would always be some road segments that would always have zero crashes, no matter what happens to other covariates, and therefore its usage can be avoided.

Additionally, and much more importantly, no packages exist to incorporate fixed effect extension to zero-inflated models or hurdle models for panel data (Rauli Susmel, 2022).

Therefore the study adopts the Poisson Regression model to predict that the primary question of interest is $\log\big(E(total\ crash)\big)$.

In the Poisson Regression model, the relationship between total crash and explanatory variables is represented as:

$$E(total\ crash|x) = e^{\beta_0 + \beta x}$$

Equation 3-3

where

x = vector of the explanatory variables

β = co-efficient of the respective explanatory variable

and total crash (Y_i) ~ Poisson (λ)

where λ = lambda, which is the parameter that controls the Poisson distribution for value λ=1,2,3,...

It is also the parameter that imposes that the mean value of the data equals the variance value (often termed *equidispersion*). However, in real-life datasets, variance is always more significant than the mean value (*overdispersion*). Overdispersion could occur due to various factors, including unobserved heterogeneity and the influence of other variables, which leads to dependence of the probability of an event on previous events, the presence of outliers, and the existence of excess zeros on the response variable. However, Simon et al. demonstrated that when computed with a robust covariance matrix (also termed quasi-Poisson regression), Poisson regression could overcome the overdispersion issue (Berrebi et al., 2021). The author suggests that such a process returns the same coefficients for the explanatory variables as the regular Poisson regression, but standard errors are much better calibrated for over-dispersion (Berrebi et al., 2021; Wooldridge, 2002).

Taking the derivative of $E(total\ crash|x)$ w.r.t $Tot\ TNC\ VMT$ results in the partial effects, also termed as *marginal impact.* The derivative is the expected number of road crashes due to adding one additional mile of TNC VMT to the base year, i.e., Y.R. 2010.

$$\frac{\delta E(total\ crash\ |x)}{\delta Tot\ TNC\ VMT} = \frac{e^{\beta_0} * e^{\beta_1 * \ln(Tot\ VMT)} * e^{\beta_2 * (CongSpeed\ Freeeways)} * e^{\beta_3 * (CongSpeed\ Freeways)^2} * e^{\beta_4 * (CongSpeed\ Otherroads)} * e^{\beta_5 * (CongSpeed\ Otherroads)^2} * e^{\beta_6 * \ln(PUDO)} * e^{\beta_7 * \ln(Tot\ TNC\ VMT)}}{\delta Tot\ TNC\ VMT}$$

Equation 3-4

$$\frac{\delta E(total\ crash\ |x)}{\delta Tot\ TNC\ VMT} = \frac{e^{\beta_0} * e^{\ln(Tot\ VMT)^{\beta_1}} * e^{\beta_2 * (CongSpeed\ Freeeways)} * e^{\beta_3 * (CongSpeed\ Freeways)^2} * e^{\beta_4 * (CongSpeed\ Otherroads)} * e^{\beta_5 * (CongSpeed\ Otherroads)^2} * e^{\ln(PUDO)^{\beta_6}} * e^{\ln(Tot\ TNC\ VMT)^{\beta_7}}}{\delta Tot\ TNC\ VMT}$$

Equation 3-5

3.4.2.3 Introducing fixed effects and time effects

The study introduces a dummy variable “year_2016_dummy” (0 for the year 2010 or 1 for the year 2016) to capture linear time trends during the term, while α_i represents the individual-specific effects (fixed effects).

Equation 3-3 can be re-written as

$$E\left(total\ crashes_{i,t} \middle| x_{i,t}\right) = e^{\beta x + \alpha_i + year_{2016_{dummy}}{}_t}$$

Equation 3-6

And more specifically,

$$\begin{aligned} E\left(total\ crashes_{i,t} \middle| x_{i,t}\right) \\ &= \beta_0 \ln(Tot\ VMT)_{(i,t)} + \beta_1 CongSpeed\ Freeway_{(i,t)} + \beta_2 CongSpeed\ Freeway^2{}_{(i,t)} \\ &+ \beta_3 CongSpeed\ Otherroads_{(i,t)} + \beta_4 CongSpeed\ Otherroads^2{}_{(i,t)} \\ &+ \beta_5 \ln(PUDO)_{(i,t)} + \beta_6 \ln(Tot\ TNC\ VMT)_{(i,t)} + \beta_7 Year_{2016_{dummy}}{}_{(i,t)} + \alpha_i + \varepsilon_{(i,t)} \end{aligned}$$

Equation 3-7

where

i = identity (i.e., road category per census tract)

t = observation year (i.e., either YR 2010 or YR 2016)

β_1 to β_7 = co-efficient of explanatory variables x participating in the regression

α_i = individual-specific effects (i.e., fixed effects)

$\varepsilon_{(i,t)}$ = individual identities error term

The β parameters use maximum likelihood estimations (MLE) and represent the outcomes at the individual road category for a given census tract per year (Hausman et al., 1984).

The resulting conditional likelihood is proportional to the right-hand side of Eq. (9.3), as mentioned by Cameron and Trivedi in their book (Cameron & Trivedi, 1998) and is restated here

$$\hat{\beta} = \left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(X_{(i,t)} - \overline{X}_i\right)\left(X_{(i,t)} - \overline{X}_i\right)'\right]^{-1} \sum_{i=1}^{n}\sum_{t=1}^{T}\left(X_{(i,t)} - \overline{X}_i\right)\left(y_{(i,t)} - \overline{y}_i\right)$$

where

$\overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{(i,t)}$ and $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{(i,t)}$

are individual-specific averages over time.

The individual-specific fixed effects can be estimated by $\hat{\alpha}_i = \overline{y}_i - \overline{X}_i{}'\hat{\beta}$.

For a short panel,

n →∞ and T is fixed,

$\hat{\beta}_{FE}$ is consistent for β, while $\hat{\alpha}_i$ is not consistent for α_i as only T observations are used in estimating each α_i.

Estimates are obtained by setting the partial derivative of the log-likelihood with respect to beta to zero, as in Eq. (8).

$$\sum_{i=1}^{n}\sum_{t=YR\ 2010}^{YR\ 2016} x_{(i,t)}\left(total\ crashes_{(i,t)} - \sum_{s=1}^{T}(total\ crashes_{(i,s)}\frac{e^{\beta x_{(i,t)}}}{\sum_{s=1}^{T} e^{\beta x_{(i,t)}}}\right)$$

At this point, the constant added to the observed explanatory variable $x_{(i,t)}$ gets canceled out in both equations. It allows the study to capture the variation *within* the variable in time for each road category per census tract. The equation is free from heterogeneity and endogeneity effects and can be ignored over a five-to-six-year panel.

3.4.2.4 Isolating the effects of each explanatory variable on total crashes

Taking exponential on both sides,

$$E(total\ crashes_{i,t}|x_{i,t}) = e^{\beta_1 \ln(Tot\ VMT)_{(i,t)}} + e^{\beta_2 CongSpeed\ Freeway_{(i,t)}} + e^{\beta_3 CongSpeed\ Freeway^2{}_{(i,t)}} + e^{\beta_4 CongSpeed\ Otherroads_{(i,t)}} + e^{\beta_5 CongSpeed\ Otherroads^2{}_{(i,t)}} + e^{\beta_6 \ln(PUDO)_{(i,t)}} + e^{\beta_7 \ln(Tot\ TNC\ VMT)_{(i,t)}} + e^{\beta_8 Year_{2016} dummy_{(i,t)}} + e^{\alpha_i}$$

Equation 3-8

$$E(total\ crashes_{i,t}|x_{i,t})$$
$$= e^{\beta_1 \ln(Tot\ VMT)_{(i,t)}} * e^{\beta_2 CongSpeed\ Freeway_{(i,t)}} * e^{\beta_3 CongSpeed\ Freeway^2{}_{(i,t)}}$$
$$* e^{\beta_4 CongSpeed\ Otherroads_{(i,t)}} * e^{\beta_5 CongSpeed\ Otherroads^2{}_{(i,t)}} * e^{\beta_6 \ln(PUDO)_{(i,t)}}$$
$$* e^{\beta_7 \ln(Tot\ TNC\ VMT)_{(i,t)}} * e^{\beta_8 Year_{2016} dummy_{(i,t)}} * e^{\alpha_i}$$

Equation 3-9

As per the laws of exponential

$e^{\ln(x)}$ = x, and therefore we can translate the above equation to

$$total\ crashes_{(i,t)}$$
$$= Tot\ VMT_{(i,t)}{}^{\beta_1} * e^{\beta_2 CongSpeed\ Freeway_{(i,t)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,t)}}$$
$$* e^{\beta_4 CongSpeed\ Otherroads_{(i,t)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,t)}} * PUDO_{(i,t)}{}^{\beta_6}$$
$$* Tot\ TNC\ VMT_{(i,t)}{}^{\beta_7} * e^{\beta_8 Year_{2016} dummy_{(i,t)}} * e^{\alpha_i}$$

Equation 3-10

We can re-write LHS in Equation 3-10 as $E\left[total\ crash_{(i,t)}|x_{(i,t)}\right] = total\ crash_{(i,t)}$

Therefore, $E[total\ crashes]$ for each of the years, i.e., YR 2010 and YR 2016, could be expressed as

$$E\left[total\ crashes_{(i,2010)}|x_{(i,2010)}\right]$$
$$= Tot\ VMT_{(i,2010)}{}^{\beta_1} * e^{\beta_2 CongSpeed\ Freeway_{(i,2010)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}$$
$$* e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}} * PUDO_{(i,2010)}{}^{\beta_6}$$
$$* Tot\ TNC\ VMT_{(i,2010)}{}^{\beta_7} * e^{\beta_8 Year_{2016} dummy_{(i,2010)}} * e^{\alpha_i}$$

Equation 3-11

$$E\left[total\ crashes_{(i,2016)} \middle| x_{(i,2016)}\right]$$
$$= Tot\ VMT_{(i,2016)}{}^{\beta_1} * e^{\beta_2 CongSpeed\ Freeway_{(i,2016)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,2016)}} * e^{\beta_4 CongSpeed\ Otherroads_{(i,2016)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,2016)}} * PUDO_{(i,2016)}{}^{\beta_6} * Tot\ TNC\ VMT_{(i,2016)}{}^{\beta_7} * e^{\beta_8 Year_{2016 dummy}{}_{(i,2016)}} * e^{\alpha_i}$$

Equation 3-12

Taking the ratio year 2016 and year 2010

$$\frac{E\left[total\ crashes_{(i,2016)} \middle| x_{(i,2016)}\right]}{E\left[total\ crashes_{(i,2010)} \middle| x_{(i,2010)}\right]} = \frac{Tot\ VMT_{(i,2016)}{}^{\beta_1}}{Tot\ VMT_{(i,2010)}{}^{\beta_1}} * \frac{e^{\beta_2 CongSpeed\ Freeway_{(i,2016)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,2016)}}}{e^{\beta_2 CongSpeed\ Freeway_{(i,2010)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}} * \frac{e^{\beta_4 CongSpeed\ Otherroads_{(i,2016)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,2016)}}}{e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}}} * \frac{PUDO_{(i,2016)}{}^{\beta_6}}{PUDO_{(i,2010)}{}^{\beta_6}} * \frac{Tot\ TNC\ VMT_{(i,2016)}{}^{\beta_7}}{Tot\ TNC\ VMT_{(i,2010)}{}^{\beta_7}} * \frac{e^{\beta_8 Year_{2016 dummy}{}_{(i,2016)}}}{e^{\beta_8 Year_{2016 dummy}{}_{(i,2010)}}} * \frac{e^{\alpha_i}}{e^{\alpha_i}}$$

Equation 3-13

$$\frac{E\left[total\ crashes_{(i,2016)} \middle| x_{(i,2016)}\right]}{E\left[total\ crashes_{(i,2010)} \middle| x_{(i,2010)}\right]} = \left(\frac{Tot\ VMT_{(i,2016)}}{Tot\ VMT_{(i,2010)}}\right)^{\beta_1} * \left(\frac{e^{\beta_2 CongSpeed\ Freeway_{(i,2016)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,2016)}}}{e^{\beta_2 CongSpeed\ Freeway_{(i,2010)} + \beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}}\right) * \left(\frac{e^{\beta_4 CongSpeed\ Otherroads_{(i,2016)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,2016)}}}{e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)} + \beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}}}\right) * \left(\frac{PUDO_{(i,2016)}}{PUDO_{(i,2010)}}\right)^{\beta_6} * \left(\frac{Tot\ TNC\ VMT_{(i,2016)}}{Tot\ TNC\ VMT_{(i,2010)}}\right)^{\beta_7} * \left(\frac{e^{\beta_8 Year_{2016 dummy}{}_{(i,2016)}}}{e^{\beta_8 Year_{2016 dummy}{}_{(i,2010)}}}\right)$$

Equation 3-14

Change in total crashes going from 2010 → 2016 can also be written as

$$\Delta\ Change\ in\ total\ crashes = \frac{total\ crashes_{2016} - total\ crashes_{2010}}{total\ crashes_{2010}}$$

Equation 3-15

Therefore, the introduction of one additional mile of $Tot\ TNC\ VMT$ equals $\Delta\ total\ crashes$ w.r.t YR 2010

$$E\ \left[\frac{\Delta total\ crashes}{\Delta Tot\ TNC\ VMT}\right] = total\ crashes_{(i,2010)} \left[\left(\frac{Tot\ VMT_{(i,2010)}}{Tot\ VMT_{(i,2010)}}\right)^{\beta_1} * \left(\frac{e^{\beta_2 CongSpeed\ Freeway_{(i,2010)}+\beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}}{e^{\beta_2 CongSpeed\ Freeway_{(i,2010)}+\beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}}\right) * \left(\frac{e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)}+\beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}}}{e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)}+\beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}}}\right) * \left(\frac{PUDO_{(i,2010)}}{PUDO_{(i,2010)}}\right)^{\beta_6} * \left(\frac{Tot\ TNC\ VMT_{(i,2010)}+1}{Tot\ TNC\ VMT_{(i,2010)}}\right)^{\beta_7} * \left(\frac{e^{\beta_8 Year_{2016} dummy_{(i,2010)}}}{e^{\beta_8 Year_{2016} dummy_{(i,2010)}}}\right)\right] - total\ crashes_{(i,2010)}$$

Equation 3-16

3.4.2.5 Advantages of the model

According to Cameron and Trivedi, "*Poisson regression is a particular type of nonlinear regression that respects the discreteness of the count variable*. (Cameron & Trivedi, 1998)" Therefore, the assumptions mentioned in section 3.4.2.1 still hold.

According to Woolridge, the fixed effects Poisson (FEP) estimator has strong robustness for estimating the parameters in the conditional mean. The FEP estimator is consistent for β_0 under the conditional mean assumption only and is entirely unrestricted of whether the variables are over-dispersed or under-dispersed, which is one of the main assumptions for the Poisson model (Wooldridge, 2002).

Uniqueness holds under general identification assumptions. Even when the conditional mean is exponential, the model structure allows dropping coefficients on time-constant explanatory variables, just as in the linear case. Interpreting in our case, including fixed effect transformation for the Poisson model, will help eliminate all variations happening at the SF County level, which at the geographic level do not change much in a five to six years timeframe. As a result, we are left with only traffic attributes that change every year, potentially reducing any biases emerging from the belief that TNCs services may correlate with something important omitted from the model.

According to Wooldridge, FEP allows any serial correlation, which may be the case for the TNC VMT and TNC PUDO variables which are components of TNC as a service.

Guimarães shows that the conditional fixed effects Negative Binomial model do not necessarily remove the individual fixed effects in count panel data; and therefore is not an accurate fixed-effects model (Guimarães, 2008).

3.4.2.6 Interpretation of the results

All the variables Tot VMT, CongSpeed Freeway, CongSpeed Otherroads, PUDO, and Tot TNC VMT on RHS of the equation are time-varying. Given that Tot VMT, PUDO, and Tot TNC VMT are a natural log, their β coefficients can be interpreted as elasticity to the change observed from 2010 to 2016.

On the other hand, because CongSpeed Freeway and CongSpeed Other roads are not logged transformations, their co-efficient can be interpreted as one unit change in the variable value leading to e^{β} times change in road crashes while holding other variables constant.

3.5 Summary

Like past research, the study explores the association between TNC service operations and road safety outcomes in a before-and-after layout; the "before" condition represents a scenario when TNC activities were non-existent, while the "after" represents a condition when they were not.

We do this using the case of San Francisco (SF) County, a county covering 47 square miles of land area at the end of the San Francisco Peninsula and represents the densest residential and commercial location in the Bay Area, with an estimated population of 815,201 as of 2021 (US Census Bureau, 2022). The study chose the SF area for one main reason

a) because the county is the first city where Uber rendered the inaugural TNC service trip in May 2010; subsequently, Lyft rolled out its service in 2012. Together they expanded and now control the dominant share of TNC trips in the SF Area. According to

Erhardt et al., which has studied the SF TNC profile in detail using the real-world dataset, between 2010 and 2016, almost half of the SF VMT increase is attributable to TNCs (G. Erhardt et al., 2019; Roy et al., 2020).

Therefore, 2010 represents the SF scenario when TNC activities were negligible in numbers (assumed to be zero), while 2016 represents a condition when they were not.

There are two main sections in the Chapter. The data section provides details of the data sources used to construct the required dataset, while the Methods present the statistical analytical framework intended to complete the analysis.

The Data sections provide details of the sources used to compile an annual estimation file. The study utilizes three data sources:

- Traffic flow estimates, which produce traffic volumes on all roads and vehicular speed, are made available by San Francisco's travel demand model, the SF-CHAMP model, for both the 2010 and 2016 periods. The estimates reflect land use and land base changes observed in the SF area during the respective years. These are valid estimates because San Francisco County Transportation Authority (SFCTA) continues to use them to analyze policy and infrastructure changes (Brisson et al., 2012; Castiglione et al., 2006).
- TNC trip data compliments the traffic flow estimates and showcases the local TNC usage profile estimated for SF county. It is generated after scrapping Uber and Lyft servers every five seconds via their application programming interface (APIs) between mid-November to mid-December of 2016, excluding dates around the Thanksgiving 2016 holiday period. The scrapped data is further processed to identify pick-up and drop-off (PUDO) locations and to define a TNC trip (Cooper et al., 2018). Erhardt et al. 2019 further enhanced the data and identified the total number of TNC volumes and pick-ups and drop-offs (PUDO) occurring on each road segment in San Francisco county by the time of day (TOD) (G. Erhardt et al., 2019).

- The same SF-CHAMP methodology also provided a Counterfactual (CF) 2016 scenario in which TNC services do not exist. The counterfactual scenario is used to test and validate the impact of TNC services on road safety outcomes
- Road crash data for 2010 and 2016 is obtained from Statewide Integrated Traffic Records System (SWITRS) website. The crash data are re-examined to identify missing information, especially latitudes and longitudes. For records missing such information, geographic coordinates are retrieved using the address geocoding process.
- The KABCO injury style classification nomenclature is conditionally modified and transformed to total crashes, fatal and injury crashes, pedestrian and bicyclist crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes, which act as our primary dependent variables.

The method section lists potential explanatory variables used to perform statistical analysis. It also provides a detailed overview of the developed model to be used for road crash prediction

A few of the variables identified are:

- Vehicle miles traveled (VMT) – a tested and acceptable variable to measure road crash frequency (Fridstrom et al., 1995; Jovanis & Chang, 1986; Song et al., 2006). Past studies find VMT to be positively associated with more road crashes (Levine et al., 1995), with the rate higher for urban roads (Karlaftis & Tarko, 1998; Hadayeghi et al., 2003; Ladrón de Guevara et al., 2004; Hadayeghi et al., 2006; Aguero-Valverde & Jovanis, 2006; M. A. Quddus, 2008; Tarko et al., 1996). It is because VMT is directly related to urban sprawl. The larger the city, the farther the citizens travel (VMT), increasing road crash exposure. The study uses Tot VMT as an exposure variable instead of an independent variable that participates in the statistical model. It ensures that Tot VMT and dependent variables are linked.
- Another independent variable is Congested Speed, which is positively associated with road crashes (Solomon, 1964; Elvik, 2001; Hauer, 2009).

However, the relationship is an inverse U-relationship (non-linear) (Shefer, 1994). To complicate further, Shefer and Reitvield's separate study also concluded that congestion has an unwanted gain, potentially in lower road fatalities (Shefer & Rietveld, 1997). According to their study, the fatality rate is lower during peak hours than at non-peak hours for freeways and urban roads.

- TNC PUDO is also included in the list of explanatory variables because the activity, by its nature, is hazardous. Past research to understand the impact of PUDOs on road safety is rare. (Schimek, 2018) indicates that almost 12-27% of bicycle crashes involve parked vehicles on the curbside, along with the detailed procedure adopted to develop the statistical model. Researchers from the University of Washington who studied the TNC driver's PUDO activity in the South Lake Union (SLU), Seattle neighborhood, found that during peak hours, the PUDO activity amounts to 29 - 39% of the total traffic volume, with many vehicles performing PUDO activities in the travel lane (Goodchild et al., 2019)
- Tot TNC VMT, derived by subtracting non-TNC VMT from Tot VMT, is also included in the list of explanatory variables. It allows the study to measure the possible heterogeneous effects of TNC services on explanatory variables

Later, the model sub-section starts with why the linear regression model and its natural logarithm (log) of both the predicted and independent variables are unsuitable for predicting explanatory variables (count data). As an alternative, it presents a Poisson Regression with a fixed effect model framework with robust standard errors to overcome issues of overdispersion, any concerned heterogeneity, and endogeneity effects such that it can be used to conduct the “before-and-after” assessment proposed.

CHAPTER 4 SPATIAL ANALYSIS AND VISUALIZATION

4.1 Background

The following chapter conducts the spatial analysis of the data used in the research. The first section will focus on road crashes and the road network individually, while the second part will explore combined trends. Wherever possible, data is presented both in tabular and geographical map format.

4.2 Road Crashes

Motor vehicle crash incidents get recorded in a 2D space and expressed in geographical terms (longitude, latitude) or cartographic coordinates (East, North) or the local plan (x,y). Geographic Information System (GIS) software suites like ArcGIS and QGIS facilitate storing and processing georeferenced data. GIS aids in representing and managing many attribute information in the road crash database. The study uses GIS to conduct preliminary spatial analysis to identify areas with a significant concentration of road crashes.

4.2.1 Countywide fatal and injury trends

Table 4-1 below presents the statistics for the primary dependent variables: total crashes, fatal and injury crashes, pedestrian and bicyclist crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes maintained by the California Highway Patrol (CHP) through the Statewide Integrated Traffic Records Systems (SWITRS) for both 2010 and 2016. The absolute difference and percentage difference between these years is also presented.

Table 4-1: Countywide road crashes by severity type

	Total crashes	**Fatal and injury crashes**	**Pedestrian and bicyclist crashes**	**Alcohol-involving (DUI) crashes**	**Property-damage only (PDO)**
Year 2010	5870	3609	1431	696	2234
Year 2016	6831	3888	1474	672	2909
Absolute diff	961	279	43	-24	675
% Diff	16.4%	7.7%	3.0%	-3.4%	30.2%

The key takeaway from the stats shown in Table 4-1 is that the jump in total crashes is due to a sharp rise in two crash types: fatal and injury crashes (approx. 8%) and property damage-only crashes, which have gone up by 30%.

These statistics may not match the annual report released by San Francisco Municipal Transportation Agency (SFMTA), which does not report crashes on freeways (excluding Van Ness Avenue, Presidio) and PDO crash reports crash. The overall collision rate for the city, however, may have changed due to increasing population and economic activities.

As indicated in Chapter 3, not every crash gets geocoded in SWITRS. The study matched the data with those maintained by the University of California Berkeley (*Transportation Injury Mapping System (TIMS)*, 2022), which applies SafeTREC geocoding methodology to SWITRS data statewide. Almost all other crashes (99%) have spatial coordinates specified except for the PDO crashes.

In order to get the locations of the crashes involving PDOs, the study undertakes its geocoding process as described in the data processing section 3.3.1.5.

Figure 4-2 below summarizes the crashes after undertaking all the procedural steps mentioned in the data processing section 3.3.2. It includes the geocoding process where the resultant crashes were tagged to the road segments using the nearest neighborhood principle. To maintain the road crashes attached to the most immediate road segment, any road crash with a distance to the nearest line (D2NL) attribute greater or equal to 10 meters is dropped from the dataset.

Table 4-2: Countywide road crashes by type after undertaking data processing

	Total crashes	**Fatal and injury crashes**	**Pedestrian and bicyclist crashes**	**Alcohol-involving (DUI) crashes**	**Property-damage only (PDO)**
Year 2010	4430	2961	1296	502	1469
Year 2016	5316	3288	1352	537	2028
Absolute diff	886	327	56	35	559
% Diff	20.00%	11.04%	4.32%	6.97%	38.05%

While these numbers are substantially less than the observed crashes, adopting such filtration steps ensures that collisions are not falsely tagged to a road segment and, therefore, does not bias the likely relationship between road crash frequency and TNC variables.

The study further splits the road crashes by facility types shown in Table 4-3. 75% of the crashes occur on Arterials and Locals roads except for pedestrian and bicyclist crash types, in which collector roads take the dubious second position.

The crash percentages are consistent with overall road crashes, which suggest that crash deaths in urban areas are more likely to occur on arterial (58%), local (12%), collector (10%), and freeways (20%) (IIHS, 2020).

Table 4-3: road crashes by facility types

Year	**Facility type**	**Total crash**	**% of total**	**Fatal and injury crash**	**% of total**	**Pedestrian and bicyclist crash**	**% of total**	**Alcohol-involving (DUI) crash**	**% of total**	**Property-damage only (PDO)**	**% of total**
2010	Freeways	448	10.2%	245	8.3%	7	0.5%	68	13.5%	243	16.5%
	Arterials	2153	49.0%	1593	53.8%	713	55.0%	219	43.6%	560	38.1%
	Collectors	807	18.4%	559	18.9%	325	25.1%	90	17.9%	248	16.9%
	Locals	982	22.4%	564	19.0%	251	19.4%	125	24.9%	418	28.5%
2016	Freeways	925	17.4%	370	11.3%	20	1.5%	90	16.8%	555	27.4%
	Arterials	2355	44.3%	1684	51.2%	757	56.0%	250	46.6%	671	33.1%
	Collectors	865	16.3%	558	17.0%	300	22.2%	83	15.5%	307	15.1%
	Locals	1171	22.0%	676	20.6%	275	20.3%	114	21.2%	495	24.4%

Table 4-4 below shows the crash split by facility type. Arterials witness the highest impacts in absolute numbers, but the percentage rise is somewhat nominal, with the sharpest increase found in PDO crash types (+20%). In comparison, collectors and local facility types witness less than half of such numbers, with locals seeing a 20% jump in total crash occurrences and PDO crashes.

Table 4-4: Crashes by facility types (2010 vs. actual 2016)

		Total Crashes	Fatal And Injury Crashes	Pedestrian And Bicyclist Crashes	Alcohol-Involving (DUI) Crashes	Property-Damage Only (PDO)
FREEWAYS	2010	448	245	7	68	243
	2016	925	370	20	90	555
	abs diff	477	125	13	22	312
	% change	106.5%	51.0%	185.7%	32.4%	128.4%
ARTERIALS	2010	2153	1593	713	219	560
	2016	2355	1684	757	250	671
	abs diff	202	91	44	31	111
	% change	9.38%	5.71%	6.17%	14.16%	19.82%
COLLECTORS	2010	807	559	325	90	248
	2016	865	558	300	83	307
	abs diff	58	-1	-25	-7	59
	% change	7.2%	-0.2%	-7.7%	-7.8%	23.8%
LOCALS	2010	982	564	251	125	418
	2016	1171	676	275	114	495
	abs diff	189	112	24	-11	77
	% change	19.25%	19.86%	9.56%	-8.80%	18.42%

4.2.2 Fatal and non-Fatal Injury Collision trends

In San Francisco, vulnerable road users - pedestrians, cyclists, and motorcyclists constituted 75% of all the fatal victims. As shown in Figure 4-1, in 2010, out of 29 people fatally killed, 16 were walking, two were bicycling, and five were motorcyclists. For 2016 these numbers were 18,4 and 3, respectively, of the total 34 fatal crashes mapped. The trend indicates a rising trend in the involvement of non-motorized road users in road crashes.

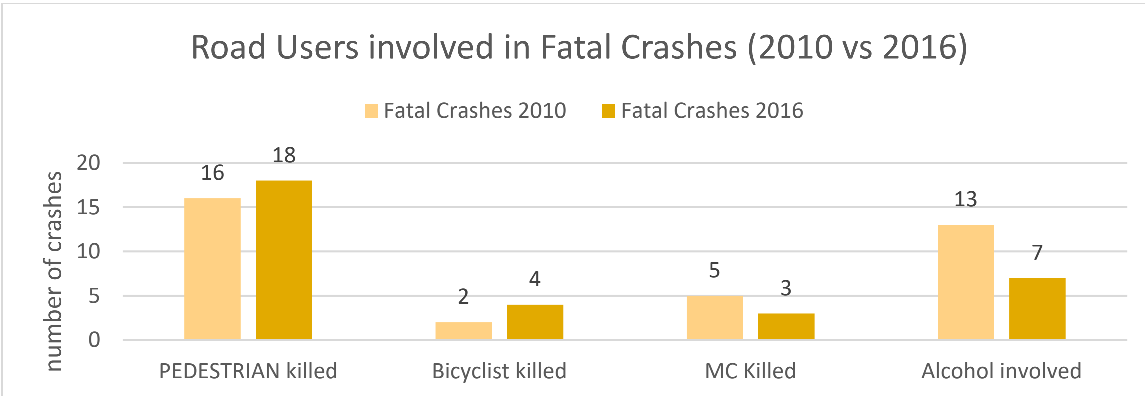

Figure 4-1: road users involved in fatal crashes (2010 vs. 2016)

Likewise, Figure 4-2 provides collision trends for non-fatal injuries (excluding PDOs). Such trends are a much more reliable indicator of long-term city road crash trends. In the, suggest a ~9% increase in such instances. However, these percentage jumps and drops are well within the observed fluctuations when compared to the trend of the past five-year non-fatal injury collision totals. It suggests that non-fatal injuries long-term average remain unchanged between 2010 and 2016.

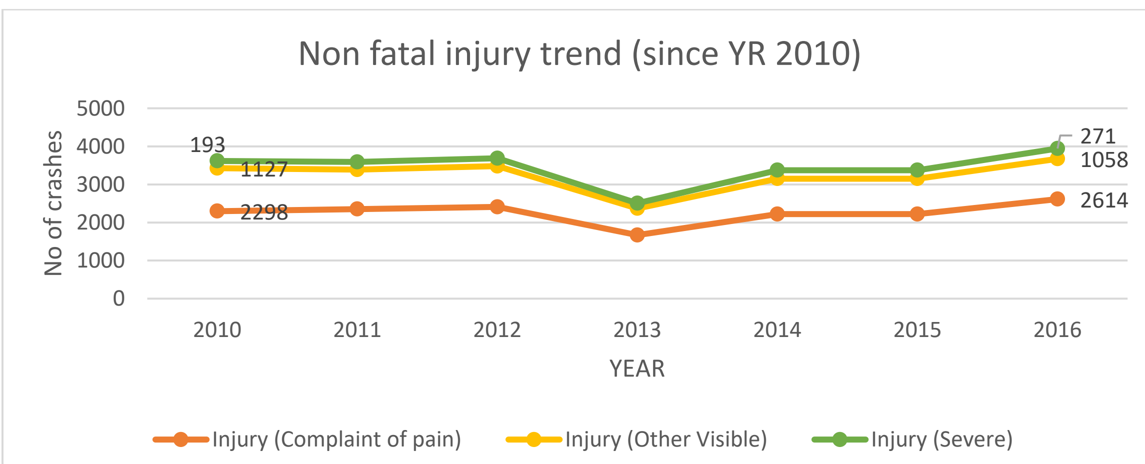

Figure 4-2: non-fatal injury trends since the year 2010

The Vision Zero Annual Report of 2017 claims that 70% of San Francisco's severe and fatal traffic injuries occur on 12% of its streets (SFMTA, 2017). The snapshot of the high-injury network map is presented in Figure 4-3.

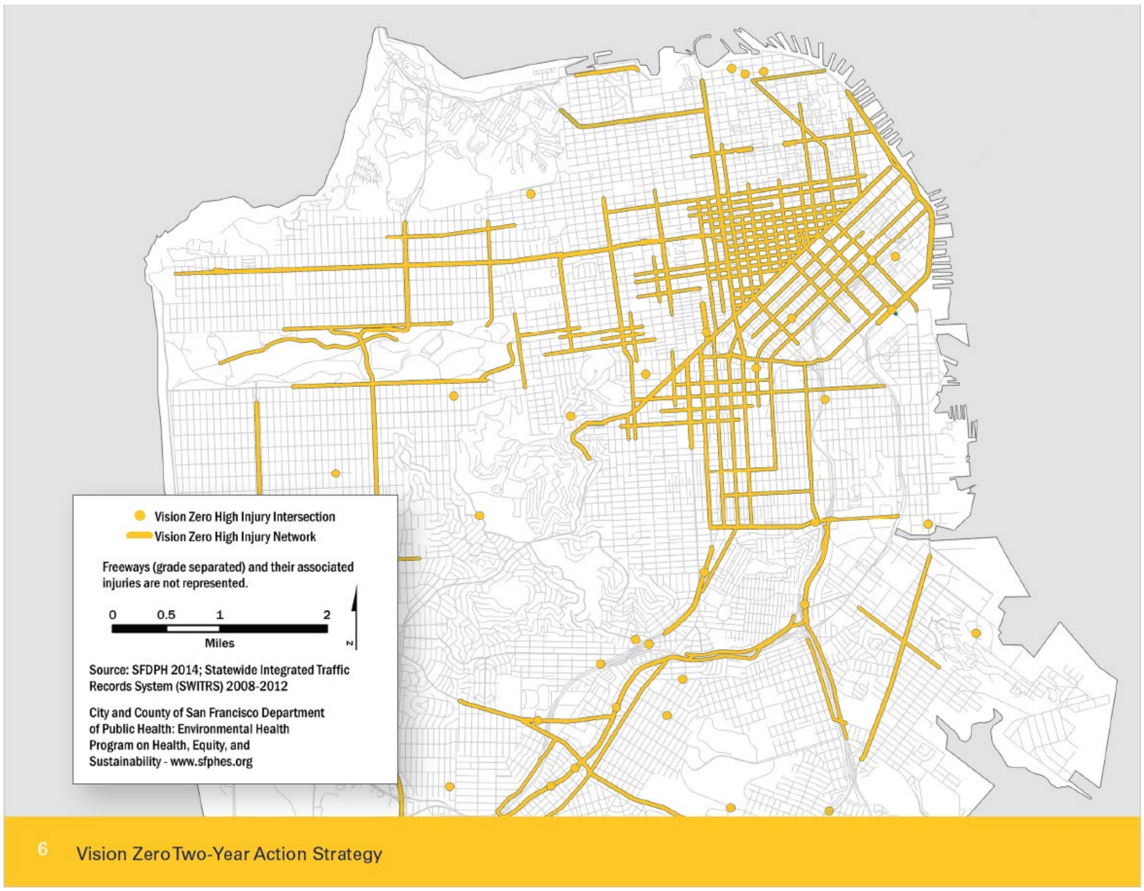

Figure 4-3: SF Vision Zero 2017 - High Injury Network (SFMTA, 2017)

4.2.3 Collision Types and causes

Figure 4-4 and Figure 4-5 below shows injury collision totals by primary collision type for 2010 and 2016. Three of the most common collisions are rear end, broadsides, and sideswipe. Sideswipe appears to increase from 20% to 25% between 2010 to 2016, while others remain stagnant.

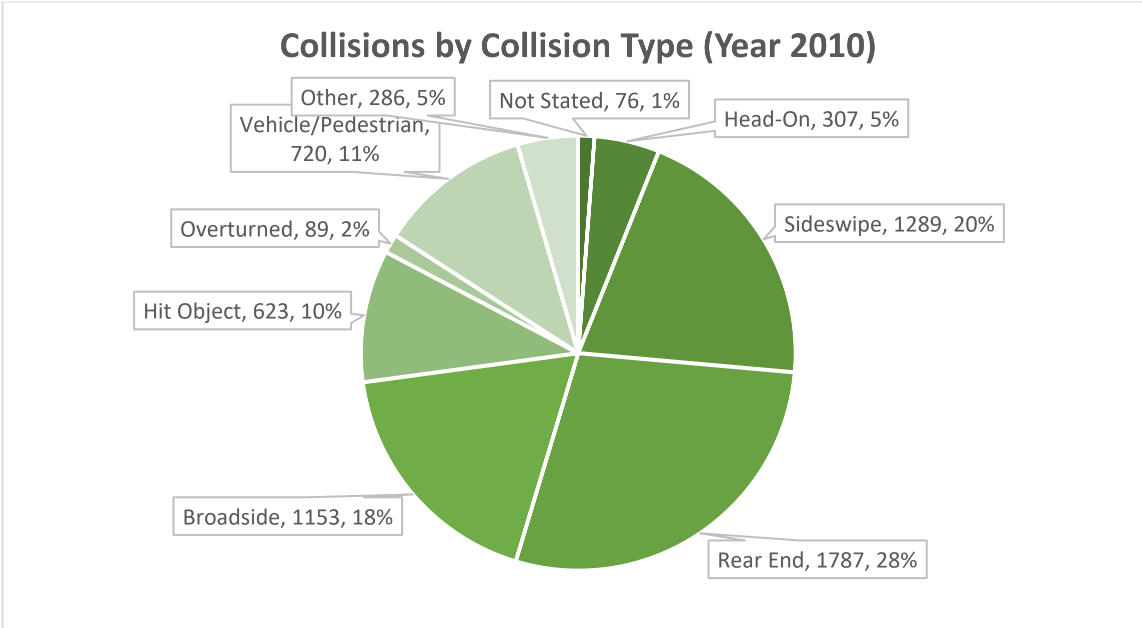

Figure 4-4: Collisions by Collision types in the year 2010

More frequent occurrences of read-end crashes indicate that the area is heavily congested. The crashes may not be fatal, but they tend to occur because of motor vehicles driving too close to each other or due to distracted driving.

Broadside crash is the name given to road crashes that happen when the front of one vehicle hits the side of the other. While they may lead to severe injuries and deaths, they are often preventable and are likely to be caused due to inattention or negligence. These crashes are usually found near road intersection areas because maneuvering is required to negotiate the area. Add to that the complexity of other road users who are also trying to utilize the road space simultaneously as motor vehicles. It creates an ideal situation, especially when drivers are not paying attention.

Like any urban area, sideswipe crashes are also abundantly found in the SF area, indicating that crash occurrence is likely not only due to drivers' unsafe conditions but also related to traffic conditions, speed, and roadway geometric features.

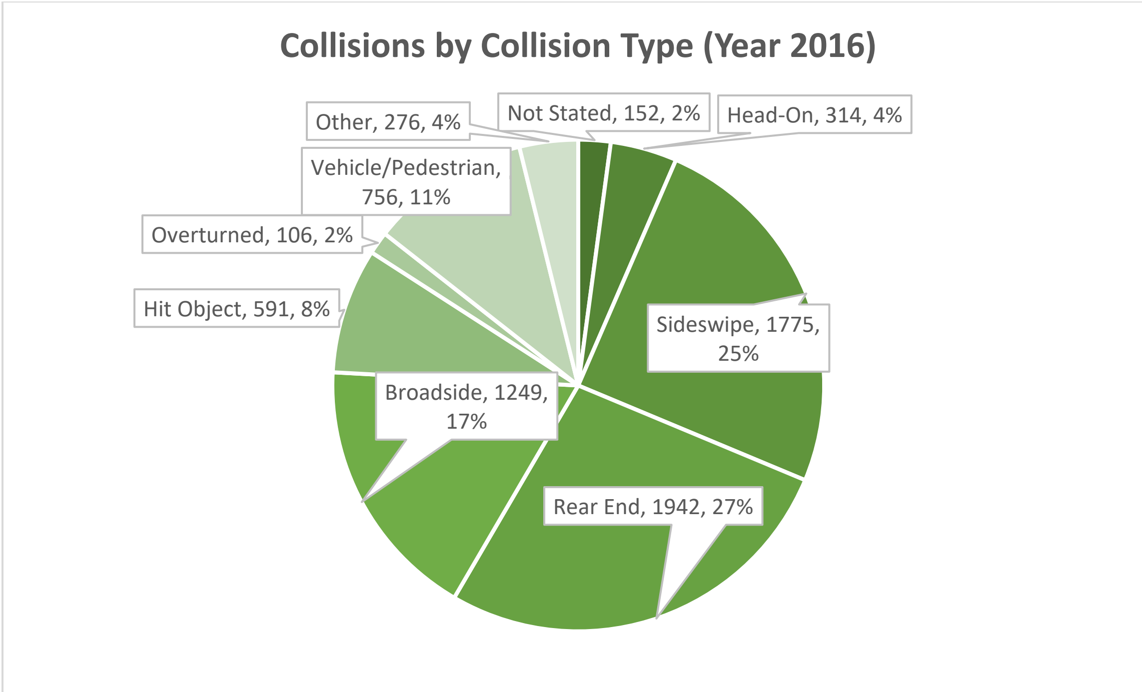

Figure 4-5: Collisions by Collision types in the year 2016

4.2.4 Primary Collision Factors (PCF)

Figure 4-6 below shows the top violations which resulted in injury collisions. The top PCF violation category is *03-Unsafe Speed*.

Unsafe Speed is not always over the speed limit, and it indicates the vehicle was traveling at a pace that was not appropriate as per traffic/weather conditions. Other factors include improper turning, unsafe lane changing, and traffic signal violation which may result in the rear end, broadsides, and sideswipe, as mentioned in the previous section.

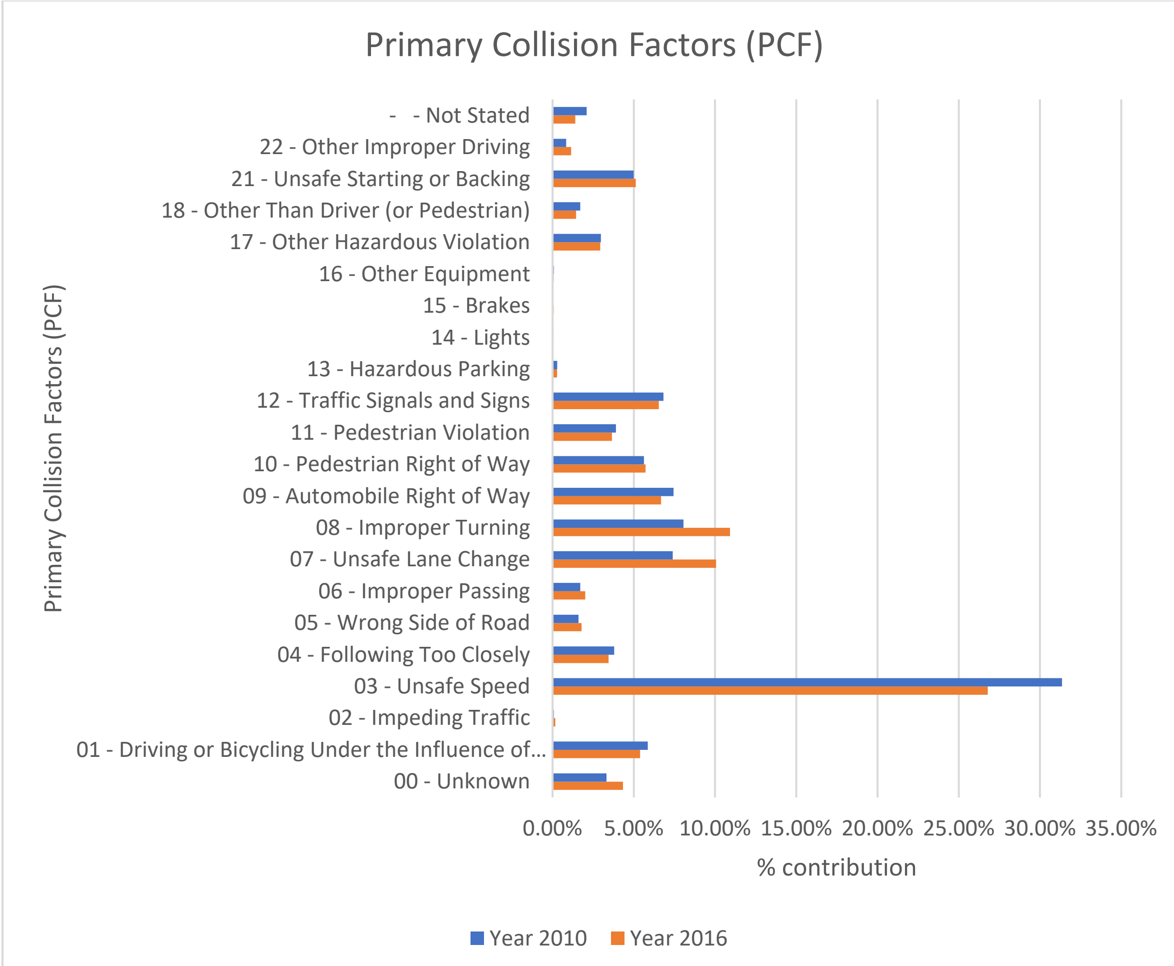

Figure 4-6: Primary collision factors

4.2.5 Census Tract

4.2.5.1 Mapping road crashes

Figure 4-7 through Figure 4-11 maps the percentage difference between various road crash frequency types between 2010 and 2016 in each census tract. There are four color variants used to prepare such figures.

a) Shades of red indicate that road crash incidents increased in 2016 w.r.t 2010. The darker the red color, the bigger is percentage increase.
b) Shades of green are the opposite of red and signify a decrease in road crash instances in 2016 w.r.t. 2010. The darker the shade of green, the more significant the percentage decrease.

c) Color white to denote that there is neither a decrease nor increase in crash instances

d) The color grey signifies that the percentage difference cannot be determined. The scenario may arise if no crashes occurred in 2010 but existed in 2016.

Figure 4-7 shows the total number of crashes occurring per census tract. In general, all census tracts have witnessed an increase in crash incidents. Census Tract containing the Golden Gate Bridge and San Francisco - Oakland Bay Bridge seems to witness a more significant number of crashes and is primarily driven by the PDO crash instances. It is intuitive, given that the locations are entry or exit points to reach San Francisco.

The downtown area in North-East also has witnessed a reasonable percentage increase in road crashes. However, it is to be noted that these areas do not witness the highest jump in road crashes.

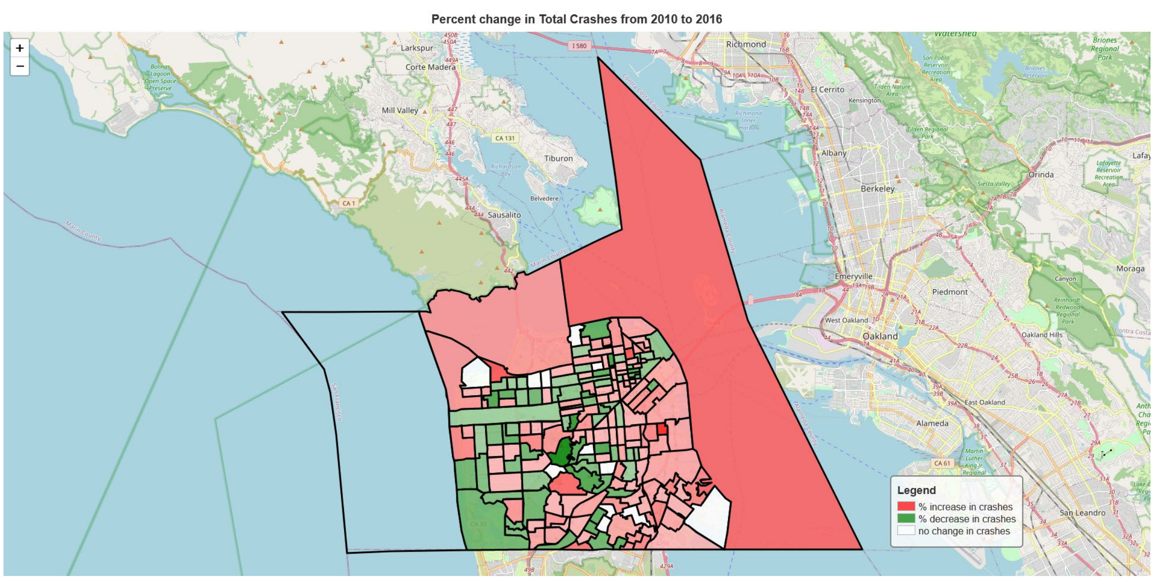

Figure 4-7: percent difference between Total Crashes between 2010 and 2016

Like the Total Crashes, the fatal and injury crashes (see Figure 4-8) also follow a similar trend, with most crashes happening in the northeast and southeast. It is also the area that has a higher concentration of freeway network.

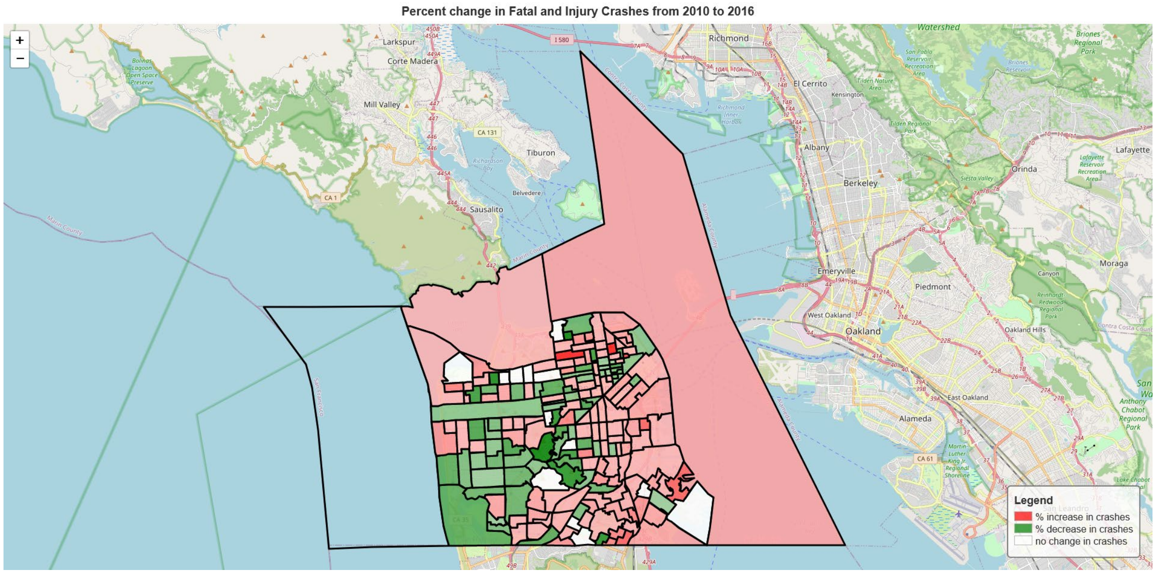

Figure 4-8: percent difference between Total Fatal and Injury Crashes between 2010 and 2016

There is no set pattern in Pedestrian and Bicyclist crashes, as shown in Figure 4-9, with many incidents occurring in isolated pockets. The most prominent areas include SOMA, Chinatown, South Park/ Oracle Park, the area surrounding Lake Merced, Balboa Park, and areas near Lombard Street downtown. These are also areas that are visited by tourists and have a strong presence of non-motorized transport infrastructure facilities.

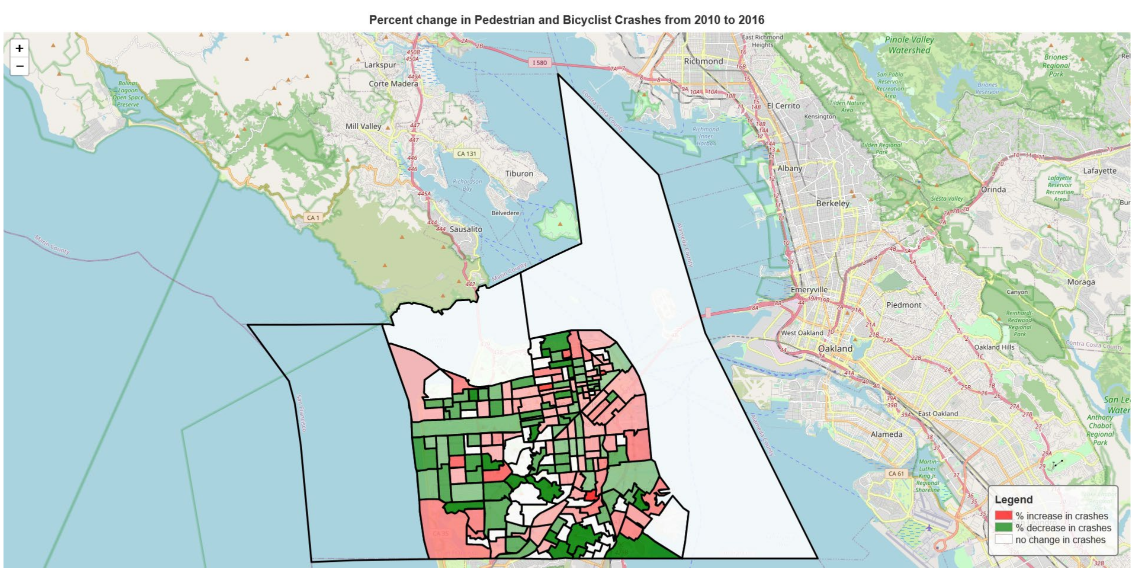

Figure 4-9: percent difference between Pedestrian and Bicyclist Crashes between 2010 and 2016

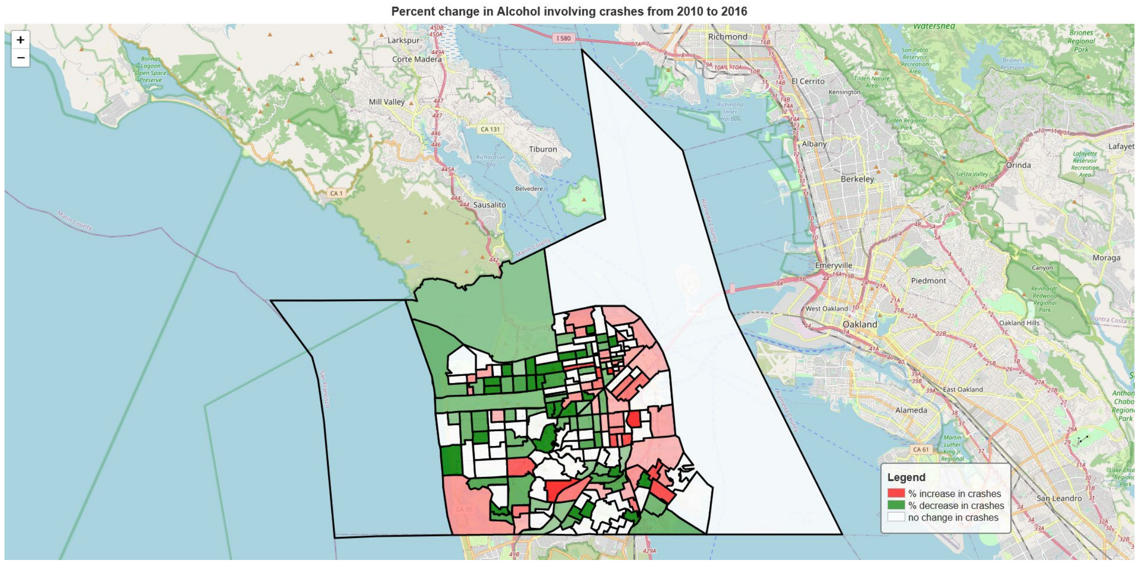

Figure 4-10: percent difference between Alcohol-involved (DUI) crashes between 2010 and 2016

Both DUI-related crashes are rare and majorly witnessing a decreasing trend. Their occurrences are prominent around downtown areas with the maximum number of leisure and tourist spaces.

Similar to Total Crashes and Fatal and Injury crashes, PDOs are also spread everywhere. The study finds a general pattern: higher-density areas and traffic flows have the largest concentration of fatal crashes and crashes that cause injury and property damage.

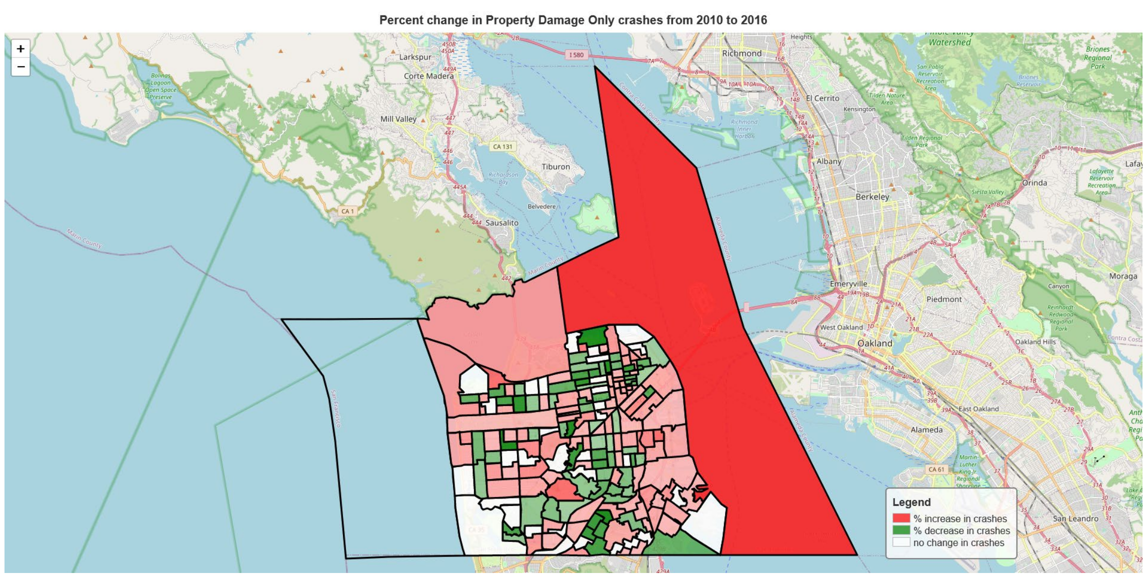

Figure 4-11: percent difference between PDO Crashes between 2010 and 2016

Table 4-5 and Table 4-6 presents the five-point summary for each road crash type in absolute and percentage terms. The min and max values for percentage changes observed between 2010 and 2016 cannot be accurately calculated because the base values in 2010 were zero in many instances.

Table 4-5: descriptive statistics for the difference between 2010 and 2016 for each road crash type

	Total Crashes	Fatal And Injury Crashes	Pedestrian And Bicyclist Crashes	Alcohol-Involving (DUI) Crashes	Property-Damage Only (PDO)
min	-16	-14	-14	-8	-12
25%	-2	-2	-2	-1	-1
50%	1	1	0	0	0
75%	8	4.5	2	1	4
max	104	39	17	14	80

Table 4-6: descriptive statistics of the percent changes (in %) observed between 2010 and 2016 for each road crashes type

	Total Crashes	Fatal And Injury Crashes	Pedestrian And Bicyclist Crashes	Alcohol-Involving (DUI) Crashes	Property-Damage Only (PDO)
min	-100.0	-100.0	-100.0	-100.0	-100.0
25%	-14.7	-20.0	-40.0	-50.0	-26.8
50%	11.1	12.5	0.0	0.0	0.0
75%	50.0	50.0	40.8	0.0	89.9
max	333.3	500.0	500.0	400.0	727.3

4.3 Road Network

Previously. Section 3.5.1 presented an overview of the process followed in the study to merge and aggregate the SF-CHAMP data into one road network for each year, i.e., 2010 and 2016. Further, the aggregate network is re-categorized into four facility types, as in Table 3-1. The total length of the road network (in miles) in each category for the years is presented below:

Table 4-7: Road network (in miles) per road category (2010 vs. 2016)

Year	Freeways	Arterials	Collectors	Locals
2010	87.07	419.82	349.48	1607.51
2016	87.07	420	350.27	1608.08
Absolute diff	0	0.18	0.79	0.57
% Diff	0.00%	0.04%	0.23%	0.04%

Table 4-7 shows that SF has a well-developed road network with little to no change in lane miles going from 2010 and 2016. The stableness of the road network is advantageous for the current study as most of the marginal changes can be absorbed by the fixed-effect variable.

However, SFCTA continues to invest in non-motorized transport (NMT) infrastructure like pedestrian walkways and bike lanes in their pledge towards sustainability and to make the network safer and more comfortable to access. In addition, in 2014, San Francisco adopted Vision Zero to eliminate all traffic deaths and severe injuries by 2024, and they have been working aggressively towards it.

While these are necessary projects, we do not believe they influence motorable road networks because many of the improvements enhance pedestrian and bicycling infrastructure. A report by SFCTA indicates that as of December 2015, 30 high-priority projects were undertaken for improvement, resulting in 1599 safety treatments installed along with more targeted enforcement and public awareness schemes (SFCTA, 2015). These safety treatments ranged from building bulb-outs and refuge islands to daylighting and pedestrian countdown signals.

4.3.1 Vehicle miles traveled (VMT)

Table 4-8 below presents the VMT on each road category. The steepest increase in VMT between 2010 to 2016 is found on Collectors roads. It is followed by Local and Arterials roads. Overall, the traffic has risen by 14%, with most of the contribution coming from Arterials, Collectors, and Local roads.

Table 4-8: VMTs per road facility type (2010 vs. 2016)

Year	Freeways	Arterials	Collectors	Locals	Total
2010	4,227,881	4,106,867	743,368	955,544	10,033,660
2016	4,633,180	4,756,220	923,290	1,132,927	11,445,617
absolute difference	405,299	649,353	179,922	177,383	1,411,957
% increase	9.6%	15.8%	24.2%	18.6%	14.1%

Compared to the stats above, Table 4-9 below presents the VMT comparison per road facility type for 2010 and CF 2016. In the counterfactual scenario, the VMT has increased modestly across all the facility types. The percentage increase does not follow the big jumps observed in the 2016 scenario.

Table 4-9: VMTs per road facility type (2010 vs. CF 2016)

Year	Freeways	Arterials	Collectors	Locals	Total
2010	4,227,881	4,106,867	743,368	955,544	10,033,660
2016	4,497,834	4,443,617	809,983	1,016,541	10,767,976
absolute difference	269,953	336,750	66,615	60,997	734,316
% increase	6.4%	8.2%	9.0%	6.4%	7.3%

The study focuses on the TNC-related VMTs and finds the rise to spread across all facility types with more prominence on Arterial, Local, and Collector roads. The information is consistent with the conclusion of (G. Erhardt et al., 2019; Roy et al., 2020), which are the primary source of the data but also with (Gehrke et al., 2018; Henao & Marshall, 2019; Schaller, 2017) study indicates that the net effect of TNCs on VMT is additive.

Table 4-10: TNC VMTs per road facility type (2016)

Year	Freeways	Arterials	Collectors	Locals
2016	184,142	473,208	94,606	107,531
% contribution w.r.t Tot VMT difference between 2010 and 2016	45.4%	72.9%	52.6%	60.6%

4.3.2 Congested speed

Table 4-11 below presents the (weighted average) congested speed across road segments for facility types for 2010 and 2016. The last column refers to the overall (weighted average) congested speed at the county level. The most significant percentage drop is on freeways, followed by collectors, while the rest categories have undergone marginal differences.

While the drop in congested speed does not seem much, it follows the positive correlation trend with Total VMT. The more miles people drive, the more vehicles are on the roadways. Higher numbers of vehicles eventually result in congestion.

Table 4-11: weighted average congested speed in miles per hour (2010 vs. actual 2016)

Year	Freeways	Arterials	Collectors	Locals	SF County
2010	42.12	15.25	17.28	15.70	27.44
2016	40.20	15.06	16.64	15.60	25.95
absolute difference	-1.92	-0.18	-0.64	-0.10	-2.84
% increase	-4.5%	-1.2%	-3.7%	-0.6%	-10.4%

Table 4-12 below presents the (weighted average) congested speed across road segments for facility types for 2010 and the CF scenario of 2016. The last column refers to the overall (weighted average) congested speed at the county level. The most significant percentage drop is on freeways, while the remaining categories have undergone marginal differences.

Table 4-12: weighted average congested speed in miles per hour (2010 vs. CF 2016)

Year	Freeways	Arterials	Collectors	Locals	SF County
2010	42.12	15.25	17.28	15.70	27.44
2016	40.70	15.20	17.03	15.71	26.65
absolute difference	-1.42	-0.04	-0.25	0.01	-1.70
% increase	-3.4%	-0.3%	-1.5%	0.0%	-6.2%

Comparing Table 4-11 and Table 4-12, it can be stated that the road network of actual 2016 is more congested than the counterfactual scenario. It follows up from the VMT stats, showing that the rise is much more considerable for 2016 than its counterfactual scenario.

Roy et al. explain the rise and decrease in congestion speed between 2010 and 2016. A few of the prominent ones are mentioned here:

- The population of SF county increased from 805,000 to 876,000 (31);
- employment grew from 545,000 to 703,000
- Looked for any significant network changes during this period, especially the rebuilding of the Presidio Parkway, the introduction of turn restrictions on Market Street, several "road diets," and bus improvements. However, these projects' likely impacts have been directly accounted for in SF-CHAMP traffic data
- Their study also investigated active construction projects during 2016 that may be responsible for the speed decrease. However, they did not find any such projects

4.3.3 Road Network with highest road crashes

Figure 4-12: Road segments with frequent road crash occurrences (YR 2010 vs. 2016)

Figure 4-12 provides a comparative picture of the road segments where the maximum number of (total crash) crashes occurred for both years. Again the northeast and southeast part are much flared up in both 2010 and 2016; however, visible increases (difference) are observed in the 2016 tally, particularly on the I-80 stretch, both inbound and outbound from the San Francisco – Oakland bridge connecting I-101 in the South, and I-280 running South West. Another prominent location that witnessed an uptick in crash occurrences is the Mission District area, which has a dense road network and the entry point to the downtown area from the South, West, and Eastern part of the county. It is indicative that crashes occur where there is a complex interaction between various road user groups, vehicles, and road infrastructure connection spots.

4.4 Where are the TNCs rendering their services?

Knowing the TNC trip's origin and destination geolocation can provide a robust understanding of how likely it is to exacerbate its negative consequence on other road users. As covered in the literature review section, the negative externalities include traffic congestion and conflicts, which may lead to traffic crash incidents.

SFCTA, in their public report, has reviewed the question extensively and in detail (SFCTA, 2017). They identified that most TNC activity occurs in the northeast quadrant, which is also the most congested part of the city. The area is well served by all modes of transportation, public transit, bicycling, and walking facilities. Particular areas include the South of Market, the Mission Street corridor, the Van Ness Avenue corridor, Pacific Heights and the Marina, Geary Street corridor, Panhandle, Inner Sunset, and the Stonestown/San Francisco State University area.

Erhardt et al., and Roy et al., found that of the 13% growth in Total VMT between 2010 and 2016, TNCs contributed to almost half of it (G. Erhardt et al., 2019; Roy et al., 2020). Figure 4-13 presents the spatial overview of the most impacted streets due to TNC activities (San Francisco County Transportation Authority, 2018). When the pattern from Figure 4-13 is compared with Figure 4-12, it can be concluded that these are the same streets witnessing an uptick in total road crashes.

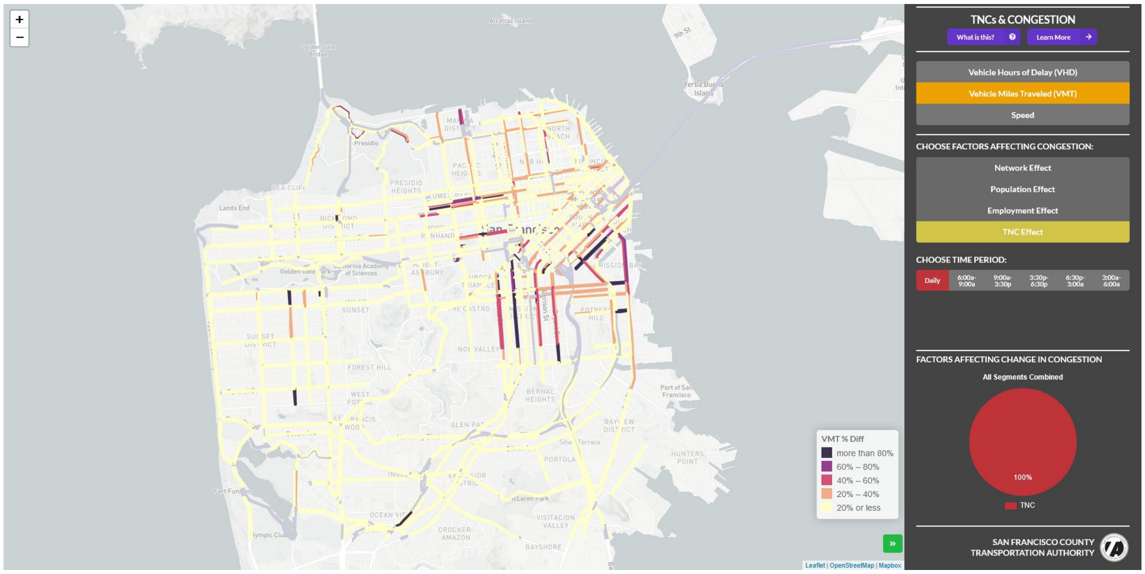

Figure 4-13: Percentage contribution of TNC activities on the VMT traveled (source: https://tncsandcongestion.sfcta.org/)

4.5 Summary

The chapter presents results from the explorative analysis of traffic crashes, traffic estimates, and TNC data to understand the spatial distribution of these activities.

Table 4-1 presents the countywide fatal and injury trends for the primary dependent variables: total crashes, fatal and injury crashes, pedestrian and bicyclist crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes. The key takeaway from the comparative trends is that the jump in total crashes (+20%) is primarily driven by the sharp rise in property damage-only crashes, which went up by 30%, and a moderate increase in fatal and injury crashes (approx. 8%). Pedestrian and bicyclist crashes (+3%) and Alcohol-involving (DUI) crashes (-3.4%) remained stagnant between 2010 and 2016. The composition (Table 4-2) is maintained even after the study undertook the geocoding process to populate spatial coordinates in the crash records without such information.

The research further decomposes road crashes by road facility types (Table 4-3 and Table 4-4). Arterials witness the highest impacts in absolute numbers, but the percentage rise is somewhat nominal, with the sharpest increase found in PDO crash types (+20%). In comparison, collectors and local facility types witness less than half of such numbers. Local roads see a 20% jump in total crash occurrences and PDO crashes. Overall, 75% of the crashes occur on Arterials and Locals roads except for pedestrian and bicyclist crash types, which are Arterials and Collector roads.

Non-motorized transport users (pedestrians and bicyclists), followed by motorcyclists, are the most vulnerable road user group and constitute 75% of all fatal victims (see Figure 4-1). The nonfatal injuries (excluding PDOs), a much more reliable indicator of long-term city road crash trends, are also assessed. However, the percentage jumps and drops are well within the observed fluctuations compared to the past five-year non-fatal injury collision totals (see Figure 4-2), indicating that non-fatal injuries long-term average remained unchanged between 2010 and 2016.

The rear end, broadsides, and sideswipes remain the most frequently occurring road crashes between motor vehicles, increasing from 20% to 25% between 2010 and 2016. At the same time, it is indicative that urban area is heavily congested, mainly because motor vehicles are driving too close to each other or due to distracted driving.

Aggregation of the crashes at the census tract level does not present any emergence of spatial pattern limiting the identification of dangerous locations within the road network. It further reaffirms that crashes result from randomness and are tightly bound to local factors when the road crash incident occurs.

San Francisco county has a very mature road network with very little or almost negligible additional lane miles added in 2016. The constantness in the road network is advantageous for the study and re-affirms our decision to use a fixed-effect estimator over random effects. The model would absorb any marginal changes in road network lane length.

Most road network changes undertaken between 2010 and 2016 have focused on improving accessibility for pedestrians and bicyclists through non-motorized transport (NMT) infrastructure. In addition, in 2014, San Francisco adopted Vision Zero to eliminate all traffic deaths and severe injuries by 2024. A report by SFCTA indicates that as of December 2015, 30 high-priority projects were undertaken for improvement, resulting in 1599 safety treatments installed along with more targeted enforcement and public awareness schemes (SFCTA, 2015). These safety treatments ranged from building bulb-outs and refuge islands to daylighting and pedestrian countdown signals. Hence, their influence on vehicular traffic estimates is likely minimal.

However, the study observed the following changes to Vehicle Miles Traveled (VMT), Congested Speed

- Overall, Total VMT rose by 14%, between 2010 and 2016, with an almost one-fourth-fold increase on Collector road, followed by Locals (+18%) and Arterials (+15%). VMT on Freeways rises by 10%. It suggests the rapid increase in vehicular movement is happening away from the freeways and within the city's core network.
- The study focuses on the TNC-related VMTs and finds the rise to spread across all facility types with more prominence on Arterial, Local, and Collector roads. The information is consistent with the conclusion of (G. Erhardt et al., 2019; Roy et al., 2020), which are the primary source of the data but also with (Gehrke et al., 2018; Henao & Marshall, 2019; Schaller, 2017) study indicates that the net effect of TNCs on VMT is additive.
- The most significant reduction in Congested Speeds happens at Freeways, which dropped by 2 miles per hour between 2010 and 2016, followed by Collector roads. While the drop in congested speed is not proportional to the rise in VMTs, it follows the positive correlation trend with Total VMT. The more miles people drive, the more vehicles are on the roadways. Higher numbers of vehicles eventually result in congestion.

Spatial comparison (see Figure 4-12) of the road segments where the maximum number of (total crash) crashes occurred for both years indicate that many road crashes occur in the northeast and southeast part. However, the visible difference is observed on the I-80 stretch, both inbound and outbound from the San Francisco – Oakland bridge connecting I-101 in the South and I-280 running South West. Another prominent location that witnessed an uptick in crash occurrences is the Mission District area, which has a dense road network and the entry point to the downtown area from the South, West, and Eastern part of the county. The spatial diversity is indicative that crashes occur where there is a complex interaction between various road user groups, vehicles, and road infrastructure connection spots.

Bay Bridge appears to be an outlier and is worth conducting statistical analysis with and without this census tract.

Most TNC activity occurs in the northeast quadrant of SF County, which is also the most congested part of the city. The identified area is well served by all modes of transportation, public transit, bicycling, and walking facilities. Particular areas include the South of Market, the Mission Street corridor, the Van Ness Avenue corridor, Pacific Heights and the Marina, Geary Street corridor, Panhandle, Inner Sunset, and the Stonestown/San Francisco State University area.

CHAPTER 5 MODEL RESULTS

5.1 Background

CHAPTER 3, section 3.4.2.4 presented the statistical model's structure to estimate the relationship between road crash frequency occurrences and independent variables. The available independent variables are congested speed – freeways, congested speed – other roads, TNC PUDO, TNC VMT, and the time-fixed effects introduced as a binary variable, i.e., year_2016_dummy.

This chapter presents the most desirable regression model these variables could provide to predict crashes. The models are compared to estimate their accuracy in predicting the dependent variable, i.e., crash frequencies. Section 5.2 presents the most suitable model and its model output. Section 5.3 further describes the measures adopted to test the robustness of the model shown in section 5.2. Other variants of the regression model with different parametric specifications are also presented along with the predicted crash estimates in section 5.4.

5.2 Main Results

Table 5-1 below shows statistical model results for all variants of dependent variables like Total Crashes, Fatal and Injury Crashes, Pedestrian and Bicyclist involving crashes, and alcohol-involving (DUI) crashes.

The study also includes PDO as the dependent variable to understand if there exists a strong relationship between TNC service and non-fatal injuries. Exploring the relationship is essential, given that PDO crashes are typically not part of the FARS report and account for more than 50% of crashes in the general population. Other traditional independent variables include Congested Speed on Freeways and other road network and their square term, the natural log of TNC-related PUDO, and the dummy variable termed year_2016_dummy to capture time-effect changes. The primary model includes Total VMT as a regressor, not an exposure variable.

Table 5-1: Main results of the statistical modeling for identified dependent variables

Model (→)	**1**		**2**		**3**		**4**		**5**	
Y (→)	**Total Crashes**		**Fatal and Injury Crashes**		**Pedestrian and Bicycle crashes**		**Alcohol-involving (DUI) crashes**		**Property-damage-only (PDO)**	
X (↓)	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**
Total VMT	0.37 *	0.159	0.361 *	0.177	0.15	0.272	0.863	0.529	0.349	0.25
Cong Spd [Freeways]	-0.636 **	0.204	-0.188	0.241	-1.23	2.079	-0.53	0.672	-0.936 *	0.422
Cong Spd [Freeways]2	0.001	0.002	-0.003	0.002	0.003	0.023	0	0.007	0.002	0.005
(Cong Spd [Other road network])	0.6	0.326	0.768 *	0.361	0.433	0.516	0.389	0.737	0.155	0.561
(Cong Spd [Other road network])2	-0.022	0.012	-0.029 *	0.013	-0.021	0.02	-0.028	0.026	-0.005	0.022
Ln (TNC Pick-up and drop-offs)	0.065 **	0.021	0.055	0.029	0.092	0.08	0.056	0.065	0.093 *	0.037
Ln (Tot TNC VMT)	-0.065 **	0.022	-0.057	0.032	-0.095	0.081	-0.074	0.075	-0.081	0.05
Year 2016 Dummy	0.054	0.147	0.022	0.178	0.009	0.265	-0.122	0.455	0.08	0.263
Log Likelihood	-1012		-875.34		-577.836		-370.314		-687.317	
Wald Chi2 (7)	379.9		51.95		6.39		11.92		79.45	
*p < 0.05, **p < 0.01										
Estimates are from the panel Fixed Effects Poisson regression model with Robust Standard errors. All models include entity and time-fixed effects.										

5.2.1 Interpretation of the model coefficients

For the total crash as the dependent variable, the coefficient of total VMT is positive and equal to 0.37. The value is significant at a 95% confidence interval. The coefficient of Total VMT for other road crash types is also positive. Given that its coefficient is not equal to one suggests that road crashes do not have a linear (direct) relationship with VMT. It also explains why road crashes and their occurrence are considered complex and dependent upon a diverse set of behavioral and engineering factors which vary according to the area, local topographical conditions, and traffic habits of the residents. As Total VMT participates in the model in the natural logarithmic form, its impacts change from a unit to a percent change. Therefore the percentage increase can be easily interpreted as a 10% increase in the Total VMT variable (not its log), resulting in an increase of crashes by 3.7%, 3.61%, 1.5%, 8.6%, and 3.49% for Total Crashes, Only Fatal and Injury involving Crashes, Only Pedestrian and Bicyclist involving crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes, respectively.

For the total crash as the dependent variable, the coefficient of congested speed freeway has a negative value of -0.636. Its square term has a negative coefficient value of +0.001. The relationship validates our assumption that the relationship between road crashes and speed is non-linear. A big negative coefficient term and almost negligible positive value for the square terms indicate that rate of decrease of road crashes has a negative linear slope. In other words, the rate of road crash occurrence increases non-linearly as the difference in congested speeds increases. Except for total crashes and PDO, the linear term of congested speed freeway is insignificant for other road crash types at a 95% confidence interval.

On the contrary, the congested speed on the rest of the road network has a positive coefficient value of 0.6, while its square term has a negative coefficient of -0.02. The congested speed on the rest of the road network and its square term is significant at a 95% confidence interval for only fatal, and injury crashes. Their coefficients can be interpreted similarly to those for congested speed for freeways. Contrary to congested speed freeways, the coefficient for other road networks indicates that as the congestion on these road networks increases, the congested speed decreases (i.e., average speed further goes down), thereby increasing the probability of witnessing an increase in road crash occurrence.

Opposite signs both for the linear and non-linear term of the congested speed also indicate that the rate of decline in congested speed is more gradual for freeways compared to other road networks. The rate of change is presented graphically in Figure 5-1.

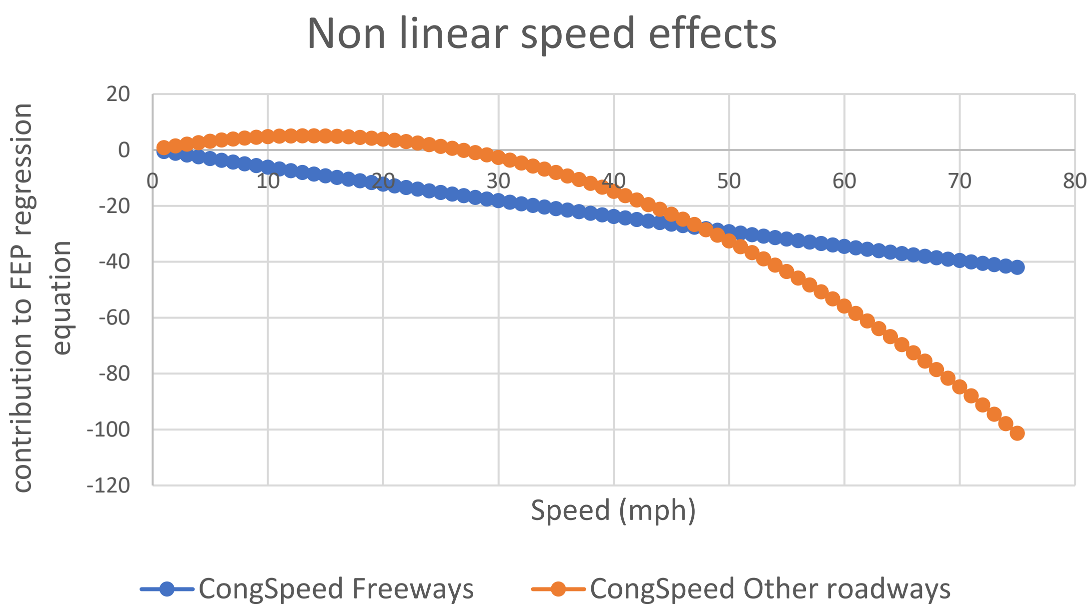

Figure 5-1: rate of change of congested speed (freeway vs. other road networks)

Intuitively, these crash factors make sense. For example, consider the multiplicative factors of the congested speed–freeways (0.53 & 0.83) and other road networks (1.78 & 2.09) for the total crash and fatal & injury crashes, respectively. The multiplicative factor for fatal and injury crashes caused (0.83*2.09) by these parameters is more significant than those estimated for total crash (0.53 * 1.78) frequency. It is because the operating speed of motor vehicles is directly proportional to the crash outcome. Higher operating speeds generate more incredible kinetic energy[3] (K.E). It means a large amount of K.E. gets dissipated from the high-speed motor vehicle to the other crash-participating object at the time of a road crash. If the energy dissipated far exceeds the level, the human body can tolerate it, resulting in a severe or fatal injury. For example, suppose a pedestrian gets hit

[3] By definition kinetic energy is the energy that a body possesses by being in motion

by a vehicle traveling at 15 mph (typical of the local road environment). The risk of bodily injury is less than 10% but rises to about 50% if the vehicle travels at 45 mph (typical on the highway) (WHO, 2008). Suppose one assumes the odds of getting into a road crash in all road category types to be equal. In that case, the resultant motor crash severity depends entirely on kinetic energy. Therefore, the crash severity resulting from crash occurrence is relatively high when the operating speed is high and is precisely what these coefficients convey (Kockelman, 2006).

On the contrary, the multiplicative factor for pedestrian, bicycle, and alcohol-involving crashes is smaller than their fatal, injury, and total crashes counterparts. It is because the odds of involving in a motor vehicle crash depend not only on the operating speed but also on the characteristics of the involved parties, i.e., one of the participants shall either be a pedestrian or bicyclist or consumed alcohol which decreases the odds of likelihood.

Similarly, the partial effects (or the probability) of getting involved in a PDO is higher on local or arterial roads than on the freeway. It goes back to the relationship between speed and crash severity. At a higher speed, the chance of a road crash that is purely a PDO crash is somewhat limited, if not rare. Instead, there is a more substantial chance that such a crash incident results in a) a road traffic injury or a fatality and b) a secondary PDO incident. However, since crash incidents get tagged as the most serious, the chances of having such a type are rare.

The chances of getting into a PDO crash are high at other road networks for various reasons, including lower speeds, stop-and-go driving, vehicles failing to yield, differential speeds among the participating vehicles, more significant instances of speed limit violations, or instances of driving too closely. Lower speeds at high population density areas increase the probability of getting involved in fender-bender road crashes (Ewing & Hamidi, 2015).

The partial effects of congested speed–freeways between alcohol-involving crashes aligned with expected trends. According to NHTSA, around 88% of all alcohol-involving crashes occur away from interstate roads (NHTSA, 2022b).

The relationship between congested speed – on other road networks and alcohol-involving crashes (DUI) is somewhat surprising and opposite to the generally accepted trend that drinking likely occurs more on weekends and at night. These opposite trends suggest that these results are unstable and should not be relied upon for further analysis.

There are a few possible reasons for such contrasting trends:

- Alcohol-involving (DUI) crashes increased over the years by minimal numbers even after more vigorous enforcement under SF Vision Zero Scheme. It may also be the result of underreporting or the result of the data aggregation process adopted in conducting the current study (crash incident drops out if the nearest road segment it connects is greater than 10m). The variability in the data tends to make the analysis much more unstable, making it more challenging to establish a meaningful trend.
- Low-traffic volume conditions and high free flow speeds drive alcohol-impaired drivers to overestimate their driving skills, creating more opportunities for crash incidents (Elvik et al., 2012). These components partially explain the positive coefficient on alcohol crashes. However, in addition to the driving conditions, drivers also consider urban settings, distance to travel, and regulatory policies. It may be related to enforcement (increased during holidays, nights, or weekends), penalties, and availability of alternative modes of transport, general awareness among the public, or simply the transportation choices in the city, explaining the negative coefficient on alcohol crashes involving only nights, weekends, and holidays.
- It may indeed be that TNC services may have indeed reduced DUI incidents and bookings. A report published by the National District Attorney Association (NDAA), which studied the relationship between Lyft service trips (volume) and DUI incidents in three cities of California, including San Francisco,

indicates a downward trend in DUI total incidents and bookings. Our study results are also partially supported by Greenwood & Wattal's study, which examined California's DUI data records between 2009-2015 and found UberX to reduce DUI rates by 3.6% (Greenwood & Wattal, 2017). However, it contradicts the conclusion of (Brazil & Kirk, 2020, 2016; Dills & Mulholland, 2017; Kontou & McDonald, 2020b), which found no significant relationship between DUI incidents and TNCs.

Moving on to the TNC PUDO, the coefficients are positive, weak in absolute values, and significant at 95% confidence for only total crashes. Using the natural logarithmic of PUDO in the model form transforms its coefficient into elasticity. Therefore, the percentage increase can be easily interpreted as a 10% increase in the PUDO variable (not its log), increasing the crash occurrences by 0.65%, 0.55%, 0.92%, 0.56%, and 0.93% for Total Crashes, Only Fatal and Injury Crashes, Only Pedestrian and Bicyclist involving crashes, alcohol-involving crashes, and property damage only crashes respectively. The weak relationship might be related to the limitation of the NHTSA crash database, which restricts crash entry records to those involving motor vehicles, i.e., where motor vehicles are in transport (in motion). It may also suggest that pick-up and drop-off activity has little effect on crash outcomes.

In contrast, the coefficients for Tot TNC VMT are negative and somewhat opposite to the coefficient of PUDO. The coefficients are statistically significant at 95% confidence only for total crashes. Like the PUDO, the natural logarithmic of Tot TNC VMT allows us to interpret the coefficient as elasticity. Therefore, a 10% increase in the Tot TNC VMT variable (not its log) decreases the crash occurrences by 0.65%, 0.57%, 0.95%, 0.74%, and 0.81% for Total Crashes, Only Fatal and Injury Crashes, Only Pedestrian and Bicyclist involving crashes, alcohol-involving crashes, and PDO, respectively. The negative association could be a combination of two or more aspects, namely a) the selection of TNC drivers, which bares entry of unskilled drivers, b) the driving skills of TNC drivers getting superior with every additional mile driven, c) the feedback system placed by the TNC service providers to rate the drivers after the end of a TNC trip, with minimum drive score to be maintained. The feedback covers all aspects of driver personality – including driver hygiene, driving habits, speeding, rash driving, and much more. Also, making it mandatory for the drivers to maintain a specific threshold score pushes the TNC drivers to be obliged with the requirements. d) driving more vehicle miles makes the TNC drivers superior and more efficient in handling fatigue than ordinary drivers. e) since TNC drivers are the vehicle owners and therefore have a reasonable incentive to drive and maintain the vehicle gently.

We include both TNC PUDO and TNC VMT in the model specification because they have different effects. PUDO is specifically about vehicle interactions at the curbside. In contrast, TNC VMT may be a replacement for private VMT with another kind of driver or a replacement for miles driven while impaired by alcohol with miles driven by a TNC driver. However, the two factors are also correlated, so we include both variables with some risk of biasing the magnitude of these coefficient estimates through collinearity. In robustness checks presented later, we examine how these coefficients would change if we included only one or the other.

The 2016-year dummy variable tests for spatially uniform changes across the two time periods. For example, if the mix of vehicles in the fleet changes between two years in a way that affects crash outcomes, we might expect those changes to be relatively uniform across all parts of the city. It can be due to various reasons like travel behavior changes, vehicle standards (& technology adoption), law-enforcement of traffic laws, and (positive) changes in post-crash care treatments. The 2016 dummy coefficients vary in sign and magnitude for the different types of crashes, but all are insignificant at a 95% confidence interval.

5.2.2 Comparing the estimated vs. observed crashes

The final model is presented in Equation 5-1, where the β's represent the corresponding coefficients for the respective variable for the crash types, as summarized in Table 5-9. These coefficients and the (actual) observed values for each independent variable participating in the model can be used to estimate the crash frequency for each road category per census tract.

Suppose we want to calculate the predicted crashes for road category "2" in Census Tract "010100". The process is as follows:

The first step is to get all relevant information related to the model parameters (independent variables). Given that the road category is "2", the congested speed–freeways would be zero. Other observable values for independent parameters for the corresponding census tract road category are shown in Table 5-2

$$E\left[\frac{\Delta total\ crashes}{\Delta Tot\ TNC\ VMT}\right] = total\ crashes_{(i,2010)}\left[\left(\frac{Tot\ VMT_{(i,2010)}}{Tot\ VMT_{(i,2010)}}\right)^{\beta_1} * \left(\frac{e^{\beta_2 CongSpeed\ Freeway_{(i,2010)}+\beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}}{e^{\beta_2 CongSpeed\ Freeway_{(i,2010)}+\beta_3 CongSpeed\ Freeway^2{}_{(i,2010)}}}\right) * \left(\frac{e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)}+\beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}}}{e^{\beta_4 CongSpeed\ Otherroads_{(i,2010)}+\beta_5 CongSpeed\ Otherroads^2{}_{(i,2010)}}}\right) * \left(\frac{PUDO_{(i,2010)}}{PUDO_{(i,2010)}}\right)^{\beta_6} * \left(\frac{Tot\ TNC\ VMT_{(i,2010)}+1}{Tot\ TNC\ VMT_{(i,2010)}}\right)^{\beta_7} * \left(\frac{e^{\beta_8 Year_{2016 dummy}{}_{(i,2010)}}}{e^{\beta_8 Year_{2016 dummy}{}_{(i,2010)}}}\right)\right] - total\ crashes_{(i,2010)}$$

Equation 5-1 - Final model representation form

Table 5-2: Observable values of the Total Crash model parameters for the years 2010 and 2016 for CT="010100"

Census Tract	**road category**	**Accident Year**	**Total Crash**	**Tot VMT**	**Cong Speed Freeway**	**Cong Speed Freeway SQR**	**Cong Speed Other roads**	**Cong Speed Other roads SQR**	**PUDO**	**Tot TNC VMT**	**Year 2016 Dummy**
010100	2	2010	17	10746.87	0	0	13.33	177.68	0	0	0
010100	2	2016	25	12175.44	0	0	13.12	172.04	1682.50	1566.595	1

Plugging respective values in Equation 5-1 yields

$$\Delta\, Total\ Crash = 17 * \left[\left(\frac{12175.44 - 10746.87}{10746.87}\right)^{1} * \left(\frac{e^{-0.636*0+0.001*(0)^2}}{e^{-0.636*0+0.001*(0)^2}}\right) * \left(\frac{e^{0.6*(13.12-13.33)\,+\,(-0.022*(13.12-13.33)^2)}}{e^{0.6*(13.33)\,+\,(-0.022*(13.33)^2)}}\right) * \left(\frac{1682.50}{1}\right)^{0.065} * \left(\frac{17042.97}{1}\right)^{-0.065} * \left(\frac{e^{0.006*(1-0)}}{e^{-0.006*0}}\right)\right] - 17$$

$$\Delta\, Total\ Crash = 18.87 - 17 = 1.87$$

5.3 Modeled Scenarios

The scenarios divide the process into five unique iterative steps, each built incrementally upon the previous scenario. An (additional) variable is introduced at each iterative step, and its impact on the overall road crash prediction is measured. The estimated coefficients in Table 5-9 are used to predict the variable's contribution to road crashes. Introducing one new variable at each iterative step helps isolate the variable's contribution to the prediction model. For each scenario, there are three possible solutions.

1. Firstly, if the predicted road crash is larger than the previous iteration's predictions, it can be said that introducing the variable leads to increased road crashes.
2. Second, if the predicted road crash is smaller than the previous iteration's predictions, the variable is said to reduce the crash occurrences.
3. Lastly, suppose there is no change in the predicted road crashes compared to the previous scenario. In that case, it can be safely said that the variable has no contribution to the crash incident occurrence.

Table 5-3 briefly overviews these five scenarios and the inputs used for each scenario. A detailed description of the construction of these five scenarios can be found in Roy's Ph.D. dissertation and the subsequent papers related to the same research dataset (G. Erhardt et al., 2019; G. D. Erhardt et al., 2022; Gregory D. Erhardt, Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione, 2019; Roy, 2019; Roy et al., 2020).

Table 5-3: Summary of Scenarios Tested to isolate the impact of each variable in the primary statistical model

Scenario	**Scenario description**	**Traffic Volumes**	**Time Effects**	**Speeds**	**TNC VMT**	**PUDO**
A	2010	2010	None	None	None	None
B1	CounterFactual (CF) 2016 - traffic volumes	CF 2016 (No TNC)	None	None	None	None
B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)	CF 2016 (No TNC)	Yes	None	None	None
B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD	CF 2016 (No TNC)	Yes	CF 2016	None	None
C	2016 - traffic volumes + TE + CF CSPD	2016 with TNC	Yes	CF 2016	None	None
D	2016 - traffic volumes + TE + CSPD	2016 with TNC	Yes	2016 with TNC	None	None
E1	2016 - traffic volumes + TE + CSPD + TNC VMT	2016 with TNC	Yes	2016 with TNC	2016 TNC VMT	None
E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO	2016 with TNC	Yes	2016 with TNC	2016 TNC VMT	2016 PUDO

Scenario A is the first scenario and the base condition for 2010. It assumes no TNC service is in operation. It is the same dataset used to represent the 2010 condition in the primary model. The dataset is a base condition against which all subsequent scenarios (B through E) will be measured and compared. The SF-CHAMP travel demand model estimates the scenario and reflects 2010 socioeconomic conditions. The scenario excludes all attributes for 2016.

Scenario B is a counterfactual (CF) scenario to actual conditions (Scenario E). It incorporates the network changes (including road and transit network), population growth, demographics, and employment changes between 2010 to 2016. As a result, the congested speeds on the freeways and the remaining road network are different from the actual observed. We split the scene into three sub scenarios B1, B2, and B3. It allowed us to understand the impact of each additional component on the road crash changes. In Scenario B1, we introduced only the counterfactual 2016 Tot VMT into the equation and measured its impact. Next, Scenario B2 includes the time-variant variables as a dummy variable year_2016_dummy. The study introduces the Counterfactual Congested Speed variable segregated by freeways and other road networks in Scenario B3. All sub-scenarios assume no TNC operations and no TNC-related PUDO activity on the road segments in 2016.

Scenario C builds upon the previous scenario (Scenario B3). It replaces Scenario B traffic volumes for 2016 from counterfactual volumes, which did not account for TNC volumes, with actual (observed) trips that include out-of-service (OOS) TNC volumes with directional links in the SF-CHAMP road network. The scenario assumes no TNC-related PUDO activity on the road segments. There are three components to the scenario setup. Deadhead or OOS TNC vehicles purely add traffic to the network. Additional traffic to the network will increase VMT, which increases the chances of road crash exposure. At the same time, increased VMT will likely introduce congestion reducing the travel speed, and as a result, it may minimize road crash occurrence. Lastly, the in-service TNC trips (those transporting passengers), which are the net result of the trips generated due to replacing the taxi or car trips, or trips attracted from sustainable modes like walking, cycling, transit, and induced demand may or may not result in an increase VMT and hence road crash exposure.

Scenario D builds upon the previous scenario and replaces the counterfactual congested speeds with observed speeds for 2016. The scenario maintains TNC-related PUDO activity on the road segments to be zero.

Like Scenario B, Scenario E consists of two sub-scenarios, E1 and E2. Scenario E1 introduces Tot TNC VMT as one of the additional predictor variables w.r.t Scenario D. Given that Tot TNC VMT is already part of Tot VMT change, crashes are considered indirect of TNC VMTs. Similarly, Scenario E2 builds upon E1 and is the final scene of the modeling simulation. Here, the disruptive effect of curbside TNC PUDO on traffic flow gets introduced. The scenario resembles the actual (dataset) ground condition for 2016.

5.3.1 Modeled Scenarios Results

Table 5-4, Table 5-5, and Table 5-6 present the Total VMT and Congested Speed for 2010, 2016 CF, and 2016. These metrics get reported at the San Francisco County level by taking weighted averages of the variables reported for all road category segments w.r.t Tot VMT.

Table 5-4: Comparison in Total VMT in the year 2010, 2016 CF and 2016

Total VMT (in miles)			
	Freeways	Other roads	Total
2010	4,227,880.55	5,805,779.34	10,033,659.89
2016 CF	4,497,834.44	6,270,141.41	10,767,975.85
2016 Observed	4,633,179.87	6,812,437.30	11,445,617.17

Table 5-5: Percentage of Total VMT by road category types

Total VMT (% distribution)		
	Freeways	Other roads
2010	42.14%	57.86%
2016 CF	41.77%	58.23%
2016 Observed	40.48%	59.52%

Table 5-6: (Weighted Average) Congested Speed in different road categories in the year 2010, 2016 CF and 2016

Congested Speed (in miles per hour)		
	Freeways	Other roads
2010	42.11	16.75
2016 CF	40.70	16.57
2016 Observed	40.20	16.25

Table 5-7 shows the modeled results for the tested scenarios for Total Crashes

Table 5-7: Iterative Scenarios for y = total crashes

Total Crashes							
Scenario	sub-scenarios	Description	Observed		Additional Crashes		Variable Contribution
			Year 2010	Year 2016	Year 2010	Year 2016	
A		2010	4430				
B	B1	CounterFactual (CF) 2016 - traffic volumes				152	152
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)				409	257
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD				1115	706
C		2016 - traffic volumes + TE + CF CSPD				1325	210
D		2016 - traffic volumes + TE + CSPD				1747	422
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT				-666	-2413
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO Effect		5316		889	1555

Figure 5-7 describes how the predicted crashes move with each additional variable added to the model structure for total crashes.

Starting from the base, Scenario A represents the 2010 crash situation. The blue column represents the observed crashes for the year.

Scenario B, i.e., Scenario B1, introduces only the counterfactual Tot VMT into the previous model equation. As a result, we see a net increase in total crashes by 152 (+3% w.r.t Scenario A). In Scenario B2, total crash counts increase to 409 due to the time-effect component. Direct attribution of the time-effect component is 409 – 152 = 257 (+6%) additional crashes compared to Scenario B1. Finally, when the counterfactual speed gets added to sub-scenario B2, the model equation predicts 1115 total crashes. It is a 15% rise (1115 – 409 = 706) in crashes w.r.t Scenario B2. The net estimated total crashes equal 4430 + 1115 = 5545 crashes. Scenario B (B1+B2+B3) predicts a 25% increase in crash count compared to base Scenario A. It is an acceptable trend, given that the only significant change from the base condition is an increase in Tot VMT and road congestion (reduction in vehicle speeds), positively correlating with road crash occurrence.

In Scenario C, the counterfactual traffic flow for 2016 gets replaced with observed traffic flows. Because Total VMT rises dramatically, it further increases the road crash numbers. The predicted total crashes for 2016 equals 4430 +1325 = 5755. The net difference between Scenario C and Scenario B is 5755 - 5545 = 210 (+4%), much higher than those predicted for the counterfactual scenario by SF-CHAMP. The further increase does not come as a surprise, given that an increase in Tot VMT will likely increase the total crashes, just as observed in Scenario B. So, we can infer that the direct impact of Tot VMT is to increase the occurrence of road crashes.

In Scenario D, we replace the counterfactual speed for 2016 with observed speed rates. The predicted crashes jump to 1747, and the total crashes for 2016 are estimated to be 4430 + 1747 = 6177. The net difference between Scenario D and Scenario C is 6177–5755 = +422 (+7%). We can conclude that the contribution of observed 2016 congested speed results in 422 additional crashes compared to Scenario C crash estimates.

Like Scenario B, we split Scenario E into sub-scenarios, E1 and E2. Scenario E1 introduces the Tot TNC VMT as an additional parameter. As explained earlier, introducing Tot TNC VMT as an independent variable in addition to the Tot VMT component allows us to capture the impact of TNC fleets on road crashes. Interestingly, the total predicted crashes dropped to 666, and the total crashes for 2016 are estimated to be 4430 – 666 = 3764. The net difference between Scenario E1 and Scenario D is 3764 – 6177 = -2413 (-39%). We can conclude that the indirect contribution of TNC VMT is negative and helps reduce overall road crashes.

Lastly, Scenario E2 represents the actual ground conditions for 2016, where all relevant variables participating in the statistical model act in tandem. It is arrived at by adding the TNC service components like PUDO to the Scenario E dataset. Interestingly, the total change in crashes flips from negative to positive. It now equals 4430 + 889 = 5319 crashes. The net difference between Scenario E2 and Scenario E1 is 5319 – (3764) = 1555 (+41%) crashes. Since TNC PUDO is the only additional variable to the previous scenario, we can safely conclude that TNC service operations, especially its PUDO activities, increase crashes w.r.t Scenario E1. Overall, Scenario E decreases crash by (889 - 1747)

858 compared to Scenario D, indicating that TNC service operations dampen the overall road crash scenario.

Compared to observed road crashes for 2016 (5316), the model predicts 4430 + 889 = 5319, which is near perfect prediction. The model scenarios can be visualized as a waterfall chart, as shown in Figure 5-7. The figure's first and last (blue) bars represent the observed crash values for 2010 and 2016, respectively, while the remaining bar represents the model scenario from A to E, including sub-scenarios. The red color of the bar indicates that the crashes have fallen compared to the previous scenario, while the green bar indicates an increase w.r.t earlier scene.

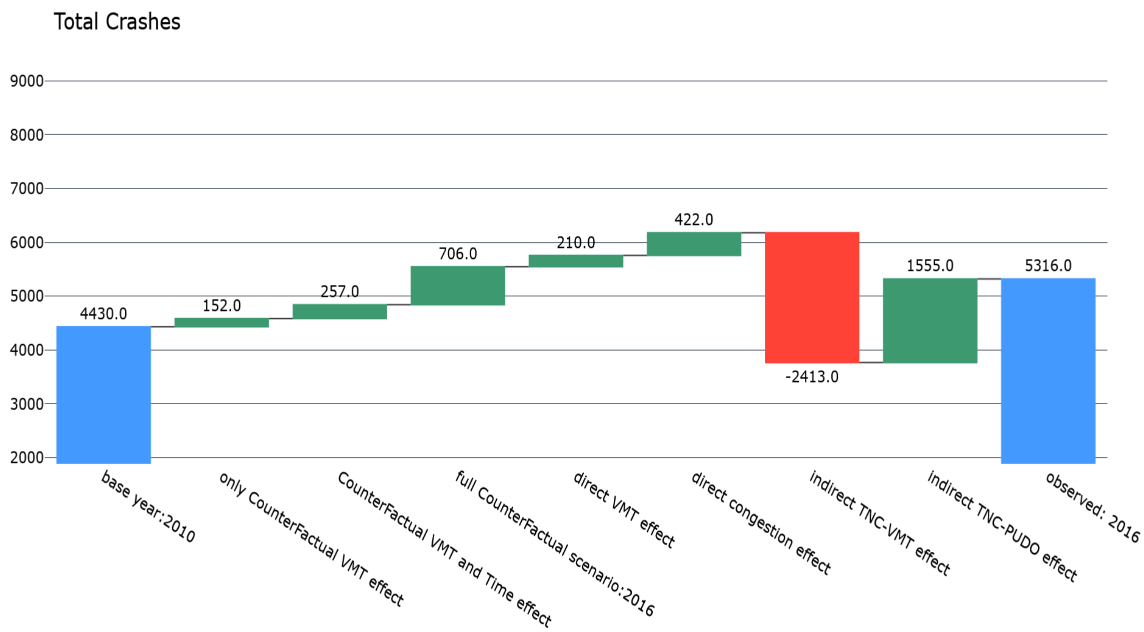

Figure 5-2: Scenario results for the dependent variable is equal to Total Crash

Labels in the chart correspond to the description presented in Table 5-8..

Table 5-8: Labels corresponding to modeled scenarios (and sub-scenarios)

Scenario	**sub-scenarios**	**Description**	**Label**
A		2010	base year: 2010
B	B1	CounterFactual (CF) 2016 - traffic volumes	only the counterfactual VMT effect
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)	counterfactual VMT and time-effect
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD	full counterfactual scenario: 2016
C		2016 - traffic volumes + TE + CF CSPD	direct VMT effect
D		2016 - traffic volumes + TE + CSPD	direct congestion effect
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT	indirect TNC-VMT effect
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO Effect	indirect TNC-PUDO effect
Observed Crashes 2016		Actual 2016 crashes	Observed 2016

The study tried to change the order of introducing sub-scenarios within B (B1, B2, B3) and E (E1 and E2) and found no particular change in the magnitude of road crashes. Introducing TNC PUDO always results in a net increase in crashes, while TNC VMT results in a net decrease.

Waterfall charts for other road crash types are presented in Figure 5-8, Figure 5-9, and Figure 5-10.

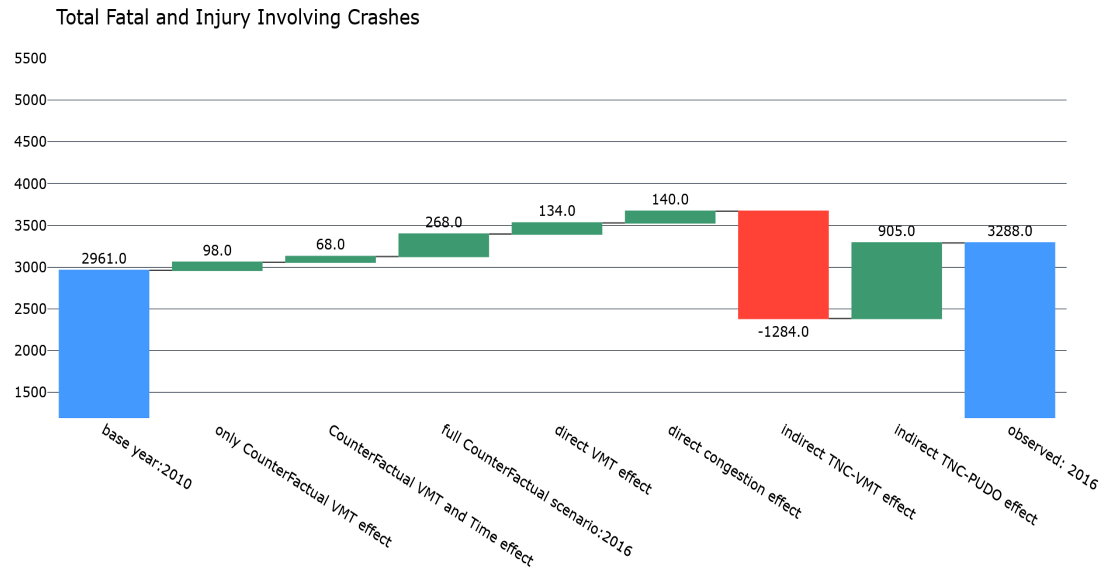

Figure 5-3: Scenario results for the dependent variable is equal to the count of all Fatal and Injury Involving Crash

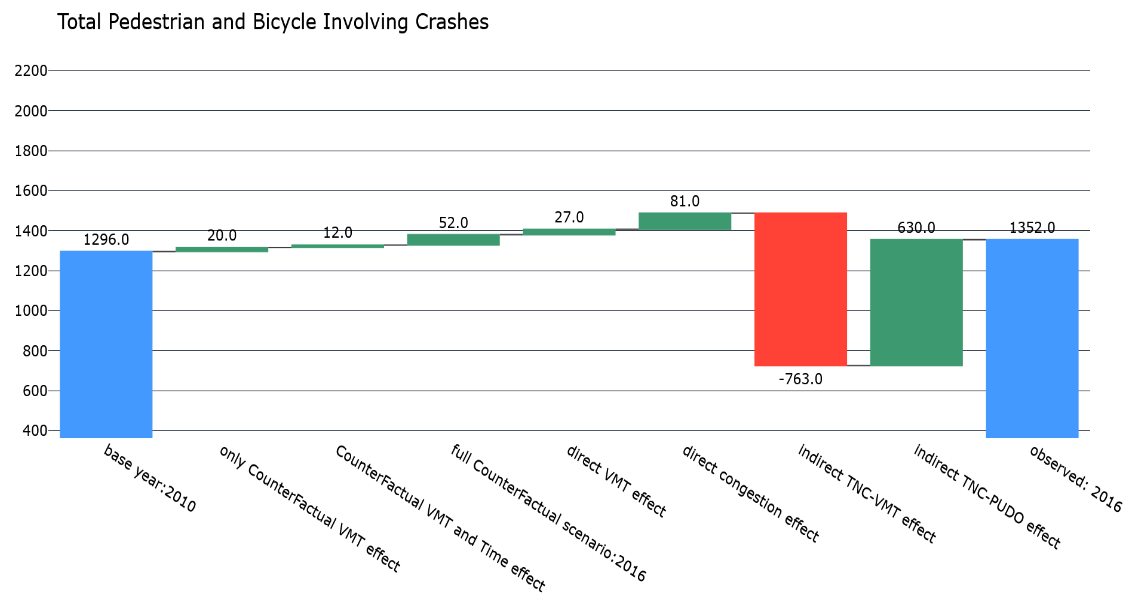

Figure 5-4: Scenario results for the dependent variable is equal to the count of all Pedestrian and Bicyclist involving crashes

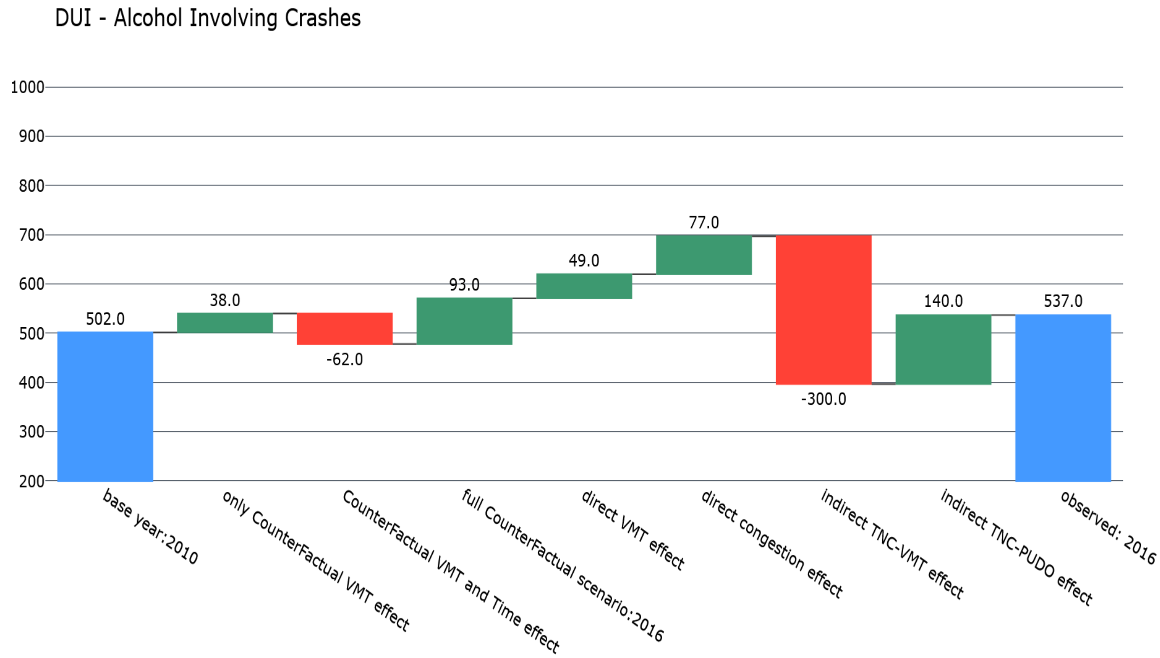

Figure 5-5: Scenario results for the dependent variable is equal to the count of all alcohol-involved crashes (DUI-Alcohol)

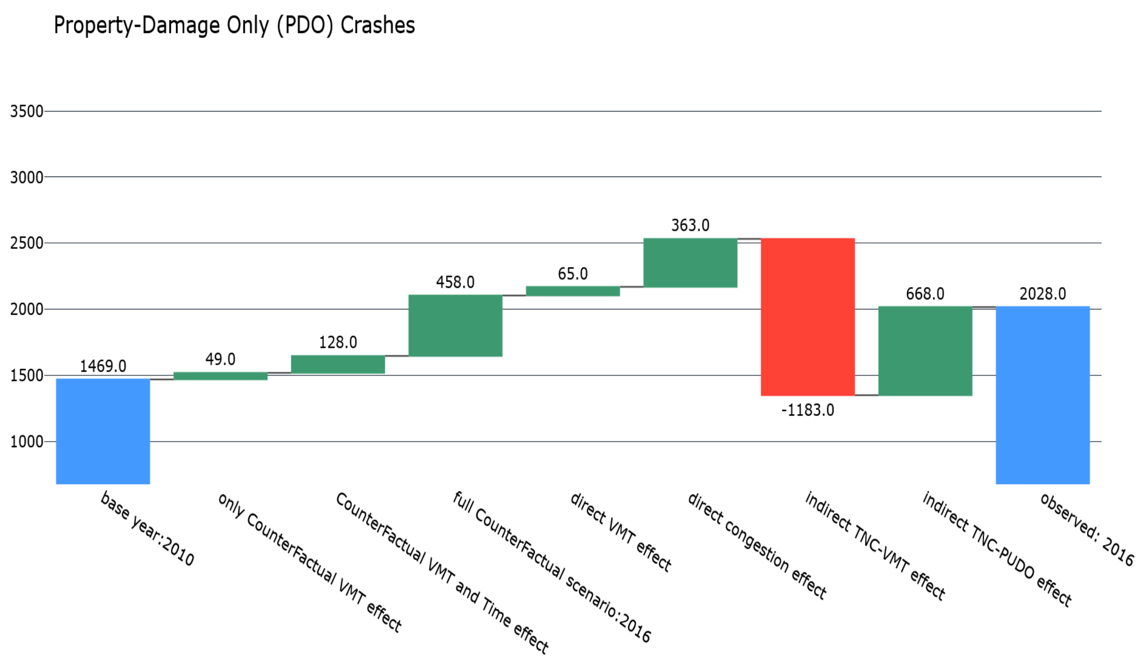

Figure 5-6: Scenario results for the dependent variable is equal to the count of all Pedestrian and Bicyclist involving crashes

Broadly, we can make several observations from these scenarios:

1. Between 2010 and 2016, VMT increases by about 7% due to population growth, employment growth, and network changes, leading to greater exposure to crashes. This increased exposure leads to more crashes through the exposure variable. Due to background (non-TNC) traffic changes, traffic speeds also change over this period, with a non-linear effect on crashes that varies by facility type. In addition, our models include some systemwide changes between 2010 and 2016 accounted for in the 2016 dummy variable (effectively a time effect). These systemwide changes vary in magnitude for the different crash types but are all statistically insignificant. Scenario B accounts for these three effects together, resulting in between 0.1% to 15% more crashes, depending on the crash type. This scenario shows what we would expect to happen in the absence of TNCs.
2. TNCs added about 7% to the total VMT in 2016. This VMT adds to the exposure, increasing crashes, as shown in Scenario C.
3. By adding traffic to the road, TNCs slow vehicle speeds. The effect of this is not necessarily straightforward. It can lead to more vehicle interactions and, therefore, more crashes, but they may be less severe due to the lower travel speeds. Our results show that the net effect of these TNC-related speed changes leads to increased crashes for all four crash types. Scenario D shows these effects. However, the speed coefficients are only statistically significant for total, fatal, and injury crashes.
4. TNCs have a direct effect on crashes through both the TNC PUDO variable that accounts for vehicle interactions during pick-up and drop-off operations, and the TNC VMT variable that accounts for differences in driving that may occur along the route, such as replacing an alcohol-impaired driver with a TNC driver. These coefficients have opposite signs and partially offset each other, but Scenario E shows that their combined effect is to reduce crashes for all crash types. However, except for total crashes, the coefficients are not statistically significant.

Overall, the results point to TNCs increasing crashes through higher exposure and speed changes, with those differences offset by the TNC direct effect. While we have sound reasons to believe that the exposure effect is accurate, we have less confidence in the estimated speed effect and TNC direct effect. For the speed effects, the estimated coefficients are significant for total crashes and fatal and injury crashes but not for bicycle and pedestrian crashes or alcohol-involved crashes. For the TNC direct effect, the estimated coefficients are only significant for one out of the eight estimated coefficients.

5.4 Additional tested statistical model and results

5.4.1 Treating Total VMT as the exposure variable

Table 5-9 displays the model results for various crash frequency models when Total VMT is considered an exposure variable. Past literature indicates that having Total VMT as an exposure variable is meaningful and practical as it measures the extent of road users' exposure to the overall level of travel risk given the road conditions each year (Qin et al., 2004; Stewart, DE, 1998). It essentially means that the coefficient of Total VMT is fixed to value 1.

The coefficients for most of the variables between the primary model results in Table 5-1 and Table 5-9 are almost identical, further reinforcing the validity of the primary model. The log-likelihood scores of all these model specifications are close to each other, with few models increasing or decreasing around the main result. It indicates that all these models are robust and simply a matter of choice based on individual preferences.

Table 5-9: Main results of the statistical modeling for identified dependent variables

Model (→)	6		7		8		9		10	
Y (→)	**Total Crashes**		**Fatal and Injury Crashes**		**Pedestrian and Bicycle crashes**		**Alcohol-involving (DUI) crashes**		**Property-damage-only (PDO)**	
X (↓)	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**
Total VMT	1 (EXPOSURE)		1 (EXPOSURE)		1 (EXPOSURE)		1 (EXPOSURE)		1 (EXPOSURE)	
Cong Spd [Freeways]	-0.636 **	0.205	-0.183	0.231	-0.76	1.99	-0.533	0.673	-0.952 *	0.428
Cong Spd [Freeways]^2	0.001	0.002	-0.002	0.002	0.002	0.022	0	0.007	0.003	0.005
(Cong Spd [Other road networks])	0.748 *	0.335	0.904 *	0.373	0.468	0.585	0.416	0.759	0.329	0.561
(Cong Spd [Other road networks])^2	-0.028 *	0.012	-0.034 *	0.014	-0.023	0.021	-0.029	0.026	-0.012	0.022
Ln (TNC Pick-up and drop-offs)	0.04	0.15	0.026	0.026	0.009	0.075	0.051	0.061	0.069	0.036
Ln (Tot TNC VMT)	-0.055 *	0.02	-0.044	0.03	-0.025	0.075	-0.073	0.074	-0.074	0.051
Year 2016 Dummy	0.006	0.022	-0.031	0.181	-0.137	0.267	-0.13	0.453	0.035	0.265
Log Likelihood	-1020.8975		-881.4961		-583.04199		-370.35815		-690.66452	
Wald Chi2 (7)	217.41		56.58		26.16		14.42		39.19	
*p < 0.05, **p < 0.01										
Estimates are from the panel Fixed Effects Poisson regression model with Robust Standard errors. All models include entity and time-fixed effects.										

Corresponding waterfall charts for Total Crashes, Fatal and Injury Crashes, Pedestrian and Bicycle Crashes, Alcohol-involved Crashes, and Property Damage Only Crashes with Total VMT as exposure variable is presented in Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10 respectively.

Figure 5-7: Scenario results for the dependent variable is equal to Total Crash (when Tot VMT is utilized as the exposure variable)

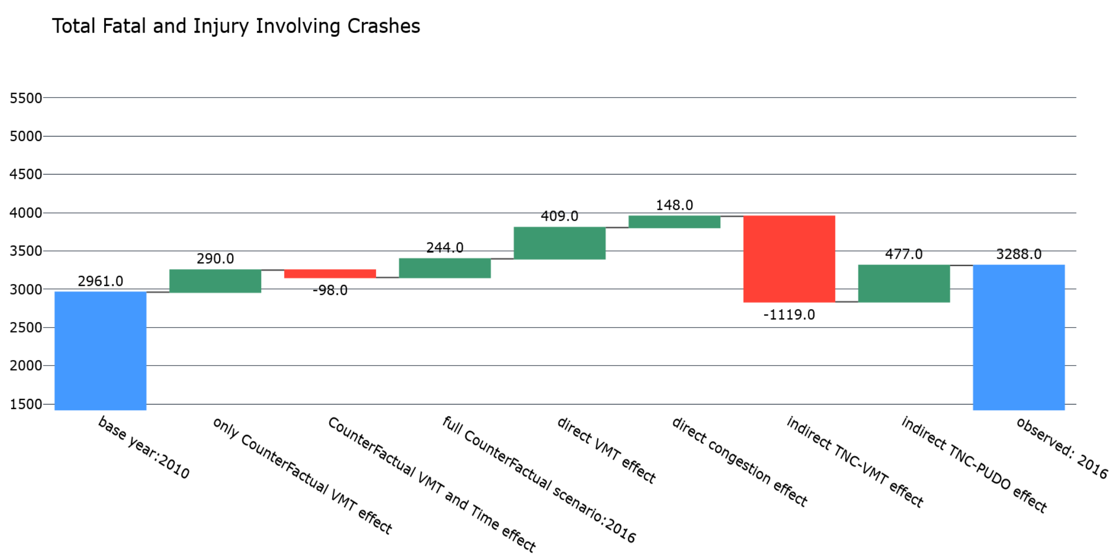

Figure 5-8: Scenario results for the dependent variable is equal to the count of all Fatal and Injury Involving Crash (when Tot VMT is utilized as the exposure variable)

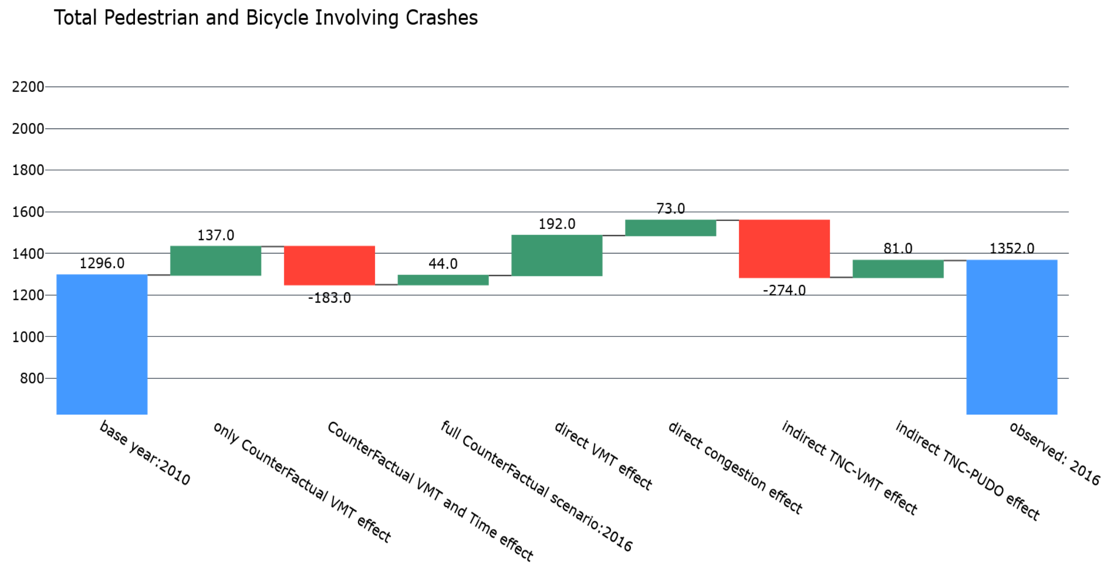

Figure 5-9: Scenario results for the dependent variable is equal to the count of all Pedestrian and Bicyclist involving crashes (when Tot VMT is utilized as the exposure variable)

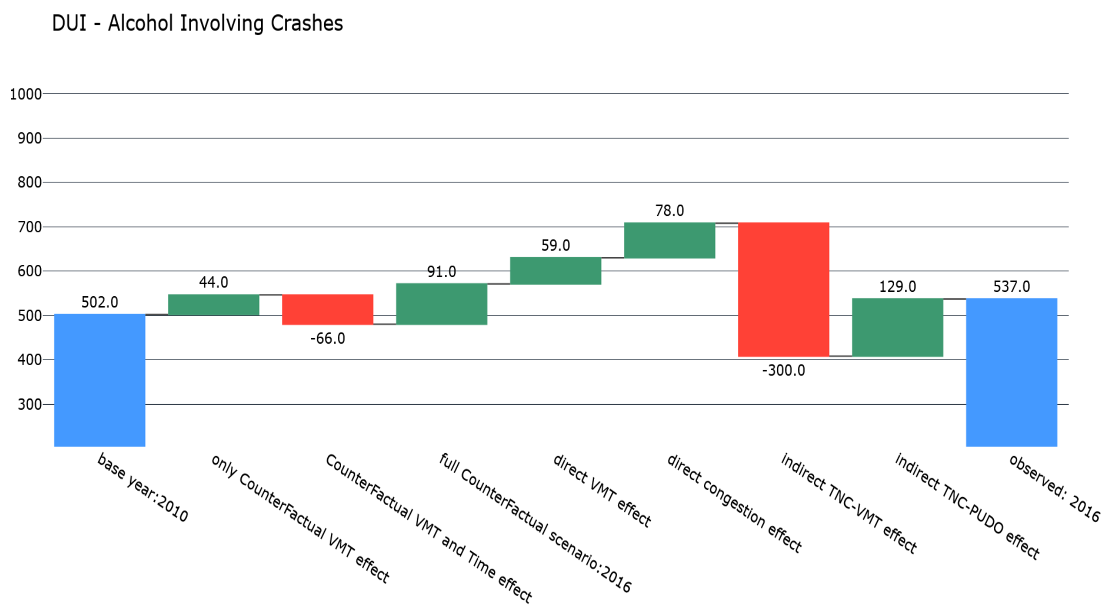

Figure 5-10: Scenario results for the dependent variable is equal to the count of all alcohol-involved crashes (DUI-Alcohol) (when Tot VMT is utilized as the exposure variable)

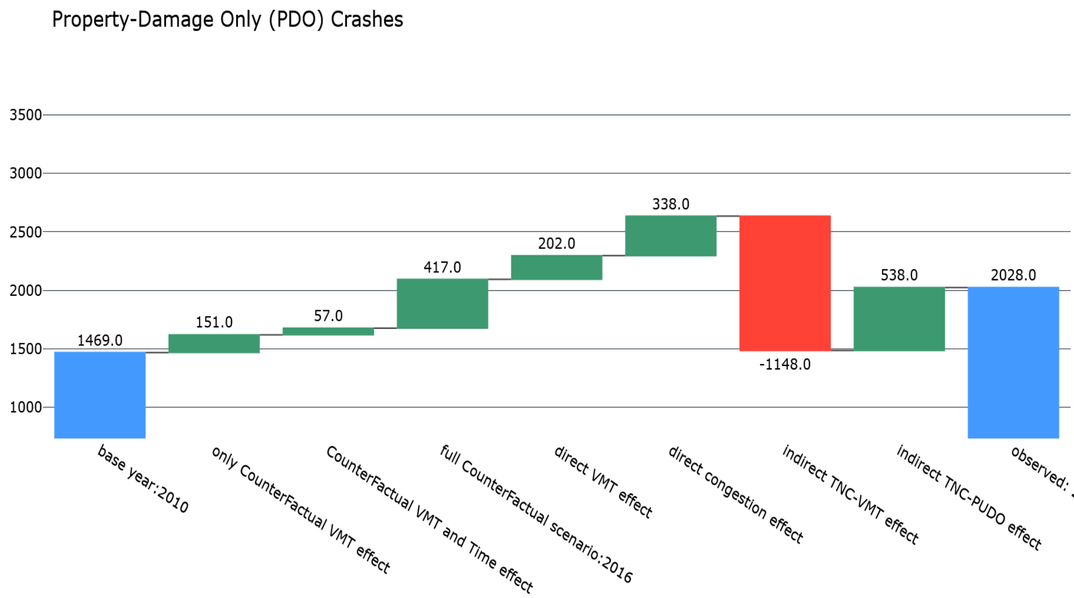

Figure 5-11: Scenario results for the dependent variable is equal to the count of all Property-Damage Only (PDO) Crashes (when Tot VMT is utilized as the exposure variable)

5.4.2 Keeping either TNC VMT or TNC PUDO to represent TNC activity in the model

One may argue that the influence of TNC VMT is already part of Total VMT and hence likely be colinear. Although collinearity does not bias the estimates, it significantly increases the standard errors and makes individual variables insignificant.

Table 5-10 provides the model's results containing the log transformation of TNC PUDO, while Table 5-11 only contains the log transformation of TNC VMT.

Table 5-10 controls the TNC PUDO and presents the model output results when Y equals Total crashes, Fatal and Injury crashes, Pedestrian and Bicyclist crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes. The TNC PUDO coefficients are generally positive but weak in absolute values and insignificant at 95% confidence, except for PDO crashes. Using the natural logarithmic of PUDO in the model form transforms its coefficient into elasticity. Therefore, the percentage increase can be easily interpreted as a 10% increase in the PUDO variable (not its log), increasing the crash occurrences by 0.4%, 0.2%, 0.1%, 0.3%, and 0.7% for the target-dependent variables. The weak relationship might be related to the limitation of the NHTSA crash database, which restricts crash entry records to those involving motor vehicles, i.e., when motor vehicles are in transport (in motion). The negative coefficient on PUDO for pedestrian and bicyclist crash occurrence might be related to the database's inability to capture crashes resulting from passengers' loading/unloading (PUDO) behavior.

On the contrary, Table 5-11 presents the results with TNC VMT only. The TNC VMT coefficients are generally negative, weak in absolute values, and insignificant at 95% confidence. The coefficients of TNC VMT are almost identical to TNC PUDO found in Table 5-10 but opposite in signs indicating very strong collinearity between them. It is evident, given that no TNC miles exist without any TNC vehicle picking up or dropping off passengers and vice-versa. Using the natural logarithmic of TNC VMT in the model form transforms its coefficient into elasticity. Therefore, the percentage increase can be easily interpreted as a 10% increase in the PUDO variable (not its log), increasing the crash occurrences by -0.3%, -0.2%, -0.1%, -0.5%, and -0.5% for the target-dependent variables. The log-likelihood is also greater in values, indicating that they are not the best models to explain the relationship. Hence, modeled scenarios are not performed.

Table 5-10: Other Model Structure I: Only PUDO included to represent TNC activity

Model (→)	**6**		**7**		**8**		**9**		**10**	
Y (→)	**Total Crashes**		**Fatal and Injury Crashes**		**Pedestrian and Bicycle crashes**		**Alcohol crashes**		**PDO**	
X (↓)	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**
Total VMT	0.42 **	0.159	0.41 *	0.176	0.24	0.265	0.91	0.522	0.39	0.25
Cong Spd [Freeways]	-0.41 *	0.201	-0.02	0.178	-0.56	1.888	-0.28	0.663	-0.64	0.401
Cong Spd [Freeways]^2	0	0.002	0	0.002	0	0.022	0	0.008	0	0.005
(Cong Spd [Other road network])	0.53	0.308	0.7 *	0.345	0.22	0.466	0.32	0.722	0.07	0.539
(Cong Spd [Other road network])^2	-0.02	0.011	-0.02	0.013	-0.01	0.017	-0.02	0.025	0	0.02
Ln (TNC Pick-up and drop-offs)	0.04	0.019	0.02	0.022	0.01	0.037	0.03	0.057	0.07 *	0.032
Year 2016 Dummy	-0.21	0.113	-0.19	0.133	-0.14	0.239	-0.46	0.336	-0.3	0.182
Log likelihood	-1016.19		-877.235		-578.6131		-371.023		-690.162	
Wald Chi2 (6)	154.75		67.95		6		13.09		65.93	
*p < 0.05, **p < 0.01										
Estimates from panel Fixed Effects Poisson regression model with Robust Standard errors. All models include entity and time-fixed effects.										

Table 5-11: Other Model Structure II: Only TNC VMT included to represent TNC activity

Model (→)	11		12		13		14		15	
Y (→)	**Total Crashes**		**Fatal and Injury Crashes**		**Pedestrian and Bicycle crashes**		**Alcohol-involving (DUI) crashes**		**Property-damage-only (PDO)**	
X (↓)	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**	**Coefficient**	**robust std. err**
Total VMT	0.51 **	0.155	0.47 **	0.169	0.26	0.264	1.01 *	0.507	0.57 *	0.235
Cong Spd [Freeways]	-0.35	0.202	0.01	0.172	-0.51	1.845	-0.3	0.659	-0.51	0.399
Cong Spd [Freeways]^2	0	0.002	0	0.002	0	0.021	0	0.008	0	0.005
(Cong Spd [Other road network])	0.43	0.349	0.62	0.372	0.19	0.476	0.26	0.764	-0.02	0.6
(Cong Spd [Other road network])^2	-0.02	0.013	-0.02	0.014	-0.01	0.018	-0.03	0.026	0	0.023
Ln (Tot TNC VMT)	-0.03	0.022	-0.02	0.025	-0.01	0.038	-0.05	0.064	-0.05	0.048
Year 2016 Dummy	0.18	0.147	0.1	0.172	0.02	0.266	0.01	0.438	0.33	0.311
Log Likelihood	-1016.98		-877.37		-578.59		-370.80		-691.78	
Wald Chi2 (7)	169.21		60.57		5.94		11.79		74.69	
*p < 0.05, **p < 0.01										
Estimates are from the panel Fixed Effects Poisson regression model with Robust Standard errors. All models include entity and time-fixed effects.										

5.4.3 Segregating road segments into five main road facility types

Alternatively, it may be the case that TNC operations may be more sensitive to one facility type than others, such that its road safety contribution (positive or negative) may be more severe than others. In order to test this claim, the study reclassifies the road network into five classes: Freeways, Major Arterial, Minor Arterial, Collectors, and Locals, as summarized in Table 5-12. The purpose of reclassification is to understand and identify if the impact of TNC service on road safety varies by facility type.

Table 5-12: reclassified facility types

(reclassified) Facility Type (FT)	**FT code**	**Description**
Freeways	1,2,3,5	Roads classified as Fwy-Fwy Connector or Freeway or Expressway or Ramps
Major Arterials	7, 15	Major Arterial and Super Arterial roads
Minor Arterials	12	Minor Arterial
Collectors	4	Collector roads
Locals	11	All local roads

We ran these specifications in three ways

a) split each participating variable, i.e., Tot VMT, Congested Speed, TNC-VMT, and TNC-PUDO, into facility type as specified in Table 5-12.

For all other crash types, the TNC PUDO is positive, while TNC VMT is negative. Except for freeways where no TNC PUDO is reported (and therefore has a coefficient value of zero), the coefficients of TNC PUDO and TNC VMT are almost identical in absolute values. The significance also follows a similar trend, with Collector road variables being extraordinarily significant and having the biggest coefficient, followed by Major Arterial for all road crash types. TNC PUDO and TNC VMT are significant in fatal and injury crashes on minor arterial roads. Except for congested freeway speed, all other category speed values and square terms are significant at 95% confidence. When these variable coefficients are used to predict the crashes, as explained in section 5.3, the predicted values are close to observed values, indicating robustness.

Table 5-13: Estimated model results when the road network is segregated into five major road facility types

Models (→) **Independent Variables (↓)**	**Total Crashes**		**Count Fatal and Injury**		**Count Pedestrian and Bicyclists**		**Alcohol involved crashes**	
Total VMT (Freeways)	-0.443	1.187	-0.88	1.513	27.977	14.422	1.439	6.624
Total VMT (Major Arterial)	0.615	0.393	0.896	0.47	0.937	0.661	1.732	1.038
Total VMT (Minor Arterial)	0.285	0.338	0.119	0.345	0.854	0.689	0.908	1.481
Total VMT (Collectors)	0.08	0.31	-0.065	0.356	-0.112	0.508	-0.644	1.097
Total VMT (Locals)	0.35	0.286	0.074	0.337	-0.028	0.455	2.146 *	0.914
Cong Spd [Freeways]	-0.447	0.251	0.149	0.254	2.583	1.722	-0.613	0.84
Cong Spd [Freeways]2	-0.001	0.003	-0.005	0.003	0.002	0.018	0.001	0.011
Cong Spd [Major Arterial]	1.079 *	0.416	1.436 **	0.448	1.612 **	0.556	1.639	1.025
Cong Spd [Major Arterial]2	-0.034 *	0.014	-0.046 **	0.016	-0.051 *	0.02	-0.063	0.038
Cong Spd [Minor Arterial]	1.591 *	0.761	2.095 *	0.831	1.63	0.842	-0.87	2.312
Cong Spd [Minor Arterial]2	-0.063 *	0.027	-0.083 **	0.031	-0.066 *	0.029	0.033	0.086
Cong Spd [Collectors]	4.312 *	1.685	7.445 ***	1.78	5.044 *	2.53	-7.028	6.464
Cong Spd [Collectors]2	-0.175 *	0.068	-0.299 ***	0.073	-0.201 *	0.1	0.225	0.227
Cong Spd [Locals]	-0.043	1.347	1.224	1.929	3.966	2.915	-6.778	4.538
Cong Spd [Locals]2	0.012	0.06	-0.046	0.086	-0.175	0.131	0.319	0.203
Ln (TNC Pick-up and drop-offs) [Freeways]	0	0	0	0	0	0	0	0
Ln (TNC Pick-up and drop-offs) [Major Arterial]	0.276 **	0.094	0.306 **	0.099	0.486 **	0.145	0.267	0.246
Ln (TNC Pick-up and drop-offs) [Minor Arterial]	0.147	0.109	0.274 *	0.124	0.16	0.149	0.23	0.295
Ln (TNC Pick-up and drop-offs) [Collectors]	0.511 ***	0.12	0.653 ***	0.16	0.597 **	0.203	0.414	0.437
Ln (TNC Pick-up and drop-offs) [Locals]	0	0.097	0.216	0.143	0.195	0.188	-0.365	0.413
Ln (Tot TNC VMT) [Freeways]	-0.047	0.026	-0.007	0.033	0.273	0.16	-0.11	0.09
Ln (Tot TNC VMT) [Major Arterial]	-0.251 **	0.079	-0.283 **	0.087	-0.441 **	0.138	-0.289	0.228
Ln (Tot TNC VMT) [Minor Arterial]	-0.146	0.107	-0.263 *	0.123	-0.172	0.155	-0.236	0.281
Ln (Tot TNC VMT) [Collectors]	-0.5 ***	0.112	-0.638 ***	0.151	-0.591 **	0.199	-0.409	0.432
Ln (Tot TNC VMT) [Locals]	-0.001	0.085	-0.2	0.128	-0.196	0.164	0.264	0.374
Year 2016 Dummy	0.117	0.162	0.088	0.199	0.074	0.323	0.15	0.546
Log Likelihood	-1134.2432		-968.3234		-625.731		-395.984	
Wald Chi2 (7)	385.83		144.98		62.65		25.59	
*p < 0.05, **p < 0.01, ***p<0.001								
Estimates from panel Fixed Effects Poisson regression model with Robust Standard errors. All models include entity and time-fixed effects.								

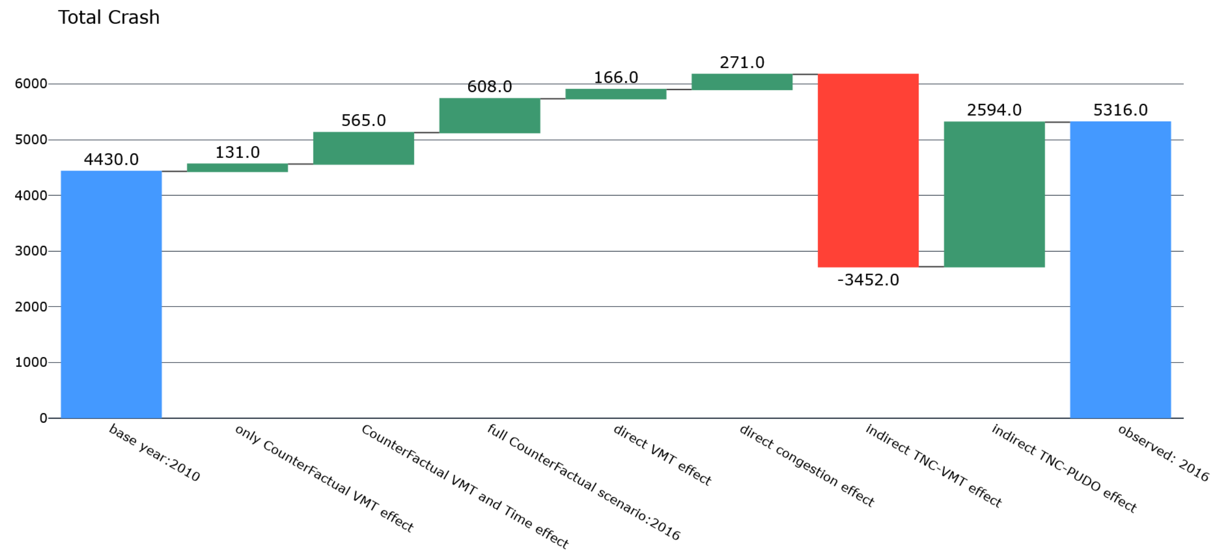

Figure 5-12: Modeled Scenario results for Total Crash using coefficients as presented in Table 5-13

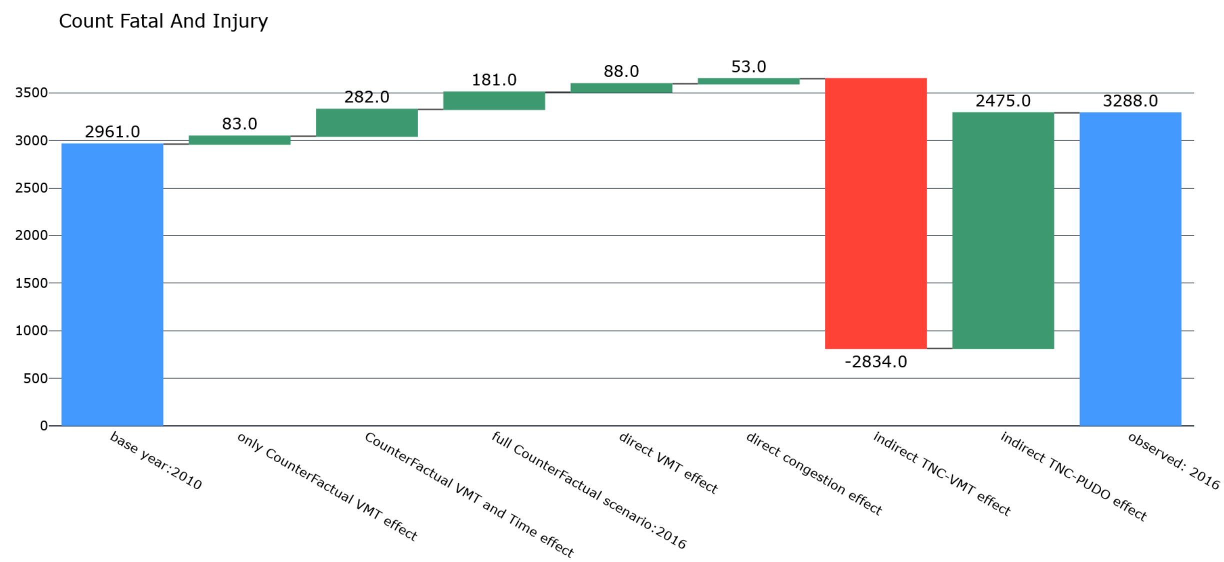

Figure 5-13: Modeled Scenario results for Fatal and Injury Crashes using coefficients as presented in Table 5-13

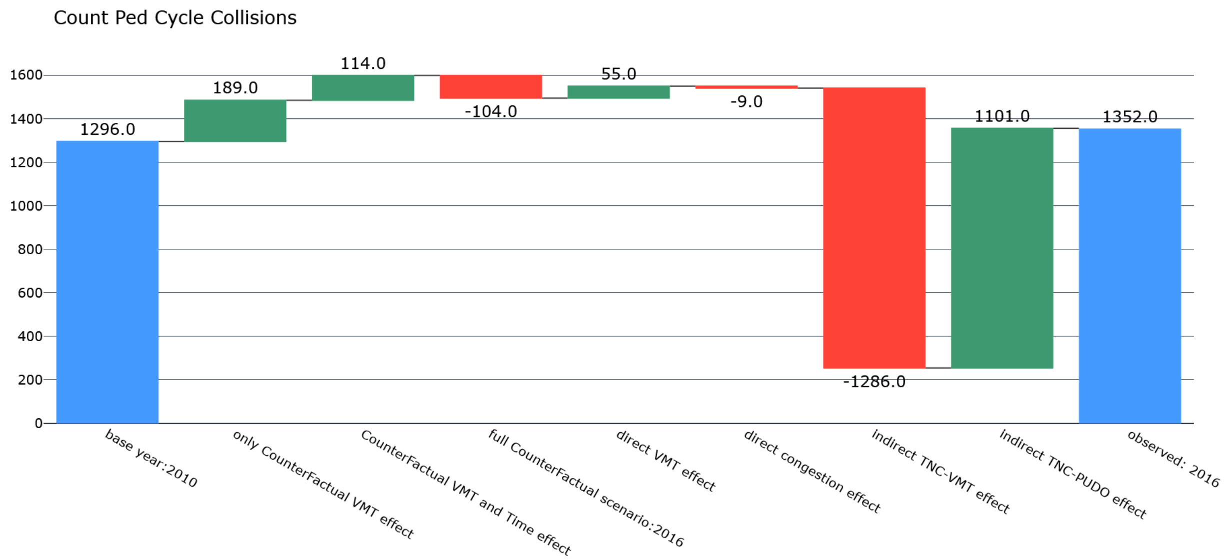

Figure 5-14: Modeled Scenario results for Pedestrian and Bicyclist Crashes using coefficients as presented in Table 5-13

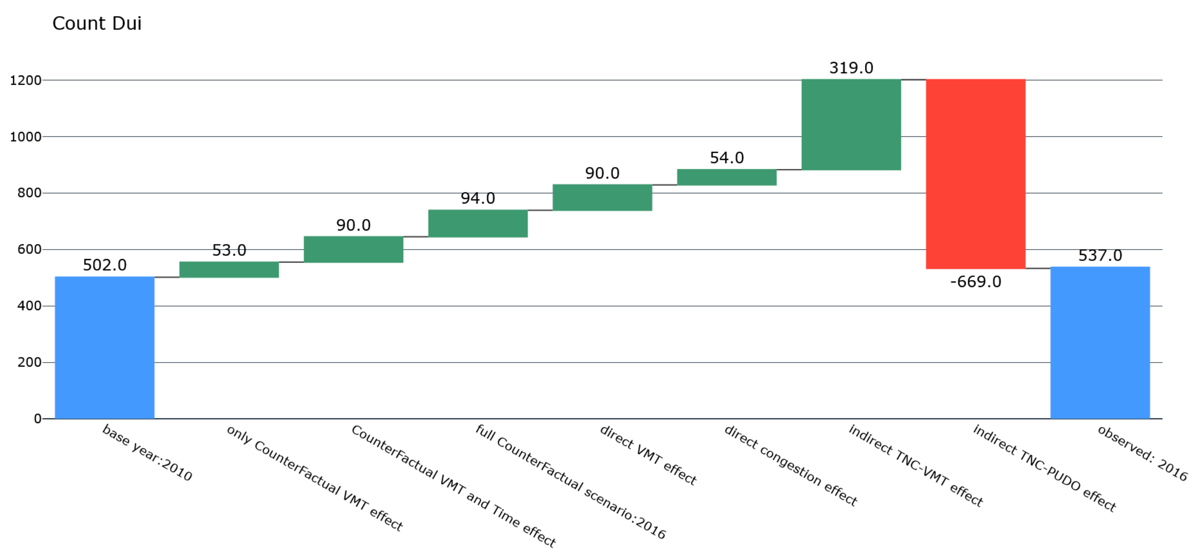

Figure 5-15: Modeled Scenario results for Alcohol involving crashes using coefficients as presented in Table 5-13

b) drop all attributes belonging to freeway facility type as they contributed the least number of road network miles and crashes.

Models (→) / Independent Variables (↓)	Total Crashes		Count Fatal and Injury		Count Pedestrian and Bicyclists		Alcohol involved crashes	
Total VMT (Major Arterial)	0.615	0.394	0.898	0.47	0.938	0.661	1.741	1.039
Total VMT (Minor Arterial)	0.291	0.337	0.111	0.347	0.847	0.69	0.895	1.48
Total VMT (Collectors)	0.076	0.309	-0.059	0.358	-0.109	0.509	-0.634	1.1
Total VMT (Locals)	0.345	0.286	0.081	0.337	-0.023	0.455	2.158 *	0.915
Cong Spd [Major Arterial]	1.068 *	0.415	1.449 **	0.449	1.627 **	0.557	1.665	1.027
Cong Spd [Major Arterial]^2	-0.033 *	0.014	-0.047 **	0.016	-0.051 **	0.02	-0.064	0.038
Cong Spd [Minor Arterial]	1.62 *	0.757	2.056 *	0.836	1.602	0.845	-0.937	2.304
Cong Spd [Minor Arterial]^2	-0.064 *	0.027	-0.082 **	0.031	-0.065 *	0.029	0.035	0.086
Cong Spd [Collectors]	4.306 *	1.684	7.461 ***	1.782	5.046 *	2.53	-7.117	6.489
Cong Spd [Collectors]^2	-0.175 *	0.068	-0.299 ***	0.074	-0.2 *	0.1	0.229	0.227
Cong Spd [Locals]	-0.047	1.354	1.233	1.926	3.98	2.916	-6.776	4.529
Cong Spd [Locals]^2	0.012	0.06	-0.047	0.086	-0.175	0.131	0.319	0.203
Ln (TNC Pick-up and drop-offs) [Major Arterial]	0.272 **	0.094	0.312 **	0.099	0.491 **	0.145	0.275	0.245
Ln (TNC Pick-up and drop-offs) [Minor Arterial]	0.146	0.109	0.274 *	0.125	0.16	0.149	0.23	0.295
Ln (TNC Pick-up and drop-offs) [Collectors]	0.509 ***	0.119	0.656 ***	0.16	0.597 **	0.203	0.416	0.437
Ln (TNC Pick-up and drop-offs) [Locals]	0.007	0.098	0.206	0.143	0.187	0.188	-0.38	0.413
Ln (Tot TNC VMT) [Major Arterial]	-0.242 **	0.08	-0.295 **	0.088	-0.451 **	0.138	-0.306	0.227
Ln (Tot TNC VMT) [Minor Arterial]	-0.14	0.107	-0.27 *	0.125	-0.177	0.155	-0.248	0.281
Ln (Tot TNC VMT) [Collectors]	-0.492 ***	0.113	-0.648 ***	0.151	-0.597 **	0.199	-0.424	0.433
Ln (Tot TNC VMT) [Locals]	-0.002	0.085	-0.198	0.128	-0.194	0.164	0.265	0.374
Year 2016 Dummy	0.08	0.174	0.14	0.213	0.111	0.323	0.229	0.544
Log Likelihood	-1061.2704		-909.18877		-617.0016		-360.361	
Wald Chi2 (7)	103.08		99.35		52.38		21.74	
*p < 0.05, **p < 0.01, ***p<0.001								
Estimates from panel Fixed Effects Poisson regression model with Robust Standard errors. All models include entity and time-fixed effects.								

Even after dropping the freeway road category from the analysis, the coefficients value of TNC PUDO, TNC VMT, and their significance largely follow the trends observed in Table 5-13.

c) For each facility type, we tested four model structures, namely i) non-linear speed (in addition to TNC VMT and TNC PUDO), ii) non-linear speed and only TNC PUDO, and iii) non-linear speed and only TNC VMT.

Unfortunately, few co-efficient values of the participating regressing variables were absurdly large in each model setup, prompting us not to utilize the results for model scenarios and predictions. The details of each model, including coefficients and standard errors of the regressing variables, log-likelihoods, and Chi-Square scores, are presented in Appendix A6.

For each model specification, we provide log-likelihood values, which measure how well a particular model fits the data. Akaike's Information Criterion can utilize the log-likelihood values to choose the best model from the set. The basic AIC formula $AIC = -2(log - likelihood) + 2K$.

5.5 SUMMARY

The chapter provides an after-implementation summary of the statistical model presented in CHAPTER 3. It does this for all dependent variables like Total Crashes, Fatal and Injury Crashes, Pedestrian and Bicyclist involving crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes.

For the total crash as the dependent variable, all the identified independent variables' coefficients are presented in the tabular format in Table 5-1. Following are the observations for each independent variable.

For the dependent variable, Total Crashes, the Total VMT holds a positive coefficient of 0.37. Given that its coefficient is not equal to one suggests that road crashes do not have a linear (direct) relationship with VMT and hence cannot be used as an exposure variable. The congested speed freeway holds a negative value of -.636, while its square term has a negative coefficient value of +0.001. On the contrary, the congested speed on the rest of the road network has a positive coefficient value of 0.748, while its square term has a negative coefficient of -0.028. The congested speed on the rest of the road network and its square term are significant at a 95% confidence interval. It indicates that as the congestion on the other road segments increases, the congested speed decreases (i.e., average speed further goes down), which leads to a greater probability of witnessing an increase in road crash occurrence. A big negative coefficient term and almost negligible positive value for the square terms indicate that rate of decrease of road crashes has a negative linear slope. In other words, the rate of road crash occurrence increases non-linearly as the difference in congested speeds increases. Except for total crashes and PDO, the linear term of congested speed freeway is insignificant for other road crash types at a 95% confidence interval. The congested speed on the rest of the road network and its square term are significant at a 95% confidence interval only for fatal and injury crashes.

The models also indicate that the coefficient of TNC PUDO is positive, weak in absolute values, and significant at 95% confidence for total crashes and PDO crashes. The natural logarithmic of PUDO in the model transforms its coefficient into elasticity and, therefore, can be easily interpreted as a percentage increase or decrease. The study finds that a 10% increase in the PUDO variable (not its log) increases the crash occurrences by 0.65%, 0.55%, 0.92%, 0.56%, and 0.93% for Total Crashes, Only Fatal and Injury Crashes, Only Pedestrian and Bicyclist involving crashes, alcohol-involving crashes, and property damage only crashes respectively. One primary reason for observing such a weak relationship might be how crashes are registered inside the NHTSA crash database. NHTSA restricts crash entry records to only those involving moving motor vehicles, i.e., where motor vehicles are in transport (in motion). It may also suggest that pick-up and drop-off activity has little effect on crash outcomes.

In contrast, coefficients for Tot TNC VMT are negative and only significant for total crashes. To a more considerable extent, the coefficients of TNC VMT offset the coefficient of PUDO, such that the net effect of TNC operations cannot be accurately estimated. Similar to the PUDO, the natural logarithmic of Tot TNC VMT allows us to interpret the coefficient as elasticity. Therefore, a 10% increase in the Tot TNC VMT variable (not its log) decreases the crash occurrences by 0.65%, 0.57%, 0.95%, 0.74%, and

0.81% for Total Crashes, Only Fatal and Injury Crashes, Only Pedestrian and Bicyclist involving crashes, alcohol-involving crashes, and PDO, respectively.

The time-variant effect through 2016 dummy coefficients varies in signs and magnitude but is insignificant for all explained variables at a 95% confidence interval, indicating that its impact is marginal.

To understand and estimate the contribution of each independent variable in road crashes, the study divides the process into five unique iterative steps, each built incrementally upon the previous scenario (shown in Table 5-3).

Scenario A is the first scenario and the base condition for 2010. It assumes no TNC service is operating and reflects the 2010 condition in the primary model.

Scenario B is a counterfactual scenario to actual conditions (Scenario E) and shows what we expect to happen without TNCs. This counterfactual scenario differs from the actual condition as it considers the network changes (including road and transit networks), population growth, demographics, and employment changes between 2010 and 2016. We split the scene into three sub scenarios B1, B2, and B3. It allowed us to understand the impact of each additional component on the road crash changes. In Scenario B1, we introduced only the counterfactual 2016 Tot VMT into the equation and measured its impact. Next, Scenario B2 includes the time-variant variables as a dummy variable year_2016_dummy. The study introduces the Counterfactual Congested Speed variable segregated by freeways and other road networks in Scenario B3. All sub-scenarios assume no TNC operations and no TNC-related PUDO activity on the road segments in 2016..

Scenario C builds upon the previous scenario. It replaces Scenario B traffic volumes for 2016 from counterfactual volumes, which did not account for TNC volumes, with actual (observed) trips. Because VMT increased by 14% between 2010 and 2016, the crashes also rose sharply.

For Scenarios B and C, the rise is along the expected lines, i.e., more exposure (VMT) results in increased road risks, thereby increasing the possibility of getting involved in a road crash incident.

Scenario D builds upon the previous scenario and replaces the counterfactual congested speeds with observed speeds for 2016. The scenario assumes the impact of TNC-related PUDO activity and TNC-related VMT variable to zero. Our results show that the net effect of these TNC-related speed changes leads to increased crashes for all four crash types. Scenario D shows these effects. However, the speed coefficients are only statistically significant for total, fatal, and injury crashes.

Like Scenario B, Scenario E consists of two sub-scenarios, E1 and E2. Scenario E1 introduces Tot TNC VMT as one of the additional predictor variables w.r.t Scenario D. Tot TNC VMT gets introduced as one of the explained variables instead of being part of Tot VMT, in addition to the already accounted factors like TNC Volumes and congestion observed due to TNC service operation on the SF road network. At the same time, E2 builds upon E1 and is the final scene of the modeling simulation. In E2, the disruptive effect of curbside TNC PUDO on traffic flow gets introduced. The coefficients of these variables have opposite signs and partially offset each other. Overall, Scenario E shows that TNC-specific variables jointly reduce crashes for all crash types; however, except for total crashes, the coefficients are not statistically significant.

The study also tested the different model structures a) by varying the regressing variables, especially keeping either TNC PUDO or TNC VMT to represent TNC activity, b) by re-categorizing road network into five major road facility types, namely Freeways, Major Arterial, and Minor Arterial, Collectors, and Locals roads. While the predicted crashes for the latter were no better when compared to the primary model, the prior model specification could not better the results.

Overall, the trends across all these models is found to be similar. TNC activities increase crashes through higher exposure and speed changes, with those differences directly offsetting the TNC's direct effect.

CHAPTER 6 DISCUSSION AND CONCLUSIONS

6.1 Summary

In little more than a decade of operations, Transportation Network Companies (TNC) have changed the urban mobility landscape in the USA. As TNCs emerged in the USA, there is an interest in understanding their effects on urban transportation. One such area where TNCs could potentially change the expected outcomes if not deployed effectively is the area related to road safety risk and road crashes in general.

The dissertation aims to measure the effect of TNC operations on road safety outcomes using San Francisco (SF) county as a case study. The area was selected because TNC services had continuously operated on SF streets since May 2010, when Uber first started offering such trips to the world. It is also the headquarters of Uber and Lyft, two leading TNC service providers, covering a market share greater than 95%.

The study compares 2010 safety outcomes when TNCs were negligible in the SF area to safety outcomes for the exact locations in 2016 for which San Francisco County Transportation Authority (SFCTA) has successfully collected data, including TNC activity. We evaluate these outcomes for four types of crashes: total crashes, fatal and severe injury crashes, crashes involving pedestrians and bicyclists, and crashes involving alcohol.

We control for the changes occurring at the network level in vehicle miles traveled, vehicle speed, and road network changes. TNC service-related activities like pick-up and drop-offs (PUDO) and TNC Vehicle Miles Travelled (Tot TNC VMT) are also considered. Fixed-effects Poisson Regression model is used to model the relationship between identified crash types and explanatory variables mentioned before. Previous studies have shown that the Fixed-effects Poisson Regression model estimates are robust to overdispersion. The dissertation reveals that a 10% increase in PUDO activity increases the crash occurrences by 0.65%, 0.55%, 0.92%, 0.56%, and 0.93% for Total Crashes, Fatal and Injury Only Crashes, Pedestrian and Bicyclist only involving crashes, alcohol-involving (DUI) crashes and property damage only (PDO) crashes. However, the association between PUDO and respective crash types is not statistically significant at a

95% confidence interval except for Total Crashes and Property-damage only (PDO) Crashes. Similarly, a 10% increase in the Tot TNC VMT activity decreases the crash occurrences by 0.65%, 0.57%, 0.95%, 0.74%, and 0.81% for Total Crashes, Fatal and Injury Only Crashes, Pedestrian and Bicyclist only involving crashes, alcohol-involving (DUI) crashes, and property damage only (PDO) crashes. Except for Total Crashes incidents, the Tot TNC VMT shows a statistically significant association with Total Crashes type only at a 95% confidence interval.

To further decompose the extent to which the explanatory variables contribute to change in road crashes observed in San Francisco between 2010 and 2016, we split the analytical framework into five unique iterative steps termed "Scenarios." Each scenario is built incrementally on the prior with a series of control brought to the changes in vehicle miles traveled, vehicle speed, and road network that occur over time. This hypothetical scenario generation and estimate of the road crash exercise allowed us to pin the expected variable contribution. An overview of the counterfactual scenarios is shown in Table 6-1 below.

Table 6-1 - Counterfactual scenarios used to decompose explanatory variable's contribution to road crashes

Scenario	**Sub scenario**	**Scenario description**	**Traffic Volumes**	**Time Effects**	**Speeds**	**TNC VMT**	**PUDO**
A		2010	2010	None	None	None	None
B	B1	CounterFactual (CF) 2016 - traffic volumes	CF 2016 (No TNC)	None	None	None	None
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)	CF 2016 (No TNC)	Yes	None	None	None
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD	CF 2016 (No TNC)	Yes	CF 2016	None	None
C		2016 - traffic volumes + TE + CF CSPD	2016 with TNC	Yes	CF 2016	None	None
D		2016 - traffic volumes + TE + CSPD	2016 with TNC	Yes	2016 with TNC	None	None
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT	2016 with TNC	Yes	2016 with TNC	2016 TNC VMT	None
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO	2016 with TNC	Yes	2016 with TNC	2016 TNC VMT	2016 PUDO

6.2 Findings

Our analysis shows that TNCs may have both a direct and an indirect effect on crashes.

The TNC direct effect is measured in two ways. First, curbside pick-ups and drop-offs may increase vehicle-vehicle, vehicle-pedestrian, and vehicle-bicyclist conflicts, especially on busy arterials, leading to more crashes. Our estimated coefficients on TNC pick-ups and drop-offs are positive, consistent with this hypothesis but not statistically significant. Conversely, TNCs may reduce crashes along the route traveled by replacing higher-risk VMT with lower-risk VMT, either because the drivers are more experienced or because they replace trips that someone under the influence of alcohol would otherwise drive. Our estimate coefficients on TNC VMT are negative, consistent with the lower-risk replacement hypothesis but only statistically significant for total crashes. The scenarios show that the combined effect of both is to reduce crashes across all crash types, but due to their insignificance, we have little confidence in this result.

TNCs may also affect crashes indirectly, either through a change in exposure or through their effect on vehicle speeds.

In this analysis, we use VMT as the measure of exposure and assume that crashes are proportional to exposure, all else being equal. Between 2010 and 2016, VMT in San Francisco increased by 14%. Previous research showed that about half of this increase is due to population growth, employment growth, and network changes, while the other half is due to the introduction of TNCs (G. Erhardt et al., 2019; Roy et al., 2020). Therefore, it is reasonable to expect that TNCs increased crashes by about 7% over this period due to higher exposure.

Our analysis shows that the effect of vehicle speeds on crashes is non-linear and different by facility type. We measured congested speeds using the volume-delay functions in the SF-CHAMP travel demand model five times a day. For each type of crash, we estimated coefficients on the speed and the square of speed separately for freeways versus other facility types. On freeways, the estimated speed coefficients are negative, and the coefficients on speed squared are positive with a value close to zero, suggesting that higher speeds lead to fewer crashes. The only significant estimate is for total crashes. This result may be because when speeds are high on freeways, there are fewer vehicle interactions and, therefore, fewer crashes.

Conversely, on arterials, collectors, and locals, the estimated speed coefficients are positive, and the squared speed coefficients are negative and higher in magnitude, indicating a non-linear relationship. The coefficients are significant for fatal and injury crashes, which also is the highest magnitude among others crashes. The result suggests that more congestion may reduce the severity of crashes by limiting vehicle speeds or that more congestion can lead to more crashes through increased conflicts. The effect varies depending on the operating conditions. By adding traffic to the road, TNCs slow vehicle speeds through more congestion. When we apply the models in our scenario testing, our analysis suggests that this leads to more crashes for all types. We have some confidence in this result for total crashes and fatal and injury crashes but not for pedestrian and bicyclist crashes, alcohol-involved crashes, or property damage-only crashes

Revisiting the objectives of this research as presented in CHAPTER 1, we sought to measure whether the emergence of TNCs is associated with a decrease or increase in four types of crashes. We find that between 2010 and 2016 in San Francisco:

- TNCs indirectly increased total crashes by 4% due to higher exposure and 7% due to changes in vehicle speeds. The direct effect of TNCs on crashes offsets these increases, reducing crashes by 14%, but this effect depends upon the model specification and is insignificant in other specifications tested.
- TNCs indirectly increased fatal and severe injury crashes by 3% due to higher exposure and 3% due to changes in vehicle speeds. We do not detect a statistically significant direct effect of TNCs.
- For Pedestrians and Bicyclists involving crashes, we do not detect a significant indirect effect due to higher exposure or any changes in vehicle speeds. Neither the direct effect of TNCs was significant.

- Similarly, we do not detect a significant indirect effect for alcohol-involving crashes due to higher exposure or any changes in vehicle speeds. Neither the direct effect of TNCs was significant.
- TNCs indirectly increase property damage-only crashes by 1% due to higher exposure and 2% due to changes in vehicle speeds. The direct effect of TNCs on crashes offsets these increases, reducing crashes by 9%.

The differences across crash types are likely due in part to the smaller sample sizes for the latter categories. Because we have fewer observed crashes that involve pedestrians and bicyclists and alcohol, it is harder to obtain a statistically significant result.

Overall, the results suggest that TNCs are a minor factor in road safety outcomes, at least within the limits of what we can measure with the available data. This finding is broadly consistent with past research on the topic. The conclusions from the independent scientific studies which explored the potential linkages of TNC operations and their impact on road traffic safety outcomes are surprisingly mixed (Barrios et al., 2019; Brazil & Kirk, 2016; Dills & Mulholland, 2018b; Greenwood & Wattal, 2017; Kirk et al., 2020; Morrison et al., 2018). Past research has hypothesized various reasons for such mixed results. One possible reason for such mixed results lies in the non-availability of TNC trip-level data. Few researchers believe that the cost of TNC trips is still the primary factor of whether people use such services as an alternate mode of transport, and until it falls below the total perceived cost (private vehicles or other public transportation model options), people are unwilling to make such a shift. However, this study found a limited effect with spatially detailed data available and still did not find a significant effect in either direction. Our results are consistent with previous research (Kontou & McDonald, 2020a) investigated the influence of Ride Austin, a local TNC service in Travis County, Texas, using real-world data available from RideAustin's open records concludes no significant relationship between TNCs and road crash data (Kontou & McDonald, 2020a).

6.3 Recommendations

Our study found that TNCs may increase crashes indirectly due to higher exposure and changes to vehicle speeds, but these increases are offset by TNCs' direct effect on crash outcomes. TNC's pick-up and drop-off (PUDO) activity positively correlates with road crashes. It is unsurprising, given that "*TNC operates in the most congested areas at the most congested time of the day"*(ITDP, 2019).

Therefore, one of our study's first recommendations is that the city actively regulates curbside road spaces to ensure that multimodal needs (including urban freight delivery service providers, micro-mobility services, transit, pedestrian, and bicyclists) for safe and efficient curbside access are not compromised.

The city administration should work with the TNC providers (data) to identify streets with high TNC usage rates. Based on the strategic goal and priorities, these hotspots' curb space should be re-designed or introduced with passenger loading/unloading (PLC) zones. It will reduce PUDO activities or double parking in travel lanes, bus stops, or bike lanes, reducing road conflicts with other users and streamlining traffic movement.

Efforts should be made to digitize all road assets, including curb spaces. To optimize curb space allocation, mechanisms should be implemented to update these digital assets routinely (or dynamically). These digital inventories should be standardized, and data should be available to public and private actors. It will ensure widespread acceptance of such curbside rules and, therefore greater success rate. For example, in Washington D.C., the authority converted traditional parking spaces around popular dining and entertainment areas with passenger load/unloading zones (ITE, 2014). In 2020, with TNC providers' help, San Francisco digitally geofenced congested streets like Market Street so that the passengers were redirected to nearby side-street locations for pick-ups and drop-offs (Rodriguez, 2020). A study conducted by the University of Washington found that, on average, introducing dedicated passenger loading zones reduced the total time taken by TNC drivers to drop off and pick up passengers by 42seconds in South Lake Union (SLU) area, which is the main campus for Amazon and served by extensive TNC trips (Goodchild et al., 2019).

Second, the California Public Utility Commission (CPUC), responsible for managing TNC licenses within the state, should envisage a mechanism by which TNC-related road crashes are matched with the police, hospital, and emergency department records. Such integration will help agencies understand road safety trends within their area limits. It will also ensure no underreporting of road crashes in all categories. It will help the authorities to develop comprehensive data-led road safety countermeasures.

As of 2023, Judge Mason's landmark judgment on August 2022 mandates TNC service operators to submit annual accident and incidents information to CPUC on or before 19th September of the following year. The redacted version of the crash information excluding sensitive personal information regarding the TNCs driver and passenger is available to the general public from 2021 onwards. However, it lacks spatial information about the crash (coordinates), crucial for developing a data-driven and evidence-based road safety management scheme at all levels (local, state, and federal agencies). Accurate crash data goes a long way in establishing road crash and severity trends. It helps decision-makers develop and implement scientific techniques at a lower cost, leading to higher returns.

Another recommendation will be mandating the background of the vehicle number plates of TNC fleets to be distinctly different from passenger or commercial vehicles. For example, as per the Government of India's Motor Vehicle Act, 1988, all vehicles plying on the road must be registered with the Regional Transport Office (RTOs) (GoI, 1988). In addition, specific rules regarding automobile registration plates exist and are implemented following its 1989 amendment. For example, all private or non-commercial vehicles (two-wheelers and four-wheelers) are required to bear white number plates (GoAP, 2023). A vehicle bearing a white license plate with black lettering signifies that it is for private/personal use only and cannot be used to transport goods or carry passengers (GoAP, 2023). On the other hand, a yellow-colored number plate with black lettering indicates a commercial vehicle. Commercial vehicles have different tax structures than private vehicles and require commercial driving licenses (GoAP, 2023). It can be an alternative low-cost measure until proper laws involving TNCs are enacted.

6.4 Limitations and Future Research

It is important to note that this work is subject to several limitations and offers rich future research opportunities.

On a geographical front, the results and analysis are limited to San Francisco county and, therefore, only depict the association between TNC and road safety for said area. It may be the case that choosing a geographical location with different urban forms and topology may produce a different result.

The second is the analytical design. The research evaluates the objectives using a classic quasi-experimental design framework. The 2010 data represent the traffic flow conditions before the TNC services launch, while the 2016 data represents the after TNC services begin operations. Changes in values get compared within the individual (in our case, road category per census tract) over time, not between the individual and a control. The setup is a low-cost, convenient, and valid alternative to a more survey-intensive study. However, such a design setup faces particular limitations:

- It lacks a control group that can be used to provide some evidence that changes occurring over time were not the result of natural temporal trends or of unmeasured events that occurred during the same time of the study.
- Another issue in before and after studies is the statistical phenomenon of regression to the mean (RTM). It may be the case that the number of crashes observed between these two years is purely random, and we are observing simply noise. However, RTM can be accounted for and minimized by adding more data points (at least three years) for pre- and post-intervention periods.

The third limitation is the non-variability of traffic flows and vehicle speeds. The study uses the annual average daily vehicle miles traveled as an exposure measure, which can be too aggregate to capture the safety effects of traffic flow variations and daily operations. Flow characteristics such as speed variation and congestion level play a significant role in road crash occurrence and should be explored.

The fourth limitation relates to road crash information. There are three aspects related to road crash data.

- Involvement of TNC in a crash – The crash data do not record whether any parties were TNC drivers or passengers. If these data were recorded, it would allow for more direct measurement of the rate of crashes involving TNCs versus the general population. As it stands, the best this study can do is to infer the effect of TNCs on the overall crash rates.
- Underreporting of crashes - One major cause is the underreporting of road crash data. Sciortino et al., in their San Francisco-focused study, noted that police records underestimated the number of injured pedestrians by 21% (Sciortino et al., 2005). While their conclusions are long outdated, the possibility of such an event is not obscure. One of the primary reasons for such underreporting may be related to reporting. Contrary to the general belief, not all road crashes must be reported. Some percentages of crashes never get included in the official records because the incidents are notified to the respective authorities. These include a private settlement between involved parties for insurance purposes, no third-party participation (single-vehicle crashes), or no injuries after the motor vehicle crash. Crashes involving motorcyclists, pedestrians, bicyclists, and PDOs have the highest under-reporting rates (Abay, 2015; Alsop & Langley, 2001; Amoros et al., 2006; Elvik & Mysen, 1999; M. Imprialou & Quddus, 2019; Salifu & Ackaah, 2012; Watson et al., 2015). Underreported crashes can affect the crash frequency count model and injury severity analysis (M. Imprialou & Quddus, 2019; P. E. Ma, 2009; Yamamoto et al., 2008; Yasmin & Eluru, 2013; Ye & Lord, 2011). For this study, which compares two points in time, it is especially important that any be consistent in both time periods. Unfortunately, we do not have a basis to evaluate how consistently crashes were recorded over these two periods.
- Crash location and time – Around 44% of all crashes in 2010 did not have locational attributes. For 2016, this percentage rises to 47%. In general, the more severe the road crash, the higher the probability is that the incident is

accurately reported, although property damage-only (PDO) crashes are estimated to be approximately ~70% of all crashes (NHTSA, 2022c). Errors in reported crash locations generate complexity in data pre-processing. For instance, each crash must be accurately linked to the road link to develop an annual road crash frequency model. Failure to do so may lead to incorrect tagging, which may bias the model result. While no information is available on the impact of wrong crash location on road safety analysis, past studies indicate that it affects the estimation of the model coefficients (M.-I. Imprialou, 2015) and post-construction performance of countermeasures proposed (Brown et al., 2015)

- Crash severity records – Not all crashes have an equal societal impact; therefore, there is a need to classify crash injuries correctly (M. Imprialou & Quddus, 2019). While the public authorities follow strict criteria (in the current study case, the California Collision Investigation Manual 2003), studies have shown that injury severity classifications are not always accurate, leading to potential misclassifications. For instance, McDonald et al. found that 15% of crashes reported as "slight" injury crashes are found to be "life-threatening," according to hospital data (McDonald et al., 2009). Amoros et al. found that such misclassification is not random and targets specific crashes or user characteristics (Amoros et al., 2006). The classification bias and crash under-reporting may affect our count regression analysis involving crash severity.

Finally, the study is limited by the quantity of data. Crashes are relatively rare occurrences, so for many observations, the number of crashes may increase by a small number but results in a significant percentage change, such as from 0 to 1 for one observation and 1 to 0 for a similar observation. Such changes can occur for any number of reasons that are seemingly random and go beyond what our models can reasonably explain. A large enough sample can overcome this variation in the data, which may be one reason why the models for total crashes and fatal and injury crashes include more coefficient estimates that are statistically significant than the other models. If this study is repeated with more data, either from more cities or from more years within the same city, it might be possible to detect more meaningful effects, and it might be possible further to segment those effects by road type or crash type.

6.5 Policy Implications

This study found that TNCs may increase crashes indirectly due to higher exposure and changes to vehicle speeds, but these increases may be offset by TNCs' direct effect on crash outcomes. However, this finding—particularly the direct effect—comes with a low degree of confidence. The evidence suggests that TNCs are a minor factor in road safety outcomes.

This finding is of interest to engineers, planners, and policymakers who seek to improve road safety. It suggests that TNCs are neither a significant cause of the safety problems on American roads nor an effective solution to those problems. Those aiming to reduce traffic crashes would do well to focus on known solutions like adopting a safe system approach as recommended by the U.S. Department of Transportation (DOT) and is part of their National Roadway Safety Strategy to reduce roadway fatalities and serious injuries (FHWA, 2022; USDOT, 2022). Regarding roads, the design should allow humans to make mistakes and greatly reduce injury impacts. It can be achieved by managing vehicle speeds, reducing road user conflicts, improving vehicle safety technology such as driver assistance and collision avoidance, vehicle design to reduce the size and weight of vehicles, or reducing exposure through reduced VMT.

APPENDICES

A1: Primary Modeled Scenarios with intermediary steps – Fatal and Injury Crashes

Fatal and Injury							
Scenario	**sub-scenarios**	**Description**	**Observed**		**Additional Crashes**		**Variable Contribution**
			Year 2010	**Year 2016**	**Year 2010**	**Year 2016**	
A		2010	2961				
B	B1	CounterFactual (CF) 2016 - traffic volumes				98	98
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)				166	68
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD				434	268
C		2016 - traffic volumes + TE + CF CSPD				568	134
D		2016 - traffic volumes + TE + CSPD				708	140
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT				-576	-1284
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO Effect		3288		329	905

A2: Primary Modeled Scenarios with intermediary steps – Pedestrian and Bicycle crashes

Pedestrian and Bicyclists							
Scenario	**sub-scenarios**	**Description**	**Observed**		**Additional Crashes**		**Variable Contribution**
			Year 2010	**Year 2016**	**Year 2010**	**Year 2016**	
A		2010	1296				
B	B1	CounterFactual (CF) 2016 - traffic volumes				20	20
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)				32	12
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD				84	52
C		2016 - traffic volumes + TE + CF CSPD				112	27
D		2016 - traffic volumes + TE + CSPD				193	81
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT				-570	-763
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO Effect		1352		61	630

A3: Primary Modeled Scenarios with intermediary steps – Alcohol-involving Crashes

DUI							
Scenario	**sub-scenarios**	**Description**	**Observed**		**Additional Crashes**		**Variable Contribution**
			Year 2010	**Year 2016**	**Year 2010**	**Year 2016**	
A		2010	502				
B	B1	CounterFactual (CF) 2016 - traffic volumes				38	38
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)				-24	-62
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD				68	93
C		2016 - traffic volumes + TE + CF CSPD				118	49
D		2016 - traffic volumes + TE + CSPD				194	77
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT				-105	-300
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO Effect		537		34	140

A4: Primary Modeled Scenarios with intermediary steps – Property-damage-only (PDO) Crashes

PDO							
Scenario	sub-scenarios	Description	Observed		Additional Crashes		Variable Contribution
			Year 2010	Year 2016	Year 2010	Year 2016	
A		2010	1469				
B	B1	CounterFactual (CF) 2016 - traffic volumes				49	49
	B2	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE)				177	128
	B3	CounterFactual (CF) 2016 - traffic volumes + Time Effects (TE) + CF CSPD				634	458
C		2016 - traffic volumes + TE + CF CSPD				700	65
D		2016 - traffic volumes + TE + CSPD				1063	363
E	E1	2016 - traffic volumes + TE + CSPD + TNC VMT				-120	-1183
	E2	2016 - traffic volumes + TE + CSPD + TNC VMT + PUDO Effect		2028		548	668

A5: Modeled Scenarios with facility types classified into five classes

I. Freeways

Total Crashes								
Models (→) / **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	-0.1	0.82	-0.173	1.25	0.00	0.00	-0.173	1.25
Congested Speed	-0.488 ***	0.08	-0.467	0.25	0.00	0.00	-0.467	0.25
Congested Speed2			0	0.00	0.00	0.00	0	0.00
Ln(PUDO)					0.00	0.00		
Ln(Tot TNC VMT)	-0.084 *	0.04	-0.082	0.04			-0.082	0.04
year_2016_dummy	0.453	0.36	0.443	0.38	0.00	0.00	0.443	0.38
Log pseudolikelihood	-72.73		-72.73				-72.73	
Wald Chi2()	382.78		387.62				387.62	

Count Fatal and Injury Crashes								
Models (→) / **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.158	1.259	-1.27	1.599	0	0	-1.27	1.599
Congested Speed	-0.269 **	0.104	0.175	0.248	0	0	0.175	0.248
Congested Speed2			-0.006	0.003	0	0	-0.006	0.003
Ln(PUDO)					0	0		
Ln(Tot TNC VMT)	0.017	0.051	0.049	0.058			0.049	0.058
year_2016_dummy	-0.246	0.433	-0.42	0.475	0	0	-0.42	0.475
Log psuedolikelihood	-59.597		-58.798				-58.798	
Wald Chi2()	50.94		64.29				64.29	

Count Pedestrian and Bicyclist Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	38.317 *	15.14	38.782 *	15.95	0.00	0.00	38.782 *	15.95
Congested Speed	4.019 *	1.68	3.88	2.56	0.00	0.00	3.88	2.56
Congested Speed2			0.002	0.02	0.00	0.00	0.002	0.02
Ln(PUDO)					0.00	0.00		
Ln(Tot TNC VMT)	0.855	0.51	0.849	0.52			0.849	0.52
year_2016_dummy	-3.959	2.90	-3.952	2.89	0.00	0.00	-3.952	2.89
Log psuedolikelihood	-7.99		-7.99				-7.99	
Wald Chi2()	9.06		9.47				9.47	

Alcohol Involved Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	0.977	4.35	0.907	6.71	0.00	0.00	0.907	6.71
Congested Speed	-0.558	0.33	-0.542	0.89	0.00	0.00	-0.542	0.89
Congested Speed2			0	0.01	0.00	0.00	0	0.01
Ln(PUDO)					0.00	0.00		
Ln(Tot TNC VMT)	-0.058	0.21	-0.056	0.24			-0.056	0.24
year_2016_dummy	-0.313	1.88	-0.322	2.03	0.00	0.00	-0.322	2.03
Log psuedolikelihood	-35.55		-35.55				-35.55	
Wald Chi2()	5.58		5.57				5.57	

II. Major Arterial

Total Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	0.76	0.42	0.613	0.39	0.879 *	0.40	1.075 **	0.39
Congested Speed	0.105	0.13	1.144 **	0.43	0.397	0.33	0.293	0.31
Congested Speed2			-0.036 *	0.01	-0.013	0.01	-0.012	0.01
Ln(PUDO)	0.174 *	0.09	0.304 **	0.09	0.038	0.04		
Ln(Tot TNC VMT)	-0.165 *	0.08	-0.302 **	0.09			-0.015	0.04
year_2016_dummy	0.147	0.32	0.343	0.32	-0.341	0.26	-0.049	0.33
Log psuedolikelihood	-291.94		-288.65		-293.33		-293.74	
Wald Chi2()	15.46		21.49		9.16		9.79	

Count Fatal and Injury Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	1.088 *	0.50	0.904	0.47	1.142 *	0.49	1.37 **	0.46
Congested Speed	0.077	0.15	1.488 **	0.46	0.683	0.36	0.554	0.34
Congested Speed2			-0.048 **	0.02	-0.024	0.01	-0.022	0.01
Ln(PUDO)	0.154	0.10	0.33 **	0.10	0.041	0.05		
Ln(Tot TNC VMT)	-0.148	0.10	-0.33 **	0.11			-0.016	0.05
year_2016_dummy	0.053	0.38	0.292	0.38	-0.462	0.28	-0.14	0.36
Log psuedolikelihood	-256.65		-252.29		-256.32		-256.67	
Wald Chi2()	14.40		27.53		15.70		14.35	

Count Pedestrian and Bicyclist Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	1.183	0.70	0.967	0.68	1.259	0.68	1.645 *	0.68
Congested Speed	0.235	0.20	1.878 **	0.61	0.511	0.46	0.391	0.48
Congested Speed2			-0.06 **	0.02	-0.016	0.02	-0.017	0.02
Ln(PUDO)	0.339 *	0.14	0.586 ***	0.16	0.067	0.07		
Ln(Tot TNC VMT)	-0.328	0.17	-0.612 **	0.19			-0.019	0.08
year_2016_dummy	0.244	0.64	0.754	0.66	-0.669	0.41	-0.2	0.59
Log psuedolikelihood	-164.34		-161.49		-165.98		-166.40	
Wald Chi2()	17.63		23.36		10.55		8.42	

Alcohol involved Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	1.996	1.03	1.738	1.05	1.923	1.10	2.115	1.11
Congested Speed	-0.117	0.34	1.658	1.05	0.925	0.80	0.946	0.75
Congested Speed2			-0.064	0.04	-0.041	0.03	-0.043	0.03
Ln(PUDO)	0.034	0.22	0.273	0.25	0.012	0.12		
Ln(Tot TNC VMT)	-0.036	0.23	-0.301	0.27			-0.043	0.13
year_2016_dummy	-0.214	0.95	0.209	0.94	-0.476	0.71	-0.119	0.93
Log psuedolikelihood	-106.35		-105.12		-105.63		-105.58	
Wald Chi2()	6.07		7.81		7.96		8.01	

III. Minor Arterial Road

Total Crashes								
Models (→) / **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.488	0.38	0.464	0.34	0.463	0.34	0.488	0.33
Congested Speed	0.119	0.24	2.447 ***	0.70	2.459 ***	0.70	1.989 **	0.62
Congested Speed2			-0.087 ***	0.02	-0.087 ***	0.02	-0.072 ***	0.02
Ln(PUDO)	-0.007	0.10	0.143	0.11	0.15 *	0.06		
Ln(Tot TNC VMT)	0.095	0.13	0.008	0.13			0.135	0.07
year_2016_dummy	-0.573	0.44	-0.939 *	0.47	-0.927 *	0.37	-0.943 *	0.47
Log pseudolikelihood	-206.41		-201.50		-201.51		-202.31	
Wald Chi2()	9.35		34.90		32.79		27.15	

Count Fatal and Injury Crashes								
Models (→) / **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.284	0.39	0.221	0.34	0.247	0.34	0.267	0.35
Congested Speed	0.041	0.25	2.58 **	0.80	2.294 **	0.80	1.694 *	0.71
Congested Speed2			-0.096 **	0.03	-0.084 **	0.03	-0.066 *	0.03
Ln(PUDO)	0.098	0.10	0.269 *	0.12	0.125	0.07		
Ln(Tot TNC VMT)	-0.057	0.13	-0.167	0.15			0.074	0.08
year_2016_dummy	-0.207	0.50	-0.551	0.52	-0.806	0.43	-0.571	0.51
Log pseudolikelihood	-180.00		-175.93		-176.66		-177.89	
Wald Chi2()	4.50		18.61		13.92		10.38	

Count Pedestrian and Bicyclist Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	1.104	0.70	1.097	0.69	1.095	0.70	1.112	0.68
Congested Speed	0.185	0.30	2.559 **	0.82	2.571 **	0.81	2.043 **	0.70
Congested Speed2			-0.091 **	0.03	-0.091 ***	0.03	-0.073 **	0.02
Ln(PUDO)	-0.006	0.14	0.153	0.15	0.159	0.09		
Ln(Tot TNC VMT)	0.109	0.17	0.007	0.17			0.145	0.09
year_2016_dummy	-0.802	0.63	-1.139	0.65	-1.127 *	0.54	-1.149	0.64
Log pseudolikelihood	-124.98		-123.12		-123.12		-123.46	
Wald Chi2()	6.13		25.66		25.66		19.97	

Alcohol involved Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	0.772	1.55	0.813	1.51	0.911	1.47	0.881	1.48
Congested Speed	-0.119	0.57	-1.351	2.45	-1.624	2.51	-2.066	2.37
Congested Speed2			0.047	0.09	0.06	0.09	0.071	0.09
Ln(PUDO)	0.292	0.29	0.234	0.30	-0.022	0.16		
Ln(Tot TNC VMT)	-0.348	0.33	-0.322	0.32			-0.11	0.17
year_2016_dummy	0.522	1.14	0.713	1.21	0.095	1.04	0.687	1.21
Log pseudolikelihood	-69.32		-69.19		-69.65		-69.47	
Wald Chi2()	2.11		2.20		1.37		1.90	

IV. Collector Road

Total Crashes								
Models (→) **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.17	0.36	0.13	0.34	0.34	0.38	0.529	0.37
Congested Speed	-0.098	0.35	4.409 *	1.74	1.759	1.59	1.803	1.58
Congested Speed2			-0.176 *	0.07	-0.072	0.07	-0.073	0.07
Ln(PUDO)	0.434 ***	0.12	0.55 ***	0.12	-0.012	0.06		
Ln(Tot TNC VMT)	-0.511 ***	0.12	-0.621 ***	0.13			-0.089	0.06
year_2016_dummy	0.666 *	0.32	0.67 *	0.32	0.032	0.31	0.479	0.33
Log pseudolikelihood	-253.22		-250.40		-259.90		-258.35	
Wald Chi2()	20.71		26.58		3.92		6.27	

Count Fatal and Injury Crashes								
Models (→) **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.082	0.43	-0.003	0.38	0.232	0.45	0.479	0.44
Congested Speed	-0.096	0.47	7.63 ***	1.84	4.358 **	1.61	4.414 **	1.59
Congested Speed2			-0.303 ***	0.08	-0.174 **	0.07	-0.175 **	0.07
Ln(PUDO)	0.467 **	0.16	0.686 ***	0.16	0.003	0.08		
Ln(Tot TNC VMT)	-0.542 **	0.16	-0.755 ***	0.16			-0.094	0.08
year_2016_dummy	0.618	0.42	0.658	0.42	-0.114	0.40	0.453	0.43
Log pseudolikelihood	-215.08		-209.07		-218.15		-217.04	
Log pseudolikelihood	12.44		28.81		8.13		8.95	

Count Pedestrian and Bicyclist Crashes								
Models (→) / **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.007	0.57	-0.056	0.53	0.126	0.57	0.299	0.58
Congested Speed	0.111	0.63	5.099 *	2.57	2.096	2.55	2.338	2.46
Congested Speed2			-0.198	0.10	-0.08	0.10	-0.088	0.10
Ln(PUDO)	0.442 *	0.20	0.599 **	0.21	-0.042	0.09		
Ln(Tot TNC VMT)	-0.556 **	0.21	-0.701 **	0.22			-0.144	0.09
year_2016_dummy	0.825	0.62	0.806	0.60	0.148	0.56	0.795	0.60
Log pseudolikelihood	-156.46		-154.93		-1593.35		-158.20	
Wald Chi2()	8.20		12.30		2.29		4.81	

Alcohol involved Crashes								
Models (→) / **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	-0.677	1.06	-0.693	1.09	-0.536	1.10	-0.372	1.11
Congested Speed	-1.058	0.88	-6.594	6.43	-7.73	6.38	-8.334	6.36
Congested Speed2			0.205	0.23	0.248	0.23	0.273	0.23
Ln(PUDO)	0.513	0.43	0.4	0.44	0.086	0.16		
Ln(Tot TNC VMT)	-0.423	0.47	-0.336	0.47			0.04	0.17
year_2016_dummy	-0.384	1.04	-0.241	1.06	-0.529	0.95	-0.306	1.05
Log pseudolikelihood	-76.44		-76.13		-76.39		-76.50	
Wald Chi2()	2.69		3.34		2.72		2.47	

V. Local Road

Total Crashes								
Models (→) **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.308	0.29	0.318	0.29	0.325	0.27	0.358	0.25
Congested Speed	0.235 **	0.08	-0.064	1.40	-0.082	1.37	-0.131	1.40
Congested Speed2			0.013	0.06	0.014	0.06	0.016	0.06
Ln(PUDO)	0.052	0.11	0.047	0.11	0.038	0.06		
Ln(Tot TNC VMT)	-0.014	0.08	-0.008	0.09			0.028	0.05
year_2016_dummy	-0.117	0.32	-0.119	0.32	-0.117	0.32	-0.068	0.28
Log pseudolikelihood	-314.72		-314.70		-314.70		-314.74	
Wald Chi2()	31.06		45.28		44.34		45.12	

Count Fatal and Injury Crashes								
Models (→) **Independent Variables (↓)**	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Ln(Tot VMT)	0.094	0.34	0.052	0.35	0.238	0.32	0.271	0.30
Congested Speed	0.17	0.09	1.198	1.95	0.661	1.90	0.805	1.93
Congested Speed2			-0.045	0.09	-0.019	0.09	-0.026	0.09
Ln(PUDO)	0.227	0.17	0.247	0.17	0.021	0.08		
Ln(Tot TNC VMT)	-0.183	0.13	-0.205	0.13			-0.018	0.07
year_2016_dummy	-0.076	0.47	-0.065	0.48	-0.003	0.48	0.234	0.42
Log pseudolikelihood	-269.79		-269.64		-270.41		-270.41	
Wald Chi2()	18.83		33.27		32.38		32.04	

Count Pedestrian and Bicyclist Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	0.128	0.43	-0.067	0.46	0.135	0.41	0.16	0.40
Congested Speed	0.04	0.11	3.851	2.93	3.114	2.73	3.355	2.85
Congested Speed2			-0.17	0.13	-0.135	0.12	-0.146	0.13
Ln(PUDO)	0.174	0.24	0.26	0.24	0.025	0.12		
Ln(Tot TNC VMT)	-0.109	0.16	-0.209	0.17			-0.019	0.09
year_2016_dummy	-0.333	0.76	-0.237	0.75	-0.144	0.74	0.122	0.60
Log pseudolikelihood	-175.02		-174.21		-174.55		-174.55	
Wald Chi2()	2.52		6.92		6.61		6.48	

Alcohol involved Crashes								
Models (→)	**linear speed**		**non-linear speed**		**Only TNC PUDO**		**Only TNC TNC VMT**	
Independent Variables (↓)								
Ln(Tot VMT)	2.046 *	0.95	2.19 *	0.94	1.927 *	0.82	1.775 *	0.76
Congested Speed	0.452 *	0.21	-6.774	4.51	-6.184	4.26	-6.206	4.40
Congested Speed2			0.319	0.20	0.291	0.19	0.291	0.20
Ln(PUDO)	-0.278	0.44	-0.417	0.46	-0.128	0.21		
Ln(Tot TNC VMT)	0.122	0.36	0.268	0.38			-0.047	0.17
year_2016_dummy	0.403	1.18	0.431	1.20	0.394	1.22	-0.059	1.09
Log pseudolikelihood	-111.16		-109.71		-109.94		-110.10	
Wald Chi2()	6.54		9.27		9.36		9.06	

BIBLIOGRAPHY

Abay, K. A. (2015). Investigating the nature and impact of reporting bias in road crash data. *Transportation Research Part A: Policy and Practice*, *71*, 31–45. https://doi.org/10.1016/j.tra.2014.11.002

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, *38*(3), 618–625. https://doi.org/10.1016/j.aap.2005.12.006

Albalate, D., & Fageda, X. (2021). On the relationship between congestion and road safety in cities. *Transport Policy*, *105*, 145–152. https://doi.org/10.1016/j.tranpol.2021.03.011

Alemi, F., Circella, G., Handy, S., & Mokhtarian, P. (2018). What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. *Travel Behaviour and Society*, *13*, 88–104. https://doi.org/10.1016/j.tbs.2018.06.002

Allison, P. (2012, August 7). Do We Really Need Zero-Inflated Models? *Statistical Horizons*. https://statisticalhorizons.com/zero-inflated-models/

Alsop, J., & Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis & Prevention*, *33*(3), 353–359. https://doi.org/10.1016/S0001-4575(00)00049-X

Amoros, E., Martin, J.-L., & Laumon, B. (2006). Under-reporting of road crash casualties in France. *Accident Analysis & Prevention*, *38*(4), 627–635. https://doi.org/10.1016/j.aap.2005.11.006

Anderson, M. L., & Davis, L. W. (2021). *Uber and Alcohol-Related Traffic Fatalities* (Working Paper No. 29071; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w29071

Andrew J Hawkins. (2015, October 23). *Uber covers 75 percent of the US, but getting to 100 will be really hard—The Verge* [Technology journalism]. Uber Covers 75 Percent of the US, but Getting to 100 Will Be Really Hard. https://www.theverge.com/2015/10/23/9603324/uber-coverage-rural-areas-dominance-plan

Balkin, S., & Ord, K. (2001). Assessing the impact of speed-limit increases on fatal interstate crashes. *Journal of Transportation and Statistics*, *4*.

Barkana, Y., Zadok, D., Morad, Y., & Avni, I. (2004). Visual field attention is reduced by concomitant hands-free conversation on a cellular telephone. *American Journal of Ophthalmology*, *138*(3), 347–353. https://doi.org/10.1016/j.ajo.2004.04.016

Barreto, Y., Silveira Neto, R. da M., & Carazza, L. (2021). Uber and traffic safety: Evidence from Brazilian cities. *Journal of Urban Economics*, *123*, 103347. https://doi.org/10.1016/j.jue.2021.103347

Barrios, J. M., Hochberg, Y. V., & Yi, H. (2019). *The Cost of Convenience: Ridehailing and Traffic Fatalities* (SSRN Scholarly Paper ID 3361227). Social Science Research Network. https://papers.ssrn.com/abstract=3361227

Bayam, E., Liebowitz, J., & Agresti, W. (2005). Older drivers and accidents: A meta analysis and data mining application on traffic accident data. *Expert Systems with Applications*, *29*(3), 598–629. https://doi.org/10.1016/j.eswa.2005.04.025

Beede, K. E., & Kass, S. J. (2006). Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis & Prevention*, *38*(2), 415–421. https://doi.org/10.1016/j.aap.2005.10.015

Berrebi, S. J., Joshi, S., & Watkins, K. E. (2021). On bus ridership and frequency. *Transportation Research Part A: Policy and Practice*, *148*, 140–154. https://doi.org/10.1016/j.tra.2021.03.005

Blazquez, C., Laurent, J. G. C., & Nazif-Munoz, J. I. (2021). Differential impacts of ridesharing on alcohol-related crashes by socioeconomic municipalities: Rate of technology adoption matters. *BMC Public Health*, *21*(1), 2008. https://doi.org/10.1186/s12889-021-12066-z

Bolitho, J. (2013). *A multi-stage, multi-faceted approach to addressing 'car dooring' in inner Melbourne*. Australasian College of Road Safety Conference. https://acrs.org.au/files/papers/50%20Bolitho_NPR.pdf

Brazil, N., & Kirk, D. (2020). Ridehailing and alcohol-involved traffic fatalities in the United States: The average and heterogeneous association of uber. *PLOS ONE*, *15*(9), e0238744. https://doi.org/10.1371/journal.pone.0238744

Brazil, N., & Kirk, D. S. (2016). Uber and Metropolitan Traffic Fatalities in the United States. *American Journal of Epidemiology*, *184*(3), 192–198. https://doi.org/10.1093/aje/kww062

Brett Helling. (2018, July 25). *Understanding the Lyft Driver Rating System—Ridester*. https://www.ridester.com/lyft-driver-rating/

Brisson, E. M., Sall, E., & Ang-Olson, J. (2012). Achieving Goals of San Francisco, California, for Greenhouse Gas Reductions in Transportation Sector: What Would It Take? *Transportation Research Record*, *2287*(1), 89–97. https://doi.org/10.3141/2287-11

Brown, K., Sarasua, W. A., Ogle, J. H., Mammadrahimli, A., Chowdhury, M., Davis, W. J., & Huynh, N. (2015). *Assessment of Crash Location Improvements in Map-Based Geocoding Systems and Subsequent Benefits to Geospatial Crash Analysis*.

Browne, R. (2019, November 25). *Uber stripped of its London license as regulator says it put passengers at risk*. CNBC. https://www.cnbc.com/2019/11/25/uber-stripped-of-its-london-license-in-huge-blow-dealt-by-tfl.html

Browne, R. (2020, September 14). *Uber fights London ban in court for a second time*. CNBC. https://www.cnbc.com/2020/09/14/uber-fights-london-ban-in-court-for-a-second-time.html

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, *40*(4), 1282–1293. https://doi.org/10.1016/j.aap.2008.01.009

Cameron, A. C., & Trivedi, P. K. (1998). *Regression Analysis of Count Data*. 434.

Castiglione, J., Hiatt, R., Chang, T., & Charlton, B. (2006). Application of Travel Demand Microsimulation Model for Equity Analysis. *Transportation Research Record*, *1977*(1), 35–42. https://doi.org/10.1177/0361198106197700105

CHP-SWITRS. (2019). https://iswitrs.chp.ca.gov/Reports/jsp/index.jsp

Clewlow, R., & Mishra, G. (2017). *Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States*.

Cooper, D., Castiglione, J., Mislove, A., & Wilson, C. (2018). Profiling Transport Network Company Activity using Big Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(42), 192–202. https://doi.org/10.1177/0361198118798459

CPUC. (2016). *Insurance Requirements for TNCs*. https://www.cpuc.ca.gov/regulatory-services/licensing/transportation-licensing-and-analysis-branch/transportation-network-companies/tnc-insurance-requirements

Cumming, B. (2012). *Conflict path analysis: Analysing & managing the cyclist-driver interface*. 15.

Dalziel, J. R., & Job, R. F. S. (1997a). *Taxi drivers and road safety* (p. 162) [Report to the Federal Office of Road Safety]. Univeristy of Sydney. http://www.taxi-library.org/taxi-drivers-and-road-safety.pdf

Dalziel, J. R., & Job, R. F. S. (1997b). Motor vehicle accidents, fatigue and optimism bias in taxi drivers. *Accident Analysis & Prevention*, *29*(4), 489–494. https://doi.org/10.1016/S0001-4575(97)00028-6

Desktop Help 10.0—Map Projections. (2013). ArcGIS Desktop Help 10.0. https://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/What_are_map_projections/003r00000001000000/

Dills, A. K., & Mulholland, S. E. (2017). *Ride-Sharing, Fatal Crashes, and Crime* (SSRN Scholarly Paper ID 2783797). Social Science Research Network. https://papers.ssrn.com/abstract=2783797

Dills, A. K., & Mulholland, S. E. (2018a). Ride-Sharing, Fatal Crashes, and Crime: Ride-Sharing, Fatal Crashes, and Crime. *Southern Economic Journal*, *84*(4), 965–991. https://doi.org/10.1002/soej.12255

Dills, A. K., & Mulholland, S. E. (2018b). Ride-Sharing, Fatal Crashes, and Crime. *Southern Economic Journal*, *84*(4), 965–991. https://doi.org/10.1002/soej.12255

Elvik, R. (2001). Area-wide urban traffic calming schemes: A meta-analysis of safety effects. *Accident Analysis & Prevention*, *33*(3), 327–336. https://doi.org/10.1016/S0001-4575(00)00046-4

Elvik, R., Christensen, P., Amundsen, A. H., & Transportøkonomisk institutt (Norway). (2006). *Speed and road accidents: An evaluation of the power model*. Institute of Transport Economics.

Elvik, R., & Mysen, A. (1999). Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. *Transportation Research Record*, *1665*(1), 133–140. https://doi.org/10.3141/1665-18

Elvik, R., Sogge, C. V., Lager, L., Amundsen, F. H., Pasnin, L. T., Karlsen, R., & Fosli, K. (2012). Assessing the efficiency of priorities for traffic law enforcement in Norway.

Accident Analysis & Prevention, *47*, 146–152. https://doi.org/10.1016/j.aap.2012.01.017

Erhardt, G. D., Mucci, R. A., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2020). *Do transportation network companies increase or decrease transit ridership? Empirical evidence from San Francisco*. submitted for publication in Transportation.

Erhardt, G. D., Mucci, R. A., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2022). Do transportation network companies increase or decrease transit ridership? Empirical evidence from San Francisco. *Transportation*, *49*(2), 313–342. https://doi.org/10.1007/s11116-021-10178-4

Erhardt, G., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Science Advances*, *5*, eaau2670. https://doi.org/10.1126/sciadv.aau2670

Evans, D. (2021, August 31). Uber and Lyft rides are more expensive than ever because of a driver shortage. *CNBC*. https://www.cnbc.com/2021/08/31/why-uber-and-lyft-rides-are-more-expensive-than-ever.html

Ewing, R., & Hamidi, S. (2015). *Urban sprawl leads to more fatal crashes, while denser areas experience higher rates of accidents which cause injuries. | USAPP*. https://blogs.lse.ac.uk/usappblog/2015/02/27/urban-sprawl-leads-to-more-fatal-crashes-while-denser-areas-experience-higher-rates-of-accidents-which-cause-injuries/

Ewing, R., Rolf, P., & Chen, D. (2002). *MEASURING SPRAWL AND ITS IMPACT* (p. 42). Smart Growth America.

Ewing, R., Schieber, R. A., & Zegeer, C. V. (2003). Urban Sprawl as a Risk Factor in Motor Vehicle Occupant and Pedestrian Fatalities. *American Journal of Public Health*, *93*(9), 5.

Ewoldsen, B. (2020, August 17). *COVID-19 trends impacting the future of transportation planning and research | National Academies*. https://www.nationalacademies.org/trb/blog/covid-19-trends-impacting-the-future-of-transportation-planning-and-research

Federal Highway Administration. (2019). *Office of Highway Policy Information—Policy | Federal Highway Administration*. https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm?CFID=141526723&CFTOKEN=3840459bf50ef5a1-239AAFED-D2C2-5F97-D4E139A8C152B64E

Fehr & Peers. (2019). *Estimated TNC Share of VMT in Six US Metropolitan Regions*. https://drive.google.com/file/d/1FIUskVkj9lsAnWJQ6kLhAhNoVLjfFdx3/view

Feigon, S., & Murphy, C. (2018). *Broadening Understanding of the Interplay Between Public Transit, Shared Mobility, and Personal Automobiles* (TCRP 195). Transportation Research Board. https://doi.org/10.17226/24996

Feikert-Ahalt, C. (2020, January 13). *England: Uber Faces Operating Ban in London | Global Legal Monitor* [Web page]. //www.loc.gov/law/foreign-news/article/england-uber-faces-operating-ban-in-london/

FHWA. (2022). *Making our Roads Safer through a Safe System Approach | FHWA* (FHWA-HRT-22-002). FHWA. https://highways.dot.gov/public-roads/winter-2022/01

Flor, M., Ortuño, A., & Guirao, B. (2022). Does the Implementation of Ride-Hailing Services Affect Urban Road Safety? The Experience of Madrid. *International Journal of Environmental Research and Public Health*, *19*(5), 3078. https://doi.org/10.3390/ijerph19053078

Flor, M., Ortuño, A., Guirao, B., & Casares, J. (2021). Analysis of the Impact of Ride-Hailing Services on Motor Vehicles Crashes in Madrid. *Sustainability*, *13*(11), Article 11. https://doi.org/10.3390/su13115855

Foss, R. D., Goodwin, A. H., McCartt, A. T., & Hellinga, L. A. (2009). Short-term effects of a teenage driver cell phone restriction. *Accident Analysis & Prevention*, *41*(3), 419–424. https://doi.org/10.1016/j.aap.2009.01.004

Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Thomsen, L. K. (1995). *MEASURING THE CONTRIBUTION OF RANDOMNESS, EXPOSURE, WEATHER, AND DAYLIGHT TO THE VARIATION IN ROAD ACCIDENT COUNTS*. 20.

Gehrke, S. R., Felix, A., & Reardon, T. (2018). *Fare choices: A survey of ride-hailing passengers in Metro Boston*. Boston Metropolitan Area Planning Council.

Gimond, M. (2022). Intro to GIS and Spatial Analysis. In *Chapter 9 Coordinate Systems | Intro to GIS and Spatial Analysis*. https://mgimond.github.io/Spatial/chp09_0.html

Girasek, D. C., & Taylor, B. (2010). An exploratory study of the relationship between socioeconomic status and motor vehicle safety features. *Traffic Injury Prevention*, *11*(2), 151–155. https://doi.org/10.1080/15389580903531598

GoAP. (2023, February). *Registration | State Codes | TRANSPORT DEPARTMENT GOVERNMENT OF ANDHRA PRADESH - INDIA*. State Codes. https://www.aptransport.org/html/registration-statecodes.html

GoI. (1988). *The Motor Vehicles Act, 1988 | Ministry of Road Transport & Highways, Government of India*. https://morth.nic.in/motor-vehicles-act-1988

Goodchild, A., Machado, J., Chiara, G. D., Ranjbari, A., & Mackenzie, D. (2019). *Curbing Conflicts: Curb Allocation Change Project | Supply Chain Transportation and Logistics Center at the University of Washington*.

https://depts.washington.edu/sctlctr/research/publications/curb-allocation-change-project

Greenwood, B. N., & Wattal, S. (2017). Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. *MIS Quarterly: Management Information Systems*, *41*(1), 163–187. https://doi.org/10.25300/MISQ/2017/41.1.08

Gregory D. Erhardt, Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen and Joe Castiglione. (2019, May 8). *Do transportation network companies decrease or increase congestion? | Science Advances*. https://advances.sciencemag.org/content/5/5/eaau2670

Guimarães, P. (2008). The fixed effects negative binomial model revisited. *Economics Letters*, *99*(1), 63–66. https://doi.org/10.1016/j.econlet.2007.05.030

Hadayeghi, A., Shalaby, A. S., & Persaud, B. (2003). Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. *Transportation Research Record*, *1840*(1), 87–95. https://doi.org/10.3141/1840-10

Hadayeghi, A., Shalaby, A. S., Persaud, B. N., & Cheung, C. (2006). Temporal transferability and updating of zonal level accident prediction models. *Accident; Analysis and Prevention*, *38*(3), 579–589. https://doi.org/10.1016/j.aap.2005.12.003

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, *39*(2), 372–379. https://doi.org/10.1016/j.aap.2006.08.013

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

Hausman, J., Hall, B., & Griliches, Z. (1984). *Econometric Models for Count Data with an Application to the Patents-R&D Relationship* (NBER Technical Working Papers

No. 0017). National Bureau of Economic Research, Inc. https://eml.berkeley.edu/~bhhall/papers/HausmanHallGriliches%20E84.pdf

Henao, A., & Marshall, W. E. (2019). The impact of ride-hailing on vehicle miles traveled. *Transportation*, *46*(6), 2173–2194. https://doi.org/10.1007/s11116-018-9923-2

Holdgraf, C., & Wasser, L. (2020, September 11). *Geographic vs projected coordinate reference systems—GIS in Python*. Earth Data Science - Earth Lab. https://www.earthdatascience.org/courses/use-data-open-source-python/intro-vector-data-python/spatial-data-vector-shapefiles/geographic-vs-projected-coordinate-reference-systems-python/

Horrey, W. J., & Wickens, C. D. (2006). Examining the Impact of Cell Phone Conversations on Driving Using Meta-Analytic Techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(1), 196–205. https://doi.org/10.1518/001872006776412135

Hu, M.-C., Pavlicova, M., & Nunes, E. V. (2011). Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial. *The American Journal of Drug and Alcohol Abuse*, *37*(5), 367–375. https://doi.org/10.3109/00952990.2011.597280

Huang, J. Y., Majid, F., & Daku, M. (2019). Estimating effects of Uber ride-sharing service on road traffic-related deaths in South Africa: A quasi-experimental study. *Journal of Epidemiology and Community Health*, *73*(3), 263–271. https://doi.org/10.1136/jech-2018-211006

Humphreys, D. K., Degli Esposti, M., Williams, F. M., Kondo, M. C., & Morrison, C. (2021). Assessing the impact of a local community subsidised rideshare programme on road traffic injuries: An evaluation of the Evesham Saving Lives programme. *Injury Prevention*, *27*(3), 232–237. https://doi.org/10.1136/injuryprev-2020-043728

IIHS. (2018, May 8). *Study highlights rising pedestrian deaths, points toward solutions*. IIHS-HLDI Crash Testing and Highway Safety. https://www.iihs.org/news/detail/study-highlights-rising-pedestrian-deaths-points-toward-solutions

IIHS. (2020). *Fatality Facts 2020: Urban/rural comparison*. IIHS-HLDI Crash Testing and Highway Safety. https://www.iihs.org/topics/fatality-statistics/detail/urban-rural-comparison

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. *Accident Analysis & Prevention*, *130*, 84–90. https://doi.org/10.1016/j.aap.2017.02.022

Imprialou, M.-I. (2015). *Developing Accident-Speed Relationships Using a New Modelling Approach*.

ITDP. (2019). *Ride Fair: A Policy Framework for Managing Transportation Network Companies*. ITDP. https://www.itdp.org/wp-content/uploads/2019/03/2019.03.13.TNC-Policy.V9.pdf

ITE. (2014). *Study and pilot projects*. District Department of Transportation (DDOT). https://www.ite.org/pub/?id=C29F4D5E-FE34-2037-3B96-DE312E1DBBFF

Ivan, J. N., Wang, C., & Bernardo, N. R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis & Prevention*, *32*(6), 787–795. https://doi.org/10.1016/S0001-4575(99)00132-3

Jackson, C. K., & Owens, E. G. (2011). One for the road: Public transportation, alcohol consumption, and intoxicated driving. *Journal of Public Economics*, *95*(1), 106–121. https://doi.org/10.1016/j.jpubeco.2010.09.010

Jeffrey Wooldridge. (2019). *Introductory Econometrics: A Modern Approach. Seventh Edition* (Seventh). Cengage Learning.

Johansson, P. (1996). Speed limitation and motorway casualties: A time series count data regression approach. *Accident; Analysis and Prevention*, *28*(1), 73–87. https://doi.org/10.1016/0001-4575(95)00043-7

Jones, R. G. (2015). *Uber Claims Credit for Drop in Drunk Driving Accidents. But Where's the Evidence?* ProPublica. https://www.propublica.org/article/uber-claims-credit-for-drop-in-drunk-driving-accidents-but-wheres-evidence

Jovanis, P. P., & Chang, H.-L. (1986). MODELING THE RELATIONSHIP OF ACCIDENTS TO MILES TRAVELED. *Transportation Research Record*, *1068*, Article HS-039 853. https://trid.trb.org/view/288394

Karlaftis, M. G., & Tarko, A. P. (1998). Heterogeneity considerations in accident modeling. *Accident Analysis & Prevention*, *30*(4), 425–433. https://doi.org/10.1016/S0001-4575(97)00122-X

Kirk, D. S., Cavalli, N., & Brazil, N. (2020). The implications of ridehailing for risky driving and road accident injuries and fatalities. *Social Science & Medicine*, *250*, 112793. https://doi.org/10.1016/j.socscimed.2020.112793

Kockelman, K. (2006). *Safety Impacts and Other Implications of Raised Speed Limits on High-Speed Roads*. Transportation Research Board. https://doi.org/10.17226/22048

Kontou, E., & McDonald, N. (2020a). *Associating Ridesourcing with Road Safety Outcomes: Insights from Austin Texas*.

Kontou, E., & McDonald, N. C. (2020b). *Associating Ridesourcing with Road Safety Outcomes: Insights from Austin Texas*. 24.

Kwayu, K. M., Kwigizile, V., Lee, K., Oh, J.-S., & Nelson, T. (2022). Automatic topics extraction from crowdsourced cyclists near-miss and collision reports using text mining and Artificial Neural Networks. *International Journal of Transportation Science and Technology*, *11*(4), 767–779. https://doi.org/10.1016/j.ijtst.2021.10.005

Ladrón de Guevara, F., Washington, S. P., & Oh, J. (2004). Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. *Transportation Research Record*, *1897*(1), 191–199. https://doi.org/10.3141/1897-25

Lee, T. B. (2022, December 22). I drove for Lyft for a week and learned its business model is broken. *Full Stack Economics*. https://www.fullstackeconomics.com/p/i-drove-for-lyft-for-a-week-and-learned

Lenzo, K. (2016, February 23). *Understanding Uber's five-star rating system*. CNBC. https://www.cnbc.com/2016/02/23/understanding-ubers-five-star-rating-system.html

Lesch, M. F., & Hancock, P. A. (2004). Driving performance during concurrent cell-phone use: Are drivers aware of their performance decrements? *Accident Analysis & Prevention*, *36*(3), 471–480. https://doi.org/10.1016/S0001-4575(03)00042-3

Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis & Prevention*, *27*(5), 675–685. https://doi.org/10.1016/0001-4575(95)00018-U

Li, L., Zhu, L., & Sui, D. Z. (2007). A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography*, *15*(4), 274–285. https://doi.org/10.1016/j.jtrangeo.2006.08.005

Lord, D., Manar, A., & Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention*, *37*(1), 185–199. https://doi.org/10.1016/j.aap.2004.07.003

Lyft. (2023a). *Driver and passenger ratings—Lyft Help*. https://help.lyft.com/hc/en-us/all/articles/115013079948-Driver-and-passenger-ratings

Lyft. (2023b). *Lyft Driver Rewards—Terms of Service*. https://www.lyft.com/terms/driver-rewards

Ma, J., & Kockelman, K. (2006). Crash frequency and severity modeling using clustered data from Washington State. *2006 IEEE Intelligent Transportation Systems Conference*, 1621–1626. https://doi.org/10.1109/ITSC.2006.1707456

Ma, P. E. (2009). *Bayesian analysis of underreporting Poisson regression model with an application to traffic crashes on two-lane highways*.

Maag, U., Vanasse, C., Dionne, G., & Laberge-Nadeau, C. (1997). Taxi drivers' accidents: How binocular vision problems are related to their rate and severity in terms of the number of victims. *Accident Analysis & Prevention*, *29*(2), 217–224. https://doi.org/10.1016/S0001-4575(96)00075-9

Macek, K. (2022). *Pedestrian Traffic Fatalities by State—2021 Preliminary Data (January-December).pdf* (p. 37). Governors Highway Safety Association. https://www.ghsa.org/resources/Pedestrians22

Maples, W. C., DeRosier, W., Hoenes, R., Bendure, R., & Moore, S. (2008). The effects of cell phone use on peripheral vision. *Optometry - Journal of the American Optometric Association*, *79*(1), 36–42. https://doi.org/10.1016/j.optm.2007.04.102

McCarley, J. S., Vais, M. J., Pringle, H., Kramer, A. F., Irwin, D. E., & Strayer, D. L. (2016). Conversation Disrupts Change Detection in Complex Traffic Scenes: *Human Factors*. https://doi.org/10.1518/hfes.46.3.424.50394

Mccartt, A. T., Hellinga, L. A., & Bratiman, K. A. (2006). Cell Phones and Driving: Review of Research. *Traffic Injury Prevention*, *7*(2), 89–106. https://doi.org/10.1080/15389580600651103

McCartt, A. T., Hellinga, L. A., & Geary, L. L. (2006). Effects of Washington, D.C. Law on Drivers' Hand-Held Cell Phone Use. *Traffic Injury Prevention*, *7*(1), 1–5. https://doi.org/10.1080/15389580500412853

McDonald, G., Davie, G., & Langley, J. (2009). Validity of Police-Reported Information on Injury Severity for Those Hospitalized from Motor Vehicle Traffic

Crashes. *Traffic Injury Prevention*, *10*(2), 184–190. https://doi.org/10.1080/15389580802593699

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *BMJ*, *331*(7514), 428. https://doi.org/10.1136/bmj.38537.397512.55

Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, *26*(4), 471–482. https://doi.org/10.1016/0001-4575(94)90038-8

Miaou, S.-P. (1996). *MEASURING THE GOODNESS-OF-FIT OF ACCIDENT PREDICTION MODELS* (FHWA-RD-96-040). Article FHWA-RD-96-040. https://trid.trb.org/view/478498

Miaou, S.-P. (2001). *Estimating roadside encroachment rates with the combined strengths of accident- and encroachment-based approaches* (FHWA-RD-01-124). https://rosap.ntl.bts.gov/view/dot/14637

Miaou, S.-P., Hu, P. S., Wright, T., Davis, S. C., Rathi, A. K., & Oak Ridge National Laboratory. Center for Transportation Analysis. (1993). *Development of Relationship Between Truck Accidents and Geometric Design: Phase I* (FHWA-RD-91-124). https://rosap.ntl.bts.gov/view/dot/54172

Miaou, S.-P., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention*, *25*(6), 689–709. https://doi.org/10.1016/0001-4575(93)90034-T

Miaou, S.-P., Song, J. J., & Mallick, B. K. (2003). *Roadway Traffic Crash Mapping: A Space-Time Modeling Approach | Bureau of Transportation Statistics*. https://www.bts.gov/archive/publications/journal_of_transportation_and_statistics/volume_06_number_01/paper_03/index

Mieno, T. (2022). *3.4 Spatial Intersection (transformative join) | R as GIS for Economists*. https://tmieno2.github.io/R-as-GIS-for-Economists/spatial-intersection-transformative-join.html

Miranda-Moreno, L. F., Morency, P., & El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. *Accident Analysis & Prevention*, *43*(5), 1624–1634. https://doi.org/10.1016/j.aap.2011.02.005

Morrison, C. N., D'Ambrosi, G., Kamb, A., MacManus, K., Rundle, A. G., & Humphreys, D. K. (2021). Rideshare Trips and Alcohol-Involved Motor Vehicle Crashes in Chicago. *Journal of Studies on Alcohol and Drugs*, *82*(6), 720–729.

Morrison, C. N., Jacoby, S. F., Dong, B., Delgado, M. K., & Wiebe, D. J. (2018). Ridesharing and Motor Vehicle Crashes in 4 US Cities: An Interrupted Time-Series Analysis. *American Journal of Epidemiology*, *187*(2), 224–232. https://doi.org/10.1093/aje/kwx233

Morrison, C. N., Kirk, D. S., Brazil, N. B., & Humphreys, D. K. (2022). Ride-Hailing and Road Traffic Crashes: A Critical Review. *American Journal of Epidemiology*, kwac033. https://doi.org/10.1093/aje/kwac033

Morrison, C. N., Mehranbod, C., Kwizera, M., Rundle, A. G., Keyes, K. M., & Humphreys, D. K. (2020). Ridesharing and motor vehicle crashes: A spatial ecological case-crossover study of trip-level data. *Injury Prevention*, injuryprev–2020–043644. https://doi.org/10.1136/injuryprev-2020-043644

Multimodal Community Advisory Committee. (2018, June 27). *Austin Strategic Mobilit Plan Status Update*. https://austintexas.gov/sites/default/files/files/Transportation/ASMP/ASMP_MCAC_27062018.pdf

NAIC. (2015). *Transportation Network Company Insurance Principles for Legislators and Regulators* (White Paper White Paper). The National Association of Insurance Commissioners; https://naic.org/store/free/TNC-OP-15.pdf. https://naic.org/store/free/TNC-OP-15.pdf

NAIC. (2016). *2016 Proceedings of the NAIC*. https://naic.soutronglobal.net/Portal/Public/en-US/RecordView/Index/12119

Nazif-Munoz, J. I., Batomen, B., & Nandi, A. (2022). Does ridesharing affect road safety? The introduction of Moto-Uber and other factors in the Dominican Republic. *Research in Globalization*, *4*, 100077. https://doi.org/10.1016/j.resglo.2021.100077

NHTSA. (2022a). *Distracted Driving | NHTSA* [Text]. National Highway Transportation Safety Admnistration. https://www.nhtsa.gov/risky-driving/distracted-driving

NHTSA. (2022b). *2020 ALCOHOL-IMPAIRED DRIVING Traffic Safety Fact Sheet-1.pdf*. NHTSA.

NHTSA. (2022c). *Traffic Safey Facts Annual Report, June 2022*. NHTSA. https://cdan.nhtsa.gov/tsftables/National%20Statistics.pdf

NHTSA. (2022d). *Early Estimate of Motor Vehicle Traffic Fatalities for the First Half (January – June) of 2022.pdf*.

NYC Department of Transportation. (2018). *New York City Mobility Report*. NYC Department of Transportation. https://www.nyc.gov/html/dot/downloads/pdf/mobility-report-2018-print.pdf

OECD. (2020). *Transport—Road accidents—OECD Data*. TheOECD. http://data.oecd.org/transport/road-accidents.htm

Overview | Geocoding API. (2022, July 14). Google Developers. https://developers.google.com/maps/documentation/geocoding/overview

Patten, C. J. D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: Cognitive workload and attention resource allocation. *Accident Analysis & Prevention*, *36*(3), 341–350. https://doi.org/10.1016/S0001-4575(03)00014-9

Peck, J. L. (2017). *New York City Drunk Driving After Uber*. 41.

Pernia, J., Lu, J. J., & Peng, H. (2004). *SAFETY ISSUES RELATED TO TWO-WAY LEFT-TURN LANES* (Final Technical Report). Article Final Technical Report. https://trid.trb.org/view/697941

Qin, X., Ivan, J. N., & Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis & Prevention*, *36*(2), 183–191. https://doi.org/10.1016/S0001-4575(02)00148-3

Quddus, M. (2013). Exploring the Relationship Between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS. *Journal of Transportation Safety & Security*, *5*(1), 27–45. https://doi.org/10.1080/19439962.2012.705232

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention*, *40*(4), 1486–1497. https://doi.org/10.1016/j.aap.2008.03.009

Rauli Susmel. (2022). *Count Data Models*. https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf

Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, *45*, 168–178. https://doi.org/10.1016/j.tranpol.2015.10.004

Retallack, A. E., & Ostendorf, B. (2019). Current Understanding of the Effects of Congestion on Traffic Accidents. *International Journal of Environmental Research and Public Health*, *16*(18), 3400. https://doi.org/10.3390/ijerph16183400

Richtel, M. (2014, December 21). Distracted Driving and the Risks of Ride-Hailing Services Like Uber. *Bits Blog*. https://bits.blogs.nytimes.com/2014/12/21/distracted-driving-and-the-risks-of-ride-hailing-services-like-uber/

Rodriguez, J. F. (2020, January 27). *Uber and Lyft pickups moving to side streets as Better Market Street project takes effect*. San Francisco Examiner. https://www.sfexaminer.com/archives/uber-and-lyft-pickups-moving-to-side-streets-as-better-market-street-project-takes-effect/article_cfe46ee0-66ec-5467-a8a1-9f6c12200aa3.html

Roy, S. (2019). *Quantifying the Impact of Transportation Network Companies (TNCs) on Traffic Congestion in San Francisco* [University of Kentucky Libraries]. https://doi.org/10.13023/ETD.2019.186

Roy, S., Cooper, D., Mucci, A., Sana, B., Chen, M., Castiglione, J., & Erhardt, G. D. (2020). Why is traffic congestion getting worse? A decomposition of the contributors to growing congestion in San Francisco-Determining the Role of TNCs. *Case Studies on Transport Policy*. https://doi.org/10.1016/j.cstp.2020.09.008

Salifu, M., & Ackaah, W. (2012). Under-reporting of road traffic crash data in Ghana. *International Journal of Injury Control and Safety Promotion*, *19*(4), 331–339. https://doi.org/10.1080/17457300.2011.628752

San Francisco County Transportation Authority. (2018). *TNCs and Congestion*. https://www.sfcta.org/emerging-mobility/tncs-and-congestion

Schaller, B. (2001). *Taxi and Livery Crashes in New York City, 1990-99*. Schaller Consulting. http://www.schallerconsult.com/taxi/crash.pdf

Schaller, B. (2017). *Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City*. Schaller Consulting. http://www.schallerconsult.com/rideservices/unsustainable.pdf

Schaller, B. (2018). *The New Automobility: Lyft, Uber and the Future of American Cities*. https://trid.trb.org/view/1527868

Schimek, P. (2018). Bike lanes next to on-street parallel parking. *Accident Analysis & Prevention*, *120*, 74–82. https://doi.org/10.1016/j.aap.2018.08.002

Sciortino, S., Vassar, M., Radetsky, M., & Knudson, M. M. (2005). San Francisco pedestrian injury surveillance: Mapping, under-reporting, and injury severity in police and hospital records. *Accident Analysis & Prevention*, *37*(6), 1102–1113. https://doi.org/10.1016/j.aap.2005.06.010

Seth, M. by. (2020a, January 7). *Safety Report From Uber Leaves Out Most Accidents*. San Francisco Public Press. https://sfpublicpress.org/safety-report-from-uber-leaves-out-most-accidents/

Seth, M. by. (2020b, January 8). *California Agency Is Hiding Uber and Lyft Accident Reports*. San Francisco Public Press. https://sfpublicpress.org/california-agency-is-hiding-uber-and-lyft-accident-reports/

Seth, M. by. (2020c, January 20). *Officials Demand Disclosure of Ride-Hail Accident Data*. San Francisco Public Press. https://sfpublicpress.org/officials-demand-disclosure-of-ride-hail-accident-data/

Seth, M. by. (2020d, January 29). *Commission Says Ride-Hailing Secrecy Footnote to Be Addressed Soon*. San Francisco Public Press. https://sfpublicpress.org/commission-says-ride-hailing-secrecy-footnote-to-be-addressed-soon/

SF-CHAMP Modeling. (2002, October 1). SFCTA. https://www.sfcta.org/sf-champ-modeling

SFCTA. (2015). *State of Vision Zero 2015*.

SFCTA. (2017). *TNCs Today: A Profile of San Francisco Transportation Network Company Activity*. https://www.sfcta.org/projects/tncs-today

SFMTA. (2017). *VISION ZERO SAN FRANCISCO Two-Year Action Strategy 2017–2018*.

Shefer, D. (1994). Congestion, air pollution, and road fatalities in urban areas. *Accident Analysis & Prevention*, *26*(4), 501–509. https://doi.org/10.1016/0001-4575(94)90041-8

Shefer, D., & Rietveld, P. (1997). Congestion and Safety on Highways: Towards an Analytical Model. *Urban Studies*, *34*(4), 679–692. https://doi.org/10.1080/0042098975970

Solomon, D. (1964). *Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle*. https://safety.fhwa.dot.gov/speedmgt/ref_mats/fhwasa1304/2_40.htm

Song, J. J., Ghosh, M., Miaou, S., & Mallick, B. (2006). Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis*, *97*(1), 246–273. https://doi.org/10.1016/j.jmva.2005.03.007

Stewart, DE. (1998). *Methodological Approaches for the Estimation, Evaluation, Interpretation, and Accuracy Assessment of Road Travel 'Basic Risk', 'Relative Risk,' and 'Relative Risk Odds-Ratio' Performance Measure Indicators: A 'Risk Analysis and Evaluation System Model' for Measuring, Monitoring, Comparing, and Evaluating the Level(s) of Safety on Canada's Roads and Highways"*.

Stewart, T. (2022). *Overview of motor vehicle crashes in 2020* (DOT HS 813 266; p. 43). National Highway Traffic Safety Administration.

Strayer, D. L., & Drews, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human Factors*, *46*(4), 640–649. https://doi.org/10.1518/hfes.46.4.640.56806

Tarko, A. P., Sinha, K. C., & Farooq, O. (1996). Methodology for Identifying Highway Safety Problem Areas. *Transportation Research Record*, *1542*(1), 49–53. https://doi.org/10.1177/0361198196154200108

Tay, R., & Choi, J. (2016). Factors associated with crashes involving taxi owners and non-owners: A case of moral hazard and adverse selection? *Accident Analysis & Prevention*, *87*, 78–82. https://doi.org/10.1016/j.aap.2015.11.028

Tirachini, A. (2019). Ride-hailing, travel behaviour and sustainable mobility: An international review. *Transportation*. https://doi.org/10.1007/s11116-019-10070-2

Transportation Injury Mapping System (TIMS). (2022). Transportation Injury Mapping System (TIMS). https://tims.berkeley.edu/about.php

Treffner, P. J., & Barrett, R. (2004). Hands-free mobile phone speech while driving degrades coordination and control. *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*(4), 229–246. https://doi.org/10.1016/j.trf.2004.09.002

Troglauer, T., Hels, T., & Christens, P. F. (2006). Extent and variations in mobile phone use among drivers of heavy vehicles in Denmark. *Accident Analysis & Prevention*, *38*(1), 105–111. https://doi.org/10.1016/j.aap.2005.07.008

Uber. (2017). *Uber's US Safety Report*. https://www.uber.com/us/en/about/reports/us-safety-report/

Uber. (2019). *Uber Safety Report 2019-2020*. Uber Technologies.

Uber. (2023a). *Driver Rewards with Uber Pro*. Uber. https://www.uber.com/au/en/drive/uber-pro/

Uber. (2023b). *How Star Ratings Work | Driver App*. Uber. https://www.uber.com/us/en/drive/basics/how-ratings-work/

Uber & MADD. (2015). *More options. Shifting mindsets. Driving better choices.* Uber. https://newsroom.uber.com/wp-content/uploads/madd/

U.S. Census Bureau. (2022, March 24). *Resident Population in San Francisco County/city, CA*. FRED, Federal Reserve Bank of St. Louis; FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/CASANF0POP

USDOT. (2022, October). *What Is a Safe System Approach? | US Department of Transportation*. https://www.transportation.gov/NRSS/SafeSystem

Vey, A. H. (1939). Improvements to Reduce Traffic Accidents. *Transactions of the American Society of Civil Engineers*, *104*(1), 474–484. https://doi.org/10.1061/TACEAT.0005186

Violanti, J. M. (1998). Cellular phones and fatal traffic collisions. *Accident Analysis & Prevention*, *30*(4), 519–524. https://doi.org/10.1016/S0001-4575(97)00094-8

Violanti, J. M., & Marshall, J. R. (1996). Cellular phones and traffic accidents: An epidemiological approach. *Accident Analysis & Prevention*, *28*(2), 265–270. https://doi.org/10.1016/0001-4575(95)00070-4

Vogt, A. (1999). *Crash Models for Rural Intersections: Four-lane by Two-lane Stop-controlled and Two-lane by Two-lane Signalized* (FHWA-RD-99-128). U.S. DOT. https://rosap.ntl.bts.gov/view/dot/35701/dot_35701_DS1.pdf

Vogt, A. & J.G, Bared. (1998). *Accident Models for Two-lane Rural Roads: Segments and Intersections*. https://trafficlab.eu/wp-content/uploads/Accident-models-for-Two-Lane.pdf

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident; Analysis and Prevention*, *60*, 71–84. https://doi.org/10.1016/j.aap.2013.07.030

Ward, J. W., Michalek, J. J., & Samaras, C. (2021). Air Pollution, Greenhouse Gas, and Traffic Externality Benefits and Costs of Shifting Private Vehicle Travel to

Ridesourcing Services. *Environmental Science & Technology*, *55*(19), 13174–13185. https://doi.org/10.1021/acs.est.1c01641

Washington, S., Metarko, J., Fomunung, I., Ross, R., Julian, F., & Moran, E. (1999). An inter-regional comparison: Fatal crashes in the southeastern and non-southeastern United States: preliminary findings. *Accident Analysis & Prevention*, *31*(1–2), 135–146. https://doi.org/10.1016/S0001-4575(98)00055-4

Watson, A., Watson, B., & Vallmuur, K. (2015). Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accident Analysis & Prevention*, *83*, 18–25. https://doi.org/10.1016/j.aap.2015.06.011

WHO. (2008). *Global Road Safety Partnership: Speed management: A road safety manual for decision-makers and practitioners. 2008*.

WHO. (2018, December). Why the U.S. Leads the Developed World on Traffic Deaths. *Streetsblog USA*. https://usa.streetsblog.org/2018/12/13/why-the-u-s-trails-the-developed-world-on-traffic-deaths/

WHO. (2020, February 7). *Road traffic injuries*. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. https://ipcig.org/evaluation/apoio/Wooldridge%20-%20Cross-section%20and%20Panel%20Data.pdf

Xu, P., Huang, H., Dong, N., & Abdel-Aty, M. (2014). Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. *Accident; Analysis and Prevention*, *70*, 110–120. https://doi.org/10.1016/j.aap.2014.02.012

Yamamoto, T., Hashiji, J., & Shankar, V. N. (2008). Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident; Analysis and Prevention*, *40*(4), 1320–1329. https://doi.org/10.1016/j.aap.2007.10.016

Yasmin, S., & Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis & Prevention*, *59*, 506–521. https://doi.org/10.1016/j.aap.2013.06.040

Ye, F., & Lord, D. (2011). Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit. *Transportation Research Record: Journal of the Transportation Research Board*, *2241*(1), 51–58. https://doi.org/10.3141/2241-06

Zegeer, C.V., Stewart, J.R., Huang, H.H., & Lagerwey, P.A. (2002). *Safety Effects of Marked Vs. Unmarked Crosswalks at Uncontrolled Locations: Executive Summary and Recommended Guidelines* (FHWA-RD-01-075). FHWA, U.S. DOT. https://www.fhwa.dot.gov/publications/research/safety/04100/04100.pdf

VITA

VEDANT SHRINIWAS GOYAL

EDUCATION

University of Kentucky, Lexington, Kentucky

Doctor of Philosophy in Civil Engineering anticipated May 2023

Dissertation Title: Do Transportation Network Companies affect road safety outcomes: a spatially detailed analysis in San Francisco
Advisor: Dr. Gregory Erhardt

University of Leeds, Leeds, United Kingdom

Master of Science (Eng.) Transport Planning and Engineering September 2011

Symbiosis International University, Pune, India

Master of Science Geoinformatics June 2008

Shivaji University Kolhapur, India

Bachelor of Science, Civil Engineering May 2006

EXPERIENCE

Lawrence Berkeley National Laboratory, Berkeley, California

Transportation Systems Postdoc February 2023 - present

Department of Civil Engineering, University of Kentucky, Lexington, Kentucky

Graduate Research Assistant August 2019-January 2023

Bloomberg Philanthropies Initiatives for Global Road Safety (BIGRS), Mumbai, India

Road Design and Transportation Coordinator January 2016 – August 2019

Deloitte Touché Tohmatsu India Private Limited Delhi, India

Manager May 2015 – January 2016

Deutsche Gesellschaft Für Internationale Zusammenarbeit (GIZ) Delhi, India

Urban Transport Advisor April 2012 – May 2015

Reliance Infrastructure Private Limited (REL) Mumbai, India

Assistant Manager January 2008 – February 2012

PUBLICATIONS

- Goyal, V., Erhardt, G., Kressner, J., Watkins, K., Berrebi, S., Brakewood “Simulating the Effect of Strategies to Increase Transit Ridership by Reallocating Bus Service.”, Poster Presentation, 2022, Transportation Research Board, Washington, D.C.
- Goyal, V., Erhardt, G., Hoque, J.M., Berrebi, S., Watkins, K., Brakewood, “Why has Public Transit Ridership Declined in the United States, Transportation Research Part A: Policy and Practice, https://doi.org/10.1016/j.tra.2022.04.006.
- Goyal, V., Watkins, K., Berrebi, S., Erhardt, G., Hoque, J.M., Brakewood, C., Ziedan, A., Darling, W., Hemily, B., and Kressner, J. 2021. “Recent Decline in Public Transportation Ridership: Analysis, Causes, and Responses. TCRP Research Report 231.” Transportation Research Board, Washington
- Goyal, V., Erhardt, G., Stamatiadis, N., Bollinger, C., “Do transportation network companies increase or decrease road crashes? Evidence from San Francisco”, working paper