

Misconduct on Public Transit: An Exploratory Analysis Using the Comments Formerly Known as Tweets

Egbe-Etu Etu, PhD
Jordan Larot

Asha Weinstein Agrawal, PhD
Dang Minh Nhu Nguyen

Imokhai Theophilus Tenebe, PhD



MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the [Mineta Consortium for Transportation Mobility \(MCTM\)](#) and the [Mineta Consortium for Equitable, Efficient, and Sustainable Transportation \(MCEEST\)](#) funded by the U.S. Department of Transportation, the [California State University Transportation Consortium \(CSUTC\)](#) funded by the State of California through Senate Bill I and the Climate Change and Extreme Events Training and Research (CCEETR) Program funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. This document is disseminated in the interest of information exchange. MTI's research is funded, partially or entirely, by grants from the U.S. Department of Transportation, the U.S. Department of Homeland Security, the California Department of Transportation, and the California State University Office of the Chancellor, whom assume no liability for the contents or use thereof. This report does not constitute a standard specification, design standard, or regulation.

Report 23-36

Misconduct on Public Transit: An Exploratory Analysis Using the Comments Formerly Known as Tweets

Egbe-Etu Etu, PhD

Asha Weinstein Agrawal, PhD

Imokhai Theophilus Tenebe, PhD

Jordan Larot

Dang Minh Nhu Nguyen

November 2023

A publication of the
Mineta Transportation Institute
Created by Congress in 1991
College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 23-36	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Misconduct on Public Transit: An Exploratory Analysis Using the Comments Formerly Known as Tweets		5. Report Date November 2023	
		6. Performing Organization Code	
7. Authors Egbe-Etu Etu, PhD ORCID: 0000-0003-0457-6296 Asha Weinstein Agrawal, PhD ORCID: 0000-0003-2328-0263 Imokhai Tenebe, PhD ORCID: 0000-0002-5862-6511 Jordan Larot ORCID: 0009-0006-8924-5363 Dang Minh Nhu Nguyen ORCID: 0000-0002-1025-1212		8. Performing Organization Report CA-MTI-2317	
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		10. Work Unit No.	
		11. Contract or Grant No. 69A3551747127	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplemental Notes 10.31979/mti.2023.2317			
16. Abstract This project developed a simple methodology for using Twitter data to explore public perceptions about misconduct on public transit in California. The methodology allows future researchers to analyze tweets to answer questions such as: How frequent are tweets related to assault, abuse, or other misconduct on public transit? What concerns arise most frequently? What are the types of behaviors discussed? We collected and analyzed data from Twitter posts in California about various types of public transit misconduct from January 2020 to March 2023 to identify the nature and frequency of reported misconduct. Our findings reveal that harassment, uncivil behavior, and assault are the commonly reported concerns; far fewer tweets mention obscene behavior, threats, or theft. It appears that at times the victims had been targeted on the basis of their race, gender, or sexual identity, or because they were transit employees. The tweets indicate that both genders are victimized, though women were targeted more often than men (57.5% vs. 42.5%). As for the alleged perpetrators of transit misconduct, more than three-quarters were male (78%). Transit agencies and researchers can use the results of these analyses to strategically improve safety measures for the benefit of passengers and transit operators.			
17. Key Words Transit, Safety, Security, Tweets, Theft		18. Distribution Statement No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 28	22. Price

Copyright © 2023

by **Mineta Transportation Institute**

All rights reserved.

DOI: 10.31979/mti.2023.2317

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219
Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2317

ACKNOWLEDGMENTS

We would like to thank Akshath Mirukula, a high school student with a research interest in public transportation, who assisted the authors in cleaning the data. This work was supported by the U.S. Department of Transportation and Mineta Transportation Institute at San José State University (grant #: 69A3551747127). The funding agency had no role in the study design, analysis, or publication decision.

CONTENTS

Acknowledgments	vi
List of Figures.....	vii
List of Tables.....	viii
Executive Summary	1
1. Introduction.....	3
2. Methods	5
2.1 Data Extraction	5
2.2 Data Preprocessing.....	6
2.3 Data Cleaning and Analysis	6
3. Results	8
3.1 Tweets by Time.....	8
3.2 Tweets by Transit Mode, Transit Agency, and County.....	8
3.3 Tweets by Gender	10
3.4 Tweets by Type of Public Transit Misconduct	10
4. Summary & Conclusions.....	14
Bibliography	17
About the Authors.....	18

LIST OF FIGURES

Figure 1. Most Commonly Mentioned Transit Agencies	9
Figure 2. Frequency of Tweets per County in California	10

LIST OF TABLES

Table 1. Public Transit Misconduct Classification Scheme 7

Executive Summary

This project developed a methodology to use Twitter data to explore public perceptions of misconduct in public transit systems in California. This simple methodology allows researchers to analyze tweets to answer questions such as: how frequent are tweets related to assault, abuse, or other misconduct on public transit? What concerns arise most frequently? What types of behavior are discussed? Transit operators and researchers can use the results of the analyses to strategically improve safety measures for the benefit of passengers and operations staff.

We extracted three years and three months' worth of tweets originating in California that include language describing harassment, assault, or other forms of uncivil behavior on public transit. The dataset includes descriptions of both incidents where the writer was the victim and incidents where the writer described events happening to others. The dataset was constructed in two steps: data extraction, to retrieve relevant tweets for the project, the named transit agencies associated with the tweets, and the geographic location of the tweets, and data preprocessing, to clean the data. To analyze the tweets, we labeled them using six categories: assault, harassment, theft, threat, obscenity, and uncivil behavior; and we conducted a qualitative assessment to explore the nature of the incidents described.

From the 317 tweets, we learned that the great majority describe events in the urban centers of Los Angeles and San Francisco, and that both buses and rail were frequently mentioned. The number of tweets varied considerably by year, with fewer in 2021 than in the years before and after. Further, a qualitative analysis of the tweets revealed that harassment, uncivil behavior, and assault are common concerns. Some victims were clearly targeted on the basis of their race, gender, or sexual identity or because they were a transit employee, such as a bus operator. The qualitative analysis also showed that both genders are victimized, though women were targeted more often than men (57.5% vs. 42.5%). Note that some tweets were coded for both genders.

Finally, we identified the following strengths and limitations of the methodology:

- The greatest value of the analysis comes from analyzing the specific content of the relevant tweets. The tweets in our dataset can be considered mini-stories that each describe an event of misconduct, allowing researchers to pinpoint details about the kinds of people who are victims (or perpetrators) and the specific misconduct behaviors. The detailed accounts provide powerful evidence to illustrate the nature of the crimes and misbehavior that impact transit riders, thus providing transit agencies with a better understanding of the problems they are working to address.
- The methodology can be used to learn about experiences in large metro areas or to explore state-wide patterns, but not to evaluate experiences on any specific smaller transit agency. We extracted only a modest number of tweets statewide, and tweets were concentrated in

two urban areas with some of the highest transit ridership in California: Los Angeles and San Francisco Counties.

- Some level of manual analysis is required to pull relevant tweets from the full set of tweets originally extracted from Twitter (now referred to as X). Most tweets that were extracted using our combinations of the relevant keywords do not describe first-hand experiences of transit misconduct. While machine learning offers the potential for future data extraction and cleaning refinements, there will inevitably be many tweets that are off-topic for the study. For clarity, we will continue to refer to the platform as Twitter in this analysis, as it was called Twitter at the time of data extraction and analysis.

1. Introduction

This project developed a methodology to use Twitter data to explore public perceptions of misconduct in public transit systems in California. This simple methodology will allow researchers to analyze tweets to answer questions such as: how frequent are tweets related to assault, abuse, or other misconduct on public transit? What concerns arise most frequently? What are the types of behaviors discussed? Transit operators and researchers can use the results of these analyses to strategically improve safety measures for the benefit of passengers and operations staff.

In the United States, public transportation is an essential infrastructure, with a well-developed system that includes buses, light rail, trains, shuttles, and ferries. Public transit is a lifeline for millions of citizens, connecting them to people, places, and possibilities (APTA, 2023). Many Americans see public transit as a crucial form of transportation because of its relatively low cost, as compared to vehicle ownership and use (Polzin, 2016; Beirão and Cabral, 2007; Litman, 2015). Public transit also provides access to employment sites, schools, and essential destinations for those who cannot drive themselves due to disabilities or health conditions, or because they are too young. Furthermore, public transit benefits society at large by reducing traffic congestion and emissions.

However, despite the numerous benefits of public transportation, it poses some significant challenges for riders. Regarding safety and security, some of the most pressing issues riders face are the possibilities of assault, verbal harassment, bullying, or abuse (Ding et al., 2020). “Assault” refers to intentional physical harm or the threat of physical harm inflicted on another person in a public transit setting. It can involve acts of violence, such as punching, hitting, spitting, or pushing, and intimidation, such as threatening behavior or gestures. These confrontations can occur between passengers or between passengers and transit operators. Incidents vary greatly in severity, ranging from minor unpleasant encounters to severe, life-threatening events.

Harassment, assault, and other crimes have a profound impact on the safety and the physical and mental well-being of passengers. At a minimum, incidents create an unpleasant and stressful environment for victims and also for witnesses. Some riders who lose trust in public transit as a safe travel option limit their use to certain times or locations, and others stop riding altogether.

Reported assault and harassment events on public transit have significantly increased in recent years. An analysis of the Los Angeles Police Department data shows that between 2020 and 2022, there was a 700-case rise in the total number of crimes committed in the city’s public transportation system, with an increase of about 300 incidents involving assault, harassment, bullying, or abuse (LAPD, 2023). These incidents range from verbal harassment to physical assaults with severe consequences for victims (e.g., emotional trauma and physical injury). Additionally, many victims lose confidence in using public transportation as a result. Victims of assault, harassment, and bullying on public transportation in California come from diverse backgrounds and can include anyone who uses the system. However, certain groups may be at greater risk than others, including the elderly, women, Hispanics, gender minorities, and low-income persons (Garcia et al., 2022;

Lubitow et al., 2020; Levin and Wachs, 1986). These findings underscore the need for continued efforts to address and prevent incidents of assault, harassment, and bullying on public transportation to ensure the safety and security of all passengers.

2. Methods

We created a dataset for analysis of three years and three months' worth of tweets originating in California that include language describing harassment, assault, or uncivil behavior on public transit. The dataset includes tweets describing cases where the writer was the victim and tweets describing experiences where the writer described events happening to others. The dataset was constructed in two steps: data extraction, which identified relevant tweets for the project, and data preprocessing, which involved cleaning the data. To analyze the tweets, we labeled them using six categories: assault, harassment, theft, threat, obscenity, and uncivil behavior.

2.1 Data Extraction

We extracted the dataset using the Twitter Application Programming Interface (API) for academic research, searching from January 1, 2020, to March 31, 2023. The search was limited to tweets originating from California. This process generated a set of 3,388 tweets. The extracted tweets include metadata information such as an identification number, tweet date, time, and language. The Twitter API provided us with each tweet's language and geographic information about where each tweet originated from (the city, state, latitude, and longitude).

The next step was to narrow down the tweets to those likely to be relevant to the project. This was done by searching for tweets that contain at least one word or term from each of two distinct sets of keywords:

- *Keywords about public transit:* This set contains 50 words or phrases that either describe a transit mode or name one of California's ten largest public transit agencies, as measured by annual ridership. The modal keywords and phrases include "bus," "train," "ferry," "trolley," and "subway." The named agencies are Santa Clara Valley Transportation Authority, Los Angeles County Metropolitan Transportation Authority (Metro), San Francisco Municipal Transportation Agency (SFMTA), Alameda-Contra Costa Transit District (AC Transit), Orange County Transportation Authority, San Diego Metropolitan Transit System, Sacramento Regional Transit District, and Bay Area Rapid Transit (BART). We included the formal agency name and commonly used abbreviations for each agency. For example, for the San Francisco Metropolitan Transportation Authority, we searched for tweets that contained the formal name "Muni" or "SFMTA."
- *Keywords about assault, harassment, or uncivil behaviors:* This set contains 236 words such as "fight," "spit," "threw," "shoved," "catcall," "rape," "theft," "abuse," "bully," "grope," "porn," and "punch." We included multiple forms of each word (or concept). For verbs, the keywords include multiple tenses (e.g., "punch," "punches," "punched," "punching"). For nouns, the keywords include singular and plural forms (e.g., "theft" and "thefts").

2.2 Data Preprocessing

We first processed and cleaned the tweets by removing special characters, digits, emoticons, and hyperlinks so that the dataset contained only words. This process is necessary because when the tweets are downloaded, the API converts all symbols to Unicode format, making the raw data hard to process and analyze. We tackled this issue by implementing a data cleaning function in our Python script that reconverts these Unicode symbols to their original, recognizable format, ensuring the tweets are ready for further analysis. For example, “wasn,Äôt” was replaced with “wasn’t,” as in this example:

Original Tweet: *He was making blatant threats towards another passenger that was getting on the same bus and that poor man wasn,Äôt even responding to any of his threats or rude comments. The man was yelling, cussing, threatening the entire bus ride. I complained to the @FlixBus worker and,Ä¶*

Edited Tweet: *He was making blatant threats towards another passenger that was getting on the same bus and that poor man wasn’t even responding to any of his threats or rude comments. The man was yelling, cussing, threatening the entire bus ride. I complained to the @FlixBus worker*

2.3 Data Analysis

We manually reviewed the 3,388 cleaned tweets to identify those directly relevant to the study. Relevant tweets were defined as those describing a personal experience with transit misconduct, whether as a victim or witness. We defined “transit misconduct” as any one of six types of behaviors: (1) assault, (2) theft, (3) harassment, (4) obscenity, (5) threat, and (6) uncivil behavior. See Table 1 for a description of each. These classifications were derived from the literature on transit harassment and safety (Ding et al., 2020; Ceccato et al., 2022).

Table 1. Public Transit Misconduct Classification Scheme

Category	Description
Assault	A physical attack
Harassment	Behavior that creates an unpleasant or hostile situation by uninvited and unwelcome verbal conduct, touching, gestures, or body language
Obscenity	Offensive sexual comments or behaviors, including viewing pornographic content or indecent exposure
Theft	The act of stealing or unlawful taking of property
Threat	A stated intention to inflict harm, injury, or damage
Uncivil behavior	Scary or offensive behaviors that the Twitter user did not perceive as directed at anyone in particular, such as playing loud music or shouting randomly

After this labeling process, we identified 317 tweets that meet the study criteria as relevant for analysis. The other tweets were discarded for a wide variety of reasons. The sorts of tweets that make up the bulk of the discarded tweets are tweets that complain about the quality of the transit service (i.e., late buses), tweets that do not describe an actual case of misconduct (e.g., “threw him under the bus”), and tweets that comment on stories the writer had read or heard about in the news or on social media.

We employed a triad of analytical methods to delve into the nature and extent of public transit misconduct. We used descriptive analysis to summarize and categorize the characteristics of the misconduct tweets, looking at the misconduct categories of the tweets, the genders of the victims and perpetrators, the named transit agencies associated with the tweets, and the geographic location of the tweets.

3. Results

The data extraction and cleaning processes described above resulted in 317 tweets about public transit misconduct in California. The dataset was analyzed to explore four topics:

- Trends in the number of tweets over time
- The location of the misconduct described (mode of transit, agency, and county)
- The genders of the victims and perpetrators
- The types of misconduct described

3.1 Tweets by Time

The frequency of tweets varied modestly by year. The largest number of tweets comes from 2020 (124, or 44% of the full dataset), compared to a slightly lower number in 2022 (99, or 35% of the total). An average of eight tweets were posted monthly, with February witnessing peak activity (33 tweets). A weekday breakdown demonstrates that Saturday accumulated the highest tweet count (56), with the average weekday recording 45 tweets. The types of reported misconduct varied considerably more per year than the overall number of tweets, at least for certain types of behavior. For example, 2021 saw a considerable drop in tweets about uncivil behavior and harassment compared to the other two years.

3.2 Tweets by Transit Mode, Transit Agency, and County

It was possible to code almost all of the tweets (92%) according to a mode of public transportation. Slightly over half the tweets mention buses (52%, 154), and most of the rest mention trains (46%). Only a single tweet mentions ferry.

Our analysis also identified references to specific transit agencies (Figure 1). Eleven such agencies were mentioned, with BART, SFMTA, and Metro being the most commonly cited.

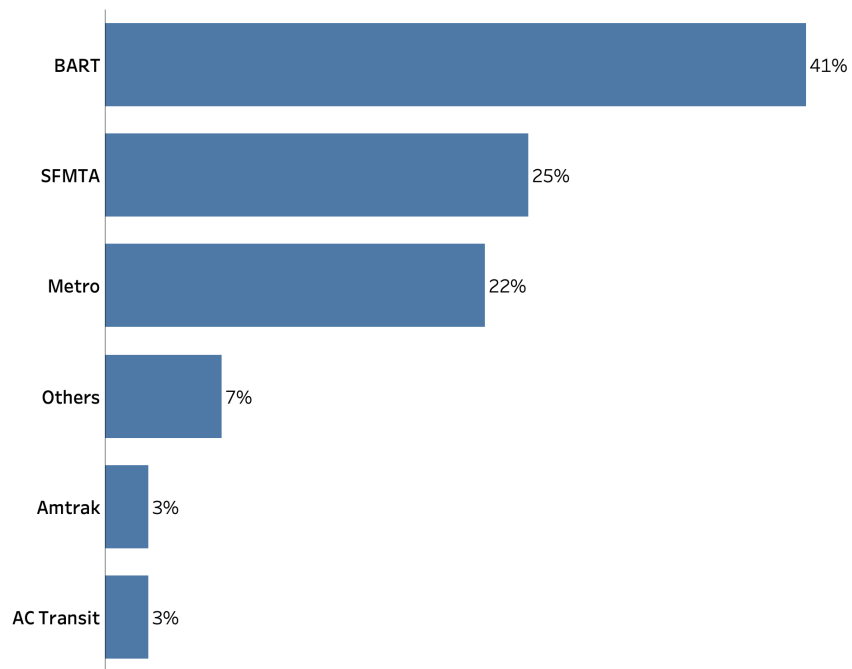


Figure 1. Most Commonly Mentioned Transit Agencies

Given that so many tweets mention BART, SFMTA, and LA Metro, it is not surprising that the majority of tweets (68%) originated from the Los Angeles and San Francisco County areas (Figure 2). Los Angeles County, contributing 146 tweets, emerged as the epicenter of public transit misconduct-related discussions on Twitter. San Francisco County accounts for the second-highest number of tweets (70).

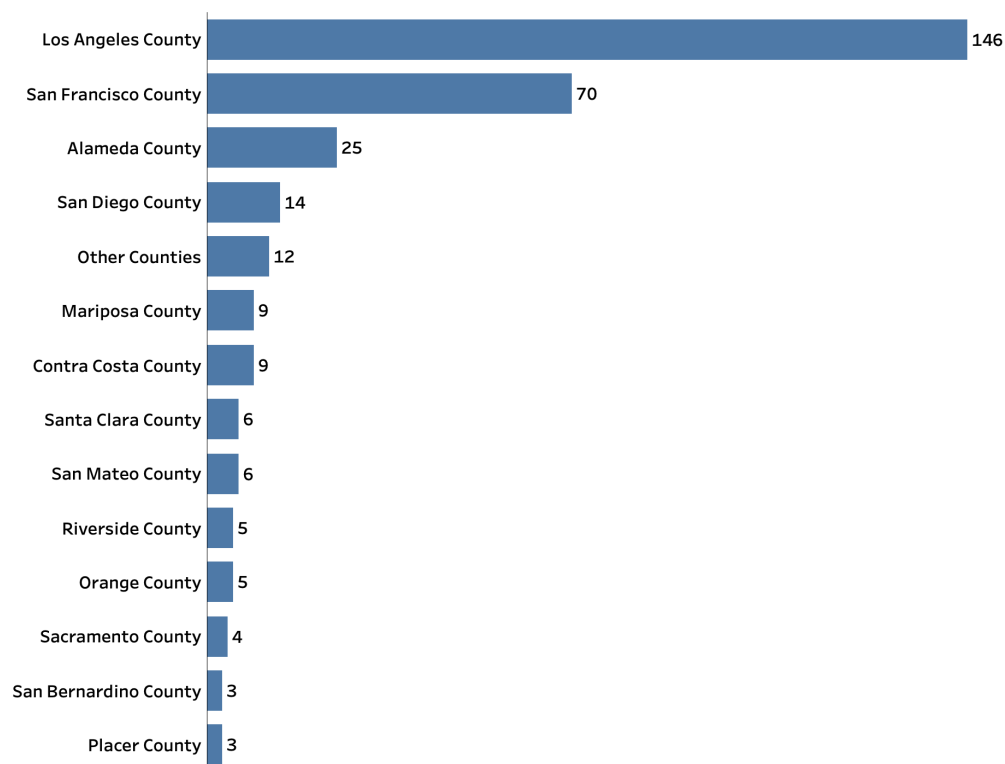


Figure 2. Frequency of Tweets per County in California

3.3 Tweets by Gender

Additionally, we explored the tweet content to identify the perceived gender of both victims and perpetrators. It was possible to assign gender to the victim in 30% of the tweets and to the perpetrator in 64% of the tweets. Among the tweets that specify the victim’s gender, 57% refer to a woman, 42% refer to a man, and 1% refer to both a woman and a man. As for the alleged perpetrators of transit misconduct, almost four out of five, or 78% (159 out of 203), were male. We did not observe tweets specifying genders other than male or female, but analysis of future tweets can explore whether this is possible to observe.

3.4 Tweets by Type of Public Transit Misconduct

Harassment, uncivil behavior, and assault were the most frequently mentioned misconduct behaviors, accounting for more than three-quarters of the tweets. A few tweets were classified as threats, obscene behaviors, or theft.

The following tweets illustrate the comments written about behaviors we classified as “harassment,” namely, behavior that creates an unpleasant or hostile situation by uninvited and unwelcome verbal conduct, touching, gestures, or body language. Specific types of harassing behaviors include yelling (G), sexual comments (E), racist language or name-calling (A, D), and staring (B, C). Some victims appear to have been targeted for their race or ethnicity (A), sexual orientation (F), or gender

(D). Many of the victims and perpetrators mentioned were bus operators (G), though more often the tweets describe the victims and perpetrators as passengers.

- A. On the train yesterday some guy yelled out to me, “Konnichiha, Konnichiha!” When I ignored it then he went on to say how Chinese ppl shouldn’t even be here or some shit, blah blah blah lmao. Little does he know, I’m neither Japanese nor Chinese. I’m Korean.
- B. Bart can be the fastest way to get around the Bay Area but this sucks when you got some creepy dude straight looking at you.
- C. There’s this old lady sitting right in front of me on the bus and she keeps staring at me.
- D. I went outside today to get groceries and a lady got in my face near the bus stop and called me a “stupid-ugly no-titty-having-ass-bitch” and then told me to have a nice day.
- E. On my way home from work in my uniform n Some drunk guy on the bus asked if i was beauty from beauty in the beast I got a sick creepy vibe, so I said no I’m the beast.
- F. Just got called a “fag” at the train station so I blew a kiss at them, hope that made him uncomfortable.
- G. Funny coincidence that I see a “@SFMTA is Hiring” on the same bus route with the driver who yelled at me for having my phone out while boarding.

Descriptions of assault, which we defined as a physical attack, are illustrated by the next set of tweets. Victims were groped (J), hit or beaten (H, K, L, M), and pushed (I, N). Some of the tweets suggest that the victims were targeted because of their gender (J) or race (I, N).

- H. During evening rush hour on the crowded @metrolosangeles station at 7th St/Metro Center station in DTLA, a woman directly next to me was assaulted by a man who immediately fled the station. She was knocked to the ground and taken away in an ambulance. It looked to be 100% random.
- I. Just witnessed a girl push an elderly Asian guy out of her way so she could get off the bus before him. #StopAsianHate
- J. So, on today, three men groping #Women against their permission.
- K. On multiple occasions I felt pretty unsafe. Example: On the A line a rider sitting next to me was smacked in the back of the head unprovoked by another rider. The Metro ambassadors were nice to see but honestly there aren’t enough for how shady/expansive the system is.

- L. The girl had a little brother who was also getting beat up on the bus for no reason. The bully got arrested and the victim's mother will press charges.
- M. San Francisco bus driver assaulted with bat over mask order.
- N. Welp, I just elbowed from a racist dude on the train of which I was doing not at all. He was wearing a ski mask, sunglass, and had his bike blocking the door. He fisted pumped a dude (who was nice to consult with me if I was okay or not after the racist left the train).

Many tweets described uncivil behaviors, which we define as frightening or offensive behaviors that the Twitter user did not perceive as directed at anyone. The following set of tweets illustrates the behaviors observed, which include playing loud music (P), putting feet on seats (O), strong body odor (S), shouting randomly (Q, R, S), and disrupting service (R). Some of the people are described in ways that suggest they were suffering from a mental illness or addiction (R, T).

- O. Here in my mid 30s, I have become the guy who tells strangers to get their "dirty feet" off the seats on public transit. Grandpa Pete will yell at you.
- P. This bus driver just yelled for someone to turn the volume down on their phone (playing obnoxious videos/music); I am about it. Should be a requirement on everything. #headphonesarerequiredALWAYS
- Q. Ah yes, the classic BART situation. A man yelling angrily.
- R. @sfmta_muni Aggressive passenger opened door while train was in motion, caused it to stop. He is now running back and forth, screaming for train to move.
- S. I need a bottle of axe, fabreeze, bath and body works or something, there are some stinky people on this ride. And a very loud angry dude yelling gibberish. I tip my hat to the brave souls who do public transit. This is why I bought a car the minute I could.
- T. This homeless dude on the metro 720 is rubbing his hands on the seats and then licking his hands.

Tweets related to obscenity, as illustrated by the list below, primarily describe passengers reading or watching pornography (U, V) or indecent exposure and public masturbation (W, X).

- U. This man sitting in front of me on the train is reading some porn novel. He's like 85 and white!
- V. Ayo, someone come get your grandpa. This guy's watching porn with the sound all the way up at the BUS STOP

W. Me to fellow crew scared to use public transport: oh, the public transport really isn't that bad in LA. I've done it loads!! Also, me: *witnesses two fights, gets flashed by a homeless bloke who is full on wanking on the train, then another guy who injects smack into his foot*

X. So, I'm on BART on my way into work finally. I sit in a seat at the back next to someone who I thought was sleeping. Halfway from Fruitvale to Lake Merritt I notice this man is staring at me. I slightly look over and notice he's rubbing his penis while staring at me.

Tweets about theft most commonly mention stolen phones (Y), though a handful of tweets describe other types of theft (Z).

Y. Was on the Muni today and a (I'm guessing) 12-year-old whipped my phone out of my hand and bolted out of the bus. Unlucky for him, I've been building my running stamina and bolted behind him through cars whizzing through on Geary.

Z. When I was in law school I saw a mass robbery at that train station.

Finally, tweets about threats (stated intentions to inflict harm) include the following examples:

AA. Can't stop thinking about how a man was standing over me and my friend screaming, "I will murder you!" and not a single person on the bus stood up for us. Not. One. Person.

BB. @SDGIS @metrolosangeles On the trolley, I once called MTS security because someone was making threats on the train and security popped on two stops later and grabbed the guy. Also, text security when I saw a seat either covered in urine or soda (couldn't tell which) the driver walk back and checked it.

4. Summary & Conclusions

This study was conducted to assess the feasibility of using tweet analysis to better understand the nature of misconduct occurring on public transit in California. From the 317 tweets, we learned that the great majority describe events in the urban centers of Los Angeles and San Francisco, and that buses and rail are frequently mentioned. The number of tweets varied considerably by year, with fewer in 2021 than in the years before and after.

Further, through a qualitative analysis of the tweets, we learned that harassment, uncivil behavior, and assault are the most commonly reported concerns; far fewer tweets mentioned obscene behavior, threats, or theft. It appears that at times the victims had been targeted on the basis of their race, gender, or sexual identity, or because they were transit employees, such as a bus operator. The tweets described both genders as victimized, though women were targeted more often than men (57.5% vs. 42.5%). As for the alleged perpetrators of transit misconduct, more than three-quarters were male (78%).¹

From these findings, we offer the following key conclusions about the feasibility of the methodology.

- *The greatest value of the analysis comes from analyzing the specific content of the relevant tweets.* Tweets are essentially mini-stories that each describe an event of misconduct, allowing researchers to pinpoint details about the kinds of people who are victims (or perpetrators) and the specific misconduct behaviors. The detailed accounts provide powerful evidence to illustrate the severe nature of the crimes and misbehavior and expand a transit agency's understanding of the types of behaviors and situations to be addressed. For example, while our findings indicate that a majority of the identified perpetrators were male, a noteworthy 20% were female. This suggests that, while males are more frequently identified as aggressors, misconduct is not solely confined to one gender.

As such, the tweets complement other research methods, including traditional passenger surveys or qualitative interviews. Passenger surveys provide important generalizable evidence about how many passengers have safety concerns but very few specific details about the events of concern. Conversely, interview studies can provide an excellent understanding of specific events, but cost and feasibility constraints typically limit the studies to learning from very few people. The Twitter analysis offers a middle ground: an inexpensive method to learn about the specific experiences reported by hundreds of people.

- The methodology can be used to learn about experiences in large metro areas or to explore state-wide patterns, but not to evaluate experiences on any specific smaller transit agency.

¹ The dataset did not contain tweets specifying genders other than male or female.

We extracted only a modest number of tweets statewide—fewer than expected, given that we collected more than three years’ worth of data for a state that has more transit riders than all but a handful of other states. Also, most tweets were concentrated in two urban areas with some of the highest transit ridership in California: Los Angeles and San Francisco Counties. The limited number of tweets about events on small transit systems makes it unlikely that the methodology would yield enough data for a meaningful analysis of a smaller system.

- Some level of manual analysis is required to pull relevant tweets from the full set of tweets originally extracted from Twitter. Most tweets that were extracted using our combinations of the relevant keywords do not describe first-hand experiences of transit misconduct. Although machine learning offers the potential for more efficient data extraction and cleaning, there will inevitably be many tweets that are off-topic for the study. For example, it is difficult to imagine that machine learning will be able to reliably identify that a quote using the colloquial expression “thrown under the bus” is not an example of transit misconduct.

To further develop the methodology and learn more about patterns of transit misconduct, we plan to take the following steps to address the limitations of the current process.

- It will be possible to identify more examples of transit misconduct within California by extracting tweets in other languages (e.g., Spanish) and by expanding the set of keywords with the public's informal names for specific transit systems.
- A second strategy to generate a larger and richer dataset is to apply the study methodology to other states with high transit ridership, such as New York and Illinois.
- Additional refinement of the misconduct categories would permit quantitative analysis to better determine the frequency of different types of behaviors. Many tweets describe negative behaviors that were easy to assign to one of the misconduct categories, but others were difficult to code with confidence. For example, a tweet that discusses someone being “pushed” might be an assault (if the victim fell) or harassment (if the victim was lightly pushed and felt threatened but was not at all injured). For future studies, we will refine the misconduct categories and develop more extensive training materials for the researchers who code the tweets.
- We will seek up-to-date information on the characteristics of people who tweet (and those who do not) to provide clearer context for the findings. Only a small proportion of people tweet in the United States, so it is important to acknowledge that the data is not representative of all transit riders.
- Part of the dataset for this study was collected during the peak of the COVID-19 pandemic, which could have uniquely influenced transit passenger interactions due to reduced

ridership, heightened social tensions, and altered perceptions of problematic behavior, such as increased sensitivity to personal space breaches. Future work should consider examining passenger interactions in both pre-pandemic and post-pandemic periods to discern the broader generalizability of our findings.

- New changes in Twitter’s terms of service, API rules, and data access policies may affect researchers’ ability to extract additional tweets. For example, changes at Twitter during 2023 limited the number of tweets researchers can extract, and the free academic accounts for researchers have been replaced with pricey subscription models. However, as of the writing of this report, transit agencies with registered and verified accounts on Twitter can extract tweets to replicate our analysis at no cost.

Bibliography

- American Public Transportation Association. Public Transportation Facts. Accessed July 30 2023. Retrieved from <https://www.apta.com/news-publications/public-transportation-facts/>
- Beirão, G., & Cabral, J. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy*, 14(6), 478–489.
- Ceccato, V., Gaudalet, N., & Graf, G. (2022). Crime and safety in transit environments: A systematic review of the English and the French literature, 1970–2020. *Public Transport*, 14, 105–153. <https://doi.org/10.1007/s12469-021-00265-1>
- Ding, H., Agrawal, A.W., & Loukaitou-Sideris, A. (2020). Sexual Harassment and Assault in Transit Environments: A Review of the English-language Literature. *Journal of Planning Literature*, 35(3), 267–280.
- García, I., Albelson, M., Puczkowskyj, N., Khan, S. M., & Fagundo-Ojeda, K. (2022). Harassment of low-income women on transit: A photovoice project in Oregon and Utah. *Transportation Research Part D: Transport and Environment*, 112, 103466.
- Levine, N., & Wachs, M. (1986). Bus crime in Los Angeles: II—Victims and public impact. *Transportation Research Part A: General*, 20(4), 285–293.
- Litman, T. (2015). *Evaluating public transit benefits and costs*. Victoria, BC, Canada: Victoria Transport Policy Institute.
- Los Angeles Police Department. Crime Data from 2020 to Present. Accessed April 15 2023. <https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8>
- Lubitow, A., Abelson, M. J., & Carpenter, E. (2020). Transforming mobility justice: Gendered harassment and violence on transit. *Journal of Transport Geography*, 82, 102601.
- Muñoz-Dueñas, P., Martínez-Comesaña, M., Martínez-Torres, J., & Bastos-Costas, G. (2023). Estimating mobility of tourists. New Twitter-based procedure. *Heliyon*, 9(2).
- Polzin, S. E. (2016). *Implications to public transportation of emerging technologies*. National Center for Transit Research.

About the Authors

Egbe-Etu Etu, PhD

Dr. Etu is an Assistant Professor of Business Analytics at San José State University. He is also a Research Associate at the Mineta Transportation Institute. His research interest centers on the development of use-inspired machine learning models to solve challenging business problems in healthcare, manufacturing, and transportation. He is a member of the Industrial Engineering and Operations Management, Institute of Industrial & Systems Engineering, and SAVE International.

Asha Weinstein Agrawal, PhD

Dr. Agrawal is Director of the Mineta Transportation Institute National Transportation Finance Center and also Professor of Urban and Regional Planning at San José State University. Her research and teaching interests in transportation policy and planning include transportation finance, bicycle and pedestrian planning, travel survey methods, and transportation history.

Imokhai Tenebe, PhD

Dr. Tenebe is a Research Associate at Mineta Transportation Institute. Previously, he was a Safety Engineer with the Texas Commission on Environmental Quality in Austin, Texas, and participated in research at Texas State University, San Marcos, Texas. He obtained his PhD in Water Resources and Environmental Engineering and has two masters in the same area of expertise from Nigeria and the United States. He has published over 80 articles on several subjects with interests in water resources, healthcare, pollution, transportation, and data science.

Jordan Larot

Jordan Larot is an undergraduate student majoring in business analytics at San José State University with an expected graduation date of May 2024. His research interests include machine learning, data mining, and business intelligence.

Dang Minh Nhu Nguyen

Dang Minh Nhu Nguyen is a recent graduate of San José State University. Nguyen received her B.S. in Applied Mathematics, with a concentration in Statistics, in June 2022. Her research interests include technology, healthcare, business, finance, and aerospace.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Founder, Honorable Norman Mineta***
Secretary (ret.),
US Department of Transportation

Chair, Jeff Morales
Managing Principal
InfraStrategies, LLC

Vice Chair, Donna DeMartino
Retired Transportation Executive

Executive Director, Karen Philbrick, PhD*
Mineta Transportation Institute
San José State University

Rashidi Barnes
CEO
Tri Delta Transit

David Castagnetti
Partner
Dentons Global Advisors

Maria Cino
Vice President
America & U.S. Government
Relations Hewlett-Packard Enterprise

Grace Crunican**
Owner
Crunican LLC

John Flaherty
Senior Fellow
Silicon Valley American
Leadership Form

Stephen J. Gardner*
President & CEO
Amtrak

Ian Jefferies*
President & CEO
Association of American Railroads

Diane Woodend Jones
Principal & Chair of Board
Lea + Elliott, Inc.

Priya Kannan, PhD*
Dean
Lucas College and
Graduate School of Business
San José State University

Will Kempton**
Retired Transportation Executive

David S. Kim
Senior Vice President
Principal, National Transportation
Policy and Multimodal Strategy
WSP

Therese McMillan
Retired Executive Director
Metropolitan Transportation
Commission (MTC)

Abbas Mohaddes
CEO
Econolite Group Inc.

Stephen Morrissey
Vice President – Regulatory and
Policy
United Airlines

Toks Omishakin*
Secretary
California State Transportation
Agency (CALSTA)

April Rai
President & CEO
Conference of Minority
Transportation Officials (COMTO)

Greg Regan*
President
Transportation Trades Department,
AFL-CIO

Rodney Slater
Partner
Squire Patton Boggs

Paul Skoutelas*
President & CEO
American Public Transportation
Association (APTA)

Kimberly Slaughter
CEO
Systra USA

Tony Tavares*
Director
California Department of
Transportation (Caltrans)

Jim Tymon*
Executive Director
American Association of
State Highway and Transportation
Officials (AASHTO)

Josue Vaglienty
Senior Program Manager
Orange County Transportation
Authority (OCTA)

* = Ex-Officio
** = Past Chair, Board of Trustees
*** = Deceased

Directors

Karen Philbrick, PhD
Executive Director

Hilary Nixon, PhD
Deputy Executive Director

Asha Weinstein Agrawal, PhD
Education Director
National Transportation Finance
Center Director

Brian Michael Jenkins
National Transportation Security
Center Director

