

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

Project 2317  
November 2023

Egbe-Etu Etu, PhD  
Jordan Larot

Asha Weinstein Agrawal, PhD  
Dang Minh Nhu Nguyen

Imokhai Theophilus Tenebe, PhD



## Introduction

Public transit is one of the safest ways to travel. Still, misconduct on public transit affects riders, transit staff, and public perception. 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 critical 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.

## Study Methods

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) 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.

## Findings

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 were 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 common concerns. The victims had at times been targeted on the basis of their race, gender, or sexual identity or because they

were transit employees, such as a bus operator. The qualitative analysis also showed that both men and women are victimized, though women were targeted more often than men (57.5% vs. 42.5%).

The tweets of misconduct on public transit provide powerful evidence to illustrate the specific nature of the crimes and misbehavior that impact transit riders.

### Policy/Practice Recommendations

The greatest value of the analysis comes from analyzing the specific content of the relevant tweets. These tweets are essentially miniature 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 specific 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. 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.

### About the Authors

**PI Egbe-Etu Etu, PhD**, is an Assistant Professor of Business Analytics at San José State University (SJSU). Before joining SJSU, Dr. Etu received his PhD in Industrial and Systems Engineering from Wayne State University, Detroit, in 2021. His research interest centers on the development of use-inspired machine learning models to solve challenging business problems in transportation, healthcare, and manufacturing.

**Asha Weinstein Agrawal, PhD**, is the 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**, is a Research Associate at Mineta Transportation Institute. 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 transportation, water resources, healthcare, pollution, and data science.

### To Learn More

For more details about the study, download the full report at [transweb.sjsu.edu/research/2317](https://transweb.sjsu.edu/research/2317)



MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.