UNDERSTANDING ROAD ACCIDENT PATTERNS USING EXPLORATORY DATA MINING: A CASE STUDY OF NYC

by

AFOLABI PAUL ADENIYI

A THESIS

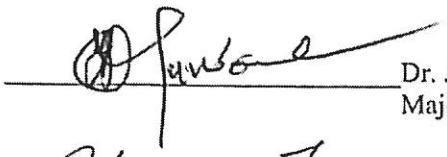
Submitted in partial fulfillment of the requirements
for the degree of Master of Urban and Regional Planning
in the Department of Community and Regional Planning
in the School of Graduate Studies

Alabama A&M University
Normal, Alabama 35762

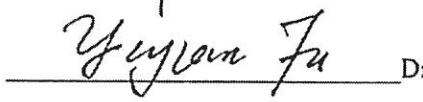
December 2024

Submitted by AFOLABI PAUL ADENIYI in partial fulfillment of the requirement for the degree of MASTER OF URBAN AND REGIONAL PLANNING specializing in TRANSPORTATION PLANNING.

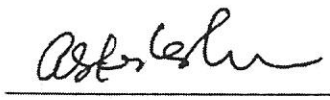
Accepted on Behalf of the Faculty of the Graduate School by the Thesis Committee:



Dr. Jacob Oluwoye
Major Advisor



Dr. Yujian Fu



Dr. Aschalew Kassu

Dr. Florina Dutt

Dr. Kadhi Tau
Dean of the School of Graduate Studies

Date

To Almighty God and my wonderful family, Njideka, Oluwajomiloju, and Temiloluwa Adeniyi, for the endless love and support you all have given me, I give you my sincere thanks and gratitude.

UNDERSTANDING ROAD ACCIDENT PATTERNS USING EXPLORATORY DATA MINING: A CASE STUDY OF NYC

Adeniyi, Afolabi, Paul, M.U.R.P., Alabama A&M University, 2024. 94 pp.

Thesis Advisor: Dr. Jacob Oluwoye

This thesis examines road accident patterns in New York City (NYC) from 2013 to 2023 using exploratory data mining techniques to uncover key predictors and provide actionable insights aligned with Vision Zero's goal of reducing traffic fatalities and injuries. Through Exploratory Data Analysis (EDA) and Micro-Level Analysis, the study identifies significant contributors to accidents, including unsafe speed, alcohol involvement, driver age groups (notably young and elderly drivers), and environmental factors like slippery pavement. While total accidents declined significantly post-2019, injury and fatality rates have increased, reflecting a shift in accident severity. A predictive multivariate regression model, developed through correlation and regression analysis, achieves high accuracy with an Adjusted R-Squared of 0.993, explaining 99.3% of the variance in total accidents. The model demonstrates strong alignment between observed and predicted accident counts, with prediction errors typically within ±3%, validating its reliability for practical applications. Despite its robustness, minor prediction deviations and collinearity issues highlight areas for improvement in future research. The analysis also reveals the pandemic's influence on traffic patterns, with reduced mobility leading to fewer accidents. These findings emphasize the value of data-driven strategies for proactive road safety planning. By identifying critical risk factors and forecasting trends, this research offers actionable recommendations for policymakers, urban planners, and traffic safety

authorities to develop targeted interventions, improve traffic management, and enhance road safety in New York City.

KEYWORDS: Traffic Accidents, Knowledge Discovery, New York City, Predictive Modeling, Traffic Safety, Vision Zero, Driver Behavior, Environmental Factors

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

AI	Alcohol Involvement
BU	Backing Unsafely
DA1	Driver Age 16-25
DA2	Driver Age 26-35
DA3	Driver Age 36-45
DA4	Driver Age 46-55
DA5	Driver Age 56-65
DA6	Driver Age 66-75
DA7	Driver Age 76-85
DA8	Driver Age 86-95
DID	Driver Inattention/Distraction
FYRW	Failure to Yield Right-of-Way
NYC	New York City
PS	Pavement Slippery
TA	Total Accident
TI	Total Injured
TK	Total Killed
US	Unsafe Speed

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to the Almighty God for His infinite wisdom, guidance, strength, and support throughout this journey. Truly, it has been by His grace. To my Lord Jesus, I offer my heartfelt thanks and praise.

I extend my deepest gratitude to Dr. Jacob Oluwoye, whose mentorship and encouragement were instrumental in my decision to pursue a master's degree in urban and regional planning. His guidance has continually challenged me to strive for excellence so that I will be prepared for opportunities that awaits me as a planner within the professional sphere.

I am also deeply grateful to the distinguished members of my committee, Dr. Yujian Fu, Dr. Aschalew Kassu, and Dr. Florina Dutt, for their thoughtful guidance, constructive feedback, and immense support which have significantly enriched this research.

Special thanks are extended to Mr. Bedemi Atolagbe-Olaoye, Mr. Samuel Oladapo, Mr. Mckinley Curtis III, and Dr. (Mrs.) Oluwoye for their mentorship, sound advice, prayers, and financial support during this journey.

To my beloved wife, Njideka, and our sons, Oluwajomiloju and Temiloluwa Adeniyi, your unwavering love, patience, and support have been my bedrock throughout this program. I am profoundly grateful for your encouragement and sacrifices. Lastly, I extend my heartfelt thanks to my extended family for their prayers, encouragement, and

steadfast support which have been a source of strength and motivation throughout this journey. Thank you all.

CHAPTER 1

INTRODUCTION

Road traffic accidents constitute a significant public health issue, consistently ranking among the leading causes of death and injury worldwide (WHO, 2018). In urban environments, where population density and vehicular traffic converge, the impact of road accidents is especially pronounced. NYC, known for its bustling streets and dense traffic, characterizes such environments where road safety is both a persistent challenge and a priority for municipal policy.

The importance of this study is underscored by the global urgency to enhance road safety. According to the WHO (2018), road traffic injuries are the eighth leading cause of death globally, surpassing HIV/AIDS and tuberculosis. In the context of NYC, the Vision Zero initiative reflects the city's systemic approach to addressing the dangers of road traffic accidents by aiming for zero fatalities or serious injuries (NYC Vision Zero, 2021). This proactive stance towards road safety underscores the relevance and timeliness of employing advanced analytical techniques to understand and prevent road accidents.

Data mining offers a robust set of tools capable of extracting hidden patterns and new knowledge from vast quantities of data which traditional statistical methods might not fully capture. This thesis leverages exploratory DM methodologies which have demonstrated significant efficacy in uncovering factors contributing to road accidents

across various settings (Takale et al., 2022). Such techniques enable the examination of complex interactions between numerous variables including, but not limited to, crash date, crash time, crash location, age, sex, contributing factors, and driver behavior.

Moreover, the advancement in big data technologies enhances the capacity to analyze and interpret large datasets, leading to more precise and predictive analytics (Feng et al., 2020). In NYC, where data is abundant but also complex due to the city's diverse transportation dynamics, such approaches are not just suitable but necessary. By applying these data-driven methodologies, this research study aims to contribute to a deeper, more nuanced understanding of the precursors to road accidents, offering a data-informed foundation for interventions aimed at reducing their occurrence.

This study, titled Understanding Road Accident Patterns Using Exploratory Data Mining: A Case Study of NYC, leverages the power of exploratory data mining techniques to analyze comprehensive accident data collected in NYC from 2013 to 2023. By utilizing descriptive statistics, visualization (including hierarchical clustering),correlation and multivariate regression, this research delves into key predictors of road accidents, including driver behavior, environmental conditions, and demographic variables. Data mining allows for identifying complex relationships that traditional methods may overlook, offering a more granular understanding of accident dynamics in NYC.

The primary objectives of this research are fourfold: (1) to identify and quantify the factors most strongly associated with road accidents in NYC; (2) to uncover unique patterns and dependencies among accident-related variables; (3) to create predictive models that can assist in forecasting accident frequencies, and (4) to translate insights derived from exploratory data mining techniques into actionable strategies. These objectives are aligned

with broader safety goals, aiming to equip city planners, policymakers, and public safety officials with actionable insights that can drive targeted interventions, reduce accident rates, and enhance traffic safety in NYC.

This thesis contributes to the literature on urban road safety by employing data analytics to model accident patterns in NYC, a city that serves as an example of the challenges faced by other large urban areas. Through rigorous analysis, this research offers insights into high-risk behaviors, vulnerable age demographics, and critical environmental conditions affecting accident rates. Ultimately, this study aims to support NYC's Vision Zero goals by recommending data-informed safety strategies that are responsive to identified accident trends and predictive of future risks.

Statement of the Problem

The increasing frequency and severity of road accidents in urban settings, particularly in New York City, represent a critical issue affecting public safety and mobility. Despite various initiatives aimed at improving road safety, such as NYC's Vision Zero, accidents persist, leading to fatalities, injuries, and extensive economic costs. While existing studies have analyzed road accidents using traditional statistical methods, these approaches often fall short in accounting for the complex, multifaceted nature of urban road safety dynamics.

In NYC, unique factors, including high population density, diverse demographic groups, and a complex transportation network propose specific challenges to road safety. Traditional methods often overlook interactions between variables such as the influence of environmental conditions, driver age, and behavioral factors on accident occurrences.

Additionally, while data mining techniques have demonstrated substantial potential in enhancing road safety analysis, the application of these methodologies to NYC unique road accident data has not been adequately explored. Furthermore, there is a limited exploration of how situational changes like those introduced by the COVID-19 pandemic have altered traffic patterns and accident rates in NYC.

To address these gaps, this thesis study applies exploratory data mining techniques to NYC's Road Accident Data from 2013 to 2023. By analyzing large volumes of data and extracting hidden patterns, this research aims to provide a more comprehensive understanding of accident predictors, including driver behavior, environmental conditions, and demographic variables. This approach not only enhances the accuracy of accident predictions but also supports the development of data-informed strategies to reduce accidents and align with Vision Zero's goals of minimizing traffic fatalities and serious injuries.

Research Questions

1. What are the primary contributing factors to road accidents in New York City from 2013 to 2023?
2. How have the severity, trends, and recurring patterns of road accidents in New York City evolved between 2013 and 2023?
3. Can the insights derived from exploratory data mining techniques be translated into practical and actionable strategies for reducing road accidents in New York City?