

Assessing Motor Vehicle Collisions Severity in Toronto*

Jason Yang

September 25, 2024

This paper examines statistical techniques to analyze motor crash data in Toronto, using either real-world or simulated datasets, with the goal of identifying factors contributing to crashes. The models developed in this analysis suggest that the number of accidents per year has declined over time, likely due to advancements in technology and improved public safety measures.

1 Introduction

Toronto, one of the largest and most densely populated cities in North America, consistently ranks among the top 10 cities globally for the highest levels of traffic congestion (Usnews.com). With a growing number of population from an increase in immigrants, the traffic-related has become a significant problem for everyone. Along with the challenges of congestion, Toronto also faces a high number of motor vehicle collisions each year, leading to a significant public safety concerns.

This paper aims to analyze motor vehicle collision data in Toronto by collecting the Data from Open World Toronto. The data used in this study were sourced from Open Toronto, but the accuracy, completeness, and timeliness of the data are not guaranteed, and it should not be compared with other crime data sources. The purpose is to assess the causes of motor vehicle collisions in Toronto and identify patterns or factors that contribute to these incidents. By analyzing the data through statistical methods, the study aims to uncover key variables such as weather conditions, time of day, road types, and driver behavior that may influence the likelihood of crashes.

*Code and data are available at: https://github.com/jasonbot123/Toronto_Motor_Collisions

2 Data

2.1 Overview

This paper uses the Motor Vehicle Collisions involving Killed or Seriously Injured Persons data from Open Data Toronto, accessed using the `opendatatoronto` package (Gelfand 2022.) This dataset includes all Toronto’s traffic collisions where a person was killed or seriously injured since 2006. According to “Our World In Data”, in 2019, motor vehicle collisions accounted for 2.3% of global deaths, resulting in 1.3 million fatalities annually and approximately 3,500 deaths per day. With these numbers being calculated, the broader context behind this dataset involves the efforts of Toronto’s city planners, traffic safety agencies, and policymakers to reduce the number of severe accidents in Toronto.

Additionally, this dataset contains a wide range of variables that capture various aspects of motor vehicle collisions. The variables include unique identifiers for each collision, details about the time and location, road and environmental conditions, information on the individual involved, such as driver’s action and the severity of injuries.

Despite the dataset’s richness and detailed aspects of the vehicle collision, not all variables were utilized in the current analysis, allowing us to focus on the most relevant factors for the purpose of this paper. In this paper, R (R Core Team) was used for all statistical computing and visualizing associated with the data. For simplicity reasons, the data was cleaned for any missing information and the most relevant variables were selected for statistical analysis.

2.2 Results