Assessing Motor Vehicle Collisions Severity in Toronto*

Jason Yang

September 25, 2024

This paper examines statistical techniques to analyze motor vehicle collision 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. These insights emphasize the need for city planners of Toronto to address the act of reckless driving and promote public safety.

Table of contents

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

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 (Navarre (2024)). 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 significant public safety concerns.

This paper aims to analyze motor vehicle collision data in Toronto by collecting the Data from Open World Toronto (Gelfand (2022)). 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.

In this paper, models were made for each categorical variable that could've contributed for each motor vehicle collision observation. By observing each model, references were drawn to conclude whether those variables had significant impacts of the cause of each collision. These findings could then inform Toronto's city planners to make better decisions and improvise at reducing the number of motor vehicle 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 (Dattani et al. (2023)). 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 sever 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 collisions, 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 (2023)) 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 (Wickham et al. (2023)).

2.2 Features

The dataset titled Motor Vehicle Collisions Involving Killed or Seriously Injured (KSI) Since 2006 contains 18,957 observations across 50 variables. Due to the complexity of analyzing such a large dataset, only the most relevant variables were selected for statistical analysis. This was achieved using the select() function from the dplyr package in R, allowing for a more focused exploration of the key factors contributing to serious collisions.

Among the 50 variables recorded in this dataset, the most relevant variables chosen for this paper are:

- `DATE`: Date Collision Occurred
- `TIME`: Time Collision Occurred
- `DISTRICT`: City District
- `VISIBILITY`: Environmental Condition
- `INVAGE`: Age of Involved Party
- `DIVCOND`: Driver Condition
- `ACCLASS`: Classification of Accident
- `INJURY`: Severity of Injury
- `SPEEDING`: Speeding Related Collision
- `AG_DRV`: Aggressive and Distracted Driving Collision

2.3 Data Visualization

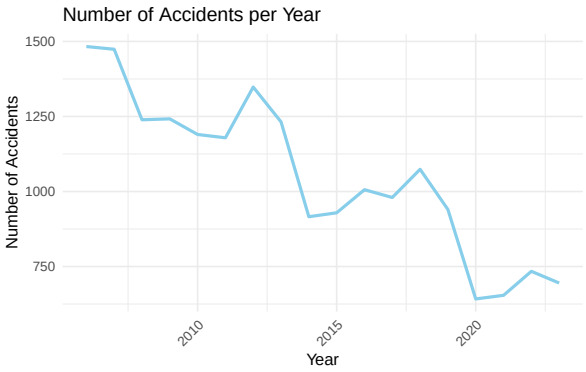


Figure 1

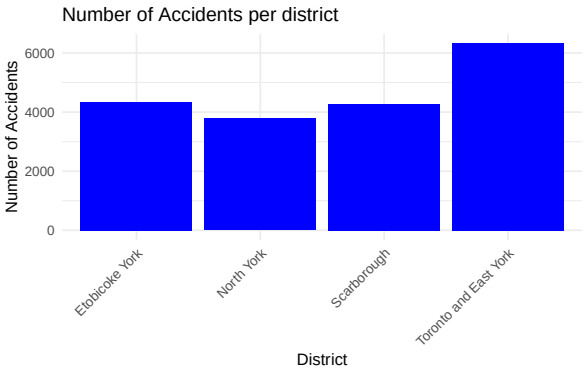


Figure 2

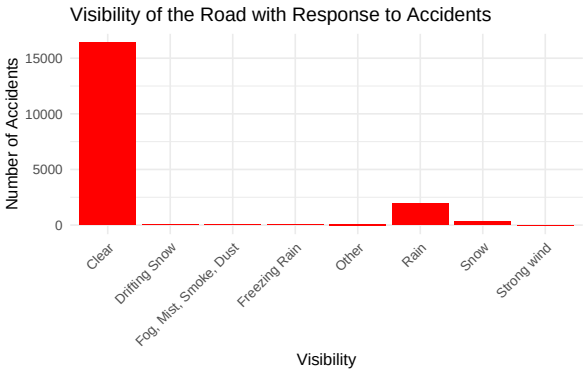


Figure 3

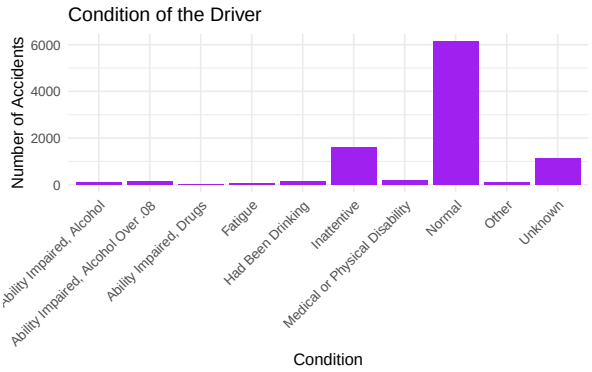


Figure 4

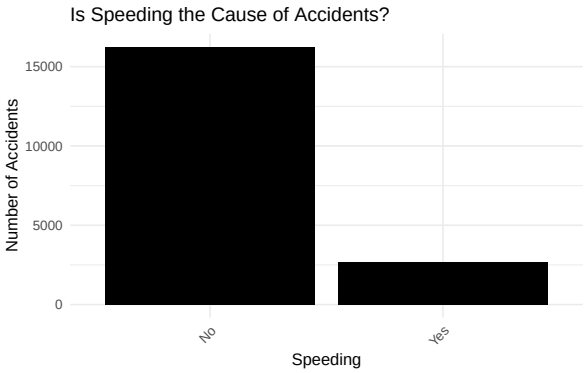


Figure 5

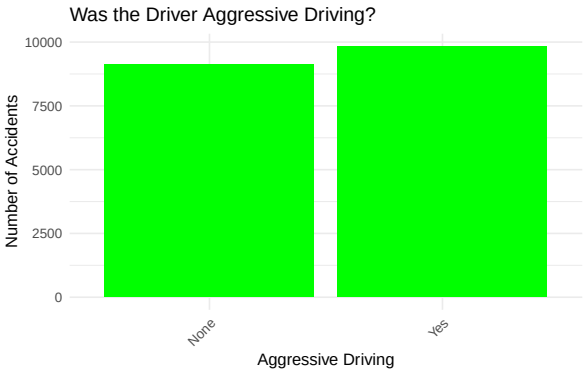
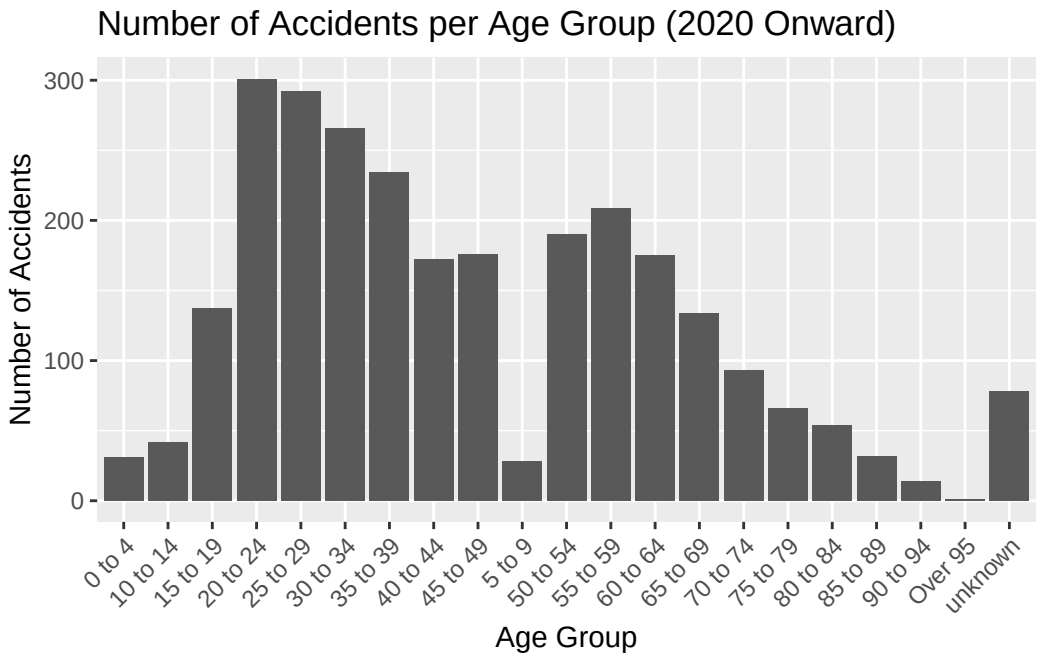


Figure 6



2.4 Results

As shown in figure 1, over the past 18 years, the number of motor vehicle collisions in Toronto has decreased significantly. The highest number of incidents occurred in Toronto and East York. However, Figure 2 shows that the other three regions average around 3,500 incidents, which is not far behind the numbers for Toronto and East York. The reason for a slight higher number for Toronto and East York may due to the congestion within the downtown Toronto and the higher chance of unexpected accidents to occur.

The bar graph for 'Number of Accidents per Age Group (2020 Onward)' clearly shows that the data is right-skewed, with a higher number of accidents occurring among individuals in their 20s to 40s. At this age, people are more likely to make awful decisions that may result in unfortunate events. Figure 6 shows that approximately 50% of the incidents are attributed to aggressive driving. The remaining 50% are caused by factors such as speeding, impaired vision, and inattentiveness, as illustrated in Figures 3, 4, and 5.

3 Discussion

The purpose of this study was to analyze motor vehicle collision data from Toronto, with a particular focus on identifying the factors that contribute to severe injuries and fatalities, as well as exploring the impact of driver behavior, road conditions, and environmental factors on accident outcomes.

The findings of this study have important implications for public safety and traffic policy in Toronto. Policies that target speeding and aggressive driving, such as the installation of more speed cameras or stricter enforcement of traffic laws, could further reduce the number of severe accidents. Additionally, improving road visibility through better lighting and infrastructure upgrades may help reduce accident severity, especially under poor weather conditions. These insights could be valuable for policymakers aiming to improve road safety and reduce fatalities in urban areas.

3.1 Limitations

The accuracy and completeness of the data cannot be guaranteed, which may have introduced some bias into the analysis. Future research could explore the impact of more granular factors, such as specific driver behaviors or weather patterns, as well as the effect of new vehicle technologies on reducing collisions.

References

Dattani, Saloni, Fiona Spooner, Hannah Ritchie, and Max Roser. 2023. "Causes of Death." *Our World in Data*.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Navarre, Brianna. 2024. "10 Cities with the Worst Traffic in the World." https://www.usnews.com/news/cities/slideshows/cities-with-the-worst-traffic-in-the-world.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.