Vehicle Collision Analysis in Toronto

GROUP 19

Qi Zhao
Weijing Zhang
Jiabo Wang
Nancy Pham

1. Introduction

One of the goals in achieving smart cities is to improve public safety, which is the main topic of this project. In 2016, 160,315 injuries happened in Toronto along with 1898 fatalities due to vehicle collisions [1]. This prompts the need for changes to be made for eliminating the amount of injuries and collisions in the city of Toronto. The Toronto Police Department has established an open dataset that records collisions involving fatalities and serious injuries occurred in Toronto from 2007 to 2017. This project aims to explore and analyze this set of data in order to make proposals to improve traffic safety for the city of Toronto.

1.1 Datasets

Collision Killed and Seriously Injured (KSI) Dataset

The Toronto Collision Cases dataset includes the geographic location of all reported collisions in Toronto and specifies the date and time of the collisions. The data is provided by the Toronto Police Service Public Safety Data Portal. This dataset is a subset of the Killed and Seriously Injured (KSI) dataset from 2007-2017. These events include any serious or fatal collisions occurring in the city of Toronto.

Among some of the information provided, the scope of this project will make use of: the classification of collision, the time and date when the collision happened, its location (longitude, latitude), the ward (geographical area) where the collision took place, weather and road conditions, injury information, reasons for collisions, and driver conditions.

Population Dataset

Selecting the population data from the City of Toronto Neighbourhood Profiles.

2. Objective

The main objective of this project is to build a model that will help predict the types of collisions that occur in the city. The idea behind this objective is that this model will be helpful for the city to input data, such as the hour of the day, the weather conditions, the geographical area, the driving manoeuvres etc. and have a prediction of the most likely type of collision to occur. By being able to predict what kinds of collisions will happen, the city can be more prepared, to make appropriate plans to help prevent these collisions.

3. Exploratory Data Analysis

To better understand the vehicle collisions that occur in the city of Toronto, it is important to look at some different aspects of the KSI dataset.

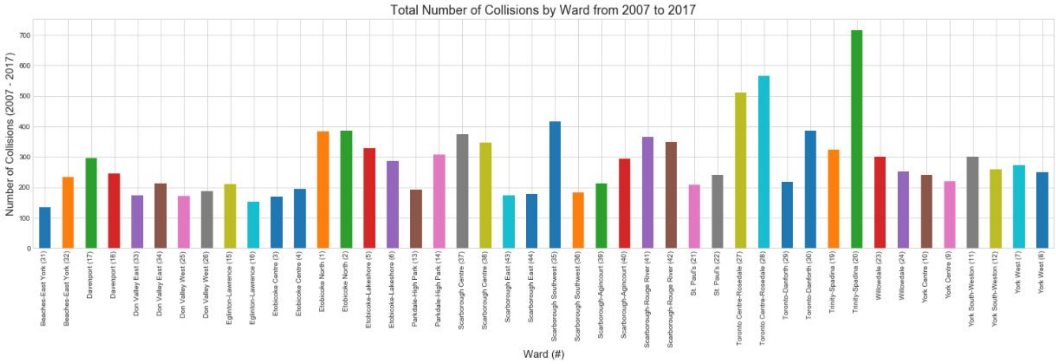

Figure 1. Total number of collisions by ward from 2007 to 2017

3.1 Data Analysis

The KSI dataset can be explored to how trends of collisions in Toronto. Inevitably, there are areas of the city where collisions occur more frequently. Figure 1 shows what the distribution of the number of collisions happened in different areas.

In order to further understand our dataset, we also performed time-series analysis and predicted the number of collisions for the next 12 months based on historical data. While the total number decreases over the years (which makes sense as car design is improving and the city has already taken some preventative measures), there is still something interesting to note: a fairly strong seasonal trend is revealed following the time-series prediction. It has shown that more collisions resulting in KSI victims accurred during the summer.

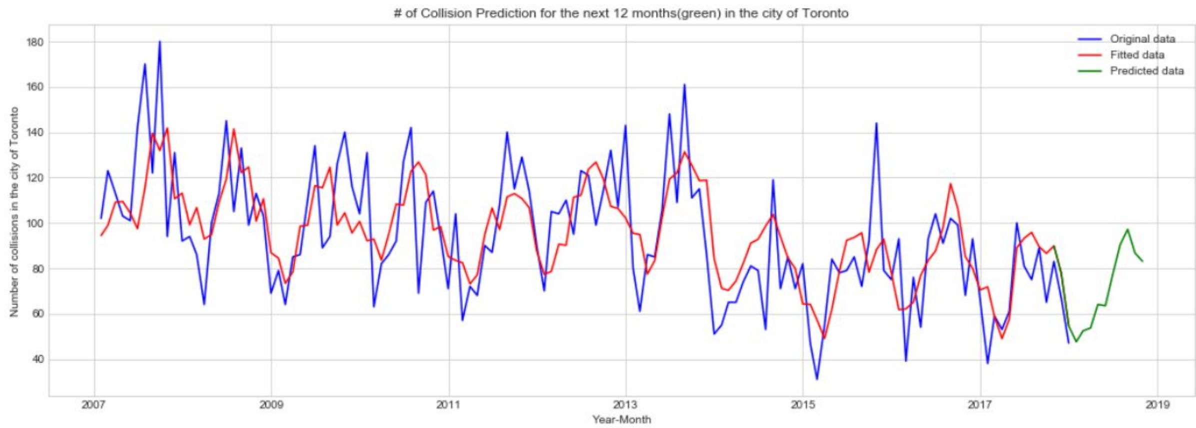

Figure 2: Number of Collision Prediction for the next 12 months in the city of Toronto

3.2 Problem Proposals

3.2.1 Prediction Model

After looking at the KSI dataset, some things that stood out were that there was a fairly clear divide between the types of collisions that occur, which can be grouped into pedestrian collisions, cyclist collisions, and vehicle-related collisions. Given all of these records from the past ten years, and the fact that Toronto is a rapidly growing city, it would be useful if there were a way to use the data to help the city in better understanding what to plan for the future by knowing the types of collisions that might occur in the future.

Accuracy for Random Forest for Vehicle Collision Classification: 71.76%

	precision	recall	f1-score	support
Cyclist Collisions	0.49	0.25	0.33	312
Pedestrian Collisions	0.72	0.74	0.73	1237
Vehicle Collisions	0.74	0.79	0.77	1556
avg / total	0.71	0.72	0.71	3105

Figure 3: Accuracy Matrix for Random Forest Algorithm

It was decided that the prediction model would help to classify which of the three types of collisions (pedestrian, cyclist, vehicle) and that the input features that should be used for this prediction model were features that city planners could plan for beforehand. This meant excluding features such as driver conditions (impairment and alertness), driver actions (how the driver reacted to the impending collision), and factors relating to what kind of driving was involved (speeding, running red lights, etc.). The features that were included were: light conditions (e.g. whether it was dusk/dawn, daytime, or nighttime), road conditions, visibility, what kind of manoeuver (turns, straight driving, reversing,

etc.), the type of road (e.g. highway, main street, side street, etc.), and where (includes features for both the geographical area, and whether it happened at the intersection or mid-block).
After training and testing different models, including logistic regression, decision trees, and random forests, the optimal model with the highest accuracy was the random forest. The results are shown in Figure 3. Through verification and optimization, this model should become reliable enough to be used for future years, to help in planning for the city.
Looking at the predictive model, it gives an idea of what kinds of factors might affect the types of collisions and how we can mitigate the predicted vehicle collisions. For example, Figures 4 and 5 are two examples of how it is possible to analyze the factors that contribute to the collisions. Figure 4 shows that vehicle collisions occur less when there is less lighting, and pedestrian collisions occur more when there is less light. From Figure 5, the trend can be seen that pedestrian collisions increase in the winter months, while vehicle collisions and cyclist collisions seen more in the summer months.

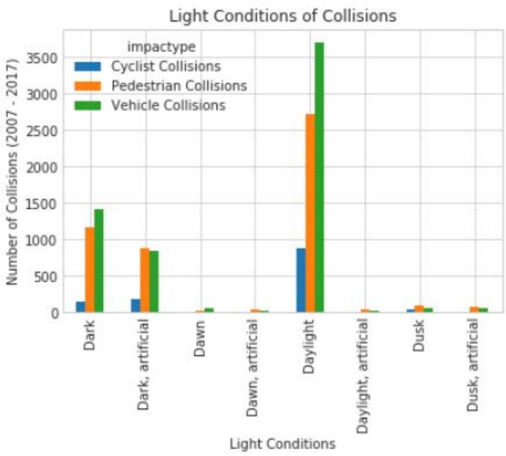

Figure 4: Light Conditions of Collisions

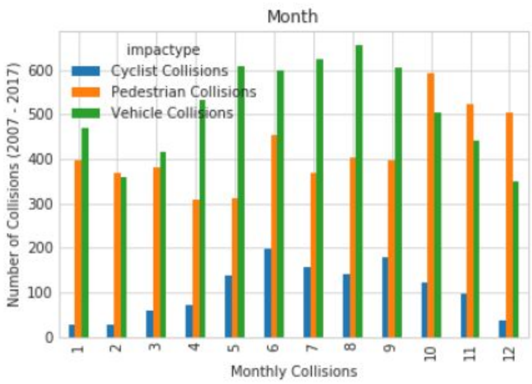

Figure 5: Monthly Collisions

Also, if we use certain features as input, our model will have an output showing that in corresponding conditions, what type of collision will most likely happen. We can apply this to each ward (areas) in Toronto, for example, ward 1in the condition of dark light, mid-block road and snow day will be more likely to have pedestrian collision, or ward 20 on Monday at 6pm in the condition of intersection road will be more likely to have vehicle collision.

3.2.2 Collision Position Analysis and Problem

Based on the analysis of coordinates of each occurred collision in the city of Toronto, we mapped all the points according to longitude and latitude, as well as imported the police division dataset to scatter the positions of police stations (figure 6 left). It was found that the area distribution of the police station was not good enough where some collisions spots was still relatively far away from the nearest police division. Besides, another problem is that the number of police stations was either not enough in those areas that were more likely to have collisions.
Similarly, we also imported the hospital data to scatter the hospital positions in the plot of fatal collision points visualized by tool gmplot in the google map (figure 6 right). It was found that the placement of hospitals has the similar problem as the police division that centralized fatal collision points in the city of Toronto need the ambulance and first-aid as soon as possible in order to save more people’s life.

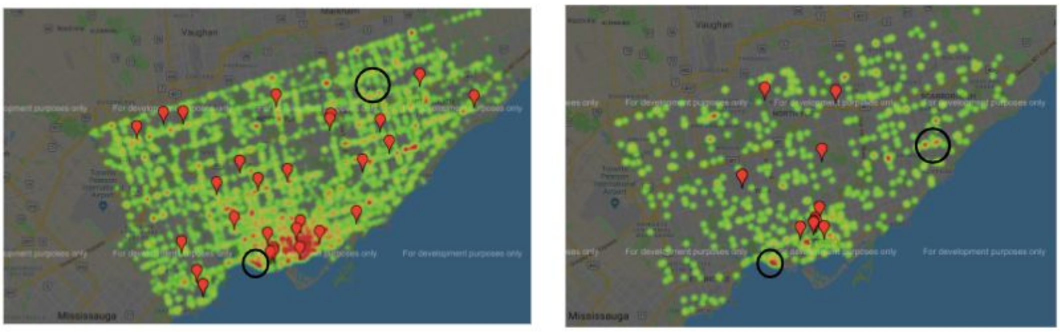

Figure 6. Plotting collisions with police stations(left) and hospitals (right) on Google Maps

3.2.3 Age Range Analysis and Problem

Combing the population data, the 15-24 age range has the highest percentage of collisions, which is shown below as figure 7. Besides that, figure 8 and figure 9 show that this age range also has the largest proportions of alcohol and speeding-related collisions. Young drivers' behavior is a serious problem that may cause more collisions.

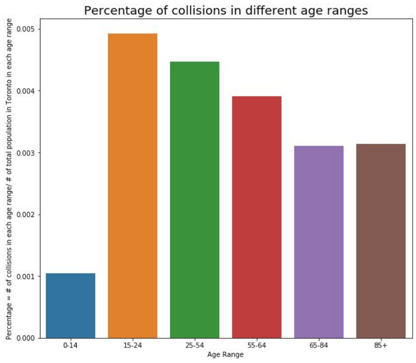

Figure 7.Percentage of Collisions in different age ranges

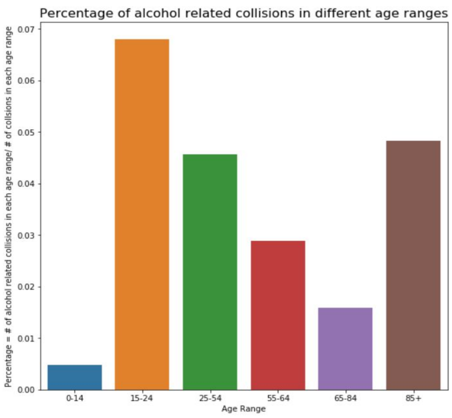

Figure 8.Percentage of alcohol-related collisions

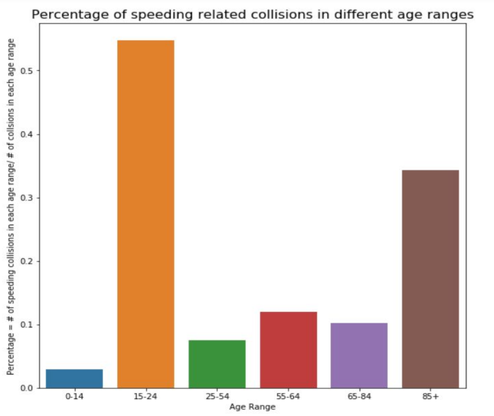

Figure 9.Percentage of speeding-related collisions

4. Discussion - Solutions to problems

4.1 Solution to the Prediction Model Problem

From previous part, we analyzed our model and find some features that will help us mitigate the predicted collisions. Generally, we can arrange more street lights on certain area or certain street to mitigate the visibility-related problem and we can introduce 3D crosswalk or arrange more footbridges to the city to mitigate the pedestrian collisions in winter that may due to the snowfall. However, we still need more detailed analysis and more data set, which may be our future work to do.

4.2 Solution to the Collision Position Problem

Based on the scattering plots of collision and police divisions, we would suggest that it would be helpful to locate new police divisions in the Scarborough-Agincourt and Etobicoke North. A significant number of collisions happened in these two wards, but there is a lack of police division. Adding more police and increasing the number of patrol policemen would be an improvement for these wards.

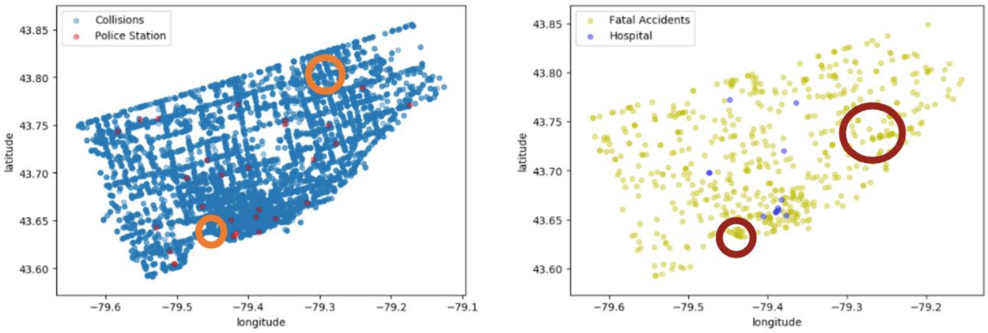

Figure 10. Locations of Police Divisions and Hospitals

Studies worldwide have shown that death could be prevented in many cases in which people died before reaching a hospital [2]. As seen in Figure 10, Parkdale-High Park and Scarborough-Rouge River are two collisions 'hotspots' where many collisions occur but do not have a hospital in close proximity. The proposal is to build hospitals for injured people so that they receive timely treatment.

4.3 Solution to the Age Range Problem

Another problem we are trying to solve is related to young driver's behavior, cause the age range of 15-24 will be more likely to involve in a collision, especially alcohol and speeding-related collision.

The minimum age to get the driver's license in Ontario is 16 years old. According to the results, it is better to increase the minimum age to 18 to prevent more collisions, which is shown in the report [3].

Throughout Canada, the maximum BAC for fully licensed drivers is 0.08. Driving with BAC over 0.08 is a criminal offence and the penalties are severe. In Ontario, people will also face serious consequences if the BAC is between 0.05 and 0.08 [4]. It is hard to control the alcohol driving problems, so the government can decrease the BAC level to 0.05 for safer driving. Since 2008, 10 countries have improved their drink–driving laws to meet best practice (BAC =0.05 g/dl or less), helping protect 186 million people [5].

Apart from that, Toronto government could set the law specific to the drivers aged between 15-24.The New Zealand government set an 18-month restricted permit for 15- 24 years drivers that allowed no driving from 22:00 to 05:00, no passengers under 20 years and a blood alcohol concentration (BAC) limit of 0.03 grams per deciliter (g/dl). An evaluation showed that the graduated driver license had contributed to an 8% reduction in crashes involving serious injury among young novice drivers [7].

5. Conclusion

This report conducted a detailed analysis of various types of killed and seriously injured collisions in Toronto from 2007 to 2017 and is aimed to propose preventative measures to mitigate this problem in the future. In order to solve the problem, several algorithms were implemented to choose the best prediction model, including logistic regression, decision tree, and random forest. The random forest model performed well for this case and was used for the final model. Time-Series analysis was applied to predict the number of collisions in the city of Toronto in the next year. After analyzing the data, the proposed changes to decrease the number of KSI collisions are setting police divisions and hospitals in certain areas and making new rules aim at 15-24 years old drivers.

6. Reference

[1] .Canadian Motor Vehicle Traffic Collision Statistics: 2016 - Transport Canada. (2018). Retrieved from
https://www.tc.gc.ca/eng/motorvehiclesafety/canadian-motor-vehicle-traffic-collision-statistics-2016.html

[2]. Mock CN, nii-Amon-Kotei D, Maier RV. Low utilization of formal medical services by injured persons in a developing nation: health service data underestimate the importance of trauma. Journal of Trauma, 1997, 42:504–513.

[3].Norghani M et al. Use of exposure control methods to tackle motorcycle accidents in Malaysia. Serdang, Road Safety Research Centre, Universiti Putra Malaysia, 1998 (ResearchReport 3/98)

[4]. Impaired driving. (2018). Retrieved from
http://www.mto.gov.on.ca/english/safety/impaired-driving.shtml

[5].World Health Organization. (2013). Global Status Report on Road Safety 2013: Supporting a Decade of Action (p. 16).

[6].Begg D, Stephenson S. Graduated driver licensing:the New Zealand experience. Journal of Safety Research,2003, 34:3−4.

[7]. Homel RJ. Random breath testing in Australia: a complex deterrent. Australian Drug and Alcohol Review, 1988, 7:231−241.