Evaluation of the Efficacy of Red Light Cameras on Road Safety in Toronto

DATA 601 Final Report
Date: October 17th, 2024
Group Members: Anika Achari, Franco Doi, Lalith Nandakumar, Peter Yee, Stephen Bonsu

Table of Contents

1. Introduction

Roads are among the most important forms of infrastructure around the world. They are a hallmark of any efficient and effective society due to their role in facilitating human movement, connectivity, convenience and access to essential services. As a result, it is imperative to foster road safety to ensure the continued use of roads by all individuals and the safety of all road users. The population of urban areas increased from 751 million people in 1950 to 4.2 billion in 2018, and projections indicate that 68% of the global population will reside in urban areas by 2050 (1). Toronto is one of several large cities around the world that has started to experience this type of population growth in recent years (2). According to 2024 estimates, the population of Toronto's metro area is 6,431,000, which is a 0.93% increase from 2023 (2). Unsurprisingly, Toronto is the fastest growing city in North America (2). As large cities continue to experience unchecked population growth, it is inevitable that this growth will be accompanied by traffic problems (1-2). Additionally, ensuring road safety will be critical to effectively managing rapidly growing urban areas.

Tasked with ensuring sustained road safety in the face of projected population increases, cities across Canada have implemented the use of red light cameras (RLCs) as a means of improving road safety. These cameras operate by penalizing motorist failures to abide by red light, bus lane, speed limit, and stop sign regulations (3). In 2023, the City of Toronto committed to expanding their RLC pilot program by increasing the number of cameras from 149 to 298 (4-5). The installation of these cameras has led to reductions in angle collisions that often result in fatalities (4-5).

One study by Scientific American that focused on the evaluation of RLCs in Houston concluded that RLCs "may not make streets safer" (6). This study did not uncover evidence that would suggest that RLCs reduce the total number of accidents, traffic-related injuries or the likelihood of incurring an incapacitating injury (6). As a result, we felt it was absolutely necessary to evaluate the efficacy of RLCs through comprehensive data analysis to confirm their role in traffic safety improvement. Additionally, conducting data analysis is critical to determining if such cameras should be adopted and deployed on a wider-scale or if other, potentially superior road safety strategies should be explored.

Toronto has utilized RLCs in a limited capacity since 2000 (4-5). In 2023, the City of Toronto began the process of deploying RLCs throughout the metro center (4-5). Hence, we wanted to evaluate the efficacy of RLCs in Toronto. Our findings regarding the efficacy of these RLCs could potentially inform Toronto city officials and Canadian federal government officials of RLC efficacy before a wide-scale deployment of such cameras begins across the country. We chose to evaluate the role of RLCs in maintaining safety at major intersections by exploring traffic collisions at intersections with RLCs and traffic collisions based on traffic volume at intersections with RLCs compared to intersections without RLCs.

2. Datasets

Our analysis of RLC efficacy relies on three datasets derived from the City of Toronto Open Data Portal. The following three datasets were used:

2a. Police Annual Statistical Report - Traffic Collisions

This dataset was acquired from the Open Data Portal and is refreshed annually by Toronto Police Services (7). It includes all Motor Vehicle Collision (MVC) occurrences; which includes property damage (PD) collisions, fail to remain (FTR) collisions, injury collisions and fatalities. A collision is defined as the contact from a motor vehicle that results in property damage, injury or death. This data exists at the occurrence level, and thus, each record can be associated with multiple offenses and victims. This dataset does not include collision occurrences that have been determined as unfounded. FTR collisions are defined as collisions that occur when individuals involved in a motor vehicle collision fail to remain on the scene and fail to provide their information at the scene of the collision. PD collisions are defined as collisions that occur when a motor vehicle collision results in damaged property or when the value of damages is less than $2,000 for all parties involved in the collision. Personal injury collisions are defined as those collisions that result in the individual(s) involved sustaining personal injuries. Fatal collisions are defined as collisions that result in a fatality within 30 days from a motor vehicle collision. This dataset consists of 21 columns.

2b. Red Light Cameras

This dataset was acquired from the Open Data Portal and is refreshed daily by Toronto Transportation Services (8). This dataset includes all intersections in Toronto at which red light cameras (RLCs) are located. RLCs in Toronto are triggered when vehicles enter an intersection after the light has turned red. When this occurs, the RLC will take two photographs of the motor vehicle, both of which are time-stamped. The first photograph is taken as a motor vehicle approaches the stop line and the second photograph is taken as a motor vehicle moves through the intersection. These RLCs are intended to modify driver behavior to reduce instances of individuals running red lights. This dataset consists of 28 columns.

2c. Traffic Volumes at Intersections for All Modes

This dataset was acquired from the Open Data Portal and is refreshed daily by Toronto Transportation Services (9). This dataset contains traffic volume data for the City of Toronto and comes in two forms: Automatic Traffic Record (ATR) Counts and Turning Movement Counts (TMC). ATR data reflects the total volume of traffic on a specific street moving in a specific direction, whereas TMC data pertains to total volume of traffic observed in each leg of an intersection (i.e. North, East, South and West). Additionally, TMC data includes the observed turning movement by mode (i.e. type of motor vehicle).

3. Objectives & Guiding Questions

To conduct our analysis into the efficacy of RLCs in the City of Toronto, we developed a set of two objectives and associated guiding questions to serve as the foundation for this work. The objectives and associated guiding questions are outlined below:

Objectives

Objective #1: Determine if intersections with RLCs experience fewer traffic collisions compared to intersections without RLCs.

Objective #2: Determine if intersections that experience increased traffic volume also experience a greater number of collisions compared to intersections that experience lower traffic volume.

Guiding Questions

Guiding Question #1: Do intersections with RLCs experience fewer traffic collisions compared to intersections without RLCs?

Guiding Question #2: Do intersections that experience increased traffic volume also experience a greater number of collisions compared to intersections that experience lower traffic volume.

4. Hypotheses

We have developed a set of hypotheses based on the aforementioned objectives and guiding questions. Both hypotheses have been outlined below:

Hypothesis #1: RLC infrastructure is costly and has been developed with the specific intention of reducing traffic collisions. Hence, we expect that intersections with RLCs will experience comparatively fewer collisions than intersections without RLCs. Additionally, we anticipate that collision numbers will differ greatly across intersections in Toronto, based on the location of the intersection and the volume of traffic that passes through the intersection. This hypothesis is linked to the first objective and first guiding question.

Hypothesis #2: We expect to observe a greater number of collisions at intersections that experience higher levels of traffic volume, based on literature that suggests that accident frequency increases as traffic congestion increases (1-2). This hypothesis is linked to the second objective and second guiding question.

5. Methodology

5a. Data Collection

Three separate datasets were first acquired from the City of Toronto Open Data Portal, two of which are provided by Toronto Transportation Services and one provided by Toronto Police Services (see section 2).

5b. Data Cleaning

The datasets were first inspected for missing data and irrelevant data; which was removed. For instance, the Traffic Volumes at Intersections for All Modes dataset (see section 2c) consisted of traffic volumes through both streets and intersections. As a result, we dropped all data pertaining to traffic volume through streets. Data cleaning was performed using the Pandas and NumPy libraries in Python.

5c. Data Transformation

Each of the three datasets was downloaded from the Open Data Portal in a csv file format. Geospatial data in each of the three datasets existed in the form of longitude and latitude features. These were extracted and used to create a new feature entitled ‘geo_coordinates.’

A feature called ‘red_light_intersection’ was created to categorize whether each intersection has an RLC by referencing the RLCs dataset. In addition, another feature called ‘not_red_light_intersection’ was created to denote intersections without RLCs, likewise, referencing the RLCs dataset. A ‘distance_to_intersection’ feature was added to denote the distance to the closest intersection, only if the collision was determined to be within the threshold of an intersection for which traffic volume data had been collected.

The data was summed across time periods of collection and across various modes (i.e. vehicle types) to acquire the total volume of motor vehicle traffic through a particular intersection at a particular time. The Police Annual Statistical Report - Traffic Collisions dataset (see section 2a) was sorted by location and all collisions proximal to an intersection were associated with one another. After associating traffic collisions with intersections and sorting the data, it was partitioned into intersections with RLCs and intersections without RLCs.

5d. Exploratory Data Analysis (EDA)

After performing extensive data cleaning and transformation to ensure the data was in an appropriate format, we proceeded with exploratory data analysis (EDA) to gain insight regarding the two main variables of interest: traffic volume and traffic collisions at each intersection. EDA for traffic volume and traffic collision was completed for intersections with and without RLCs.

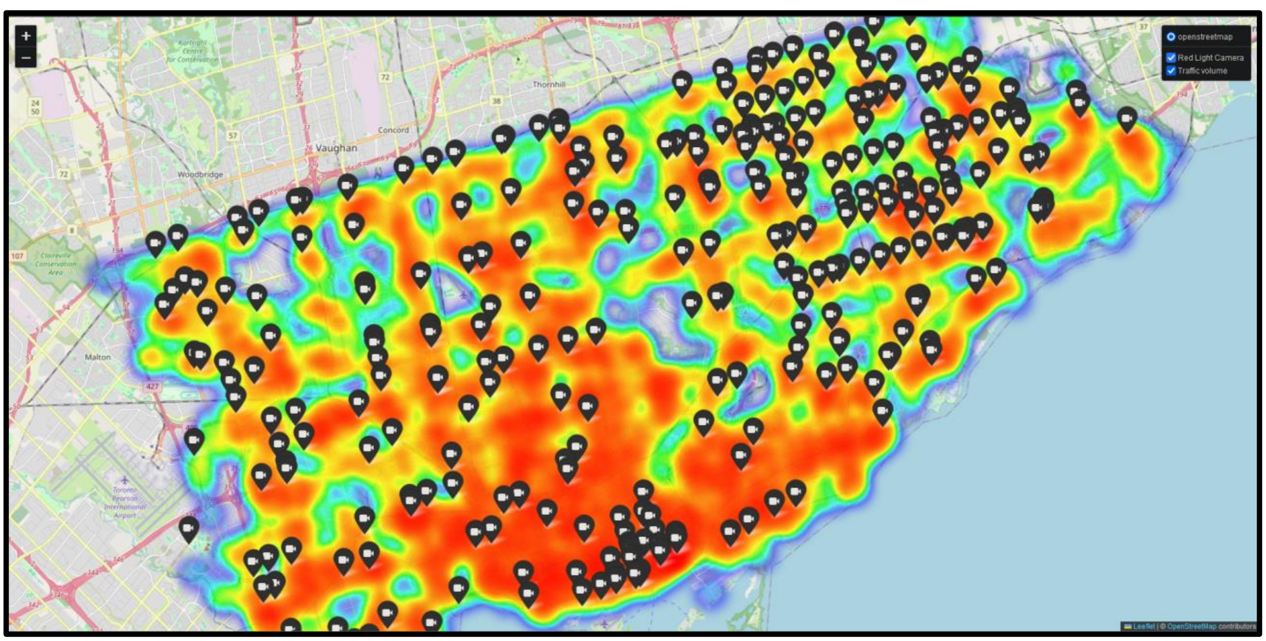

Figure 1: Heat map illustrating traffic volume at red light camera locations.

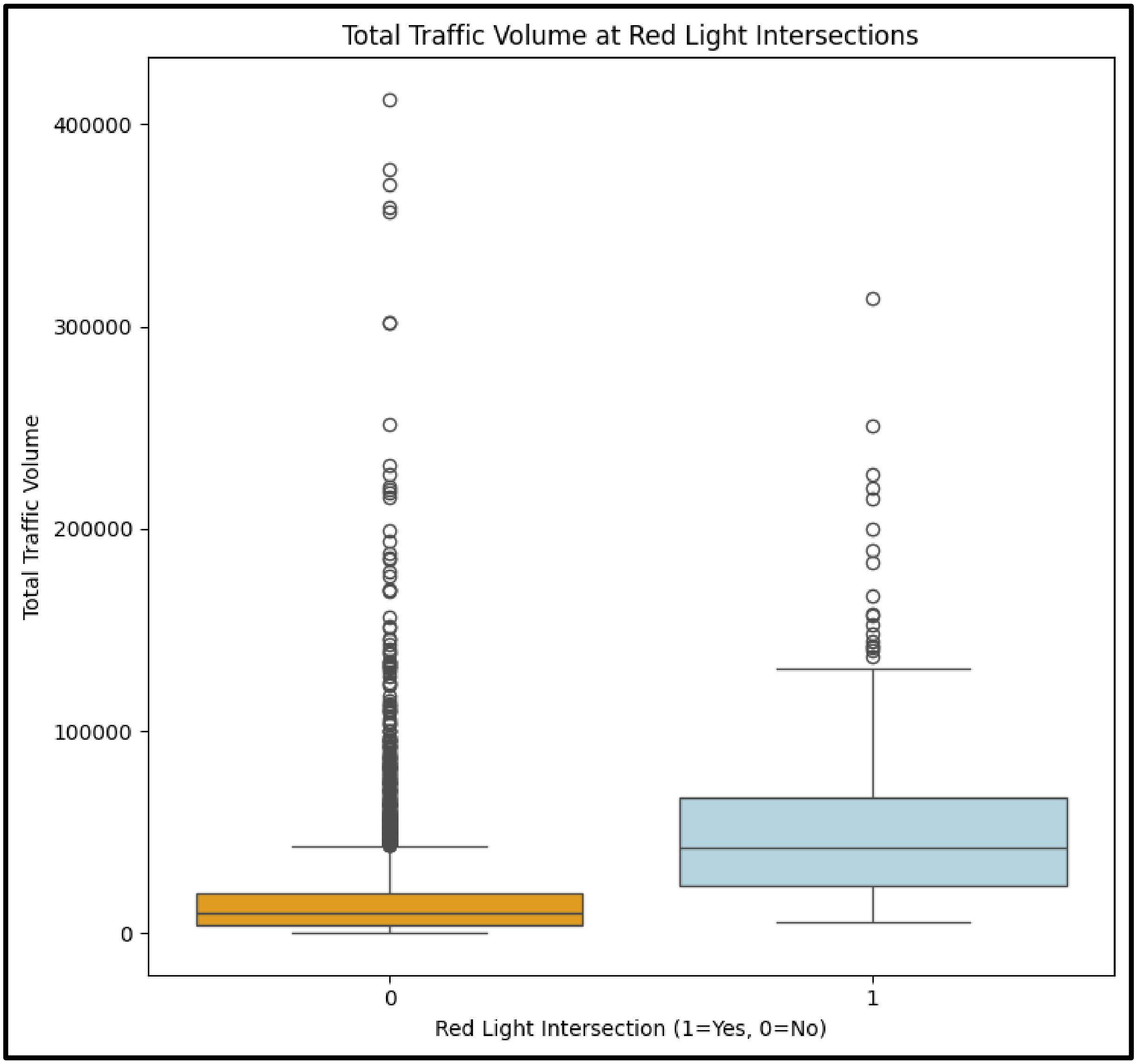

Figure 2: Boxplot of Traffic Volume for Non-Red Light Camera Intersections and Red Light Camera Intersections

The boxplot above (Figure 2) demonstrates the spread of traffic volume for non-RLC intersections compared to RLC intersections. It is evident that the spread of traffic volume is higher for RLC camera intersections compared to non-RLC intersections. Figure 1, the heatmap of traffic volume with RLC locations, illustrates that RLCs tend to be situated in high traffic volume hotspots. Additionally, from the boxplot (Figure 2), we can see that the boxplot for RLC intersections has a higher spread. This suggests that the City of Toronto installed RLCs at intersections that experience higher traffic volumes, with the hopes that these cameras would eliminate reckless driving behaviors.

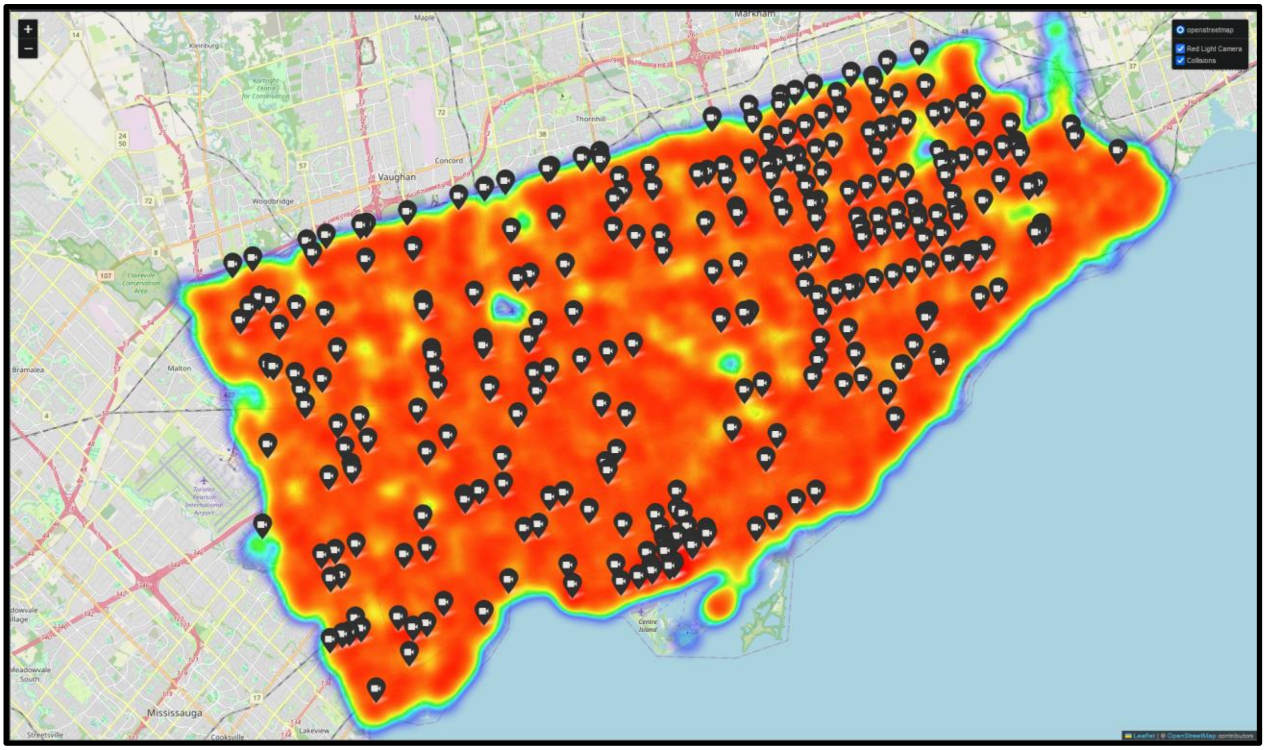

Figure 3: Heat map illustrating traffic volume at red light camera locations.

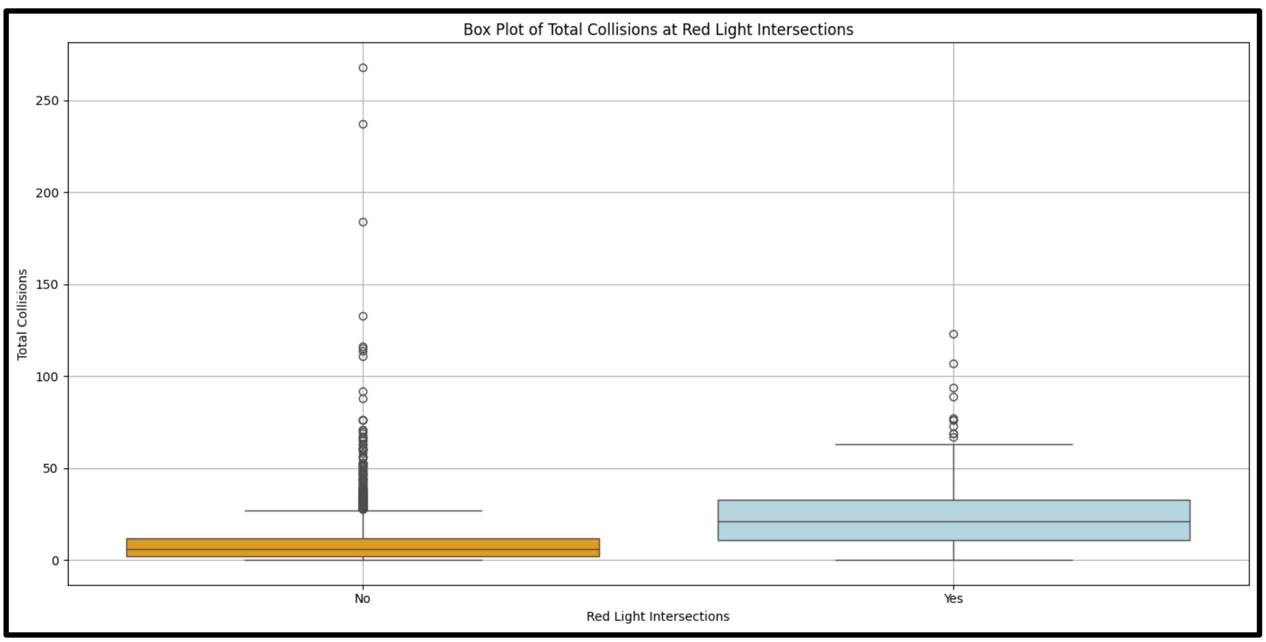

Figure 4: Boxplot of Traffic Collisions for Non-Red Light Camera Intersections and Red Light Camera Intersections

In Figure 3, we can observe that RLCs are situated in hotspots of traffic collisions compared to the surrounding areas. This is further corroborated by the box plot (Figure 4), which exhibits a higher collision distribution for RLC intersections. This confirms our assumptions regarding RLCs being installed at intersections experiencing higher traffic volumes and higher collisions. The findings from the two initial heatmaps and two box plots, suggests that we may observe a strong statistical relationship between traffic volume and traffic collisions.

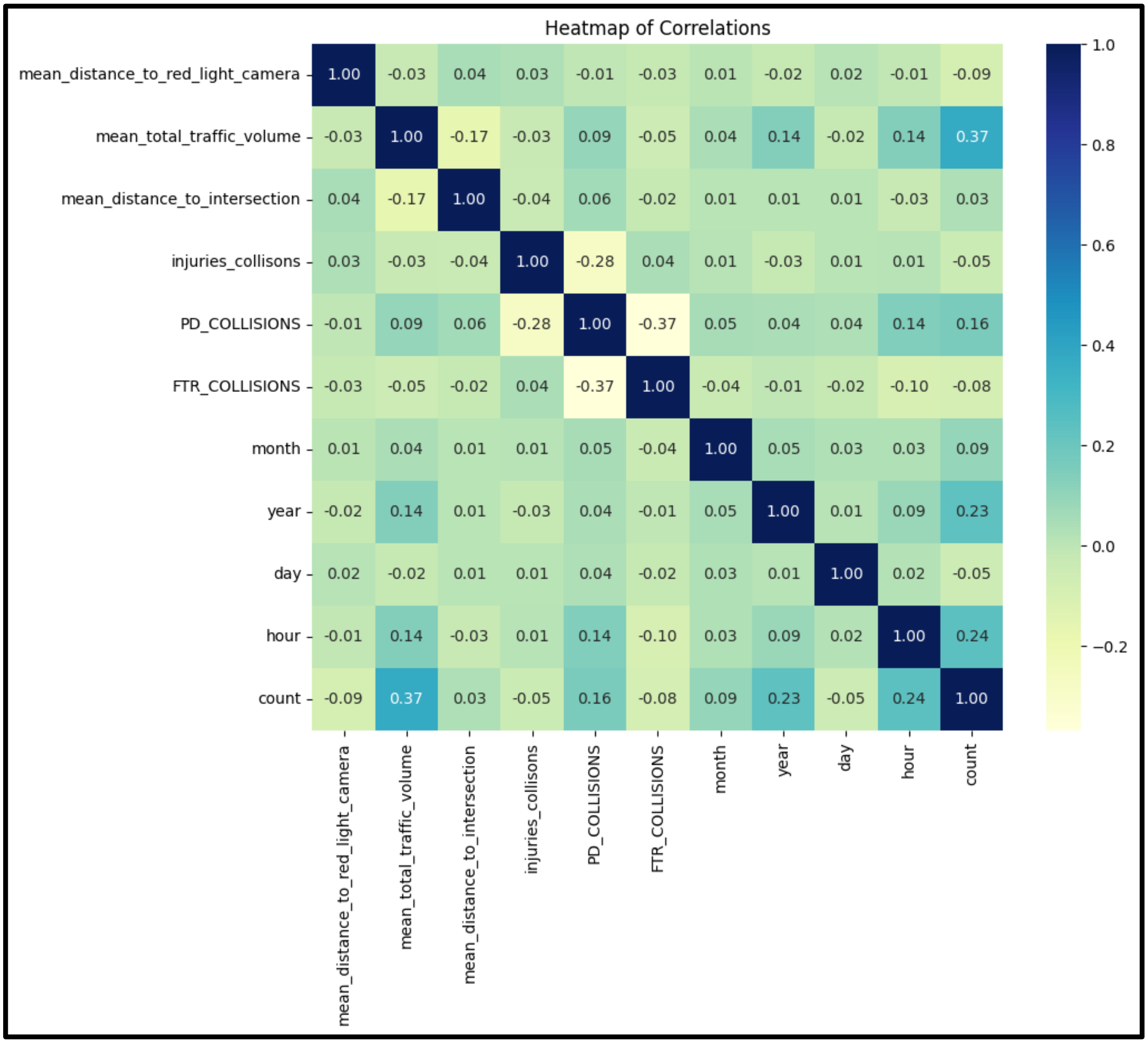

Figure 5: Heatmap of correlations for key columns of interest.

The above (Figure 5) is a heatmap of correlations for our chosen data. Inspection of this heat map reveals that two of our temporal data columns, year and hour of the day, demonstrate the highest correlation with the "count" variable, which is the count of traffic collisions at each intersection. While this heatmap does demonstrate that traffic volume is correlated with traffic collisions, further analysis was required to better understand how different temporal aspects contribute to traffic collisions.

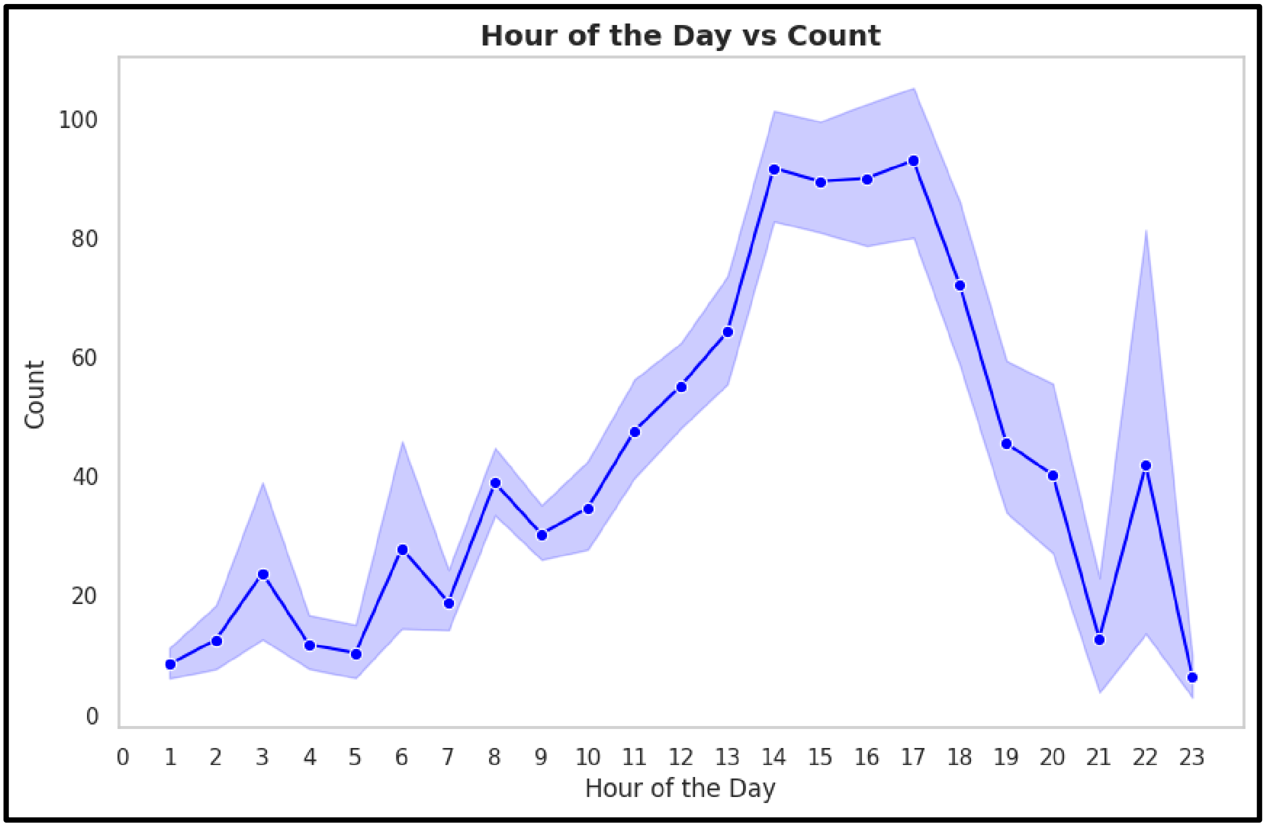

Figure 6: Distribution of motor vehicle collisions throughout the day.

The figure above illustrates the count of motor vehicle collisions by hour of the day. It can be observed that a majority of motor vehicle collisions occur between the hours of 1 to 6 pm (Figure 6). This is unsurprising however, as this period of each day coincides with peak business hours and returning rush-hour traffic. However, it is equally interesting to note that there is no observable peak in collisions during the morning hours that coincides with the typical morning rush-hour.

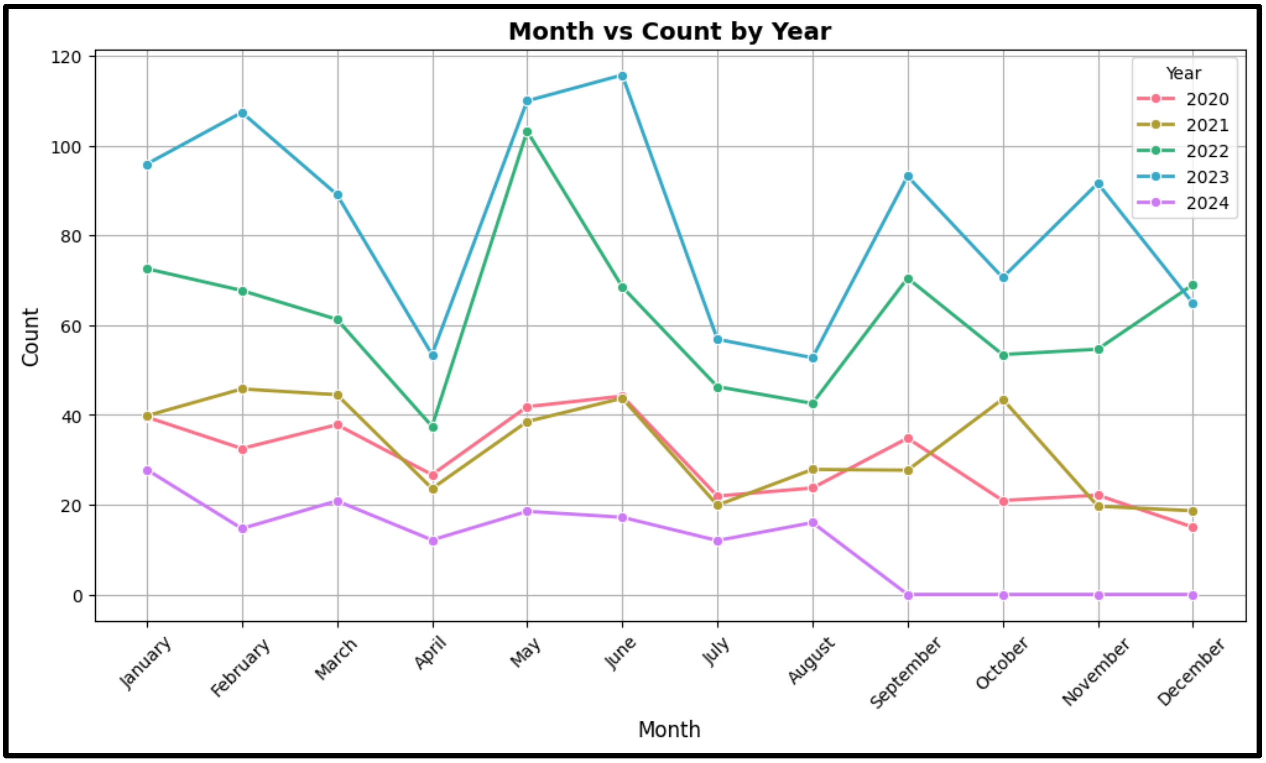

Figure 7: Distribution of collisions by calendar month from 2020-2024.

The plot above illustrates the count of collisions in Toronto by month between 2020 to 2024. It provides key insight regarding patterns of Toronto traffic. It can be observed that collisions are lower in each year from 2020 to 2022 compared to 2023 (Figure 7). It is likely that the lower number of collisions observed from 2020 to 2022 is due to the Coronavirus pandemic and the work-from-home policies that were active during this time. It should also be noted that the data for 2024 is incomplete as it was pulled in September 2024. Additionally, in April and July for each of 2020, 2021 and 2022, there is a sharp decrease in collisions. This might be explained by events that force streets to close such as block parties, and spring/summer festivals that frequently happen during these months. Further investigation would be required however, to confirm that this is indeed the reason.

Although the temporal aspects of the data that we have discussed above, are not the focus of this project, these aspects of the data are nonetheless interesting and inform future analysis, and continued work on this topic.

5e. Regression Analysis

To test our two hypotheses regarding (Hypothesis #1) RLC intersections experiencing fewer collisions and (Hypothesis #2) intersections with higher traffic volume leading to higher collision numbers (see section 4), we utilized Ordinary Least Squares (OLS) for polynomial regression. Additionally, we were interested in evaluating whether there is a significant difference between a regression model using RLC intersections and a regression model using non-RLC intersections. For OLS regression, traffic collision was the dependent variable, while traffic volume and the square of traffic volume were the independent variables. It should be noted that in all three models we created, the squared traffic volume coefficient was negligible and was thus omitted due to its insignificance on these models.

Table 1: Important regression values for three OLS models pertaining to all intersections, RLC intersections and non-RLC intersections, respectively.

Regression Model	Adjusted R-Squared	P-Value	Regression Intercept	Traffic Volume Coefficient
All Intersections	0.195	2.44e-142	33.8	0.0020
Red Light Camera Intersections	0.056	2.56e-4	119.9	0.0016
Non-Red Light Camera Intersections	0.132	1.26e-85	34.4	0.0015

It can be observed that the accuracy is quite low for the first regression model pertaining to all intersections, with an adjusted R-squared value 0.195 (Table 1). This is a weak model with regards to accuracy. It does indicate however, that several other factors might contribute to the occurrence of traffic collisions at an intersection than simply the number of cars that pass through an intersection (i.e. traffic volume). This is unsurprising, given that extensive research has been conducted in an effort to better understand the causes of traffic collisions. A single factor would likely never completely explain traffic collisions and in the future, other contributing factors would need to be considered to create a more accurate regression model. Our models for RLC intersections and non-RLC intersections demonstrated even lower accuracy, though this may be due to the reductions of the sample sizes for these models (Table 1).

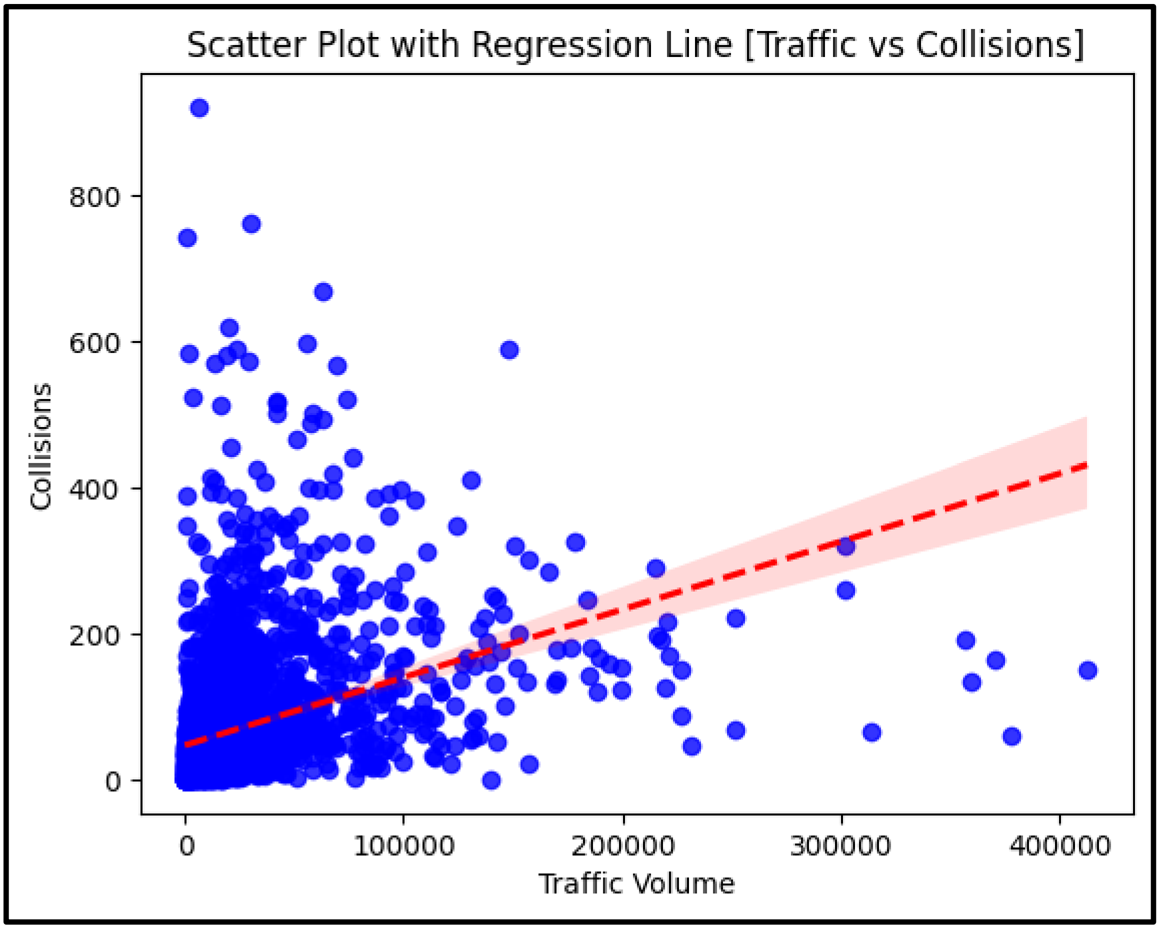

Figure 8: Scatterplot of Traffic Volume vs. Traffic Collisions for All Intersections with Regression Line.

In addition to the low accuracies, the P-values are also low for all three regression models. This allows us to confirm our hypothesis that traffic volume demonstrates a statistically significant positive linear relationship with traffic collisions (see Objective 1, Guiding Question 1, Hypothesis 1).

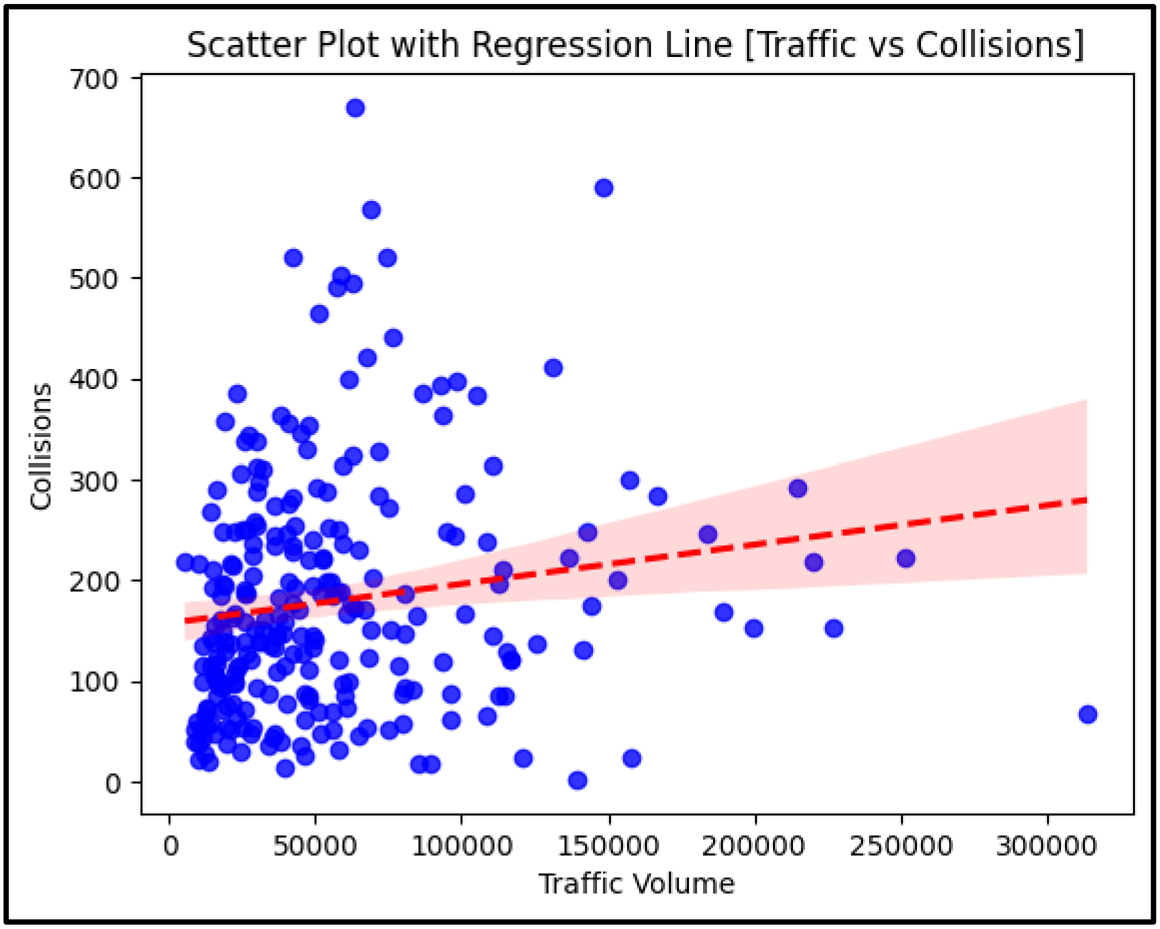

Figure 9: Scatterplot of Traffic Volume vs. Traffic Collisions for Red Light Intersections with Regression Line.

The coefficients of the regression models for RLC and non-RLC intersections may be the most interesting aspect of our regression analysis. The intercept of the RLC intersection regression model is much higher than the intercept of the non-RLC intersection regression model. This is unsurprising because our analysis of both RLC and non-RLC intersections revealed that RLC intersections had a high spread of motor vehicle collisions. However, despite the difference in the intercepts for both regression, the coefficient for the linear variable for traffic volume is nearly identical. This suggests that collisions occur at the same rate per traffic volume for both RLC and non-RLC intersections. Additionally, the entirety of the difference in the collision count spread at each intersection type, is entirely accounted for by the intercept of this regression model. Although the two models exhibited low accuracy, the lack of a difference with respect to traffic volume should be noted and considered in dialogue about the efficacy of RLCs at traffic intersections.

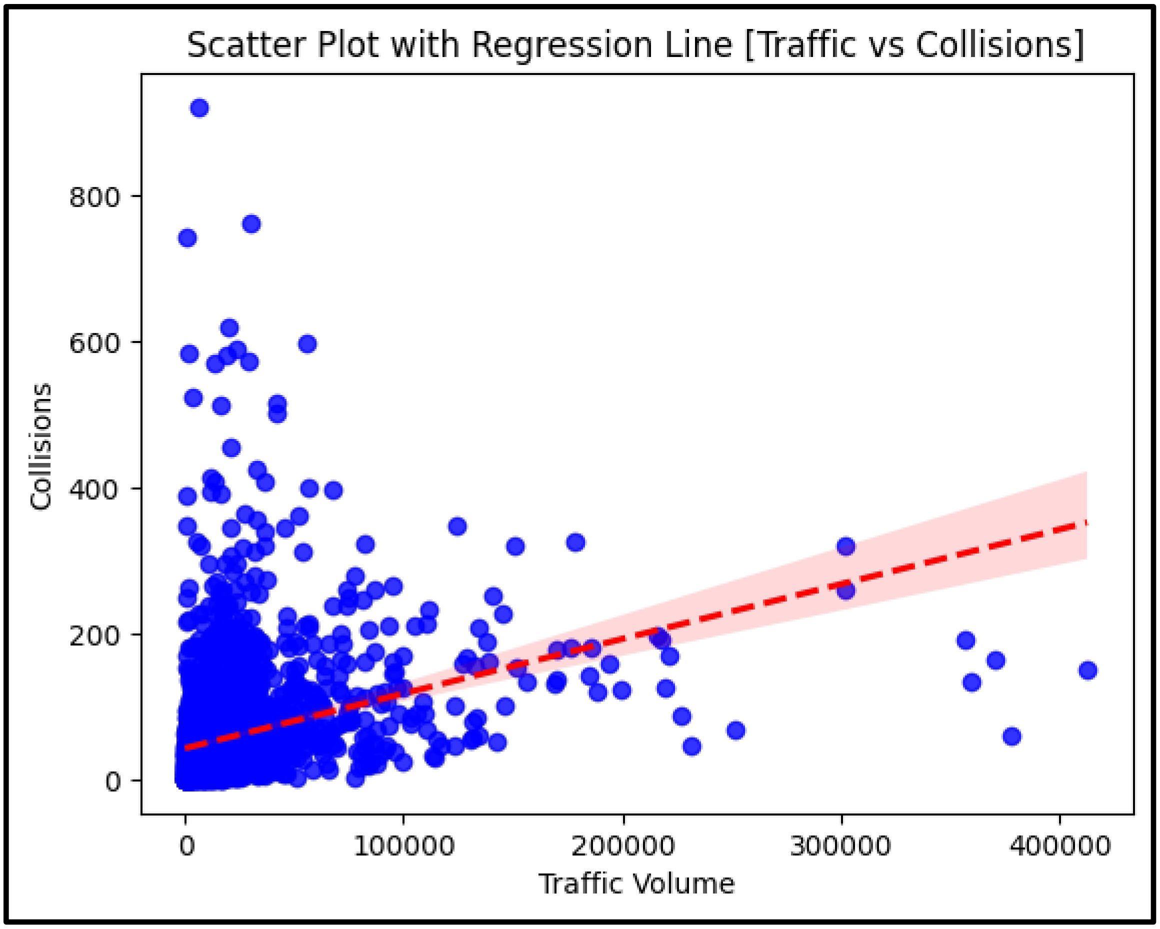

Figure 10: Scatterplot of Traffic Volume vs. Traffic Collisions for Non-Red Light Camera Intersections with Regression Line.

6. Discussion

From the comparison of our two regression models pertaining to RLC intersections and non-RLC intersections, we can concur with the Scientific American study (6) regarding the effect of RLCs on traffic in Houston. Likewise, our analysis uncovered that implementing RLCs may not make roads safer or lead to drastic improvements in road safety (6). Given that the differences between these regression models ultimately measure differences in behavior at these types of intersections, we are forced to conclude that RLCs may not make roads safer.

Firstly, the driving behaviors that RLCs are intended to combat to reduce crashes such as running red lights, could be offset by other driving behaviors that are also encouraged by the existence of RLCs that actually lead to increased rear-end collisions, such as stopping abruptly to avoid being ticketed. Another interpretation is that the effect of RLCs is more widespread and analyzing collisions on such a granular scale fails to capture the full extent of RLC impact on road safety. This might be explained by the fact that people adjusted their driving behavior at all intersections after RLCs were implemented, not just at intersections with RLCs. This model would therefore be ill-equipped to detect this kind of impact. Data from before these RLCs were implemented would need to be compared to current data, but this is challenging, as traffic patterns are subject to tremendous change over the course of years. As a result, even with before-RLC data and after-

RLC data, it would be difficult to determine if RLCs have a significant effect on improving road safety. The final interpretation of our models is that traffic volume is not the best parameter by which to evaluate the effect of RLCs on traffic collisions. This can be explained by the fact that regardless of the traffic volume at an intersection, only a few cars may actually be involved in reckless behavior that leads to collisions. It might not be the volume that directly contributes to collisions. With all of these valid interpretations of our result, along with the low accuracy of our models, without further analysis, the strongest conclusion that can be made is that red light cameras may not have a significant impact on road safety.

Despite this potentially flawed road safety model, we can confirm the positive relationship between traffic volume and traffic collisions (see Objective 2, Guiding Question 2, Hypothesis 2). This provides insight with regards to which road safety policies are more likely to have a measurable impact on reducing collisions on the road. Policies that reduce the number of cars on the road are likely to be the most effective, as these will be more likely to reduce traffic collisions as well. Policies like this are already widely implemented around the world, as cities have recognized the need to provide alternative modes of transportation along with the supporting infrastructure to be a viable replacement for former drivers. Notably, the Netherlands has reduced the rate of traffic-related deaths from 20% higher than the United States in 1975 to 60% less than the United States by the mid 2000s by implementing policies to redesign their infrastructure to incentivize walking, cycling, and using public transport as viable alternatives to driving that reduced the amount of vehicle traffic (10). These big infrastructure policies are more expensive and disruptive than installing RLCs at existing traffic intersections, but do deliver measurable results. Extensive infrastructure reform may not have much political will and therefore, be difficult to enact and pay for, but at the rate society pays for deadly traffic collisions monetarily and with the lives of its citizens, society will pay regardless.

7. Conclusions

We remain unable to definitively conclude to what extent RLCs impact road safety, as our regression models did not show any difference between the rate of collisions at RLC intersections and non-RLC intersections. Many factors contribute to motor vehicle accidents and focusing our analysis on traffic volume did not yield a significant difference. In future work, there may be other parameters of our data that we can include in our models or alternatively, another road safety model can be developed that would allow us to measure the effect of RLCs on road safety. Based on the models discussed in this report however, RLCs are not likely to significantly improve road safety.

While some studies have been able to conclude that RLCs reduce certain types of motor vehicle accidents, further research is required to determine their specific effects and limitations (4-5). Policies that seek to improve road safety by modifying driving behaviors with punitive measures may be misguided, as these policies may contribute to increased driving-related fear. Since our results have been able to confirm a positive relationship between traffic volume and traffic collisions, alternative road safety policies that reduce the amount of vehicle traffic on the road may prove to be effective. We have seen foreign infrastructure reforms create multiple viable transportation alternatives to driving thereby, reducing the number of vehicles on the road and the number of fatalities on the road.

References

1.https://www.3m.com/3M/en_US/road-safety-us/resources/road-transportation-safety-center-blog/full-story/~/cities-of-the-future/?storyid=44e91529-6a46-4f81-a608-0dda4f23bd0d

2.https://thevarsity.ca/2024/01/13/opinion-torontos-new-years-resolutions-must-facilitate-its-population-growth/

3.https://www.blackboxmycar.ca/pages/differences-between-traffic-cameras-red-lights?srsltid=AfmBOooJVIz_SnV4FK8UsHcGlvX0w_lKcb49skzGeobzZjLsnSxdIYcN

4.https://www.thestar.com/news/gta/toronto-issued-90-000-red-light-camera-tickets-last-year-a-rise-of-nearly-50/article_1def0a65-f9a9-5a37-a951-deede6294b24.html

5.https://www.toronto.ca/services-payments/streets-parking-transportation/road-safety/vision-zero/vision-zero-dashboard/red-light-camera-vision-zero/#:~:text=A%20Red%20Light%20Camera%20is,historical%20pedestrian%20and%20cyclist%20collisions.&text=*Please%20note%20the%202016%20Red,between%20(2007%2D2011).

6. https://www.scientificamerican.com/article/red-light-cameras-may-not-make-streets-safer/

7. https://open.toronto.ca/dataset/police-annual-statistical-report-traffic-collisions/

8. https://open.toronto.ca/dataset/red-light-cameras/

9. https://open.toronto.ca/dataset/traffic-volumes-at-intersections-for-all-modes/

10. https://www.scientificamerican.com/article/we-need-to-make-cities-less-car-dependent/