Road to Safety: Unveiling Strategies to Reduce New York City Traffic Accidents

Ashwin Ajit (aa794), Mohammad Islam (mai54) , Kelly Lambert (kjl222), Shriya Bansal (sb885)

Table of Contents

Abstract

In the busy city of New York, vehicle incidents are a common occurrence which increase the city's vehicular traffic. The NYC Department of Transportation needs to address this issue to ensure pedestrian and motorist safety, and spend less on settlement costs. After a thorough analysis of the two data sets, we created linear and logistic models that enabled us to analyze seasonal patterns in crash occurrences, when accidents occur the most, and contributing factors. These seasonal patterns show that the number of vehicle accidents and people injured peak during the summer months. Accidents occur most frequently between the hours of 3 and 5pm, specifically on Thursday and Friday. We also observe trends in accidents caused by driver inexperience, which in particular spikes in summer months.

Our first logistic model reveals that the most prevalent causes of accidents are driver inattention, speeding, and failure to yield right of way. An additional driver's safety course could address these causes as well as benefit the inexperienced drivers in summer months. Our second logistic model reveals that more injuries occur at lower traffic levels. Thus we recommend an increase in road infrastructure (speed bumps, traffic signals, etc.) and police presence on the roads to mitigate driver inattention at both high and low levels of traffic.

We were able to then ask predictive questions and make models that represented these well. Two questions that were investigated were what the number of accidents would look like in future years and what the probability is of an accident occurring based on traffic volumes.

It is vital that policy makers understand when accidents occur the most and what the leading causes are that way traffic fatalities decrease and all New Yorkers, drivers, bikers, and those who walk are safe traveling around the city.

Using Traffic Accidents Data to Understand Trends in NYC

In the busy city of New York, vehicle incidents are a common occurrence which increase the city's vehicular traffic. However, the damages these incidents cause go beyond inconvenience. Half of all accidents cause injuries, and in 2022 alone, these accidents caused 255 fatalities [1]. Additionally, in 2021, traffic accidents cost $130.1 million in settlements [2]. The NYC Department of Transportation needs to address this issue to ensure pedestrian and motorist safety, and spend less on settlement costs.

To investigate the frequency of NYC accidents, we explored the NYC traffic accident dataset [4], which contains information on motor vehicle accidents within the NYC boroughs of Manhattan, Brooklyn, Bronx, Queens, and Staten Island, from 2013 to 2019. Each record represents one vehicle collision with its date, time, location, vehicles, and contributing factors

such as daydreaming or intoxication. This dataset provides a great opportunity to gain insight on motor accidents and to understand the circumstances surrounding them.

While investigating this problem, there were a few limitations in the data being analyzed. To begin, for each accident that was listed, there were no details provided regarding the demographics of the drivers. Not knowing the exact age/ level of experience of the drivers, as well as their general background made it more difficult to make specific demographics-based recommendations for the future. This information would be helpful because if a clear trend could be established that those who just began driving are getting into the majority of car accidents, then a possible recommendation would be requiring more hours of supervised driving prior to taking the road test. Typically, a drivers test does not take place on major highways or in high traffic areas, so before drivers are sent out on their own, their ability to drive on these types of roads is not tested. On the other hand, if the majority of accidents are caused by the elderly, then a possible recommendation is to require those above 65 years of age to take their road test again.

Additionally, 37% of the data under contributing factors is null. This is a large portion of the listed car accidents that have unknown causes. With such information missing, it is unclear if these accidents are caused by weather/ road conditions, vehicle malfunction, human error, or lack of awareness of the driver. Without this information, making recommendations becomes challenging as we are unaware of the causes for over a third of the dataset.

Utilizing Traffic Volume Data to Enhance Understanding of NYC Accidents

In the original data set, titled 'New York City Traffic Accidents', there was no information regarding traffic volumes during the time of the accidents. After conducting further research, we found a second dataset, 'NYC Automated Traffic Volume Counts,' which collects the traffic sample volume counts at bridge crossings and major roadways. With this information, we are now able to investigate if there is a correlation between the frequency in which accidents occur and general traffic volumes in boroughs.

Through analyzing these datasets, we hope to help the local government implement the optimal reforms needed to minimize traffic accidents. Additionally, we will be able to inform drivers of the seasons, times of the day, and boroughs of New York City to proceed with more caution when driving. By increasing driver awareness and improving government policies such as VisionZero which is a program we will focus on later in the report, we can minimize the damage these accidents cause and save the lives of drivers, passengers, and pedestrians.

Exploring NYC Accidents: Research Questions for an In-Depth Analysis

We hope to be able to make strong recommendations that can inform both policy-makers and drivers to minimize accidents. As a result, we shed light on the following three research questions to gather more information about traffic accidents that occur in NYC.

1. *Are there seasonal patterns in crash occurrences based on the total percentage of accidents per month?*

This question allows us to analyze the disparity in the number of accidents among different seasons. We can use this data to see which road and weather conditions, based on the season, play the largest role in traffic accidents.

2. *When do accidents occur the most in NYC and what are the most popular contributing factors?*

To answer this question we must also investigate the following sub questions: *what hour of day do accidents occur most frequently?,* and *how does the total number of accidents differ based on what day of the week it is?* Our goal in answering these questions is to narrow down the time

range where accidents are occurring the most. By investigating the contributing factors to accidents in this time range, it will let us develop a probability for different accident types and introduce recommendations in reducing these factors.

3. *What will the number of accidents look like in future years, and what is the probability that an accident will occur based on traffic volume?*

We can use our datasets to help us forecast to see what the trend is for accidents in the future. This will help us illustrate the probability that these accidents will have an injury occurring. As a result, it allows us to see how necessary it is for change to occur and what needs to be implemented through reducing volume .

Seasonal Patterns in Vehicle Accidents

From **Figure 1** we can see that the number of vehicle accidents and number of people injured peaks during the summer months. To interpret this data, we must account for improved school year schedules. We backed up this observation by analyzing crash patterns in inexperienced drivers. Students are usually off from school or college from the beginning of June till the end of August, and we can see that there is a general increasing trend for accidents due to driver inexperience from June to August from **Figure 2** and it peaks around the end of June and middle of July. We can thus conclude that the presence of young drivers causes this spike in accidents due to inexperience. We hope that this information will allow us to focus on implementing additional safety measures that could benefit young drivers and how to potentially create preliminary policies to help reduce fatalities and injuries caused by inexperience.

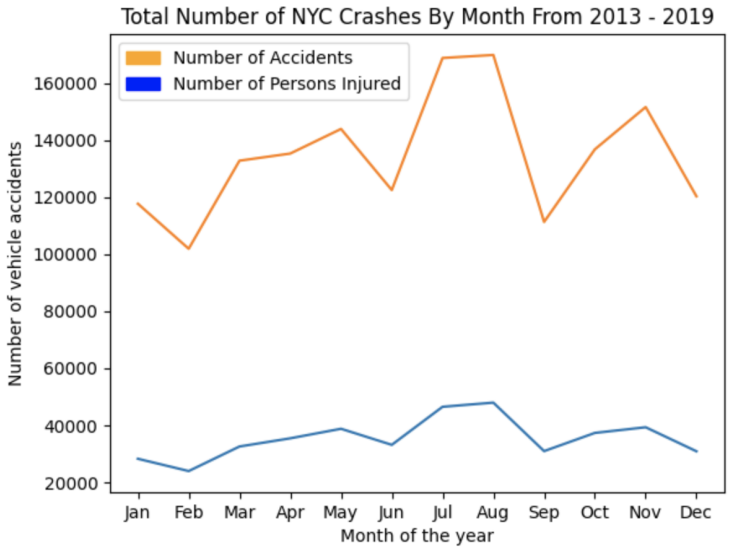

Fig. 1: Plot of the frequency of NYC vehicle accidents against the month of the year that it occurs. Accidents occur prominently during the summer months.

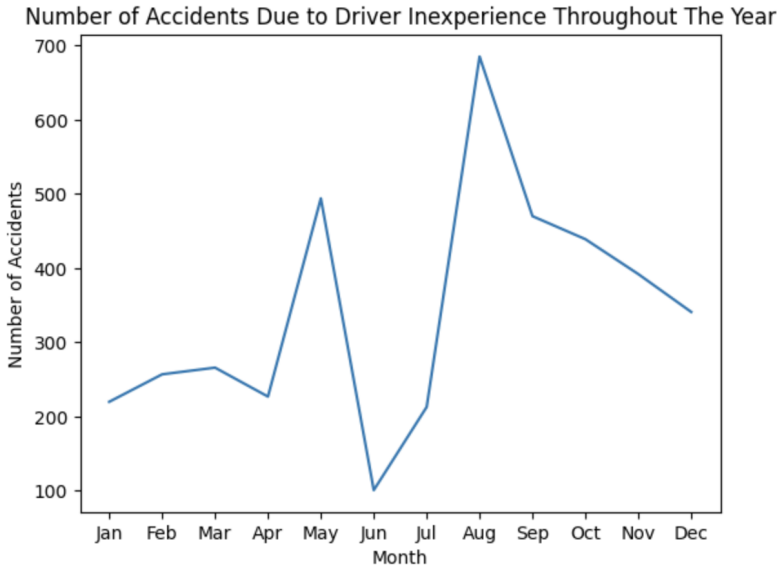

Fig. 2: Plot of the frequency of NYC vehicle accidents against the month of the year that it occurs based on driver inexperience. Accidents occur prominently during the summer months of July till September.

Exploring Contributing Factors For Accidents During Peak Accident Hours

Most Frequent Hour of Vehicle Accidents

Choosing question one for our initial analysis, we manipulate the dataset to find the hour in which each accident occurs in integers, 0-23. We then grouped the accidents into which hours they occurred and counted the number that occurred in each to plot the result. **Figure 3** shows that accidents occur most frequently during 3-5 PM in NYC. These hours lie in the middle of the evening time where the traffic on the road is at its highest point. These results agree with another analysis done by Shaaban and Ibrahim, who identify that accident frequency peaks at these hours [7]. In their study, they determined that the highest number of crashes occurs between 5-6 PM and in our data we can see that the highest number of crashes also occurs between 4-6 PM. We can then narrow the time and date range further by exploring what days of the week accidents are most prominent.

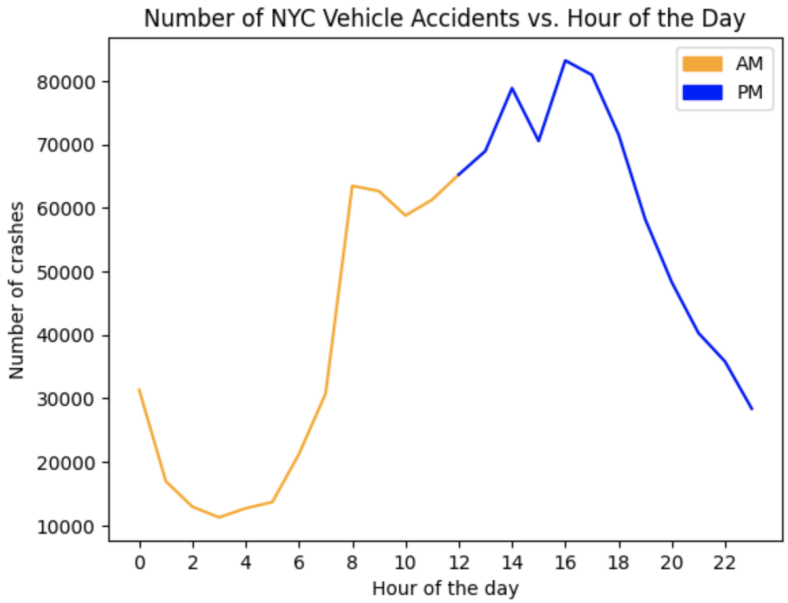

Fig. 3. Plot of the frequency of NYC vehicle accidents against when the hour in which they occur. Accidents occur prominently during 16-18 hours (4-6 pm).

Days of the Week Where Accidents Occur The Most

In **Figure 4**, we observe that the number of accidents is fairly consistent and stays relatively the same over Monday through Wednesday. However, it increases on Thursday and peaks on Friday, and then dips below the Monday through Wednesday levels on Saturday and Sunday. Traffic congestion plays a large role in the number of traffic accidents that occur and **Figures 3 and 4** allow us to see at what times the roads are the busiest. These figures allow us to comprehend when accidents are occurring the most usually and will in turn prove to be the most effective times to have potential policy changes to be implemented in.

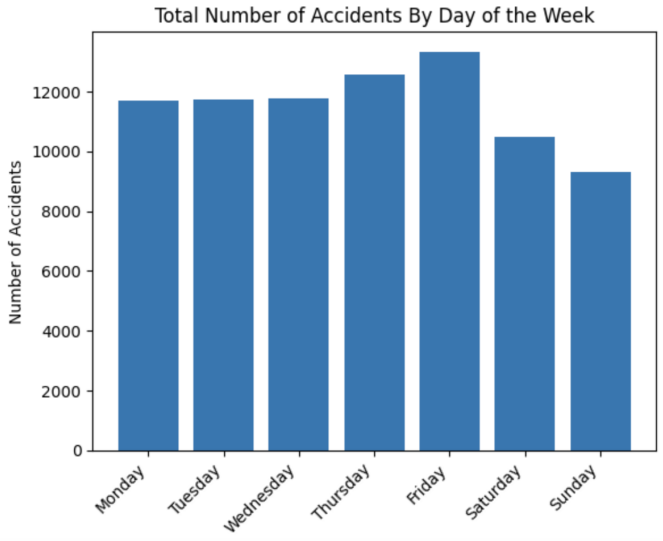

Fig. 4: Bar Graph of the frequency of NYC vehicle accidents against what day of the week they occur. Accidents occur prominently on Fridays.

Predicting the Probability of Accident Causality

From our analysis above with **Figure 3** and **Figure 4**, it is clear that the most prominent time range of an accident occurring is Thursday and Friday during 3-5 P.M. We filtered the dataset for this time range and used logistic regression, a machine learning technique to predict the probability of an outcome, on the dataset for each contributing factor respectively. Our regression split our data into 80% for training and 20% for testing where the probabilities on the graph are the result of the testing data. Looking at **Figure 5**, we can observe that the most frequent contributing factor for a crash during this high volume time range can be driver inattention, failure to yield right-of-way, unsafe speed, and traffic control disregard, with driver inattention having the most probable outcome. This brings our attention to policymakers to have these top factors in mind when thinking how to reduce the number of accidents. We now have narrowed down a time range where accidents are most prominent and probabilities of casualty for them. After finishing our investigation, we will be able to touch upon potential remedies for these contributing factors.

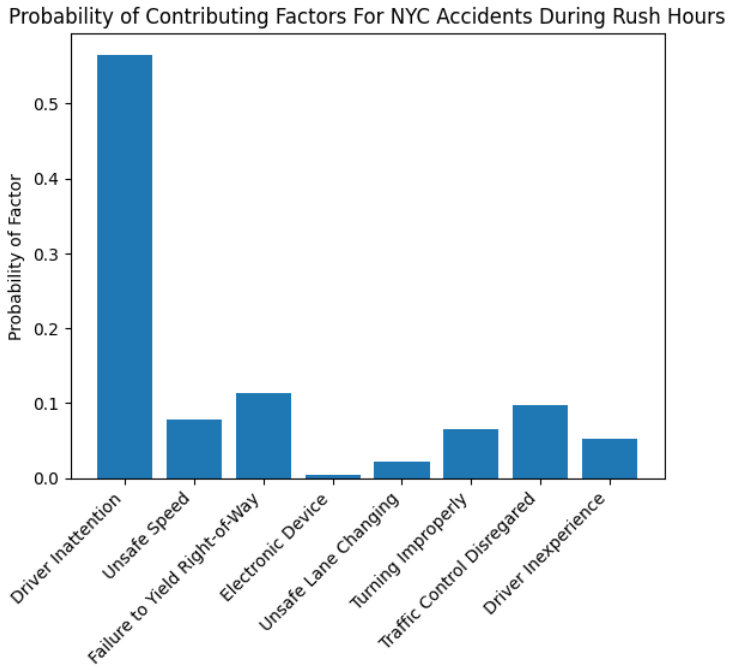

Fig. 5: The probability of popularly contributing factors during NYC rush hours on Thursday and Friday. We can observe that the most frequent cause of an accident is driver inattention.

Forecasting Accident Tolls and Injury Probability For The Next Decade

From **Figure 6** we can see the varying number of accidents year to year. Using the accident data from 2013 to 2019 as a training set, we were able to create a predictive linear regression model to forecast the total number of accidents in NYC between 2020 and 2030 as shown in **Figure 7**. There was limited numerical data when creating the predictive model, resulting in a graph that shows a linear increase throughout the years. The information that we had when creating this model was in regards to street names, vehicle types, and text information. We tackled these challenges by enumerating categorical variables that were available such as borough, contributing factors, hour of the day, and day of the week to introduce more numerical values into our dataset and be able to provide enough predictors to show a valid relationship between our forecasted number of accidents and predictors. Despite these limited factors, we

were still able to find a plausible trend from the predictors of our dataset such as vehicle types, street names, contributing factors through a slightly straight variance line through a normalized residual plot in **Figure 8**. In this linear model we can see that the number of accidents are growing by at least a thousand every single year. This suggests that change is necessary within the transportation policies or the number of injuries and fatalities will continue to rise as our previous analysis dictated a proportional relationship.

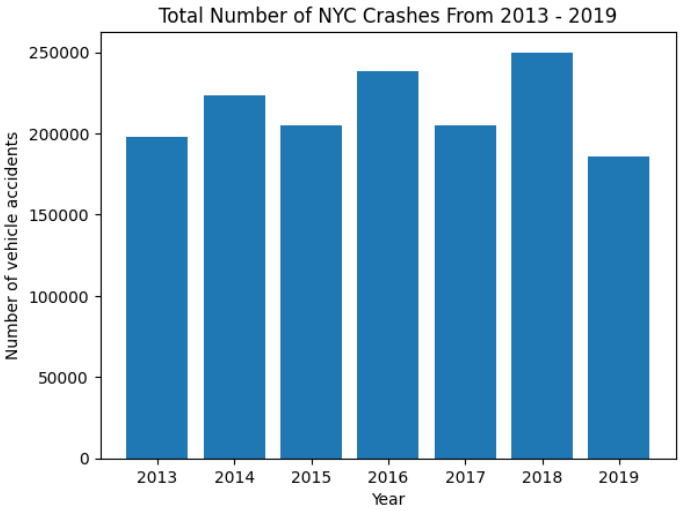

Fig. 6: Bar graph of the total number of NYC vehicle accidents that happened in each year from 2013-2019.

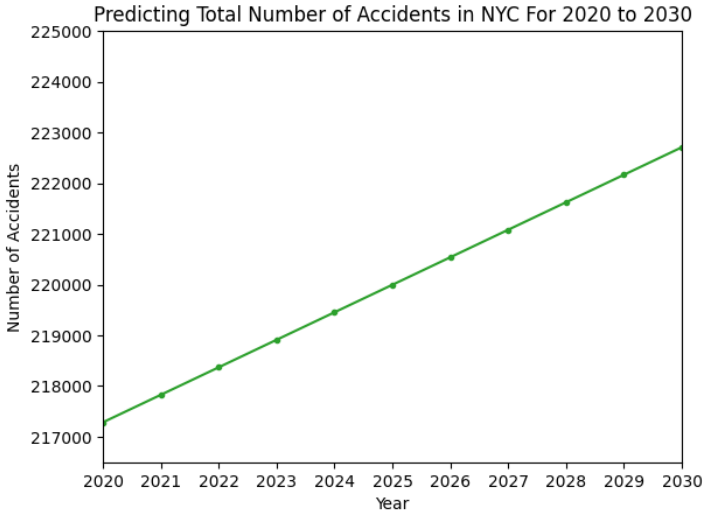

Fig. 7: Plot of the predicted frequency of NYC vehicle accidents from 2020-2030. We can see # of accidents going up every year.

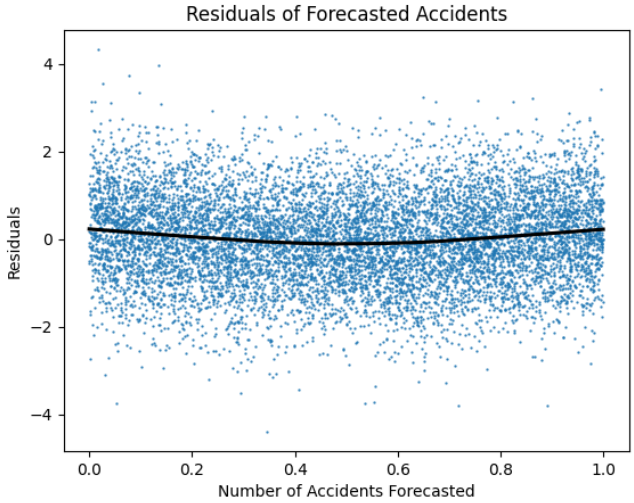

Fig. 8: Plot of the residuals for figure 7. We can see a potentially linear relationship in our model and the forecasted accidents by a slightly straight line around the value variance of 0.

Predicting The Probability of Injurious Accidents Based On Traffic Volume

In **Figures 9 and 10**, we investigate accident severity with respect to the overall traffic volume in a given borough. In our traffic volume dataset, the number of cars that pass through a given location are measured in various intersections, bridges, etc. of the city. We average these values together in each borough to get a general sense of how heavy traffic is around the time and place when a given accident occurred. We then train a logistic regression model to predict the probability of that accident having an injury when given a mean traffic volume in the

borough in which an accident occurred. We train this model on 80% of our data and test it on the other 20%. The predictions on our test data are plotted above. Our model's predictions clearly demonstrate that accidents are more likely to be injurious at lower traffic levels. Furthermore, this relationship is even more dramatic when it comes to fatalities, as we see a vaguely inverse correlation between the predictions and mean traffic volume. We investigate the underlying causes of accidents in the next section to further shed light on this relationship.

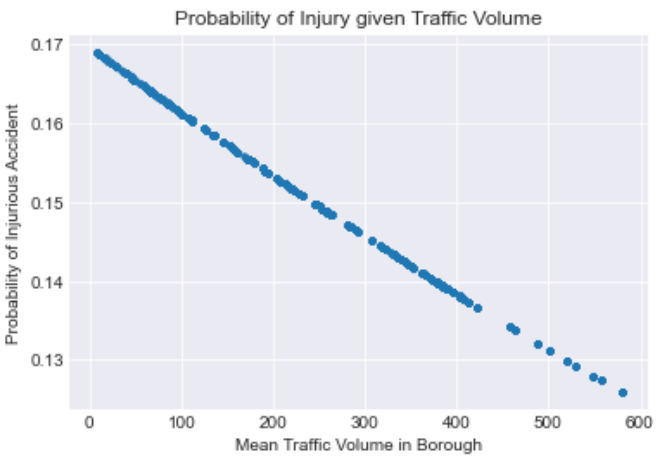

Fig 9: Plot of the probability of fatality from accidents given the mean traffic volume in the borough in which the accident occurred.

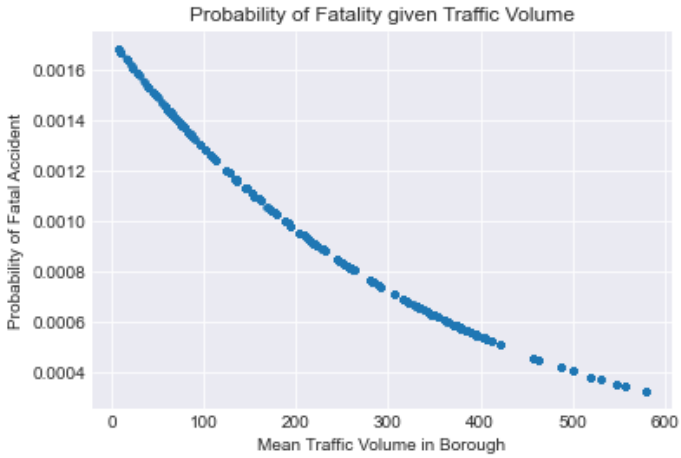

Fig 10: Plot of the probability of fatality from accidents given the mean traffic volume in the borough in which the accident occurred.

Taking Action by Recommending Policies to Address Traffic Accidents

Our goal is to grow programs such as Vision Zero because of the change that it has created in just a few years. Vision Zero is a program that was created in 2014 by the New York City Mayor Bill de Blasio. Its purpose is to eliminate traffic deaths and related injuries on NYC streets by 2024. This program used a data driven approach that identified hotspots in the city that caused the majority of traffic fatalities, and implemented a targeted plan across the city. For example, in 2019, 21.4 miles of protected bike lanes were installed. This same year, overall deaths were down 26 percent and pedestrian deaths were down 33 percent, when compared to 2013 (the year before implementation). Continuing to grow Vision Zero with policies that target the same mission will decrease the number of fatal accidents seen in the years 2020-2030. Vision Zero has proven to do as much, and with similar programs, these numbers can be lowered even further [6].

As concerned drivers and citizens, we ask that those in government roles take action. An example of policy that can be made using this information is requiring all drivers who tend to commute in these hours to take a defensive driving course. The major contributing factors of traffic control disregard, failure to yield right of way, unsafe speed, and driver inexperience could all be addressed with such a course, providing knowledge and tools they need to drive safely especially during the hours of 3 and 5pm, where accidents occur most frequently. Additionally this would help mitigate the higher accident frequency in summer months caused by inexperienced young drivers. We also recommend that policymakers generally should increase police presence on roads and build additional road infrastructure such as traffic signals or speed bumps. This would force drivers to pay more attention to the road which would mitigate the

leading cause of vehicle accidents: driver inattention. Police presence and road infrastructure would also be helpful for when traffic is low, which we discovered is when more injurious accidents occur, as drivers would have to make stops and risk getting tickets for poor driving practices regardless of the level of traffic.

References

[1]https://www.silive.com/news/2023/01/nyc-traffic-deaths-drop-in-2022-but-remain-higher-than-the-low-seen-in-2018-data-shows.html
[2]https://comptroller.nyc.gov/reports/wreckless-spending/
[3] https://www.nyc.gov/site/nypd/stats/traffic-data/traffic-fatalities.page
[4] https://www.kaggle.com/datasets/new-york-city/nypd-motor-vehicle-collisions
[5]https://rmfwlaw.com/blog/car-accidents/leading-causes-of-teenage-driver-car-accidents-in-new-york
[6] https://www.nyc.gov/html/dot/html/pr2023/vision-zero-fatalities-dropped-2022.shtml
[7] https://reader.elsevier.com/reader/sd/pii/S2352146521005809?token=4F2D1A1AD6E4949FDAAF11B295054EC4E216DDD806AA734E9BB7FA0A21BDB52494D35ABAE2A096B58EFE7AC9C9FB195D&originRegion=us-east-1&originCreation=20230504212717
[8] https://www.kaggle.com/datasets/aadimator/nyc-automated-traffic-volume-counts
[9]https://journals.ametsoc.org/view/journals/apme/46/8/jam2516.1.xml