Investigation of factors associated with traffic incidents in NYC and review of *Vision Zero* effectiveness

CUNY SPS - DATA698.4 Capstone

Sreejaya, Vuthy and Suman

1 Abstract

This paper investigates the factors associated with traffic violations and accidents in NYC and reviews the results of the *Vision Zero* plan, a program created in 2014 in an attempt to prevent the traffice incidents in NYC. Today in New York, approximately 4,000 New Yorkers are seriously injured and more than 250 are killed each year in traffic crashes. Being struck by a vehicle is the leading cause of injury-related death for children under 14, and the second leading cause for seniors. On average, vehicles seriously injure or kill a New Yorker every two hours.The Vision Zero Action Plan is the City's foundation for ending traffic deaths and injuries on our streets.

2 Keywords

"traffic tickets, motor vehicle collission, vision zero, arterial slow zones, pedestrian"

3 Introduction

In an effort to reduce traffic fatalities and injuries, NYC released the Vision Zero Action Plan in 2014. The Vision Zero action plan included many initiatives but one of the more known initiative was reducing the citywide speed limit from 30 mph to 25 mph. How effective was reduction of the speed limit at lowering injuries and fatalities? Does reducing the speed limit actually decrease the number of motor vehicle accidents and/or traffic violations? As more traffic violations can lead to increased accidents, can we identify traffic hot spots (areas with recurring traffic violations)? In this study, we would plan on applying data mining algorithms like K-means and SVD to identify clusters in the data.

Additionally, as part of this study, we would like to answer several secondary questions, such as:

- Who are involved in majority of the traffic violations?
- What are the key contributing factors to the traffic accidents?
- Where are the hot spots? Can we identify locations with excessive number of crashes of a particular type?
- When is the majority of the traffic incidents occur?
- Why are some locations more prone to accidents than others?
- How can we prevent traffic accidents from occurring?

We believe this research would help provide valuable insights into the effectiveness of the Vision Zero's intended goal of eliminating injuries and fatalities from motor vehicle accidents.

4 Literature Review

We have reviewed few articles, news items, journal papers related to the traffic incidents and factors causing the collisions. The research paper (Yuan, Zhou, and Yang, 2017) analysed motor vehicle crashes from 2006 to 2013 in the state of Iowa, and tried to predict whether an accident will occur , for each road segment in each hour. The journal article (Chang and Chen, 2005) focused on the data mining of tree-based models to analyze freeway accident frequency, leveraging the CART model and a negative binomial regression model.The research (Eisenberg, 2004) study investigates the relationship between monthly precipitation and monthly fatal crashes, using the negative binomial

regression approach. We also reviewed the journal article (Jackson and Sharif, 2005) which focuses on the impacts of the rainfall on traffic safety in Texas, by analysing the relationship between crashes and rainfall.

5 Methodology

5.1 Data Exploration

The datasets we used for this analysis are from publicly available data from NYPD motor vehicle collission, which is a breakdown of every collision in NYC by location and injury, and *traffic tickets* data from NY data site. We plan to get various *vision zero* datasets from NYC vision zero datafeeds.As part of the data exploration, we will perform *Exploratory data analysis* to better understand the relationships in the given data including feature distributions,correlations and basic summary statistics. We also plan to identify the outliers, missing data and any look for invalid data.

5.2 Data Preparation

As part of the data preparation, we plan to fix the data issues noticed in the exploratory analysis, which involves treating the outliers, missing data, invalid data etc. We also plan to identify the data classifications, and create dummy variables wherever required, and will convert the categorical data to numeric data which would help us later in model building.

5.3 Model Development

Our primary objective here is to investigate the factors associated with the traffic incidents and understand how effective is *vision zero* initiative. We would also be interested in identifying the hot spots using the clustering methods like K-Means, Hierarchical clustering.

5.4 Model Validation

6 Experimentation and Results

6.1 Data Exploration and Preparation

The below are the data sets we use to explore the NYC traffic incidents.

Dataset

NYC Traffic Tickets
NYC Motor Vehicle Collision
Vision Zero

NYC Accidents:

There are 1226374 observations from the collision dataset.

By grouping the accidents by borough and year, we can see the accident counts in each borough of NYC by year, as shown in fig. *Accidents by borough* :

Plotted the injuries by year by combining all injuries 'NUMBER OF PERSONS INJURED','NUMBER OF PEDESTRIANS INJURED','NUMBER OF CYCLIST INJURED','NUMBER OF MOTORIST INJURED':

The fatalities plot (which include all of 'NUMBER OF PERSONS KILLED','NUMBER OF PEDESTRIANS KILLED','NUMBER OF CYCLIST KILLED','NUMBER OF MOTORIST KILLED') shows the fatalities by year.

From the above plots, the fatalities appears to be slightly decreased after 2014, however no such indication in case of injuries. So, there is no clear indication if the *vision zero* is the cause for the above reduction in fatalities. We need to further evaluate the same with specific vison zero data feeds.

```
Data columns (total 29 columns):
DATE                            1226374 non-null object
TIME                            1226374 non-null object
BOROUGH                         876463 non-null object
ZIP CODE                        876369 non-null object
LATITUDE                        1010457 non-null float64
LONGITUDE                       1010457 non-null float64
LOCATION                        1010457 non-null object
ON STREET NAME                  986184 non-null object
CROSS STREET NAME               918987 non-null object
OFF STREET NAME                 192420 non-null object
NUMBER OF PERSONS INJURED       1226374 non-null int64
NUMBER OF PERSONS KILLED        1226374 non-null int64
NUMBER OF PEDESTRIANS INJURED   1226374 non-null int64
NUMBER OF PEDESTRIANS KILLED    1226374 non-null int64
NUMBER OF CYCLIST INJURED       1226374 non-null int64
NUMBER OF CYCLIST KILLED        1226374 non-null int64
NUMBER OF MOTORIST INJURED      1226374 non-null int64
NUMBER OF MOTORIST KILLED       1226374 non-null int64
CONTRIBUTING FACTOR VEHICLE 1   1220021 non-null object
CONTRIBUTING FACTOR VEHICLE 2   1055125 non-null object
CONTRIBUTING FACTOR VEHICLE 3   79276 non-null object
CONTRIBUTING FACTOR VEHICLE 4   17054 non-null object
CONTRIBUTING FACTOR VEHICLE 5   4219 non-null object
UNIQUE KEY                      1226374 non-null int64
VEHICLE TYPE CODE 1             1216638 non-null object
VEHICLE TYPE CODE 2             1024534 non-null object
VEHICLE TYPE CODE 3             76860 non-null object
VEHICLE TYPE CODE 4             16436 non-null object
VEHICLE TYPE CODE 5             4094 non-null object
dtypes: float64(2), int64(9), object(18)
```

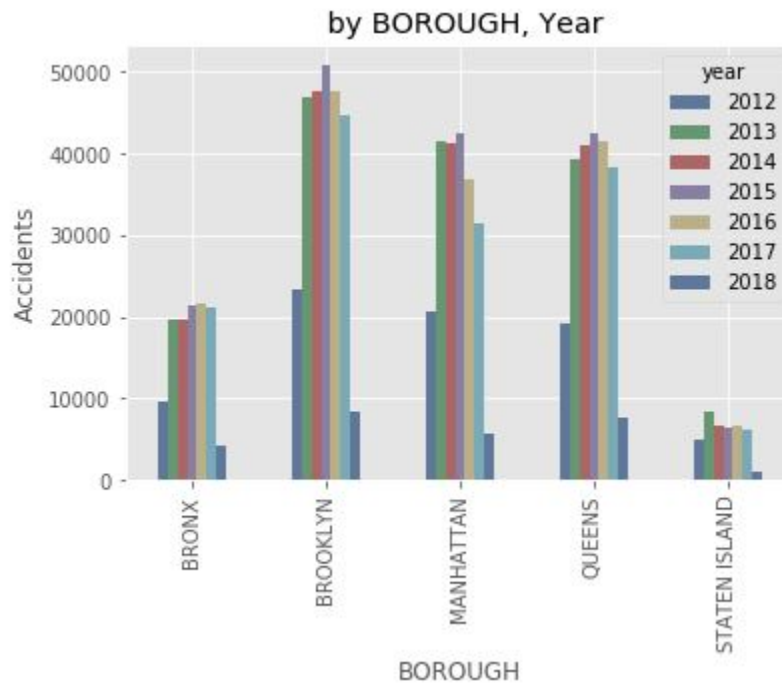Figure 1: Variables - NYC Motor Vehicle Collision

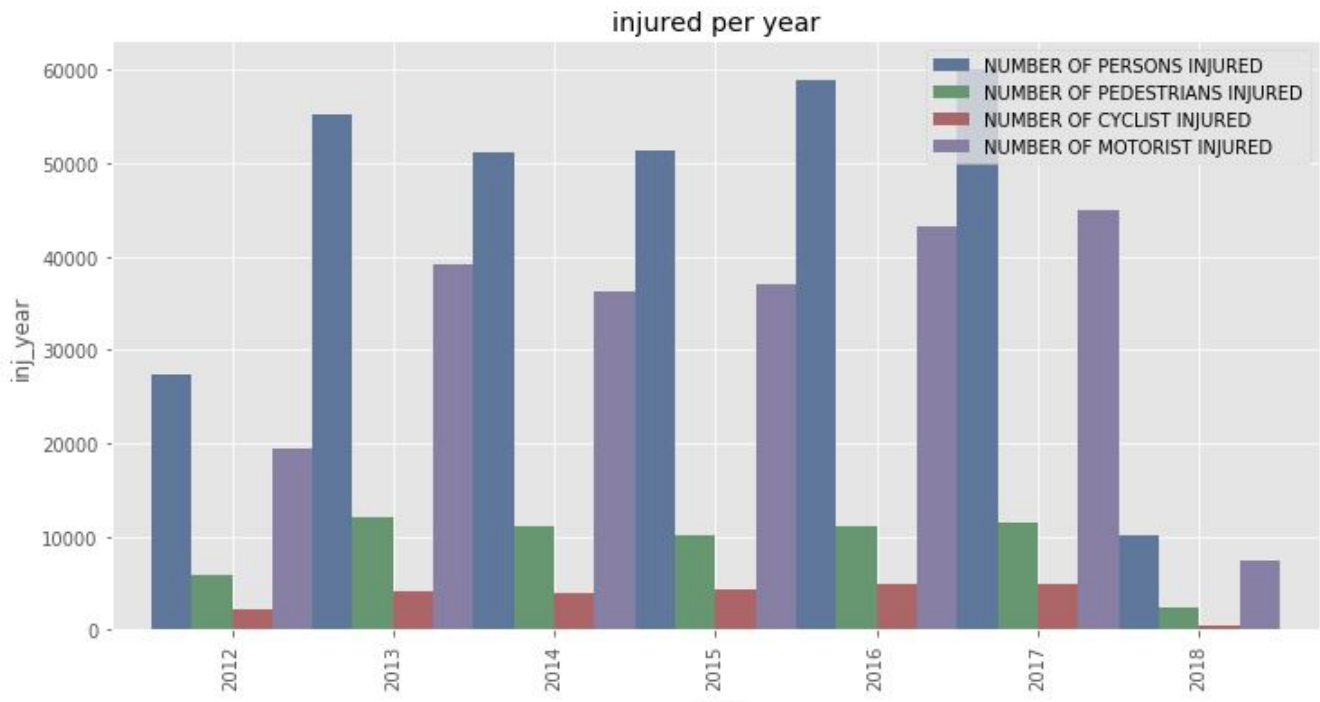Figure 2: Accidents by borough - NYC Motor Vehicle Collision

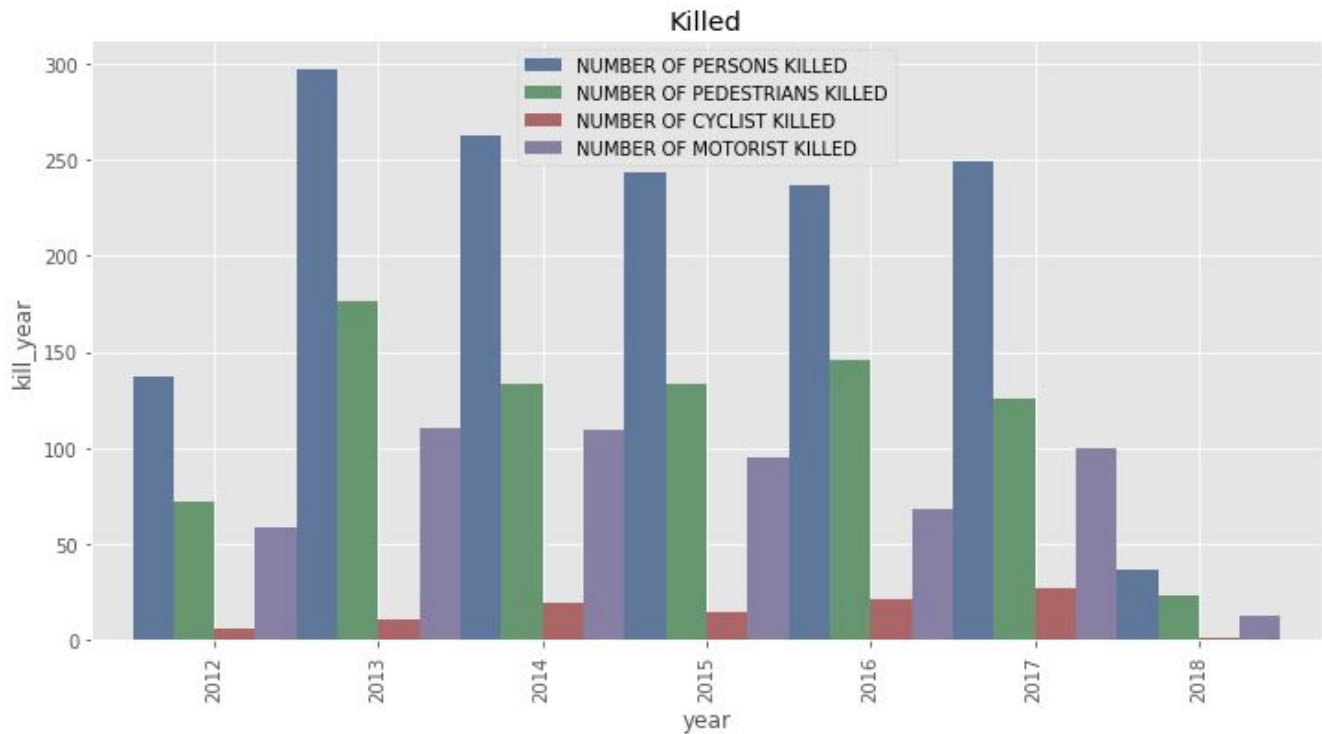Figure 3: Injuries by Year - NYC Motor Vehicle Collision

Figure 4: Fatalities by Year - NYC Motor Vehicle Collision

CONTRIBUTING FACTOR VEHICLE 1 year	Aggressive Driving	Animals Action	DUI	Defective Road	Defective signals	Disability	Driver Distraction	Driver Inexperience	Improper Driving	Mechanical defect	Other Vehicular	Pedestrian/Bicyclist/Other Pedestrian Error/Confusion	Traffic Control Device Improper	Unspecified	violaton of law	weather
2012	362	40	791	835	4	1177	21837	1252	4835	535	3231	12	409	59606	5087	524
2013	553	83	1610	1959	17	2253	45503	2384	10317	1029	6486	25	471	119122	10565	1093
2014	610	105	1911	2716	13	2324	51239	2599	12278	1023	8623	29	573	108551	11780	1182
2015	796	103	1965	3112	19	2627	64105	2752	14628	1167	12494	32	610	97234	13858	1356
2016	561	154	1608	1766	16	931	41985	2256	18417	865	6876	562	496	130293	17403	1393
2017	976	203	2643	2654	18	59	57140	4252	44250	1509	11623	1526	684	57673	40189	2268
2018	150	29	455	949	6	11	9361	537	7382	274	2029	191	158	11559	6787	473

Figure 5: Contributing Factors - NYC Motor Vehicle Collision

By grouping Vehicle contributing factors into different categories we were able to find out the major contributing factors for accidents. Human factors such as Driver Distraction,Improper Driving, violation of laws are without doubt the major reason for accidents. Road conditions and vehicle deffects play only secondary roles.

NYC Traffic Violations:

Here, we would like to briefly explore the traffic tickets dataset. NYC Violation data consists of the variables: 'Violation Charged Code', 'Violation Description', 'Violation Year', 'Violation Month', 'Violation Day of Week', 'Age at Violation','Gender', 'State of License', 'Police Agency', 'Court', 'Source'. There are approximately half million violation records during the years 2012-16.

By looking at what age group may be causing more violations, its evident that over 50% of the violations happened in the age group below 35 years.

Grouping violations by Courts in NYC from traffic tickets data, we see that there is a slight reduction of violations post vision zero initiative across all boroughs. As you can see here, Manhattan has the most traffic violations.
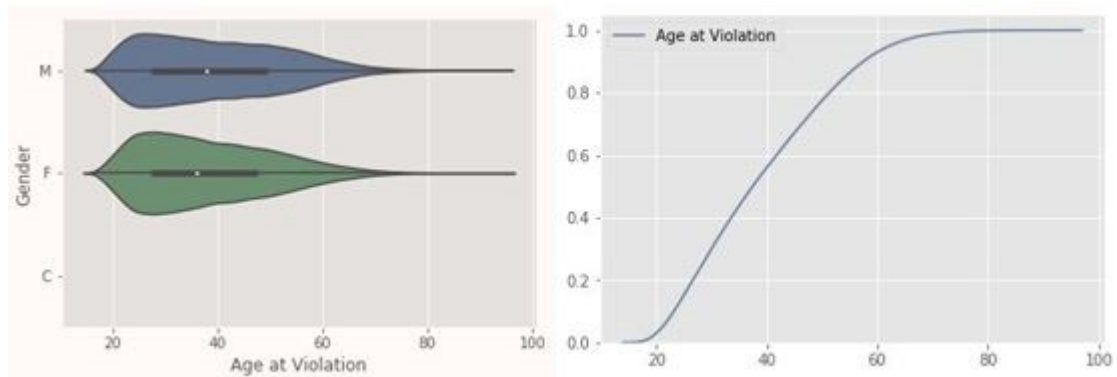
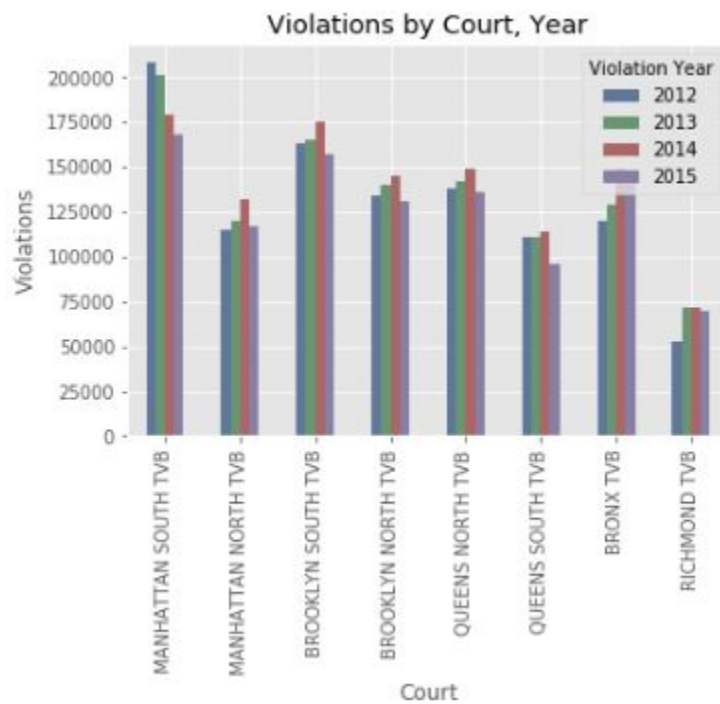Figure 6: violation age - NYC Traffic violations

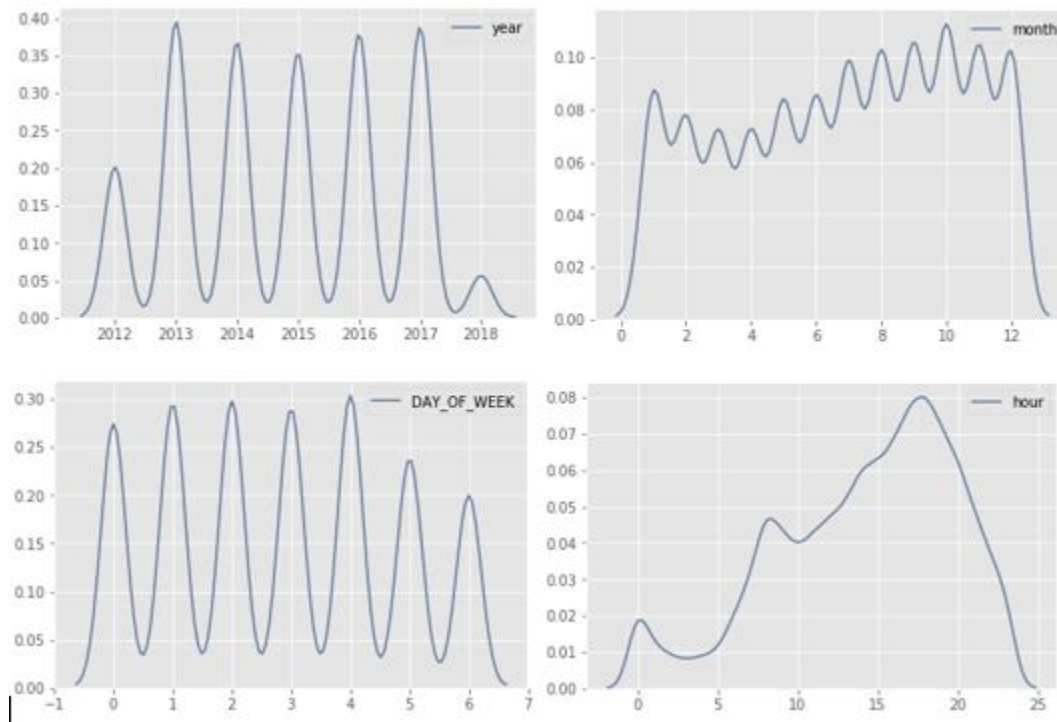Figure 7: violations by borough court - NYC Traffic violations

Figure 8: Pedestrian Incidents

6.1.1 Vision Zero Analysis

As part of our *vision zero data feeds* analysis, we will have a closer look at some specific areas and investigate if there is a change after a specific vision zero initiative.

6.1.1.1 Pedestrian Incident Analysis:

We have extracted the pedestrians and cyclists injuries and fatalities from the collision data and treated that as our pedestrian incident dataset, during the years 2012-2018.

Looking at the pedestrian incidents plot, its not showing a decrease in the incidents over the years. Also, the majority of the incidents are happening towards the later part of the day(between 3 to 4 PM), and slightly more chance on Fridays. When we looked at overall incidents acoross all months over these years, the October month has majority of the incidents.

We will try gathering the specific vision zero initative (for example, *arterial slow zones*) and analyse if there is any pattern before and after the initiative.
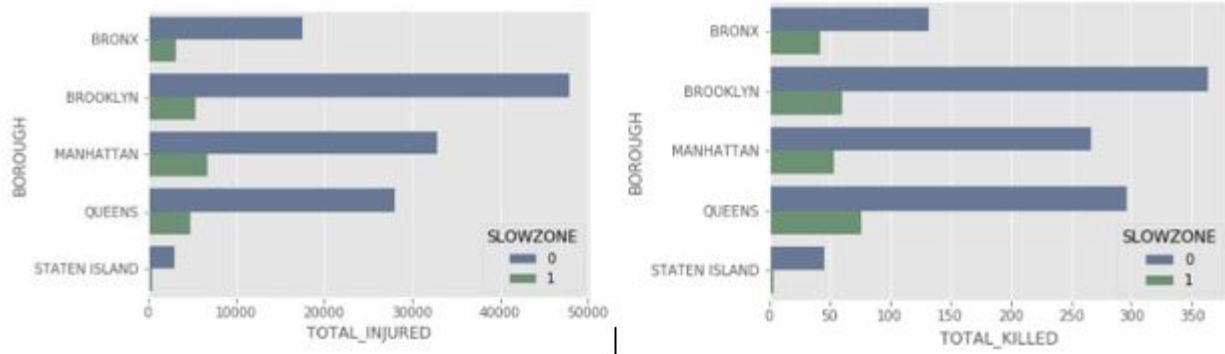
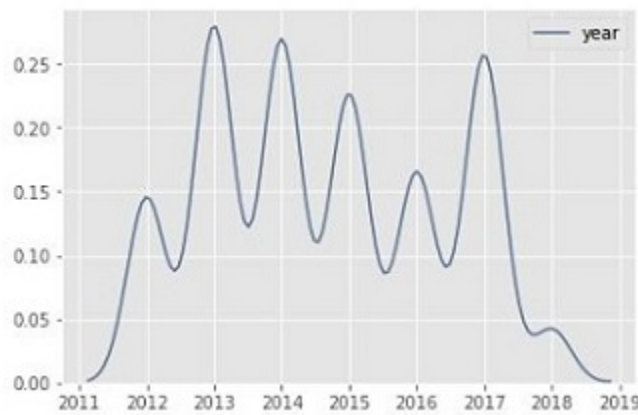Figure 9: Slowzone Vs Non Slowzone - Pedestrian Incidents

Figure 10: Slowzone - Yearly Pedestrian Incidents

Arterial Slow Zone program:

The *Arterial Slow Zone program* uses a combination of a lower speed limit, signal timing changes, distinctive signs and increased enforcement to improve safety on some of New York City's most high-crash corridors. The slow zone geo json data is available here We load the slow zone geo json data, and add a feature *SLOWZONE* to our pedestrain dataset by comparing if the incident point from the pedestrian dataset falls near the arterial slow zone or not. If the incident happend near the *SLOWZONE*, then we will mark it as 1, else 0.

There is clear reduction in the incidents in slow zones when compared with non-slow zone areas.

Taking only slowzone areas and plotting incidents year over year produced mixed results, there was a significant reduction immediately following the vision zero initiative, however, the incidents shooted up again in the year 2017!

Speed Humps:

Speed Humps are a raised area of a roadway designed to reduce vehicle speeds. Lets load the speed hump geo json (shape) file and try to mark each of the pedestrian incident with *SPEED_HUMP* as 1 (which represents the presence of speed hump) and 0 (absence of speed hump) by comparing the distance between the incident point from the speed hump location.
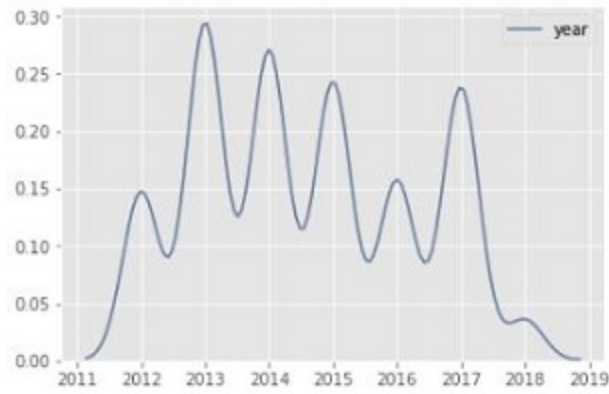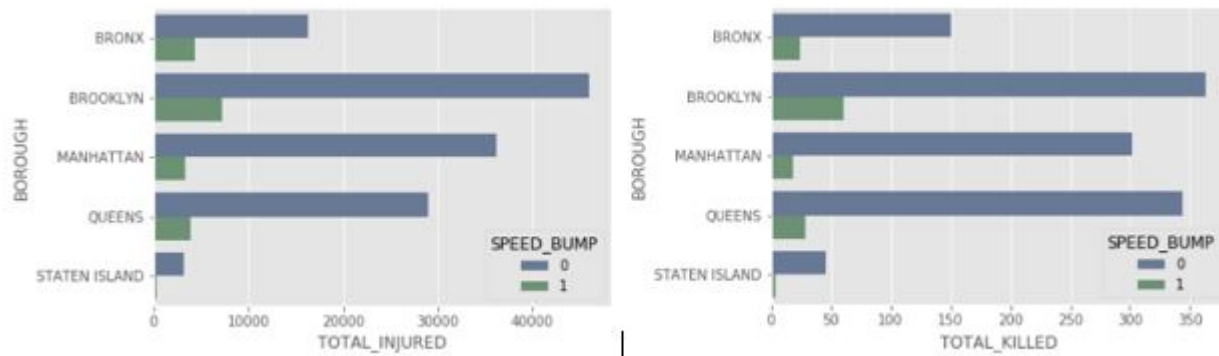
Figure 11: SPEED_HUMP - Yearly Pedestrian Incidents

Taking only SPEED HUMP locations and plotting pedestrian incidents year over year produced mixed results (very similar to what we have seen with *slow zone* analysis), there was a significant reduction immediately following the vision zero initiative, follwed by increased incidents in the year 2017!

6.2 Model Development

6.2.1 Model 1

6.2.2 Model 2

6.3 Model Validation

7 Conclusion

8 References

Chang, L. and W. Chen. "Data mining of tree-based models to analyze freeway accident frequency". In: Science Direct (2005). URL: https://www.sciencedirect.com/science/article/pii/S0022437505000708.

Eisenberg, D. The mixed effects of precipitation on traffic crashes. 2004. URL: https://www.ncbi.nlm.nih.gov/pubmed/15094418.

Jackson, T. and H. Sharif. "Rainfall impacts on traffic safety: rain-related fatal crashes in Texas". In: TandF (2005). URL: https://www.tandfonline.com/doi/full/10.1080/19475705.2014.984246?src=recsys.

Yuan, Z., X. Zhou and T. Yang. Predicting Traffic Accidents Through Heterogeneous Urban Data. 2017. URL: http://urbcomp.ist.psu.edu/2017/papers/Predicting.pdf.