Motor Vehicle Crashes Analysis using Big Data

Abijith Pradeep
ap8246@nyu.edu
New York University
USA

Damaraju Venkatesh
vd2348@nyu.edu
New York University
USA

Jaya Sabarish Reddy Remala
jr6421@nyu.edu
New York University
USA

KEYWORDS

Motor Vehicle, NYC, Crashes

ACM Reference Format:
Abijith Pradeep, Damaraju Venkatesh, and Jaya Sabarish Reddy Remala. 2023. Motor Vehicle Crashes Analysis using Big Data. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Traffic accidents are serious issues, that can possibly cause disabilities, injuries, and even fatalities. In order to decrease the number of accidents, we need to understand and analyze the traffic accidents dataset. Almost every day, someone has suffered from traffic accidents in one way or another, such as the traffic being slowed down due to an accident or collision on the same road, which may leave one or more lanes unavailable. Big Data ecosystem has the ability to store, manipulate, analyze, and mine large traffic accident datasets and can drive knowledge creation that can help decision-makers reduce the number of accidents. Upon analysis of this dataset using cleaning and statistical methods one can derive useful insights and conclusions to prevent future accidents.

2 PROBLEM FORMULATION

Given the dataset of Motor Vehicle Collisions, provide statistical answers by searching within the data to provide useful insights for motor vehicle accident prevention.

3 RELATED WORK

A study from [1] about motor vehicle collisions provided us with useful insights about pre-processing the motor vehicle collision data. The study grouped the vehicle crashes into four groups based on what type of vehicle crashed or what obstacle the vehicle crashed into. Another study by [2] developed a visualizer with which we could also visualize our datasets and the joined data. Further, a group of researchers explored database improvements for vehicle/bicycle crash analysis in [3] which also provided us with useful methods to preprocess data and join datasets.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

4 DATASET DESCRIPTION

The dataset used for analysis is obtained from the NYC Open Data program. It contains over 2 million rows with 29 columns. The dataset provided us with multiple columns about the geospatial and temporal information of the accident, like crash_date and crash_time, latitude and longitude, borough, and street_name, etc. It also has collision_id assigned which can be used to join with multiple other datasets which we have obtained from NYC Open data.

We used other datasets namely Motor Vehicle Collisions – Vehicles, and Motor Vehicle Collisions – Persons which give even more detailed information about the accident the make and model of the vehicle, its registration status, the license status of the driver, the primary and other reasons of the accident. We used these datasets to join with the main dataset we obtained to understand the causality behind the accidents involving multiple groups such as pedestrians, motorists, and cyclists.

Table 1: Collision Attributes

Attributes	Description
Collision ID	Unique identifier for each crash event
Crash Date	Date (MM/DD/YYYY) collision occurs
Crash Time	Relevant time of the day collision occurs
Borough	Location where the collision happened
Zip code	Exact zip code recorded by the police
Latitude	Geographical location
Longitude	Geographical location
Number of persons injured	Number of people injured in the collision
Number of persons killed	Number of people killed in the collision
Number of pedestrians injured	Number of pedestrians injured in the collision
Number of pedestrians killed	Number of pedestrians killed in the collision
Number of cyclists injured	Number of cyclists injured in the collision
Number of cyclists killed	Number of cyclists killed in the collision
Number of motorists injured	Number of motorists injured in the collision
Number of motorists killed	Number of motorists killed in the collision
Vehicle_Group	Group of vehicles which caused the collision

5 METHODS AND ARCHITECTURE

5.1 Datasets Joining

We have two datasets pertaining to the crash details and vehicle details where the columns merge using "collision_id" as a primary key. In our preliminary join, we used outer join to merge both table records. The first dataset contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police-reported motor vehicle collisions in NYC. Meanwhile, the second dataset comprises details on each vehicle involved in the crash. Each row represents a motor vehicle involved in a crash. The data in this table goes back to April 2016 when crash reporting switched to an electronic system. The reason for joining the above datasets is to identify the gender of the drivers.

Another reason was to determine the type of vehicles (small, light, medium, and heavy motor vehicles) involved in the crash. By determining the vehicle group mentioned before we could identify how many different types of vehicles were involved in a particular collision. In other words, how many of these vehicles shared a common collision_id.

5.2 Data Pre-processing and Cleaning

Null Values: One of the important tasks involved is finding out the null values and categories involved in different attributes of the dataset. For example, the vehicle make has multiple vehicle models and we classified them into different sub-categories. Since there are many columns and many categories of analyses that can be obtained, a row cannot be completely ignored based on one null value in a single column instead we will ignore that row when it is considered in that specific analysis. For example, when analyzing the total number of accidents that happened at 2:00 AM, a column with a null value in pedestrian_action can be still considered, but when analyzing the contributing factor count for the accident, the row cannot be considered if the value is null in that column.

Borough: Geospatial analysis plays a pivotal role in understanding and interpreting complex datasets, especially when dealing with location-specific information. In the context of a project focused on studying NYC crash data, the absence of borough values for a substantial number of rows poses a significant challenge. To address this issue, a geospatial approach has been implemented, leveraging latitude and longitude coordinates to infer missing borough values. We found out .SHP and .SHX files in Open NYC Data that classify the given latitude and longitude information to a borough. To generate these values, we created a UDF in spark that checks all the rows, if it finds a null value for borough and latitude, and longitude information present, it generates the borough information and populates the data there.

Vehicle Group: Understanding the types of vehicles involved is crucial for identifying patterns, assessing risks, and devising targeted safety measures. However, the dataset presents a challenge with variations in vehicle descriptions between two columns vehicle_make and vehicle_type. Together these columns form almost 80 different values which are too diverse to be used. This diversity in vehicle descriptions is addressed by categorizing them into standardized groups.

Table 2: Collision Data

COLLISION_ID	VEHICLE_TYPE	VEHICLE_MAKE
22	VAN	NULL
22	SPORT UTILITY / SUV	NULL
23	TAXI	NULL
23	TAXI	NULL
24	BUS	NULL
24	PASSENGER VEHICLE	NULL
25	PASSENGER VEHICLE	NULL
25	PASSENGER VEHICLE	NULL
26	PASSENGER VEHICLE	NULL
26	PASSENGER VEHICLE	NULL
27	TAXI	NULL
27	PASSENGER VEHICLE	NULL
28	TAXI	NULL
28	TAXI	NULL
29	TAXI	NULL
29	PASSENGER VEHICLE	NULL
30	TAXI	NULL
30	VAN	NULL
31	UNKNOWN	NULL
31	PASSENGER VEHICLE	NULL

Using the mentioned columns, the classification is divided into 5 groups. Heavy, Medium, Light, Small, Other, Unspecified. Grouping vehicles enables a more effective assessment of risk factors associated with each vehicle type. This information is valuable for prioritizing interventions and implementing targeted safety measures. A spark UDF is developed for this purpose which reads the data from both these columns and returns to which classification the vehicle belongs. The classification is done based on the generic properties of the vehicle such as the number of wheels, size of the vehicle, make of the vehicle can also suggest the type of vehicle (ex: 'sedan' value indicates that the vehicle is the car). The table below describes it:

Table 3: Vehicle Group Accidents

VEHICLE_GROUP	ACCIDENTS
Heavy	317,624
Light	1,851,563
Medium	1,112,349
Other	153,846
Small	83,009

Our action item from the above grouping was to determine which type of vehicle crashed into another vehicle, an obstacle, or a pedestrian/cyclist. As mentioned above we utilized another Pyspark SQL query to group the vehicles by their collision ID and counted or added 1 every time that collision_id encountered a type of vehicle. The output of the mentioned procedure is shown in Table 2. The column ACC_IN_BETWEEN describes which type of vehicle hit each other. If we see the first row, the values are 0 except

small_count	light_count	medium_count	heavy_count	other_count	COLLISION_ID	ACC_IN_BETWEEN	ACC_IN_BETWEEN_UNQ
0	0	0	1	0	148	H	H
1	1	0	0	0	463	SL	SL
0	1	1	0	0	471	LM	LM
0	1	0	1	0	496	LH	LH

Table 4: Vehicle Collision Table

for heavy_count with a value of 1. This implies that the vehicle didn't collide with another vehicle but rather a pedestrian or an obstacle.

Table 5: Small section of Unique vehicle collisions

ACC_IN_BETWEEN_UNQ	GROUPS
MO	33,492
SL	39,199
H	40,299
LO	78,748
MH	85,437
LH	152,135
M	280,592
LM	449,335
L	599,637

CONTRIBUTING FACTOR: The contributing Factor is mentioned as the primary reason why the accident could have occurred in the dataset. It is found that this contains almost 65 diverse reasons. To facilitate meaningful analysis, a classification approach has been implemented to group these factors into broader categories, providing a clearer understanding of the primary causes behind accidents. We classified them into 14 different categories as shown below. This allows for a more focused examination of the primary causes of accidents. The classification also aids in standardizing and simplifying the interpretation of contributing factors, contributing to more effective accident prevention measures.

5.3 Data Analysis

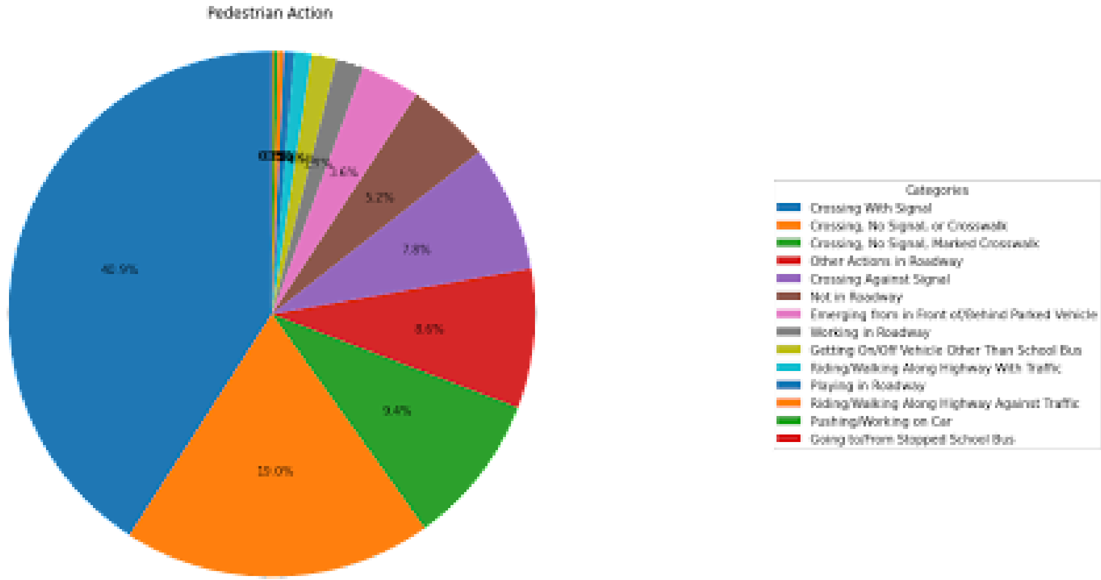
Pedestrian Data The count of accidents present here includes only the injury or death-causing accidents.

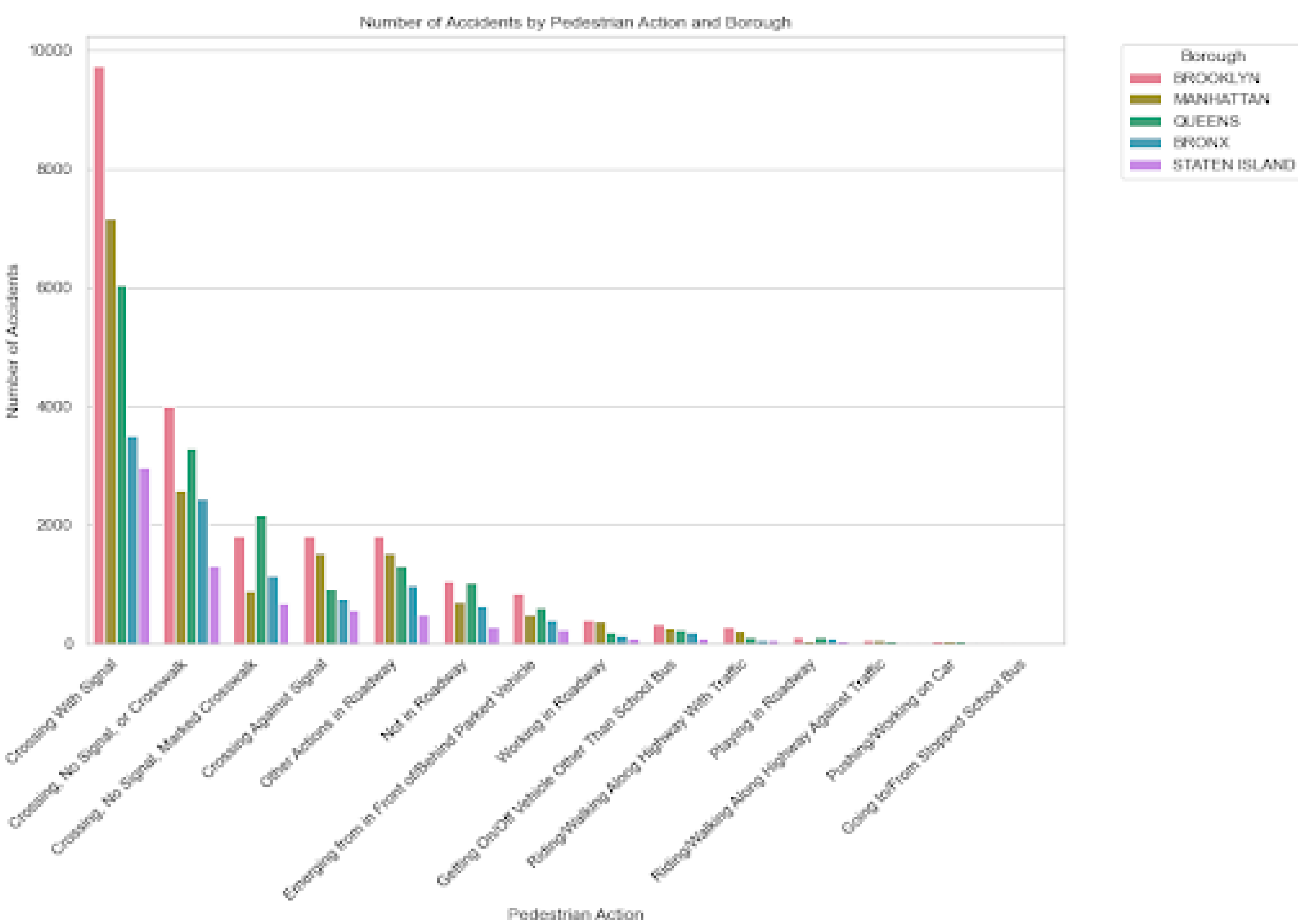
One of the main reasons why pedestrian accidents happen is "Crossing With Signal". Given that the pedestrians are crossing with a signal we can say that the error is from the side of a vehicle driver. On searching for the reasons, we found out the following are the main reasons:

Table 6: Contributing factors

Classification Type	Contributing Factors
Mechanical Failure	Brakes Defective, Steering Failure, Tire Failure/Inadequate, Accelerator Defective, Tow Hitch Defective, Tinted Windows, Windshield Inadequate
Traffic System Failure	Pavement Slippery, Obstruction/Debris, Pavement Defective, Lane marking Improper/Inadequate
Substance Abuse	Alcohol Involvement, Prescription Medication, 'Drugs(illegal)'
Technological Intervention	Cell Phone(hand-held), Cell Phone(hands-free), Using on Board Navigation Devices, Texting, Listening/Using Headphones, Other Electronic Devices
Driver Inattention	Driver Inattention/Distraction
Right-way Rule	Failure to Yield Right-of-Way
Failure to Keep Right	Failure to Keep Right
Following Too Closely	Following Too Closely, Passing Too Closely
Driver Incompetence	Backing Unsafely, Passing or Lane Usage Improper, Turning Improperly, Unsafe Lane Changing, Driver Inexperience, Aggressive Driving/Road Rage, Eating or Drinking
Traffic Rule Avoidance	Traffic Control Disregarded, Unsafe Speed
Other Vehicle Intervention	Other Vehicle Intervention
Biological Effects	Fatigued/Drowsy, Lost Consciousness, Fell Asleep, Illness, Shoulders Defective/Improper, Physical Disability
Lighting Reasons	Glare, Other Lighting Defects, Headlights Defective
Distraction	Passenger Distraction, Outside Car Distraction
Driverless/Runaway Vehicle	Driverless/Runaway Vehicle
Oversized Vehicle	Oversized Vehicle
Lane Marking Improper/Inadequate	Lane Marking Improper/Inadequate
View Obstructed/Limited	View Obstructed/Limited
Pedestrian/Bicyclist/Other	Pedestrian Error/Confusion, Other Pedestrian Error/Confusion
Traffic Control Device Improper/Non-Working	Traffic Control Device Improper/Non-Working
Animals Action	Animals Action
Vehicle Vandalism	Vehicle Vandalism
Reaction to Other Uninvolved Vehicle	Reaction to Other Uninvolved Vehicle

- Most accidents happen at the intersection points of roads.


Pedestrian Action
40.9%
19.0%
9.4%
8.6%
7.8%
5.2%
3.6%
Categories
Crossing With Signal
Crossing, No Signal, or Crosswalk
Crossing, No Signal, Marked Crosswalk
Other Actions in Roadway
Crossing Against Signal
Not in Roadway
Emerging from in Front of/Behind Parked Vehicle
Working in Roadway
Getting On/Off Vehicle Other Than School Bus
Riding/Walking Along Highway With Traffic
Playing in Roadway
Riding/Walking Along Highway Against Traffic
Pushing/Working on Car
Going to/From Stopped School Bus

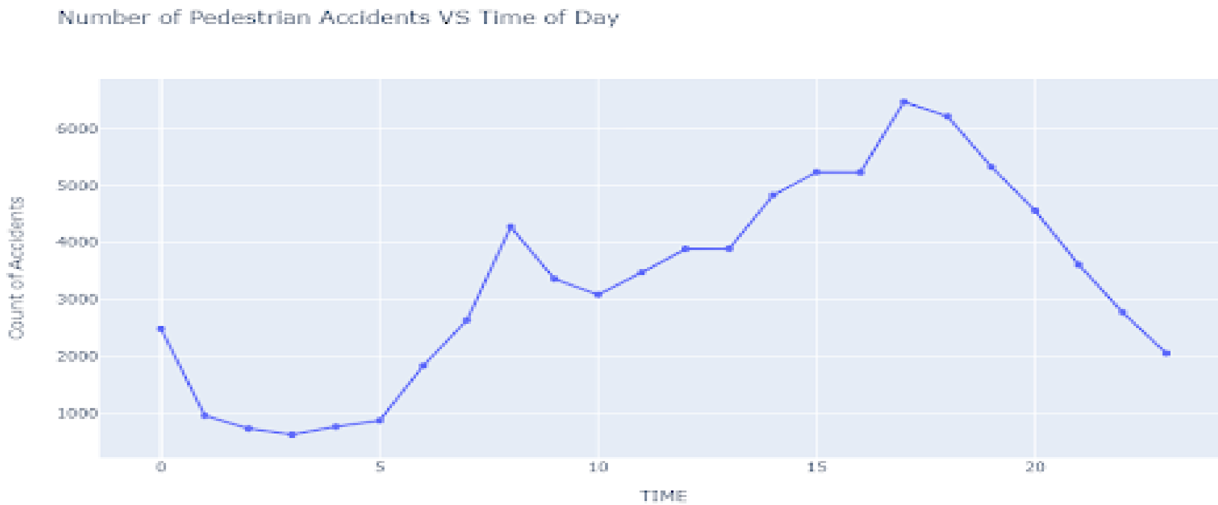

Number of Accidents by Pedestrian Action and Borough
Number of Accidents
10000
8000
6000
4000
2000
0
Borough
BROOKLYN
MANHATTAN
QUEENS
BRONX
STATEN ISLAND
Crossing With Signal
Crossing, No Signal, or Crosswalk
Crossing, No Signal, Marked Crosswalk
Crossing Against Signal
Other Actions in Roadway
Not in Roadway
Emerging from in Front of/Behind Parked Vehicle
Working in Roadway
Getting On/Off Vehicle Other Than School Bus
Riding/Walking Along Highway With Traffic
Playing in Roadway
Riding/Walking Along Highway Against Traffic
Pushing/Working on Car
Going to/From Stopped School Bus
Pedestrian Action

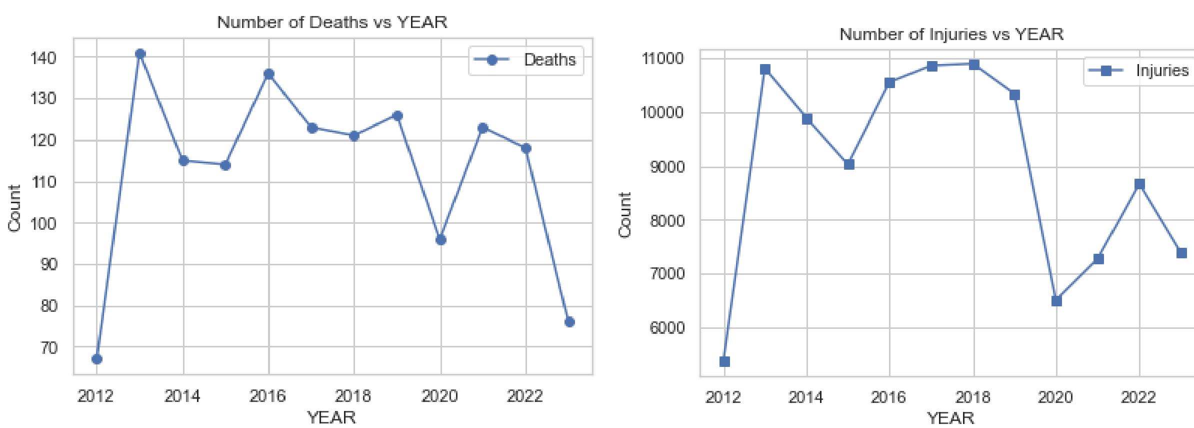
- The next contributing reason is the inattention/distraction of the driver, this is further expanded in the next sections which we have explained with numbers this reason.
- Failure to yield the right of way is another significant reason causing the accidents.
- There is confusion in the concept of the right of way, in a few states the pedestrians have the absolute authority in the right of way while in a few states, it is not. The right of way in NYC states that:
  - If stopped at a red light and the light turns green, drivers must wait if a pedestrian is still crossing.
  - Drivers must yield to both vehicles and pedestrians when entering the road from a driveway, alley, private road, or any other non-roadway.
  - A yield sign gives the right of way to pedestrians in a crosswalk. If a driver goes past a yield sign without stopping and hits a pedestrian (or another vehicle in the intersection), the accident will be deemed prima facie (automatic) evidence of failure to yield the right of way.
- The fifth rule is introduced as a part of the VISION ZERO project, where the drivers had severe consequences if they hit a pedestrian without yielding right of the way. Since these rules are not followed, we can observe that the number of accidents is increasing year over year.

Queens Anomaly More number of accidents are happening in the borough of Brooklyn and this trend is also observed in all other accident patterns. This is because the area and population of Brooklyn are far higher than all other boroughs. But one interesting point to note is that when we calculate the population-to-accident ratio it is highest in Queens Borough. Queens Boulevard noted the highest number of accidents, it has some of the most dangerous intersections in New York City causing these accidents not only for pedestrians but also for vehicles. Also, there are a few other dangerous intersections in Queens which are recorded as hotspots for accidents. Tricky intersections along with driver negligence and incompetence (several reasons are combined here) are resulting in a greater number of accidents. Also, we can see that the percentage increase in population in Queens is less when compared to that of other boroughs. Even then the increase in accident percentage suggests that there is a need for special care in the case of Queens.

We can see that the greatest number of pedestrian accidents are happening at the busiest hours in the city. These timings are the busiest because: Employee and Student Commute: The usual peak hours in NYC are from 4 PM to 7 PM when people are getting off work or leaving school. As a result, the density of vehicles on roads increases thereby increasing the chances of accidents Visibility: As the day progresses decreasing natural light could contribute to reduced visibility for both drivers and pedestrians. Diminished visibility might increase the risk of accidents. This is more relevant in winter as the sun sets before 5 PM. Morning Commute: The peak around 9 AM corresponds to the morning rush hour when people are commuting to work or school. Like the afternoon peak, increased activity during this time can contribute to a higher number of accidents. These factors when combined with the above-mentioned contributing factors such as distracted driving, non-compliance with traffic rules, etc, can lead to more number of accidents which is something we notice in the graph where the highest number of accidents are happening from 4 PM to 7 PM.

When we compare the increase in population with the increase in fatalities and injuries, we can say that the situation is improving but there needs to be additional care taken in a few areas, there should be strict monitoring and penalties imposed in case of driver negligence as this is the most contributing reason for accidents happening. But we can see a sudden increase in some injuries and deaths rising from 2016. [4] This is because in 2016 NYC Legislation declared that pedestrians have the right of way even when the flashing upraised hand signal or "countdown" is displayed. Eliminating confusion over the right of way in this phase emphasizes that traffic must prioritize keeping pedestrians safe. Now this caused confusion among pedestrians and motorists on how to cross increasing accidents.(Cited from [5]) According to the study, the speed of the vehicle plays an important role in the fatality of the accident. It states that the likelihood of pedestrian injury severity at high-speed limit zones is about 3.1 times higher for roadways. Considering this the speed limit introduced as a part of Vision Zero provided some advantages to the pedestrians.

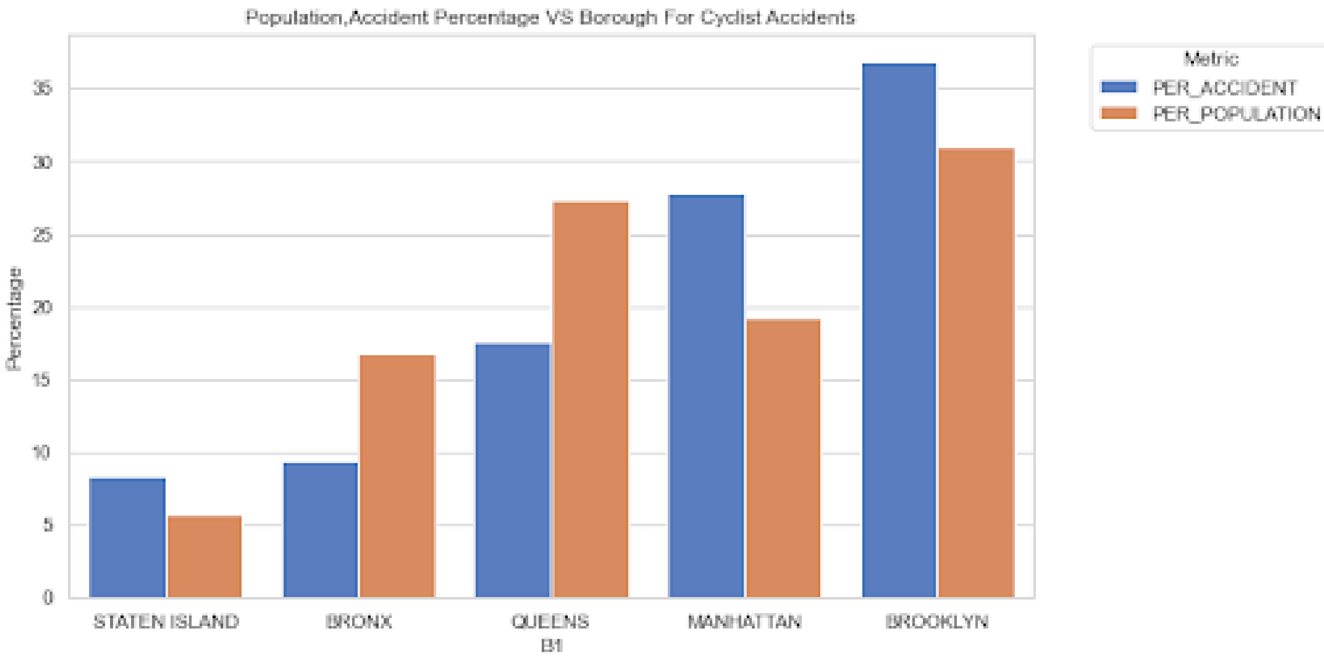


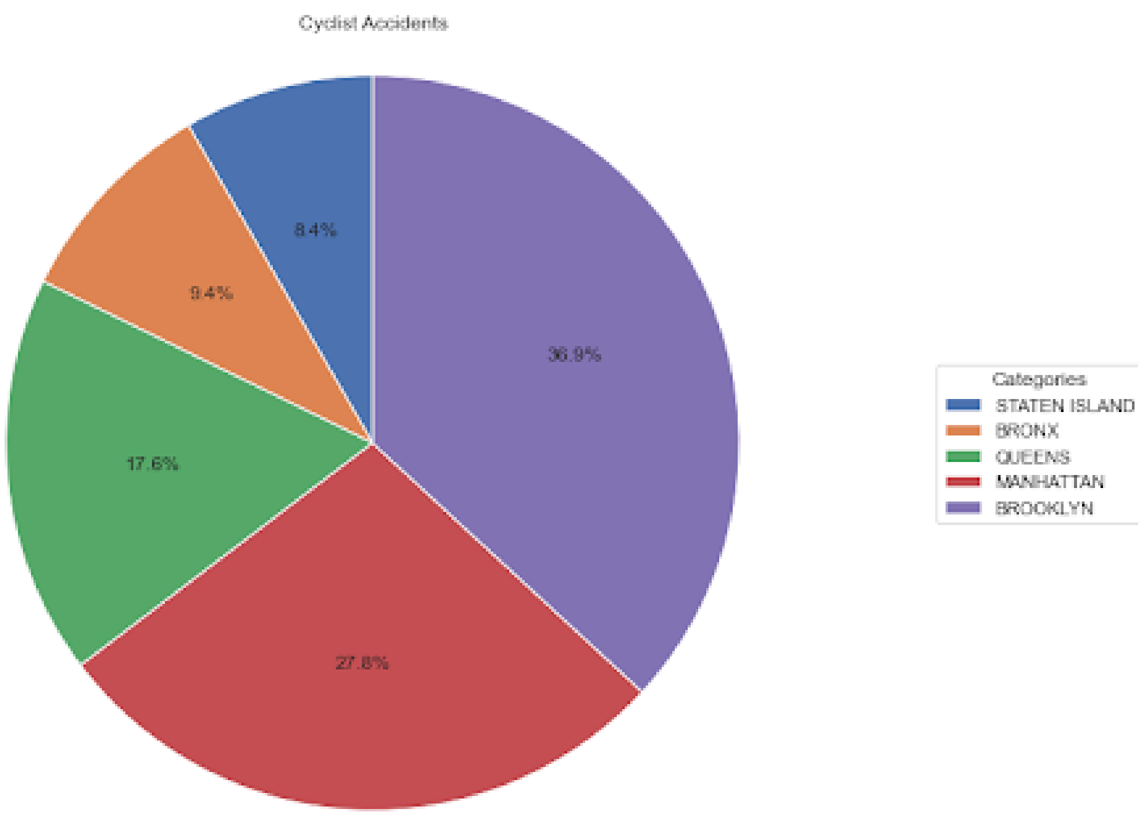


(a) Deaths vs Year **(b) Injuries vs Year**

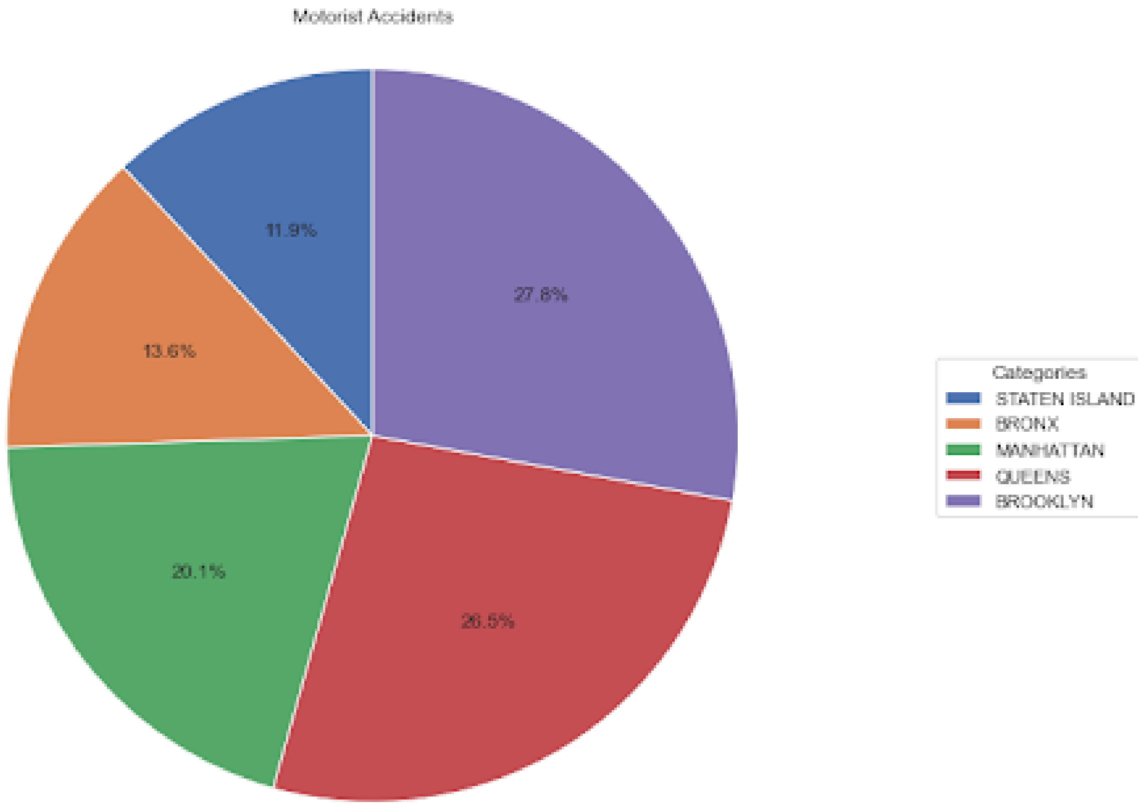
Figure 1: Yearly death and injuries of pedestrians

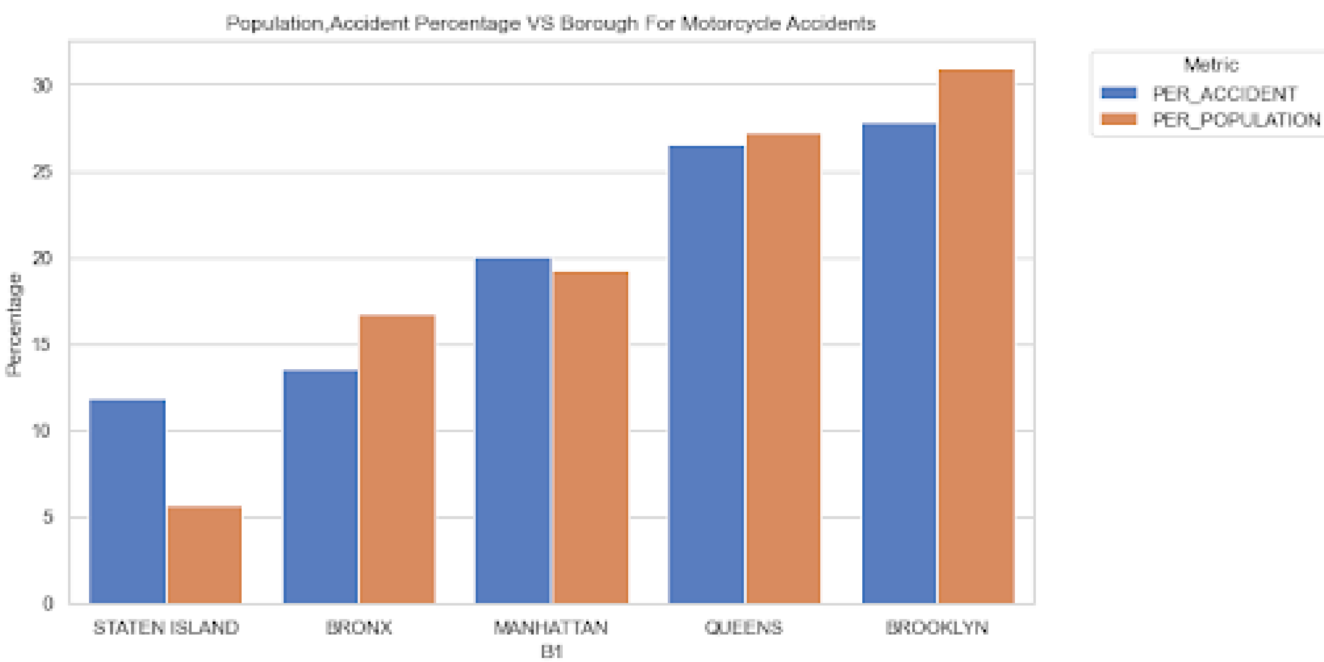
5.3.1 Cyclist Data. Although the trend of total accidents for motorists seems like that of cyclists one important point to note here is we have normalised the percentage obtained with the percentage of population. Then we can see that the highest disparity is found in the case of Staten Island and Manhattan. Manhattan is one of the most dangerous boroughs for motorists in NYC. The highest motor vehicle-dense area of NYC is Manhattan. Especially with the rise in Uber and other cab services a greater number of vehicles are traveling across the city from Manhattan. On analyzing the contributing factor, we found out that driver attention outweighed all other reasons. Along with this, we found out a few other factors which are: Busy Traffic Structure & Complex Street System: Manhattan is best known for its busy roads and congested streets, especially in the commercial and business districts. The increased traffic volume, combined with the presence of cyclists, can result in a higher number of accidents. Also, the street grid is known for its complexity, with numerous intersections, one-way streets, and various repairs and construction blockages at random in streets. Cyclists navigating through this may face challenges, and the complexity can contribute to the risk of accidents. Tourist Attraction: Manhattan is the center of tourist attraction in NYC because of places such as Times Square, Central Park, Bryant Park, and Wall Street. Tourists who are not familiar with local traffic rules, lead to potential conflicts with cyclists. Delivery and Service Vehicles: Manhattan has the highest volume of service and delivery vehicles. More number of delivery orders are made in Manhattan, especially during the afternoon and early hours of the night. Interactions between cyclists and these vehicles, especially in loading and unloading zones, can contribute to accidents. Intersection Challenges: Intersections are the highest-risk areas for accidents. Manhattan's numerous intersections, along with turning vehicles and pedestrians, can create challenging conditions for cyclists





5.3.2 Motorist Data. Inference from graph: Unlike other boroughs, we see the highest difference between the population percentage and accident percentage in Staten Island. Manhattan and Queens are almost similar in comparison with Manhattan having a slightly





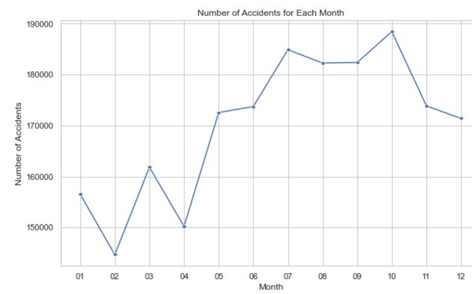
higher number of percent than the population it has. Although Brooklyn has the highest percentage of accidents we can see that the metric when normalized with population is having second least when it comes to these fatality or injury-causing accidents. preceded by the Bronx. Reasons why it could happen:[6] One of the major accident-causing reasons in Staten Island could be linked to its poor parking system. As mentioned in the article above the poor design of parking spaces coupled with large spots at malls and higher floating of heavy vehicles and SUVs resulted in a higher number of fatalities or injury-causing accidents in Staten Island. Source:[7] The reason why a greater number of accidents happening around Manhattan and less number in the Bronx could also be linked to the amount of traffic floating in these boroughs. As cited in the above article, Manhattan has the highest percentage of taxi trips in NYC, while the Bronx has the lowest with the difference being 35%. The congestion created in traffic because of these trips and since more vehicles are moving around confusing and dangerous intersections, Manhattan always has the highest probability for a vehicle collision than any other borough. Another reason we found out is the movement of heavy vehicles is highest in Queens followed by Manhattan and Brooklyn. This could also contribute majorly to the accident in Queens and Manhattan. Impact of pedestrian countdown signals:[8] A fascinating understanding from the paper is that the countdown clocks reduced pedestrian accidents when contrasted with crash rates at those equivalent convergences in earlier years. Inquisitively, be that as it may, vehicle crashes at those crossing points quadrupled. It is explained in the study that the driver's behavior to cross the signal at the T-bend before the signal changes (which can be inferred from the pedestrian signal as the countdown is about to end) is causing a greater number of accidents.

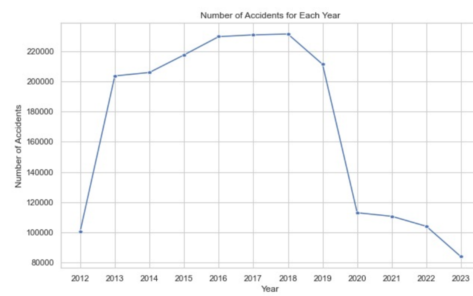
Table 7: Vehicle Groups and Types

Vehicle Group	Vehicle Type-Codes
Heavy Vehicle	Bus, truck, large commercial vehicle
Medium Vehicle	Pick-up truck, small commercial vehicle, van, livery vehicle
Light Vehicle	Sedan, SUV, Passenger vehicle, sport-utility/wagon, taxi
Small Vehicle	pedicab, Motorcycle, scooter, bicycle
Other Vehicle	Unknown type
Unspecified	Unspecified vehicle type or Null

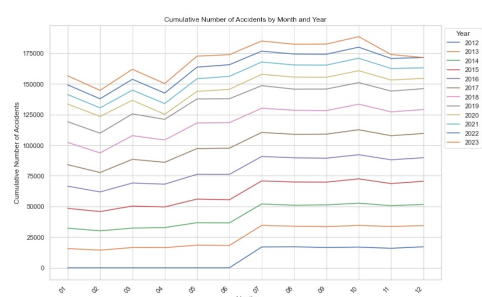
6 RESULTS

6.1 Generic Trend for vehicle crashes





Firstly, we analyzed the number of accidents happening in each month, we observed that accidents are the least in the months of winter, presumably the reason being snowfall and extreme weather causing people to travel less. While summer and fall recorded the highest number of accidents. This can be correlated to multiple reasons as these are the months when there will be tourism and the weather is pleasant enabling people to travel more. To confirm this trend we made a graph showing the monthly distribution for all the years available in the dataset. The trend remained almost the same for all years confirming this analysis.



Yearly analysis of the data showed a steady increase as time progressed until 2020, a steep decline can be noticed here, the reason being the COVID pandemic and subsequent lockdowns following it. This trend continued for the following years as mobility decreased due to various reasons such as a decline in tourism, Work From Home culture adopted by many tech companies, etc.

We tried to find the correlation between the number of persons injured/killed to the number of pedestrians and cyclists. We need to

	NUMBER OF PERSONS KILLED	NUMBER OF PEDESTRIANS KILLED	NUMBER OF CYCLIST KILLED
NUMBER OF PERSONS KILLED	1.000000	0.691424	0.279344
NUMBER OF PEDESTRIANS KILLED	0.691424	1.000000	0.021388
NUMBER OF CYCLIST KILLED	0.279344	0.021388	1.000000

	NUMBER OF PERSONS INJURED	NUMBER OF PEDESTRIANS INJURED	NUMBER OF CYCLIST INJURED
NUMBER OF PERSONS INJURED	1.000000	0.277095	0.173504
NUMBER OF PEDESTRIANS INJURED	0.277095	1.000000	-0.032314
NUMBER OF CYCLIST INJURED	0.173504	-0.032314	1.000000

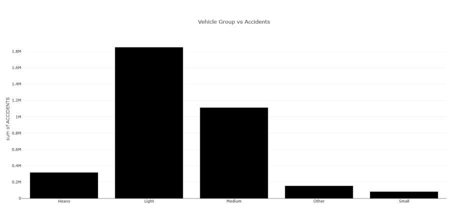
note that the number of persons killed includes other data such as the number of persons killed/injured traveling in cars, trucks, bikes, etc. It is a culmination of the total number of killed/injured people. Considerable correlation is observed in the case of the number of pedestrians killed to the number of persons killed. We must note that the data will also include cases not including any accidents involving pedestrians.

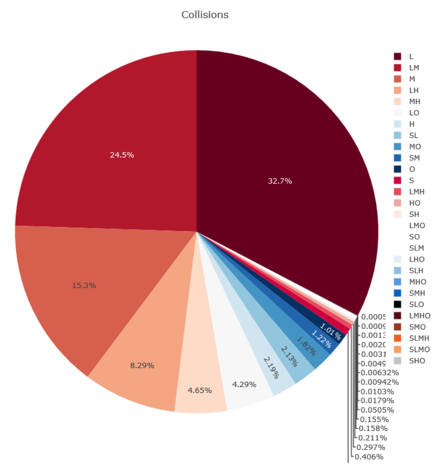
6.2 Inferences for Vehicle merged dataset

For the vehicle merged dataset we merged the original dataset and the provided vehicle dataset. The major inference was that the Light vehicles which included Sedans, SUVs, LUVs, Passenger cars, etc. had the most accidents occurred. Out of the 3.5 million different vehicles, a whopping 1.8 Million Light vehicles were involved in the crashes which are about 50% of the total vehicles. Followed by the Medium Vehicles and then the Heavy vehicles.

From Table 5, we found out the most collided vehicles will be the Light vehicles with each other or to other obstacles/persons. This explains the previous inference of having light-mode vehicles with the highest number of accidents. Of the total 1.8M collisions, almost 600k collisions involved Light-mode vehicles crashing with each other or Light-mode vehicles with another obstacle/pedestrian. That's close to 33% of the total occurred collisions.

Another inference found using the vehicle merged dataset was the grouping of drivers based on their license status. Although accidents existed for the Licensed and Permit holding drivers, there was also a certain section of Unlicensed and Unspecified drivers with some number of accidents, which implies they were illegal drivers or drivers with expired licenses. Having this subset of people drive is lawfully a crime. After analysis of the contributing_factor column for unlicensed, it was found that the drivers as expected were inattentive or careless causing a high number of accidents around 5800 accidents, which was 28% of the total unlicensed drivers.





6.3 Inferences for Contributing Factors Classification

Based on the classification of contributing factors from the table 6 that cause accidents in the NYC boroughs unveiled below from 2014 through 2023 (August). We have made almost 24 types of classifications based on the nature/reason for these accidents. We are more interested in comparing the top 7 contributing factor classifications here as they contributed almost 69% of the total accidents happening in NYC.

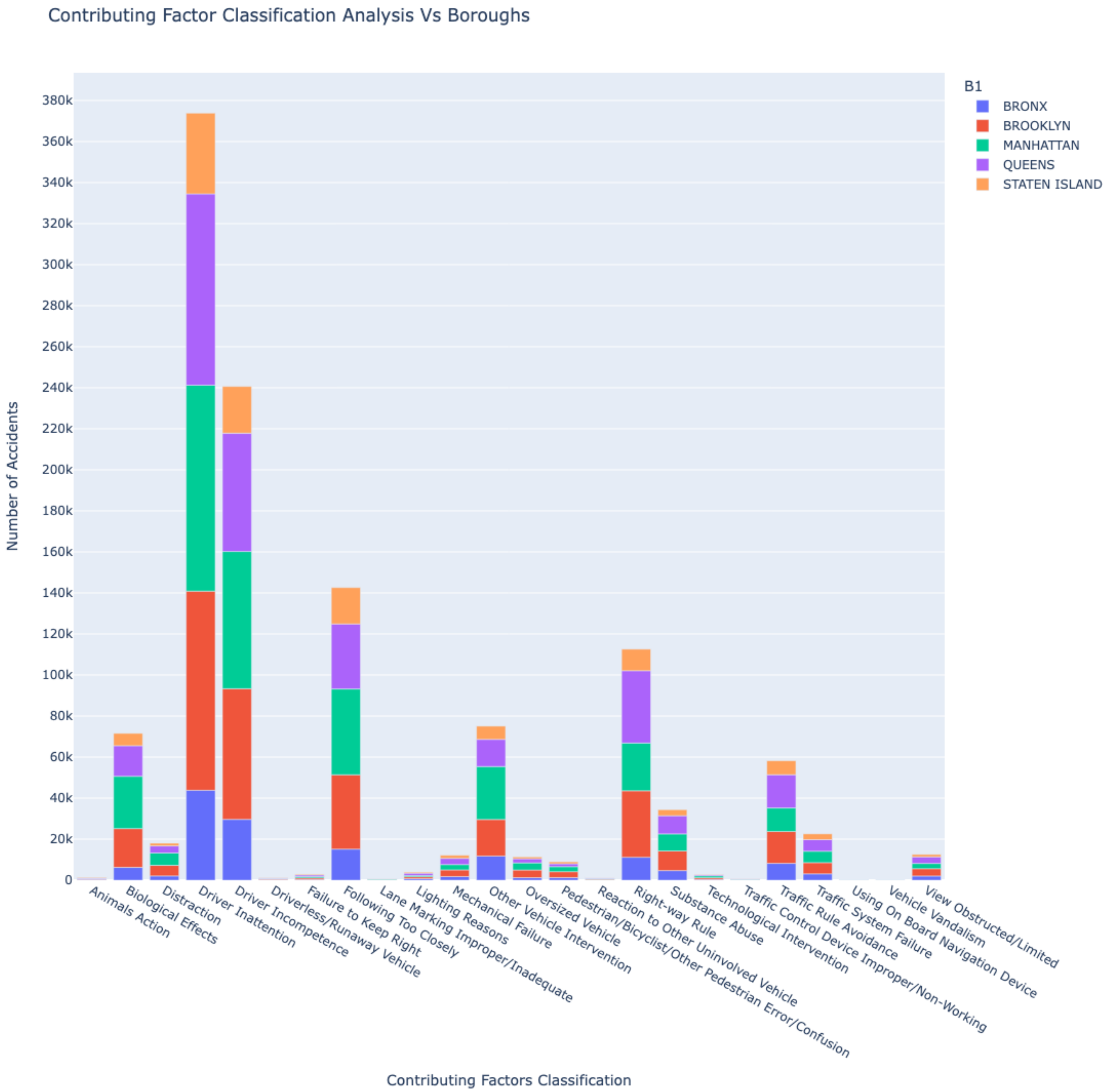
- Driver Inattention (M>B>Q>BX>S)
- Driver Incompetence (M>B>Q>BX>S)
- Following Too Closely (M>B>Q>S>BX)
- Right Way Rule (Q>B>M>BX>S)
- Other Vehicle Intervention (M>B>Q>BX>S)
- Biological Effects (M>B>Q>BX>S)
- Traffic Rule Avoidance (Q>B>M>BX>S)

M- Manhattan, B- Brooklyn, Q- Queens, BX-Bronx, S - Staten Island

Based on the above trend, Manhattan, the cause of accidents tops with the following Contributing Factors Driver Inattention, Driver Incompetence, Following closely, Other Vehicle Intervention, and Biological Effects. If we deep dive into the above causes, as our analysis states the facts are that due to an increase in the population density, being the heart of the state the people who live here prefer to have personal vehicles to commute. Having more corporate offices and the world's biggest shopping malls may lead to attracting tourists and huge no of residents during the festive season.

Based on referring to the various articles, the drivers who were responsible for accepting the Lyft services will always try to meet the expected drop-off time to satisfy their customers. To achieve that most of them will overlook the traffic signals, pedestrian rules, and bumper-bumper traffic.

6.3.1 Driver Inattention/Driver Incompetence. This is the most important contributing factor responsible for more no of accidents in NYC and even in the borough-wise analysis which is topping the list every consecutive year. As we have mentioned non-licensed drivers were causing potential damage to society with poor knowledge of



following the traffic rules, rash driving and not yielding to pedestrian rules. Under this criteria we can include the following factors Backing Unsafely, Passing or Lane Usage Improper, Turning Improperly, Unsafe Lane Changing, Driver Inexperience, Aggressive Driving/Road Rage, Eating or Drinking.

6.3.2 *Following Too Closely.* Tailgating is one of the reasons the drivers were following too closely to escape the signal wait times, single-way lanes, to meet the drop-off schedules, impatience, and traffic density at peak times. This classification starts with the same trend and now Staten Island overtakes the Bronx in these types of accidents.

6.3.3 *Biological Effects:* Under this classification, the following factors were adopted from the data set and used for the metrics Fatigued/Drowsy, Lost Consciousness, Fell Asleep, Illness, Shoulders Defective/Improper, Physical Disability. Manhattan tops the list, followed by Brooklyn, Queens, Bronx, and Staten Island.

6.4 Conclusion

A few conclusions from our side would be that pedestrians face a greater threat of injury or fatality in Queens and Brooklyn, while cyclists in Manhattan and Motorists in Staten Island. The accidents trend kept increasing till 2020 and saw a steep decline because of COVID. But again, the trend is increasing and immediate measures need to be implemented to restrict it. More accidents are happening because of drivers' inattention, too close following of vehicles, and the confusion caused by right-of-way rules. The highest number of accidents are happening in the time-frame 3 PM – 5 PM. A few suggestive measures from our side after analysis would be: • More awareness programs must be conducted to explain the drivers how important pedestrian safety is and how even slight inattention while driving could cause a fatal accident. • Based on the Vision Zero program, mid-block cross-ways have been introduced in streets to ease the process of crossing the streets, this should be implemented rapidly. • Careful monitoring and speed restriction signs must be set up in the lanes where schools and hospitals are located. • Pedestrian-safe crosswalks should be an important goal throughout NYC. • More safety initiatives needs and awareness programs need to be introduced in Queens and Brooklyn boroughs for pedestrians and in Manhattan for Cyclists. • Drivers must understand that even a slight disturbance in their concentration can lead to a fatality, therefore be a hundred percent attentive when driving • More thought must be given towards of cyclists by developing special lanes for them, most importantly in Manhattan • Trucks and other heavy vehicles should be redirected from general routes to special routes, specifically away from routes where a greater number of cyclists and school buses travel.

REFERENCES

[1] Dhanushka Abeyratne and Malka N Halgamuge. 2020. Applying big data analytics on motor vehicle collision predictions in new york city. *Intelligent Data Analysis: From Data Gathering to Data Comprehension*, 219–239.

[2] Abdullah Eyad and Ahmed Emam. 2015. Traffic accidents analyzer using big data. In *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 392–397.

[3] Anne C Lusk, Morteza Asgarzadeh, and Maryam S Farvid. 2015. Database improvements for motor vehicle/bicycle crash analysis. *Injury prevention*, 21, 4, 221–230.

[4] [n. d.] Nyc government. https://www.nyc.gov/content/visionzero/pages/legislation.

[5] Mouyid Islam. 2023. An exploratory analysis of the effects of speed limits on pedestrian injury severities in vehicle-pedestrian crashes. *Journal of Transport & Health*, 28, 101561.

[6] [n. d.] Staten island parking. https://www.silive.com/news/2023/09/heres-10-reasons-why-staten-island-parking-lots-are-the-worst-opinion.html.

[7] [n. d.] Bronx and staten island. https://medium.com/@atthabrizi41/deciphering-new-york-citys-behaviour-towards-taxis-cd4b664e5000#:~:text=Although%20NYC%20Taxis%20are%20iconic,like%20Bronx%20and%20Staten%20Island..

[8] Frank Markowitz, Stanley Sciortino, Jack Lucero Fleck, and Bond M Yee. 2006. Pedestrian countdown signals: experience with an extensive pilot installation. *Institute of Transportation Engineers. ITE Journal*, 76, 1, 43.