

# MOTOR VEHICLE ACCIDENTS IN NYC

## GROUP 01

Hira Jamshed - 21100141  
Haris Saleem - 21100088  
Mahnoor Jameel - 21100069  
Saba Rehman - 21100129

# Table of Contents

Table of Contents	<b>2</b>
Rationale	3
Preprocessing	3
Frequent Pattern Mining	4
Metrics and algorithm used	4
Over The Whole Dataset	4
Borough Wise	15
Bronx	15
Additional Findings:	16
Brooklyn	17
Manhattan	18
Queens	19
Staten Islands	21
Clustering	22
K-Means	23
Agglomerative Clustering	24

## Rationale

The goal in this deliverable was to use clustering and pattern mining techniques in order to answer relational questions, such as 'When do most accidents occur and what vehicles/factors were involved in those accidents?'.

While clustering over longitude and latitude did give us useful visualizations/results, we mainly focused on frequent pattern mining. After having covered significant depth and breadth, over time, years, and boroughs, we obtained some interesting insights. The process and results are detailed in the following sections.

## Preprocessing

Before we could continue with uncovering such patterns, we first needed to:

### → Fix typos in the vehicle type code columns'

During the trial and errors phase in pre-processing, we discovered a few more typos, such as 'uhaul' being mistakenly written as 'uhual' or 'us po' being written instead of 'us postal'. Using a similar approach as that used in the EDA phase, we used regex and corrected such inconsistencies.

### → Tag each accident with the corresponding hour bracket to ensure a mechanism of uncovering frequent hourly patterns

To help in finding out the most common times at which accidents occur and to enhance explainability in our outputs, we added a new column called 'TIME TAG (hours)' in our dataset. This column hosts information of the hour bracket in which the accident occurred, for example: *an accident occurred at 16:30:47 falls under the '16-17' category and so has been tagged this way.*

### → Added a 'CRASH YEAR' column in the dataset to help in filtering

We thought it would be interesting to also dive into yearly datasets, that is, do we get any new patterns if we look at years individually as compared to the overall dataset?

To help with this, we simply extracted the year out of the CRASH DATE data.

### → Labelled vehicle type code columns' with \_\*, where \* stands for the vehicle type code number

To increase visibility in the vehicles most frequently involved in the accidents, we introduced such labels. This helped uncover patterns such as ('sedan\_1',

'sedan\_2'), which meant that the accident involved only sedans (otherwise unclear since we can't ascribe a single occurrence to an accident having the same type of vehicle).

→ Labelled the 'killed' and the 'injured' to help differentiate within patterns

To help identify what comes out in the patterns, we label the following categories as follows:

- NUMBER OF PEDESTRIANS INJURED -> PI
- NUMBER OF PEDESTRIANS KILLED -> PK
- NUMBER OF CYCLIST INJURED -> CI
- NUMBER OF CYCLIST KILLED -> CK
- NUMBER OF MOTORIST INJURED -> MI
- NUMBER OF MOTORIST KILLED -> MK

This way, an entry that states 'PI\_0' meant that 0 pedestrians were injured.

## Frequent Pattern Mining

In this phase of implementation, we explored in a lot of depth over a wide array of features. To maintain usefulness, we made sure to include the following columns: tagged time, vehicle type code, contributing factors, and all the injured/killed columns. As described below, some of the features were excluded at times to give us more room for exploring a few centralized patterns.

The csv files for all the different granularities are present here:

[https://drive.google.com/drive/folders/1UHVW-GlfvFlpv\\_0\\_iSWChSOPgGAGRpcE?usp=sharing](https://drive.google.com/drive/folders/1UHVW-GlfvFlpv_0_iSWChSOPgGAGRpcE?usp=sharing)

### Metrics and algorithm used

We used the *Frequent Pattern (FP) Growth* algorithm as it is the most efficient one for large datasets such as ours. Through trial and error, we settled for a min\_support threshold of 0.03 - a lower threshold was largely giving us the same patterns and a larger threshold seemed to give out very few patterns. We also felt that 3% was as reasonable of a low threshold as possible.

### Over The Whole Dataset

Here we went through three sorts of granularities:

1. Find patterns in the dataset as is - without including the injured and the killed:

\_\_\_\_\_ Our Findings:

- Overall, all the timings of the day were common for accidents other than the time 23:00 - 0:00 hours
- The most common vehicles involved in accidents are:
  - Taxi
  - Sedan
  - Sports utility/ Station Wagon
  - Passenger vehicles
- The most common contributing factors are:
  - Failure to yield right-of-way
  - Passagin or lane improper usage
  - Driver inattention/ distraction
  - Backing unsafely
  - Following too closely
  - Passing too closely
  - Unsafe lane changing
- The most frequent 2-length sets were:

{Sedan, Sedan}
{Sedan, Sports Utility/Station wagon}
{Sports Utility/Station wagon, Sports Utility/Station wagon}
{Driver inattention/ distraction, Sports Utility/Station wagon}
{Driver inattention/ distraction, Sedan}
{Failure to yield right-of-way, Sedan}
{Following too closely, Sedan}
{Following too closely, Sports Utility/Station wagon}
{16-17, Sedan}
{17-18, Sedan}

Explanation:

This shows how driver inattention/distraction, following too closely, failure to yield right-of way is one of the main causes of collisions of Sedans and Sports Utility wagons. One interesting find is how 16:00 - 17:00 hours is a common time for collisions of Sedans.

- The most frequent 3-length sets were:

{Driver inattention/ distraction, Sedan, Sports Utility/Station wagon}
{Driver inattention/ distraction, Sports Utility/Station wagon, Sports Utility/Station wagon}
{Driver inattention/ distraction, Sedan, Sedan}

## 2. Find patterns in the dataset - including the injured and the killed

In this part following are the findings in addition to the ones found in the previous analysis:

- Injured and killed categories:
  - PI\_1 -> Pedestrian injured 1
  - MI\_1 -> Motorist injured 1
- The length 2 itemsets in addition to the previous ones were:

{MI_1, Sedan}
{MI_1, Sports Utility/Station wagon}

- The length 3 itemsets are exactly the same the previous section

## 3. Find patterns year wise:

### **2013:**

Our Findings:

- The most common vehicles involved:
  - Vans
  - Taxis
  - Passenger vehicles
  - Sports utility/ Station Wagon
- The most common Contributing factors:
  - Fatigues/drowsy
  - Turning improperly
  - Driver inattention/distraction
- The most common 2-length sets:

{8-9, passenger vehicle}
--------------------------

{9-10, passenger vehicle}
{12-13, passenger vehicle}
{14-15, passenger vehicle}
{15-16, passenger vehicle}
{16-17, passenger vehicle}
{17-18, passenger vehicle}
{18-19, passenger vehicle}
{19-20, passenger vehicle}
{MI_1, passenger vehicle}
{passenger vehicle, passenger vehicle}
{Sports Utility/Station wagon, passenger vehicle}
{driver inattention/distraction, passenger vehicle}
{Fatigued/drowsy, passenger vehicle}
{Sports Utility/Station wagon, Sports Utility/Station wagon}

- The only common 3-length itemset:

{Fatigued/drowsy, passenger vehicle, passenger vehicle}
---

## 2014:

### Our Findings:

- Most common vehicles:
  - Passenger vehicles
  - Van
  - Taxi
  - Other vehicles
- Most common factors:
  - Fatigued/drowsy
  - Turning improperly
  - Driver inattention/distraction
- Most common from killed/injured:
  - MI\_1
  - PI\_1

- Most common 2-length itemsets:

{MI_1, passenger vehicle}
{Fatigued/drowsy, passenger vehicle}
{driver inattention/distraction, passenger vehicle}
{passenger vehicle, passenger vehicle}
{12-13, passenger vehicle}
{13-14, passenger vehicle}
{14-15, passenger vehicle}
{15-16, passenger vehicle}
{16-17, passenger vehicle}
{17-18, passenger vehicle}

- Most common 3-length itemsets:

{Fatigued/drowsy, passenger vehicle, passenger vehicle}
---

## 2015:

### Our Findings:

- Most common vehicles:
  - Passenger vehicles
  - Sports Utility/Station wagon
  - Taxi
  - Van
  - Other vehicular
- Most common Contributing factors:
  - Fatigues/drowsy
  - Turning improperly
  - Driver inattention/distraction
  - Failure to yield right-of-way
  - A new one: Lost consciousness
- Common 2-length itemsets:

{MI_1, passenger vehicle}
{Fatigued/drowsy, passenger vehicle}



{Sports Utility/Station wagon, passenger vehicle}
{Sports Utility/Station wagon, Sports Utility/Station wagon}
{passenger vehicle, passenger vehicle}
{driver inattention/distraction, passenger vehicle}
{driver inattention/distraction, Sports Utility/Station wagon}

- The most common 3-length itemset:

{driver inattention/distraction, passenger vehicle, passenger vehicle}
--

## 2016:

### Our Findings:

- Most common vehicles:
  - Sedan
  - Sports Utility/Station Wagon
  - Passenger Vehicle
- Most common contributing factors:
  - Following too closely
  - Unsafe lane changing
  - Improper lane usage
  - Failure to yield right-of-way
  - Driver inattention/distraction
- Common two-length itemsets:

{Sports Utility/Station wagon, Sedan}
{Driver inattention/distraction, Sedan}
{Driver inattention/distraction, Sports Utility/Station wagon}
{Sports Utility/Station wagon, Sports Utility/Station wagon}
{Following too closely, Sedan}
{Following too closely, Sports Utility/Station wagon}
{Sedan, Sedan}
{MI_1, Sedan}

{16:00 - 17:00, Sedan}
{17:00 - 18:00, Sedan}
{18:00 - 19:00, Sedan}

- Common three-length itemsets:

{Driver inattention/distraction, Sedan, Sedan}
--

## 2017:

- The most common vehicles:
  - Sedan
  - Taxi
  - Sports Utility/Station wagon
- The most common Contributing Factors:
  - Backing unsafely
  - Driver inattention/distraction
  - Failure to yield right-of-way
  - Passing or lane improper usage
  - Unsafe lane changing
  - Passing too closely
- The most common 2-length itemsets:

{following too closely, Sedan}
{following too closely, Sports Utility/Station wagon}
{driver inattention, Sedan}
{driver inattention, Sports Utility/Station wagon}
{failure to yield right-of-way, Sedan}
{Sedan, Sports Utility/Station wagon}
{MI_1, Sedan}
{14-15, Sedan}
{15-16, Sedan}
{16-17, Sedan}
{17-18, Sedan}

{18-19, Sedan}
----------------

- The most common 3-length itemsets:

{driver inattention, Sedan, Sedan}
------------------------------------

{driver inattention, Sedan, Sports Utility/Station wagon}
---

**2018:**

- Most common vehicles:
  - Sedan
  - Taxi
  - Sports Utility/Station wagon
  - New one: Pickup truck
- The most common contributing factors:
  - Backing unsafely
  - Driver inattention/distraction
  - Failure to yield right-of-way
  - Passing or lane improper usage
  - Unsafe lane changing
  - Passing too closely
- The most common 2-length itemsets:

{Sedan, Sports Utility/Station wagon}
---------------------------------------

{driver inattention, Sedan}
-----------------------------

{driver inattention, Sports Utility/Station wagon}
--

{following too closely, Sedan}
--------------------------------

{following too closely, Sports Utility/Station wagon}
---

{MI_1, Sedan}
---------------

{14-15, Sedan}
----------------

{16-17, Sedan}
----------------

{17-18, Sedan}
----------------

{18-19, Sedan}
----------------

- The most common 3-length itemsets:

{driver inattention, Sedan, Sedan}
{driver inattention, Sedan, Sports Utility/Station wagon}
{driver inattention, Sports Utility/Station wagon, Sports Utility/Station wagon}

## 2019:

### Our Findings:

- Most common vehicles:
  - Sedan
  - Taxi
  - Sports Utility/Station wagon
- Most common contributing factors:
  - Following too closely
  - Unsafe lane changing
  - Improper lane usage
  - Failure to yield right-of-way
  - Driver inattention/distraction
  - Backing unsafely
- Most common two-length itemsets:

{Sedan, Sedan}
{Sedan, Sports Utility/Station wagon}
{Driver inattention, Sedan}
{Driver inattention, Sports Utility/Station wagon}
{Following too closely, Sedan}
{Following too closely, Sports Utility/Station wagon}
{Failure to yield right-of-way, Sports Utility/Station wagon}
{MI_1, Sedan}
{MI_1, Sports Utility/Station Wagon}
{16:00 - 17:00, Sedan}
{17:00 - 18:00, Sedan}

- Most common three-length itemsets:

{driver inattention, Sedan, Sedan}
{driver inattention, Sedan, Sports Utility/Station wagon}
{driver inattention, Sports Utility/Station wagon, Sports Utility/Station wagon}

## 2020:

- The most common vehicles:
  - Sedan
  - Taxi
  - Sports Utility/Station wagon
  - New one: Bike
- A new common in the injured killed category:
  - CI\_1 -> cyclist injured
- The most common contributing factors:
  - Following too closely
  - Unsafe lane changing
  - Improper lane usage
  - Failure to yield right-of-way
  - Driver inattention/distraction
  - Backing unsafely
- The most common 2-length itemsets:

{Driver inattention, Sedan}
{Driver inattention, Sports Utility/Station wagon}
{MI_1, Sedan}
{MI_1, Sports Utility/Station wagon}
{MI_1, Driver inattention}
{Following too closely, Sedan}
{Following too closely, Sports Utility/Station wagon}
{Failure to yield right-of-way, Sedan}
{14-15, Sedan}
{16-17, Sedan}
{17-18, Sedan}

- The most common 3-length itemsets:

{driver inattention, Sedan, Sedan}
{driver inattention, Sedan, Sports Utility/Station wagon}
{driver inattention, Sports Utility/Station wagon, Sports Utility/Station wagon}

## 2021:

### Our Findings:

- The most common vehicles:
  - Sedan
  - Sports Utility/Station wagon
- The most common contributing factors:
  - Following too closely
  - Unsafe speed
  - Improper lane usage
  - Failure to yield right-of-way
  - Driver inattention/distraction
  - Backing unsafely
- The most common two-length itemsets:

{Driver inattention, Sports Utility/Station wagon}
{Driver inattention, Sedan}
{Sedan, Sports Utility/Station wagon}
{Sedan, Sedan}
{Following too closely, Sedan}
{MI_1, Sedan}
{MI_1, Sports Utility/Station wagon}
{MI_1, Driver inattention}

- The most common three-length itemsets:

{driver inattention, Sedan, Sedan}
------------------------------------

{MI_1, Sedan, Sedan}
----------------------

### Conclusion for over the years:

In the starting few years the most common vehicles were passenger vehicles which then turned into Sedan in the later years. Sports utility/ station wagon stays common throughout. The Most common contributing factors were: failure to yield right-of-way, driver inattention/distraction, backing unsafely etc. And the most common accidents occur between Sedan the Sports Utility. And for all the 3-length itemsets the most common accident btw Sedan, Sports Utility, with driver inattention. In the injured/killed category MI\_1 was the most frequent.

## Borough Wise

### Bronx

#### 1. **Find patterns in Bronx data - without including the injured and the killed:**

Our Findings:

- In Bronx, the most frequent times for accident occurrences were between 9AM to 7PM
- The most common vehicles involved in accidents were:
  - Sedan
  - Sports Utility/Station Wagon
- The most common contributing factors were:
  - Driver inattention/distraction
  - Backing unsafely
  - Following too closely
  - Failure to yield right-of-way
- The most frequent two-length itemsets were:

{Sedan, Sport utility / Station wagon}
{Sedan, Sedan}
{Sport utility / Station wagon, Sport utility / Station wagon}
{Sedan, Driver inattention / Distraction}
{Sport utility / Station wagon, Driver inattention / Distraction}

Explanation: A common feature for the accidents in the Bronx is that Sedan and Station Wagons are often caught up in collisions and the cause of these collisions is mostly inattention or distractedness of the driver.

## **2. Find patterns in Bronx data - including the injured and the killed:**

Additional Findings:

- Most common vehicles:
  - Taxi
  - Passenger vehicle
- Most common contributing factors:
  - Turning improperly
  - Passing too closely
  - Passing or lane usage improper
- Injured and killed categories:
  - MI\_1 -> Motorist injured 1
  - PI\_1 -> Pedestrian injured 1
- Two-length itemsets:

{Sedan, MI_1}
{Sedan, 14:00-15:00}
{Sedan, 15:00-16:00}
{Sedan, 16:00-17:00}
{Sedan, 17:00-18:00}

- Three-length itemsets:

{Seda, Sedan, Driver inattention / Distraction}
---

### **Conclusion:**

- Motorists are most likely to get injured
- Accidents occur mostly during rush hours like in the morning during office hours or at lunch time or after work.
- Sedan and station wagons are the most common types of vehicles in accidents most probably because they are widely present.
- Driver inattention is a recurring cause for collisions



## Brooklyn

### 1. Find patterns in Brooklyn data - without including the injured and the killed:

Our Findings:

- In Brooklyn, the most frequent times for accident occurrences were between 8 AM to 11 PM and from 12 AM to 1 AM
- The most common vehicles involved in the accidents were:
  - Sedan
  - Sports Utility/Station Wagon
- The most common contributing factors were:
  - Failure to yield right-of-way
  - Passing or lane usage improper
  - Driver inattention/distraction
  - Following/passing too closely
  - Backing unsafely
- The most frequent two-length itemsets were:

{'Failure to yield right-of-way', 'Sedan'}
{'Sedan', 'Sport utility / Station wagon'}
{Sedan, Sedan}
{'Driver inattention/distraction', 'Sport utility / Station wagon'}
{ 'Sport utility / Station wagon', 'Sport utility / Station wagon'}
{'Sedan', 'Following too closely'}
{'18:00 - 19:00', 'Sedan'}
{'14:00 - 15:00', 'Sedan'}
{'17:00 - 18:00', 'Sedan'}
{'16:00 - 17:00', 'Sedan'}

- The most frequent three-length itemsets were:

{'Driver inattention/distraction', 'Sedan', 'Sedan'}
--

Explanation:

It was interesting to note that Sedan was always paired with a station wagon during accidents, and that the evening time stood out most for accidents involving Sedan.

## **2. Find patterns in Brooklyn data - including the injured and the killed:**

Additional findings:

- Most common vehicles:
  - Passenger vehicle
- Injured and/or killed categories:
  - ML\_1 -> Motorist injured 1
  - CL\_1 -> Cyclist injured 1
  - PL\_1 -> Pedestrian injured 1
- The itemsets were the same

### **Conclusion:**

- Motorists, cyclists and pedestrians are all susceptible to getting injured
- Sedan and station wagons are the most common types of vehicles in accidents most probably because they are widely present.
- Sedan drivers are involved in collisions due to inattention and failure to yield right of way
- Lunch hours to after work hours are times when most accidents occur.

## **Manhattan**

### **1. Find patterns in Manhattan data - without including the injured and the killed:**

Our Findings:

- In Manhattan, the most frequent times for accidents to occur were in between 8AM to 1 AM
- Most common vehicles involved in accidents:
  - Sedan
  - Sport Utility / Station Wagon
  - Taxi
  - Box Truck
  - Bike
- Most common contributing factors:
  - Driver inattention / Distraction
  - Failure to yield right-of-way
  - Turning improperly
  - Following too closely
  - Passing too closely
  - Backing unsafely
  - Passing or lane usage improper
  - Unsafe lane changing

- Two-length item sets:

{sedan, driver inattention / distraction}
{sport utility / station wagon, driver inattention / distraction}
{sedan, sport utility / station wagon}
{taxi, driver inattention / distraction}
{sedan, taxi}
{taxi, taxi}
{sport utility / station wagon, sport utility / station wagon}
{sport utility / station wagon, driver inattention / distraction}
{sedan, sedan}

- Three-length itemsets:

{sedan, sedan, driver inattention / distraction}
--

Explanation: In Manhattan, we see a variety of factors that contribute to accidents. Taxi and bike accidents are also more common here.

## 2. **Find patterns in Manhattan data - including the injured and the killed:**

Additional Findings:

- Injured and killed categories:
  - MI\_1 -> Motorist injured 1
  - PI\_1 -> Pedestrian injured 1
  - CI\_1 -> Cyclist injured 1
- Two-length and three-length itemsets are the same

### **Conclusion:**

- Taxis, bikes and box trucks accidents are frequent
- Motorists, Cyclists and Pedestrians are likely to get injured according to frequent patterns
- Sedan and sport utility remain the vehicles most frequently involved in collisions

## **Queens**

### 1. **Find patterns in Queens data - without including the injured and the killed:**

Our Findings:

- In Queens, the most frequent times for accidents to occur were in between 8AM to 10AM and 11AM to 8PM
- Most common vehicles:
  - Sport utility/ Station wagon
  - Sedan
- Most common contributing factor:
  - Following too closely
  - Driver inattention / distraction
  - Backing unsafely
  - Failure to yield right-of-way
- Two-length itemsets:

{sport utility / station wagon, sport utility / station wagon}
{sedan, sport utility / station wagon}
{sport utility / station wagon, driver inattention / distraction}
{sedan, driver inattention / distraction}
{sedan, sedan}

- Three-length itemsets:

{sedan, sedan, driver inattention / distraction}
--

Explanation: Accidents occur until 8PM at night. Sedan and Sport utility vehicles are mostly caught up in collisions. Driver inattention seems to be a very recurring contributing factor.

## 2. **Find patterns in Queens data - including the injured and the killed:**

Additional Findings:

- Injured and killed categories:
  - MI\_1 -> Motorist injured 1
  - PI\_1 -> Pedestrian injured 1
- Two-length itemsets:

{sport utility / station wagon, MI_1}
{sedan, MI_1}
{sedan, 14:00-15:00}

{sedan, 16:00-17:00}
{sedan, 17:00-18:00}
{sport utility / station wagon, failure to yield right-of-way}
{sedan, failure to yield right-of-way}

- Three-length itemsets:

{sedan, sedan, driver inattention / distraction}
{sedan, sport utility / station wagon, driver inattention / distraction}
{sport utility / station wagon, sport utility / station wagon, driver inattention / distraction}

### Conclusion:

- Motorists and pedestrians mostly get injured in accidents
- Driver inattention seems to be a cause in sedan-sedan, sedan-sports util and sports util-sports util accidents.
- 2PM-6PM seems to be a frequent time for sedan accidents.

## Staten Islands

### 1. Find patterns in Staten Islands data - without including the injured and the killed:

Our Findings:

- In Staten Islands, the most frequent times for accidents to occur were in between 8AM to 10AM and 11AM to 7PM
- Most common vehicles:
  - Sedan
  - Sport utility / station wagon
- Most common contributing factors:
  - Following too closely
  - Driver inattention / distraction
  - Backing unsafely
  - Failure to yield right-of-way
- Two-length itemsets:

{sedan, sedan}
----------------

{sedan, sport utility / station wagon}
{sport utility / station wagon, sport utility / station wagon}
{sedan, driver inattention / distraction}
{sport utility / station wagon, driver inattention / distraction}

- Three-length itemsets:

{sedan, sedan, driver inattention / distraction}
--

Explanation: Staten Islands is a relatively smaller borough and has the smallest population therefore, there isn't a lot of variation in the patterns found. Sedan and sport utility seem to be the most common vehicles and driver inattention the most common factor.

## 2. **Find patterns in Staten Islands data - including the injured and the killed:**

Additional Findings:

- Most common vehicles:
  - Pick-up truck
- Most common contributing factors:
  - Turning improperly
- Injured and killed categories:
  - MI\_1 -> Motorist injured 1
  - PI\_1 -> Pedestrian injured 1
- Two-length itemsets:

{sedan, MI_1}
{sport utility / station wagon, MI_1}

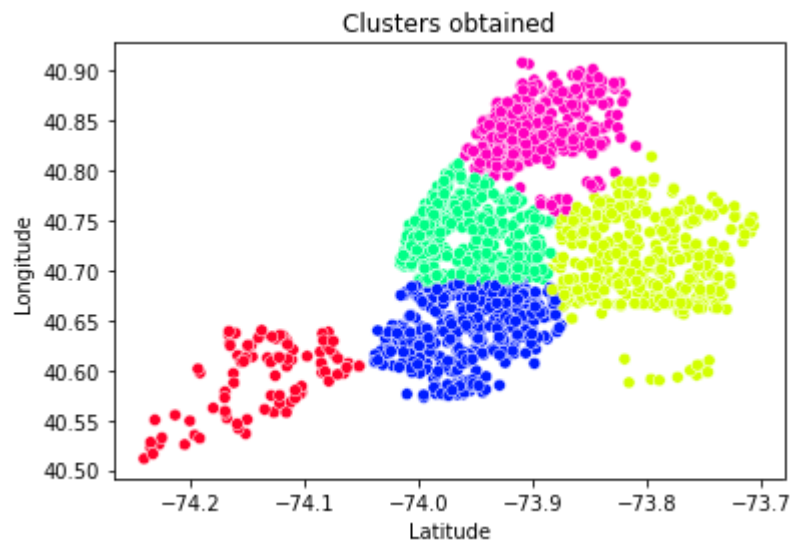
### **Conclusion:**

- Motorists and pedestrians get injured frequently
- Pickup trucks are also a common vehicle
- Accidents after 7PM are not common which might explain that Staten Islands isn't as lively at night compared to other boroughs.

## Clustering

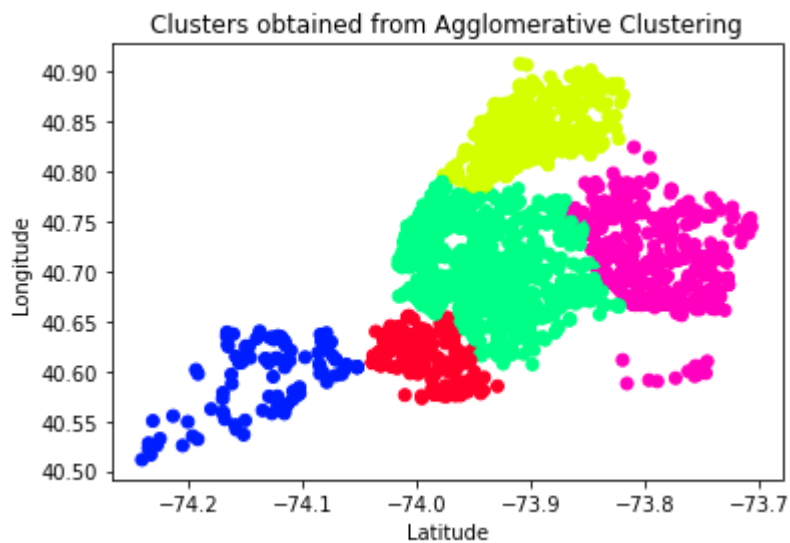
Clustering was done using Longitude and Latitude. Two types of clustering were used

## K-Means



1. We used K-means clustering to find clusters in our dataset. Optimum number of clusters we used were 5.
2. The clusters do reflect the Boroughs
  - a. Red - Staten Island
  - b. Blue - Brooklyn
  - c. Green - Manhattan
  - d. Pink - Bronx
  - e. Yellow - Queens
3. Scatter plot was not taking data other than the initial 2115 values of longitude and latitude, due to which we see some anomaly in some clusters. The reason for this is unknown to us.

## Agglomerative Clustering



1. Agglomerative Clustering shows us 5 clusters, again the points are limited to initial 2115 pair of collision locations.
2. Clusters represent
  - a. Blue - Staten Island
  - b. Red Brooklyn
  - c. Green Manhattan overlapped with Brooklyn
  - d. Yellow - Bronx
  - e. Pink - Queens
3. Although we cannot make many conclusions as this sub-data is very small. However, we can see that Brooklyn and MANhattan are more dense than others. Which represent more accidents in these Boroughs
4. Staten Island has lowest number of accidents