Big Data Analytics Symposium
Fall 2018

Analyzing Vehicle Collisions and Traffic Violations in NYC

Omkar Patinge (opp212)

Shreya Kakkar (sk7748)

Urmi Jadhav (uj256)

Analyzing Vehicle Collisions and Traffic Violations in NYC

Abstract

Every year there are a huge number of motor vehicle collisions reported across the world, leading to major loss of precious life and property. Having said that, we thought it would be insightful to draw a relation between the petty offenses and the major ones. The idea we are trying to establish is that regions where the residents showcase a general indifference towards trivial traffic rules like parking violations, skipping a red light, etc are more susceptible to life threatening road accidents. Apart from this, vehicle collisions could vary with changing times and weathers and it becomes imperative to be more careful during times when the roads are more susceptible to danger.
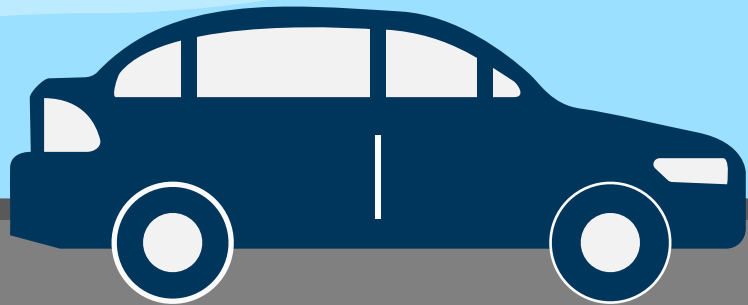
Motivation

- **Users of the analytic**

  Our analytic helps identify patterns and trends in road accidents and traffic violations. The direct user from our analytic would be the NYPD Traffic Enforcement and NY State department of motor vehicles, since this will help them identify regions that need more attention.

- **Beneficiaries**

  While the NYPD Traffic dept. will be able to make the most use of the analytic, to frame better amendments in the traffic rules, it is actually the general public that benefits from the analytic in the long run. Lesser accidents will promote a lesser mortality rate

Motivation

- **<u>Importance of the Analytic</u>**

Roadways is the most common and significant mode of transportation in the 21st century. Statistics show that every year approximately 1,093 deaths, 13,000 hospitalizations and 140,000 emergency care are made due to motor collisions in NYC. The dependance on the roadways is not expected to decrease significantly in the near future and so it is of utmost importance that we make our roads safer for the public. This being said, our analytic focuses on highlighting causes of accidents, we are also zeroing in on certain areas that need our urgent attention. The results of our analytic when combined with stricter law enforcement and a general regard for rules in the public could bring about a significant decrease in the death toll that accidents have caused.

Goodness of the Analytic

The goodness of our analytic can be measured against the following judgements :

- Our hypothesis was that the areas where the residents show a general pattern of carelessness towards the traffic rules, be it harmless rules like parking violations tend to show more tendency of violating the major rules like collisions. Our results show a strong correlation with the coefficient as high as 0.887. Despite being handicapped due to paucity of accurate data we were able to prove our results, which shows that there definitely is a relation between the two factors.

- Next, our results show that the areas with more population and urbanization show more number of collisions. Logically, this verifies our results since we expected to see such trends in our data.

- We can prove that the hotspots generated by our analytic is correct by cross-verifying it with historical articles on the web.(https://www.subinlaw.com/blog/2018/04/most-accident-prone-roads-in-new-york.shtml)

Data Sources

01 Parking Violations Issued

The dataset contains a list of all the parking violations reported in NYC by location. Each record represents a parking violation in NYC along with the violation code, borough, time and street

Size of Data: 2.1 GB

02 Traffic Tickets

The dataset contains a summary of moving violations (backing unsafely, use of the wrong lane, improper turns, not giving proper indicators, skipping traffic lights, etc.) in NYC in each borough.

Size of Data: 1.5 GB

03 Motor Vehicle Collisions

The dataset contains a breakdown of every collision in NYC by location and injury. Each record represents a collision in NYC by city, borough, precinct and cross street.
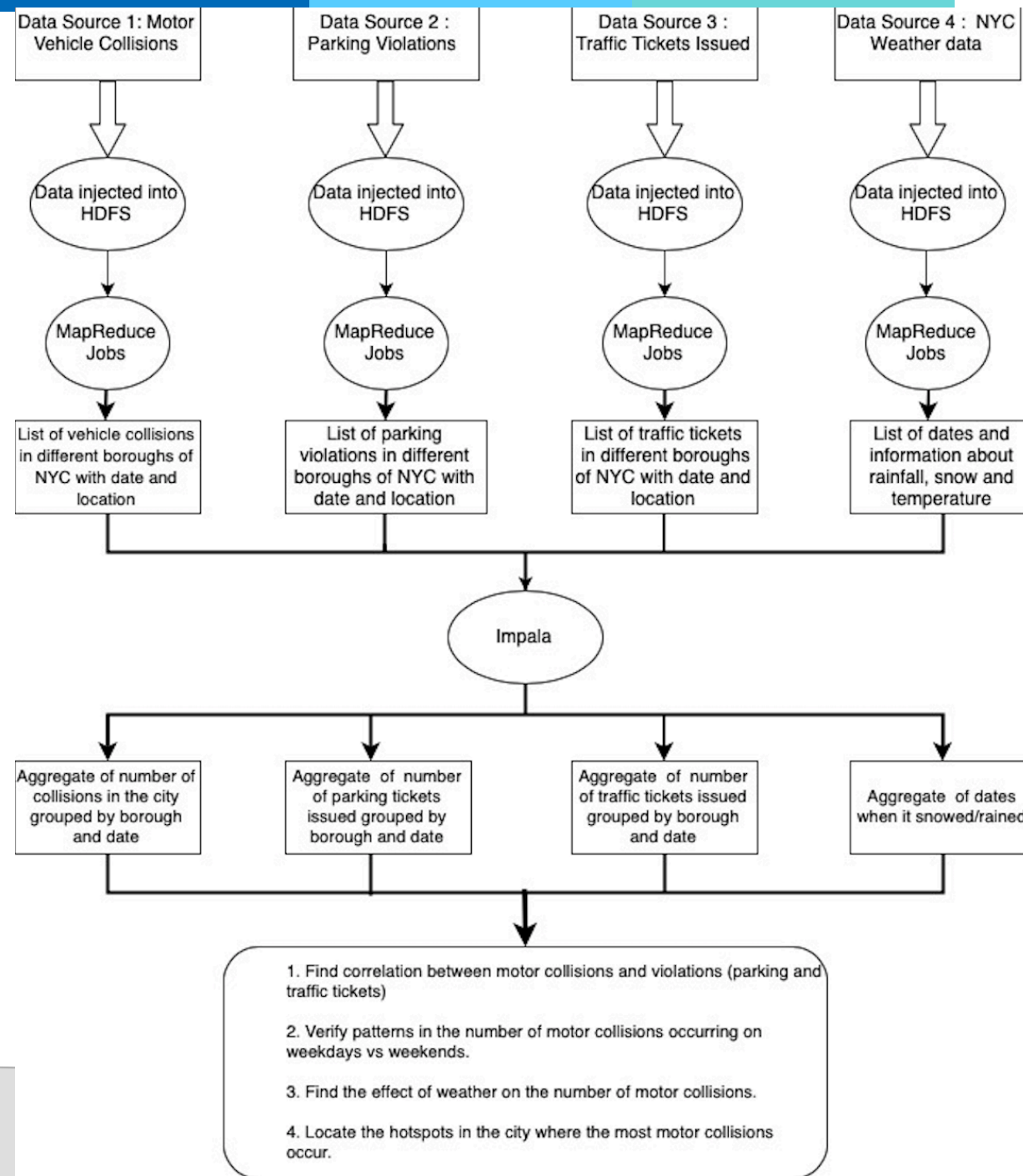
Size of Data: 500 MB

04 Weather data in New York City

The dataset contains a summary of weather through the year 2016, with the amount of precipitation, snow depth, density of rainfall and the temperature highs and lows recorded for each day.
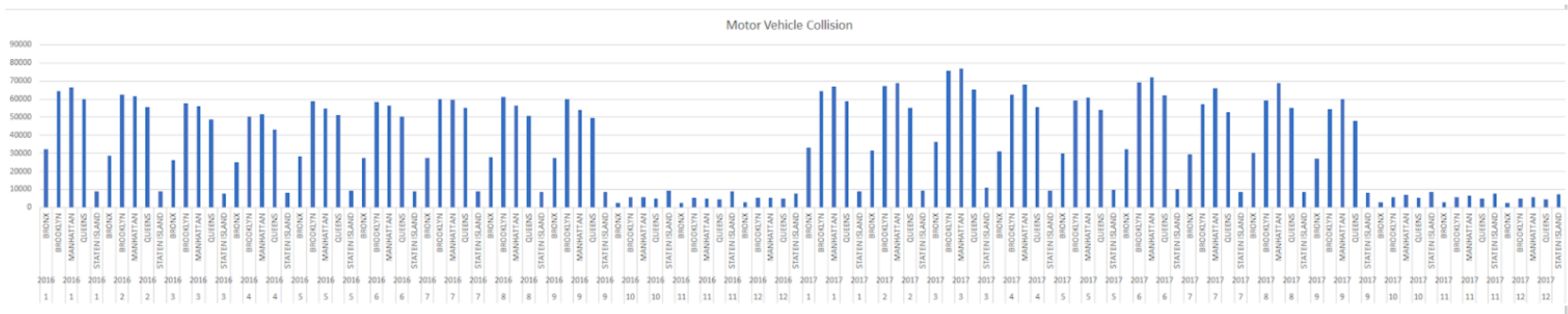
Size of Data: 12 KB

Design Diagram

Results

We found a strong correlation between the petty traffic offenses and motor collisions. A correlation coefficient of 0.887 was achieved. The bar graph below shows how our our data maps out:
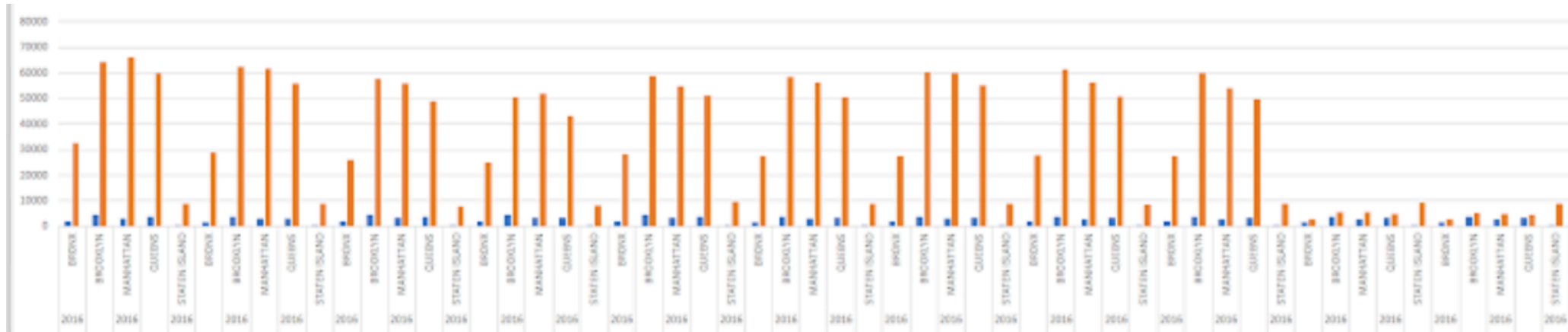

Motor Vehicle Collision(Number of collisions vs Month(grouped by Borough))

Results

We found a strong correlation between the petty traffic offenses and motor collisions. A correlation coefficient of 0.887 was achieved. The bar graph below shows how our our data maps out:
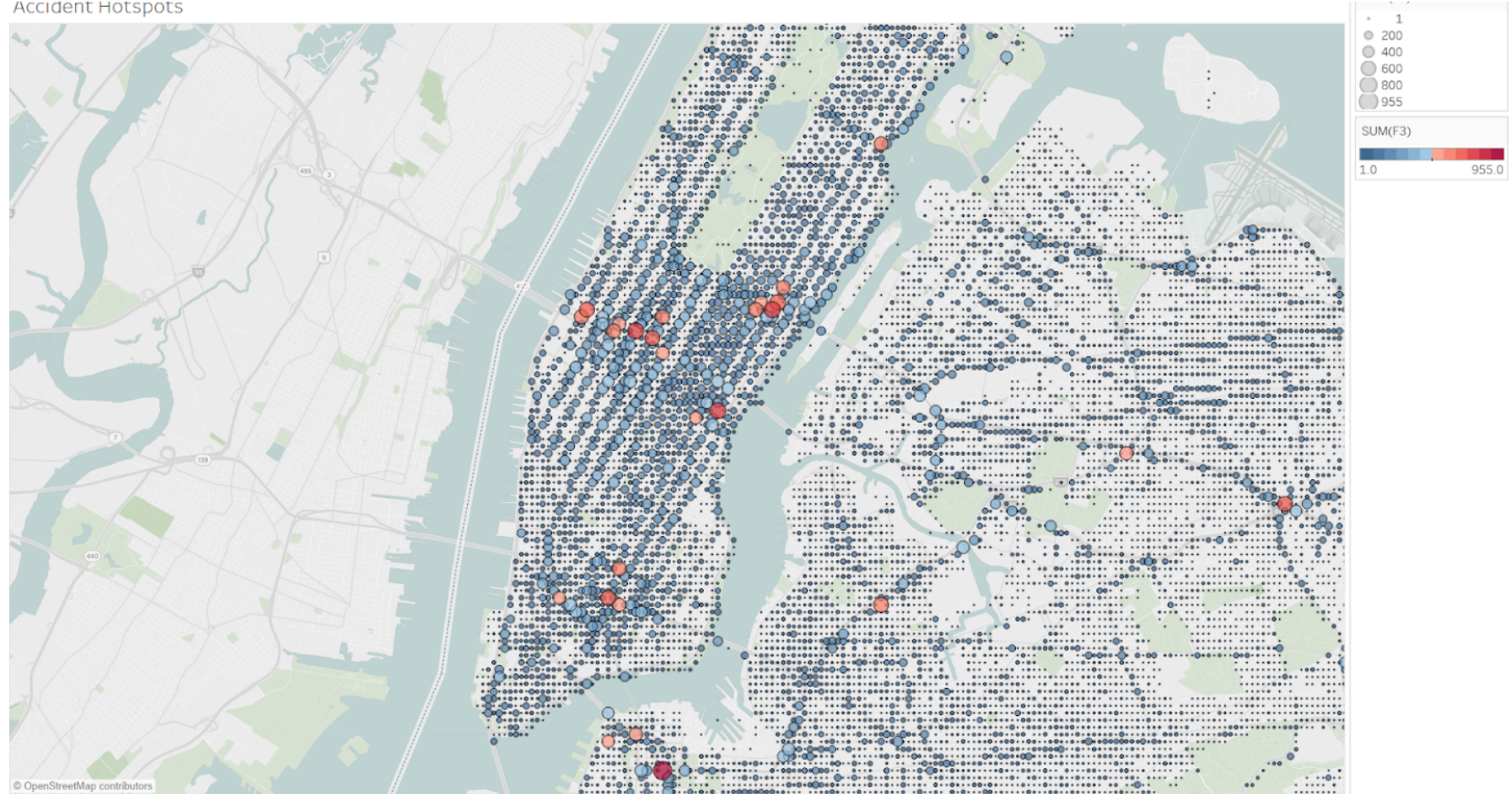

**Side by Side Comparison of Motor Collisions and Traffic Violations**

Accident HotSpots

Correlating Motor Collisions with Weather

Unfortunately we couldn't find any correlation of weather(precipitation) and motor vehicle collisions. This shocking result can be attributed to some missing data in the Motor Collision dataset. The following pie chart shows the accident ratio for days with precipitation and the ones without.

■ With Precipitation   ■ Without Precipitation
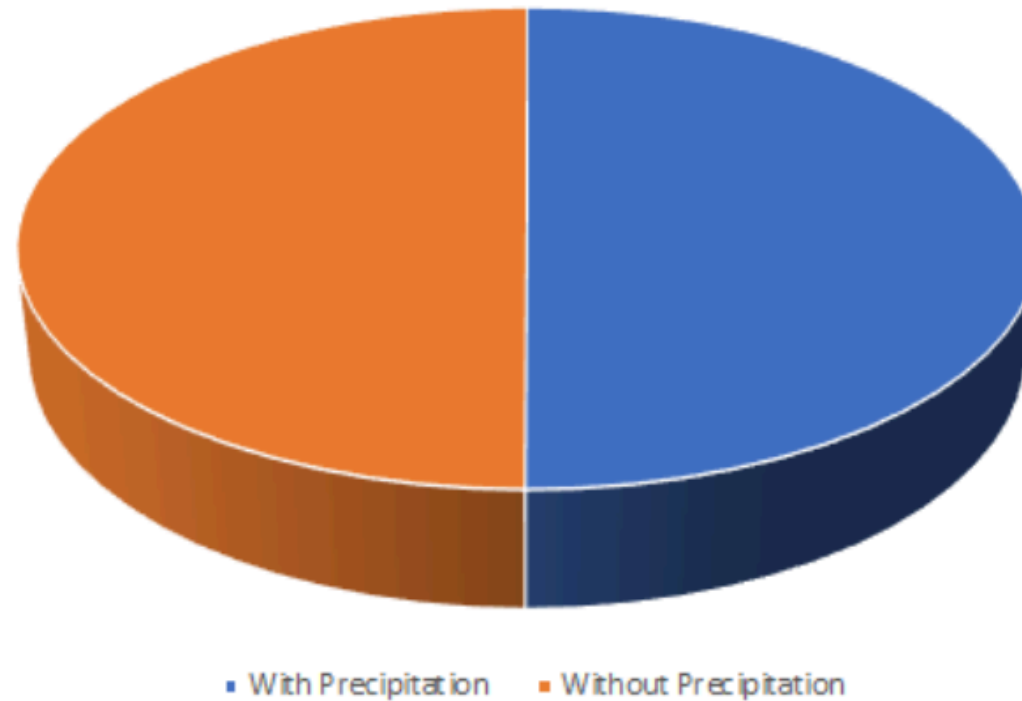
Obstacles

Inaccurate data

Like most data related projects dependant on open sourced data, our major set back was because of incorrect data. As shown in the diagram below the motor collisions recorded is exceptionally low for the months of October-December. This could be a cause of inaccurate results.
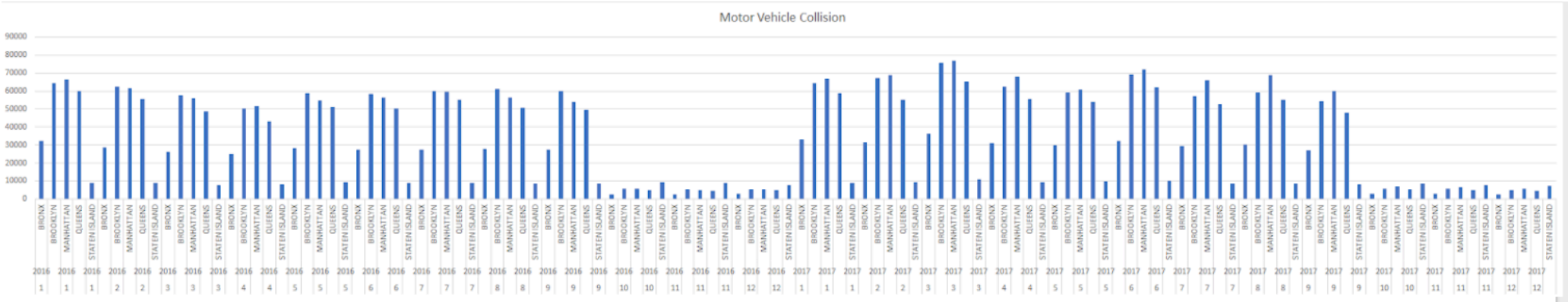
Obstacles

2. **Limitations due to absence of essential fields in the datasets**

Our datasets did not have accurate information pointing to the exact locations, which led us to change our analytic idea midway. We were constrained by this limitation and could explore the problem statement only up to a particular point.

Summary

There are a huge number of motor vehicle collisions reported across the world, leading to loss of precious life and property. The main idea of the project is to find a link between the amount of traffic violations that happen across the different boroughs and the number of motor collisions. We wish to show that if the residents exhibit a general trend of carelessness concerning the trivial traffic rules, then it is more likely that those areas will be more prone to major road accidents like motor vehicle collisions. We also try to draw a relation between the changing weather and the number of collisions, however fall short on the data front.
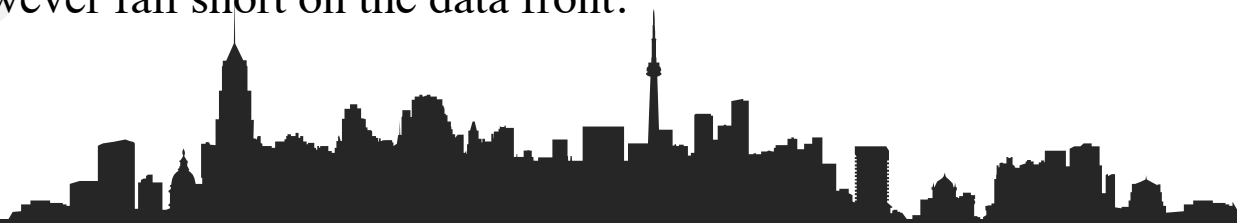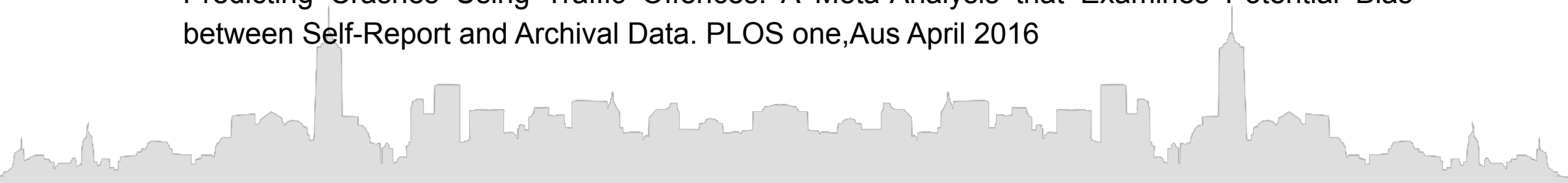
Acknowledgements

We express our sincere gratitude to Professor Mcintosh for her constant support and guidance. She always made herself available to clear all of our doubts and has made this a delightful learning experience for us!

We would also like to thank the HPC team for being kind enough to answer our queries round the clock and the graders for helping us understand the intricate details of the project.

Thank you for your help!

References

- Manik Katyal, Parag Chhadva, Shubhra Wahi & Jongwook Woo. Big Data Analysis using Spark for Collision Rate Near CalStateLA. Global Journals Inc. (USA) ISSN: 0975-4172

- Rene Richard, Suprio Ray. A tale of two cities: Analyzing road accidents with big spatial data. IEEE International Conference on Big Data (Big Data), 2017

- Addi Ait-Mlouk*, Fatima Gharnati, Tarik Agouti. Application of Big Data Analysis with Decision Tree for Road Accident. Indian Journal of Science and Technology, Vol 10(29), August 2017

- Eyad Abdullah, Ahmed Emam. Traffic Accidents Analyzer Using Big Data. International Conference on Computational Science and Computational Intelligence (CSCI), 2015.

- 5. Peter Barraclough, Anders af Wåhlberg, James Freeman, Barry Watson,Angela Watson. Predicting Crashes Using Traffic Offences. A Meta-Analysis that Examines Potential Bias between Self-Report and Archival Data. PLOS one,Aus April 2016

Thank you!