



NYPD Traffic Accident Analysis

Group 3 Tiffany Sung, Won Lee, Chelsea Matthews



Outline.

- Dataset Introduction
- Data Exploration
- Models
- Problems
- Insights



Dataset Introduction.

NYPD Motor Vehicle Collisions.
from 2012/07/01-2018/07/28



Dataset.

- Dataset: NYPD Motor Vehicle Collisions
- Source: NYC OpenData
- Numbers of observations: 1.32M
- 29 columns of inputs:
 - Date/time : 07/01/2012-07/28/2018
 - Geographical: Borough, Zip code, Latitude, Longitude, Location, On street name, cross street name, off street name
 - Injuries: Number of (person/ pedestrians/ cyclists/ motorists) injured/ killed
 - Contribution factors
 - Vehicle types



Our Problem.

How do we help NYPD reduce the number of traffic collision in New York City?
How do we improve the efficiency of reporting traffic accidents?



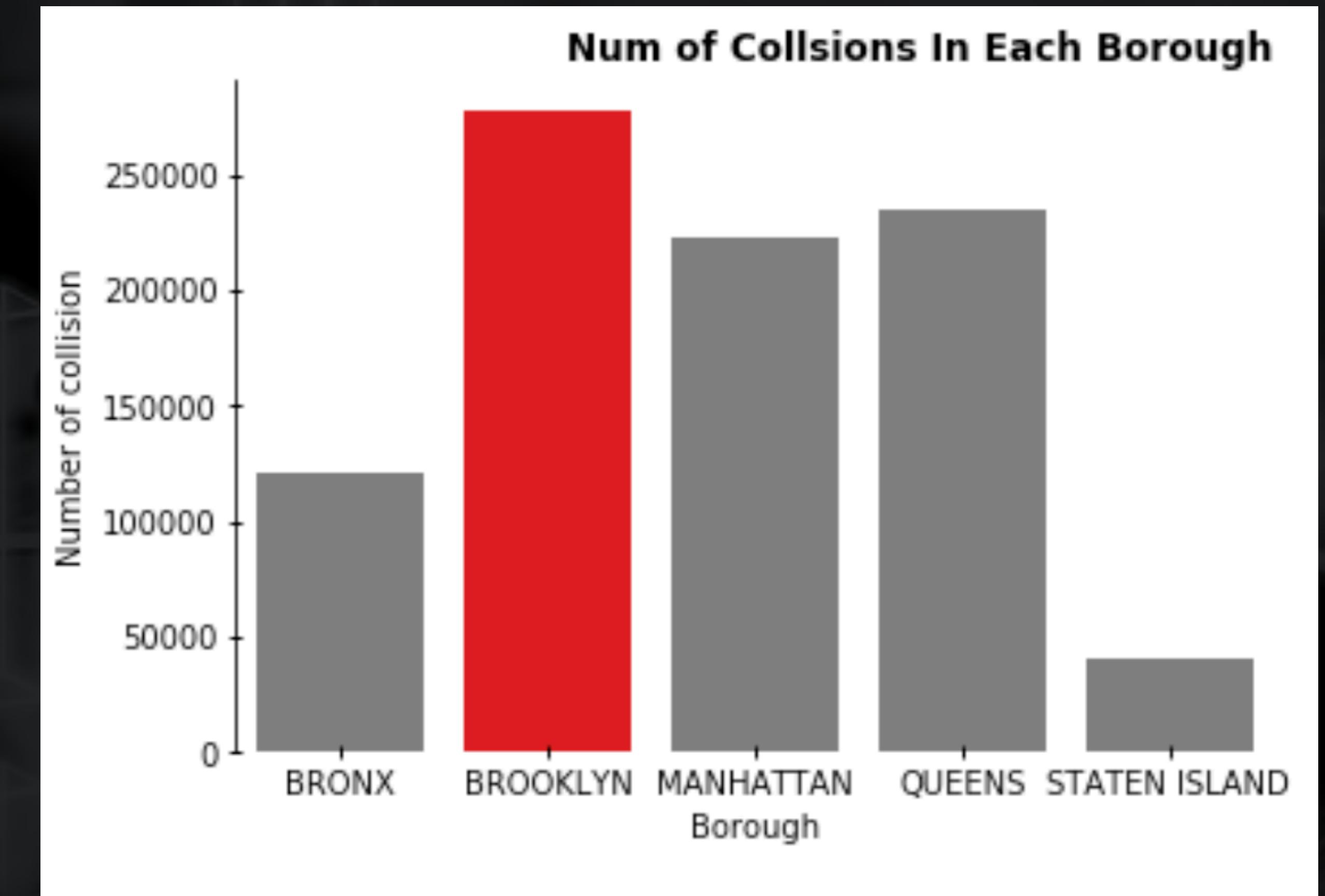
Exploratory Analysis.

To better answer our problem sets, we first take a thorough exploratory analysis of the dataset. We explore the collision numbers base on time frame, borough, and zip code. We also classified those accidents by the severity of injuries. Also, we further explore the contributing factors over the week and found some insightful patterns.



Number of Collision in each borough.

BROOKLYN	277616
QUEENS	234694
MANHATTAN	223400
BRONX	120449
STATEN ISLAND	41017





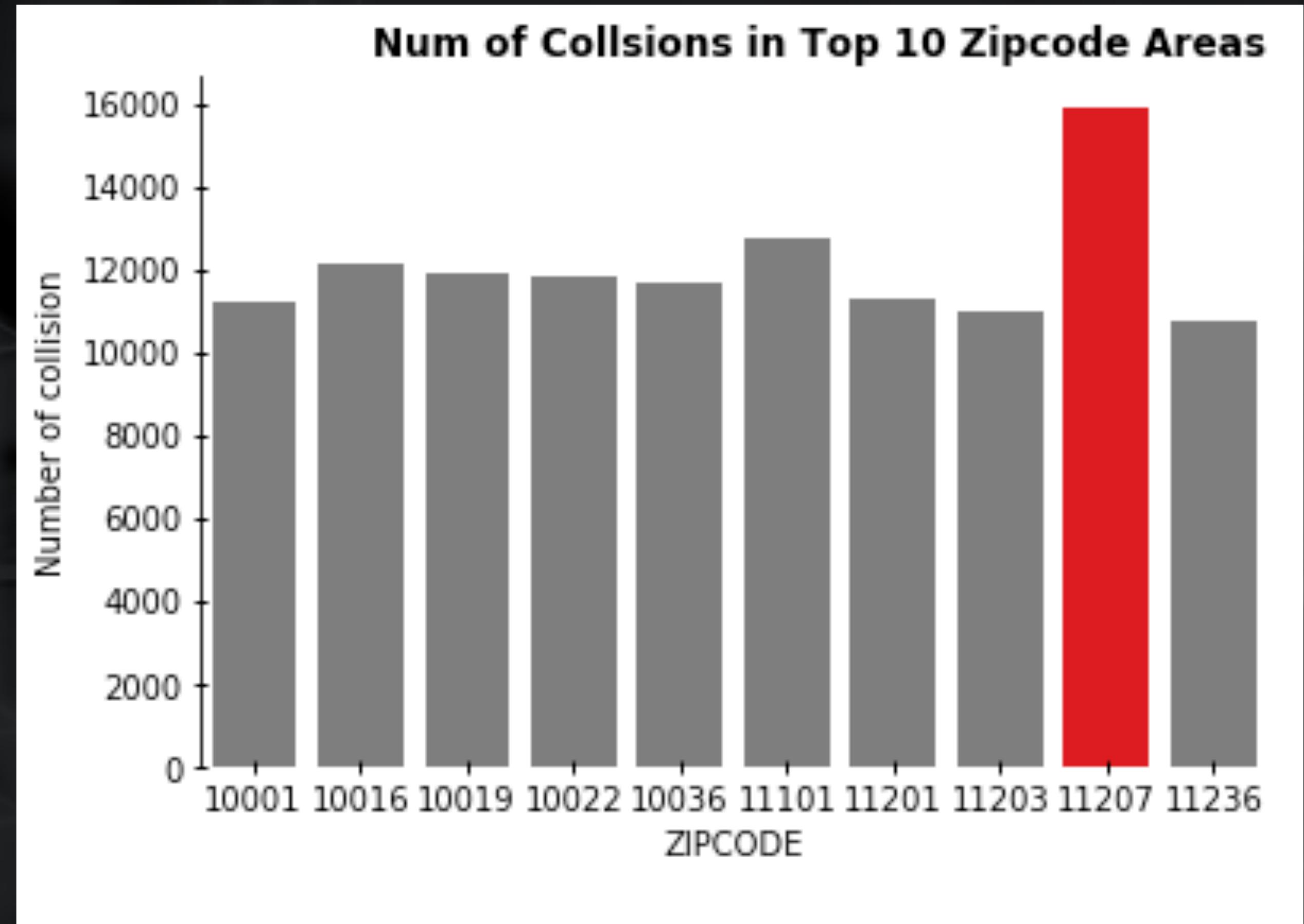
Number of Collision.

by zip code

Question: Does it means Brooklyn is the most dangerous?

From the graph, the top 3 zip code fall into Brooklyn, Queens and Manhattan.

So the collisions actually happens in all places around NYC, we can suggest that Brooklyn has most collision numbers because the area is the biggest.





Number of Injuries.

Questions: How severe were all those collisions?

From our dataset, we got number of injuries and number of persons killed in each collision, so we can assume those collision without any reported injuries or killed were 'no injuries'.

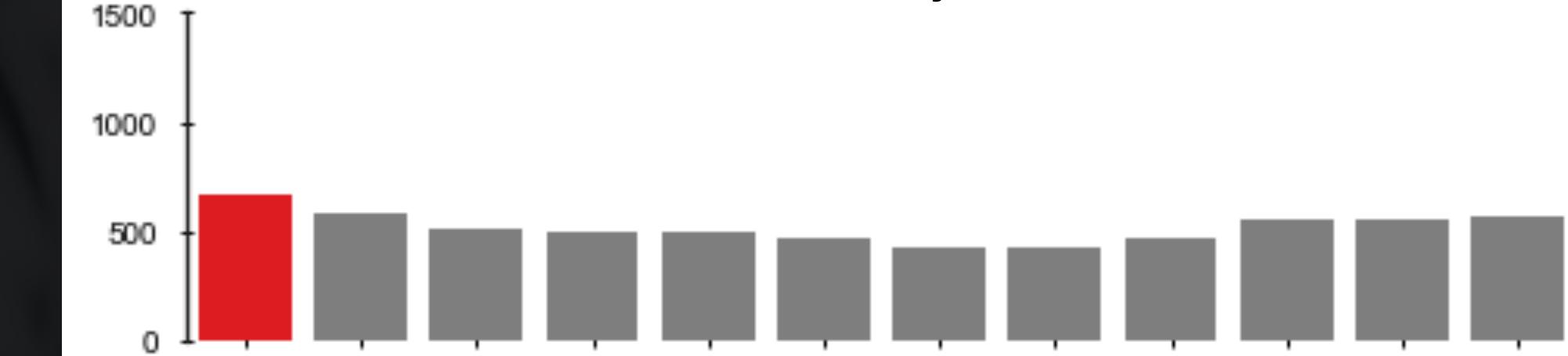
No injuries	729266
Injured	166990
Fatal	920

Number of Injuries.

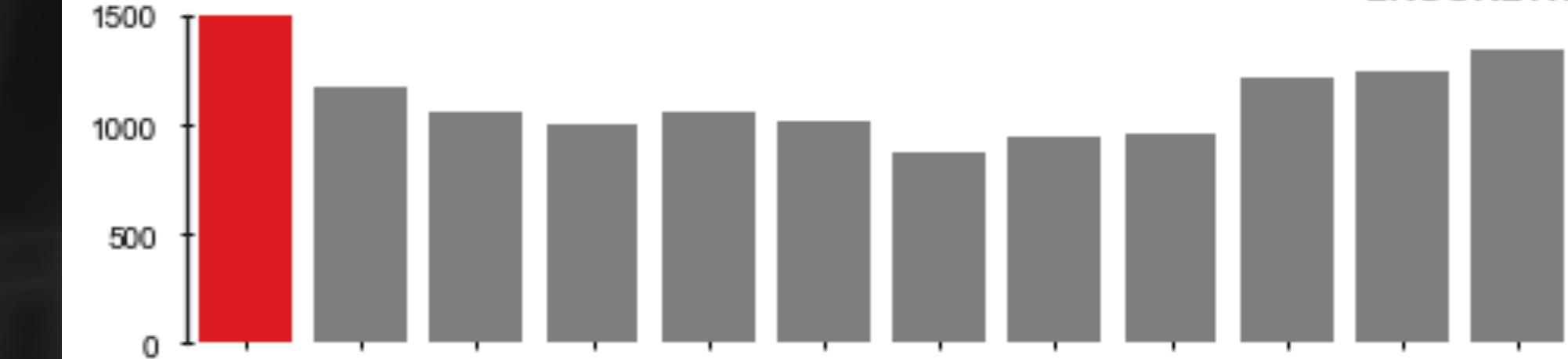
--- Pedestrians

Number of Pedestrians Injuries in Each Boro

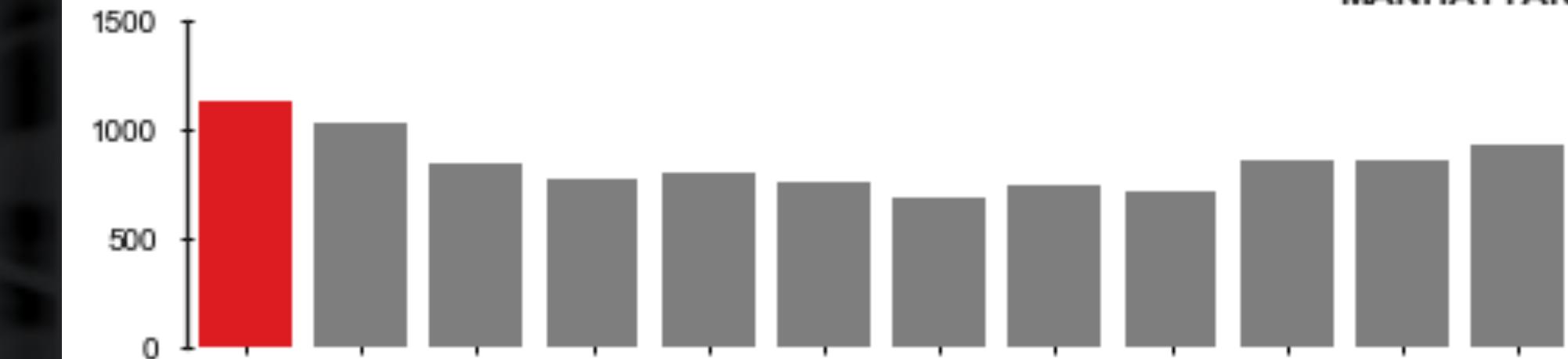
BRONX



BROOKLYN



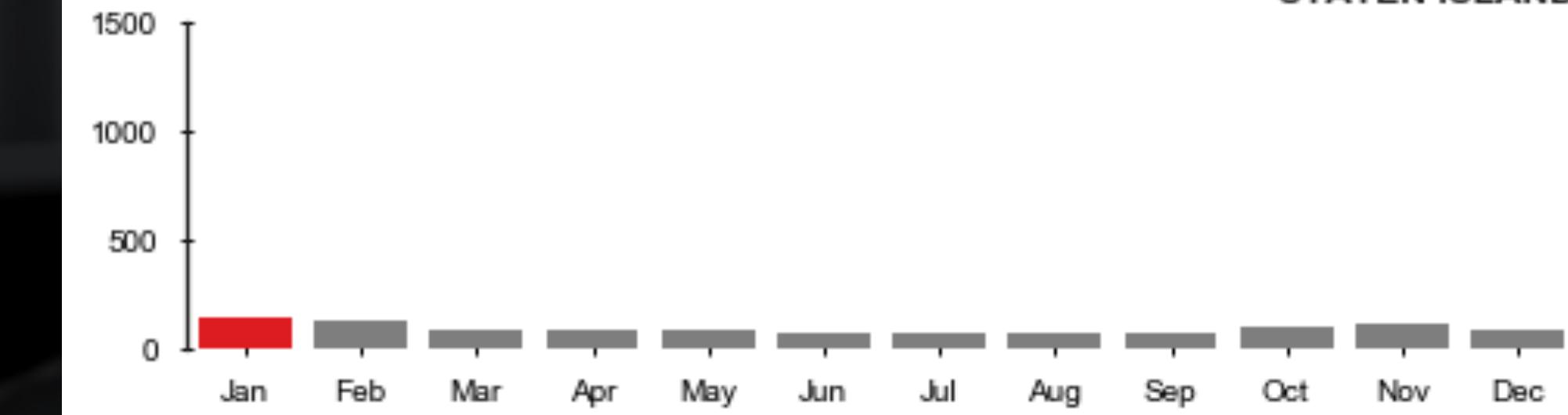
MANHATTAN



QUEENS



STATEN ISLAND

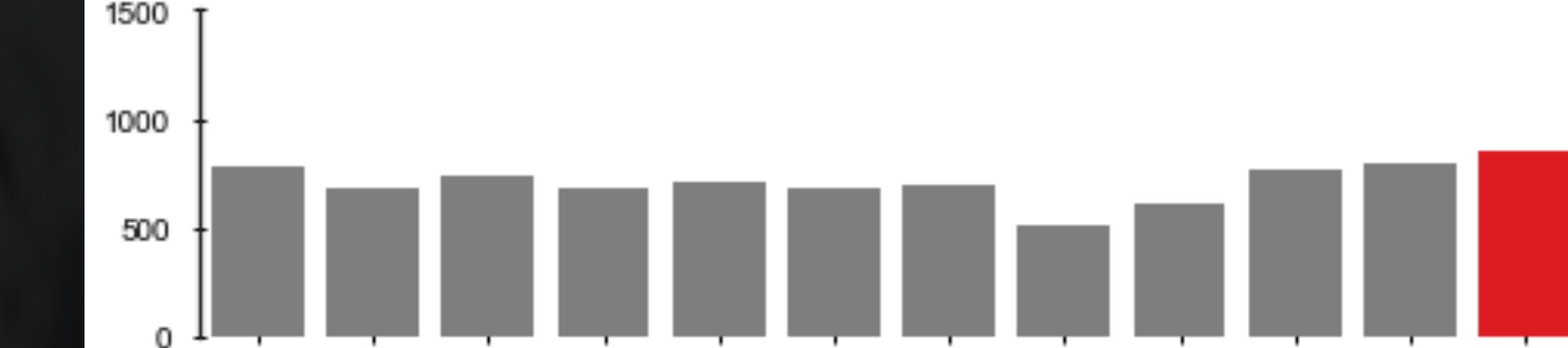


Number of Injuries.

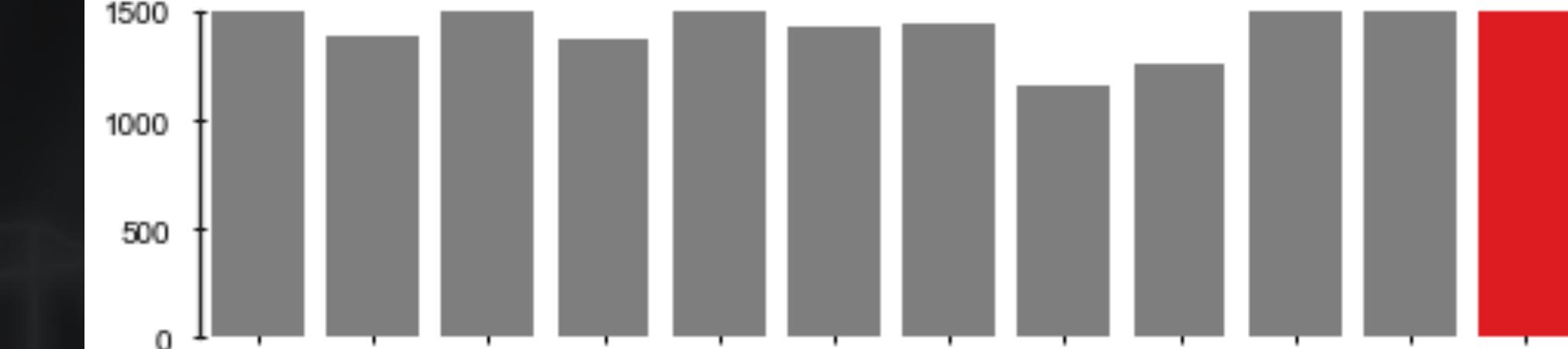
--- Cyclists

Number of Cyclists Injuries in Each Boro

BRONX



BROOKLYN



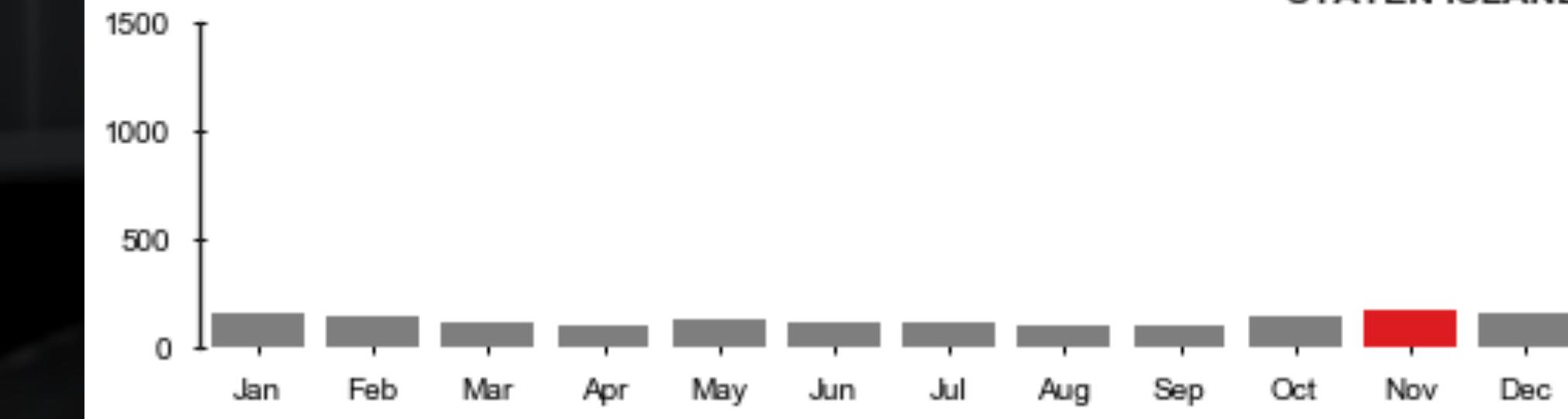
MANHATTAN



QUEENS



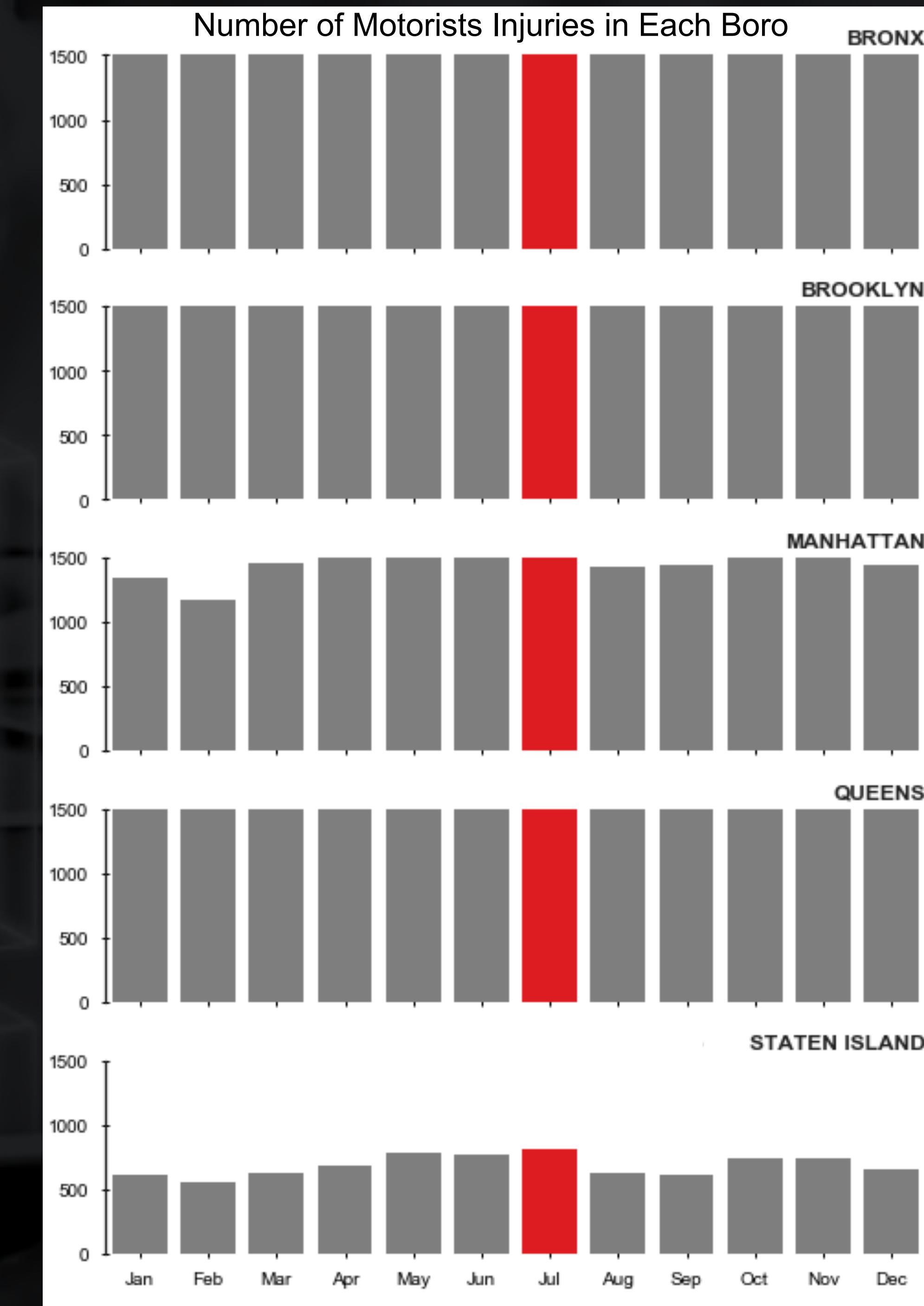
STATEN ISLAND



Number of Injuries.

--- Motorists

- The amount of injuries number remain quite constant throughout the year.
- Relatively higher amount of injuries than that of pedestrians and cyclists.
- We can suggest that motor is a more dangerous vehicle type.



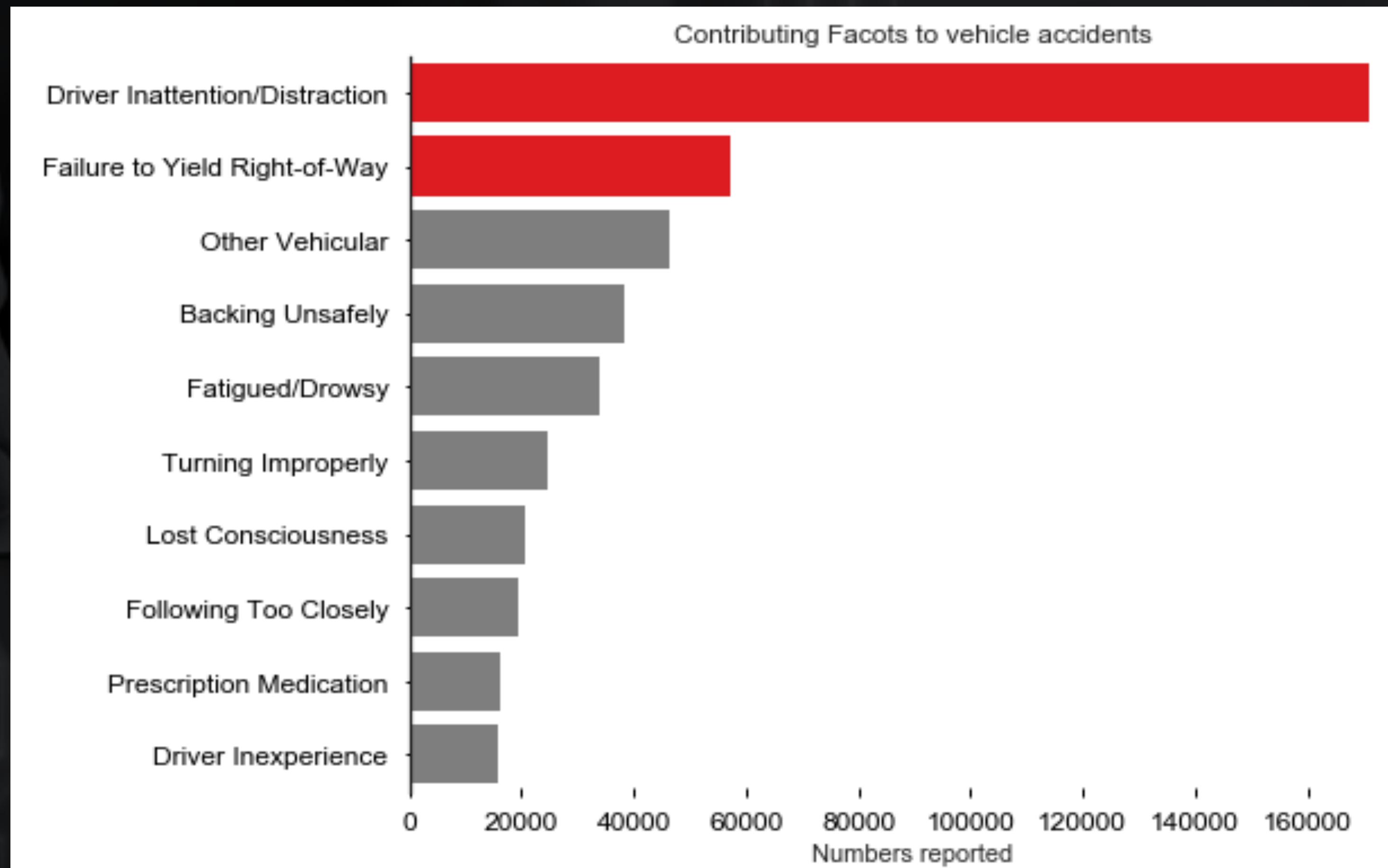


Contributing Factors

After exploring the numbers of all injuries during those collisions, we want to dig in more to find what causes all those injuries.

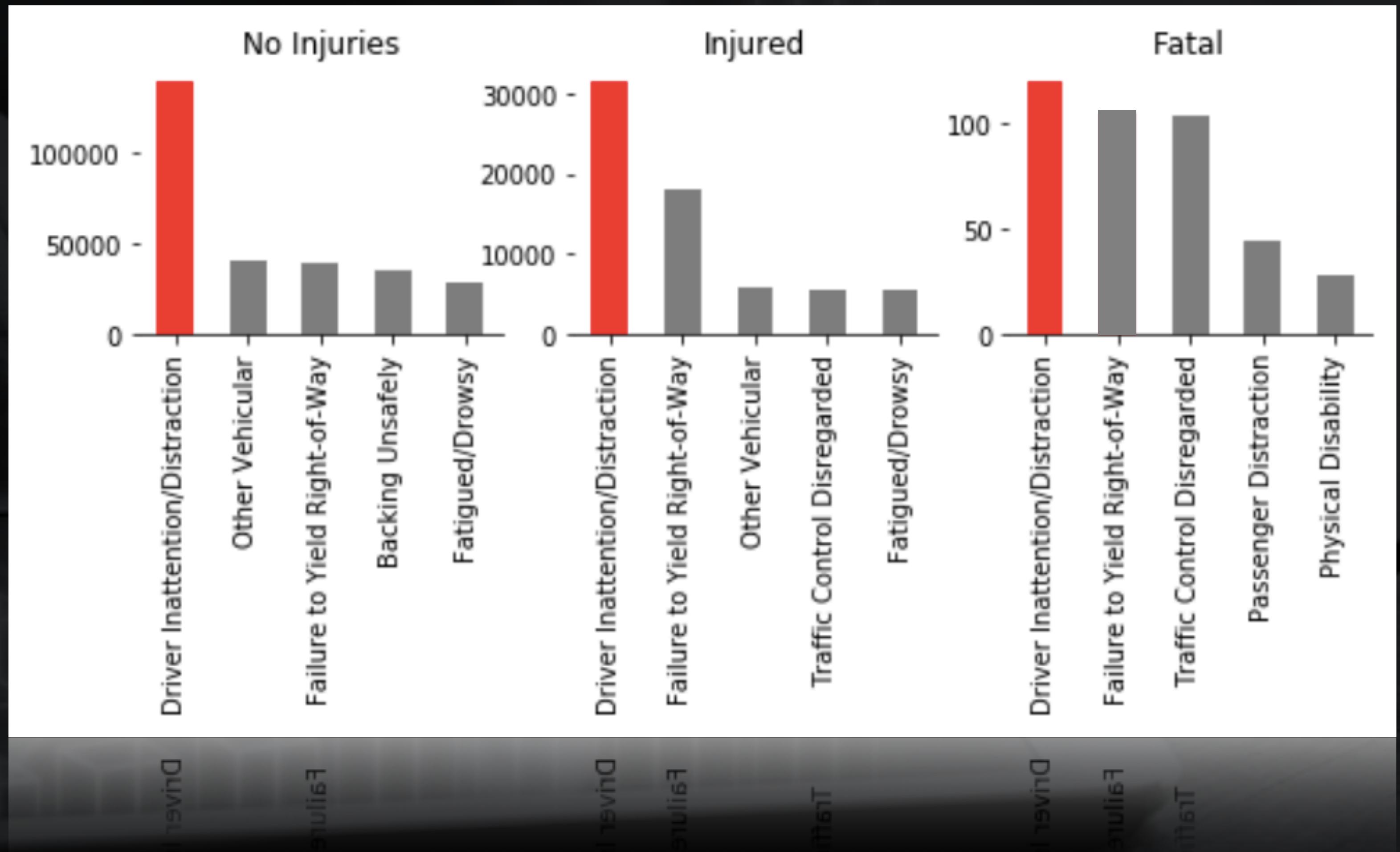


Contributing Factors





Contributing Factors

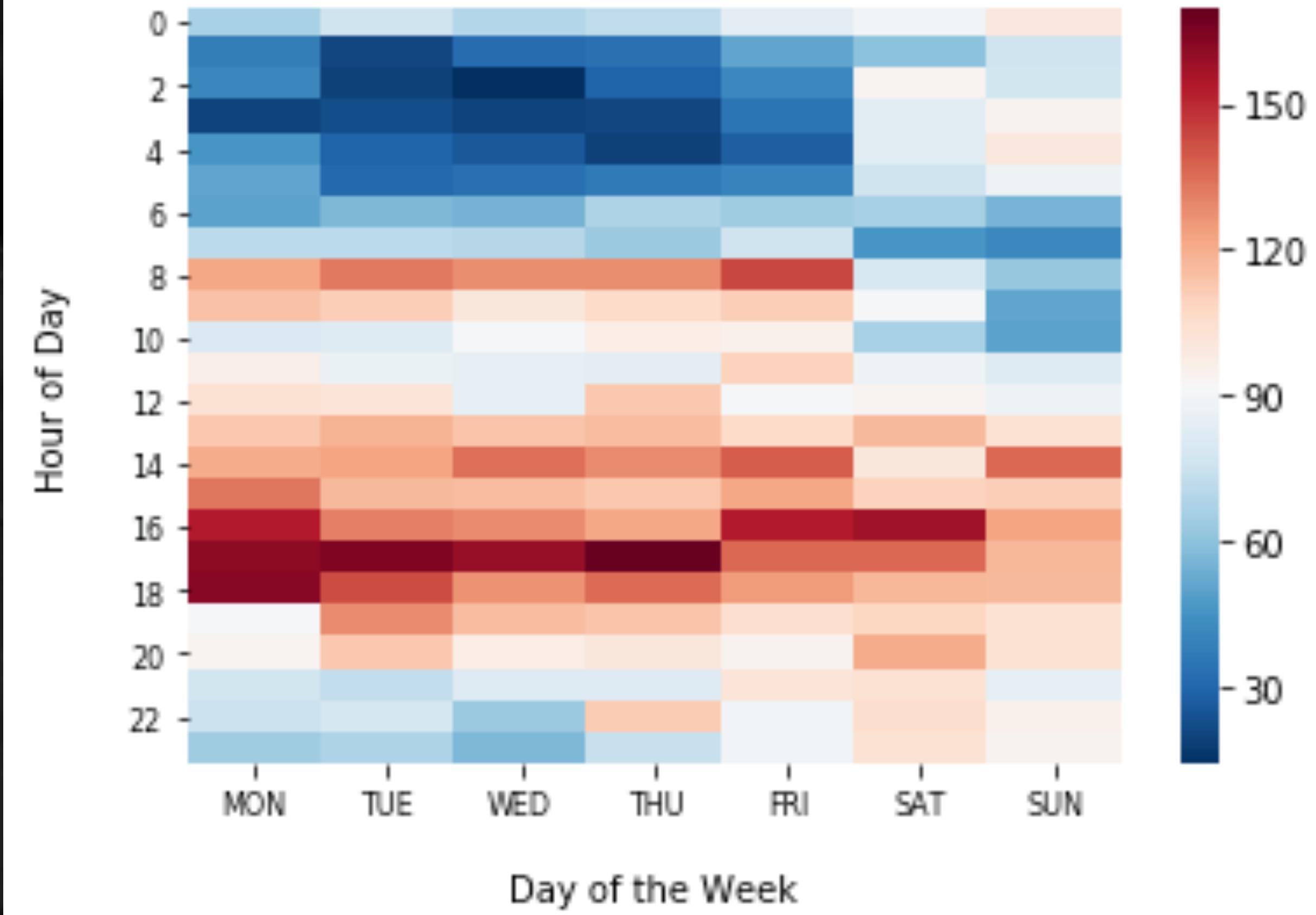




Contributing Factors

---Traffic Control Disregarded

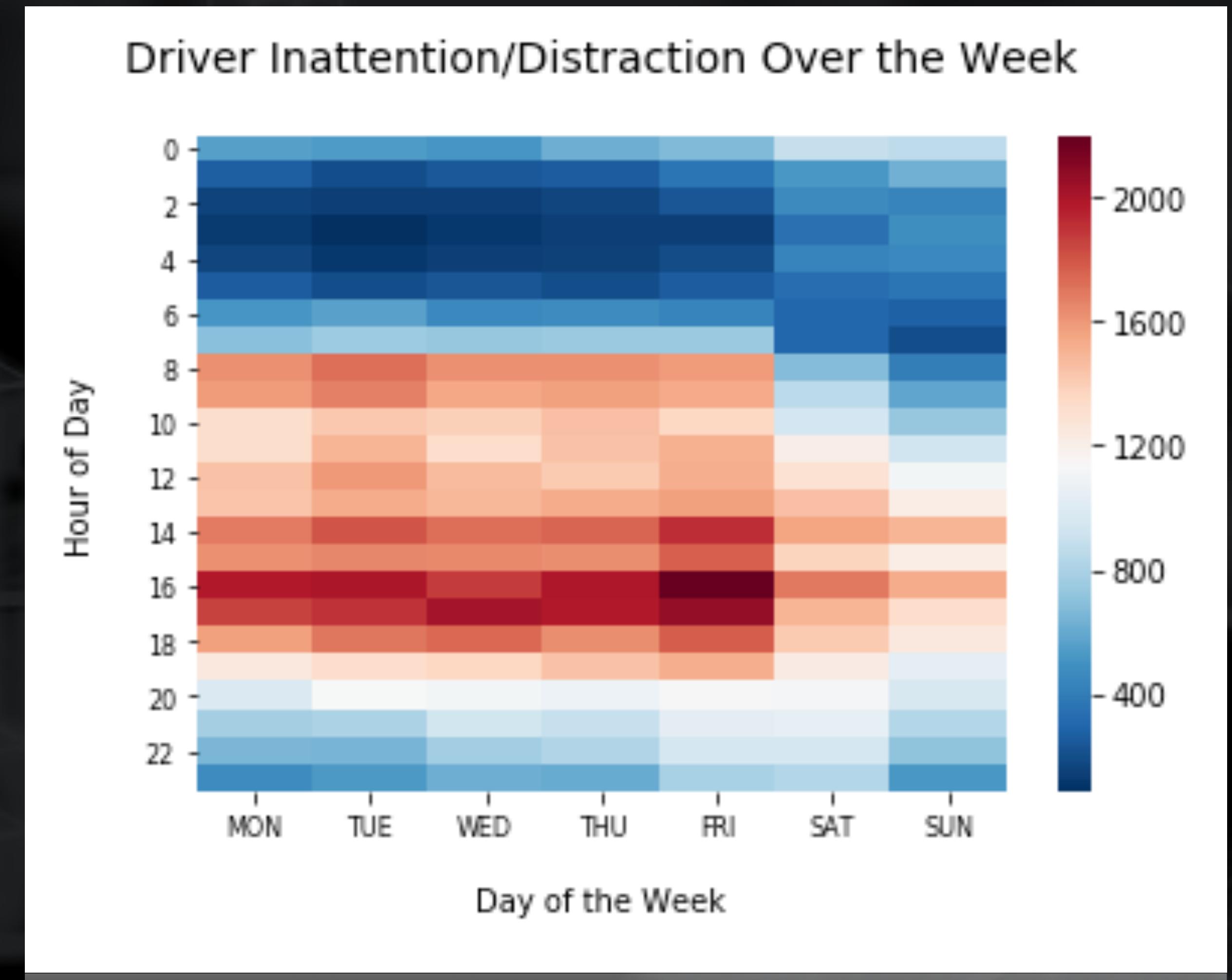
Traffic Control Disregarded Over the Week



Contributing Factors



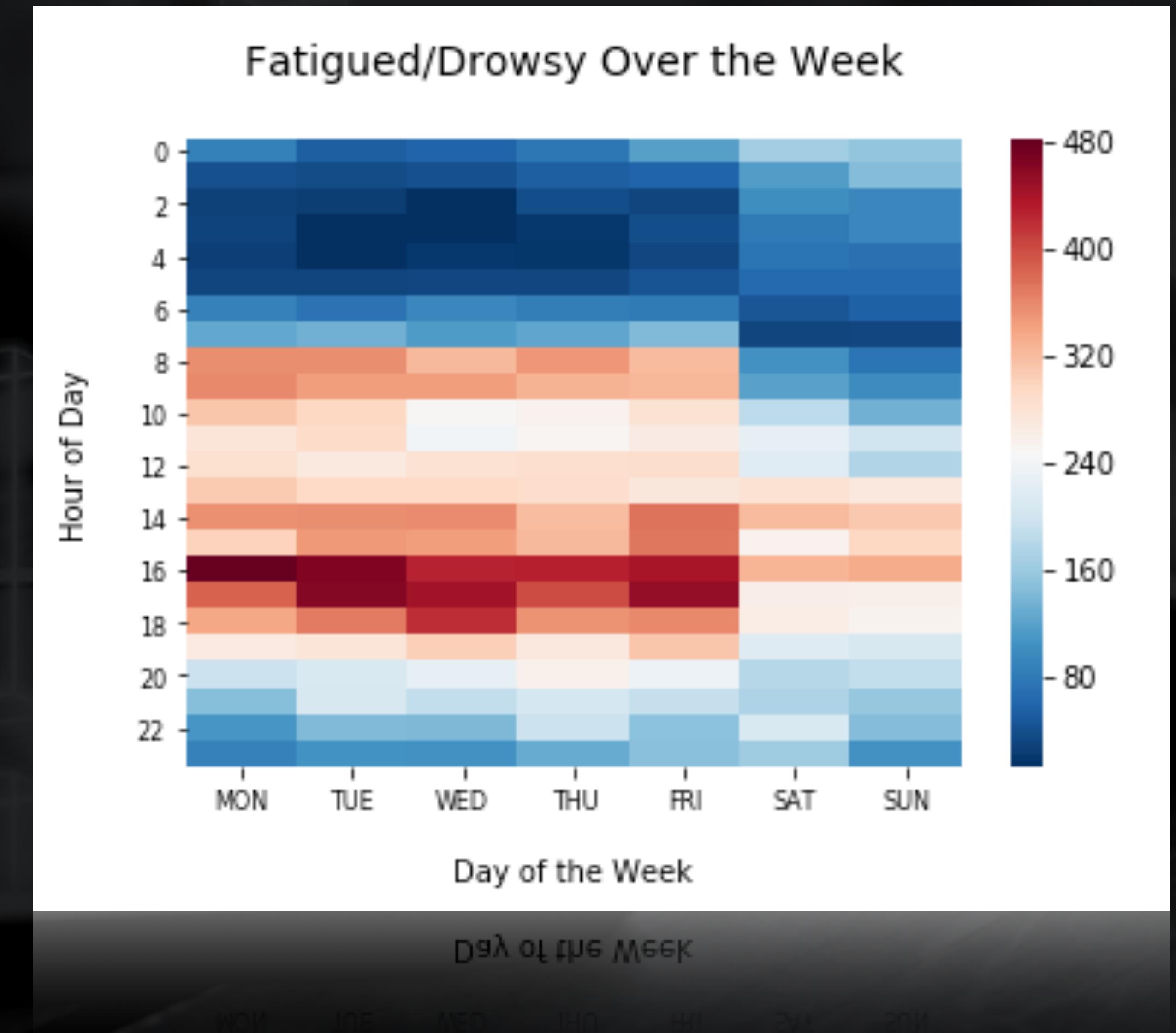
---Driver Inattention/Distraction



Contributing Factors



---Fatigued/Drowsy





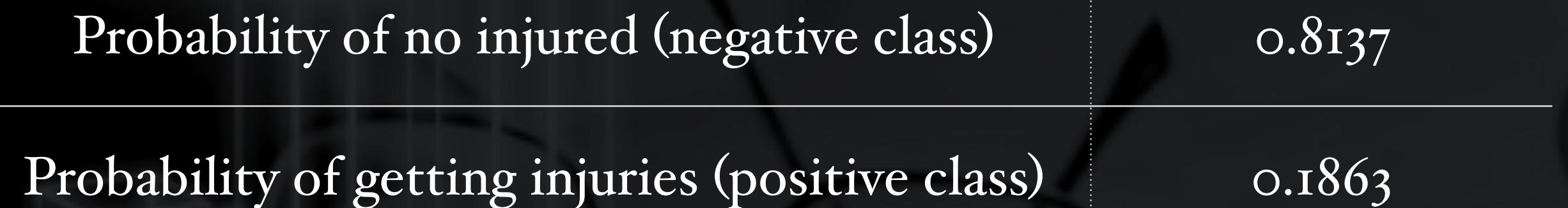
Models.

In order to find the significant features that causes potential injuries, we have tried several methodologies for instance, logistic regression, classification tree and naive bayes.



Naive Bayes.

- Training set accuracy: 0.8179
- Test set accuracy: 0.8204
- Probability of getting injure:



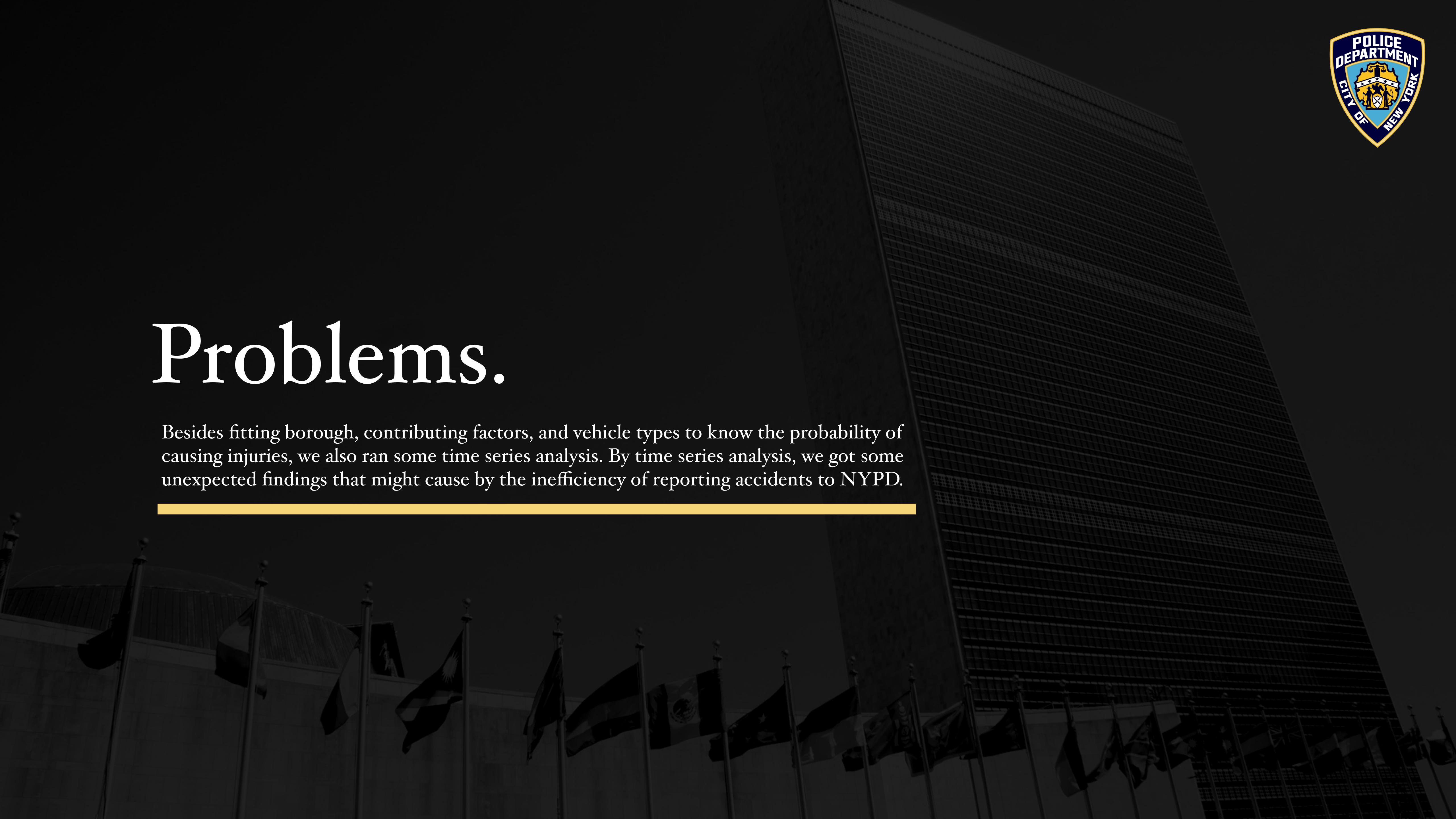
- Features Importances:

Q("VEHICLE TYPE CODE 1_Motorbike")	3.348736
Q("VEHICLE TYPE CODE 1_Bike")	2.965010
Q("VEHICLE TYPE CODE 1_Multi-Wheeled Vehicle")	2.863228
Q("VEHICLE TYPE CODE 1_BICYCLE")	2.835101
Q("VEHICLE TYPE CODE 1_Minibike")	2.575546
Q("VEHICLE TYPE CODE 1_tow t")	2.575546
Q("CONTRIBUTING FACTOR VEHICLE 1_Pedestrian/Bicyclist/Other Pedestrian Error/Confusion")	2.273467
Q("VEHICLE TYPE CODE 1_mta b")	2.170081
Q("VEHICLE TYPE CODE 1_nyc a")	2.170081
Q("VEHICLE TYPE CODE 1_Motorscooter")	2.170081
δ("VEHICLE TYPE CODE 1_Motorscooter")	5.710081
δ("VEHICLE TYPE CODE 1_nyc a")	5.710081
δ("VEHICLE TYPE CODE 1_Motorscooter")	5.710081



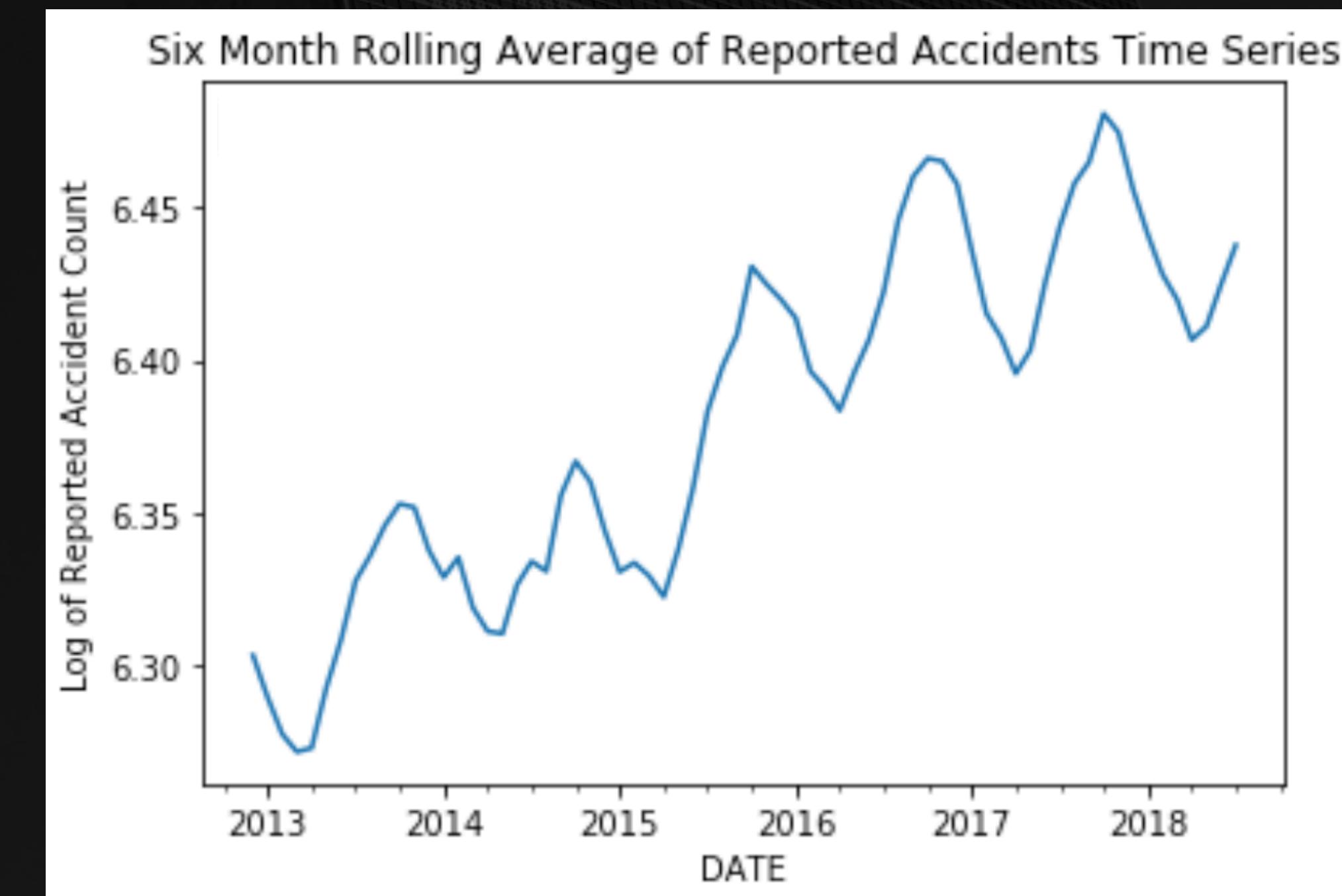
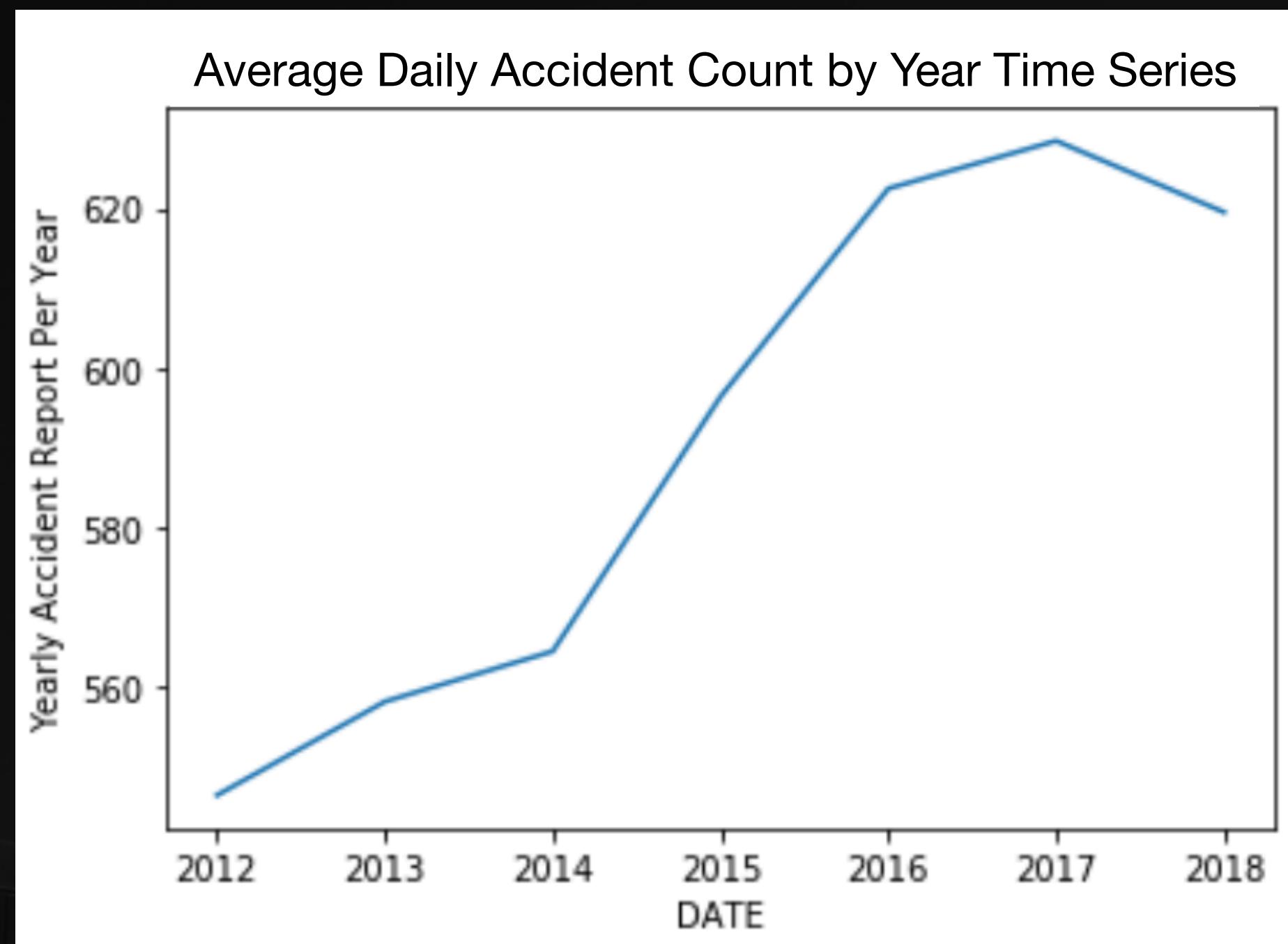
Problems.

Besides fitting borough, contributing factors, and vehicle types to know the probability of causing injuries, we also ran some time series analysis. By time series analysis, we got some unexpected findings that might cause by the inefficiency of reporting accidents to NYPD.





Accidents in NYC.



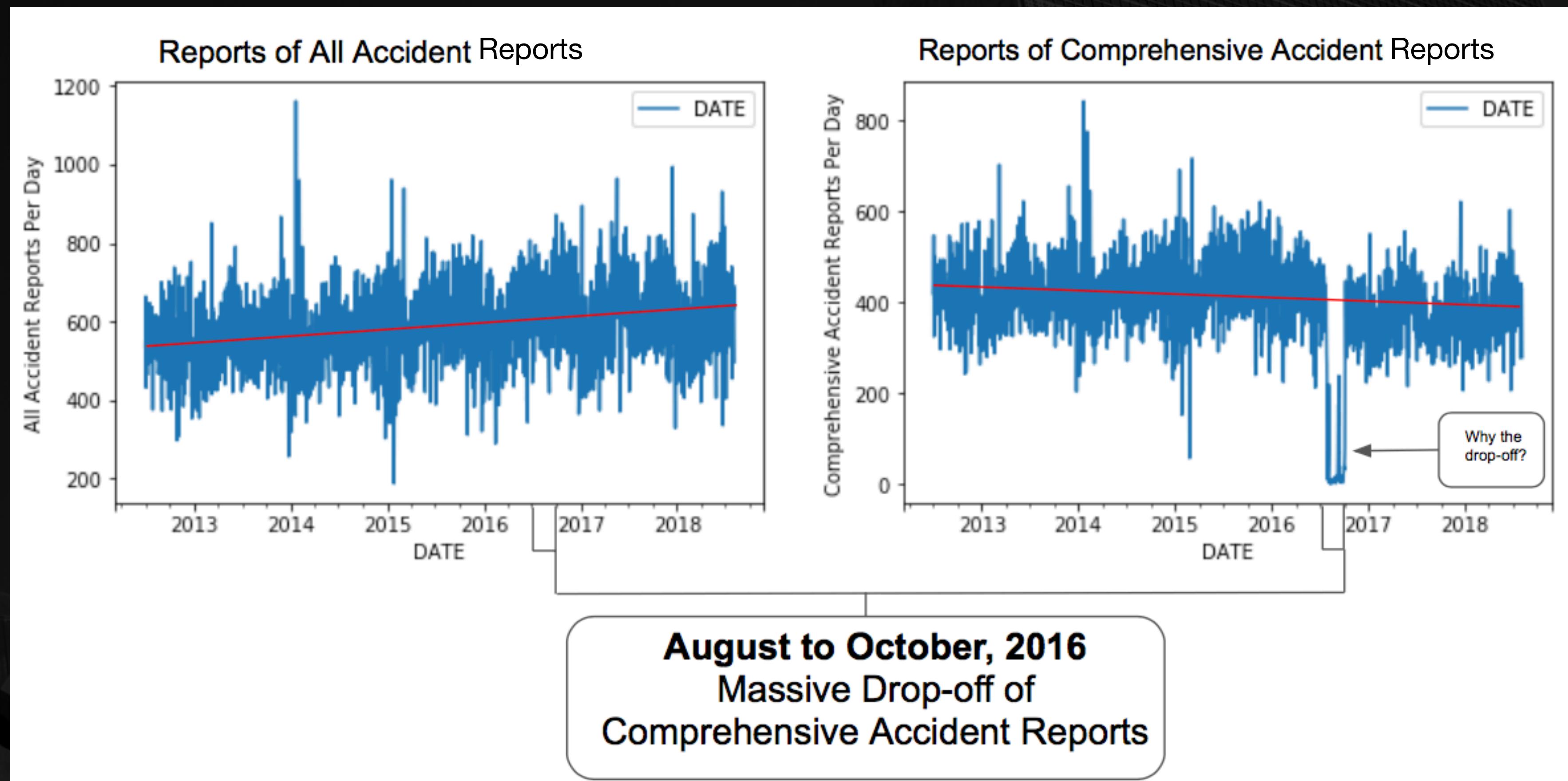
Note: NYC Pop increased by ~5% since 2010.

Source: <https://www1.nyc.gov/site/planning/data-maps/nyc-population/current-future-populations.page>

Rise in Traffic since 2016



Strange Anomaly in 2016.





Increased Traffic Congestion.

- NYPD and other emergency vehicles are slow to respond due to increased traffic congestion
- During Aug-Oct 2016, ambulances are 4.5 times less likely to be involved in accidents due to possibly being stuck in traffic.
- Despite an increasing number of accidents, many reports lack essential detail. Many only record that a person was injured or that an accident happened.

NYC Traffic Congestion Causing Big Problems For Emergency Responders

August 23, 2016 at 6:30 pm Filed Under: Emergency Responders, FDNY, Hazel Sanchez, traffic congestion

Source: <https://newyork.cbslocal.com/2016/08/23/emergency-responders-traffic-congestion/>

“ Average speeds in Manhattan are down to 8.2 mph, even with 45,000 fewer vehicles a day entering Midtown than in 2010.”
- NY Post in 2016

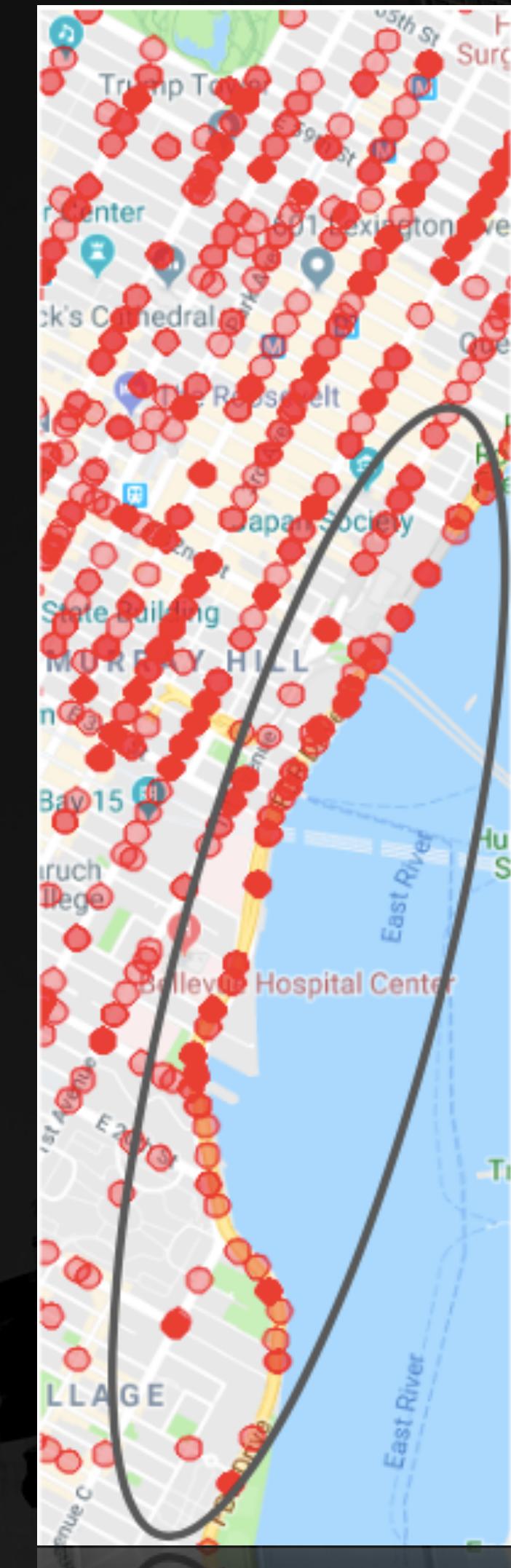


Road Closure: 71st UN General Assembly Session.

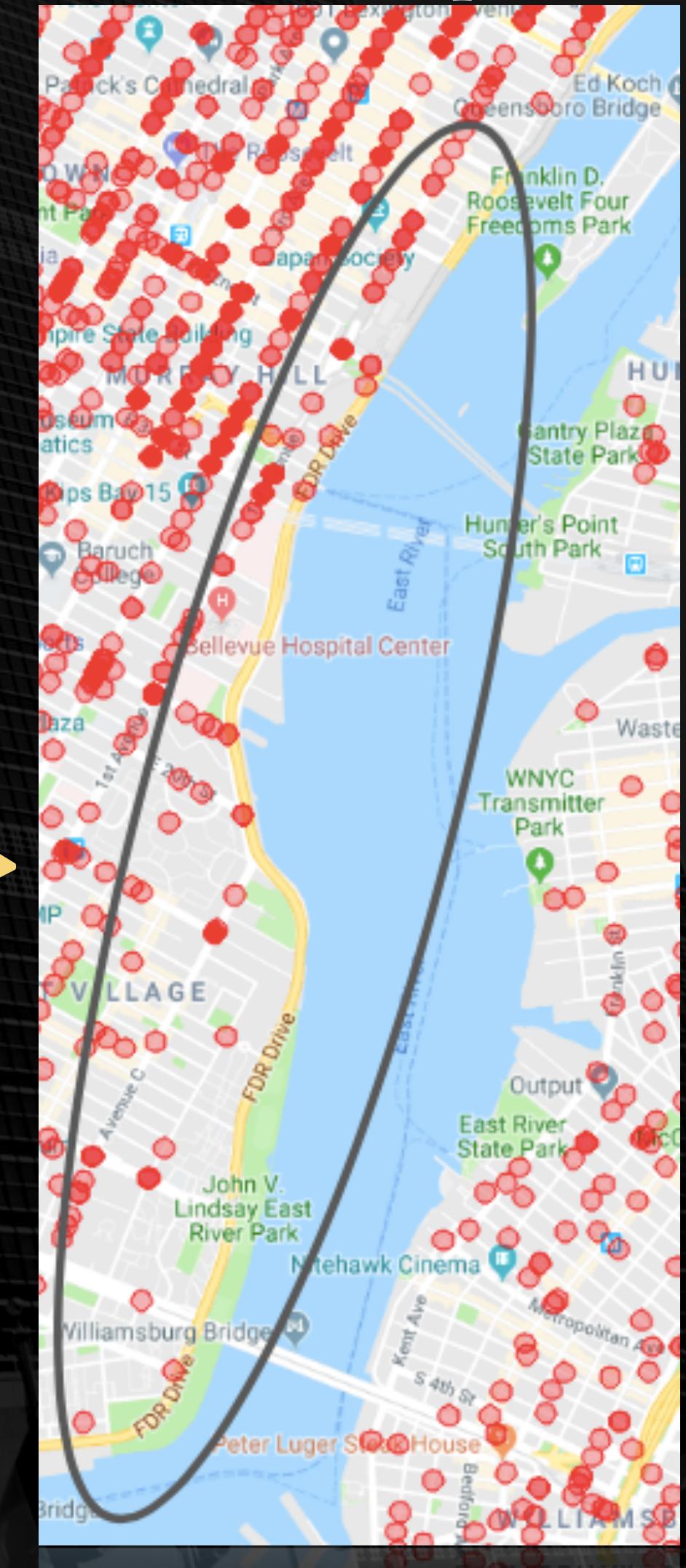
- Starts in September 2016
- Localized effect: known to coincide with massive traffic jams in Midtown
- **FDR Drive**, depicted on the right, is the main road neighboring the **UN Headquarters**



Locations of All
Accident Reports



Locations of Only
Comprehensive
Accident Reports



Source: <http://nypdnews.com/2016/09/un-general-assembly-traffic-advisory/>



Road Closure: Construction and Change.

How Bloomberg and de Blasio made traffic even worse

By Post Editorial Board

December 4, 2016 | 6:52pm | Updated

“ The traffic is being engineered, a former top NYPD official told The Post, explaining a long-term plan that began under Mayor Mike Bloomberg and hasn’t slowed with Mayor de Blasio.

“ The city streets are being engineered to create traffic congestion, to slow traffic down, to favor bikers and pedestrians, the former official said.

“ There’s a reduction in capacity through the introduction of bike lanes and streets and lanes being closed down.

- NY Post in 2016

- Effects of traffic congestion can be seen in enforced road closures.
- In places where roads are periodically being closed, there is a dramatic decrease in comprehensive accident reports.
- Correlated with rise in accidents?



Road Closure: Belt Parkway Construction.

Locations of All
Accident Reports

Locations of Only
Comprehensive
Accident Reports

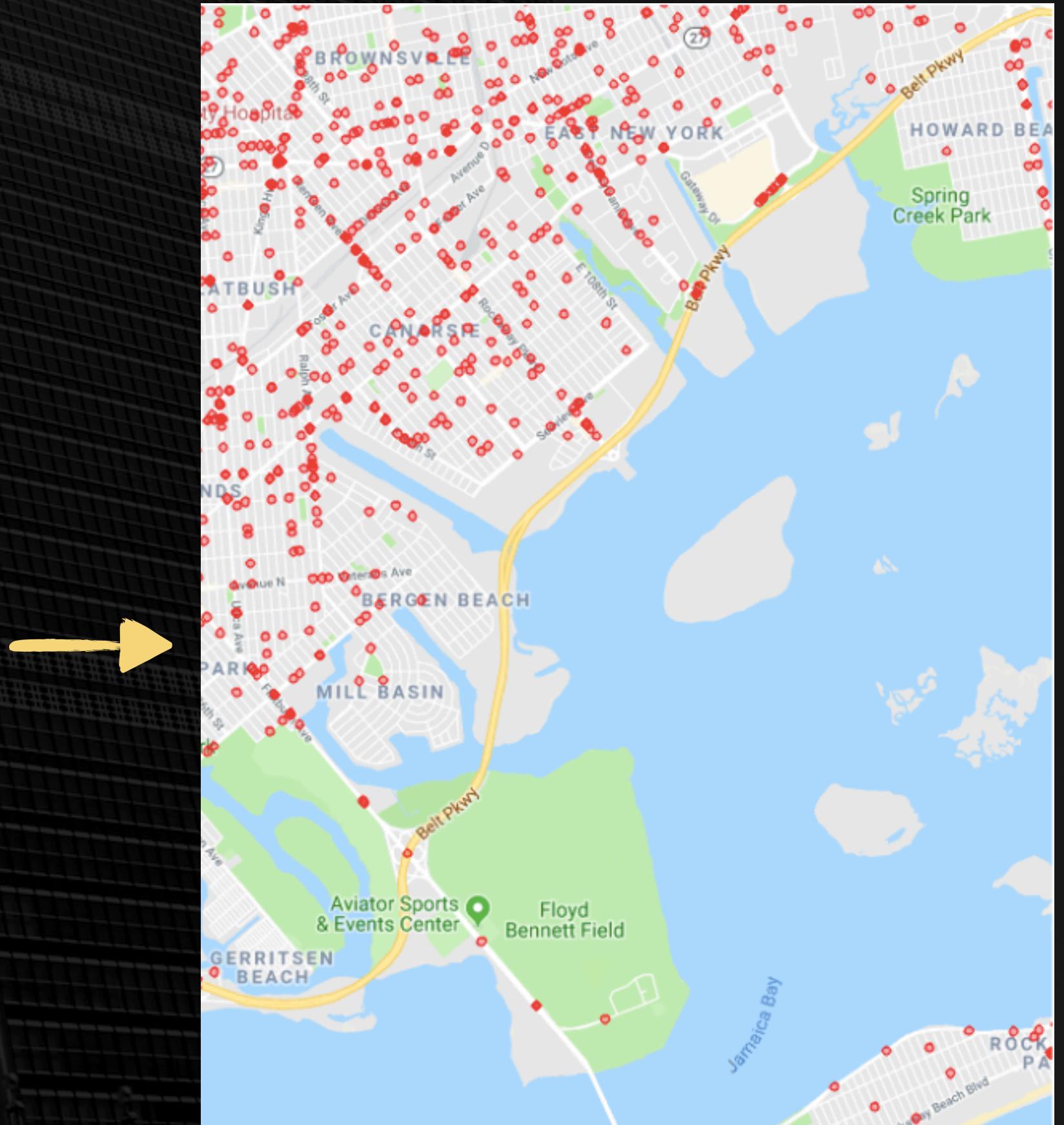
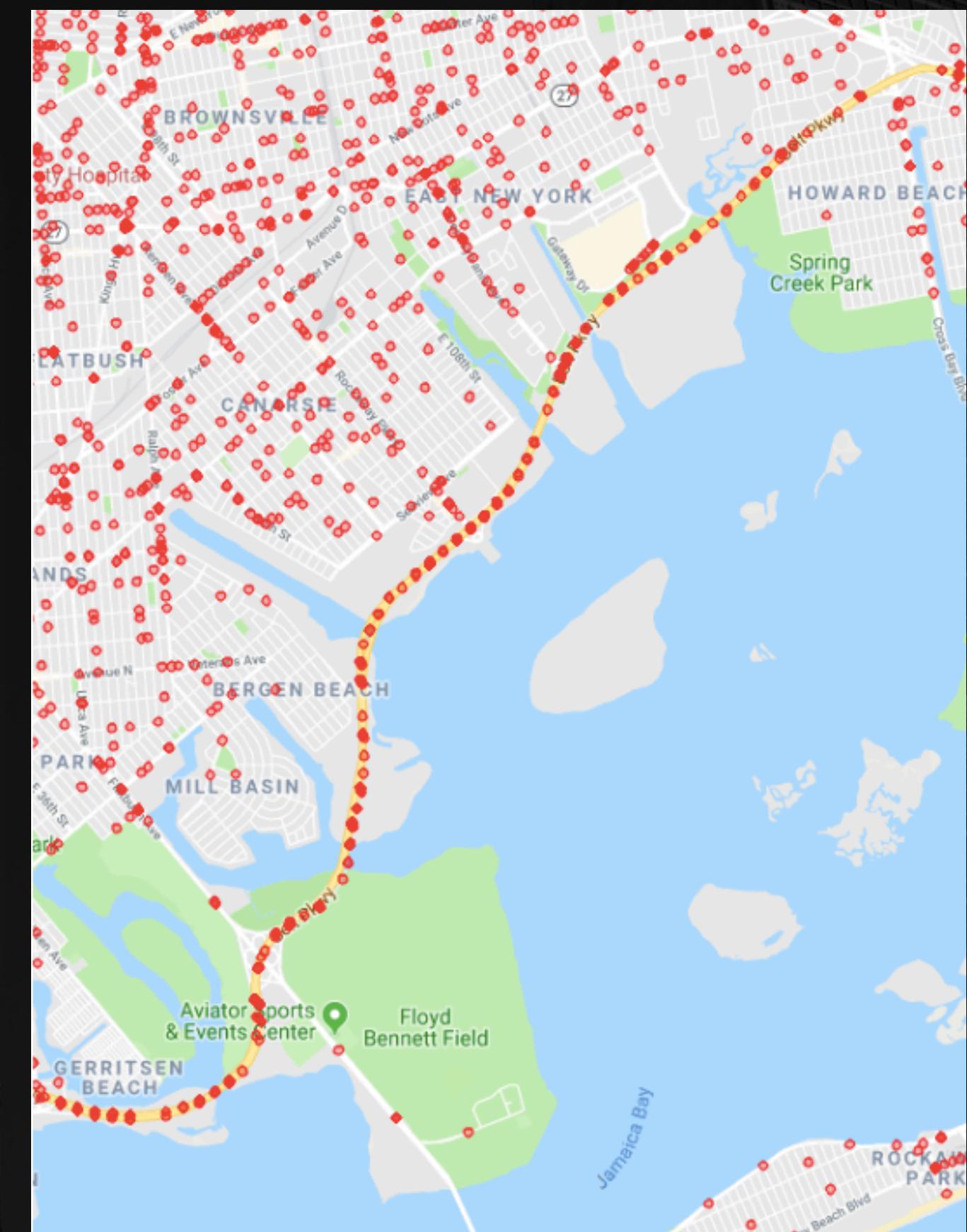
AUGUST 26, 2016 / NEWS / TRANSIT ISSUES

Building bridges: Delayed Belt Parkway spans are back on track



Department of Transportation

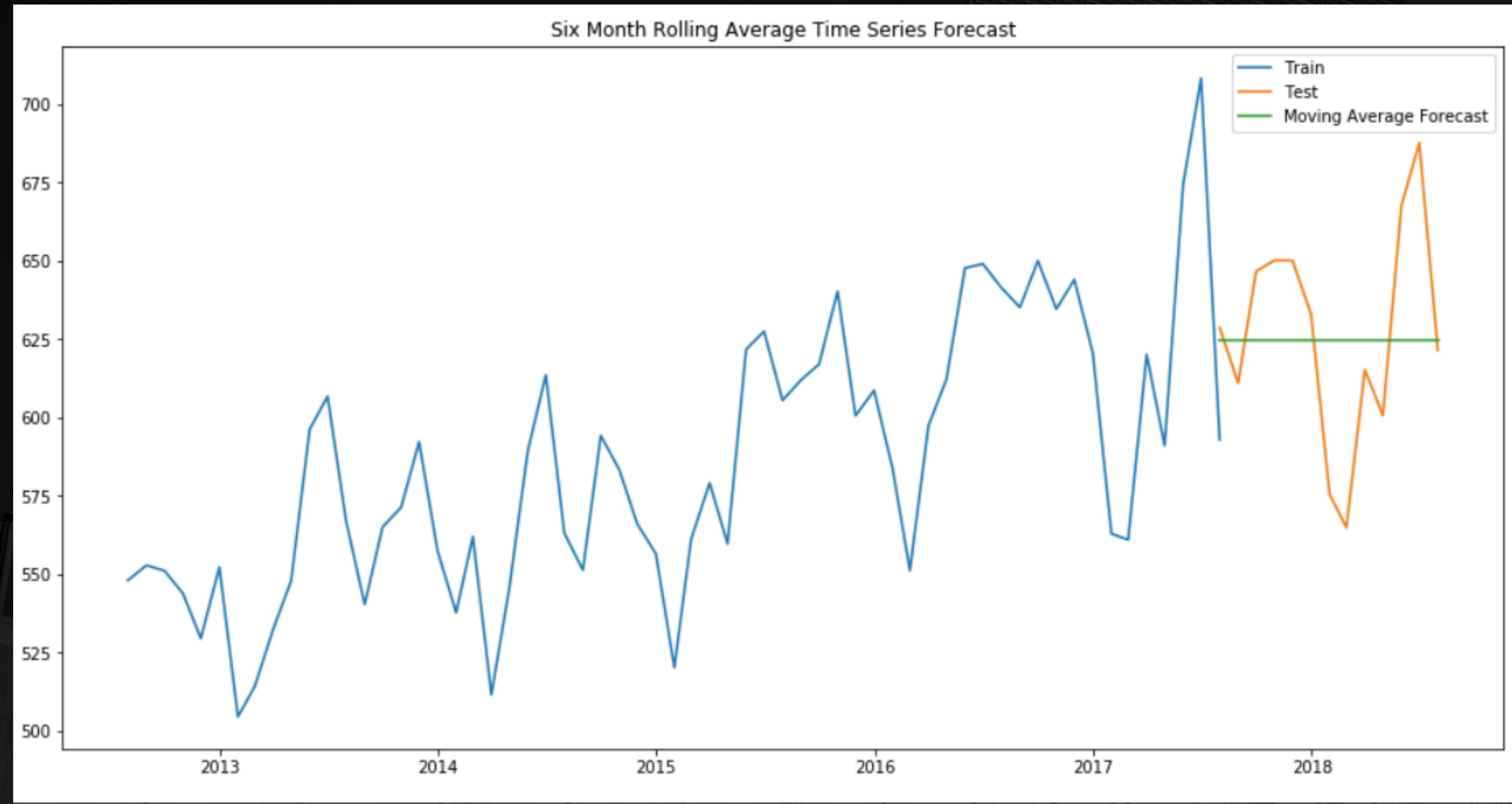
Slowly but surely: Workers are busy building the Mill Basin Drawbridge over Jamaica Bay, which is supposed to be completed in 2021.



Source: <https://www.brooklyndaily.com/stories/2016/36/mm-belt-parkway-bridges-2016-08-26-bk.html>



Time Series Forecast.





Insights.

After thoroughly exploring the data, fitting model with potential contributing factors and explaining some interesting finding of this dataset, we hope to provide some suggestions and insights to solve our problem sets.



Insights

Safer city to live in



Insurance pricing correctness



Importance of better Planning





Thank You.

Presented by
MSBA Class of 2019 Data Analysis Programming Group 3
Tiffany Sung, Won Lee, Chelsea Matthews