

# NYPD Motor Vehicle Collisions Analysis

---

Ivy Pan

# Problem Statement

How can NYC leverage motor vehicle collision data to efficiently allocate community resources as well as create new business opportunities?

# Table of Contents

1. Collisions Map
2. Collision Trends Over Time
3. Borough Trends Over Time
4. Causes of Accidents
5. Vehicles in Accidents
6. Time Characteristics of Collisions and Fatalities
7. Zip Code Analysis
8. Conclusion

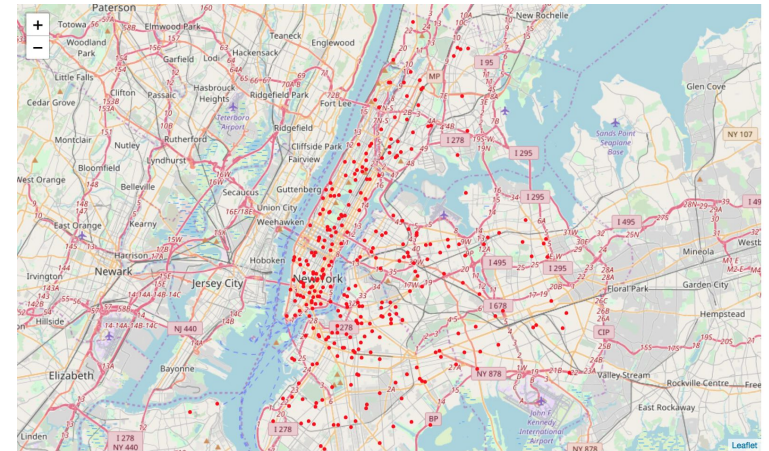
# Collisions Map

Higher concentration of collisions occur in geographically dense neighborhoods

NYC September 2017

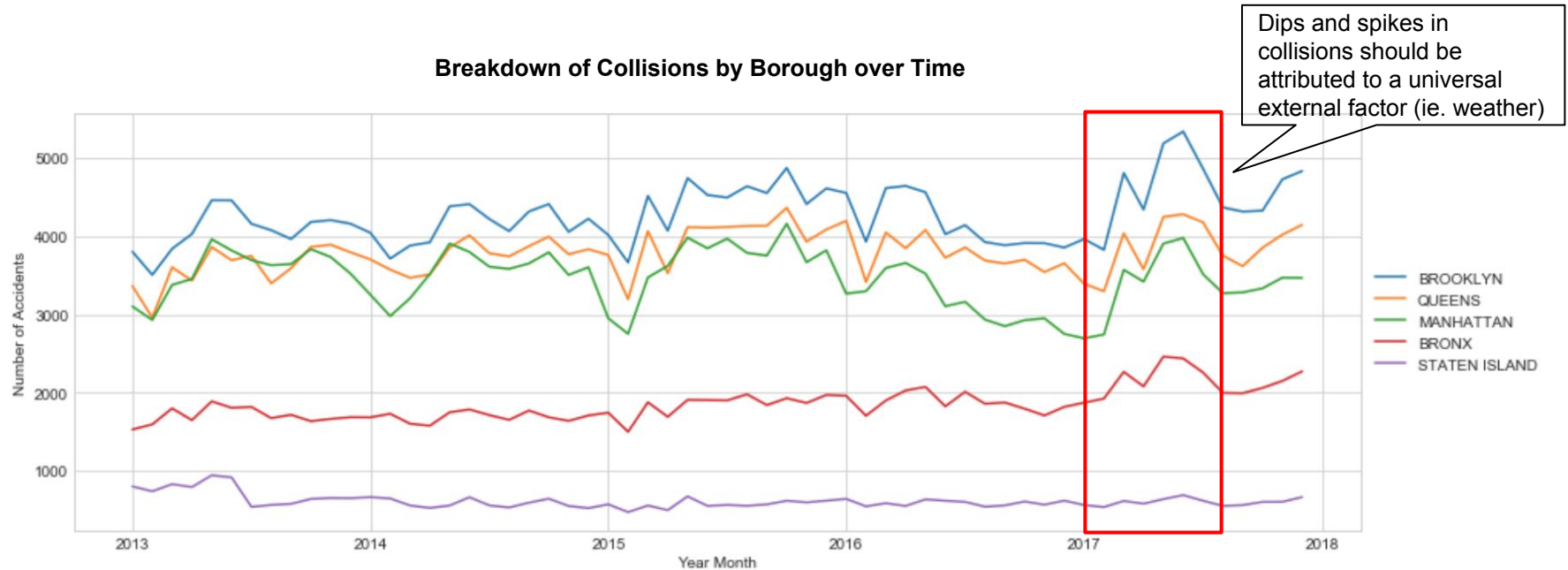


NYC September 2017 - Bicycle Fatalities



# Collision Over the Years

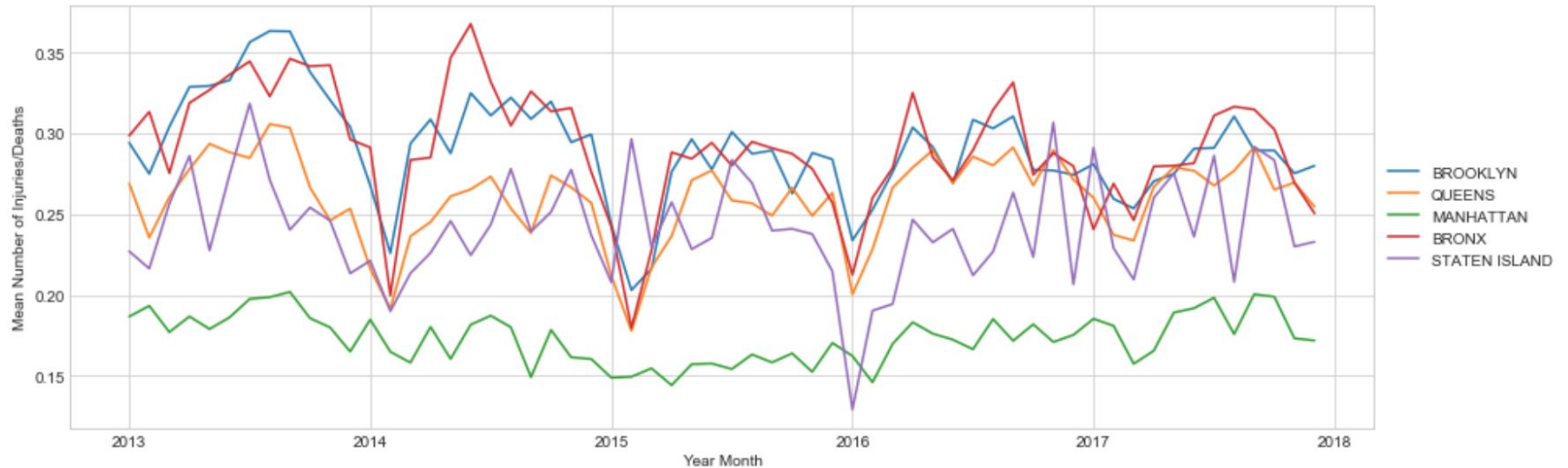
Although the volume of collisions vary by borough, the trends remain relatively constant throughout the boroughs as expected



# Higher Volume of Collisions Does Not Lead to Higher Fatalities

While Manhattan has a greater relative volume of collisions, the mean number of injuries/deaths is the lowest, whereas the Bronx exhibits the opposite characteristics

**Breakdown of Mean Number of Fatalities by Borough over Time**

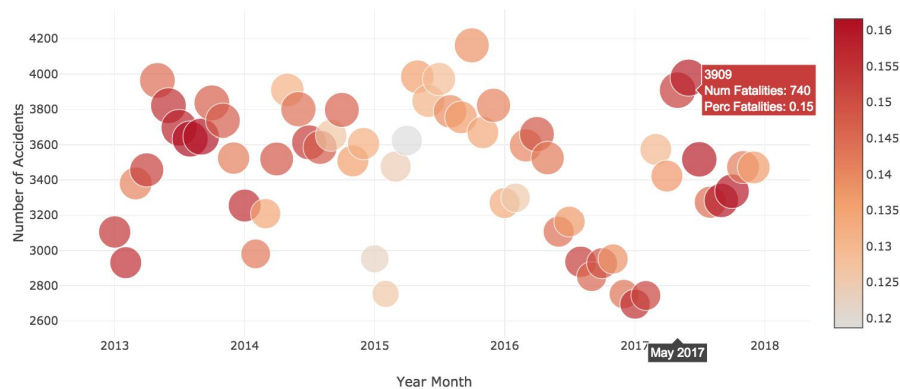


\*\*\* Correlation Coefficient: -0.008

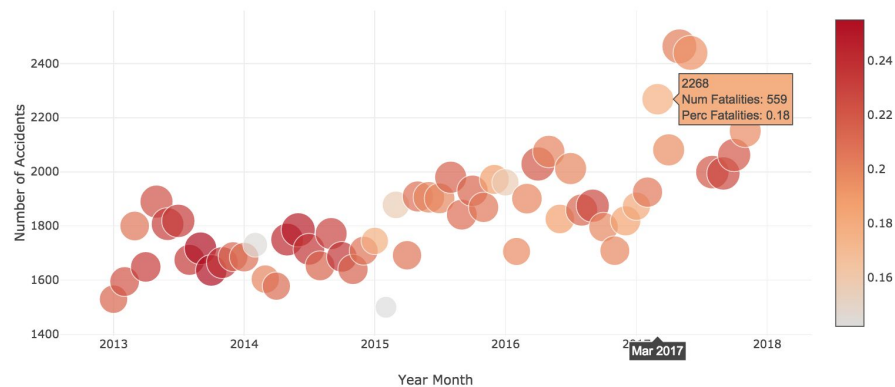
# Deep Dive Into Borough Trends Over Time

A closer look into the different factors of collisions including volume, fatality rate, and number of fatalities over time gives us a sense of how the borough is evolving and changing

MANHATTAN Accident Trends Over Time



BRONX Accident Trends Over Time

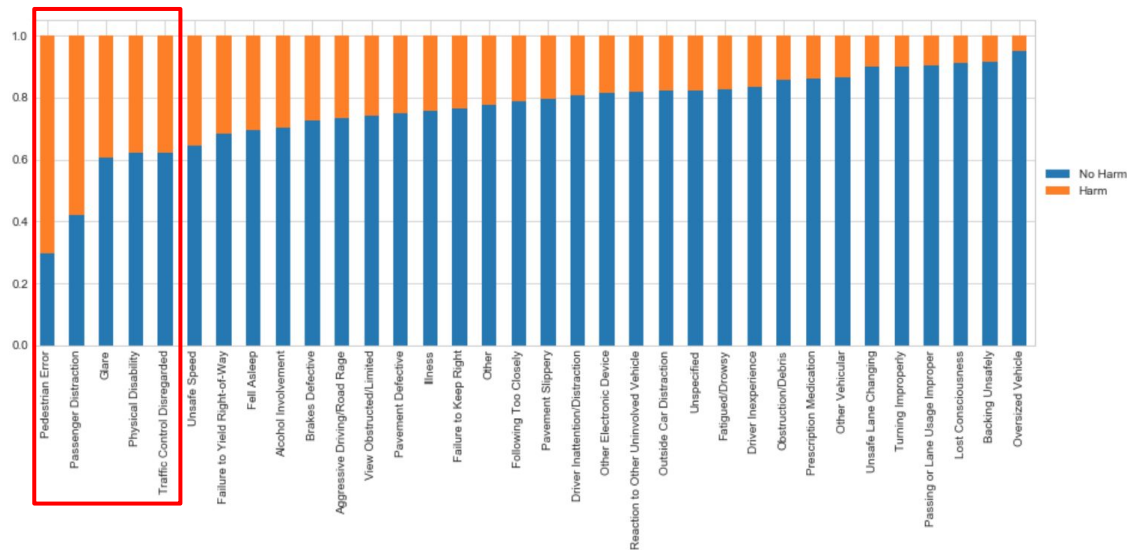


- While Manhattan has a lower average fatality rate than the Bronx (0.14 vs 0.2) while maintaining a constant volume of collisions over time, the Bronx's fatality rate is actually decreasing (therefore consistently improving) over time while Manhattan's fluctuates
- The inverse behavior of the 2 boroughs can be attributed to the growing population and traffic of Manhattan suggesting that the borough could benefit from increased presence of traffic police and targeted ads against reckless driving

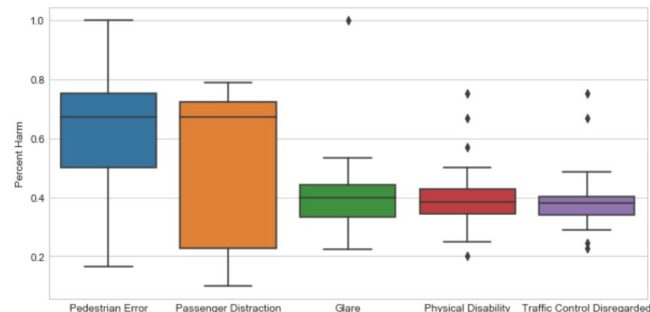
# Causes of Accidents

Three of the top causes with the highest fatalities rates can be attributed to negligence

Breakdown of Fatality Rate by Cause



Distribution of Fatality Rate of Top 5 Causes

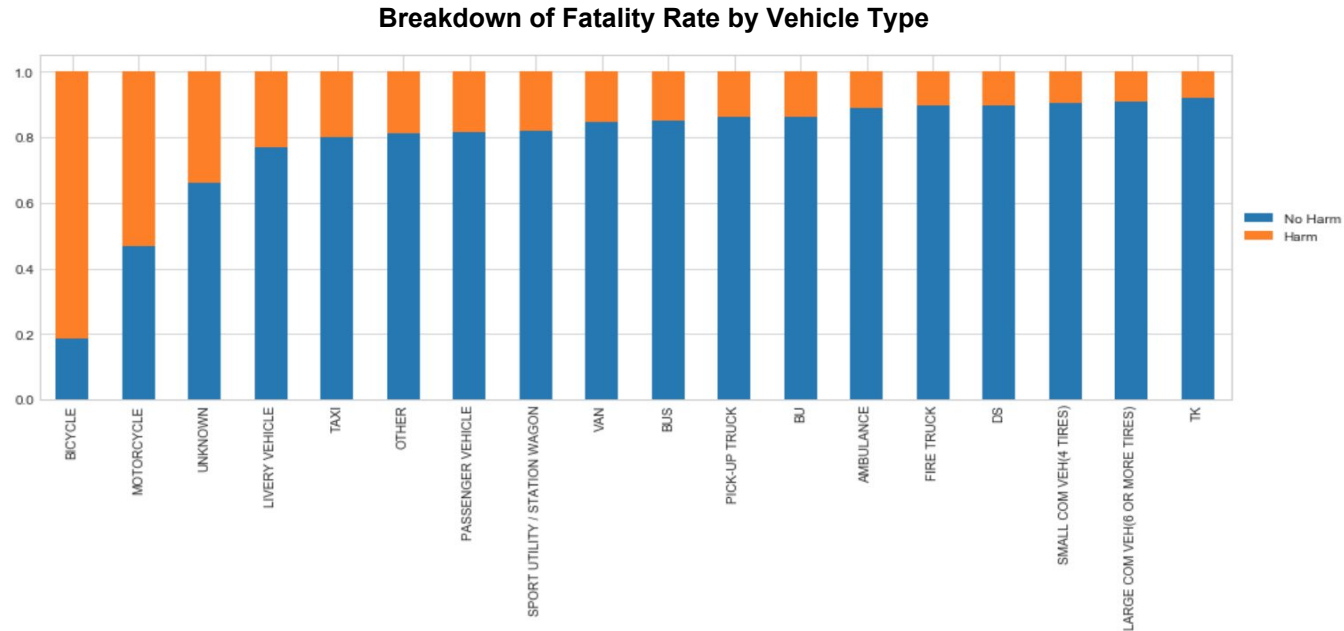


- Distribution of fatality rates are mostly normal, therefore validating the harmful effects of the top causes



# Vehicles in Accidents

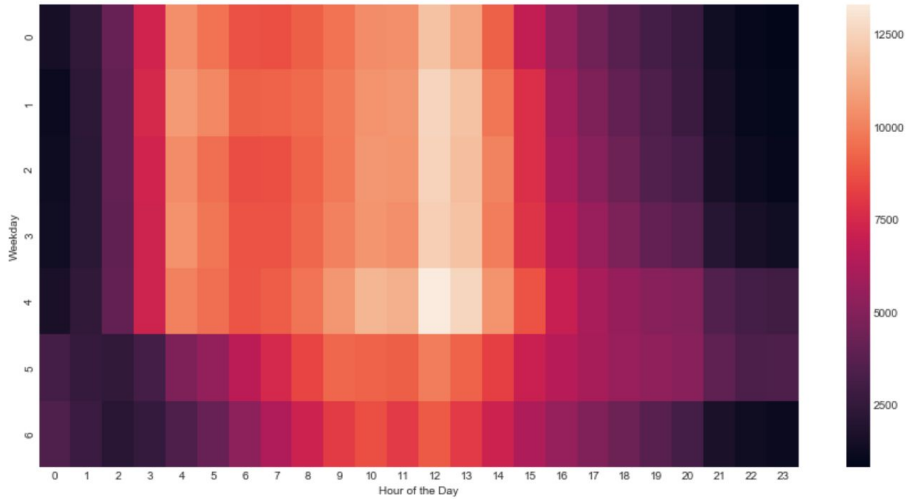
While the high fatality rates of bicycles and motorcycles are expected, transportation businesses (including taxi and delivery services) can save on insurance costs by investing in technologies to prevent accidents and enforcing stricter rules on the job



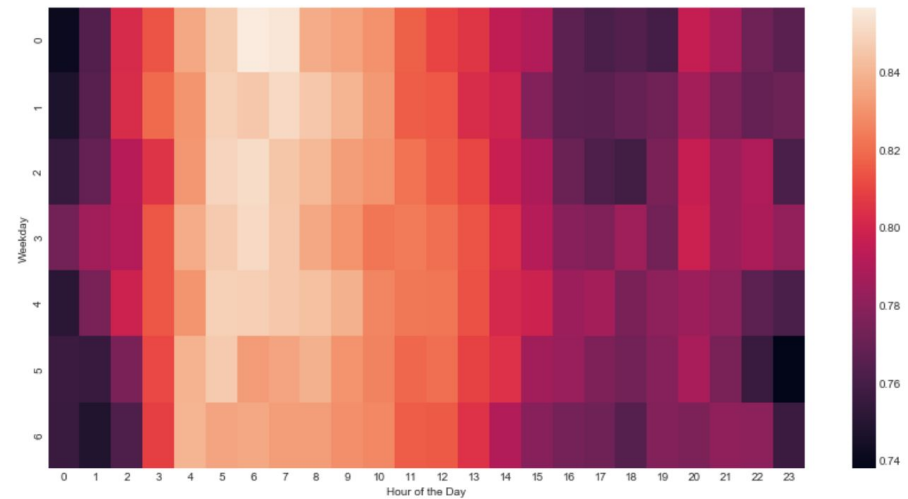
# Time Characteristics of Collisions and Fatalities

While the fatality rates are highest in the early mornings, volume of collisions are highest during the early to mid afternoon

**Volume of Collisions by Weekday and Hour**



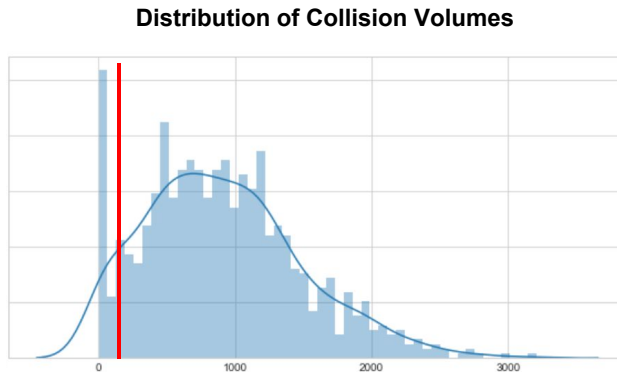
**Fatality Rate by Weekday and Hour**



- The time characteristics of collisions gives us insight into helping businesses staff members, particularly in the space of emergency services or auto repair workers

# Zip Code Analysis: Setup

- **Problem:** How can we estimate how neighborhoods are evolving by looking at motor collision trends over time?
- Methodology:
  - Investigate neighborhoods with the entire 5 years of data
  - Assumptions:
    - Demographics remain constant
    - Population growth is negligible
  - Remove bottom 10% of zip codes with the lowest collision volume
  - Calculate key metrics for each zip code
    - Collision, Fatality Rate
      - Mean, median, standard deviation
      - Momentum
      - Year over year growth



	statistic	volume
0	count	920.000000
1	mean	912.317391
2	std	576.524565
3	min	1.000000
4	25%	486.750000
5	50%	858.000000
6	75%	1243.000000
7	max	3206.000000

# Zip Code Analysis: Results

Analysis provides insight into neighborhoods that could use more resources to combat motor vehicle collisions as well as to deal with the aftereffects

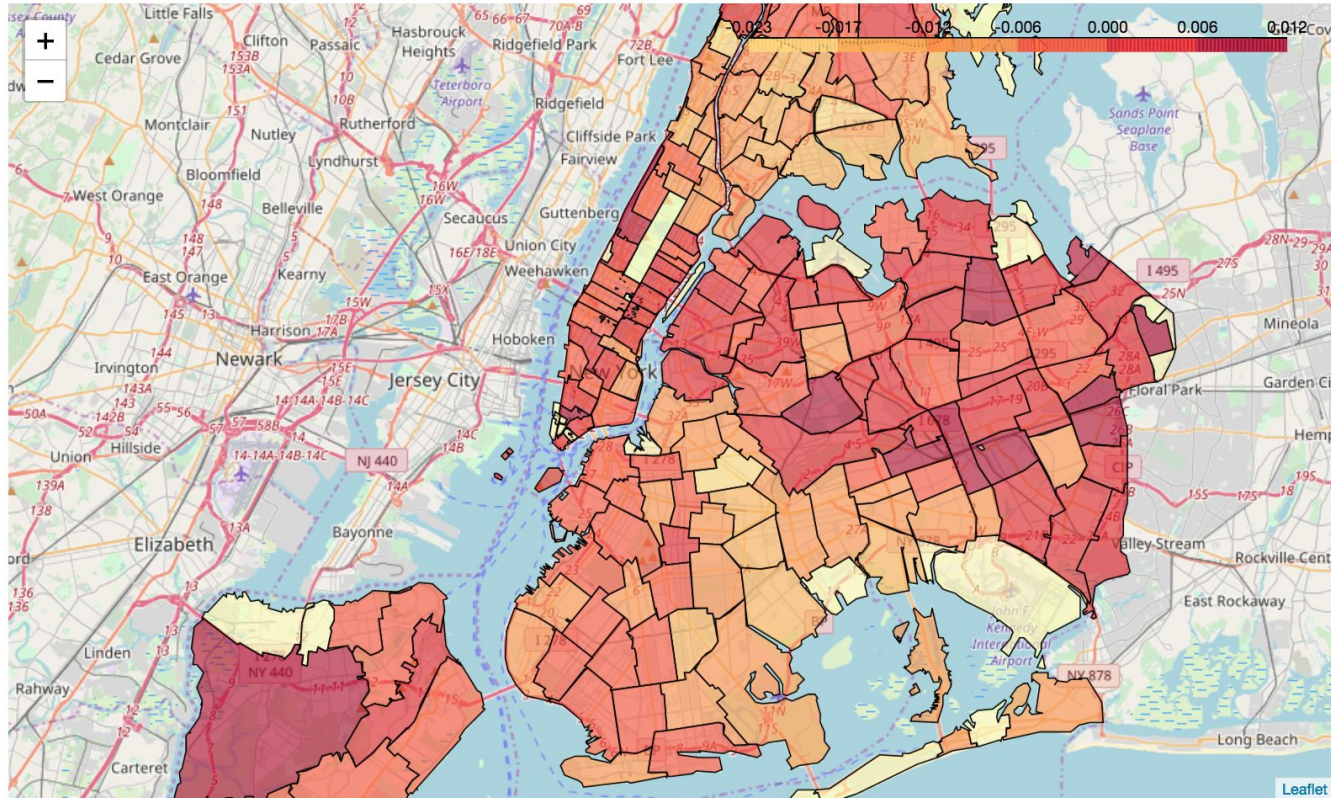
## Top Zip Codes for Extra Resources

	zip_code	collision_momen	fatality_rate_momen	collision_mean_yoy	fatality_rate_mean_yoy
118	11357.0	76.7	0.000578	0.165580	0.000123
61	10466.0	68.5	0.009292	0.055256	0.043000
140	11413.0	66.2	0.003155	0.062300	0.009387
137	11385.0	39.2	0.005343	0.029032	0.019285
72	11101.0	39.1	0.004192	0.042997	0.038942
155	11432.0	35.2	0.001225	0.031167	0.001947
28	10034.0	33.4	0.001682	0.069742	0.024857
154	11429.0	31.9	0.011149	0.065472	0.036723

	borough	number_zips
0	QUEENS	19
1	BRONX	4
2	BROOKLYN	3
3	STATEN ISLAND	2
4	MANHATTAN	2

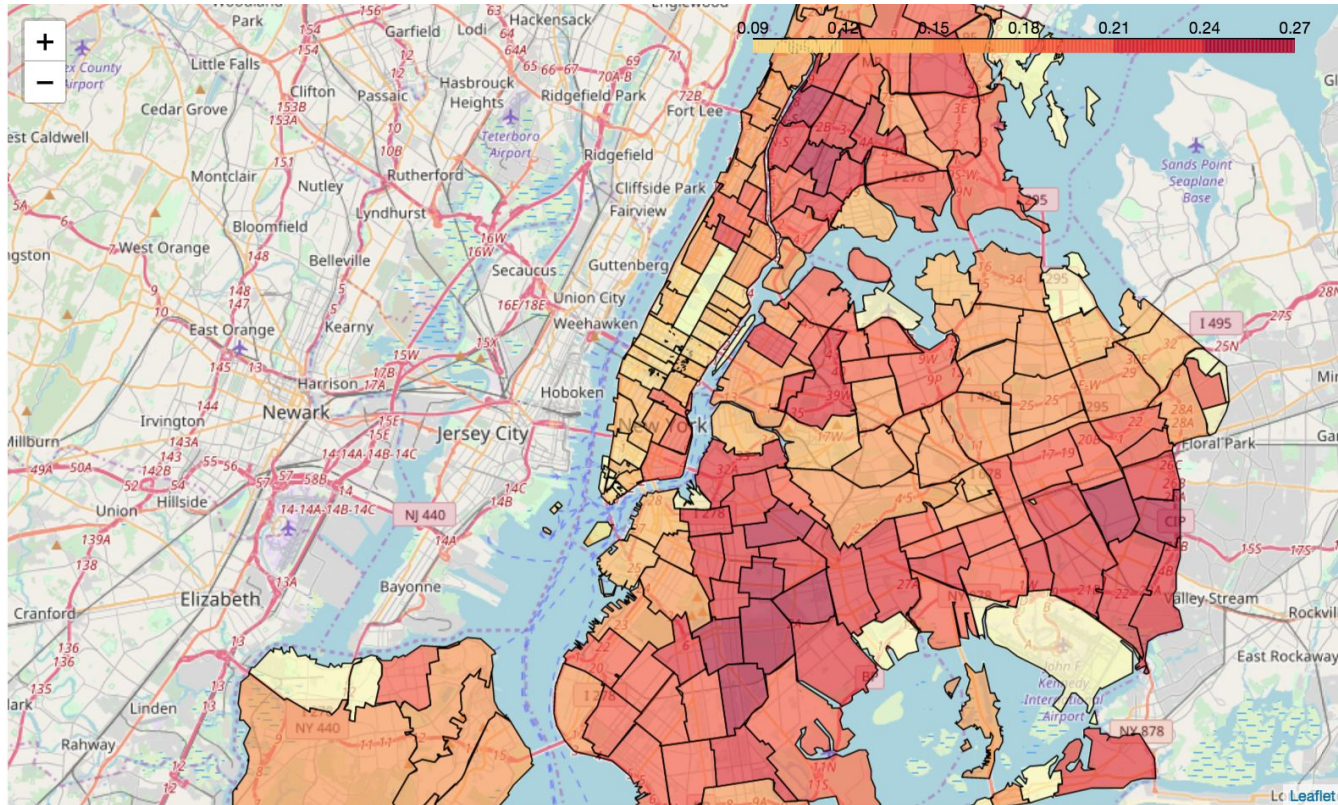
- Analysis identified 25 zip codes across the 5 boroughs with a positive trend and a positive mean year over year growth rate for collisions and fatality rates
- Most of the neighborhoods that could use resources are in Queens, understandably since it is the largest and most diverse borough
- Queens is one of the most sparsely populated and spacious boroughs, offering opportunities for construction companies, medical services, auto shops, and legal aid to both prevent and improve collision responses

# Zip Code Analysis: Fatality Rate Momentum





# Zip Code Analysis: Mean Fatality Rate



# Conclusion

- By analyzing NYPD motor vehicle collision data, we are able to better understand the potential causes and correlations of collisions along geographical, timing, factor, and vehicle dimensions
- We expanded our analysis in a more granular level (focusing on specific zip codes and neighborhoods) in order to understand evolving trends and make recommendations for increased resources and business opportunities
  - Queens and the Bronx are identified as boroughs that can provide collision-focused and collision-preventative opportunities

## Further Study

- Acquire more independent data in order to study and understand systems in depth
  - Demographics (population census, ethnicity, age, household size and income)
  - Weather (storms and snow season)
  - Motor Vehicle Insurance (price, traffic violations, license)
  - Car (model, make)
- Potential analysis
  - Build predictive analytics model to predict key metrics including volume of collisions and fatality rate
  - Use clustering analysis to hone in on drivers motor collisions within neighborhoods to better create resource and business opportunities
  - Relate vehicle model to fatality rates to drive technological enhancements of safety features and increase marketing potential