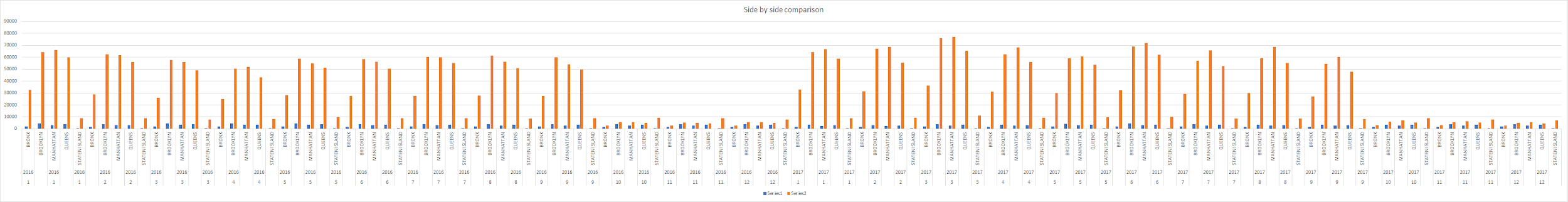
Analytics

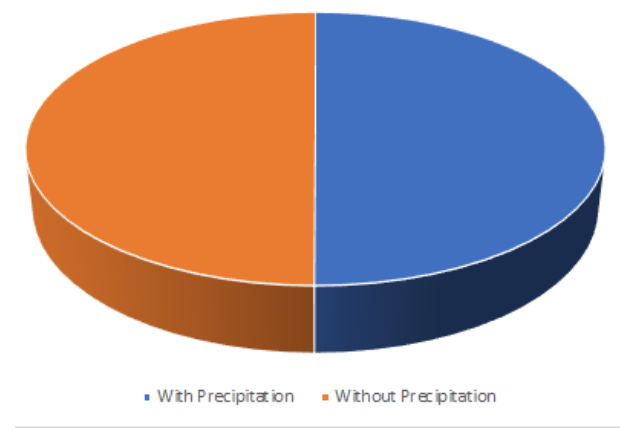
1. Correlating traffic offenses and vehicle collisions:

The hypothesis behind this analytic was that the areas where people have a general disregard for traffic rules, the chances of vehicle collisions happening in that area was higher. The traffic tickets dataset and non-moving traffic violations dataset were aggregated according to month and year boroughwise. A similar transformation was done in the motor collision dataset and a correlation coefficient of 0.88 was achieved. The graph shows the side by side comparison for the year 2016. The orange bar indicates vehicle collisions and blue bars indicate traffic violations.



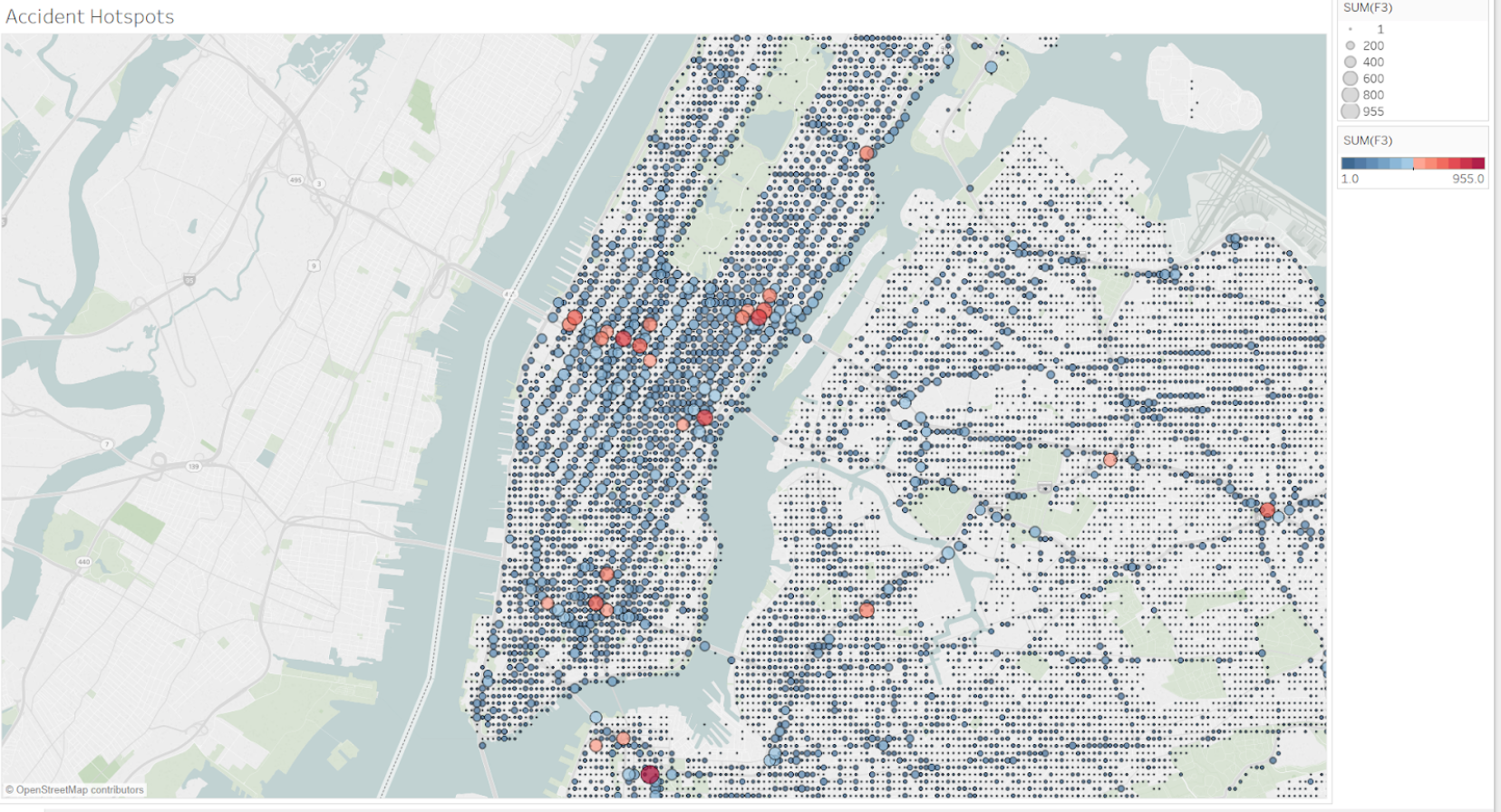
2. Effect of weather on Motor Collisions

The weather data for 2016 for NYC was taken into consideration along with the Motor Collision dataset. The Motor Collision dataset was transformed for daily accidents counts for comparing it with the weather. Unfortunately, there wasn’t any correlation between weather(precipitation) and motor vehicle collisions. This shocking result can be attributed to some missing data in the Motor Collision dataset. Also as the weather data was daily which could affect the result. The following pie chart shows the accident ratio for days with precipitation and the ones without.



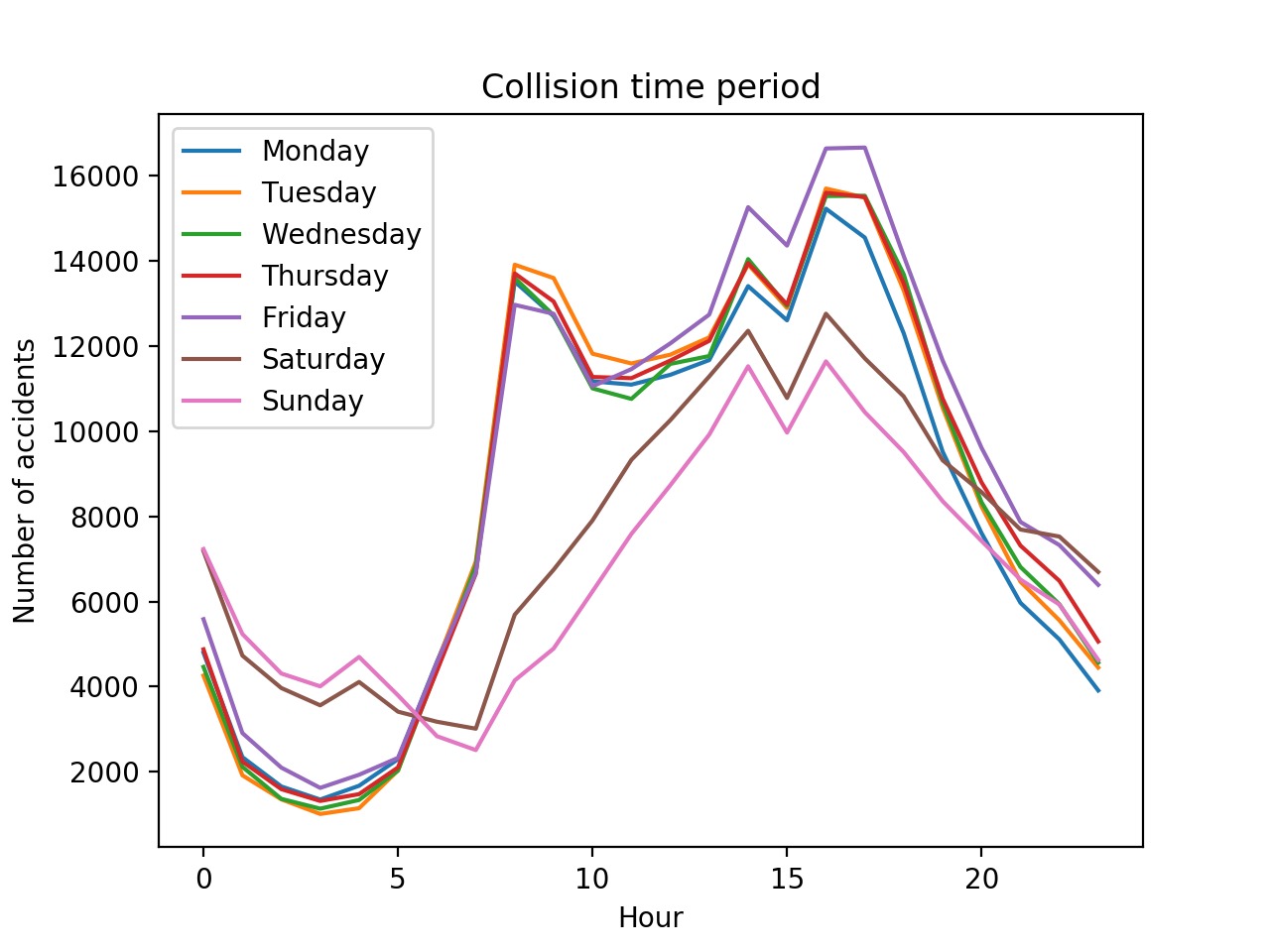
3. Collision Hotspots

The motor collisions dataset had GIS-based data of the accident areas. Hence using latitude and longitude(up to 3rd decimal place), the accident hotspots were identified up to a 100m accuracy. A lot of police reports online bolster our claim about the hotspots. The red bubbles indicate the most accident-prone areas.



4. Identifying the peak accident hours and days

The collision dataset can be transformed to give the day of the week and along with the time. Plotting of the data shows that the accident occurs in the peak travel time of 8am-9am and 3.30pm-5pm on weekdays. The graphs clearly show that the accidents are very less on weekends.



Future work

A lot of other factors can be taken into consideration to improve the result of the main analytic. Various factors like driver’s age, previous moving violations, traffic flow in each borough can be taken into consideration. Also, a better weather dataset with hourly data and proper weather information can be used for improved analytics. A detailed information of the petty traffic offenses with latitude, longitude based data will improve the analytic.

Conclusion:

The main analytic proves that the areas where people tend to flout the traffic rules are the areas with high vehicle collision rates. Hence the overall attitude of the general public in the area is very important for road safety.