Interpreting The Effect Of COVID On Motor Vehicle Accidents in the New York Metropolitan Area (SQL Integration)

Jian Ruan, Zach Yu, Carrie Hashimoto, Sophia Lyu (Undergraduate), Patrick Yee (Graduate)

March 30, 2023

1   Introduction

COVID-19 quarantine limited people to remote work and in-home activities. It's likely that the COVID-19 restriction of transportation decreased the number of car crashes in NYC. To test out our hypothesis, we first compare and contrast the car crash distribution through geographical mapping, and later through time-series analysis and free-text analysis based on New York Times articles.

For this report, we transferred our data to SQL database to significantly reduce the size of the data.frames and increase the efficiency of data analysis.

2   Geographical Distribution

Each incident is mapped as a blue dot (x = lattitude,y = longitude) based on the OpenStreetMap API for New York City. The bluer the area shows, the more accidents happened.

As shown in Figure 1, there is a sudden decrease in the number of accidents coinciding with the start of the COVID-19 pandemic, especially around the two most populous areas - Manhattan and Brooklyn. Since COVID transmitted quickly among large group of population, Manhattan and Brooklyn were the top two areas with people tested positive at the start of 2020. As a result, the quarantine policy for Manhattan and Brooklyn was more strict than other places, and people adapted to online activities to avoid infection.

3   Contributing Factors

We had previously created pie plots of the yearly relative frequency of the ten most common contributing factors to car crashes. However, these plots were difficult to interpret because some of the changes between fractions by year were fairly subtle and the color gradients aren't the most pronounced. We decided to visualize something similar using a different type of plot for this draft.

Interpretation: This line plot shows the relative frequency of crashes caused by each of the ten contributing factors that cause the most injuries by year. Ranking by the contributing factors that cause the most injuries was a way of ensuring that we considered the most relevant causes, and using a line plot allows us to look at the changes in proportion of each factor over time in addition to the importance of each factor relative to the others.

It was taken into consideration what were the top contributing factors for New York crashes (excluding the unspecified causes). Many crash factors were accounted for so the top 5 contributing factors (excluding unspecified) were examined. Within the data of top 5 contributing crash factors, the proportion of vehicle type involved was implemented by color. Below is the bar graph of the percentage of crashes occurring with the top 5 crash factors and the type of vehicle taken into account.

The bar graph shows that driver inattention/ distraction accounts significantly for the proportion of crashes in New York city. In addition, within each of the top 5 crash factors, sedan accounts for the most number of crashes, supporting the fact that sedan is involved in the most amount of crashes. We can further explore these high-risk contributing factors by visualizing raw relative counts. An array of pie charts show the
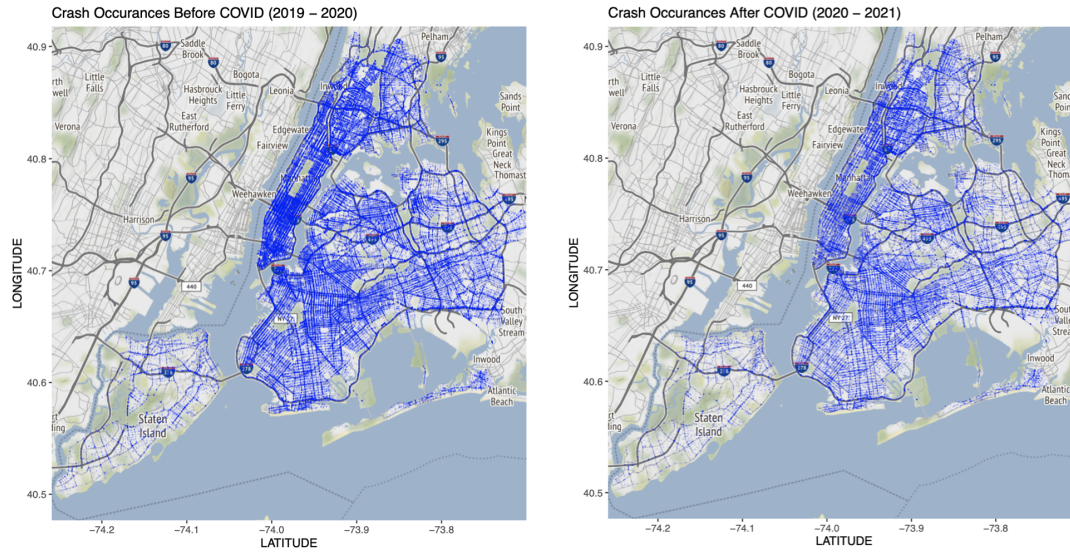
Figure 1: Accidents Distribution Before vs After COVID.

relative frequency of each of the top ten contributing factors to car crashes by borough and number of fatalities. Feature engineering was performed to combine related contributing factors into new categories such as "Fatigued/drowsy/sleepy/unconscious," and entries with unknown borough were dropped.

4 Time Series Analysis

5 Radar Analysis