NEW YORK CRASH ANALYSIS

Abstract

This study presents a thorough analysis of motor vehicle incident data in New York City, revealing numerous patterns and insights that are critical to understanding the workings of urban safety. Exposing the geographical distribution of occurrences throughout the boroughs, the study highlights Queens' disproportionate impact, since the borough with the highest number of incidents per population density is not the most densely inhabited. Furthermore, time patterns show distinct peaks throughout the morning and afternoon commute, indicating a relationship between events and rush hour traffic.

Dmitrii Ponomarev

Ec22805@qmul.ac.uk

Table of Contents

Chapter 1 Introduction

Description

This analysis focuses on motor vehicle collisions in New York city (City of New York, 2023). I'd want to investigate different hypotheses on the dataset provided below in order to uncover trends and patterns in the data that could be used to increase safety precautions during rush hours and holiday seasons.

All of the code can be found on my GitHub profile: https://github.com/Mamin-gamer/DataAnalysisModule

Dataset

Data downloaded has numerous columns showed in Figure 1.

```
Index(['CRASH DATE', 'CRASH TIME', 'BOROUGH', 'ZIP CODE', 'LATITUDE',
       'LONGITUDE', 'LOCATION', 'ON STREET NAME', 'CROSS STREET NAME',
       'OFF STREET NAME', 'NUMBER OF PERSONS INJURED',
       'NUMBER OF PERSONS KILLED', 'NUMBER OF PEDESTRIANS INJURED',
       'NUMBER OF PEDESTRIANS KILLED', 'NUMBER OF CYCLIST INJURED',
       'NUMBER OF CYCLIST KILLED', 'NUMBER OF MOTORIST INJURED',
       'NUMBER OF MOTORIST KILLED', 'CONTRIBUTING FACTOR VEHICLE 1',
       'CONTRIBUTING FACTOR VEHICLE 2', 'CONTRIBUTING FACTOR VEHICLE 3',
       'CONTRIBUTING FACTOR VEHICLE 4', 'CONTRIBUTING FACTOR VEHICLE 5',
       'COLLISION_ID', 'VEHICLE TYPE CODE 1', 'VEHICLE TYPE CODE 2',
       'VEHICLE TYPE CODE 3', 'VEHICLE TYPE CODE 4', 'VEHICLE TYPE CODE 5'],
      dtype='object')
```

Figure 1 - Columns of the dataset

Zip code, location, on street name, cross street name, and off-street name are less helpful and don't offer as much information as geographic coordinates, therefore I left them out of my research. Columns like Contributing Factor 2–5 and Vehicle Type Code 2–5 were included since some collisions involved numerous motor vehicles, although the majority of the values in those columns are missing.

I utilised a second resource to acquire more information about each of New York's boroughs, which included statistics about the population and size of each borough (www.citypopulation.de, 2023). This assisted me in concluding that, despite the fact that most collisions occur in Brooklyn, Queens has the highest density of crashes.

Hypothesis

1) The average number of incidents doesn't vary significantly across different days of the month
2) The average number of incidents doesn't vary significantly across different hours of the day
3) The average number of incidents doesn't vary significantly across boroughs
4) The density of incidents distributed the same across New York

Chapter 2 Analysis

In this section, I will be performing a comprehensive analysis of the dataset, aiming to prove or disprove my hypothesis set earlier.

Boroughs

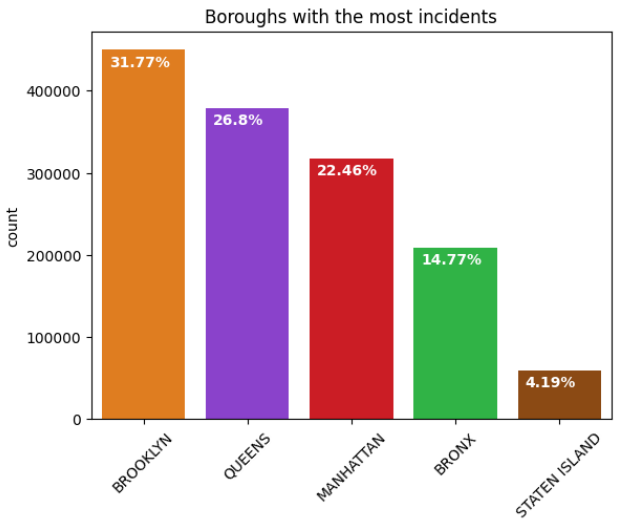

Figure 2 - Boroughs by total number of incidents

One of my hypotheses is to check whether boroughs have the same distribution of incidents. By displaying how many incidents took place in each borough, I have constructed a bar plot (Figure 2).

According to this graph, the majority of the incidents occurred in Brooklyn, which accounted for over a third of all incidents. Staten Island, on the other hand, accounts for just over 4% of all incidents, making it the safest borough to live in if you're worried about getting into a car accident.

After determining which borough had the most events, I wanted to look into why this was the case. Looking at a map of New York, it is evident that the most heavily populated areas with accidents are Midtown and Downtown Manhattan, which contain the majority of skyscrapers used for office buildings. Furthermore, Manhattan is known for its hurried pedestrians rushing to work, which is a major cause of most accidents, since they don't seem to observe where they are crossing and how quickly cars are racing to get to work as well.

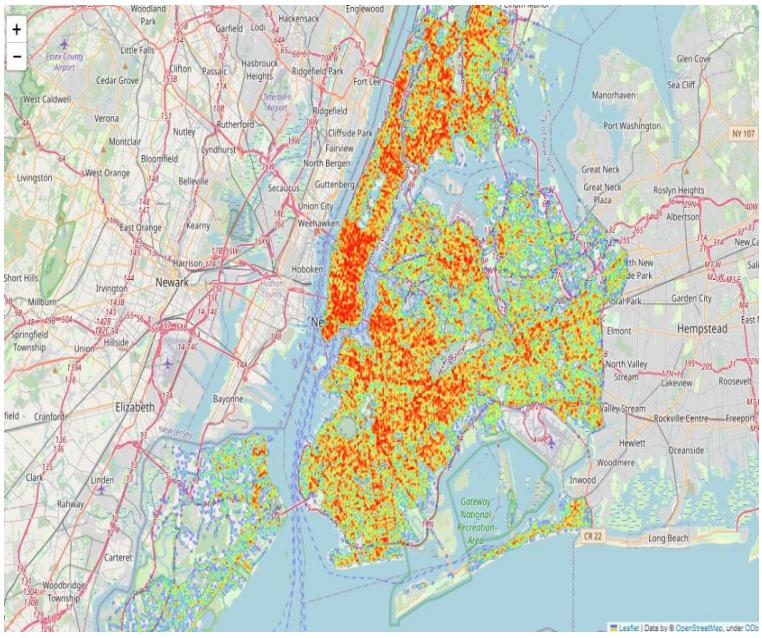

Figure 3 - Location of incidents in New York

About the same number of motorists and pedestrians have been engaged in accidents in Brooklyn and Queens as in Manhattan, given that the two boroughs are significantly larger than Manhattan in terms of area. This shows a distinct urban dynamic within Manhattan, where a significant number of occurrences involve pedestrians. The high volume of foot traffic that Manhattan's busy streets are known for may be a factor in the increased number of foot-related incidents.

It is not surprising that drivers are the most common means of transportation involved in accidents; police reports are only filed in cases when “someone is injured or killed, or where there is at least $1000 worth of damage” (City of New York, 2023). The shocking thing is that Queens and Brooklyn have more vehicle crashes than any other state. One possible explanation for this might be the heavy traffic during the commute to and from work and the airport, which can result more hurried than usual drivers causing accidents.

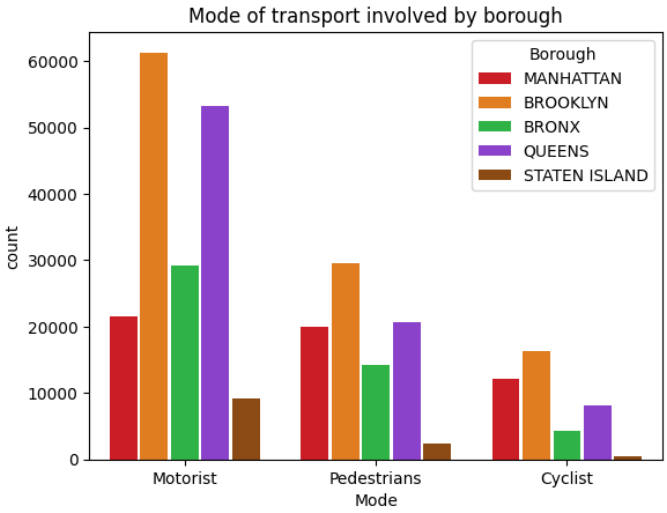

Figure 4 - Mode of transport involved by borough

Time

In this part, we look at the historical patterns of traffic accidents, specifically the relationship between time and the number of incidents. In order to identify trends and critical peak times, we conduct a thorough analysis of the hourly distribution of occurrences over the course of all years.

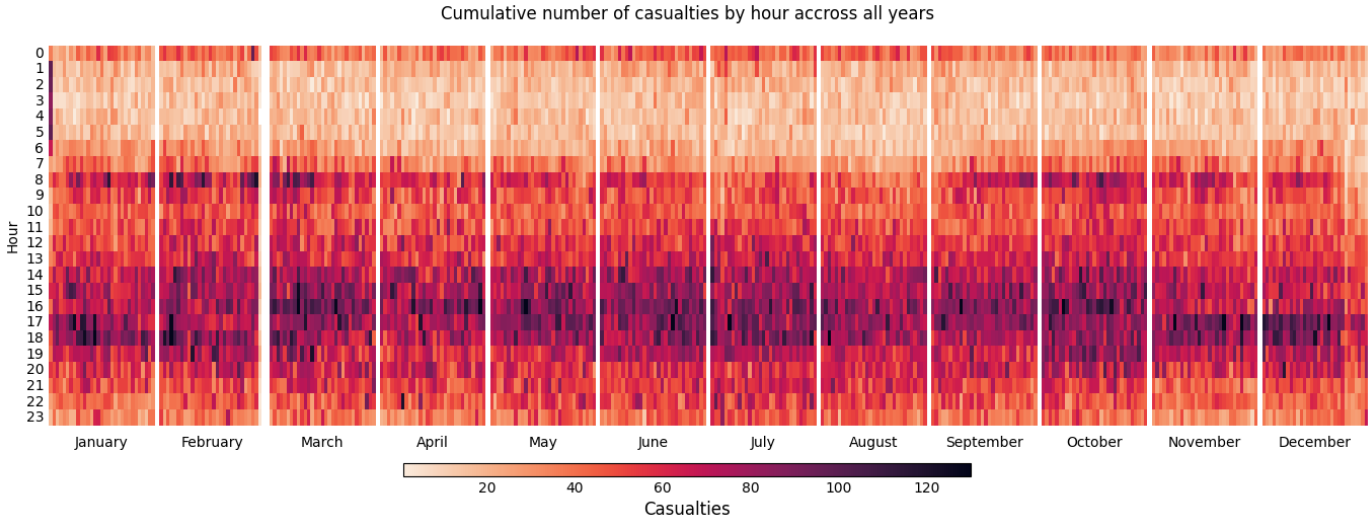

Figure 5 - Number of casualties by hour across months, years accumulated

Figure 5 shows the correlation between hour of the day and number of incidents occurred on that day for each month across all years. Interestingly, we can see patterns emerging: There is a clear trend connected with morning and afternoon rush hours. Morning rush hour is slim – around 8 am is when most of the incidents occur in the first half of the day, whereas afternoon rush hour is more spread. Most likely reason for it is because most businesses start to work at 8/9 am, but close at different times, hence people their places of work at different times and therefore having a larger spread of incident rate.

It's important to note that at midnight, there are often a lot more casualties since people are returning home from bars and clubs, which leads to an increase in both the number of intoxicated drivers and careless crossers. It's interesting to note that this graph clearly shows a trend over time

of an increase in casualties around 1 am on January 1st, when many individuals start drinking and driving after partying, which leads to an increase in incidents. At the end of December, however, there is a narrow white strip that indicates a drop in the number of casualties on December 25. The observable decline in injuries on December 25th suggests a cultural trend towards indoor activities and an overall emphasis on safety over the Christmas holiday. This cyclical explore not only reveals the effect of social behaviours on traffic incidents, but it also emphasises the importance of focused awareness initiatives and law enforcement operations during peak risk times (The Official Website of the City of New York, 2022).

Chapter 3 Conclusion

The average number of incidents doesn't vary significantly across different days of the month

Figure 5 illustrates the relationship between the average number of incidents and the number of months. Since daylight lasts longer after March, mornings are brighter, which makes it easier for pedestrians and automobiles to see one another and reduces the number of casualties during the morning rush hour. In addition, since fewer people are awake in the morning on holidays, there is a lower chance of an incident occurring at this time of year because more individuals are likely to be on vacations.

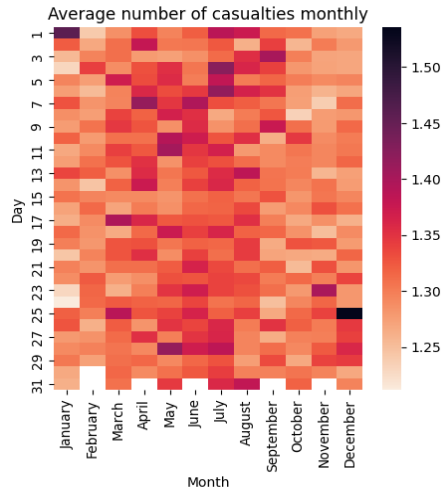

Figure 6 - Average number of casualties monthly

As shown in Figure 6, there is no significant relationship between the average number of events and the day of the month. However, it is important to notice that just a few numbers on this heatmap show a correlation: New Year, Christmas, and America's Independence Day. These days demonstrate a clear causal association between the day of the month and the average number of occurrences that happened on that day.

The average number of incidents doesn't vary significantly across different hours of the day

Shown by a Figure 5, there are clear tendencies towards majority of incidents happening in the morning and afternoon peak times. Most of incidents occur at 8 am, when people rush to work and between 3 and 7 pm when people are getting back from work.

I can conclude that there is a significant correlation between time of the day and an average number of incidents and an hour of the day caused by morning and afternoon rushes.

The average number of incidents doesn’t vary significantly across boroughs

Figure 2 showed that Manhattan has more number of incidents than any other borrow, however after closer examination, we can see that figure 8 shows that Queens, not being the most densely populated borough is the borough with the highest number of incidents per population density.

It turns out that there are noticeable differences in average number of incidents between boroughs. The main cause of this is probably the distinct separation of boroughs and the purposes for which they are occupied. As demonstrated by Kenneth M. Gold, Staten Island differs physically and culturally from the rest of New York (Columbia University Press, 2023). This is probably the cause of the island's low office density, which lowers traffic and reduces the number of auto accidents.

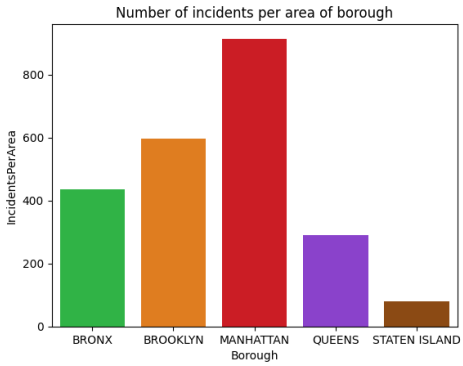

Figure 7 - Number of incidents per area of borough

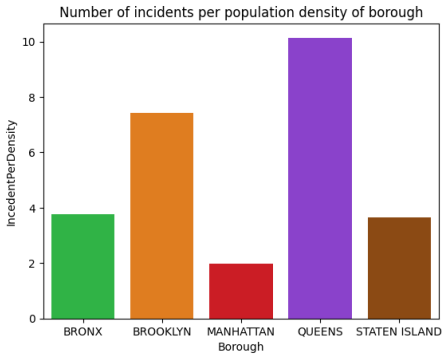

Figure 8 - Number of incidents per population density of borough

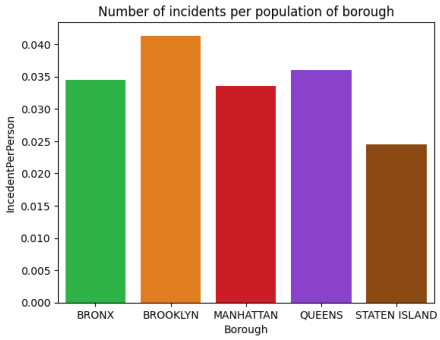

Figure 9 - Number of incidents per population of borough

The density of incidents distributed the same across New York

It turns out that Midtown and Downtown Manhattan have the greatest accident density. For starters, these areas have a high population density, which raises the likelihood of accidents, especially when people are jaywalking. Second, because these areas have numerous workplaces and people go from all around to work during rush hour, there is an elevated risk of accidents. Third, certain localities have tourist attractions, which bring more tourists from other regions and increase the chance of accidents. Finally, these areas, particularly Midtown, have plenty of clubs and other entertainment establishments. Accidents may become more common because of people drinking and driving.

Chapter 4 Table of figures

Chapter 5 References

City of New York, 2023. *Motor Vehicle Collisions - Crashes.* [Online]
Available at: https://catalog.data.gov/dataset/motor-vehicle-collisions-crashes
[Accessed 28 December 2023].

Columbia University Press, 2023. *The Forgotten Borough.* [Online]
Available at: https://cup.columbia.edu/book/the-forgotten-borough/9780231557511#:~:text=What%20sets%20Staten%20Island%20apart,populated%2C%20and%20more%20politically%20conservative.
[Accessed 11 January 2024].

The Official Website of the City of New York, 2022. *Mayor Adams Unveils Sweeping Plan for Holiday Season Car-Free Open Streets in Midtown Manhattan, Highlights Campaigns to Shop Local.* [Online]
Available at: https://www.nyc.gov/office-of-the-mayor/news/859-22/mayor-adams-sweeping-plan-holiday-season-car-free-open-streets-midtown-manhattan-
[Accessed 10 January 2024].

www.citypopulation.de, 2023. *USA: New York City Boroughs.* [Online]
Available at: https://www.citypopulation.de/en/usa/newyorkcity/
[Accessed 10 January 2024].