Using Data Viz to Motivate Safe Driving

Data tells you why common safety tips really could save your life

By Manyu Jiang 16th May 2020

Under lockdown, we all begin to miss the crowds, subways, the Met, “just everything” in New York City before the pandemic had happened, even our daily commutes.

If you sometimes have thought New York City has the worst transportation in the world one or two months ago, you’re not alone. On a daily basis, I would head out for school, jump into a car, going at a snail's pace, and often would be depressed for half of the day. Anyone who has been trapped in a car in NYC feels the helplessness. We know that a lot of delays are caused by traffic accidents, and we probably know what we should have done to avoid it.

A bird view of NYC on a sunny day (Credit: Alamy)

Nearly 12 million licensed drivers in New York have been trapped in a jam in some sense every day; in New York City, a collision happens somewhere every three minutes; 655 people are injured on average every month. NYPD started releasing datasets of every recorded vehicle collision in the city since 2011 to better inform both the public and academia, with the hopes of reducing accidents.

What we could learn from the traffic data? My research aims to analyze the causes and patterns of accidents based on the datasets with the idea of preventing them from happening. Advice about preventing accidents may sound like a cliché: watch out before you cross the street; don’t pick up your phone when driving; avoid severe weather. However, maybe common tips about

driving safely are actually supported by the data and should be taken seriously when you are on the road.

Important facts about traffic accidents happening in NYC (Credit: Manyu Jiang)

Data and method

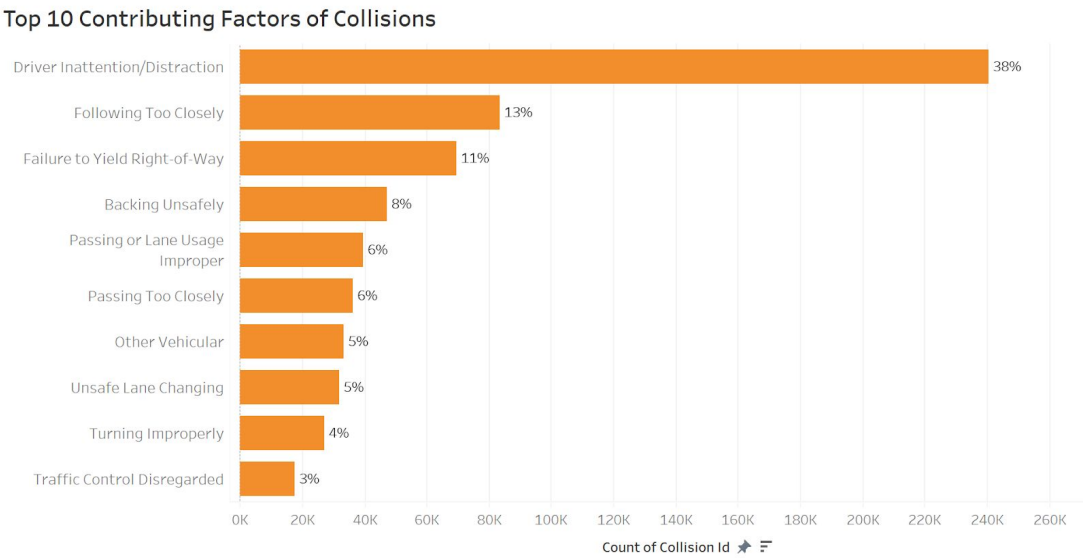
Thanks to NYC Open Data, I have access to NYPD records of every recorded motor vehicle collision. Updated monthly, this dataset includes information like vehicle type, time, and location data, whether pedestrians, motorists, or cyclists were involved in the collision, the number of injuries and fatalities, and contributing factors such as "driver distracted" or "unsafe lane changing."

I use exploratory data analysis (EDA) to summarize the main patterns and characteristics of nearly 1.05 million data points for car accidents in NYC in the recent five years, with the assistance of information visualizations. The research will prove to you, why common safety tips really could save your life.

Common tip 1: be focused!

Imagine all of these things are happening to you in one ride: as you are rushing to get through one more intersection, your phone suddenly rings. When you just reject the incoming phone call, your GPS navigation loses stable connection and the voice guidance cracks. You think it will be fun to drive your pet to the suburb, but he starts bouncing up and down on the seat and probing his head out of the window. You don't even notice that, because your brain is working hard to come up with a hook of the presentation you are going to deliver next Monday.

Concentration is never easy, especially when at the wheel. Distractions can be anywhere… inside the car, outside of it, and even in your mind. Distractions are inevitable in life, however, when on the road, even a few seconds of inattention could result in a fatality. As shown in the chart of the top 10 contributing factors of collisions, besides the commonly known reasons for bad driving habits or skills (following too close, backing unsafely, passing, or lane usage improper), the most common cause is driver inattention and distraction.

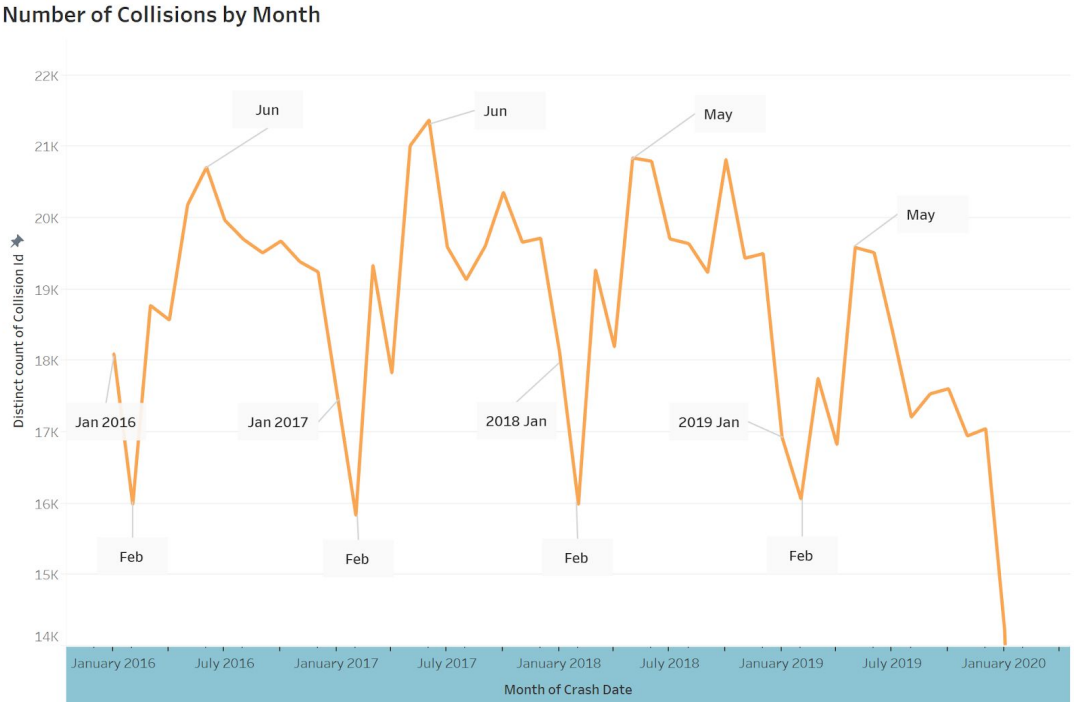


Top 10 causes of car accidents in NYC, 2015 - 2020 (Credit: Manyu Jiang)

So you should really reduce distractions while driving. BE FOCUSED! But how? Read this [complete guide] of maximizing your attention on the road.

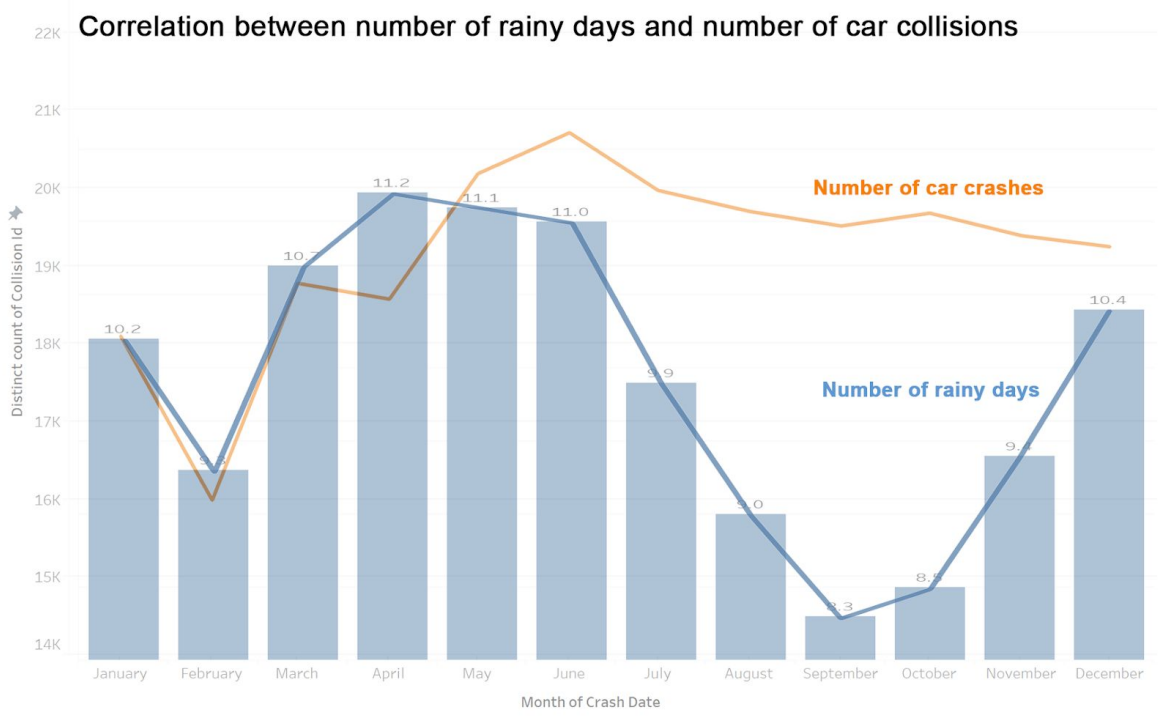
Common tip 2: be cautious of rainy days!

Rain is essential for human life, however, when it happens in Times Square, rain could bring a disaster. Rain reduces friction on the asphalt road, so when you intend to brake, the car may not stop as suddenly as you would wish. Moreover, the rain makes it really hard to see especially in heavy downpours. Therefore you might suspect that there are usually more car accidents on rainy days. The data reaches the same conclusion.



Number of car collisions by month 2016 - 2020 (Credit: Manyu Jiang)

The above chart shows the number of car collisions by month from 2016 to early 2020. We can see some similar ups and downs across the year. Februaries usually have the lowest number of collisions over the year, while May June July have a higher possibility of collisions happening. I hypothesized that this pattern was related to annual precipitation patterns. There is abundant research about the correlation between weather and traffic accidents, when extreme weather happens, such as heavy rain, snow, and storm, there tend to be more traffic collisions.



(Credit: Manyu Jiang)

Here is a correlation chart between the number of rainy days and the number of car collisions. The orange line is the number of car crashes by months, while the blue bars represent the number of rainy days by months. With exploratory data analysis, I'm not making causal claims that rainy days lead to more car crashes, but we do see the two factors are positively correlated with each other, even not statistically significant.

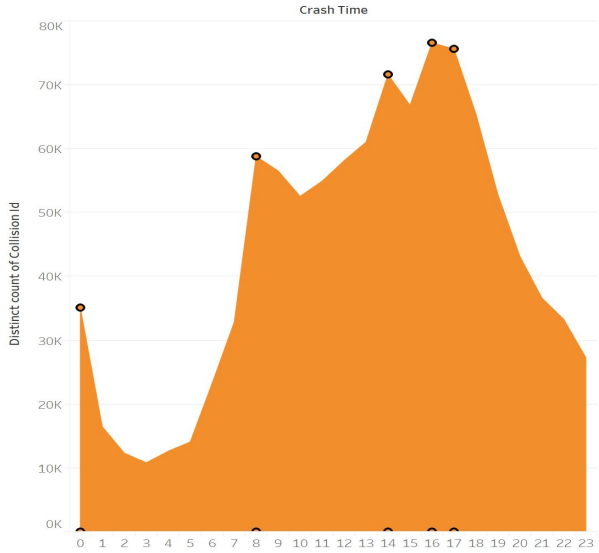
Avoiding driving in severe weather is ideal, but what if you really need to head out, for school, work, or other mandataries? There are some tips to remember. Be more cautious when braking, since the road is quite slippery, and you want to avoid slamming on your brake, so you should brake with less energy. Keep a further distance between you and the car in front of you, which gives you even more time to respond to any emergencies. Of course, the best way to avoid risk of an accident is to stay at home.

Want to learn more about safety tips? Here are ten crucial things to know before you start driving

in the rain.

Common tip 3: avoid traveling during peak traffic times

People like to start their commutes to get to work on time, not late, and not early. But when planning your schedules and work time, you may want to take peak hours into consideration. It may sound terrifying that our daily commutes are dangerous times on the road, but rush hours in the morning and evening do witness more car collisions. Driving conditions during rush hour get increasingly dangerous when the number of cars surge on the road.



Number of collisions in a day in NYC, 2015 - 2020 (Credit: Manyu Jiang)

NYC Open Data shows that the number of accidents starts increasing from 7:00 AM and reaches the first peak during the day at around 8:00 AM when people head to their workplaces and tons of cars come into the city through connecting channels and bridges. The number of car crashes continues to increase throughout the day, and reaches the highest at around 4:00 PM, when people finish their daytime job and go somewhere else. The number slows down after rush hour at 7:00 PM, but interestingly increases again after 11:00 PM then decreases after 12:00 AM. Maybe this also proves why New York City is called the city that never sleeps.

To reduce the possibility of being exposed to the dangerous times when most accidents occur, you may probably want to avoid traveling during peak traffic times.

It's always important to review the common safety tips for driving, especially through the lens of data to understand the advice is actually supported by the data. Speak it out with me: be focused, be cautious of rainy days, and avoid traveling during peak traffic times. When the city is reopen for business and we are free to drive to schools, companies, and our favorite places, the tips you just said out loud could help you be safe on the road and enjoy your trips.

Analysis of Project Evolution

In this brief analysis, I'll talk about how and why my project proposal evolved drastically from a serious research paper related to Covid-19 to an audience-friendly blog post about safe driving tips. The visualizations, writing styles, article format and presentation were modified accordingly since the topic had been changed.

Coming from a journalism background, I was interested in the news of Covid-19 research in the late February when I was first enrolled into the Writing for Data Science class, and decided to use linear regression method to detect the relationship between age and infection rate, in the hope of comparing the rate difference between different age groups. However this proposal seemed unfeasible due to the lack of original dataset about coronavirus death cases that contains information about each individual. So I tried to figure out a strategy to move forward with my project.

I'm excited about using data to tell engaging and in-depth stories and that's part of the reason why I chose this class. The experience of being late for one of our classes because of constant traffic jams on my way to the school inspired me to do some research on this. I was curious about the patterns and causes of car collisions that mostly lead to congestions and wanted to figure out some solutions to prevent them from happening. Professor Hill suggested I narrow down the topic by focusing on one story and really explore it deeply. With solid data about contributing factors of car crashes, number of accidents by month, and number of accidents at different times, I came up with the idea of explaining the rationality of three pieces of safe driving advice based on the data. This became the theme of my final project.

The visualizations were built upon this theme, too. The chart that I made the most changes was the one that illustrates collision seasonal patterns. My initial idea was to identify which time periods usually have more collisions, and what are the reasons behind it. However, my analysis and data viz were based on quarters instead of months at first, which didn't show much information about the patterns of it. Daily numbers and quarter sum-up numbers were also put on the same canvas, which made the data density very overwhelming and hard to interpret, according to peer reviews. I revised this chart in two ways: first, analyzing the data by months not by quarters to see more clues of the potential patterns; second, keeping the monthly numbers only and removing the daily numbers to increase clarity. After doing this, obvious repetitive ups and downs showed up, and this built up basis for further analysis.

My target audience is the public so I set a casual and straightforward tone in the article, and used common tips as dividers for the post. I'm glad that the data viz assisted the message I was supposed to convey in a concise way and worked efficiently as visual threads in my in-class presentation. The process of improving the project is like nurturing seeds from sprouting to growing into plants. It needs patience, passion and constant effort.