Unveiling NYC’s Vehicle Collision Landscape: A Data-Driven Exploration

Ubaydul H Sami

5/9/24

## Abstract

This project delves into the complexities of NYC’s traffic landscape by analyzing police-reported motor vehicle collisions from April 2016 onwards. We meticulously dissect various data points, including crash dates, demographics, injuries, contributing factors, and more, to uncover valuable insights.

A striking trend emerges: a surge in collisions between 2016 and 2017, followed by a significant decrease. Zooming in further, the data reveals peak accident periods during rush hour, highlighting critical times for safety interventions.

The analysis also uncovers a gender disparity, with male drivers involved in a disproportionate number of accidents. Focusing on injuries, the study identifies back, neck, knee-lower leg/foot, and head injuries as common occurrences, pinpointing areas for targeted prevention efforts. Interestingly, the data suggests that seat belts and harnesses are the primary protective measures for uninjured passengers.

The project goes beyond demographics and injuries. It unveils pedestrian/bicyclist error/confusion as the leading cause of accidents, alongside driver distraction and failure to yield right-of-way.

These findings provide a powerful roadmap for improving traffic safety in NYC. By leveraging data-driven insights, we can develop evidence-based interventions to reduce collisions, protect public health, and minimize the impact on infrastructure

## Data Description

This project analyzes traffic accidents in New York City using a comprehensive dataset of police-reported collisions from April 2016 onwards. The data, available on the NYC Open Data platform [Motor Vehicle Collisions - Person Data (NYC) ], provides valuable insights into various aspects of these incidents.

### Unit of Observation:

Each record in the dataset represents a single individual involved in a reported traffic collision. This allows us to examine the characteristics and experiences of people impacted by these accidents.

### Key Variables:

Crash Details: CRASH\_DATE and CRASH\_TIME pinpoint the exact date and time of each collision, providing a temporal context for analysis.

**People Involved:** PERSON\_TYPE (driver, occupant, pedestrian, cyclist), PERSON\_AGE, BODILY\_INJURY, SAFETY\_EQUIPMENT, PED\_ACTION, COMPLAINT, PED\_ROLE, and PERSON\_SEX offer detailed information on the individuals involved, including their role in the accident, injuries sustained, age, emotional state, location within the vehicle, safety equipment usage (for occupants), pedestrian actions (if applicable), reported complaints, pedestrian role (if applicable), and gender.

Contributing Factors: CONTRIBUTING\_FACTOR\_1 identifies the primary cause of the collision, aiding in understanding the root cause of these incidents.

## This code will not be evaluated automatically.  
## (Notice the eval = FALSE declaration in the options section of the  
## code chunk)  
  
my\_packages <- c("tidyverse", "broom", "coefplot", "cowplot",  
 "gapminder", "GGally", "ggrepel", "ggridges", "gridExtra",  
 "here", "interplot", "margins", "maps", "mapproj",  
 "mapdata", "MASS", "quantreg", "rlang", "scales",  
 "survey", "srvyr", "viridis", "viridisLite", "devtools")  
  
install.packages(my\_packages, repos = "http://cran.rstudio.com")

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Here is a snapshot of the data:

## unique\_id collision\_id crash\_date crash\_time person\_injury emotional\_status  
## 1 10249006 4229554 2019-10-26 9:43 Unspecified   
## 2 10255054 4230587 2019-10-25 15:15 Unspecified Does Not Apply  
## 3 10253177 4230550 2019-10-26 17:55 Unspecified   
## 4 6650180 3565527 2016-11-21 13:05 Unspecified   
## 5 10255516 4231168 2019-10-25 11:16 Unspecified Does Not Apply  
## 6 10253606 4230743 2019-10-24 19:15 Injured Conscious  
## bodily\_injury  
## 1   
## 2 Does Not Apply  
## 3   
## 4   
## 5 Does Not Apply  
## 6 Back  
## position\_in\_vehicle  
## 1   
## 2 Front passenger, if two or more persons, including the driver, are in the front seat  
## 3   
## 4   
## 5 Right rear passenger or motorcycle sidecar passenger  
## 6 Driver  
## safety\_equipment complaint ped\_action  
## 1   
## 2 Lap Belt & Harness Does Not Apply   
## 3   
## 4   
## 5 Lap Belt Does Not Apply   
## 6 Lap Belt & Harness Complaint of Pain or Nausea   
## contributing\_factor\_1 contributing\_factor\_2 person\_sex year month  
## 1 U 2019 10  
## 2 F 2019 10  
## 3 M 2019 10  
## 4 2016 11  
## 5 F 2019 10  
## 6 M 2019 10

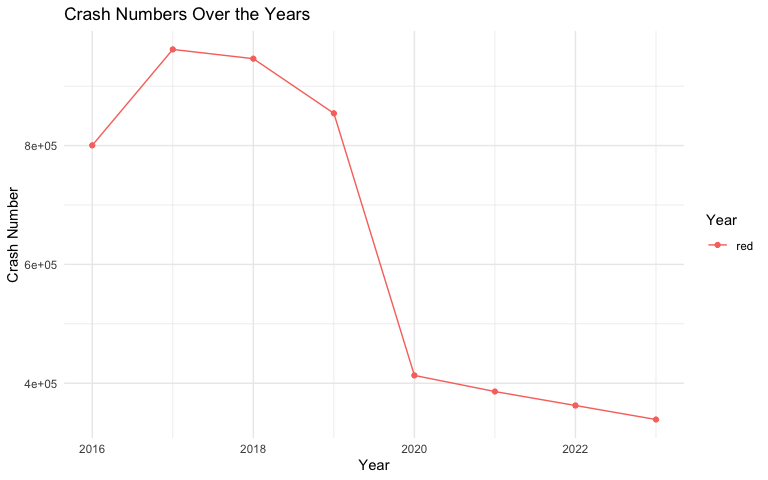
## Research Question

* Why number of Accidents are decreasing over time?
* What are the most prevalent contributing factors (e.g., speeding, distracted driving, pedestrian errors) associated with motor vehicle collisions in NYC?

## Results

### we use a time line to visualize trends of collisions over the years:

## year frequency  
## 1 2016 800353  
## 2 2017 961775  
## 3 2018 946203  
## 4 2019 854269  
## 5 2020 413193  
## 6 2021 386075  
## 7 2022 362555  
## 8 2023 338984



**Description of Figure 1:** On the graph X-axis represents Year from 2016 - 2023 and Y-axis represents number of accidents reported. From the timeline graph we see number of collision was rising from 2016 to 2017 and it starts gradually droping from 2018 to 2023. Our Goal is to find out what happened after 2017 that helped reducing the number of collisions in NYC.

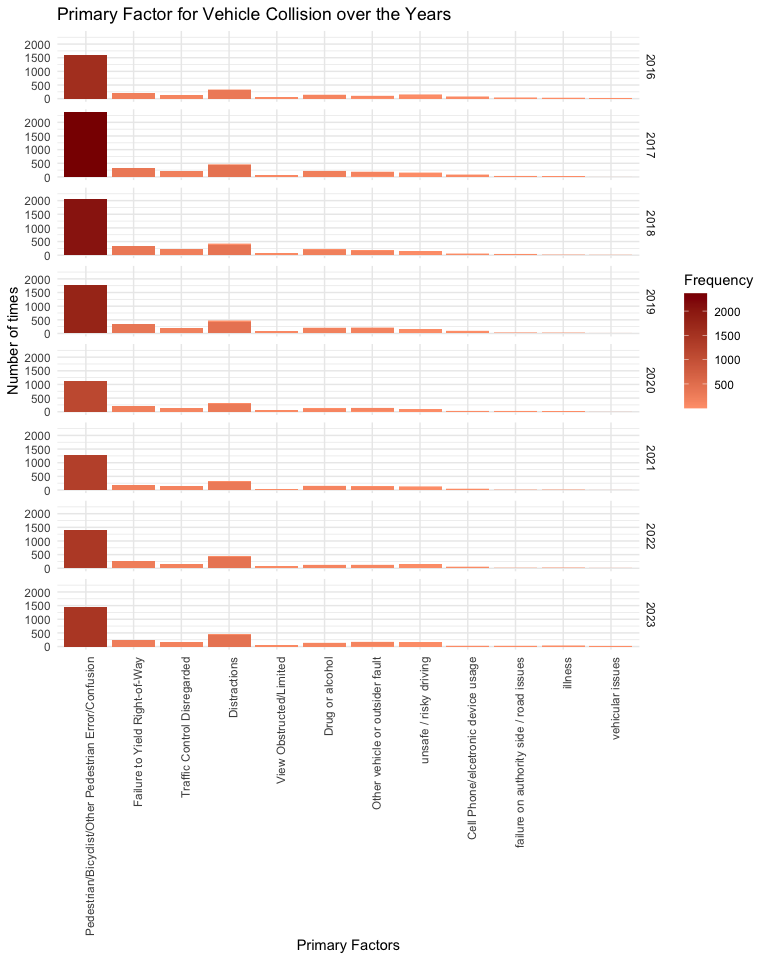
**Inorder to answer the above question we started digging into the dataset and we found column named “contributing\_factor\_1” which represnts the primary cause of the accident** We started analyzing this column value to see what causes are changed over the years.

### visualizing factors that causes crash

## contributing\_factor\_1 year frequency  
## 2 Aggressive Driving/Road Rage 2016 21  
## 3 Alcohol Involvement 2016 138  
## 4 Animals Action 2016 4  
## 5 Backing Unsafely 2016 51  
## 6 Brakes Defective 2016 2  
## 7 Cell Phone (hand-Held) 2016 46

# marging similar category

## contributing\_factor\_1 year frequency  
## 2 unsafe / risky driving 2016 21  
## 3 Drug or alcohol 2016 138  
## 4 Other vehicle or outsider fault 2016 4  
## 5 unsafe / risky driving 2016 51  
## 6 vehicular issues 2016 2  
## 7 Cell Phone/elcetronic device usage 2016 46  
## 8 Cell Phone/elcetronic device usage 2016 2  
## 9 Distractions 2016 316  
## 10 unsafe / risky driving 2016 8  
## 11 Other vehicle or outsider fault 2016 4



**Description for Figure 2:** X-axis of this Figure represents Primary causes of vehicle collisions & Y-axis represent number of collisions Facets by year (separated for each year)

## Finding:

Pedestrian or bicyclist errors stand out as the leading cause of accidents, followed closely by distractions, failure to yield right of way, traffic control disregard, drug-alcohol influence, and risky driving. Despite progress in mitigating these factors, distractions and risky driving persist as ongoing challenges.

**Trends:** While there has been a noticeable decline in most contributing factors, distractions and risky driving have shown minimal improvement between 2017 and 2023.

**Contributing Factors and Solutions:** 1. **Pedestrian or Bicyclist Error:** The implementation of additional traffic lights and dedicated bike lanes has significantly reduced accidents attributed to pedestrian or bicyclist errors. Moreover, replacing stop signs with traffic lights in congested areas and extending walk sign times have minimized confusion and enhanced safety.

1. **Traffic Control Disregard:** The deployment of speed and red light cameras, coupled with increased traffic violation penalties, has effectively curbed instances of traffic control disregard.
2. **Drug-Alcohol Influence:** A concerted effort in raising awareness about the dangers of driving under the influence, reinforced by stringent law enforcement measures and checkpoints, has contributed to a decline in drug-alcohol-related accidents.

**Persistent Challenges:** 1. **Distractions:** Despite efforts to address distractions, particularly stemming from cell phone and earphone usage, these habits persist unabated. The proliferation of mobile devices has exacerbated the problem, with individuals failing to exercise restraint even while behind the wheel.

1. **Risky Driving:** Despite widespread photo enforcement initiatives, reckless driving behaviors remain largely unchecked. In fact, the proportion of accidents attributable to risky driving has shown an alarming increase in 2023 compared to 2017.

In conclusion, while commendable progress has been made in reducing accident-causing factors, sustained efforts are imperative to effectively tackle the persistent challenges posed by distractions and risky driving.

## Limitations

Beyond the primary contributing factors, it’s essential to acknowledge other elements that could significantly impact the decrease in accidents. One such factor is the evolution of safety technology within vehicles. Over time, vehicles have incorporated advanced safety features such as alerts for unsafe speeds, lane detection systems, automatic braking mechanisms, and enhanced brake capacities. However, these advancements aren’t accounted for in the dataset, limiting our ability to analyze their influence on accident reduction.

Furthermore, the dataset lacks clarity regarding its collection methodology. It’s unclear whether the criteria for reporting accidents remained consistent over time. Without this information, we cannot definitively ascertain whether the observed reduction in collision rates is solely attributable to changes in driving behaviors or if it reflects alterations in reporting standards.

Acknowledging these limitations underscores the complexity of assessing trends in accident rates and emphasizes the need for comprehensive data collection methods and consideration of technological advancements in future analyses.