Analysis of Motor Vehicle Collisions in New York City

Fernando Emilio Martinez-Lopez

Abstract

Motor vehicle collisions are not an exception, even in cities with an outstanding public transportation system like New York City. By analyzing crash events through 2016-2021, we see that even though the number of collisions has presented a reduction, their dangerousness has dramatically increased, showing a higher number of events where persons were either injured or killed. We detected a slightly positive trend in the number of killed people in motor vehicle crashes; meanwhile, the number of injuries reduces vaguely. Most NYC crash events happen on Fridays, predominately around 5:00 PM. However, the most lethal time (highest number of killed people) is during the weekends' late-night hours.

Among NYC's boroughs, Brooklyn is constantly the borough with the largest number of tragic accidents (at least one person injured or killed), where the most predominant contributing factors of these events are: Driver was distracted, the driver was following too closely, failure to yield right-of-way, unsafe speed, and traffic control disregarded.

Introduction & Background

By analyzing crash events through 2016-2021, we will see how COVID-19 impacted crash events in New York City (NYC) and understand their fluctuations over different paradigms. My country, the Dominican Republic, has one of the highest numbers of deaths by traffic accidents every year, which motivated me to develop this work [1].

This work analyzes multiple data sources related to NYC motor vehicle collisions alongside another weather dataset for a specific timeframe. The plan is to study how vehicle crashes behave through time in the city, and especially how they fluctuated given the COVID-19 Pandemic. Additionally, I aim to study multiple trends supported by some demographics and reported crash dimensions to generate information that can be used as leverage by the relevant authorities and citizens to prevent these unfortunate matters.

Preconceived Hypothesis

COVID-19 reduced the percentage of mortality produced by motor vehicle collisions in New York City because of the lockdown.

Data sources information

For this work, we explore the four (4) data tables that compromise Motor Vehicle Collisions and historical weather information. The Motor Vehicle Collisions crash [2] table contains 1.87 million vehicle collision event details. It has 29 features, including Crash Date, Crash time, Driver Sex, and Vehicle Type. The Motor Vehicle Collisions Vehicle [3] table has details of the vehicles involved in the crash. The Motor Vehicle Collisions Person [4] table contains details for people involved in the crash, where each row represents a person (driver, occupant, pedestrian, bicyclist) involved in the event. Lastly, the weather dataset [5] contains 999 observations (each grouped by day) and 33 dimensions. This last dataset related to weather historical data is extracted from

another API connection from vissualcrossing.com. I used a trial access which allowed me to pull 999 observations, which belong to data from 2019 to 2021.

Preprocessing Data - Python

Before working on the analytics process, I had to preprocess the data. Given that the datasets contain an enormous amount of information, the only way to pull them was using the public API provided by the New York City Open Data portal (Socrata).

```
import pandas as pd
from sodapy import Socrata
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
import calendar
```

```
client = Socrata("data.cityofnewyork.us", None)
vehicles = client.get("bm4k-52h4", query = """
                                SELECT UNIQUE_ID, COLLISION_ID, DRIVER_SEX,CONTRIBUTING_FACTOR_1, CONTRIBUTING_FACTOR_2
                                WHERE CRASH_DATE BETWEEN '2016-01-01' AND '2021-12-31'
                                LIMIT 10000000""")
vehicles_df = pd.DataFrame.from_records(vehicles).astype({'UNIQUE_ID':'float', 'COLLISION_ID': 'float'})
del vehicles

person = client.get("f55k-p6yu", query = """
                                SELECT unique_id, collision_id, ped_role, person_sex, person_age
                                WHERE CRASH_DATE BETWEEN '2016-01-01' AND '2021-12-31'
                                LIMIT 10000000""")
person_df = pd.DataFrame.from_records(person).astype({'collision_id': 'float', 'person_age': 'float'})
del person

crashes = client.get("h9gi-nx95", query="""
SELECT crash_date, crash_time, collision_id, 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, borough, contributing_factor_vehicle_1, contributing_factor_vehicle_2,
zip_code, LOCATION
WHERE CRASH_DATE BETWEEN '2016-01-01' AND '2021-12-31'
LIMIT 10000000""")
crashes_df = pd.DataFrame.from_records(crashes).astype({'collision_id': 'float',
                                'number_of_persons_injured':'float', 'number_of_persons_killed': 'float',
                                'number_of_pedestrians_injured':'float', 'number_of_pedestrians_killed':'float',
                                'number_of_cyclist_injured':'float', 'number_of_cyclist_killed':'float',
                                'number_of_motorist_injured':'float', 'number_of_motorist_killed':'float'})
del crashes
crashes_df['crash_date'] = pd.to_datetime(crashes_df.crash_date + ' ' + crashes_df.crash_time)
#crashes_df = crashes_df[crashes_df.crash_date >= '2016-01-01'].drop('crash_time',axis=1)
crashes_df['Month'] = pd.DatetimeIndex(crashes_df['crash_date']).month
crashes_df['DayOfWeek'] = crashes_df['crash_date'].dt.day_name()
crashes_df['StartOfWeek'] = crashes_df.apply(lambda row: row['crash_date'] - dt.timedelta(days=row['crash_date'].weekday()), axis=1).dt.date
crashes_df['Year'] = crashes_df.crash_date.dt.year
crashes_df['Injured Persons'] = crashes_df.number_of_persons_injured.map(lambda x: 1 if x > 0 else 0)
crashes_df['Killed Persons'] = crashes_df.number_of_persons_killed.map(lambda x: 1 if x > 0 else 0)
crashes_df['Tragic Accident'] = crashes_df.apply(lambda x: 1 if x.number_of_persons_injured> 0 or x.number_of_persons_killed > 0 else 0, axis=1)
```

```
WARNING:root:Requests made without an app_token will be subject to strict throttling limits.
```

Figure 1: Preprocessing in Python

During this process, I preprocessed some features and created some others for further analysis. Most of the attributes developed were related to date/time dimensions and created one of the most critical dimensions in the following steps, **Tragic Accident**. A **tragic accident** is defined as a crash event where at least one person is either injured or killed.

Preprocessing Data – Tableau

After this step, I imported all the data sources to Tableau and created relationships between tables. To join the weather information with the rest, I had to develop a calculation at the joining step to make it work, similar to a SQL task, because the process is case-sensitive (fig. 2).

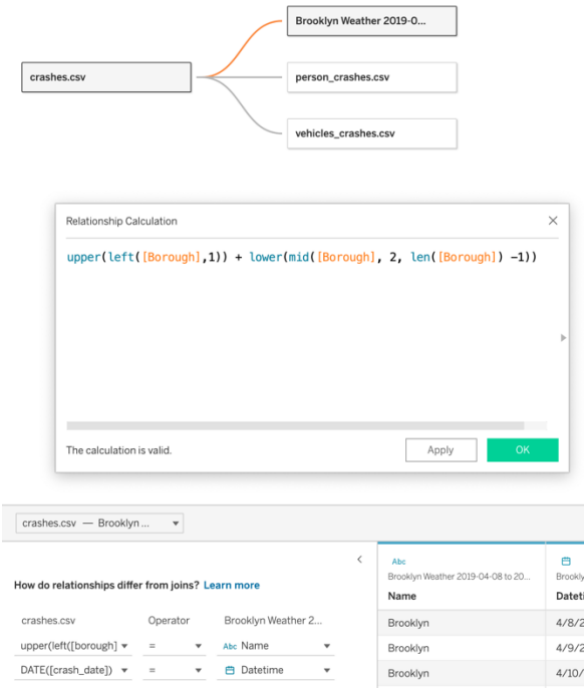


Figure 2: Data sources relationship

Similarly, to make the dashboards more dynamic, I created three (3) parameters, and two (2) features for further user interactions. List of parameters:

- *Select Dimension* (To dynamically change what plot to visualize) (fig. 3)
- Top N Contributing Factors by Borough
- Top-N Contributing Factor

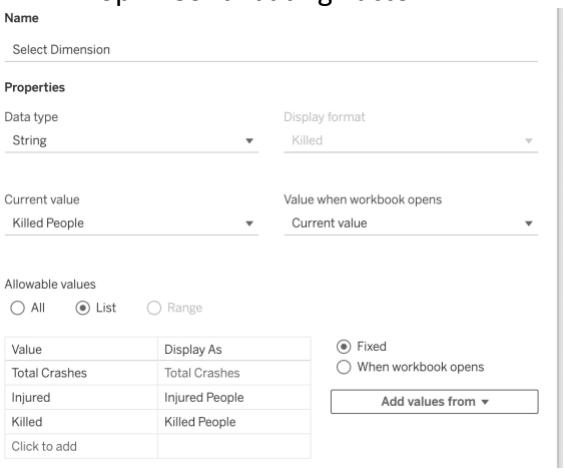


Figure 3: Parameter - Selection Dimension

Data Visualization - Story

In total, I developed 13 different visuals to tell my story, which are distributed in four (4) different Dashboards. These visuals are either the main plot of a dashboard or working as a dynamic user interactive plot, or as part of a tooltip. My Tableau Story compresses four (4) sections:

- NYC Motor Vehicle Crashes from 2016 to 2021

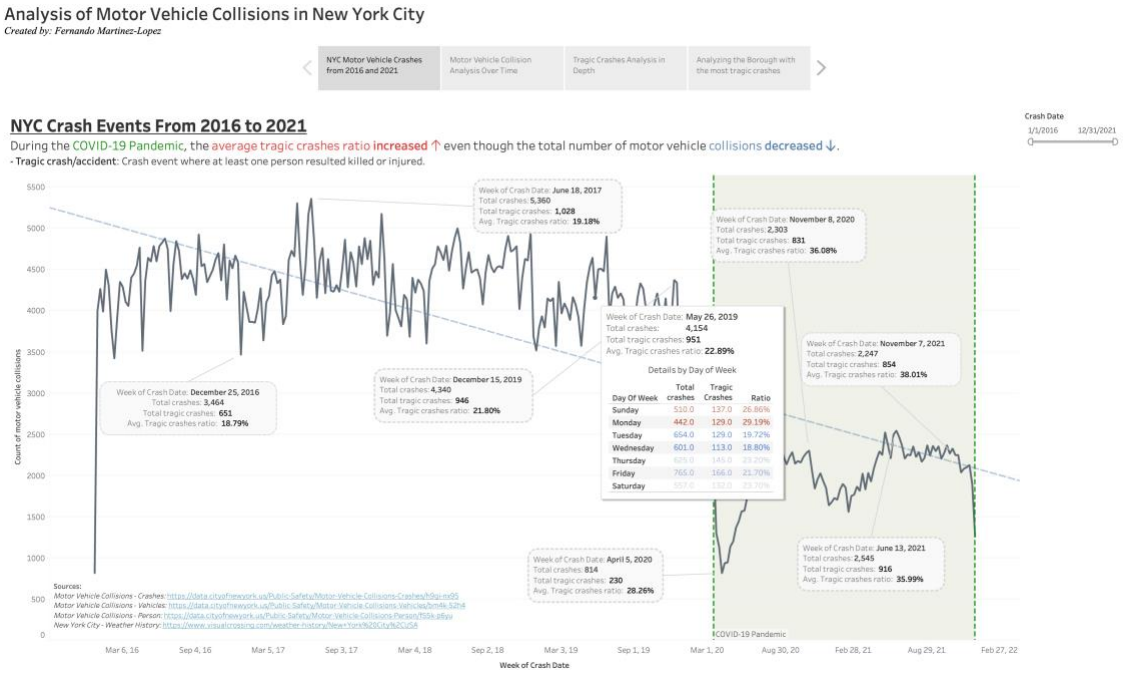
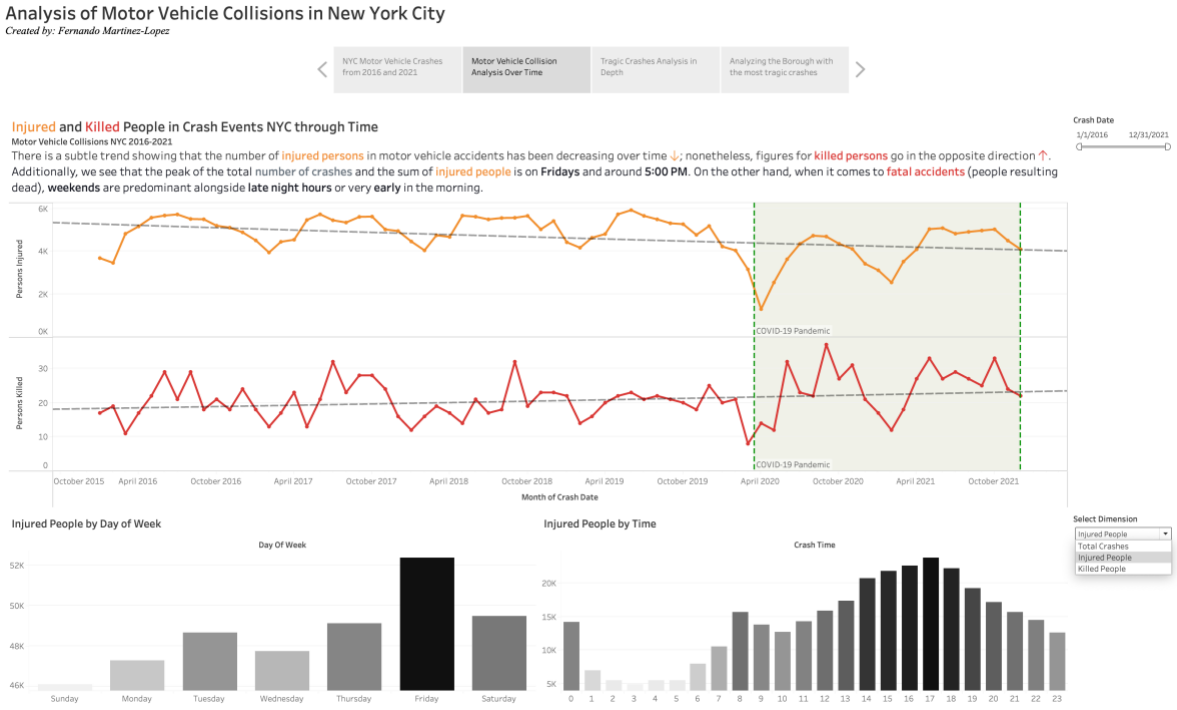
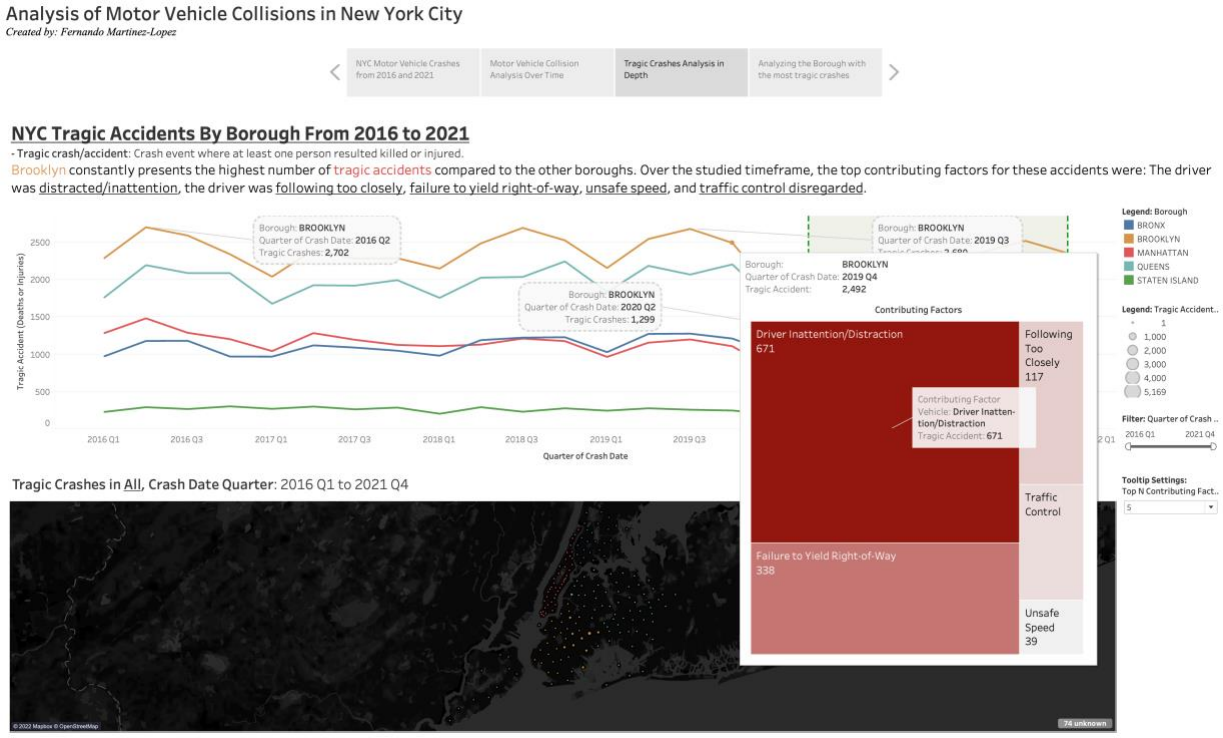


Figure 4: NYC Crashes from 2016 to 2021

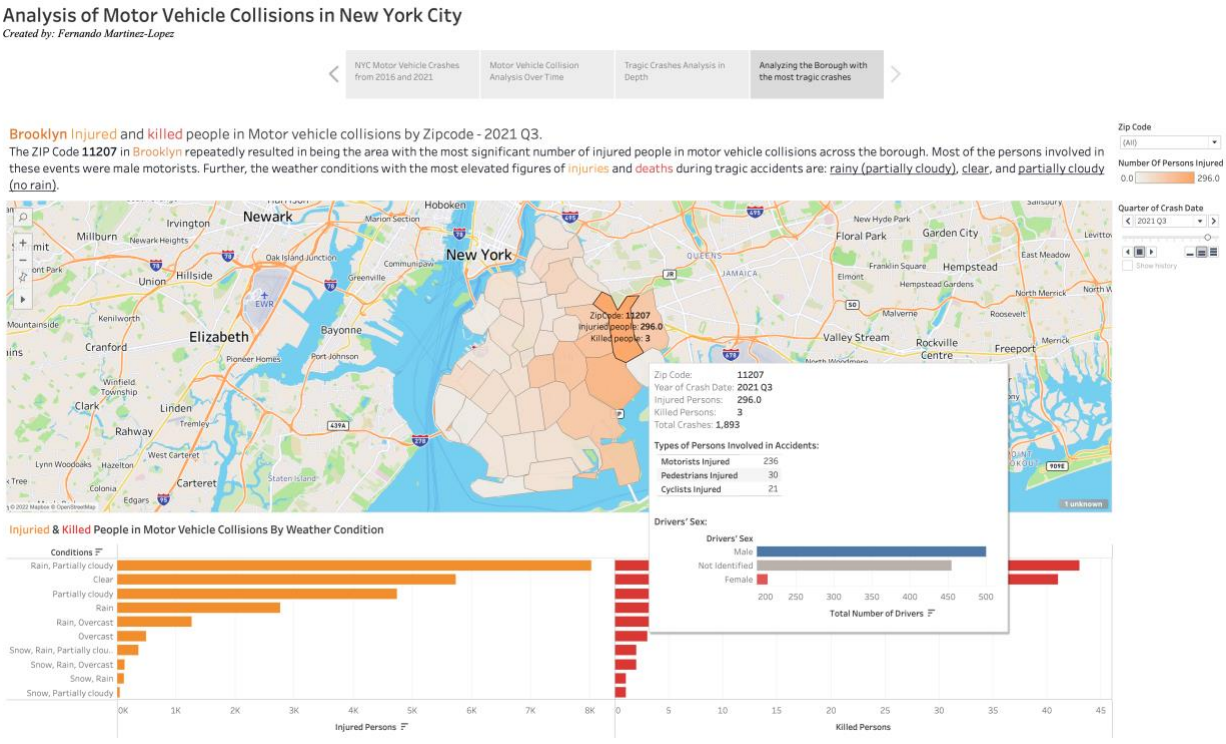
- Motor Vehicle Collision Analysis Over Time



- Tragic Crashes Analysis in Depth



- Analyzing the Borough with the most tragic crashes



Data Visualization – Story Discussion

NYC Motor Vehicle Crashes from 2016 to 2021

The story starts with an overview of all the crash events that occurred between 2016 and 2021. Even though there is a negative trend showing that the number of crash events decreases over time, when taking a deeper look, the number of tragic crashes dramatically increases after the COVID-19 Pandemic. We can infer that the reason behind this is that after the lockdown, people started driving more irresponsibly or recklessly, and the habit stayed. Before the pandemic, the Average tragic crashes ratio (tragic crashes by the total number of crashes) used to be close to 20%; however, after the lockdown, it raised near to 32% on average. This behavior continued even in late 2021, where we can see very close to 37%.

Motor Vehicle Collision Analysis Over Time

Taking a deeper look at the results of these crash events, we notice that there is a subtle trend showing that the number of injured persons in motor vehicle accidents has been decreasing over time; nonetheless, figures for killed persons go in the opposite direction. As seen in the overall view, the performance of these two (2) dimensions is not alien to the prior observations. The figures observed demonstrate that the trends rose after the lockdown, confirming that the dangerousness of the crashes increased. Additionally, we see that the peak of the total number of crashes and the sum of injured people is on Fridays and around 5:00 PM. This evidence suggests that most crashes occur when people get out of work. On the other hand, when it comes to fatal accidents (people resulting dead), weekends are predominant alongside late night hours or very early in the morning. A reason for this could be that late-night people go out and drive more irresponsibly; however, to get to these types of conclusions, we need to analyze further the accidents' contributing factors.

Tragic Crashes Analysis in Depth

Considering the implications of tragic crashes, I decided to analyze them by borough. It turns out that Brooklyn constantly presents the highest number of tragic accidents compared to the other boroughs. Over the studied timeframe, the top contributing factors for these accidents were: The driver was distracted/inattention, the driver was following too closely, failure to yield right-of-way, unsafe speed, and traffic control disregarded. Moreover, it is a constant to see that in Brooklyn, the area with the largest number of tragic accidents, belongs to ZIP Code 11207. When looking at this area on a map, we observe it has a high density of businesses of different natures but also residential zones. Because of this, we can derive that the areas at ZIP Code 11207 are more prominent to accidents because it is susceptible to be visited by numerous people who don't live there, given the number of businesses; it is a place with a high level of motor vehicle transit as being not that close to Manhattan and not that efficient public transportation available and the exposure of residents to these matters.

Analyzing the Borough with the most tragic crashes

Similarly, the ZIP Code 11207 in Brooklyn repeatedly resulted in being the area with the most significant number of injured people in motor vehicle collisions across the borough. Most of the persons involved in these events were male motorists (drivers). Further, a variable considered to be relevant for the analysis was using the weather condition to see how it influenced the accident occurrence. The weather conditions with the most elevated figures of injuries and deaths during

tragic accidents are rainy (partially cloudy), clear, and partially cloudy (no rain). As seen, none of the previous conditions, given their distribution, can directly justify this number of tragic accidents, and therefore, we can conclude that they are mainly caused by the reported contributing factor instead of an exogenous variable.

Conclusions

In conclusion, we can reject our preconceived hypothesis. The number of motor vehicle collisions has indeed reduced over time; however, its mortality/lethality has dramatically increased, averaging 32% of the tragic crash ratio after the pandemic. The NYC borough constantly presents the highest number of tragic crashes in Brooklyn, where the ZIP Code 11207 shows the highest number of these accidents and injured persons. Moreover, the most frequent contributing factors to accidents in NYC are drivers' inattention, reckless following too closely, unsafe speed, and disregarding traffic control. Weather conditions didn't represent a relevant cause. In addition, the majority of the people involved in accidents were male motorists.

Future work on this matter could include traffic information and criminality to study their relationship with motor vehicle accident occurrence.

Tableau Story

https://public.tableau.com/app/profile/fernando.martinez4669/viz/AnalysisofMotorVehicleCollisionsNYC/AnalysisofMotorVehicleCollisionsinNewYorkCity

Related Work

- Abdul Aziz, H. M., Ukkusuri, S. V., & Hasan, S. (2021). Exploring the determinants of pedestrian–vehicle crash severity in New York City. *Accident Analysis & Prevention*, *50*, 1298-1309. 10.1016/j.aap.2012.09.034
- Shaaban, K., & Ibrahim, M. (2021). Analysis and Identification of Contributing Factors of Traffic Crashes in New York City. *Transportation Research Procedia*, *55*, 1696-1703. 10.1016/j.trpro.2021.07.161
- Zhou, E., Mao, S., & Li, M. (2017). Investigating street accident characteristics and optimal safe route recommendation: A case study of New York City. *25th International Conference on Geoinformatics*, 1-7. *https://doi.org/10.1109/GEOINFORMATICS.2017.8090942*

References

[1] Listín Diario - República Dominicana es el país con mas muertes por accidents de tránsito del mundo (2022): *https://listindiario.com/economia/2022/07/14/730070/republica-dominicana-es-el-pais-con-mas-muertes-por-accidentes-de-transito-en-el-mundo*

[2] Motor Vehicle Collisions - Crashes*:* https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95

[3] Motor Vehicle Collisions - Vehicles*:* https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Vehicles/bm4k-52h4

[4] Motor Vehicle Collisions - Person: https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Person/f55k-p6yu

[5] New York City - Weather History: https://www.visualcrossing.com/weather-history/New+York%20City%2CUSA