THE GREAT VISUALIZERS

Bimal Dawadi
Marc Julot
Sushmaben Patel
Mehnaz Afrose

PROBLEM STATEMENT

- Our research is very important for New Yorkers to see (New Yorkers being residents of NYC, not the whole state). New Yorkers do not only commute with cars, but they also commute with trains. In NYC, many people still walk and are known to ride bikes. If this research deems that the risk of death and accidents are likely, state legislate may find solutions that will allow for commuters to utilize trains and bussing instead of driving. If walking and cycling has a lower death rate, state legislature may recommend more walking and cycling in the city. This project may enable the New York legislature to discount public transportation and increase the cost to non-public transportation. If consumers understand the risk of being in an accident and dying or getting injured while behind a vehicle, they be more likely to be deterred from driving.

DATA DESCRIPTION

- Any death matters when it comes to people being able to commute. If people can easily visualize the data from accidents within given time in NYC, they may be deterred or emboldened to drive. We used different visualizations tools like box plot, bar chart, pie chart, multi-line plot and other descriptive tools for this data so commuters can easily see a visual representation of the data and choose whether to take a certain risk based on the count of the accident type. This doesn't only help commuters but also the analysts and researchers to make various decisions easily.

DATA DESCRIPTION CONTINUE

#	Name	Description	Type
1	CRASH DATE	Occurrence date of collision	DATE
2	CRASH TIME	Occurrence date of collision	TEXT
3	BOROUGH	Borough where collision occurred	TEXT
4	ZIP CODE	Borough where collision occurred	TEXT
5	LATITUDE	Latitude coordinates for Global Coordinate System	NUMBER
6	LONGITUDE	Longitude coordinates for Global Coordinate System	NUMBER
7	LOCATION	Latitude, Longitude pair	LOCATION
8	ON STREET NAME	Street on which the collision occurred	TEXT
9	CROSS STREET NAME	Nearest cross street to the collision	TEXT
10	OFF STREET NAME	Street address if known	TEXT
11	NUMBER OF PERSONS INJURED	Number of persons injured	NUMBER
12	NUMBER OF PERSONS KILLED	Number of persons killed	NUMBER
13	NUMBER OF PEDESTRIANS INJURED	Number of pedestrians injured	NUMBER
14	NUMBER OF PEDESTRIANS KILLED	Number of pedestrians killed	NUMBER
15	NUMBER OF CYCLIST INJURED	Number of cyclists injured	NUMBER
16	NUMBER OF CYCLIST KILLED	Number of cyclists killed	NUMBER
17	NUMBER OF MOTORIST INJURED	Number of vehicle occupants injured	NUMBER
18	NUMBER OF MOTORIST KILLED	Number of vehicle occupants killed	NUMBER
19	CONTRIBUTING FACTOR VEHICLE 1	Factors contributing to the collision for designated vehicle	TEXT
20	CONTRIBUTING FACTOR VEHICLE 2	Factors contributing to the collision for designated vehicle	TEXT
21	CONTRIBUTING FACTOR VEHICLE 3	Factors contributing to the collision for designated vehicle	TEXT
22	CONTRIBUTING FACTOR VEHICLE 4	Factors contributing to the collision for designated vehicle	TEXT
23	CONTRIBUTING FACTOR VEHICLE 5	Factors contributing to the collision for designated vehicle	TEXT
24	COLLISION_ID	Unique record code generated by system.	NUMBER
25	VEHICLE TYPE CODE 1	Type of vehicle based on the selected vehicle category	TEXT
26	VEHICLE TYPE CODE 2	Type of vehicle based on the selected vehicle category	TEXT
27	VEHICLE TYPE CODE 3	Type of vehicle based on the selected vehicle category	TEXT
28	VEHICLE TYPE CODE 4	Type of vehicle based on the selected vehicle category	TEXT
29	VEHICLE TYPE CODE 5	Type of vehicle based on the selected vehicle category	TEXT

DATA PREPROCESSING

- Columns in our data set

```
In [62]: df.dtypes

Out[62]: CRASH DATE                          object
         CRASH TIME                          object
         BOROUGH                             object
         ZIP CODE                            object
         LATITUDE                           float64
         LONGITUDE                          float64
         LOCATION                            object
         ON STREET NAME                      object
         CROSS STREET NAME                   object
         OFF STREET NAME                     object
         NUMBER OF PERSONS INJURED          float64
         NUMBER OF PERSONS KILLED           float64
         NUMBER OF PEDESTRIANS INJURED        int64
         NUMBER OF PEDESTRIANS KILLED         int64
         NUMBER OF CYCLIST INJURED            int64
         NUMBER OF CYCLIST KILLED             int64
         NUMBER OF MOTORIST INJURED           int64
         NUMBER OF MOTORIST KILLED            int64
         CONTRIBUTING FACTOR VEHICLE 1       object
         CONTRIBUTING FACTOR VEHICLE 2       object
         CONTRIBUTING FACTOR VEHICLE 3       object
         CONTRIBUTING FACTOR VEHICLE 4       object
         CONTRIBUTING FACTOR VEHICLE 5       object
         COLLISION_ID                         int64
         VEHICLE TYPE CODE 1                 object
         VEHICLE TYPE CODE 2                 object
         VEHICLE TYPE CODE 3                 object
         VEHICLE TYPE CODE 4                 object
         VEHICLE TYPE CODE 5                 object
         dtype: object
```

DATA PREPROCESSING CONTINUE

Data Statistics

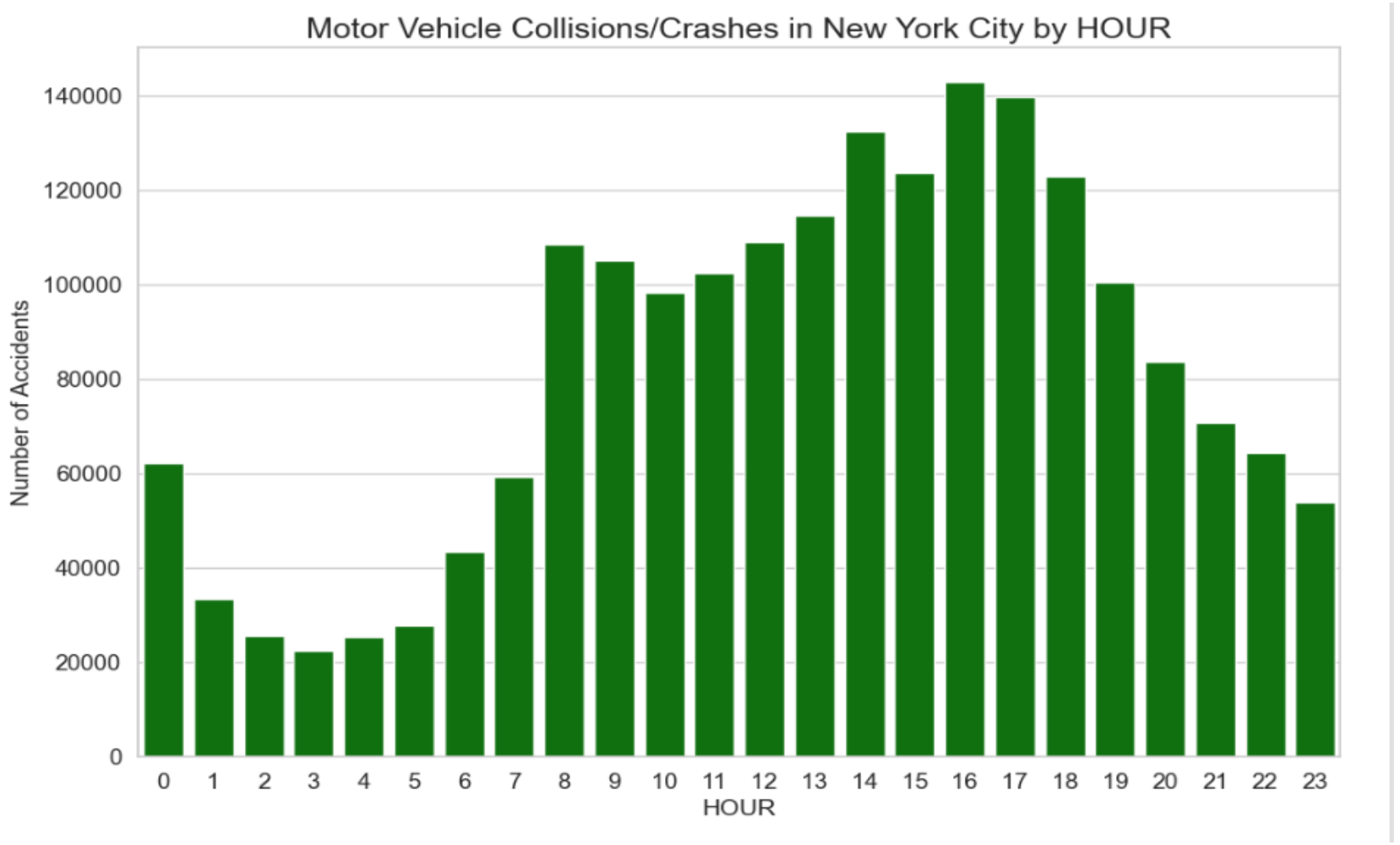
	MAX	MIN	AVG
Number of people injured	40	0	0.32
Number of people killed	8	0	0.00015
Number of pedestrians injured	27	0	0.0054
Number of pedestrians killed	6	0	0.0000712
Number of cyclist injured	3	0	0.00289
Number of cyclist killed	2	0	0.0000124
Number of motorist injured	40	0	0.0234
Number of motorist killed	4	0	0.0000623

DATA PREPROCESSING CONTINUE

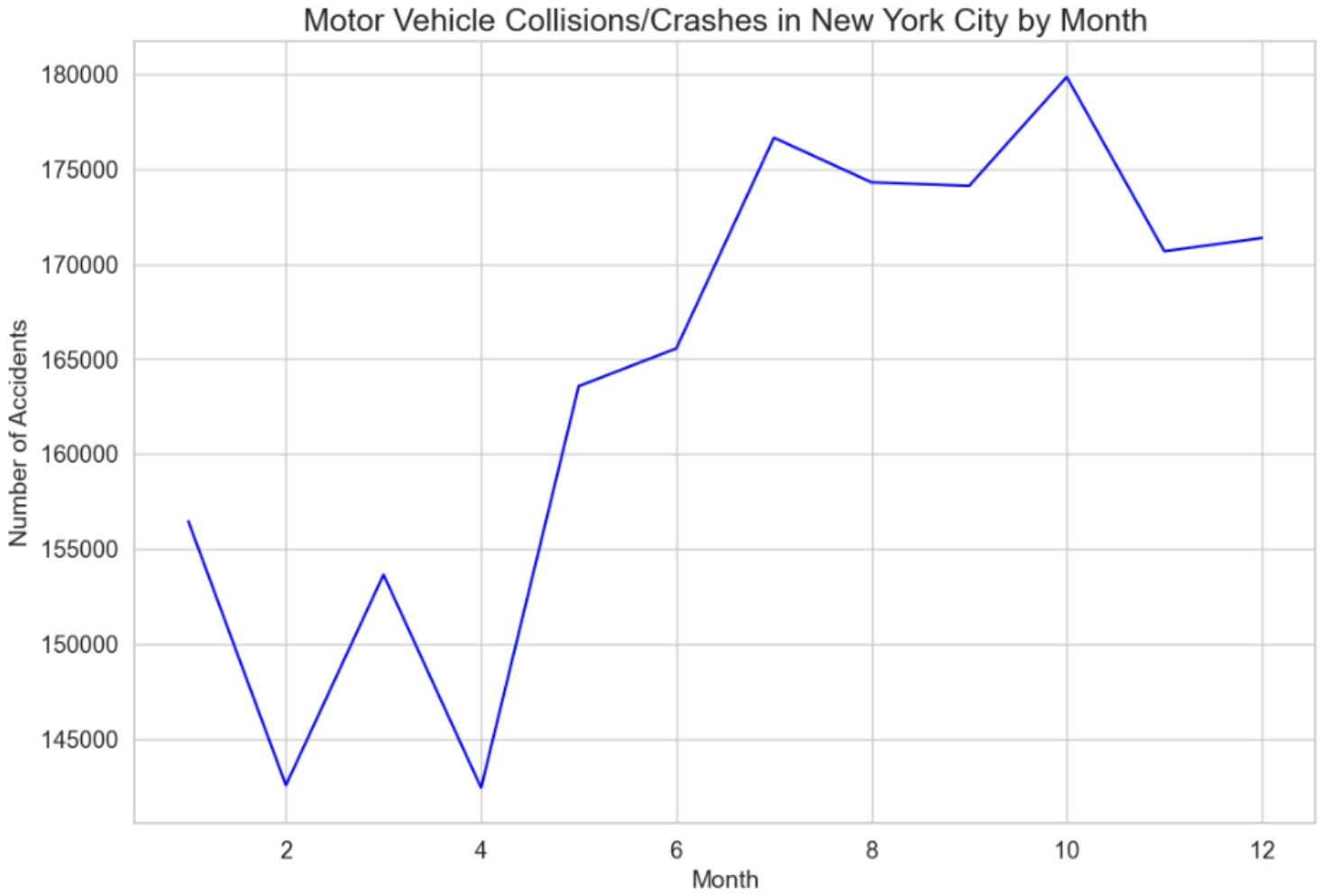
- After removing the rows with Columns (Number of Persons Injured, Number of persons killed, Number of pedestrians injured, Number of pedestrians killed, Number of cyclist injured, Number of cyclist killed, Number of Motorist injured, Number of Motorist killed) with '0' value, dataset size reduced to 250204 which means it lost 77% of the data therefor not removing rows for "0" value.

- To clean the data, we have removed the rows with columns (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) with null value.

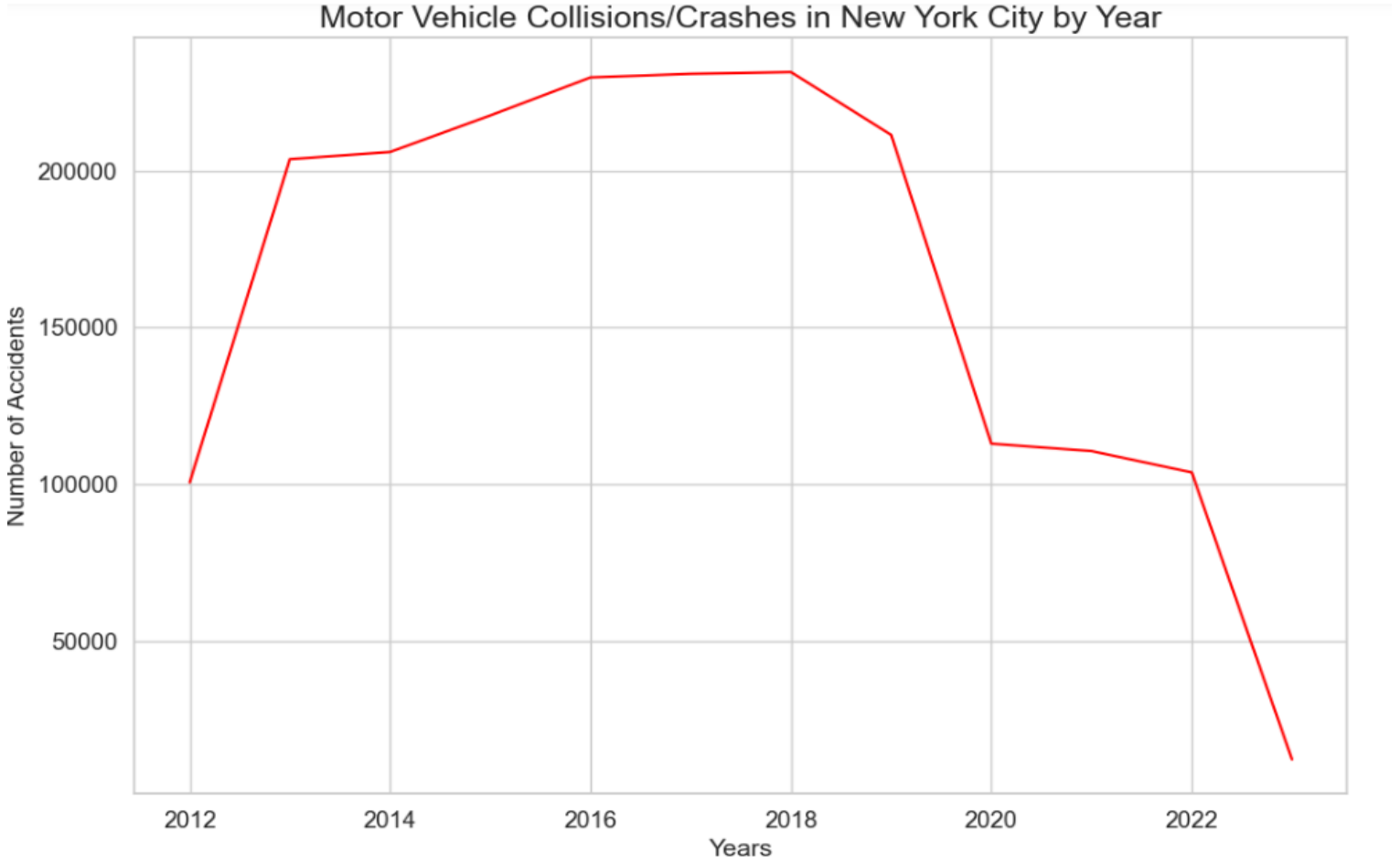
MOTOR VEHICLE COLLISIONS BY HOUR IN NEW YORK CITY



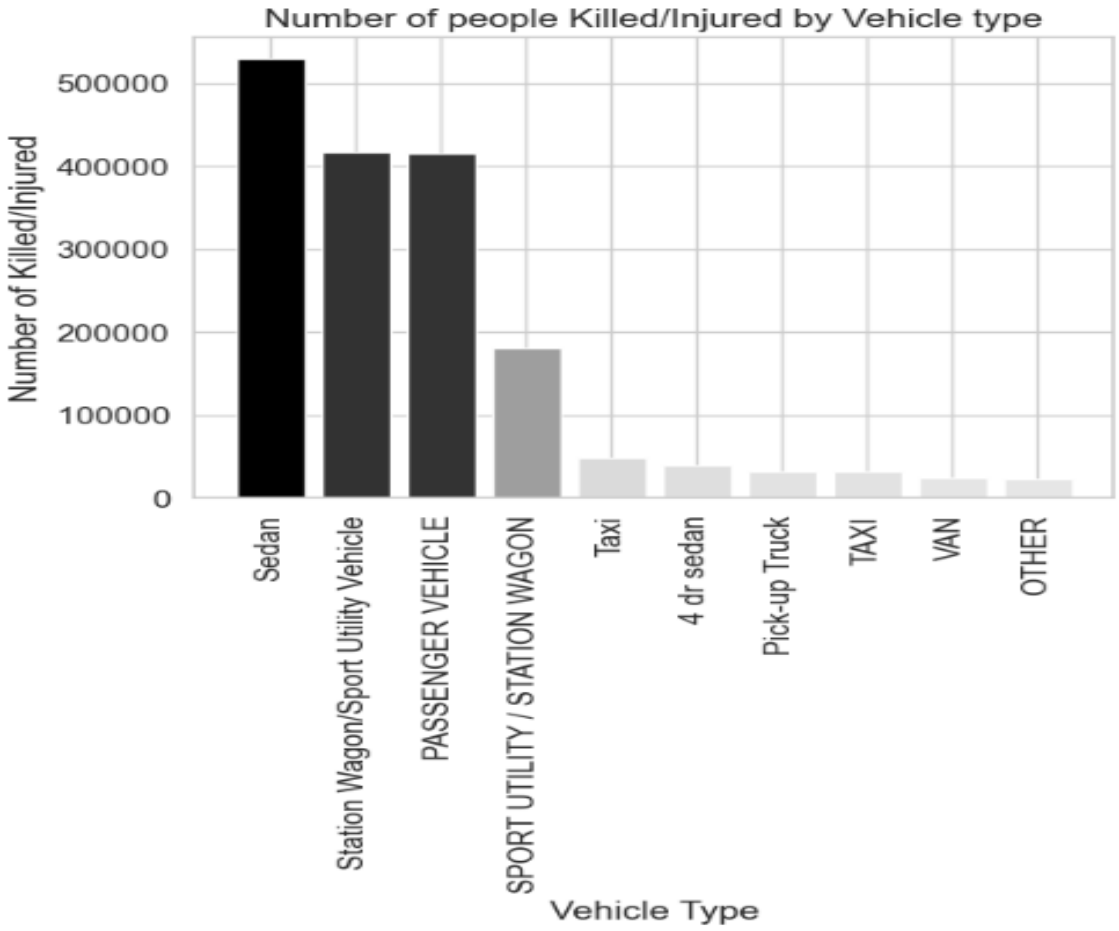
MOTOR VEHICLE COLLISIONS BY MONTH IN NEW YORK CITY



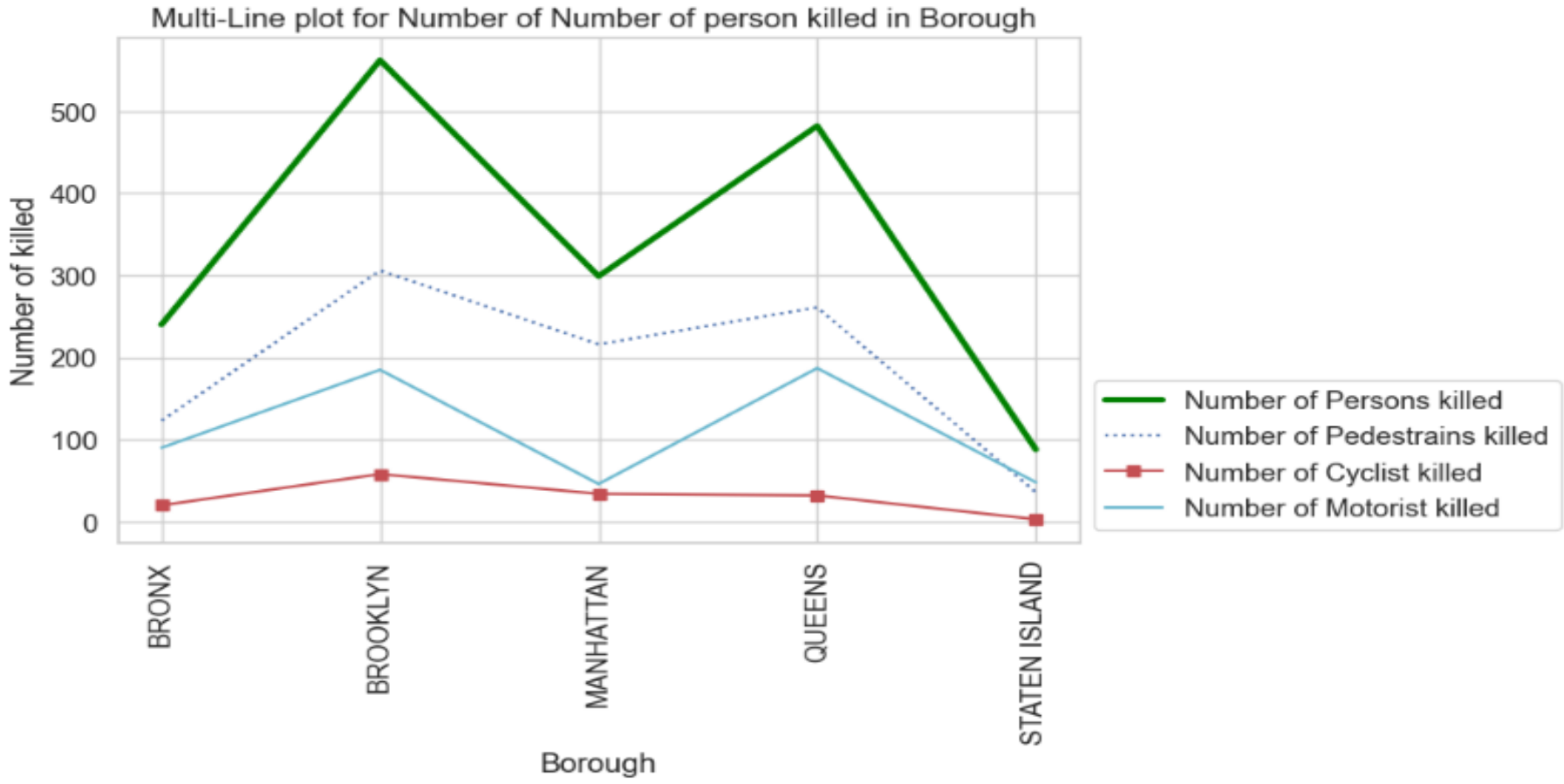
MOTOR VEHICLE COLLISIONS BY YEAR IN NEW YORK CITY



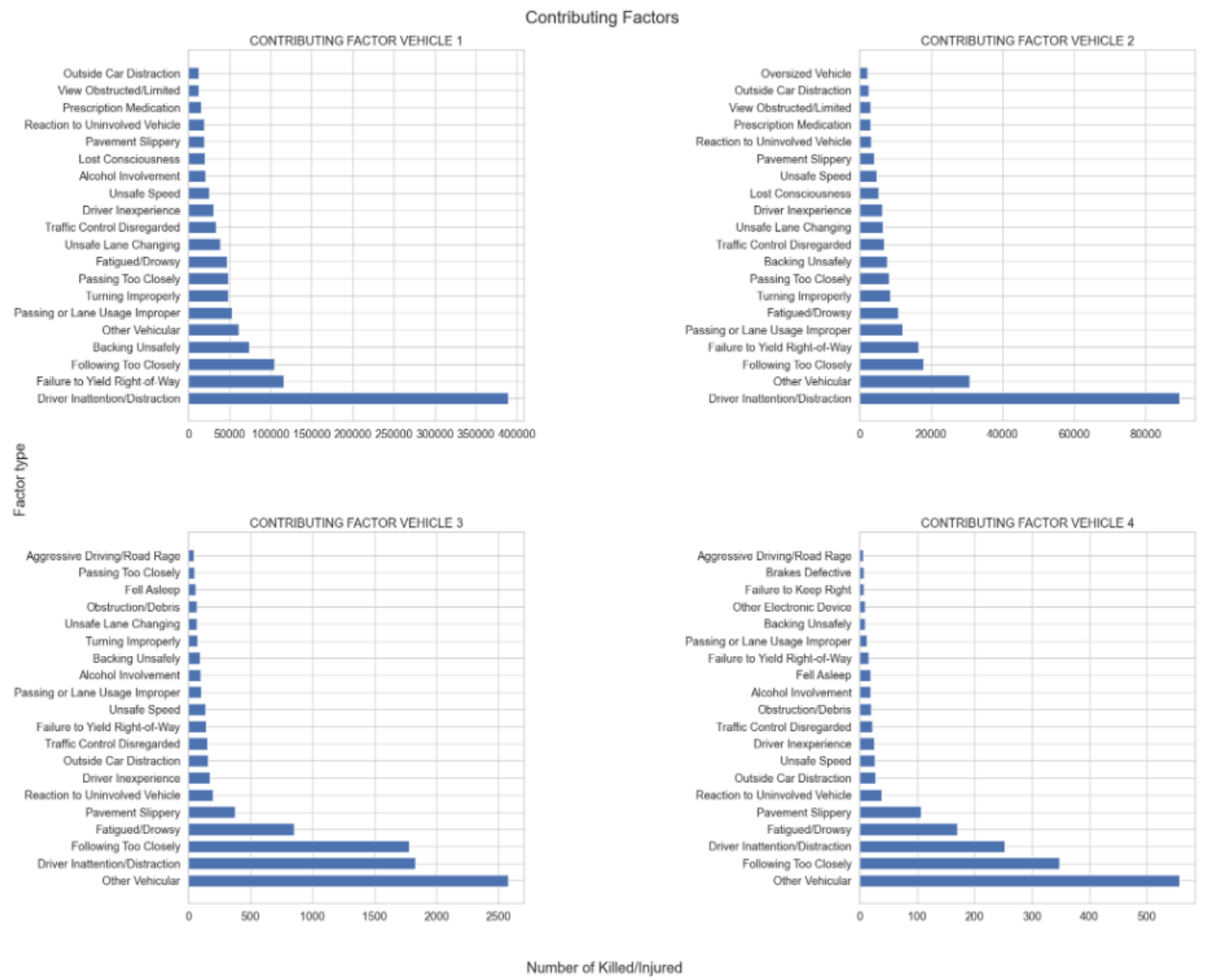
PEOPLE KILLED/INJURED BY VEHICLE TYPE



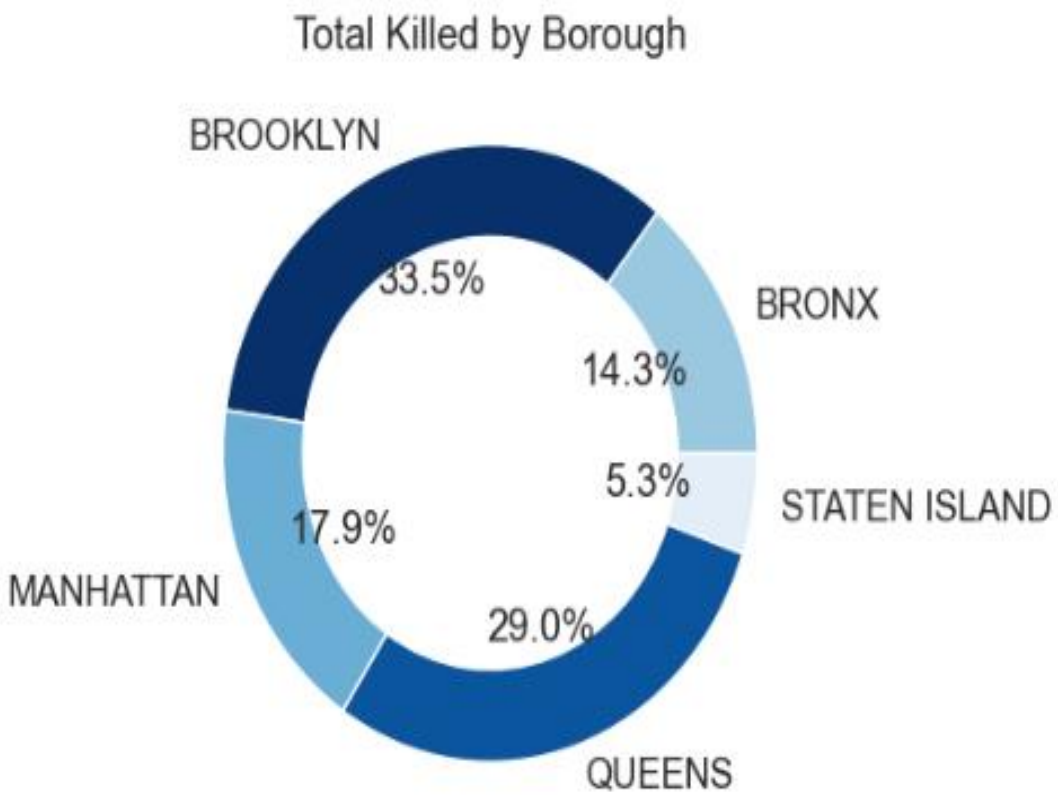
PEOPLE KILLED IN BOROUGH

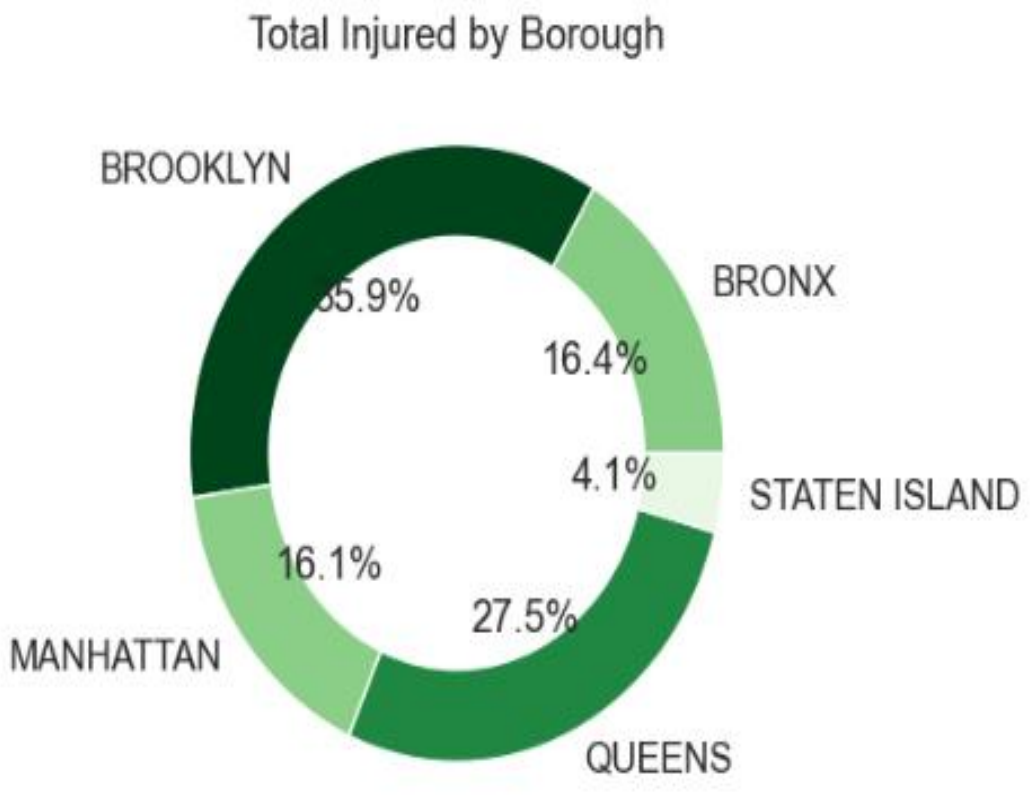


MOST COMMON ACCIDENT CONTRIBUTING FACTORS

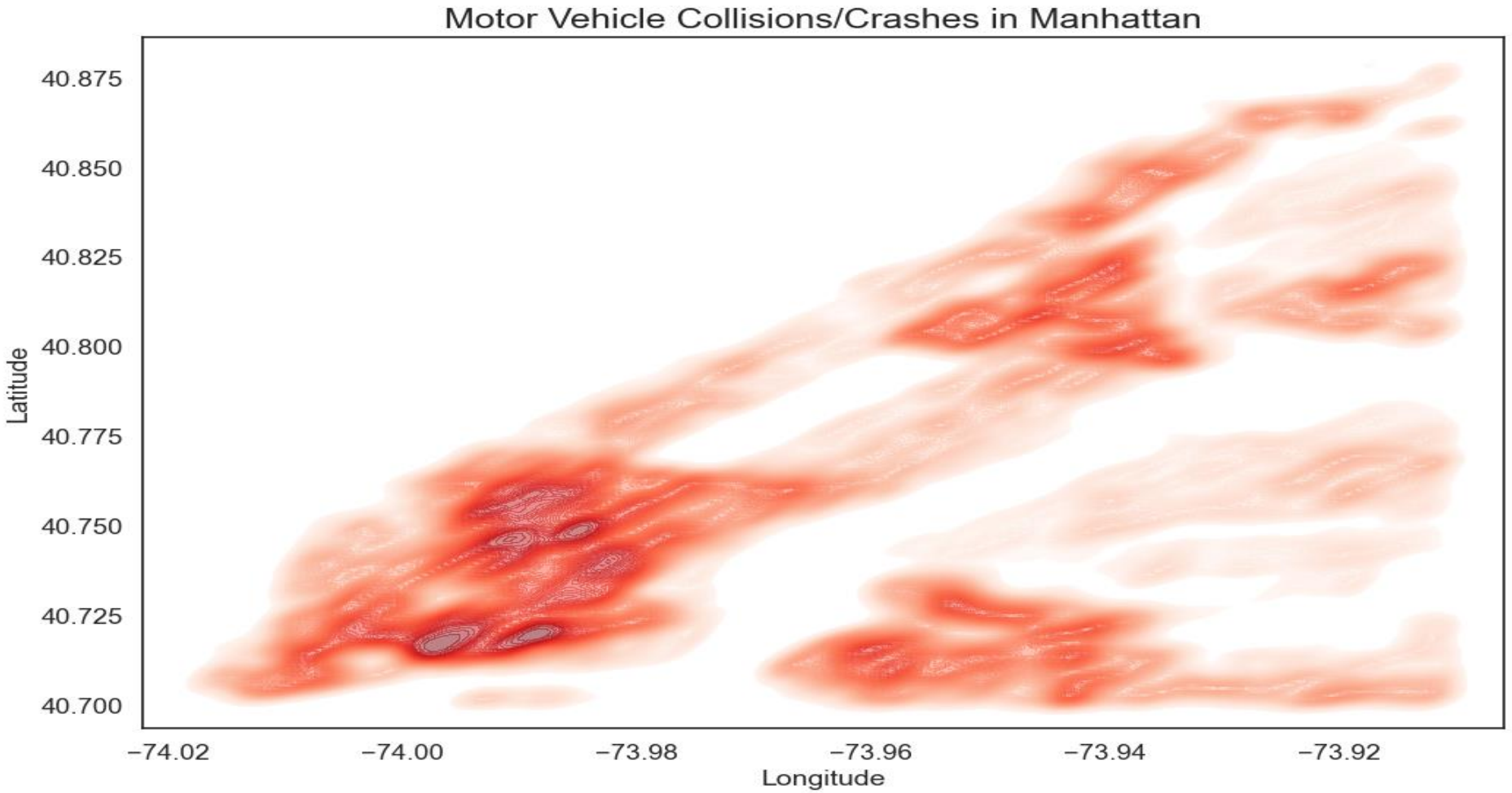


PEOPLE KILLED/INJURED IN BOROUGH





MOTOR VEHICLE COLLISION HEATMAP IN MANHATTAN CITY



CONCLUSION

- Our greatest founding is that viewers of our visualization might rather live in Staten Island where death was less likely. Viewers may not want a Sedan as it was the vehicle type to more likely to have an accident. If people choose to drive, they should choose to not be distracted while driving as that drastically increase the risk of death or injuries. People may choose to use bicycles as injury and death are the lowest. New Yorkers should be happy to know that since 2012, the number of accidents has been at an all-time low. New Yorkers should understand that October is the month that has the most accidents. New Yorkers should also be aware that 5 pm is the hour that accidents are the highest.

CONCLUSION CONT.

- We assume that in the future people will be more likely to cycle or walk since cycling and walking had the least chance of being in a car accident or death. Upon review, the state legislature may recommend more walking and cycling in the city. We hope that a project like this may enable the New York legislature to discount public transportation and increase the cost to non-public transportation. If consumers understand the amount of accident and death happening in their sate due to car accidents, they are more likely to be deterred from driving. Although it is just a small sample and may not be representative of the whole state, this data gave a glimpse into reasons behind different collisions, as well as whether the collision resulted in death.