Are Accidents Really Accidents?

Analyzing Motor Vehicle Accidents in NYC

DATASCI200 – Project 2

April 2024

Yanling Liu, Elizabeth (Lily) Maio, Diego Moss

Background

~40,000 Traffic Deaths/Year

#1 cause of death 5 - 34 year olds

Background

Are they accidents or choices?

Dataset

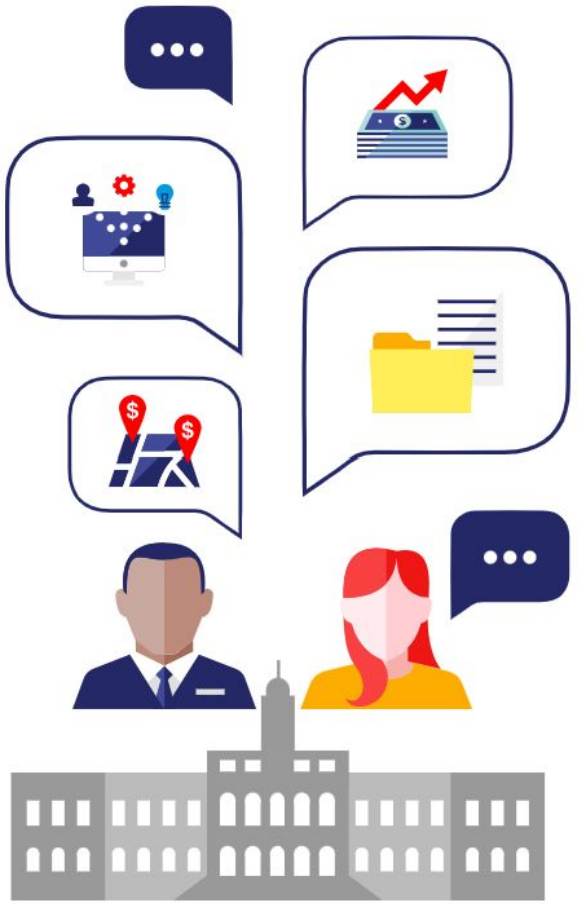
- NYC Open Data: NYC Motor Vehicle Incidents

[6]:

Data shape: (2077346, 29)

	CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS STREET NAME
0	2021-09-11	2:39	NaN	NaN	NaN	NaN	NaN	WHITESTONE EXPRESSWAY	20 AVENUE
1	2022-03-26	11:45	NaN	NaN	NaN	NaN	NaN	QUEENSBORO BRIDGE UPPER	NaN
2	2022-06-29	6:55	NaN	NaN	NaN	NaN	NaN	THROGS NECK BRIDGE	NaN
3	2021-09-11	9:35	BROOKLYN	11208.0	40.667202	-73.866500	(40.667202, -73.8665)	NaN	NaN
4	2021-12-14	8:13	BROOKLYN	11233.0	40.683304	-73.917274	(40.683304, -73.917274)	SARATOGA AVENUE	DECATUR STREET

5 rows × 29 columns



Surprises in the Data

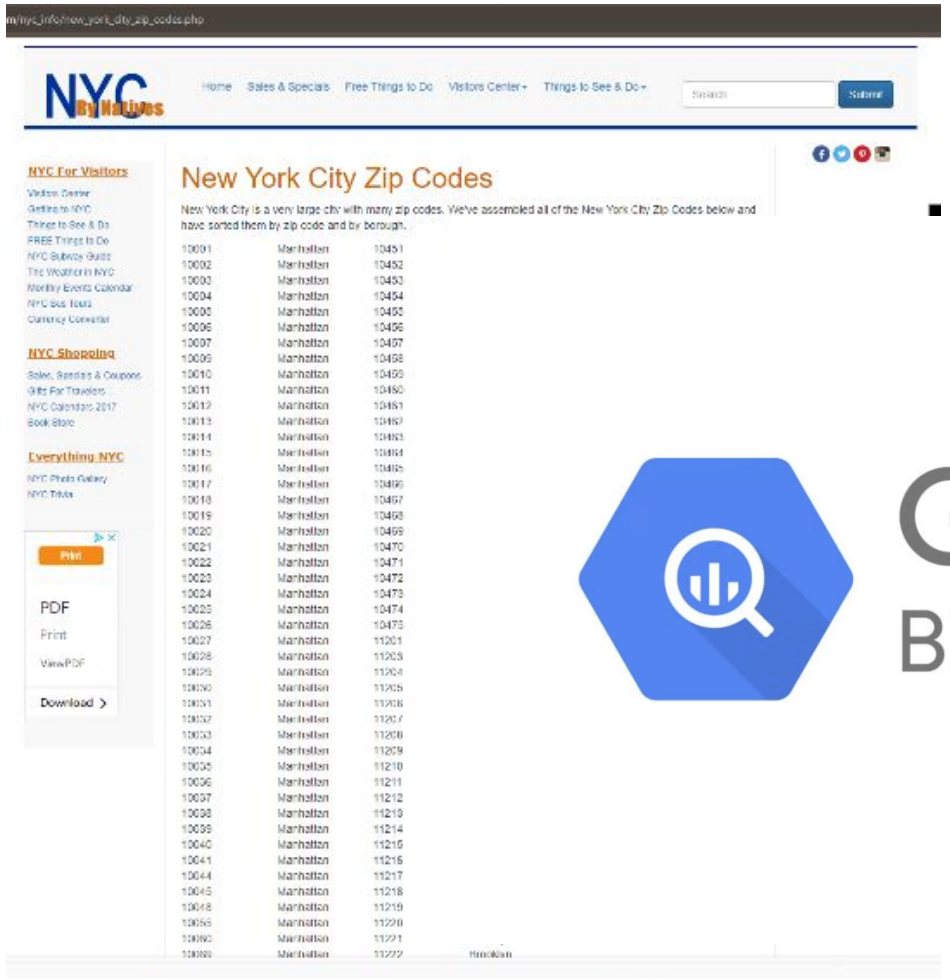
Total Sample: 1.9M

BEFORE

N/As: 0.6M ~ 30%

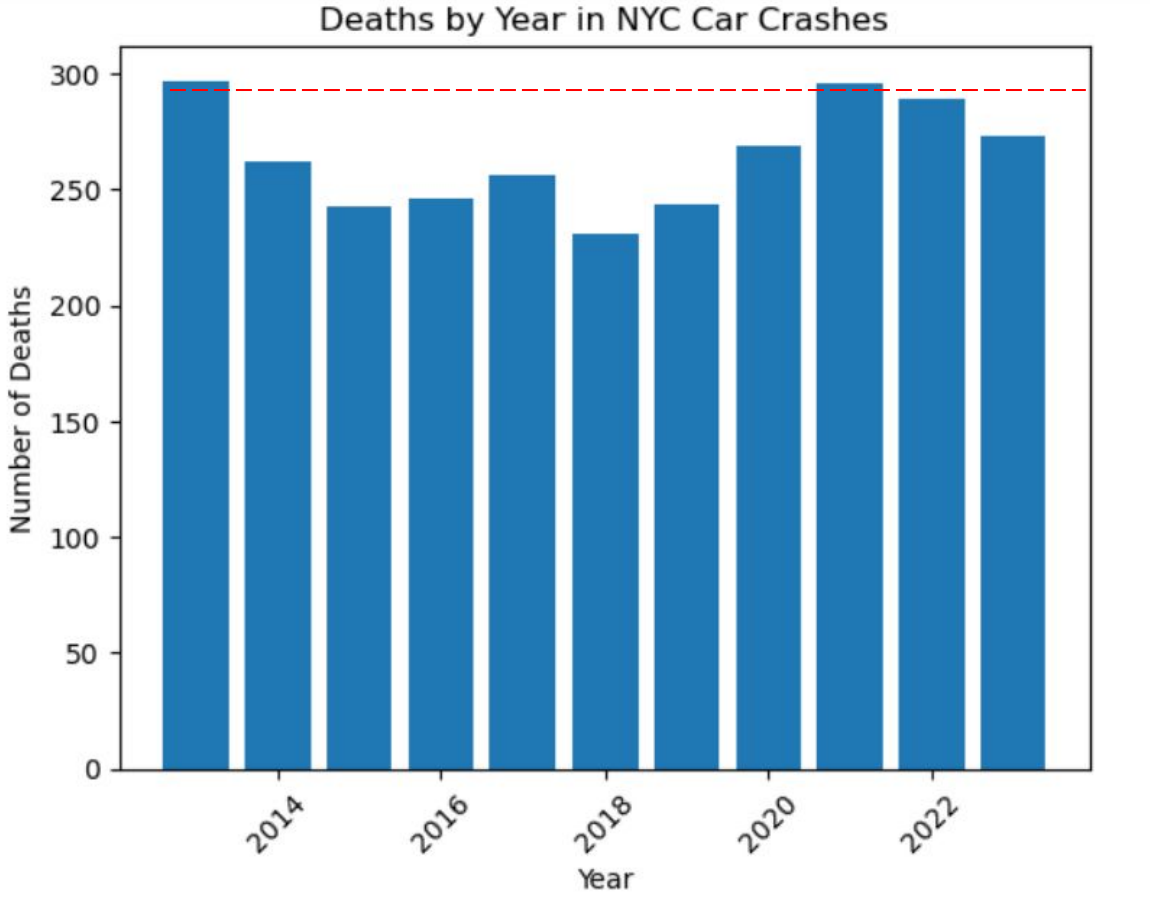
AFTER

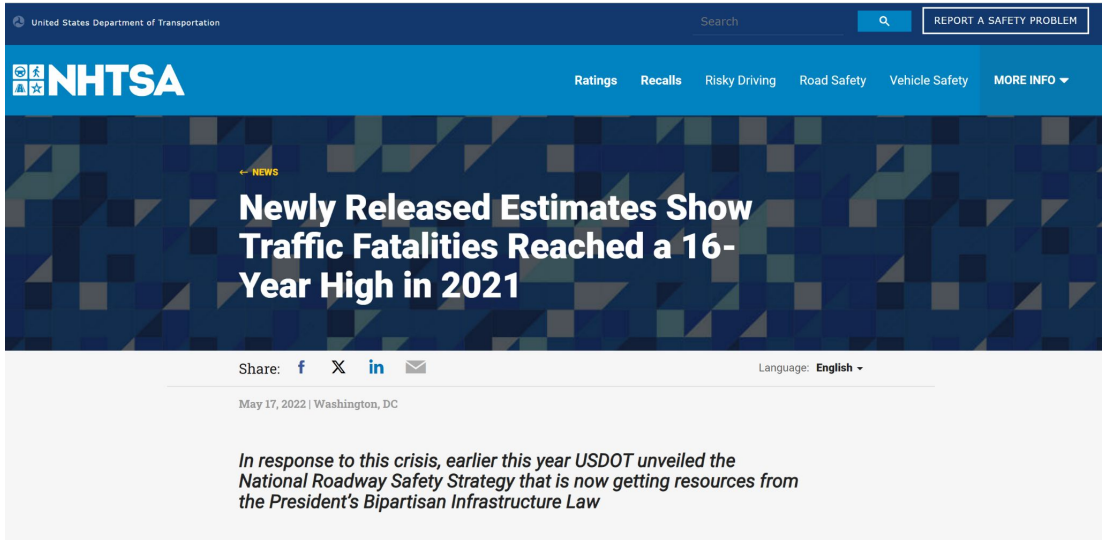
N/As: 0.2M ~ 10%



Google Big Query

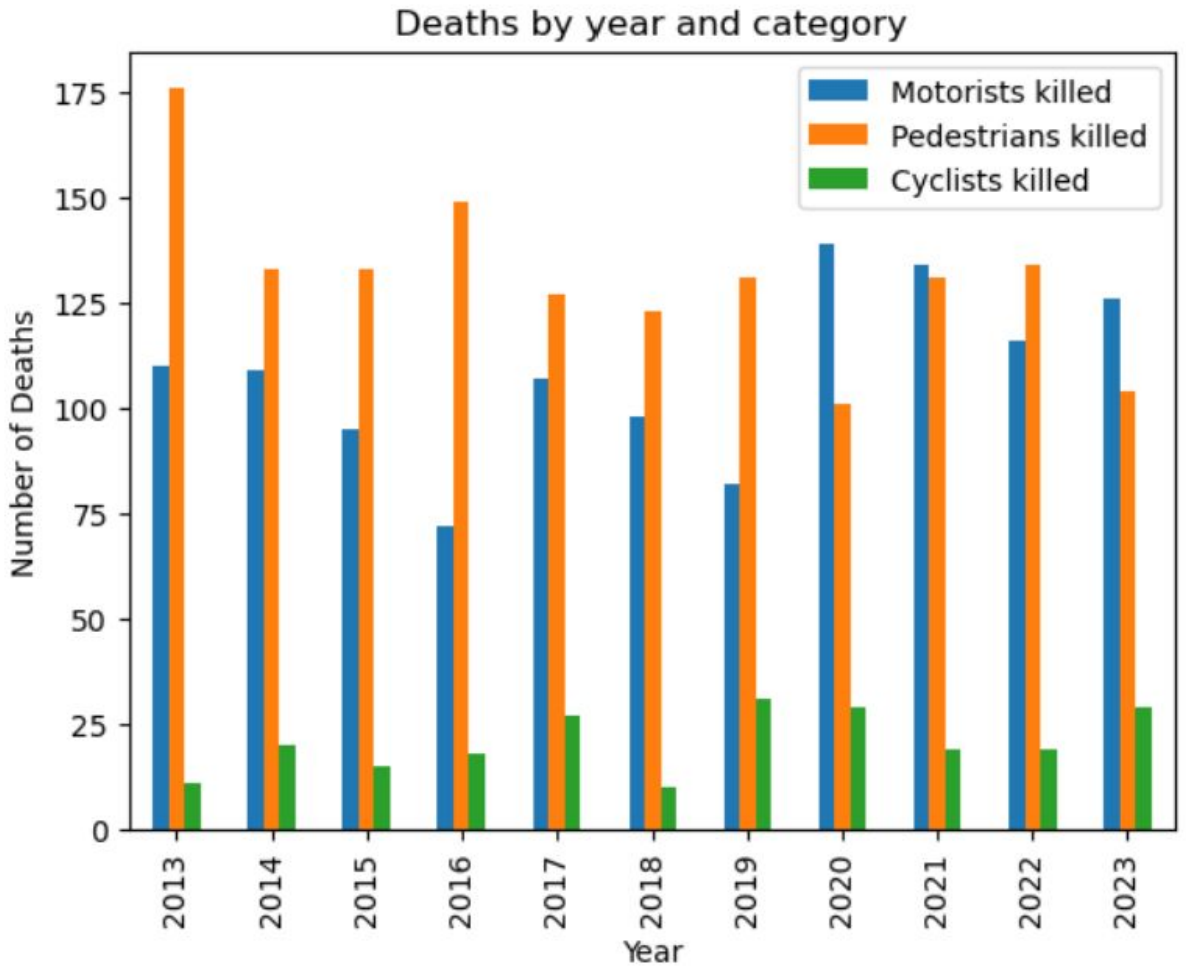
Initial Exploration/Data Validation



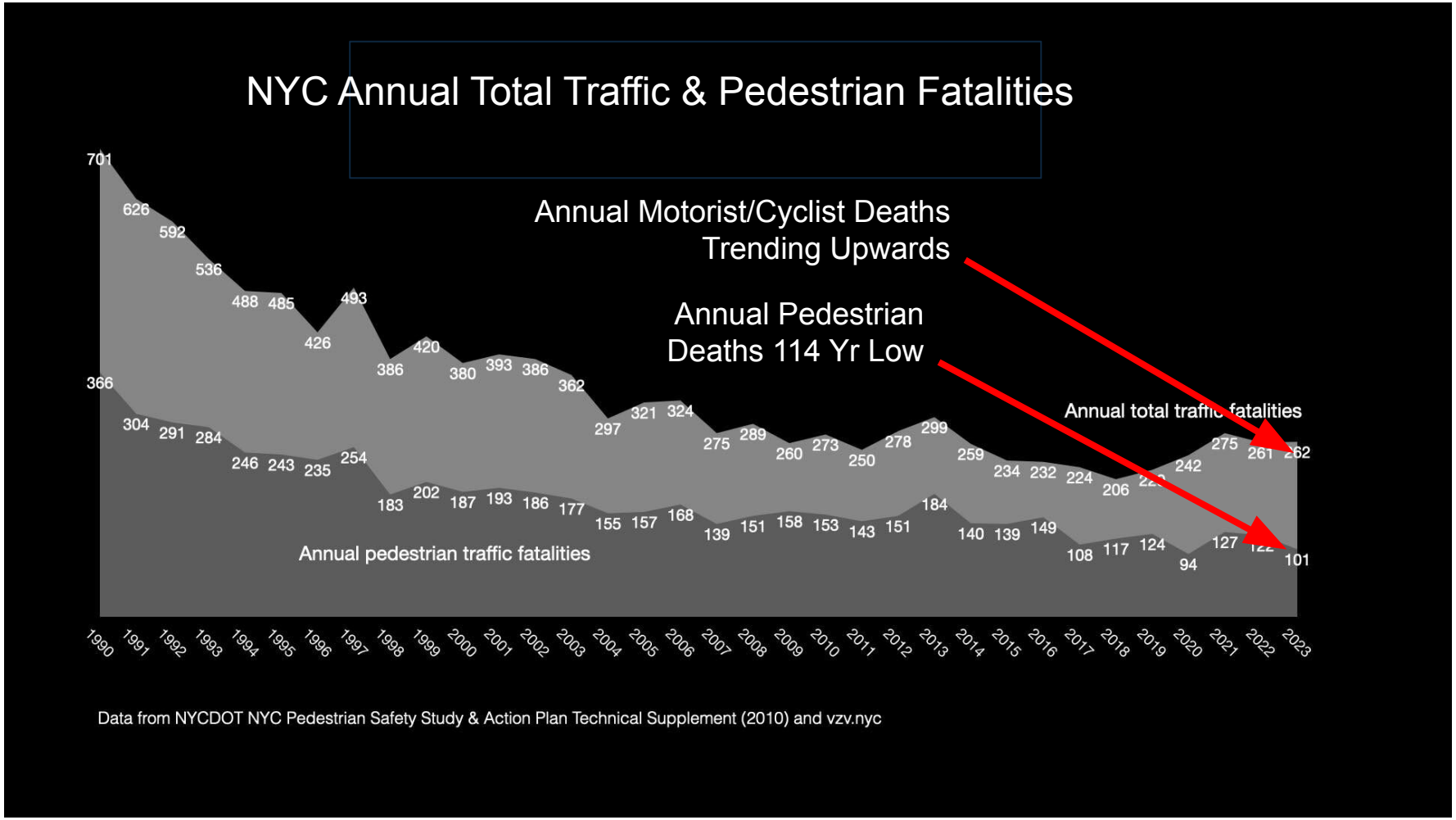


Initial Exploration/Data Validation

- Pedestrian deaths are trending down
- 2023 pedestrian deaths lowest seen in the dataset.
- Motorist and cyclist deaths are trending upwards.

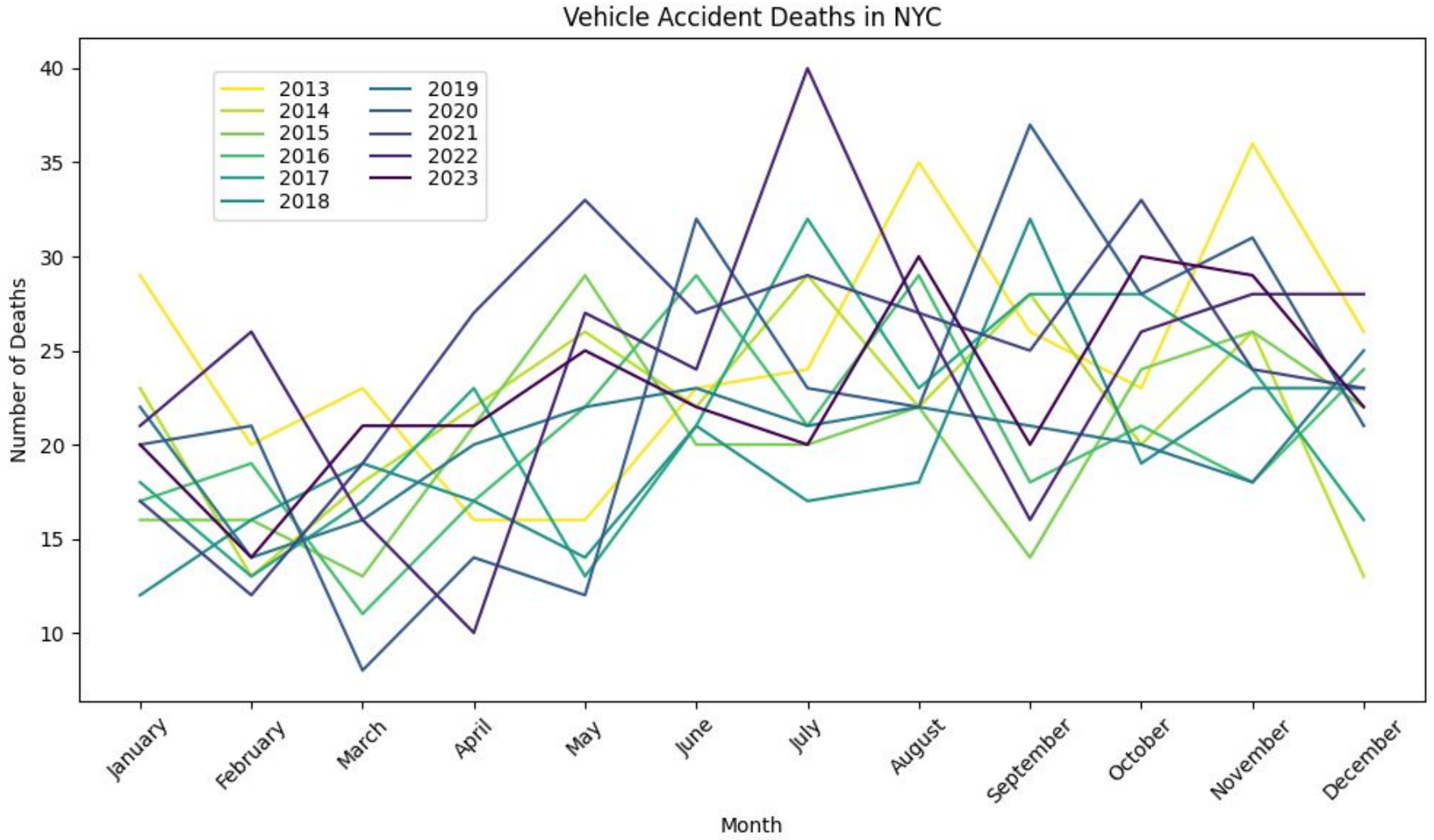


Initial Exploration/Data Validation



Bloomberg

Initial Exploration/Data Validation



Initial Exploration/Data Validation

TRANSPORTATION ALTERNATIVES

About Get Involved Resources News Support Events DONATE

Press Release Apr 19, 2022

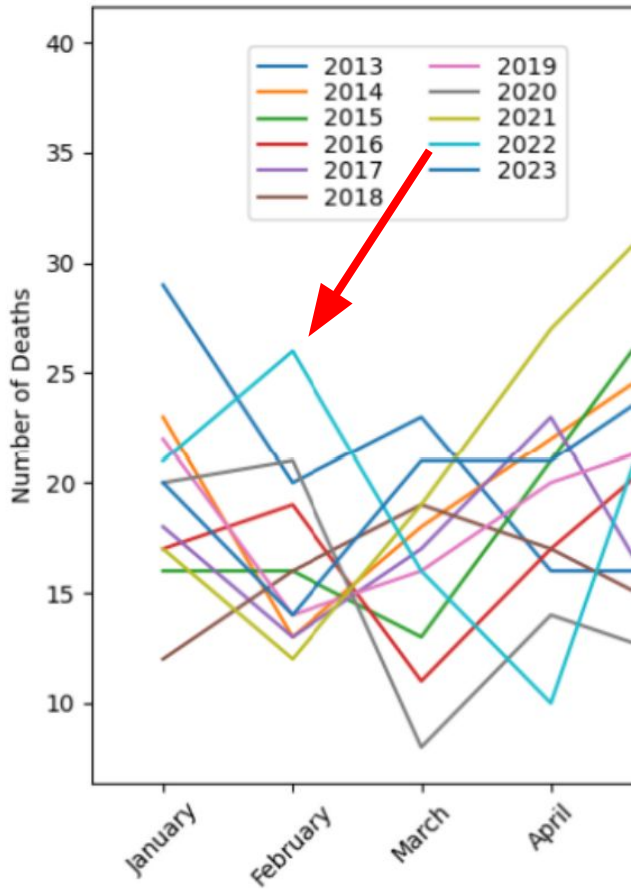
New Data Shows 44 Percent Increase in Traffic Fatalities During First Three Months of 2022, Deadliest Start to Any Year Since Vision Zero Began in 2014

Crashes killed more children during this quarter than any first quarter since before Vision Zero began

Traffic fatalities on track to rise for fourth year in a row, a first since at least 1990

Advocates urge city to fully fund NYC Streets Plan, Albany to grant home rule and pass entire Crash Victim Rights and Safety Act this session

NEW YORK — The first quarter of 2022 was the deadliest start to any year since



Analysis

We seek to understand the following questions:

1. Which borough has the highest occurrences of injuries and death from vehicle collisions? (including pedestrian, cyclist, persons overall)

2. What vehicle type is more common to cause deaths in a collision?

3. What's the correlation between characteristics (i.e. Locations, time of day, seasons in a year, vehicle type etc.) in the data and likelihood of death in a collision? What traits contribute to deaths in a collision if any?

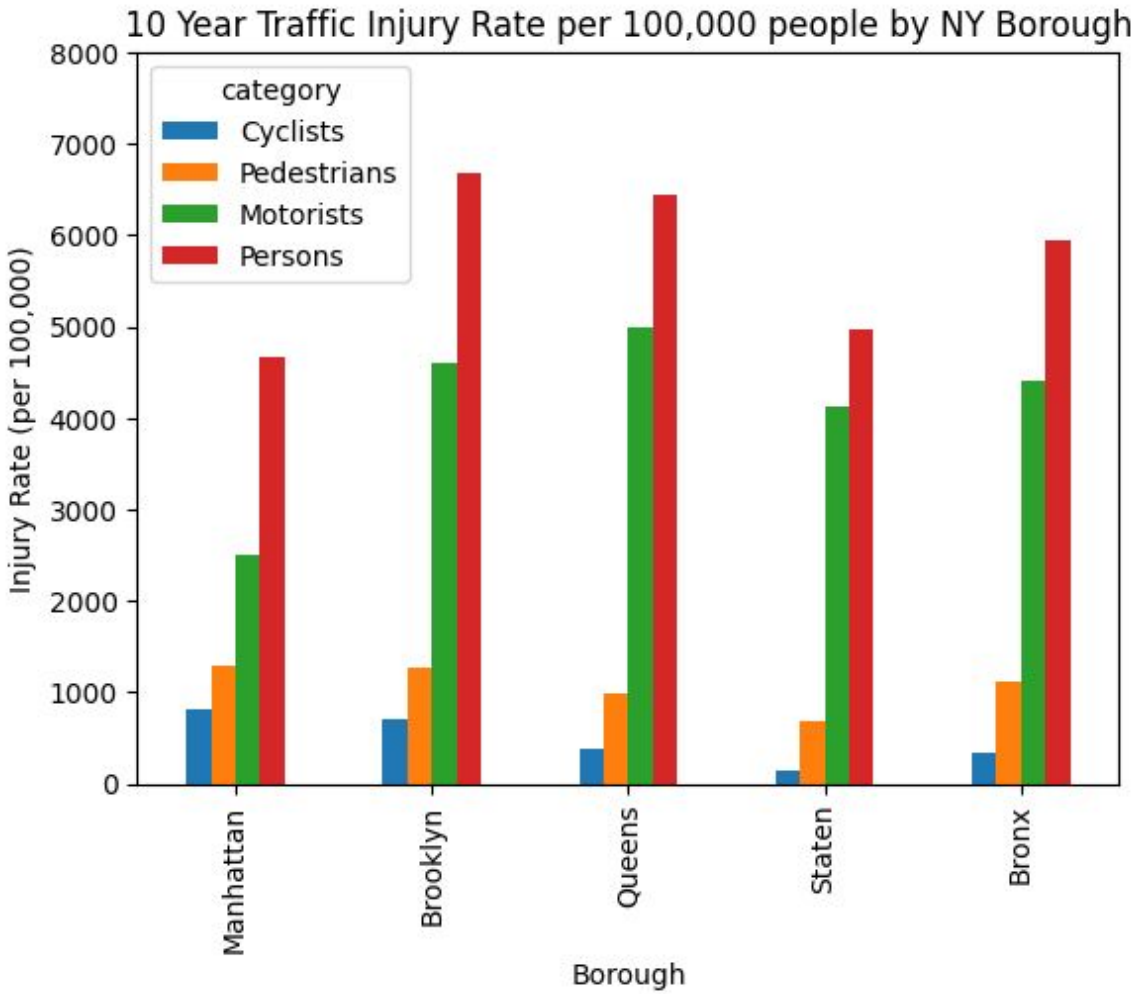
4. Is there any correlation between median income of a zip code and number of deaths from collisions?

Question 1: Borough Differences in Traffic Collisions

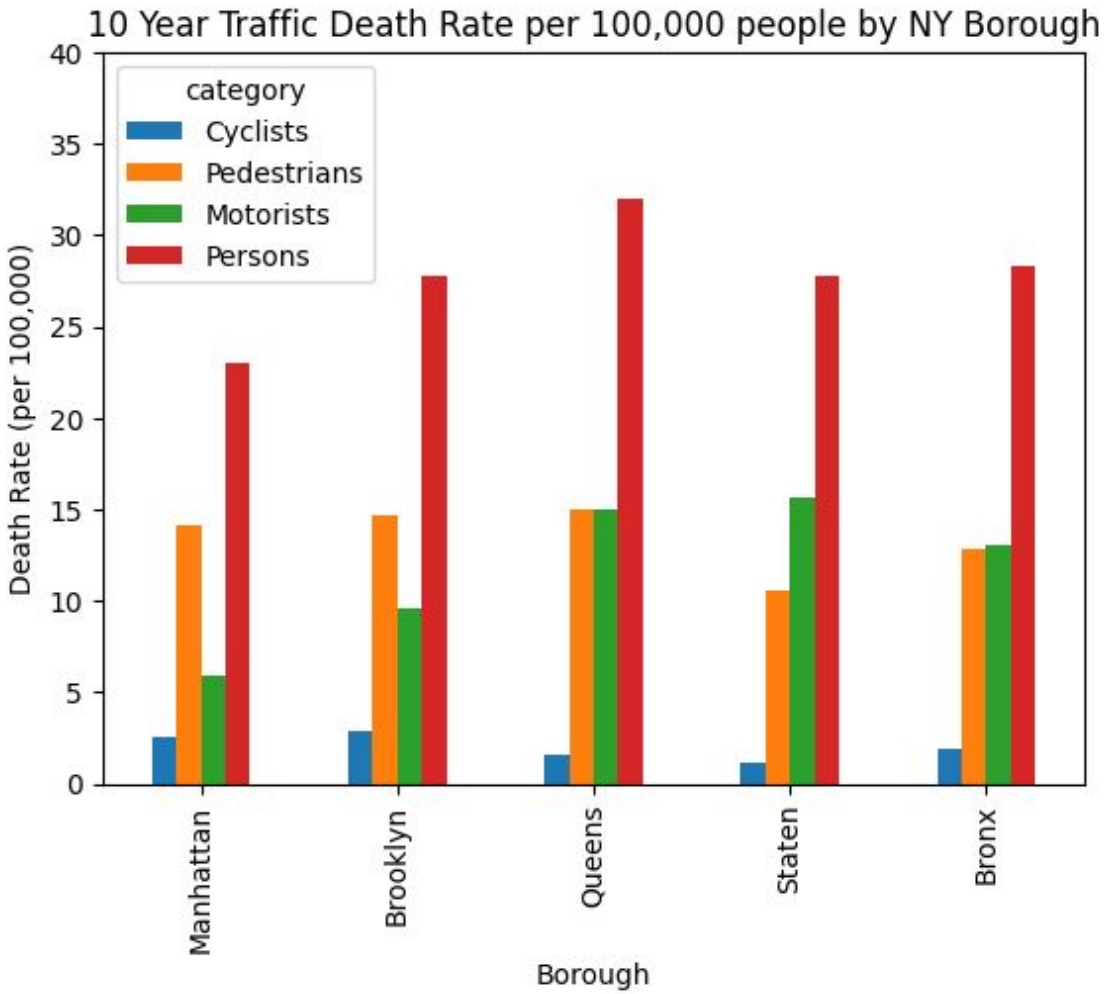
Which borough has the highest occurrences of injuries and death from vehicle collisions? (including pedestrian, cyclist, persons overall)

Question 1: Borough Differences in Traffic Collisions

INJURY



DEATH



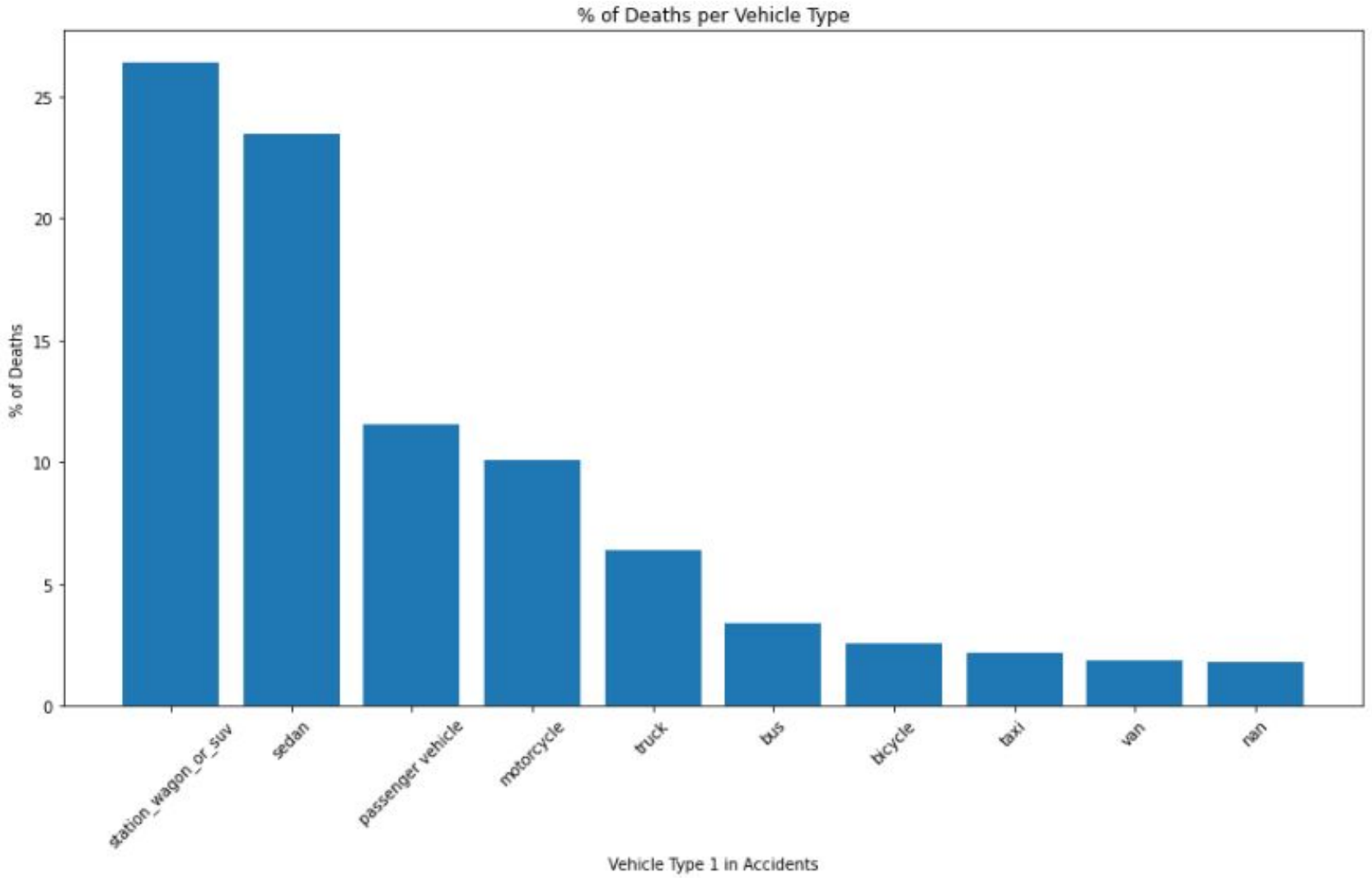
Question 2

- What vehicle type and contribution factor is more common to cause deaths in a collision? What about collisions?

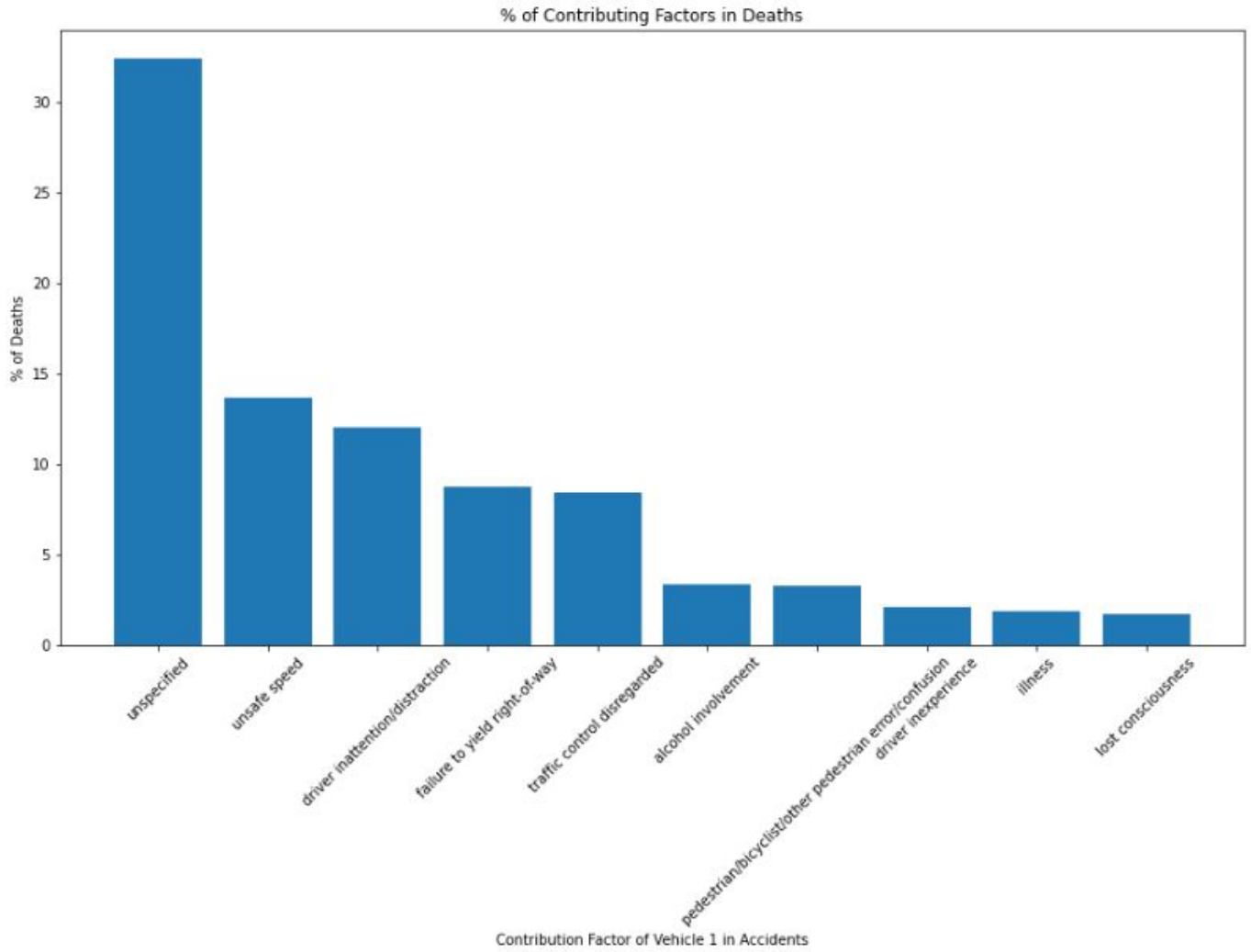
Vehicle Categorization:

- sedan: values that contain “sedan”
- station_wagon_or_suv: values that contain “station wagon”, “sport utility”, “suv”
- truck: values that contain “truck”
- van: values that contain “van”
- motorcycle: values that contain “motorcycle”, “motorbike”
- bike: values that contain “bike”, “bicycle”

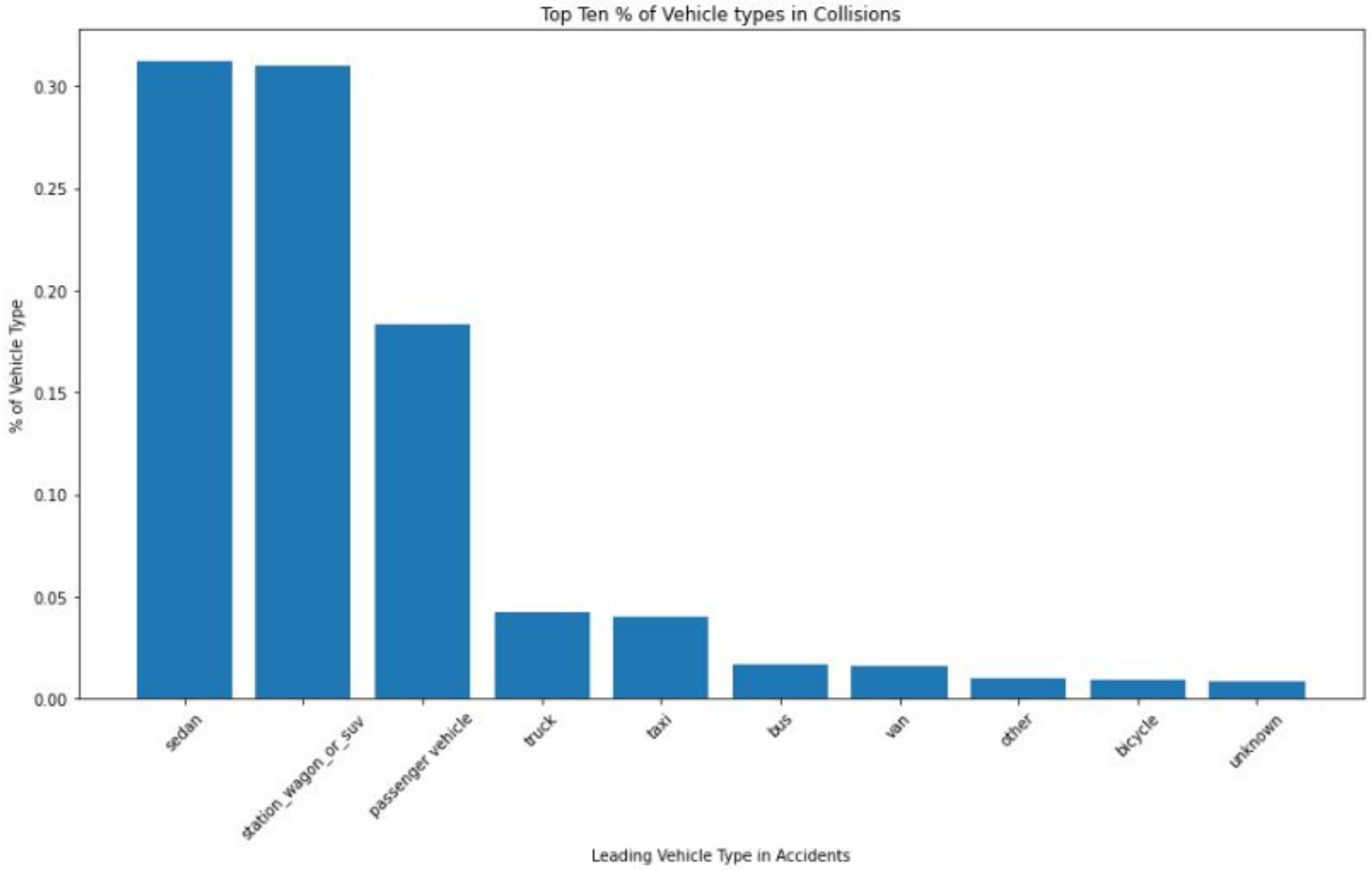
% of Deaths per Vehicle Type



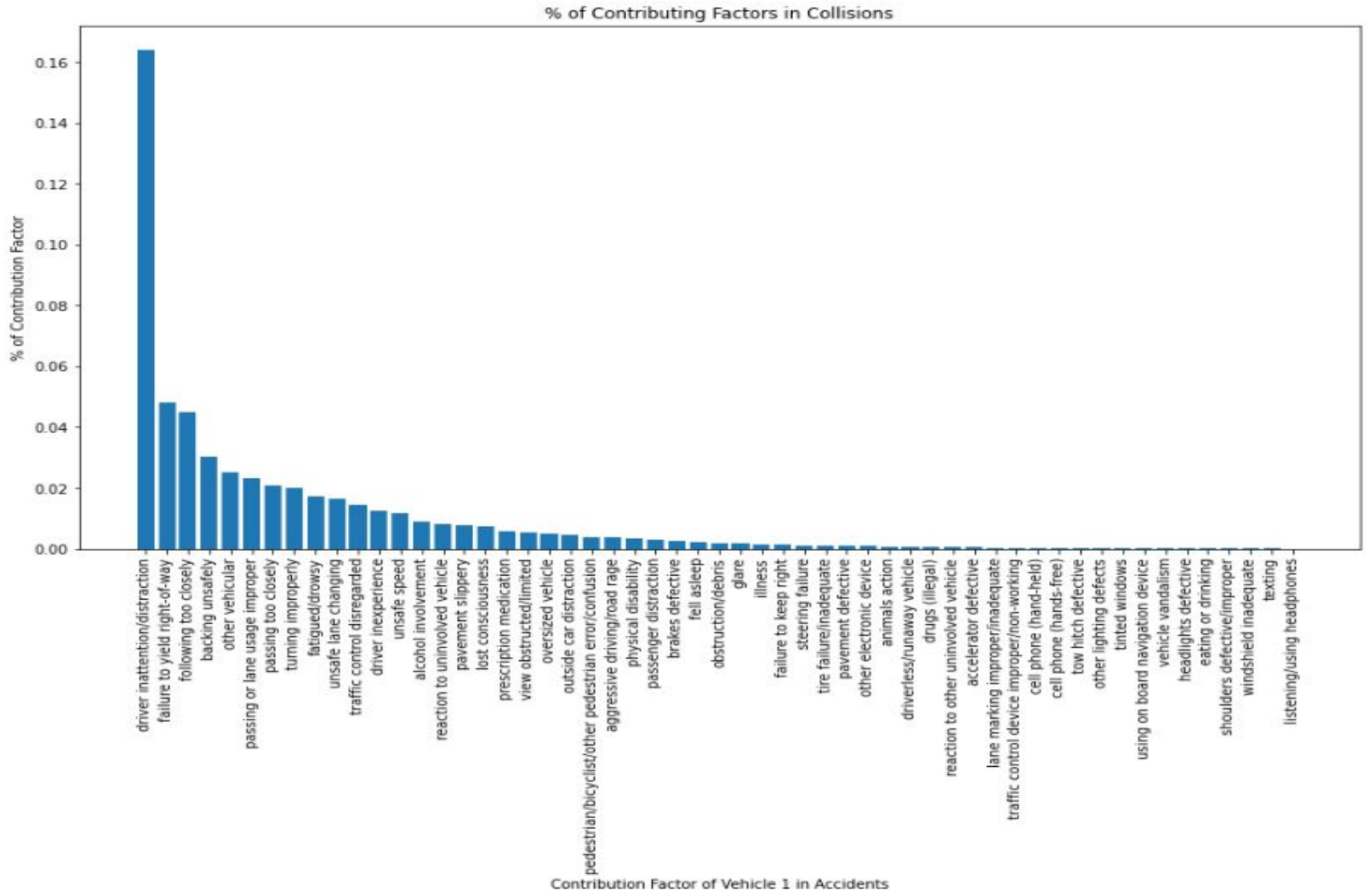
% of Contribution Factors in Deaths



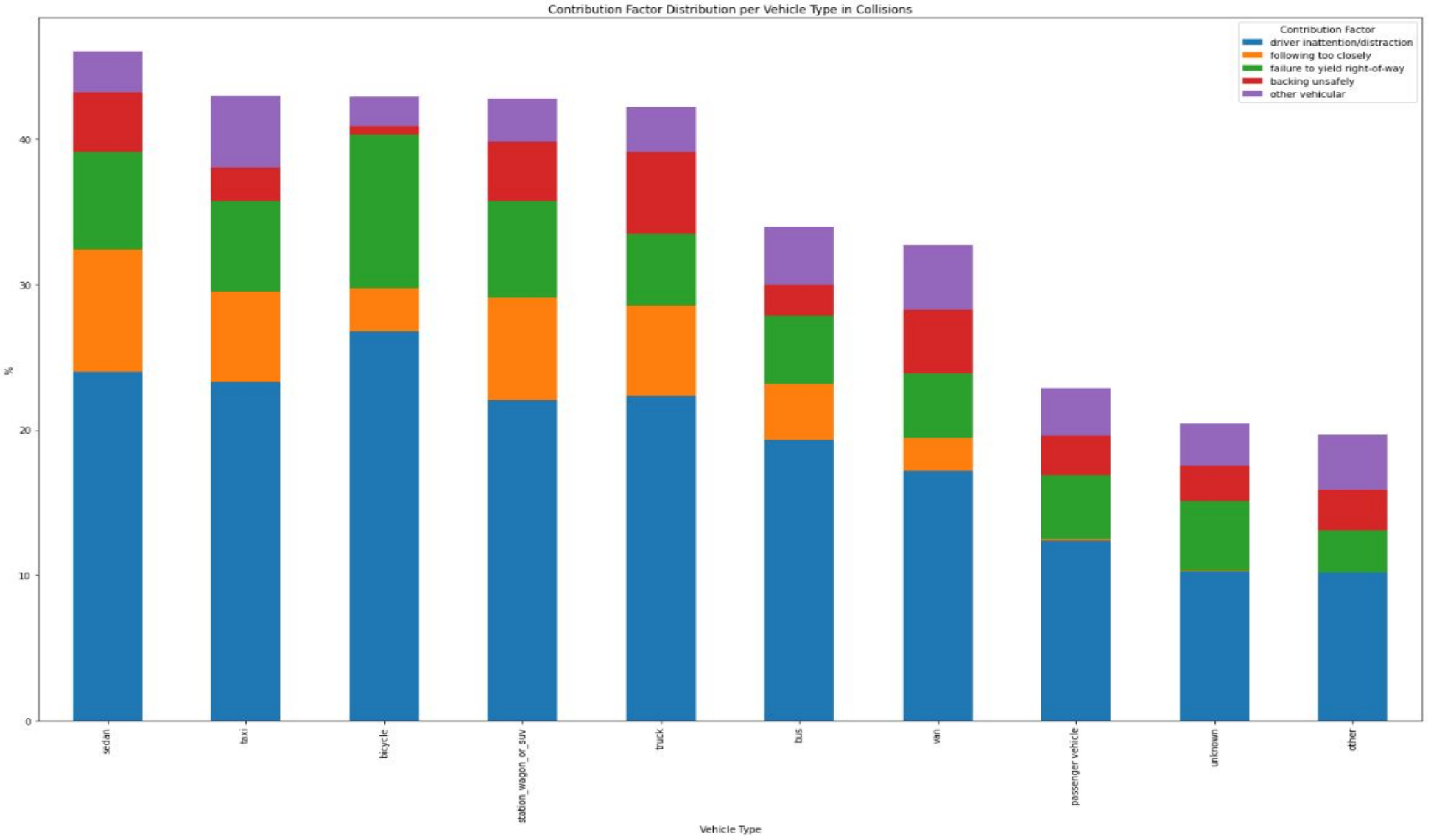
% of Vehicle Types in Collisions



% of Contribution Factors in Collisions



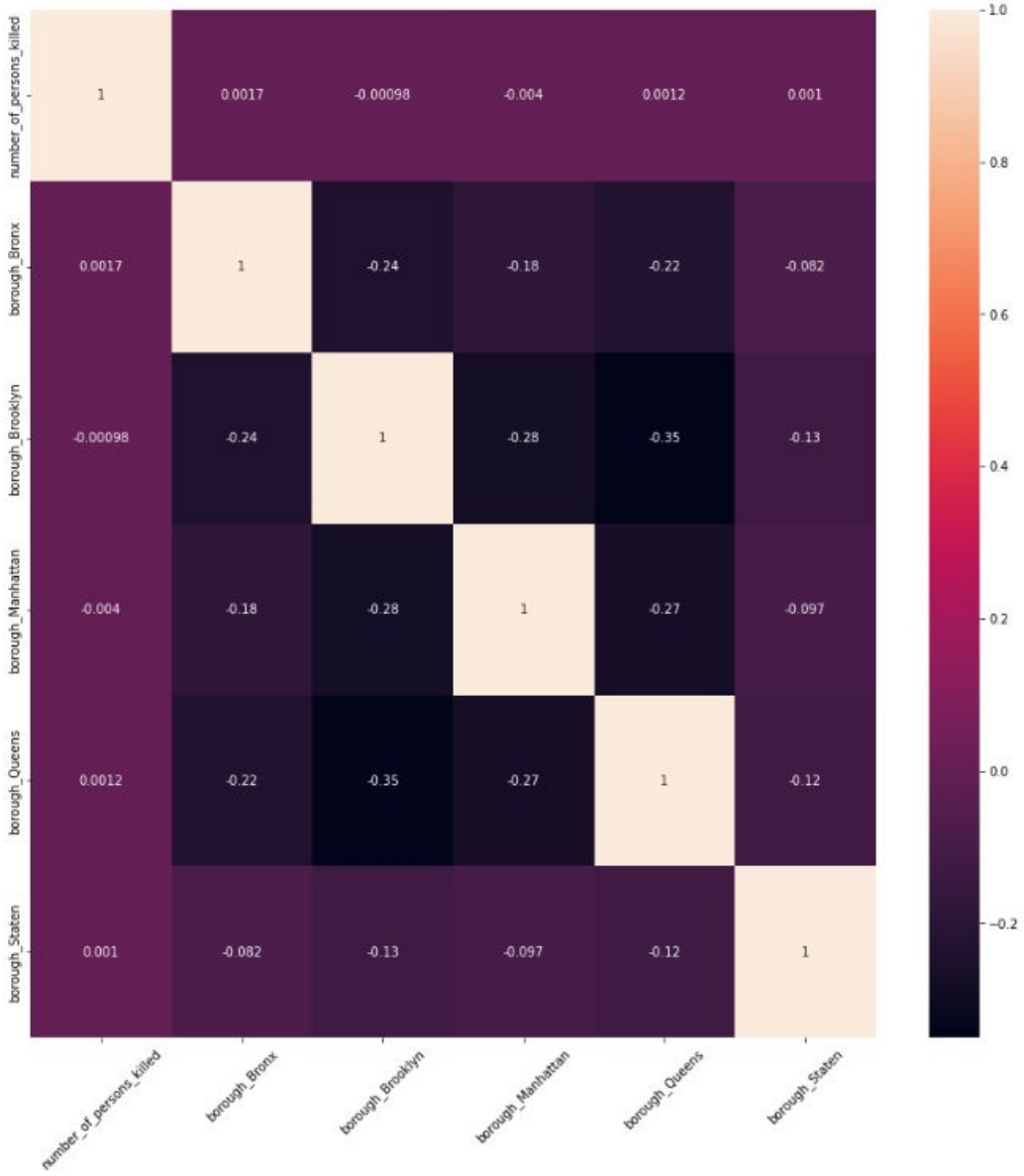
% of Contribution Factors per Vehicle Type in Collisions



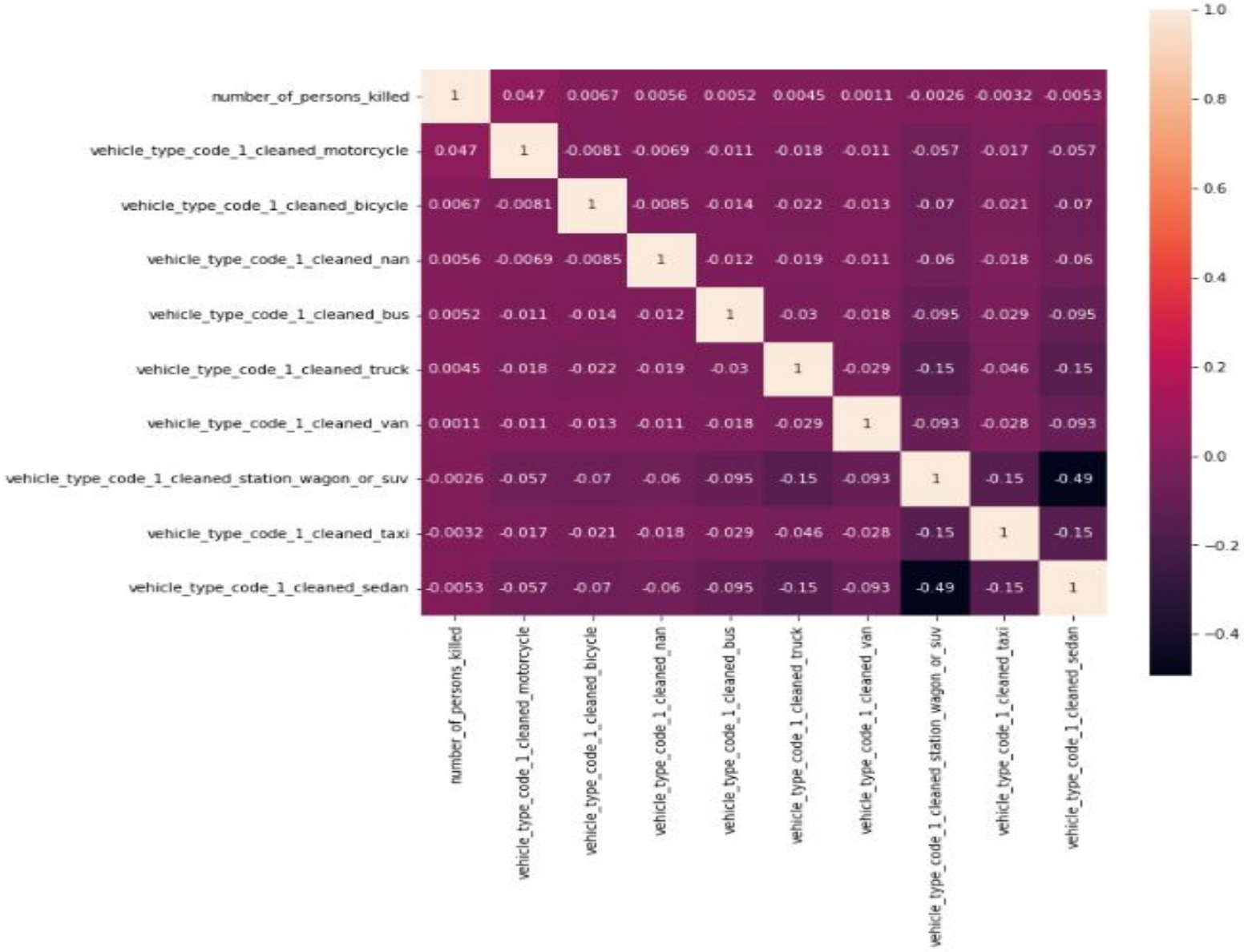
Question 3

- What's the correlation between characteristics analyzed in Question 1 & 2 (i.e. borough, vehicle type, contribution factor to collision) in the data and likelihood of death in a collision? What traits contribute to deaths in a collision if any?

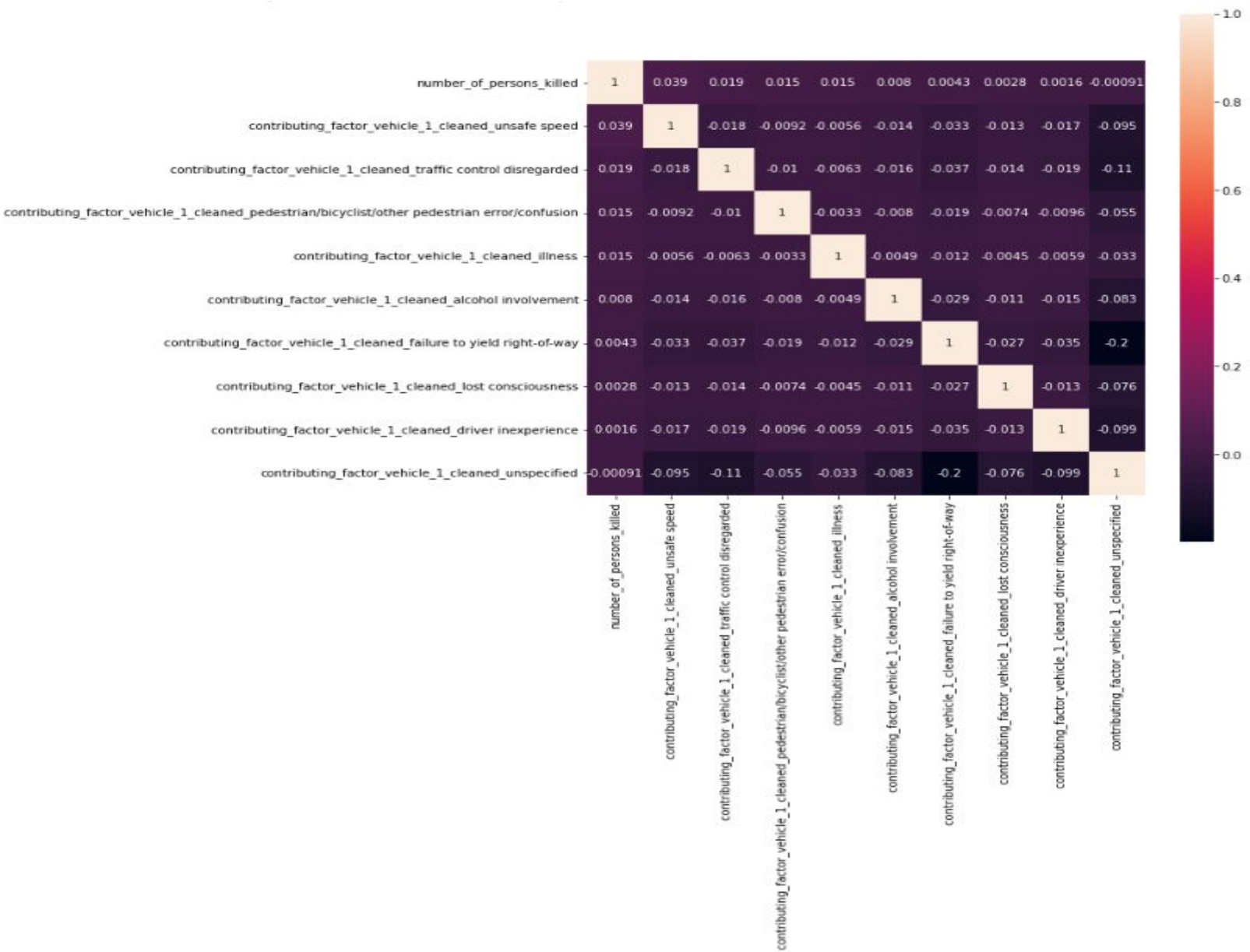
Correlation between Boroughs and Car Accident Deaths



Correlation between Vehicle Type and Car Accident Deaths



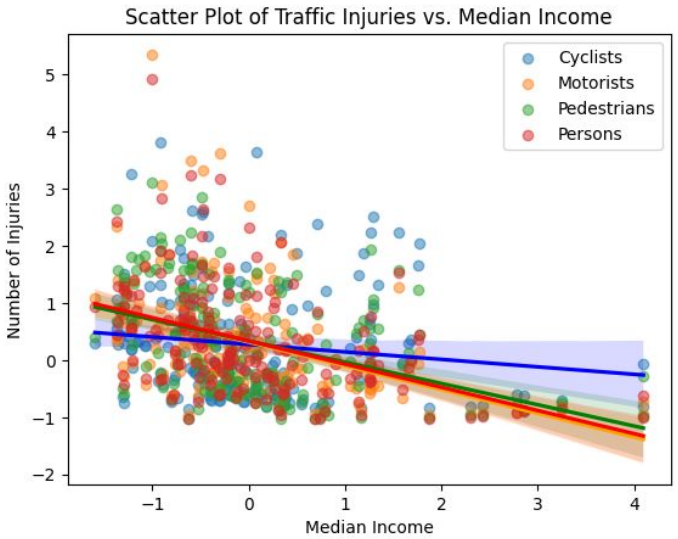
Correlation between Contributing Factor and Car Accident Deaths



Question 4: Median Income and Collisions by Zipcode

Is there any correlation between median income of a zip code and number of deaths and injuries from collisions?

Question 4: Median Income and Collisions by Zipcode



DEATH

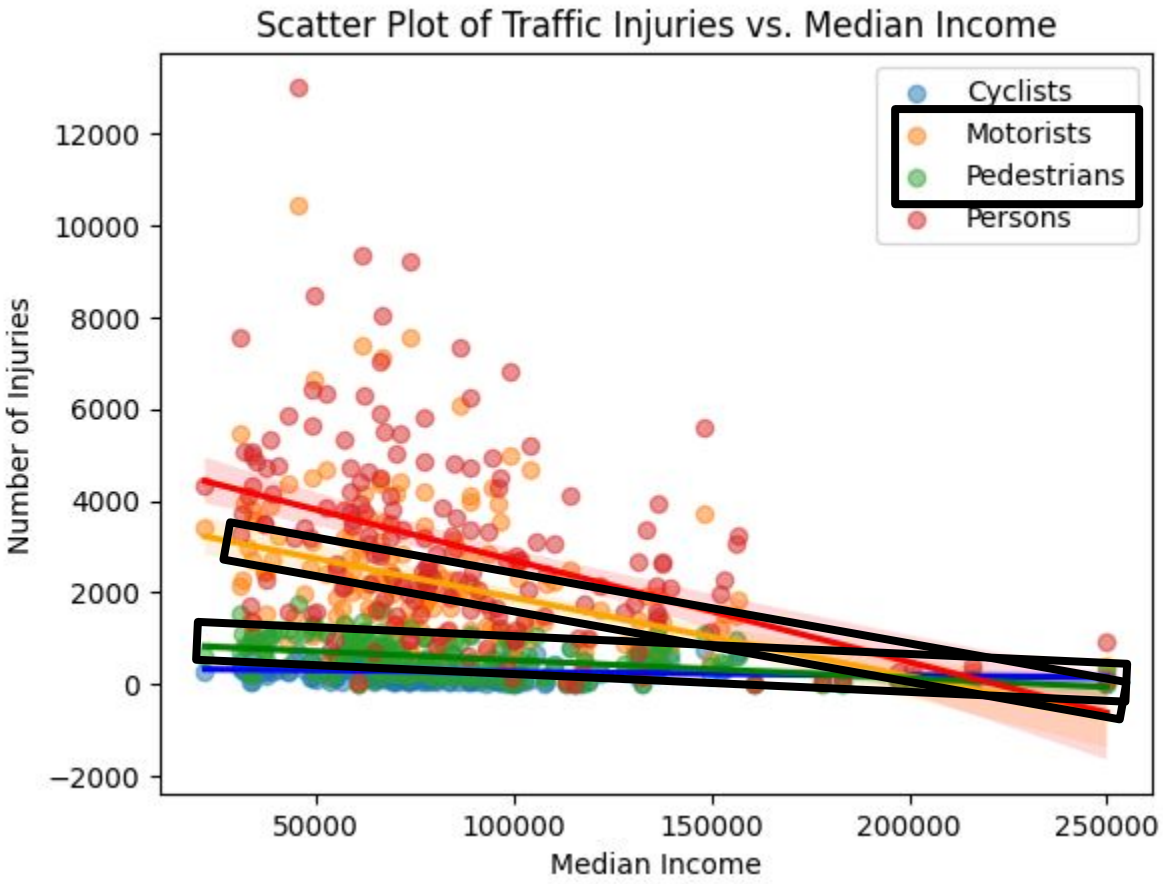
Correlations (Death Count and Zip Code Median Income):	
Total Deaths	-0.400
Pedestrian Deaths	-0.360
Cyclist Deaths	-0.150
Motorist Deaths	-0.318

INJURY

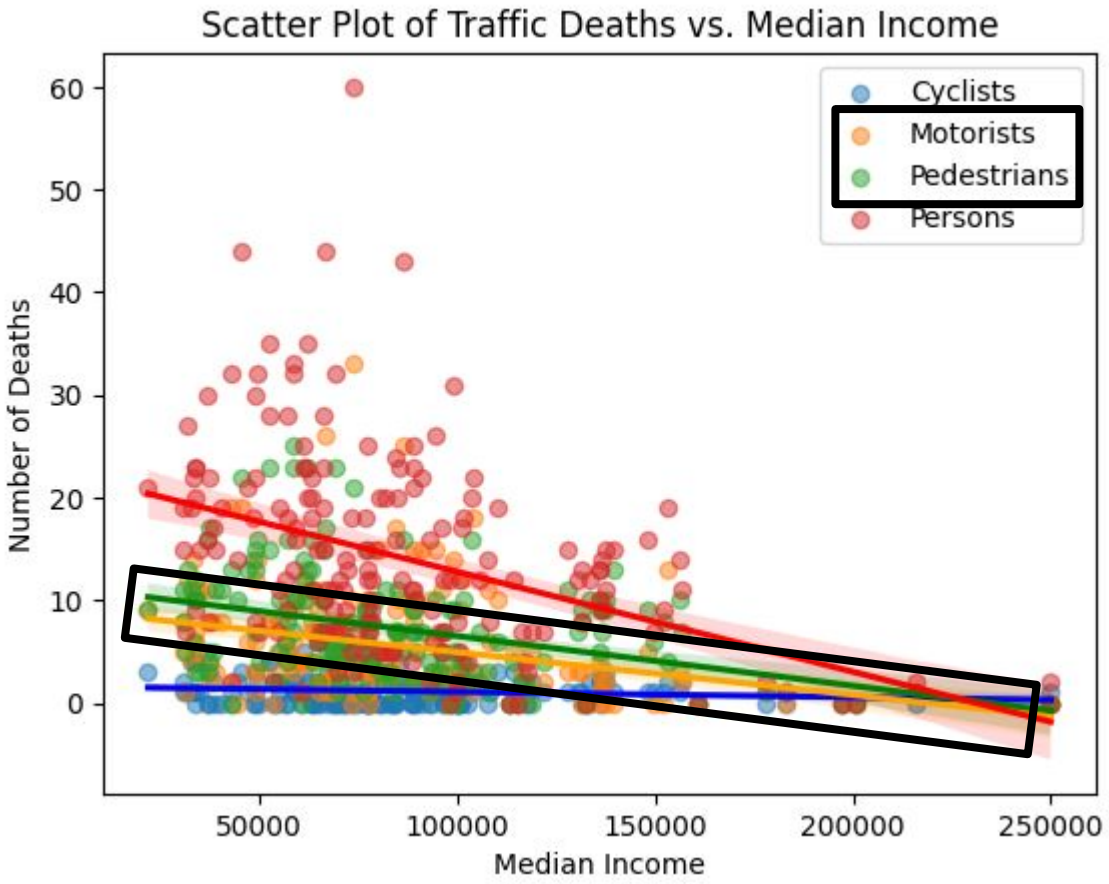
Correlations (Injury Count and Zip Code Median Income):	
Total Injuries	-0.434
Pedestrian Injuries	-0.398
Cyclist Injuries	-0.129
Motorist Injuries	-0.431

Question 4: Median Income and Collisions by Zipcode

INJURY



DEATH



Main Findings and Future Directions

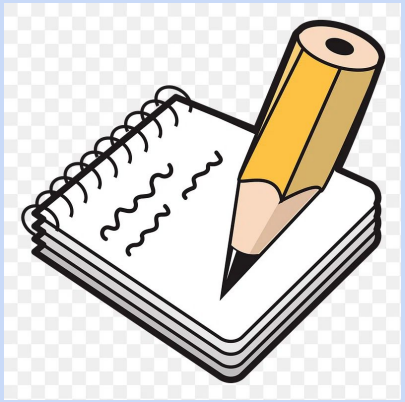
1. **Boroughs:** Manhattan has less total persons and motorist death and injury, and Staten island had low injury rates all around.
    a. However, correlations show that the boroughs were essentially not very different from each other in injury/death
    b. Further analysis is needed.

2. **Vehicle Types:** Station wagons or SUVs and Sedans are most common types in collisions,
    a. Likely because of how common these types of cars are

3. **Contributing Factors:** Driver Innattention/Distraction and speeding are most frequent causes, adding up to almost ⅕ of all reported causes in the past ten years
    a. None of the common car types or contributing factors were found to be highly associated with injury or death.

4. **Income and Injury/Death:** Zipcode median income is negatively associated with most types of traffic injury or death in new york. These correlations are sizeable, but suggests more factors are at play.

Future Directions



1. Delve Deeper

1. Borough and Zipcode level is missing detail that should be investigated with a street-level qualitative study
2. Robustness checks and inferential statistics should be applied for more concrete findings
3. Further research will zoom in and investigate factors for why poorer zipcodes have higher death and injury rates.

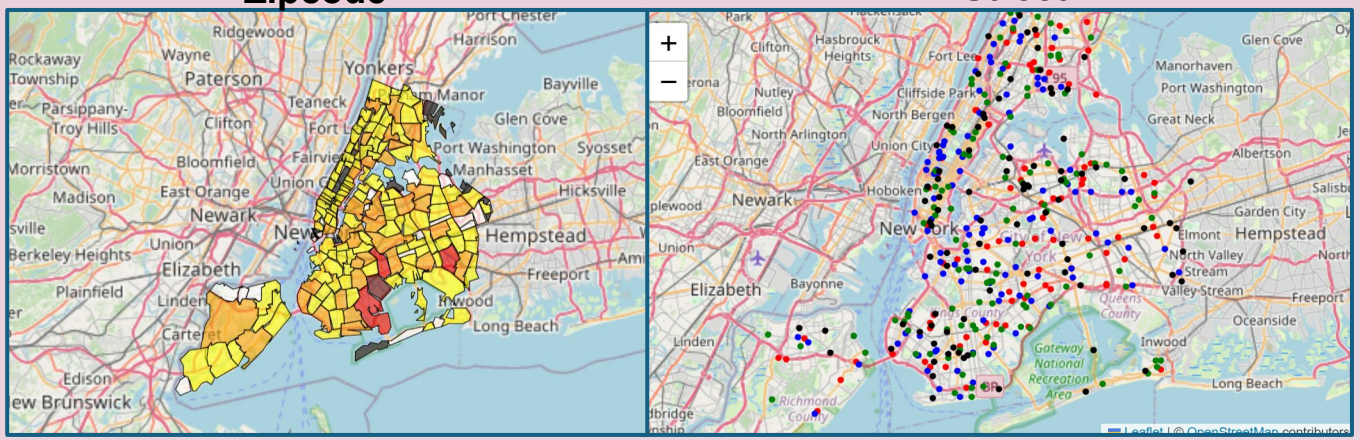
2. Work with Community Resources

> Jamaica Bay in Brooklyn is a hot spot.

> Pedestrian deaths around major intersections across the city

A qualitative investigation into problematic intersections and Jamaica Bay street systems could be applied

Zipcode **Street**



THANK YOU