Eyes on the Road:
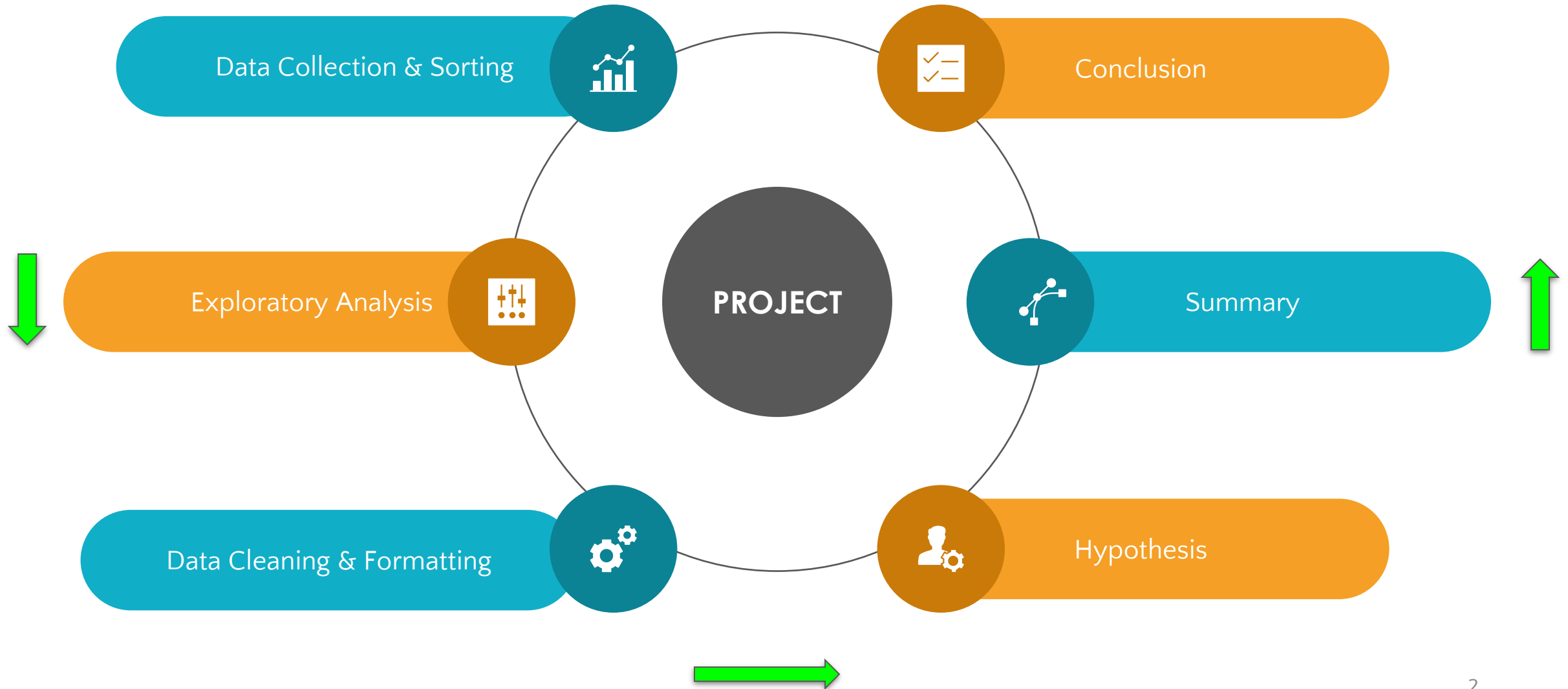## Factors that Lead to Vehicular Accidents in New York City

Group 7
Andi, Debbie, Adam & Susan

Project Stages

Data Collection & Sorting

Conclusion

Exploratory Analysis

Summary

PROJECT

Data Cleaning & Formatting

Hypothesis

Source: https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95

Import Libraries, Access API, Create DataFrame, and Sample Data

```python
# Ignore non-fatal warnings
import warnings
warnings.filterwarnings("ignore")

# Import other dependencies, including sodapy, which is identified in the API docs as crucial
import pandas as pd
from sodapy import Socrata
import matplotlib.pyplot as plt
import numpy as np
import requests
import time
import scipy.stats as st
from scipy.stats import linregress
import json
import os
import calendar
```

DataFrame Clean-up and Hygiene

```python
In [49]:   # Exclude columns that are unnecessary for the analysis
           excluded_columns = ["on_street_name", "off_street_name", "contributing_factor_vehicle_2", "vehicle_type_code2", "zip_code",
           "location", "cross_street_name", "cross_street_name", "contributing_factor_vehicle_3", "vehicle_type_code_3", "contributing_
           filtered_df = results_values_df.loc[:, ~results_values_df.columns.isin(excluded_columns)]

           # Drop rows with missing values
           filtered_df = filtered_df.dropna()

           # Drop rows where "contributing_factor_vehicle_1" is "Unspecified"
           filtered_df = filtered_df[filtered_df["contributing_factor_vehicle_1"] != "Unspecified"]

           # Reorder columns with "collision_id" as the first column so that each crash can be easily identified if needed
           column_order = ["collision_id"] + [col for col in filtered_df.columns if col != "collision_id"]
           filtered_df = filtered_df[column_order]

           # Counts by column, to make sure analysis still has a sufficient number of records after previous drops
           column_counts = filtered_df.count()
           print(column_counts)

           # Sample values to verify modifications to the data
           filtered_df.head(5)
```

```
collision_id                       456
crash_date                         456
crash_time                         456
number_of_persons_injured          456
number_of_persons_killed           456
number_of_pedestrians_injured      456
number_of_pedestrians_killed       456
number_of_cyclist_injured          456
number_of_cyclist_killed           456
number_of_motorist_injured         456
number_of_motorist_killed          456
contributing_factor_vehicle_1      456
vehicle_type_code1                 456
borough                            456
dtype: int64
```

Which condition correlates most prominently in the data with vehicular accidents? Driver Inattention or Distraction

Incidence of 'Driver Inattention/Distraction' vs. Other Factors

Driver Inattention/Distraction

32.9%

67.1%

Other Factors
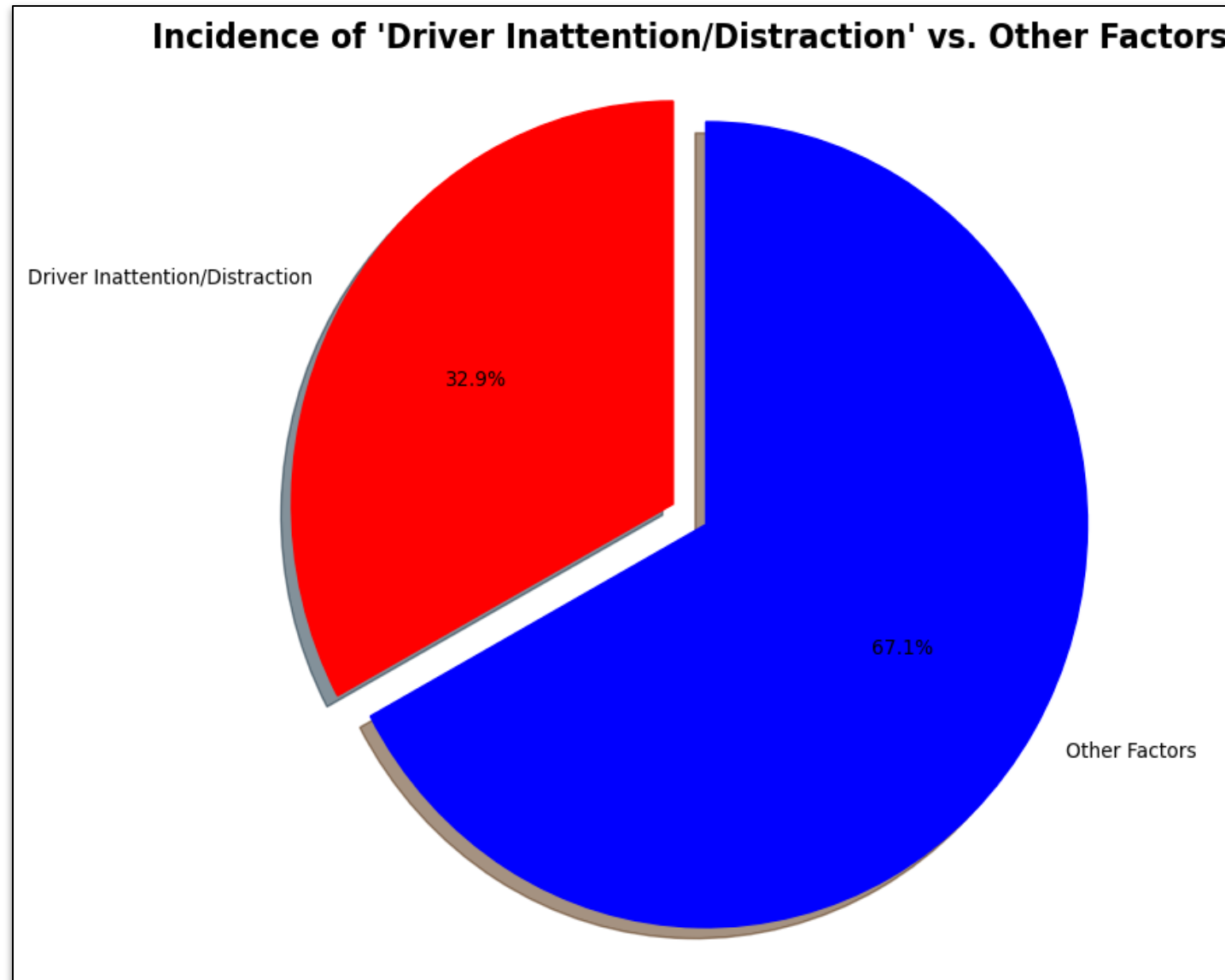
That means that a third of all the nearly 500 accidents you see here are correlated with inattentive driving!

Incidence of Contributing Factors (Percentage of the Total)
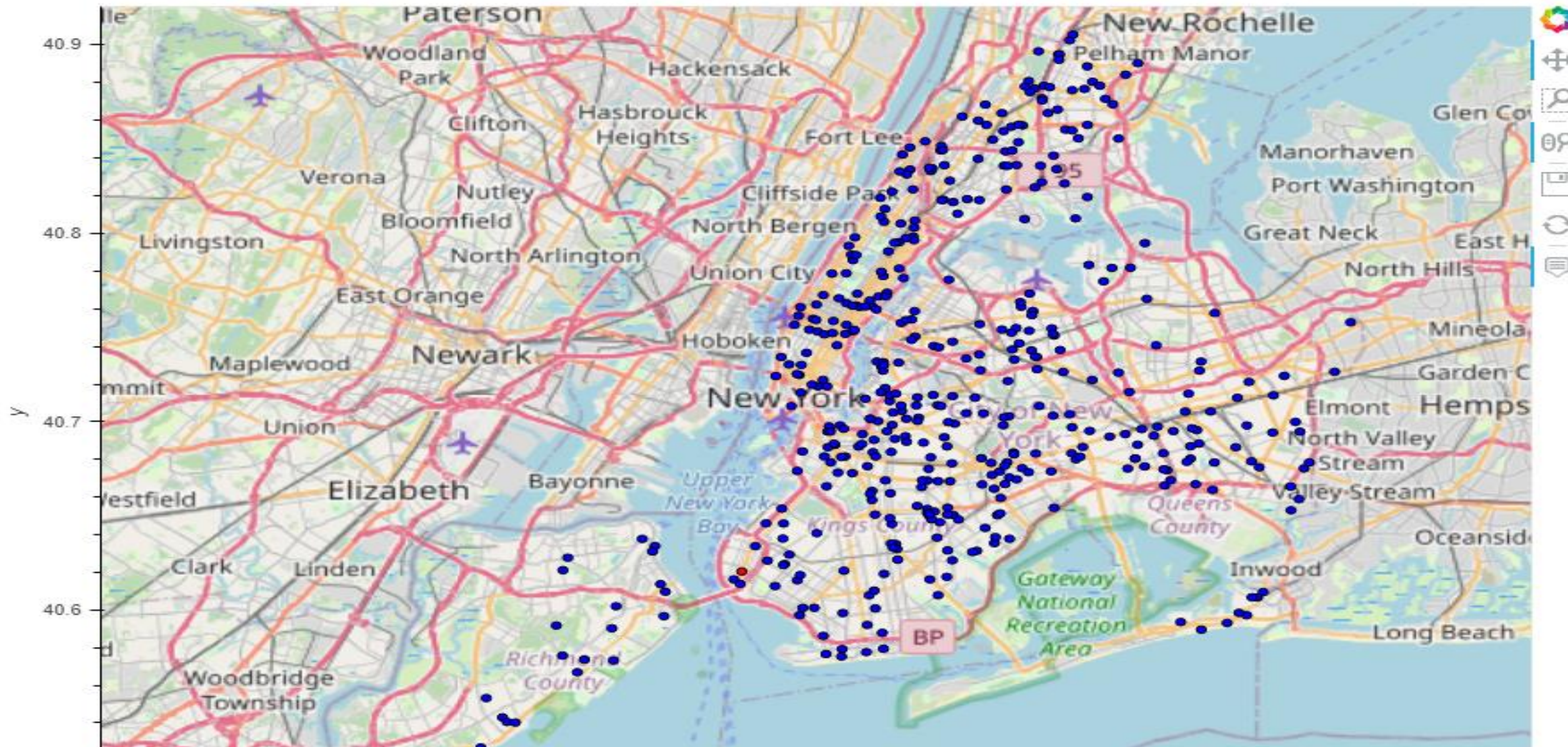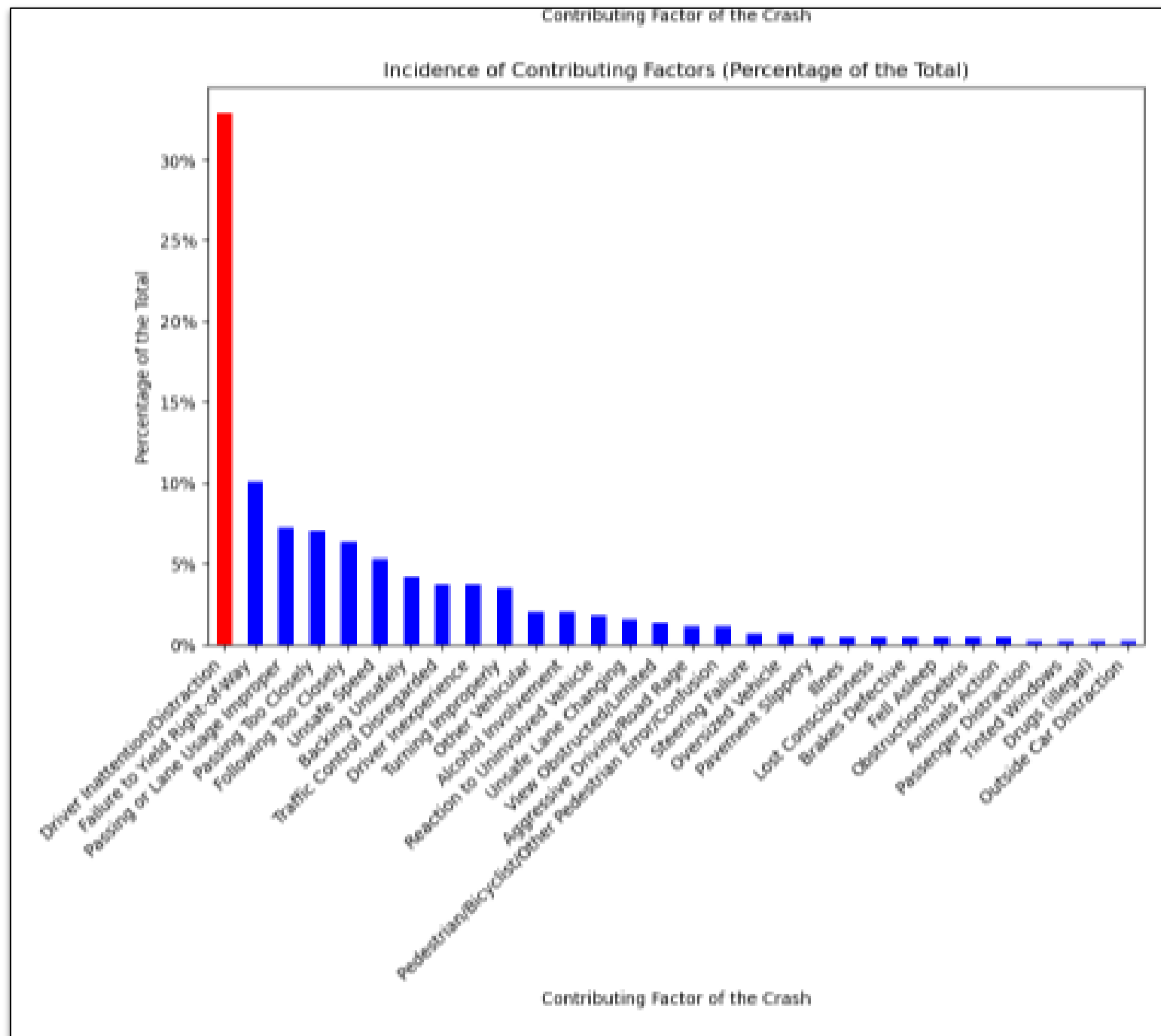
Hypotheses

Most significant vehicular crashes in New York city during the study period are associated with **inattentive driving.**
*We proposed to investigate how this phenomenon interacts with other variables:*

- Does inattentive driving remain the top correlate of accidents regardless of the persons involved?

- Having established the salience of distracted driving, we asked if accidents vary by time of day or day of the year.

- Do more indirect factors like type of vehicle and borough of New York City play a role for accidents by distracted drivers specifically?

Does the type of person involved, just the motorist or a pedestrian or cyclist, lessen the salience of distracted driving as contributing to accidents? *In a word: No, though correlates vary after it's taken into account*


Top 10 Contributing Factors for Motorists Injured

Driver Inattention/Distraction - 28
Following Too Closely - 15
Failure to Yield Right-of-Way - 15
Traffic Control Disregarded - 10

Unsafe Speed - 4
Passing or Lane Usage Improper - 4
Driver Inexperience - 4
Reaction to Uninvolved Vehicle - 3

Passing Too Closely - 3
Other Vehicular - 3
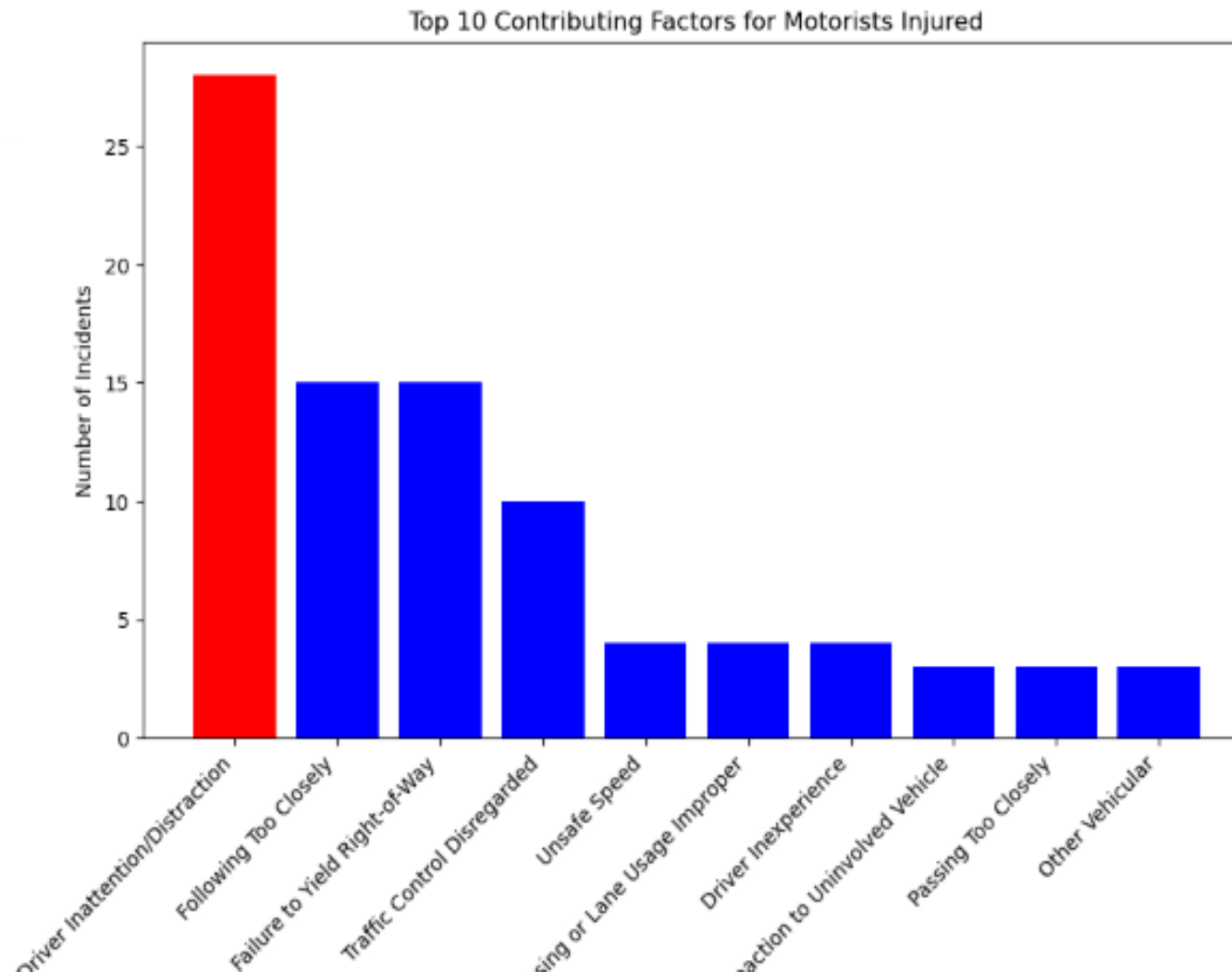
Driver Inattention/Distraction - 12
Failure to Yield Right-of-Way - 10
Backing Unsafely - 2
Unsafe Speed - 2
Aggressive Driving/Road Rage - 1
View Obstructed/Limited - 1
Tinted Windows - 1
Turning Improperly - 1
Pedestrian/Bicyclist/Other Pedestrian Error/Confusion - 1
Reaction to Uninvolved Vehicle - 1

Top 10 Contributing Factors for Pedestrians Injured
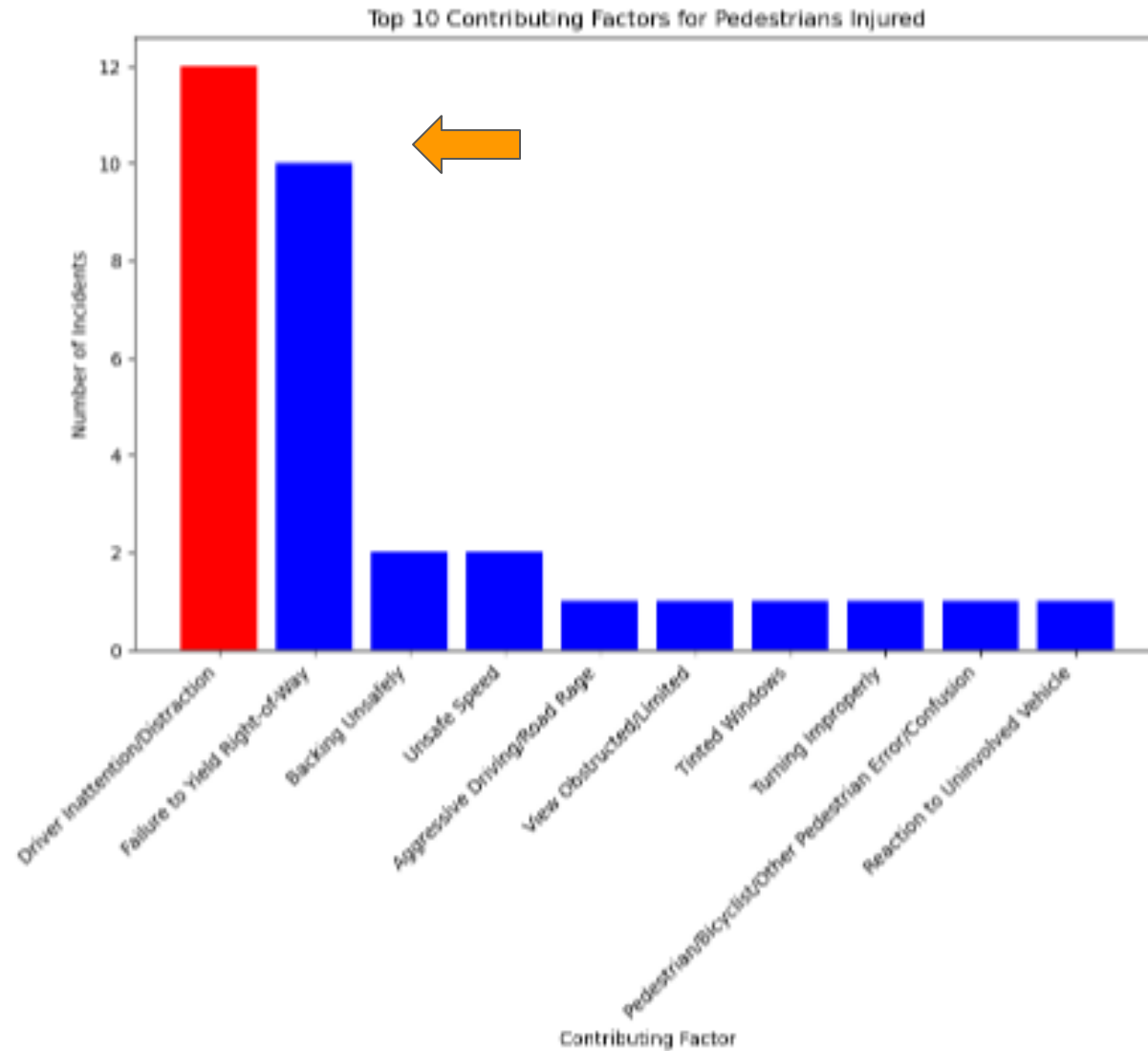
Driver Inattention/Distraction - 6
Following Too Closely - 3
Passing or Lane Usage Improper - 3
Failure to Yield Right-of-Way - 3
View Obstructed/Limited - 2
Turning Improperly - 2
Pedestrian/Bicyclist/Other Pedestrian Error/Confusion - 2
Passenger Distraction - 1
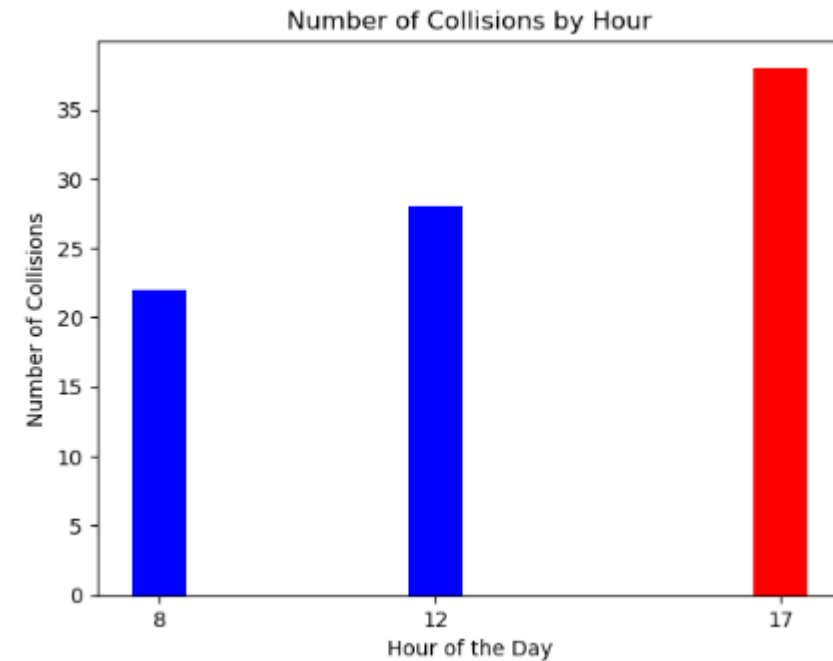Alcohol Involvement - 1
Unsafe Speed - 1

Top 10 Contributing Factors for Cyclists Injured
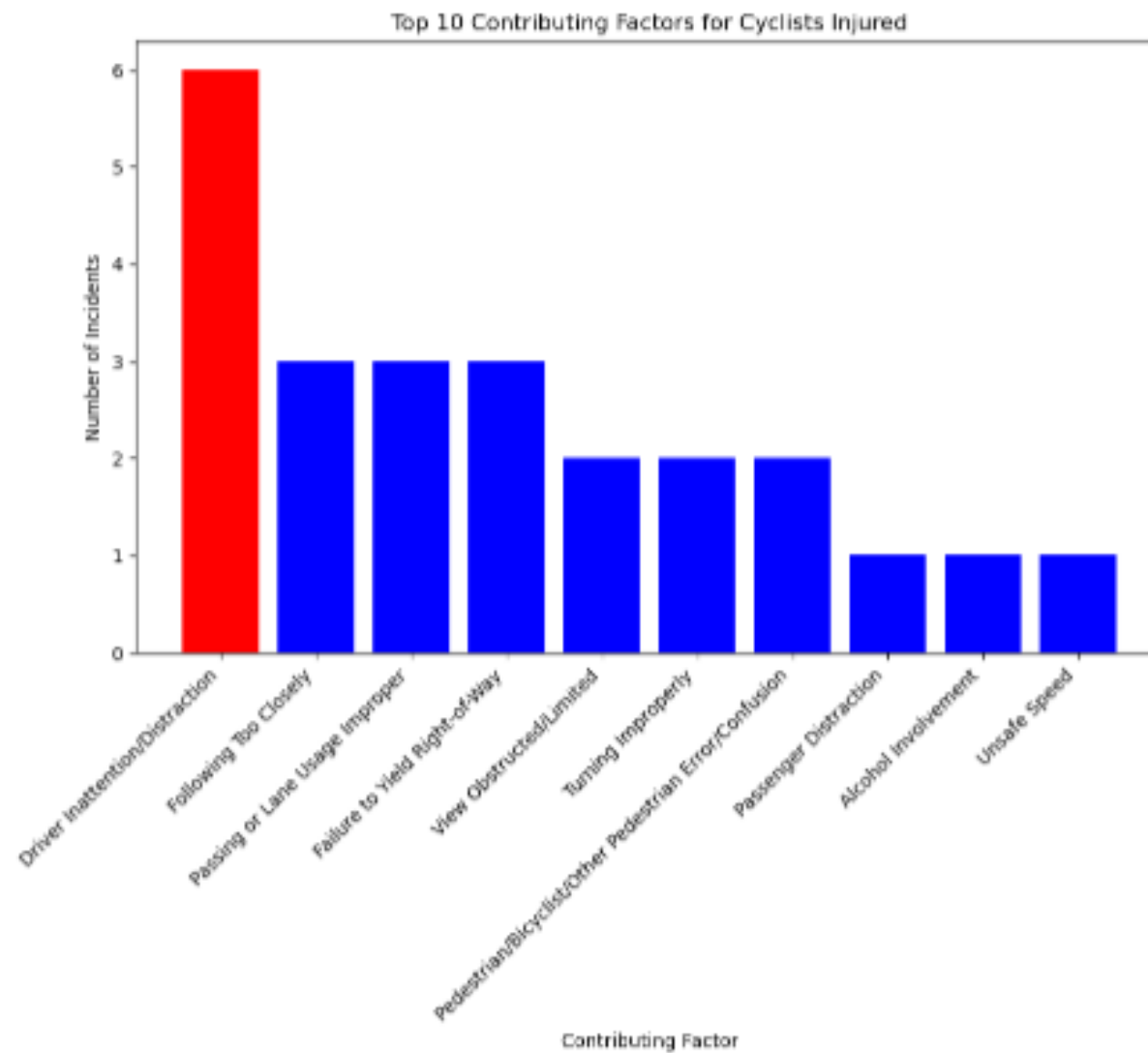
The r-value is: 0.5315547744256138

Relationship between Time of Day and Collisions

y = 0.68x + 11.18
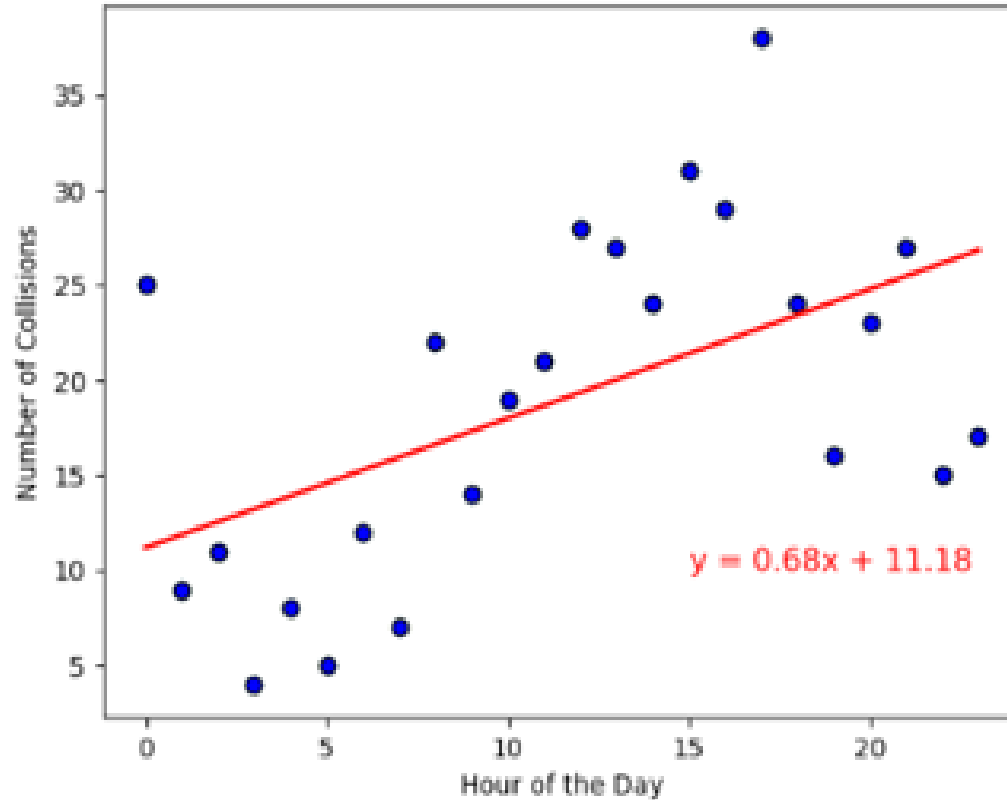
Is there a relationship between time of day (including the rush hours and the lunch hour) and the volume of collisions? *Yes, a positive one (r = 0.53)*
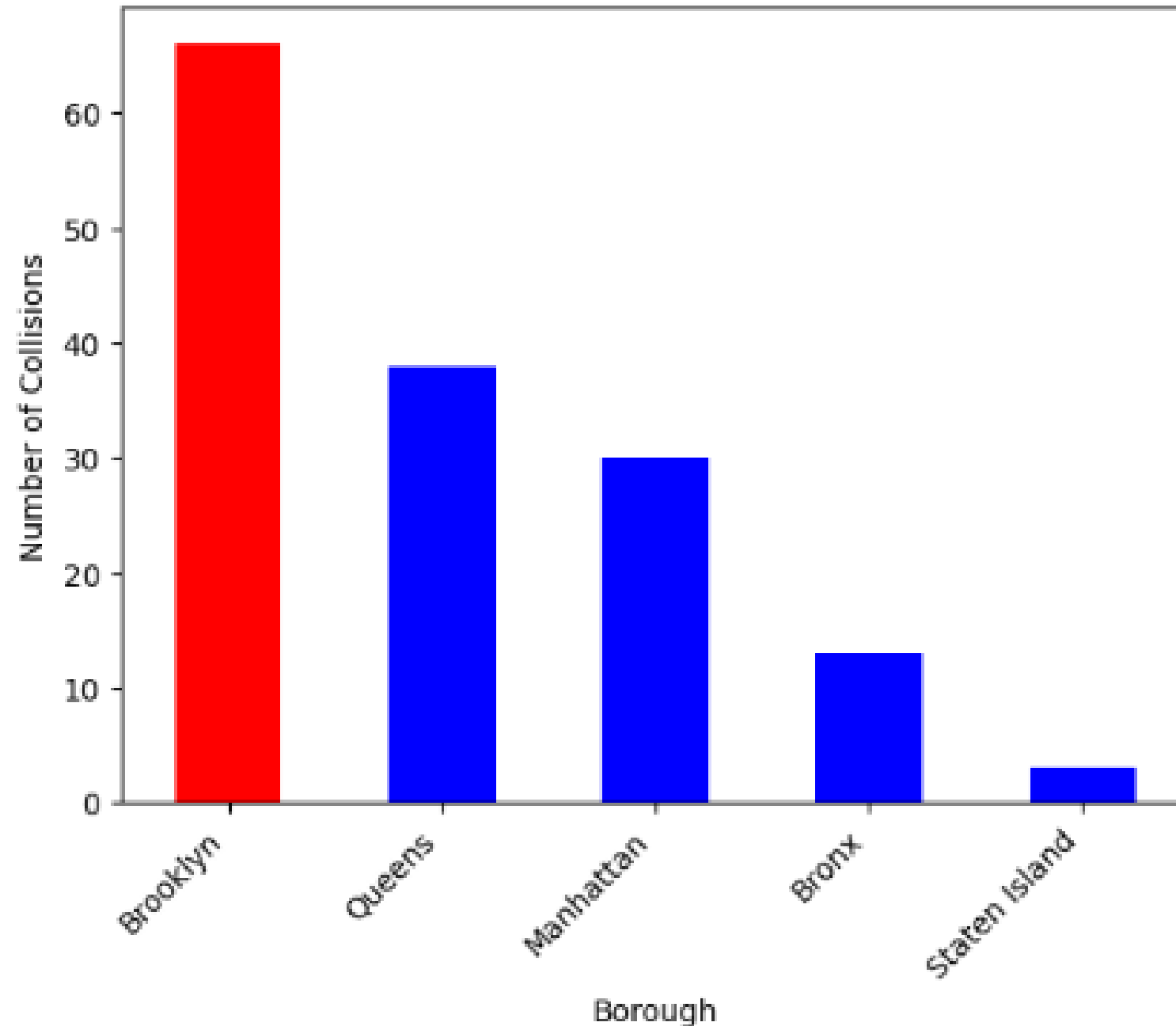
Number of Collisions by Hour

We asked if collisions were influenced by time of the year. *Collisions are markedly unipolar by month, with April accounting for a disproportionate number*

Number of Collisions by Month

Driver Inattention/Distraction Collisions by Vehicle Type
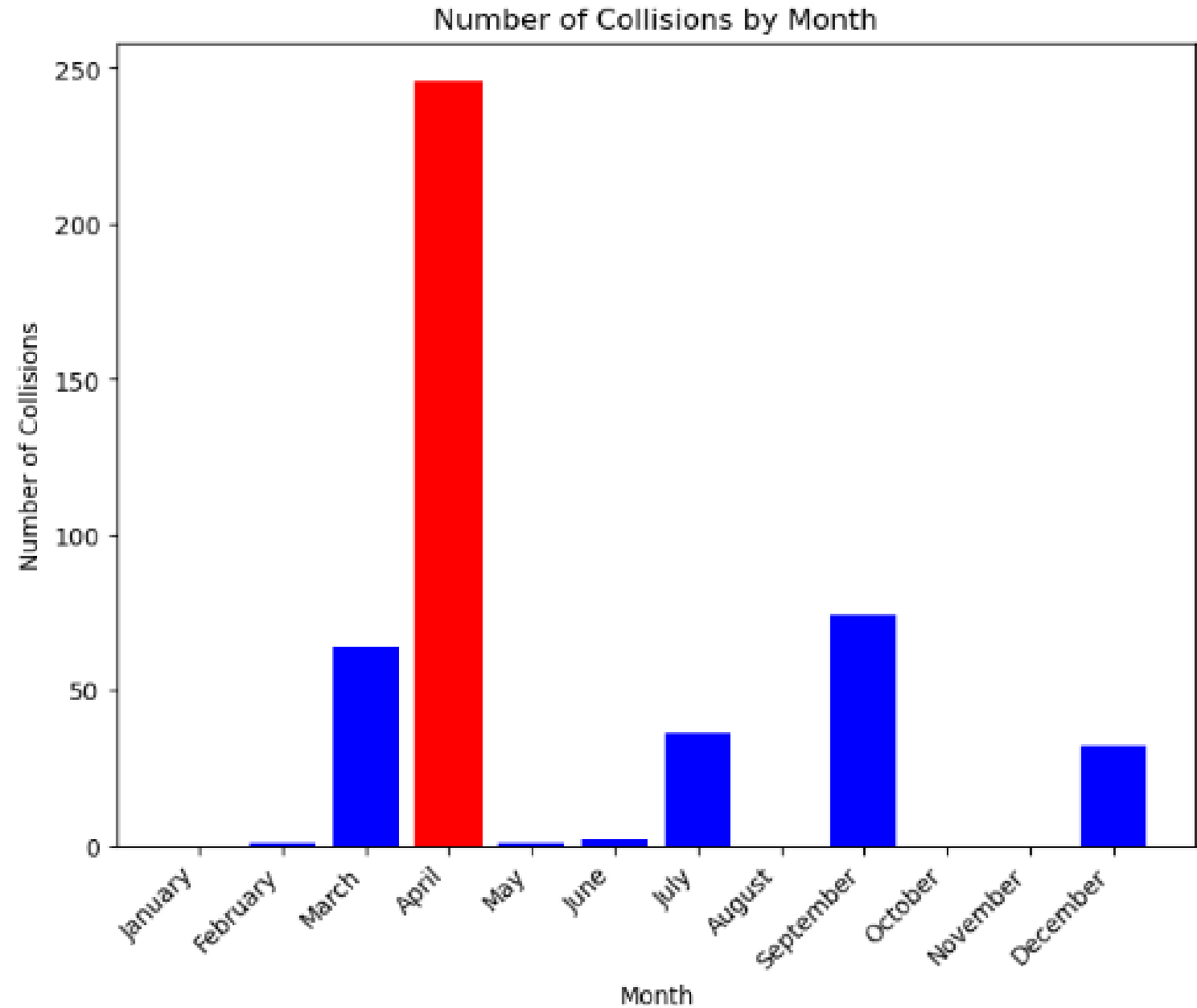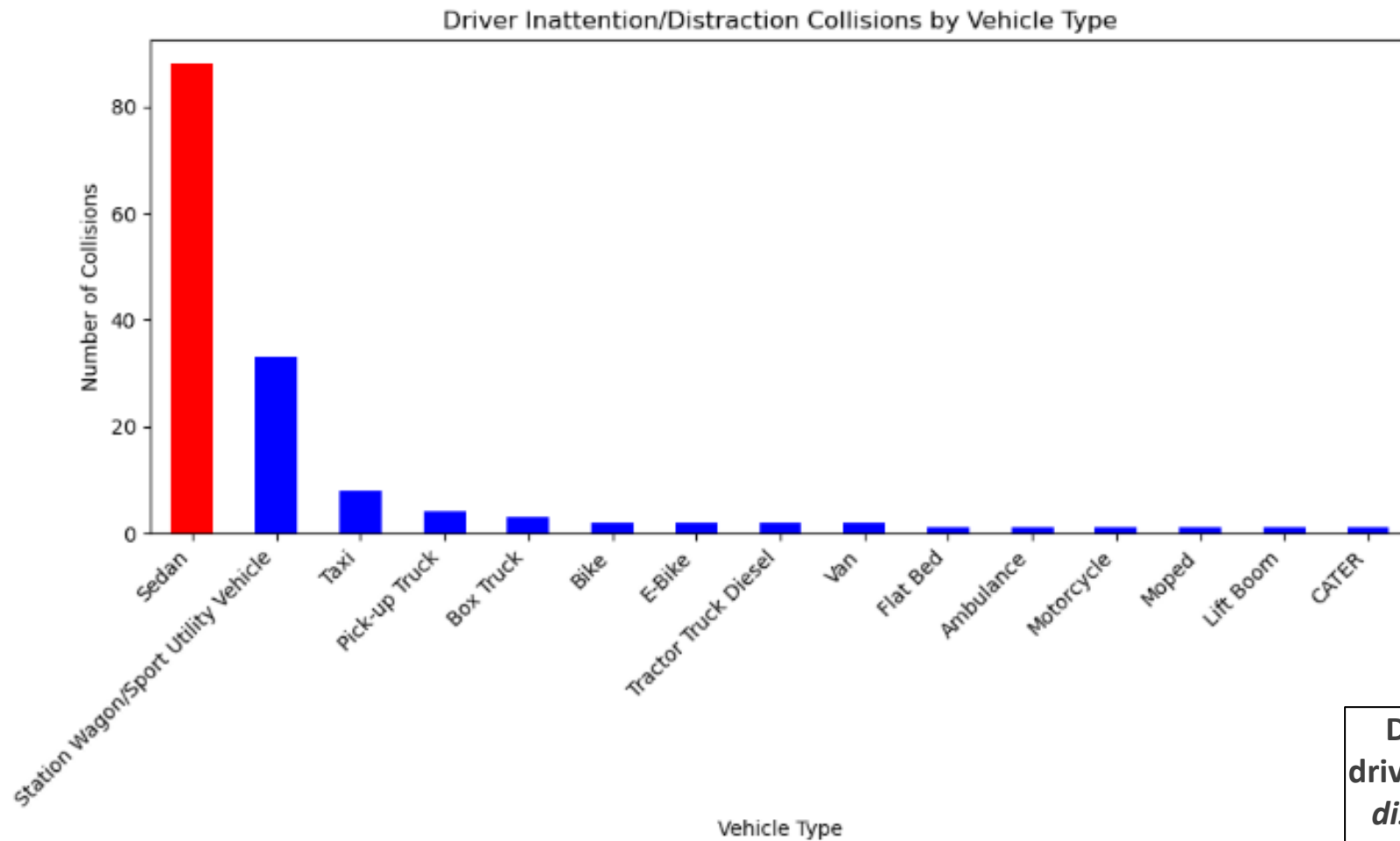
Does type of vehicle play a role in distracted driving-prompted accidents? *Most collisions from distracted drivers involve sedans, even though (presumably) taxis are on the road longer each day and thus more vulnerable to crashes.*

Driver Inattention/Distraction Collisions by New York City Boroughs

What is the influence of geography in distracted driving-correlated accidents? *The number of distracted driver-collisions by borough tracks by population, even though Manhattan contains the most cars during any typical workday*

Summary Analysis

❖ **Crash Frequency & Time Trends**

We were able to provide some clear variation in the number of crashes over time, showing high and low frequencies (yearly/monthly) trends to identify recurring patterns or seasonal variations

❖ **Crash Locations & Hotspots**

For better traffic management and safety interventions

❖ **Types of Crashes**
Vehicular, Pedestrian-involved, Cyclist, Motorist

❖ **Factors Contributing to Crashes**

Common factors : Driver Inattention, Failure to Yield Right – of-way, and Backing Unsafely

❖ **There were only four fatal vehicular crash within the seven-year period according to the dataset**

❖ **Driver inattention/distraction is the leading cause of collision compared drugs and alcohol, as this is constant across the type of persons involved.**

❖ **Brooklyn by far had the highest number of collisions followed by Queens as the runner up compared to Manhattan, even though there are more cars in Manhattan on any typical workday.**

❖ **Collision dangers become progressively greater throughout the day.**

❖ **The vehicle type that is most likely to be in a collision associated with distracted driving is a Sedan vs a Taxi, as we might otherwise predict from time on the road.**
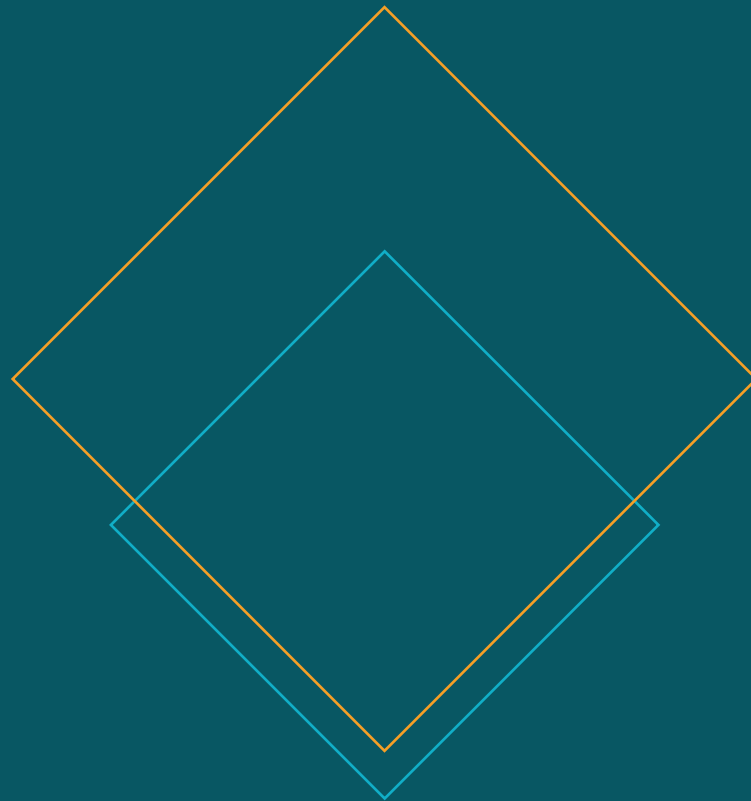
❖ **Of the seven years of data we collected, the month of April reported the highest amounts of collision (which I can speculate is due to bad weather/Spring)**

Conclusion

It was important for us to acknowledge the limitations of the analysis, such as potential data underreporting or incomplete information. Addressing these limitations and incorporating additional data sources could have further refine the analysis and provide a more comprehensive understanding of vehicular crashes in NYC. It would have been interesting to conduct a deeper dive into the meta data to isolate those specific behaviors that are closely associated with vehicle collisions (cell phone usage, texting, lateness) and if traffic jams contribute to these various distractions.

We were also curious to narrow down the dataset to see which highways, streets, avenues reported the most collision in each of the boroughs, but that was not available at this time.

Thank You!