Brooklyn Accidents Analysis & Recommendation

By Hoang To

April 2024

Executive Summary

Traffic accidents continue to be a pressing issue in New York City, posing significant risks to public safety and straining municipal resources. This analysis dived into the traffic collision data of NYC with a focus on Brooklyn and uncovered important trends and accident patterns to propose strategic recommendations to reduce accidents in Brooklyn.

The analysis revealed several trends and patterns in Brooklyn traffic accidents:
- Peak Hours: Accidents spiked during morning and evening rush hours, suggesting that congestion and increased traffic volumes contribute to heightened risk.
- High-Risk Thoroughfare: Main thoroughfares, such as Atlantic Avenue, Flatbush Avenue, and Eastern Parkway, are where the accidents tend to occur.
- Contributing Factors: Common contributing factors included driver inattention/ distraction and failure to yield right-of-way underscoring the need for educational campaigns and stricter enforcement measures.

From the analysis, I believe the main root-cause of the accidents is that the traffic congestion makes drivers become both impatient and fatigue behind the wheels, causing them into making suboptimal decisions and get into preventable accidents.

Recommendations
Based on the analysis, I recommend a multi-faceted approach aims to provide immediate relief while laying the groundwork for comprehensive, lasting change.

Short-Term Solutions:
- Rapid Flashing Beacons along major roads will immediately enhance pedestrian visibility and remind drivers to yield right-of-way.
- Congestion Pricing will promptly reduce vehicle volumes entering/exiting NYC, alleviating borough congestion.

Long-Term Solution:
- Investing in expanding and modernizing NYC's public transit networks (subways, buses, commuter rail) is crucial. Robust public transportation will alleviate congestion by reducing private vehicle reliance while promoting a safer, sustainable transportation future.

The short-term measures provide immediate relief, paving the way for the comprehensive long-term solution of world-class public transit. This multi-pronged approach tackles both urgent needs and lasting transformation towards an efficient, accessible, secure transportation ecosystem.

Analysis Preparation Pipeline and Analysis Approach

Analysis Preparation:
- **Query:**
  - Core data queried from BigQuery
- **Data cleaning and processing:**
  - Data quality check and cleaning in Python using pandas and numpy
  - Processed datatable is imported into Tableau
- **Exploratory analysis**
  - Done in both SQL and Python
- **Data visualization**
  - Use Tableau as main visualization tool
  - Python (with pandas and matplotlib) as secondary

Analysis Approach:
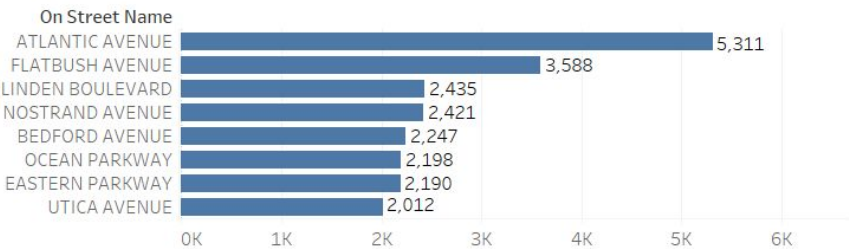In my analysis, I seek to answer the 5 W/H's:
- When does the accidents happen?
- Where does it occur?
- What are the common causes or how did this happen?
- Who are impacted by the accidents?

This would allow me to have a right context to tackle the problem and come up with the reasonable recommendation.
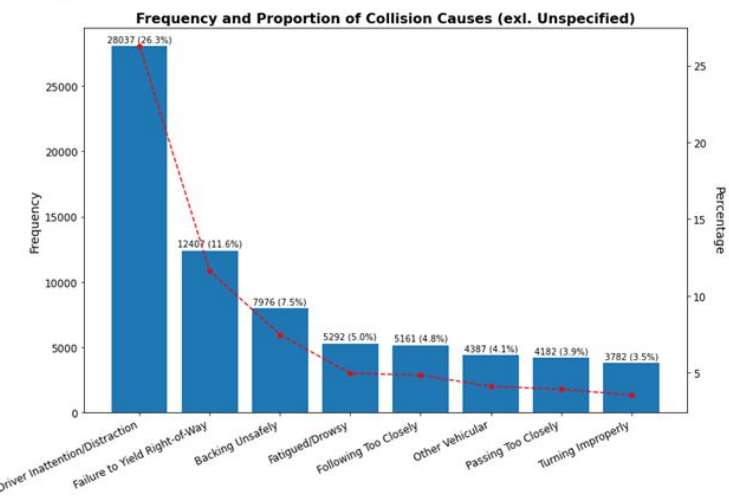
Additionally, the questions help me examine the dataset to understand the limitation and the extent that it could help answer my questions.

The Analysis Shows That Majority of Accidents and Injuries Happen During Weekday Peak Traffic Hours on Major Thoroughfare
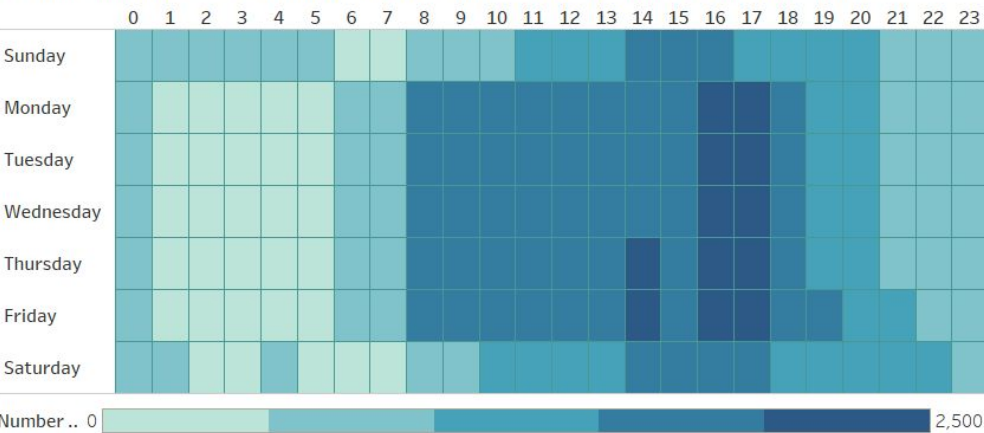
1) We see that accidents occur the most on major thoroughfares in and out of New York (e.g. Atlantic Ave, Flatbush Ave,..)


2) Most accidents occur during weekday business hours, peaking at the end of day.

3) Lastly, driver inattention/distraction is a top collision factor, accounting for 1 in 4 accidents

4) Lastly, while not all accidents result in injuries and/or deaths, it is clear that pedestrians, if involved in a collision caused by a motor vehicle, have a significantly higher risk of being fatally injured than everyone else, while chances were they did not cause the accidents.
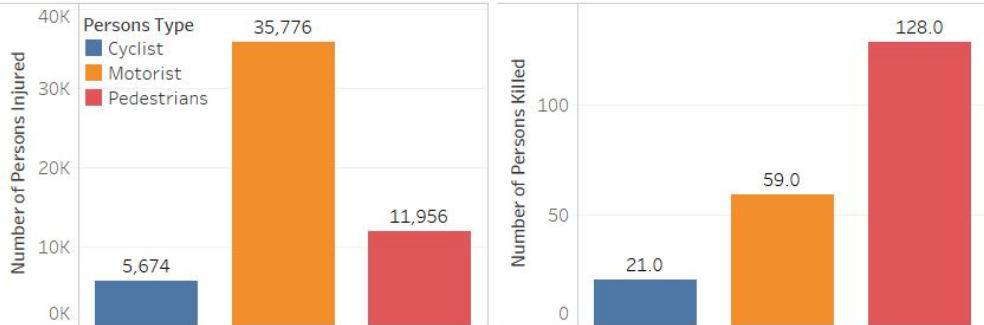
Recommendations: A Multifaceted Approach to Reduce Accidents and Injuries in Brooklyn, NY

From the analysis, I believe the main root-cause of the accidents is that the traffic congestion makes drivers become both impatient and fatigue behind the wheels, causing them into making suboptimal decisions and get into preventable accidents.

Short term solutions combine with ...

Long term investment

#1: Rapid Flashing Beacons

Installing RRFB along thoroughfares crossing will add a clear and catchy signal to drivers about pedestrian crossings, tackling the top two reasons for accidents in Brooklyn.

#2: Congestion Pricing

This will reduce the number of vehicles going into New York and in turns lessen the daily volume coming out of the city into borough such as Brooklyn.

#3: Public Transportation

Expanding and modernizing New York City's public transit networks, including subways, buses, and commuter rail, directly contributes to decongesting roads and creating a safer transportation landscape for pedestrians, cyclists, and motorists alike.

Appendix A: Overview

All the code associated with this project is stored on Github and can be accessed through the public repository:

hoangcto/brooklyncasestudy_collision (github.com)

Including:

- SQL Query file
- Jupyter Notebook with Python Data Processing
- Tableau Analysis File

Appendix B: SQL Query & Python Analysis

```sql
-- QUERY 1
-- Get the data in scope for the analysis
SELECT*
  FROM
    `bigquery-public-data.new_york.nypd_mv_collisions`
  WHERE
    borough = "BROOKLYN"
    AND cast(timestamp as date) between "2014-01-01"
AND "2017-12-31";
```

borough	contributing_fac...	contributing_fac...	contributing_fac...	contributing_fac...	contributing_fac...	cross_street_name	timestamp
BROOKLYN	Unspecified	Unspecified	Unspecified			NORTH 3 STREET	2014-06-12T16:00:00.000Z
BROOKLYN	Unspecified					BERRIMAN STREET	2017-12-18T07:00:00.000Z
BROOKLYN	Lost Consciousness					STUYVESANT AVENUE	2014-01-17T08:30:00.000Z
BROOKLYN						AVENUE J	2016-04-08T16:45:00.000Z
BROOKLYN	Unspecified					BUSHWICK AVENUE	2015-03-28T20:00:00.000Z
BROOKLYN	Unspecified					DE WITT AVENUE	2017-10-17T09:15:00.000Z
BROOKLYN	Unspecified					LAFAYETTE AVENUE	2017-11-09T07:30:00.000Z
BROOKLYN	Unspecified					50 STREET	2016-02-05T12:04:00.000Z
BROOKLYN	Unspecified					SMITH STREET	2014-10-22T13:49:00.000Z
BROOKLYN	Unspecified					NEW YORK AVENUE	2015-06-12T17:45:00.000Z
BROOKLYN	Other Vehicular					52 STREET	2014-10-31T14:55:00.000Z
BROOKLYN	Unspecified	Unspecified				FARRAGUT ROAD	2017-10-23T09:45:00.000Z
BROOKLYN	Driver Inattention/Dist...	Driver Inattention/Dist...				KEAP STREET	2017-12-29T10:32:00.000Z
BROOKLYN	Unspecified					HENRY STREET	2017-11-16T15:43:00.000Z
BROOKLYN	Driver Inattention/Dist...	Unspecified				MARCY AVENUE	2017-11-30T12:47:00.000Z
BROOKLYN	Unspecified	Unspecified				78 STREET	2014-10-10T10:00:00.000Z
BROOKLYN	Alcohol Involvement	Unspecified				HOYT STREET	2014-01-24T18:34:00.000Z
BROOKLYN	Unspecified	Unspecified				6 AVENUE	2014-09-15T07:13:00.000Z
BROOKLYN	Unspecified	Unspecified				EUCLID AVENUE	2015-11-19T13:30:00.000Z

Appendix B: SQL Query & Python Analysis

```sql
-- QUERY 2
-- What is the most common factor in a motor vehicle collision in Brooklyn?
-- This query counts the number of collisions for each collision type, in
descending order.
WITH brookyn_data_2014_2017 AS (
  SELECT*
  FROM
    `bigquery-public-data.new_york.nypd_mv_collisions`
  WHERE
    borough = "BROOKLYN"
    AND cast(timestamp as date) between "2014-01-01" AND "2017-12-31"
)
SELECT
  contributing_factor_vehicle_1 AS collision_factor,
  COUNT(*) num_collisions
FROM
  brookyn_data_2014_2017
WHERE
  contributing_factor_vehicle_1 != "Unspecified"
  AND contributing_factor_vehicle_1 != ""
GROUP BY
  1
ORDER BY
  2 DESC;
```

collision_factor	num_collisions
Driver Inattention/Distraction	24759
Failure to Yield Right-of-Way	11113
Backing Unsafely	7032
Fatigued/Drowsy	5654
Other Vehicular	3925
Following Too Closely	3778
Lost Consciousness	3534
Turning Improperly	3237
Traffic Control Disregarded	2885
Prescription Medication	2538
Passing or Lane Usage Improper	2444

Appendix B: SQL Query & Python Analysis

```sql
-- QUERY 3
-- What are the most dangerous streets for motor vehicle collisions in
Brooklyn?
-- This query counts the number of fatalities and injured by streets.
WITH brooklyn_data_2014_2017 AS (
  SELECT*
  FROM
    `bigquery-public-data.new_york.nypd_mv_collisions`
  WHERE
    borough = "BROOKLYN"
    AND cast(timestamp as date) between "2014-01-01" AND "2017-12-31"
)
SELECT
  on_street_name,
  SUM(number_of_persons_killed) AS deaths,
  SUM(number_of_persons_injured) AS injuries
FROM
  brooklyn_data_2014_2017
WHERE
  on_street_name <> ''
GROUP BY
  on_street_name
ORDER BY
  deaths DESC
LIMIT
  10;
```

on_street_name	deaths	injuries
ATLANTIC AVENUE	12	1527
LINDEN BOULEVARD	5	1124
BEDFORD AVENUE	5	916
OCEAN PARKWAY	5	667
RALPH AVENUE	4	440
EASTERN PARKWAY	4	789
FLATBUSH AVENUE	4	1045
FLATLANDS AVENUE	3	392
BROADWAY	3	487
NOSTRAND AVENUE	3	822

Appendix B: SQL Query & Python Analysis

```sql
-- QUERY 4
-- (a) When do most accidents happen in NYC's Brooklyn?
WITH brooklyn_data_2014_2017 AS (
  SELECT *
  FROM `bigquery-public-data.new_york.nypd_mv_collisions`
  WHERE borough = "BROOKLYN"
    AND CAST(timestamp AS DATE) BETWEEN "2014-01-01" AND "2017-12-31"
)
SELECT
  EXTRACT(HOUR FROM timestamp) AS hour_of_day,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 1 THEN 1 END) AS sunday,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 2 THEN 1 END) AS monday,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 3 THEN 1 END) AS tuesday,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 4 THEN 1 END) AS wednesday,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 5 THEN 1 END) AS thursday,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 6 THEN 1 END) AS friday,
  COUNT(CASE WHEN EXTRACT(DAYOFWEEK FROM timestamp) = 7 THEN 1 END) AS saturday
FROM
  brooklyn_data_2014_2017
GROUP BY
  hour_of_day
ORDER BY
  hour_of_day;
```

hour_of_day	sunday	monday	tuesday	wednesday	thursday	friday	saturday
0	901	656	612	598	634	726	870
1	662	320	258	266	263	360	598
2	560	217	188	184	205	265	471
3	521	170	136	143	147	197	427
4	587	188	142	181	175	190	535
5	523	249	202	201	242	260	442
6	415	581	542	511	508	509	408
7	318	875	893	888	858	796	393
8	660	1917	1879	1861	1903	1801	718
9	746	1855	1943	1728	1853	1762	892
10	972	1654	1681	1536	1688	1637	1080
11	1150	1621	1673	1527	1699	1615	1199
12	1291	1663	1665	1703	1710	1756	1408
13	1411	1760	1827	1728	1786	1911	1434
14	1633	1963	1933	1974	2031	2162	1673
15	1555	1941	1870	1828	1825	1907	1538
16	1627	2294	2282	2225	2192	2264	1714
17	1471	2148	2279	2180	2135	2145	1597
18	1321	1803	1947	1928	1863	1775	1501
19	1107	1397	1401	1443	1439	1519	1210
20	1025	1116	1215	1162	1156	1251	1202
21	929	888	943	993	962	1029	1038
22	827	724	770	857	840	906	1014
23	618	561	572	579	578	760	881

Appendix B: SQL Query & Python Analysis

Example of visualization in Python. Please refer to Jupyter Notebook file on GitHub

```python
#heatmap of month and dayofthemonth
dayHour = brooklyn_data.groupby(by=['weekday', 'hour']).size().unstack()

# Create a figure and axes
fig, ax = plt.subplots(figsize=(12, 6))

# Set the title
ax.set_title('Collisions heatmap: weekday & hour', fontsize=16)

# Create the heatmap
heatmap = sns.heatmap(dayHour, cmap='Blues', ax=ax)

# Rename the axes
heatmap.set_xlabel('Hour of the Day', fontsize=14)
heatmap.set_ylabel('Day of the Week (Monday = 1)', fontsize=14)

# Add a colorbar legend
cbar = ax.collections[0].colorbar
cbar.set_label('Number of Collisions', labelpad = 20, fontsize=14, rotation = 270)

plt.tight_layout()
plt.show()
```
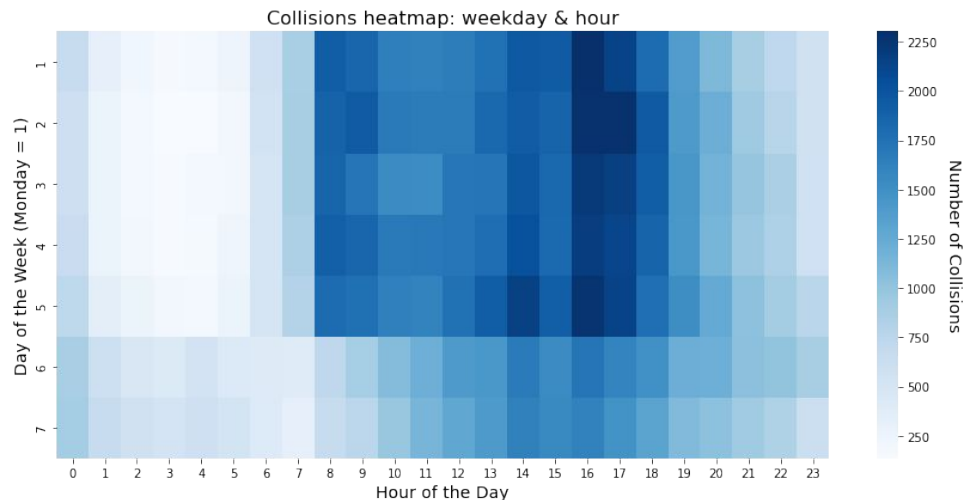

Collisions heatmap: weekday & hour

Appendix C: Tableau Dashboard

- Appendix C.1: When do the Accidents Happen?
- Appendix C.2: Where and What Causes The Accidents?
- Appendix C.3: Who Got Hurt and Tying It All Back

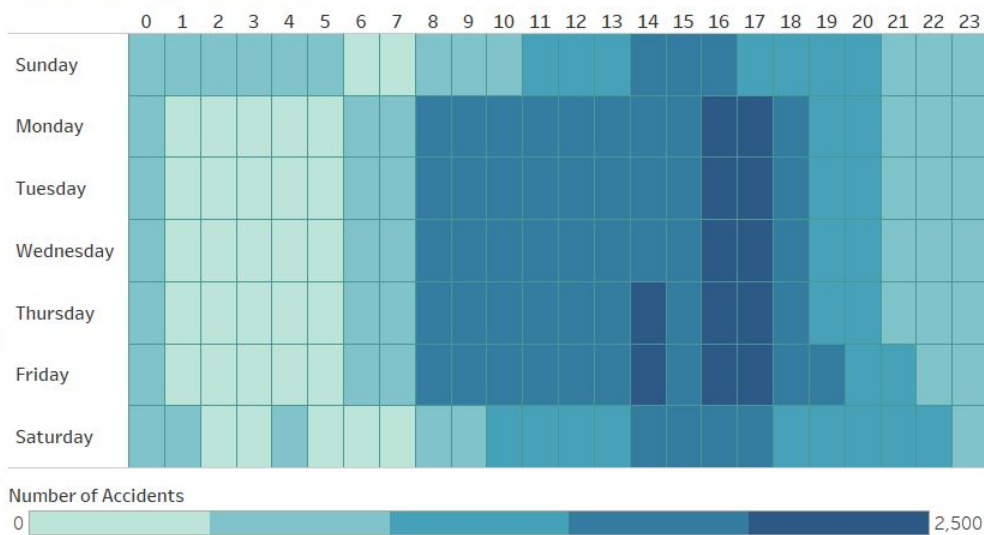# Appendix C.1 When do the Accidents Happen?

1) There is no consistent trend in which months/seasons tend to have elevated amount of accidents
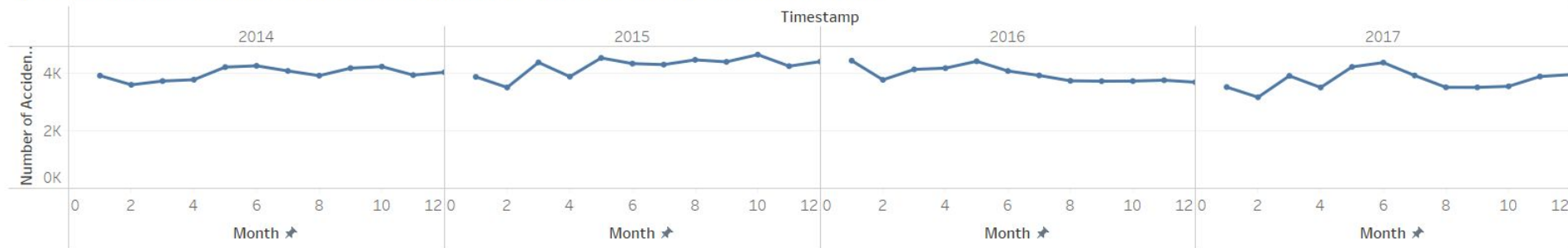
2) However, digging deeper, we see there is a big uptick in accidents happening between 8 AM and 6 PM (business hours)
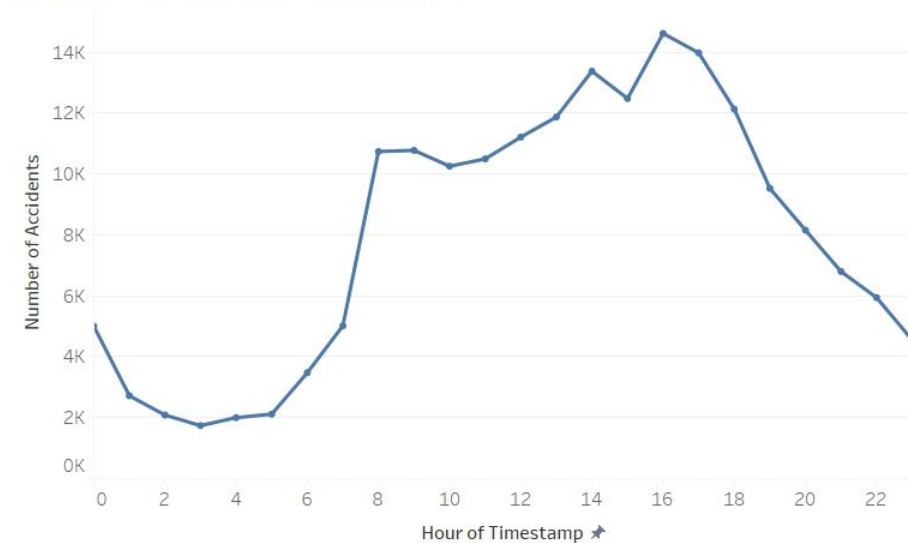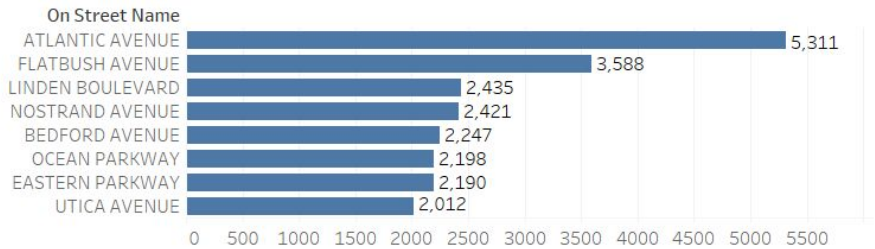

3) More specifically, accidents occur the most during peak hours on the weekday in the morning and late afternoon.
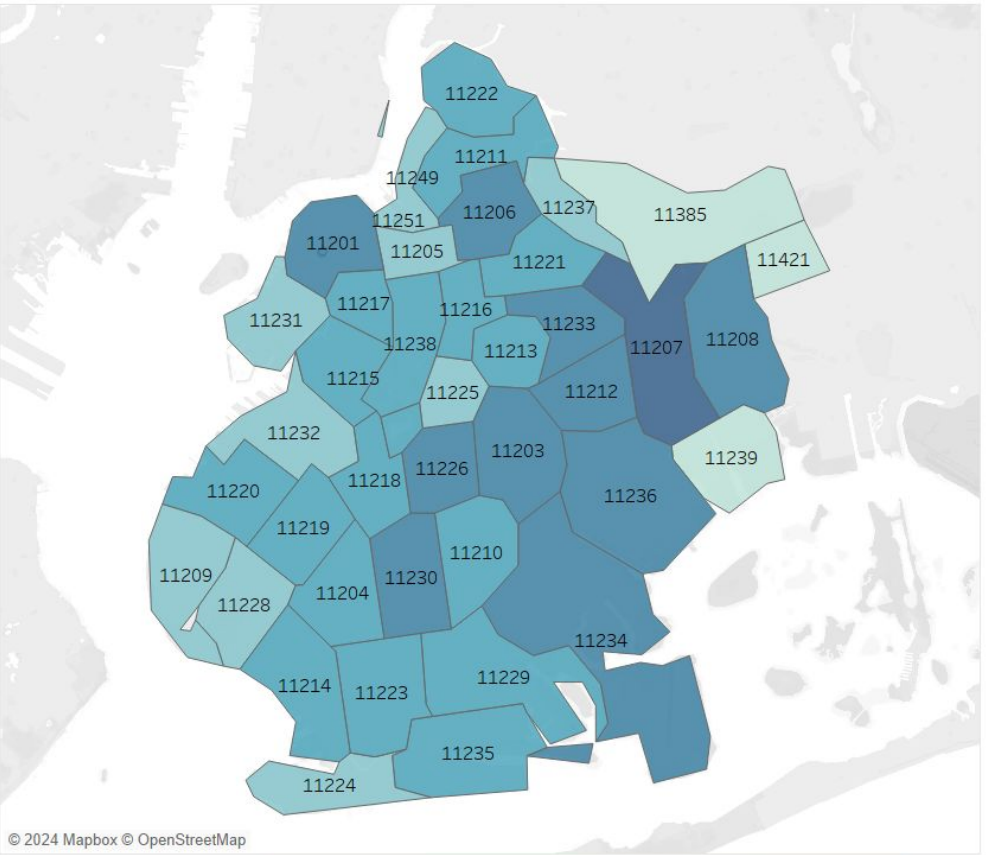
Appendix C.2 Where and What Causes The Accidents?

1) We see that accidents occur the most on major thoroughfares in and out of New York (e.g. Atlantic Ave, Flatbush Ave,..)

2) This also reflected in the Zipcode. For example, 11201 is where both Atlantic Ave and Flatbush Ave connect coming out of Brooklyn Bridge.


3) Lastly, Driver inattention/distraction is a top collusion factor for 1 in 4 accidents
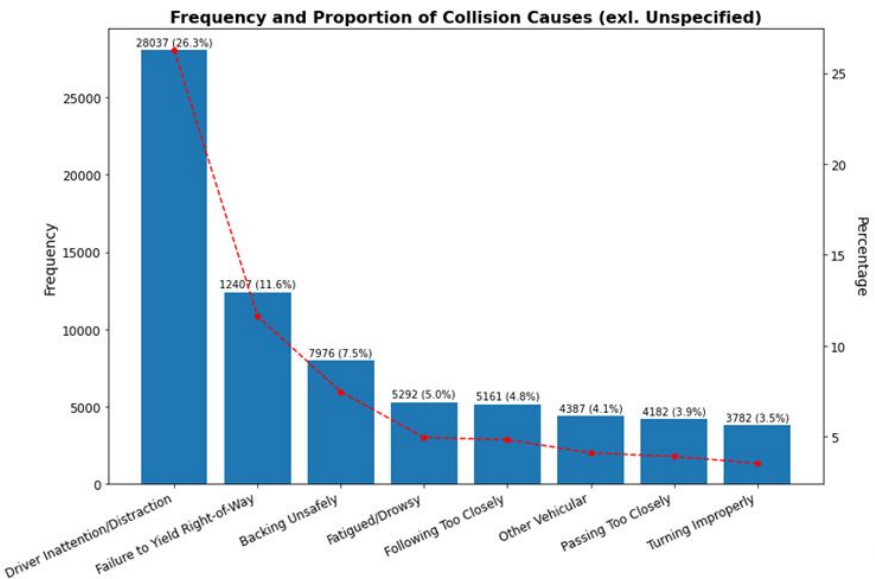
Appendix C.3 Who Got Hurt and Tying It All Back

1) Analyzed the impact of these accidents, it is clear that pedestrians are incredibly vulnerable if involved in a collision caused by a motor vehicle. The risk of them being fatally injured is significantly more than everyone else, while chances were they did not cause the accidents.
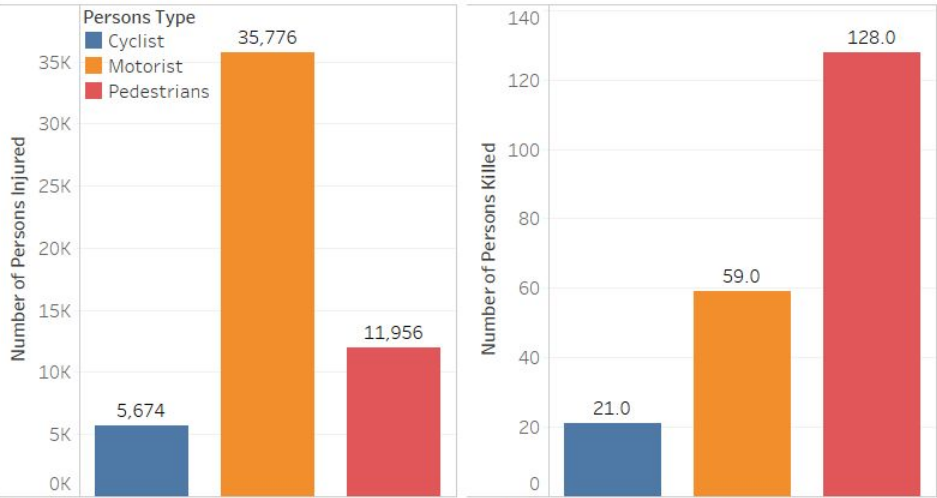
2) The accidents that involve injured pedestrians are consistent with the high-level analysis - occuring during peak hours on weekday
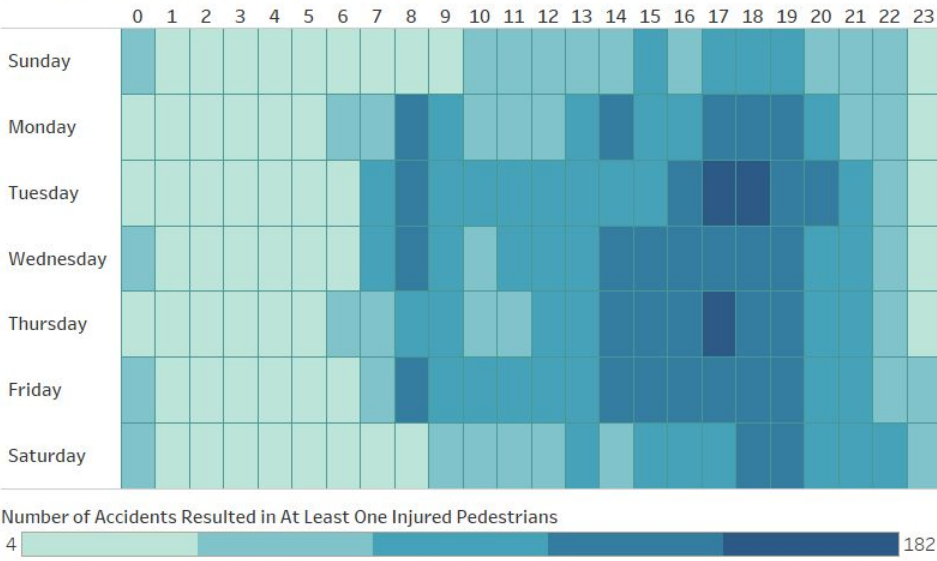
3) Similarly, we see the same street names and leading causes for the accidents. This indicated that many accidents happen due to drivers not paying enough attention to the street, especially during peak hours.