

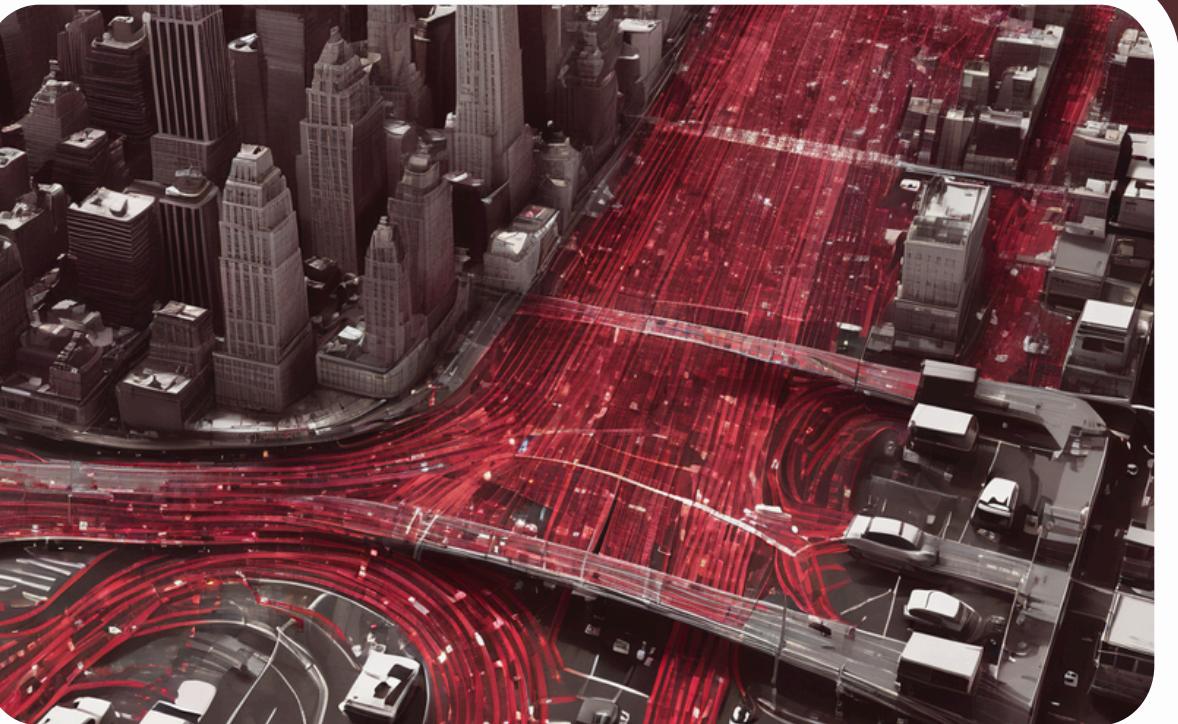
NYC TRAFFIC ACCIDENTS ANALYSIS



A comprehensive Data Study

EFFORTS BY: MRIDUL BHALLA

INTRODUCTION



Objective

To analyze NYC traffic collision data to identify patterns, trends, and hotspots.



Problem Statement

To identify patterns, trends, and hotspots and to come up with solutions to decrease the number of injuries and deaths in New York City.

| time | borough | street_name | cross_street | latitude | longitude | contributing_factor | vehicle_type | persons_injured | persons_killed | pedestrians_injured | prob |
|----------|---------------|-----------------------------------|-------------------|-----------|------------|-------------------------------|-------------------|-----------------|----------------|---------------------|------|
| 12:05:00 | Bronx | Broadway | Kimberly Place | 40.877705 | -73.90566 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 1:00:00 | Brooklyn | Avenue H | Flatbush Avenue | 40.631706 | -73.94649 | Following Too Closely | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 0:13:00 | Brooklyn | Flatlands Avenue | Fountain Avenue | 40.66045 | -73.86719 | Driver Inexperience | Passenger Vehicle | 2 | 0 | 0 | 0 |
| 3:30:00 | Brooklyn | Muffat Street | Bushwick Avenue | 40.684513 | -73.90938 | Traffic Control Disregard | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 3:25:00 | Brooklyn | Evergreen Avenue | NULL | 40.70083 | -73.93314 | Tire Failure/Inadequate | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 3:30:00 | Brooklyn | Bushwick Avenue | NULL | 40.70191 | -73.93699 | Following Too Closely | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 6:42:00 | Brooklyn | Jefferson Street | NULL | 40.703667 | -73.92609 | Fell Asleep | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 5:45:00 | Brooklyn | Oriental Boulevard | NULL | 40.57831 | -73.93927 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 5:17:00 | Manhattan | East 116 Street | NULL | 40.799007 | -73.94255 | Driver Inattention/Distrac | Passenger Vehicle | 2 | 0 | 0 | 0 |
| 6:40:00 | Queens | Long Island Express | NULL | 40.74045 | -73.844376 | Driver Inattention/Distrac | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 2:45:00 | Queens | Long Island Express | NULL | 40.74285 | -73.830284 | Unsafe Speed | Passenger Vehicle | 1 | 0 | 0 | 0 |
| 4:47:00 | Queens | 9 Avenue | NULL | 40.789284 | -73.840805 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 6:30:00 | Queens | Van Wyck Express | NULL | 40.729738 | -73.833725 | Driver Inattention/Distrac | Passenger Vehicle | 3 | 0 | 0 | 0 |
| 6:21:00 | Brooklyn | Brooklyn Avenue | Avenue I | 40.629677 | -73.94151 | Fell Asleep | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 0:38:00 | Brooklyn | Brooklyn Qns Expn Morgan Avenue | NULL | 40.723327 | -73.93927 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 2:35:00 | Brooklyn | India Street | NULL | 40.731842 | -73.95978 | Driver Inattention/Distrac | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 3:07:00 | Queens | Brooklyn Queens E | NULL | 40.738495 | -73.90274 | Driver Inattention/Distrac | Passenger Vehicle | 3 | 0 | 0 | 0 |
| 5:30:00 | Brooklyn | East 100 Street | Avenue M | 40.63862 | -73.89105 | Driver Inattention/Distrac | Passenger Vehicle | 2 | 0 | 0 | 0 |
| 9:22:00 | Brooklyn | Quentin Road | NULL | 40.60703 | -73.96938 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 8:00:00 | Bronx | White Plains Road | Rhinelander Aven | 40.8475 | -73.86696 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 7:55:00 | Queens | Rockaway Boulev Nassau Expressway | NULL | 40.65961 | -73.777834 | Driver Inattention/Distrac | Passenger Vehicle | 1 | 0 | 0 | 0 |
| 7:38:00 | Staten Island | Nonne Avenue | Purdue Court | 40.587205 | -74.15978 | Driver Inattention/Distrac | Passenger Vehicle | 1 | 0 | 0 | 0 |
| 10:57:00 | Bronx | Bainbridge Avenue | NULL | 40.879955 | -73.87888 | Unspecified | Emergency Service | 0 | 0 | 0 | 0 |
| 4:40:00 | Queens | Rockaway Freeway | Ocean Crest Boule | 40.601364 | -73.760056 | Unsafe Speed | Passenger Vehicle | 1 | 2 | 0 | 0 |
| 7:46:00 | Manhattan | Major Deegan Expr | NULL | 40.85944 | -73.915985 | Unsafe Speed | Passenger Vehicle | 2 | 0 | 0 | 0 |
| 3:15:00 | Queens | Farmers Boulevard | NULL | 40.665165 | -73.767876 | Driver Inattention/Distrac | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 3:00:00 | Queens | 230 Street | 148 Avenue | 40.656384 | -73.75305 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 2:50:00 | Queens | Little Neck Parkway | NULL | 40.745 | -73.71803 | Steering Failure | Passenger Vehicle | 1 | 0 | 0 | 0 |
| 6:45:00 | Manhattan | East 46 Street | 2 Avenue | 40.752377 | -73.97035 | Driver Inexperience | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 0:00:00 | Queens | 175 Street | NULL | 40.664538 | -73.77314 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 9:20:00 | Queens | 181 Street | 147 Avenue | 40.660138 | -73.76653 | Driver Inattention/Distrac | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 2:00:00 | Queens | 75 Street | NULL | 40.677547 | -73.86194 | Fell Asleep | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 5:30:00 | Brooklyn | Gowanus Expy (Bq) | NULL | 40.643787 | -74.018845 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 5:25:00 | Queens | Horace Harding Expr | NULL | 40.736317 | -73.857605 | Driver Inattention/Distrac | Passenger Vehicle | 1 | 0 | 1 | 0 |
| 2:21:00 | Queens | 195 Place | 90 Avenue | 40.714607 | -73.76573 | Driver Inattention/Distrac | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 2:30:00 | Queens | Linden Boulevard | NULL | 40.68882 | -73.786354 | Unspecified | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 5:10:00 | Staten Island | Bement Avenue | NULL | 40.628098 | -74.11104 | Unspecified | Passenger Vehicle | 1 | 0 | 0 | 0 |
| 4:29:00 | Queens | 64 Road | Yellowstone Boule | 40.731617 | -73.85149 | Alcohol Involvement | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 7:40:00 | Brooklyn | Ocean Parkway | Avenue D | 40.61196 | -73.96816 | Unspecified | Passenger Vehicle | 1 | 0 | 0 | 0 |
| 0:39:00 | Brooklyn | East 98 Street | Avenue M | 40.637676 | -73.8925 | Unsafe Speed | Passenger Vehicle | 0 | 0 | 0 | 0 |
| 1:02:00 | Manhattan | 2 Avenue | East 101 Street | 40.78744 | -73.94478 | Failure to Yield Right-of-Way | Passenger Vehicle | 2 | 0 | 0 | 2 |



Data Source

The datafile was downloaded from Maven Analytics Data Playground.

Link: [Maven Analytics Dataset](#)

APPROACH & METHODOLOGY

These are the steps that have been followed throughout the project.

- 01 Data Exploration - Understanding the Problem Statement and the Dataset
- 02 Data Cleaning - Preparing the Dataset and Cleaning it using MySQL
- 03 Data Analysis - Executing SQL queries to extract insights on collisions by borough, contributing factors, and severity.
- 04 Data Visualization - Using MS Excel to create a comprehensive dashboard with diverse charts.
- 05 Gaining Insights - Identifying high-risk areas, times, and common causes of collisions.
- 06 Providing Recommendations - Suggesting measures to improve road safety based on the analysis.

01

Data Exploration

The data contains Motor vehicle collisions reported by the New York City Police Department from January 2021 to April 2023. Each record represents an individual collision, including the date, time, and location of the accident (borough, zip code, street name, latitude/longitude), vehicles and victims involved, and contributing factors.

| Field | Description |
|----------------------------|--|
| Collision ID | Unique record code generated by system |
| Date | Occurrence date of collision |
| Time | Occurrence time of collision |
| Borough | Borough where collision occurred |
| Street Name | Street on which the collision occurred |
| Cross Street | Nearest cross street to the collision |
| Latitude | Latitude coordinate for Global Coordinate System WGS 1984 decimal degrees (EPSG 4326) |
| Longitude | Longitude coordinate for Global Coordinate System WGS 1984 decimal degrees (EPSG 4326) |
| Contributing Factor | Factors contributing to the collision for designated vehicle |
| Vehicle Type | Type of vehicle |
| Persons Injured | Total number of persons injured (pedestrians + cyclists + motorists) |
| Persons Killed | Total number of persons killed (pedestrians + cyclists + motorists) |
| Pedestrians Injured | Number of pedestrians injured |
| Pedestrians Killed | Number of pedestrians killed |
| Cyclists Injured | Number of cyclists injured |
| Cyclists Killed | Number of cyclists killed |
| Motorists Injured | Number of vehicle occupants injured |
| Motorists Killed | Number of vehicle occupants killed |

1. Removing Duplicates

We checked for duplicate entries to ensure data accuracy.

2. Handling Missing Values

- We found missing values in “street_name”, “cross_street” and “contributing_factor” columns.
- We converted the blank values in these columns to null values and we imputed the null values of “contributing_factor” column to “Unspecified”

3. Adding New Columns

We created 6 new columns which can be seen in the picture below.

```
-- 1. Remove Duplicates if any -----  
-- Checking for duplicates if any  
SELECT collision_id, COUNT(collision_id)  
FROM collisions_staging  
GROUP BY collision_id  
HAVING COUNT(collision_id) > 1;  
-- no duplicates found
```

| day_name | time_category | season | holiday_indicator | collision_severity | month_name |
|-----------|---------------|--------|-------------------|--------------------|------------|
| Wednesday | Afternoon | Winter | 0 | Minor | February |
| Thursday | Morning | Winter | 0 | Minor | January |
| Friday | Late Night | Winter | 1 | Severe | January |
| Friday | Early Morning | Winter | 1 | Minor | January |

Using SQL, we performed an in-depth analysis to answer various questions related to traffic collisions in NYC like:

Trends over the time

1. Trend of collisions over time
2. Are there any trends in the number of injuries or fatalities over time?
3. Are there seasonal variations in collision rates?
4. Are collision rates higher on holidays compared to regular days?
5. Most collisions happen during which time category?
6. Is there any relation between time category and collision severity?

Other Questions

7. What are the collision hotspots in NYC based on latitude and longitude?
8. What is the collision rate by Borough?
9. What are the most common contributing factors to collisions?
10. Is there any correlation between contributing factors and collision severity?
11. Which type of vehicle is involved in the most collisions?
12. Do certain vehicle types have a higher likelihood of causing fatalities?
13. What is the distribution of injuries and fatalities across different collision types (pedestrians, cyclists, motorists)?



03

Data Analysis in SQL

Kindly refer to the attached SQL File, for the entire analysis. Here are a few screenshots of the same:

```
4 -- Q1. How many collisions occurred each year?
5 • SELECT YEAR(collision_date) AS Year, COUNT(*) AS CollisionCount
6   FROM collisions_staging
7   GROUP BY YEAR(collision_date)
8   ORDER BY Year;
9   -- Maximum occurred in 2021 - 100885, followed by 2022 - 92980, and significantly lesser in 2023 - 22232
10

11 0% 30:16 |
```

result Grid Filter Rows: Search Export:

| Year | CollisionCount |
|------|----------------|
| 2021 | 100885 |
| 2022 | 92980 |
| 2023 | 22232 |

```
8 -- Q6. Collision Rate by Borough?
9 • SELECT borough, COUNT(*) AS TotalCollisions, COUNT(*) / (SELECT COUNT(*) FROM collisions_staging) AS CollisionRate
0   FROM collisions_staging
1   GROUP BY borough
2   ORDER BY CollisionRate DESC;
3   -- Borough with the highest number of collisions and collision rate is:
4   -- Brooklyn with total collisions as 73005 and collision rate as 0.3378
5
%
C | 1:55
```

Result Grid Filter Rows: Search Export:

| borough | TotalCollisions | CollisionRate |
|---------------|-----------------|---------------|
| Brooklyn | 73005 | 0.3378 |
| Queens | 56924 | 0.2727 |
| Bronx | 37781 | 0.1748 |
| Manhattan | 35956 | 0.1664 |
| Staten Island | 10431 | 0.0483 |

04

Data Visualisation in MS Excel

We used MS Excel to transform the analyzed data into a comprehensive and interactive dashboard, enabling stakeholders to easily interpret and explore the data.

- The Excel Dashboard has 3 views:

1. Dashboard View - which shows the main dashboard
2. Insights View - which showcases some key insights
3. Data View - which takes the user to the data used behind the dashboard

- Some of the Key visualisations and charts used in the dashboard are:

1. Line Chart - To showcase the collision trend over time
2. Geographical Map - To showcase the collision by Borough
3. Bar Charts - To show collisions by contributing factor and vehicle
4. Heatmap - To show collision by date and time category
5. Radial Bar Chart - To show Injuries and Deaths by road users

- Some Slicers for filtering include: Year, Season, Time Category, Collision Severity, and Borough.

- Some KPIs include: Total number of collisions, Total number of Injuries and Total number of deaths, Average number of collisions per day, No. of deaths per 1000 collisions, No. of injuries per 1000 collisions

04

Data Visualisation in MS Excel

NYC Traffic Accidents Analysis Dashboard

Total Collisions: 216097 **Total Injuries: 104334** **Total Deaths: 544**

Season: Fall Spring Summer Winter

Time Category: Afternoon Early Morning Evening
Late Night Morning Night

Severity: Fatal Minor Severe

Borough: Bronx Brooklyn Manhattan Queens Staten Island

Collision by Contributing Factor:

| Contributing Factor | No. of Collisions | No. of Deaths & Injuries |
|--------------------------------|-------------------|--------------------------|
| Driver Inattention/Distracted | 53356 | 15325 |
| Failure to Yield Right of Way | 12609 | 9580 |
| Following Too Closely | 8558 | 8558 |
| Passing or Lane Usage Improper | 15325 | 12609 |
| Passing Too Closely | 9580 | 8558 |

Collision by Vehicle:

| Vehicle Type | No. of Collisions | No. of Deaths & Injuries |
|-------------------|-------------------|--------------------------|
| Transport | 5912 | 1732 |
| Taxi | 5464 | 3538 |
| Passenger Vehicle | 182626 | 85641 |
| Bus | 3872 | 1803 |
| Bicycle | 5024 | 4661 |

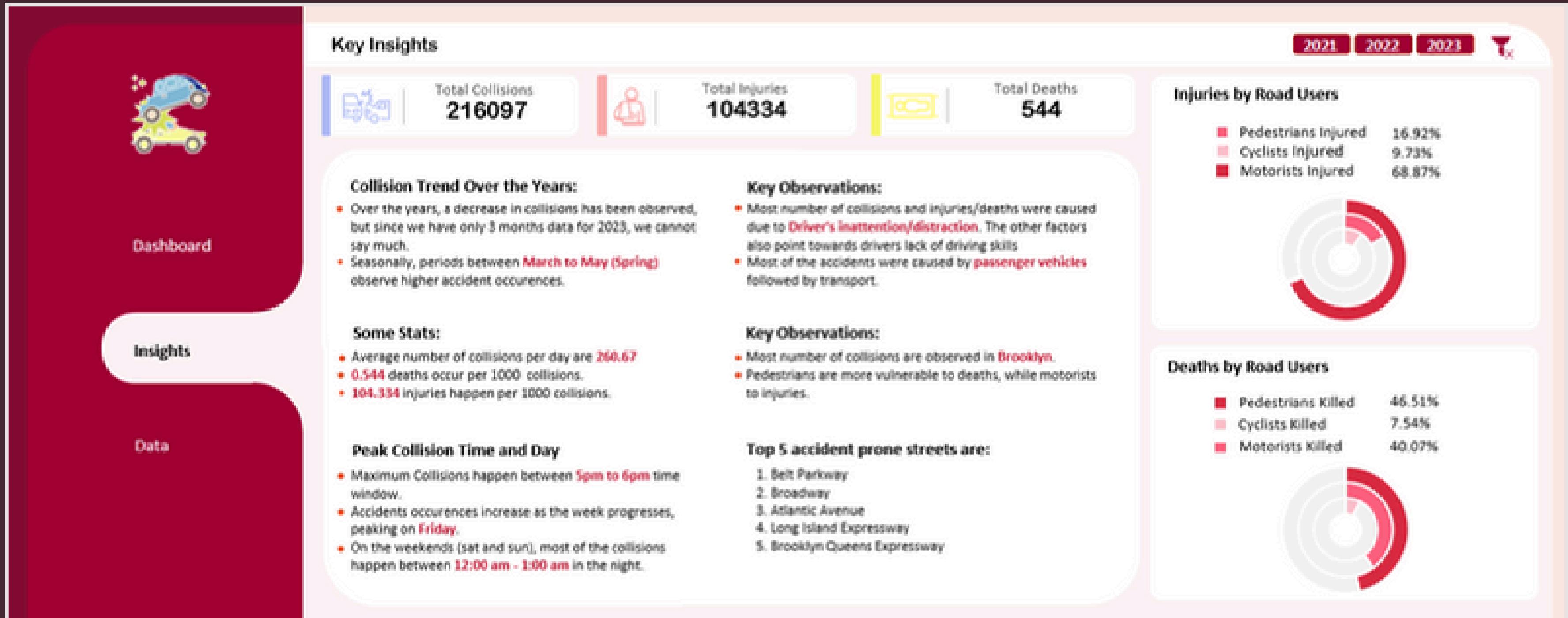
Collision by Borough:

Collision Trend Over Time:

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---------------|-------|-------|-------|-------|-------|------|------|------|------|------|------|------|
| Early Morning | 1551 | 1067 | 1131 | 1223 | 1477 | 2812 | 3035 | 2947 | 4110 | 3471 | 4634 | 4110 |
| Morning | 8297 | 8594 | 8418 | 8411 | 8405 | 5947 | 5947 | 5947 | 5947 | 5947 | 5947 | 5947 |
| Afternoon | 10626 | 11360 | 11466 | 11326 | 12183 | 9579 | 9471 | 9471 | 9471 | 9471 | 9471 | 9471 |
| Evening | 4912 | 4627 | 4699 | 4850 | 5238 | 4634 | 4634 | 4634 | 4634 | 4634 | 4634 | 4634 |
| Night | 2857 | 3334 | 3134 | 3245 | 4199 | 4148 | 3516 | 3516 | 3516 | 3516 | 3516 | 3516 |
| Late Night | 2541 | 1901 | 2000 | 2158 | 2578 | 3081 | 4201 | 4201 | 4201 | 4201 | 4201 | 4201 |

Collision by Day & Time Category:

| Category | Mon | Tues | Wed | Thurs | Fri | Sat | Sun |
|---------------|-------|-------|-------|-------|-------|------|------|
| Early Morning | 1551 | 1067 | 1131 | 1223 | 1477 | 2812 | 3035 |
| Morning | 8297 | 8594 | 8418 | 8411 | 8405 | 5947 | 5947 |
| Afternoon | 10626 | 11360 | 11466 | 11326 | 12183 | 9579 | 9471 |
| Evening | 4912 | 4627 | 4699 | 4850 | 5238 | 4634 | 4634 |
| Night | 2857 | 3334 | 3134 | 3245 | 4199 | 4148 | 3516 |
| Late Night | 2541 | 1901 | 2000 | 2158 | 2578 | 3081 | 4201 |



Our analysis revealed several critical insights into the patterns and trends of traffic collisions in NYC, which are as follows:

Collision Trends over the time

- Over the years, a decrease in collisions has been observed.
- There is a noticeable seasonal trend in collisions, with higher numbers occurring during winter months.
- Maximum Collisions occur between the 5pm - 6pm time window.
- Accident occurrences increase as the week progresses peaking on Friday.
- Average number of collisions per day is 260.67

Geographical Insights

- The highest number of collisions occur in **Manhattan**, followed by **Brooklyn**.
- The top three collision hotspots in NYC are:
 1. Latitude: **40.861862**, Longitude: **40.861862**, num of collisions: 137
 2. Latitude: **40.675735**, Longitude: **-73.896860**, num of collisions: 131
 3. Latitude: **40.658577**, Longitude: **-73.890630**, num of collisions: 106
- Top 2 accident prone streets are: **Belt Parkway and Broadway**

Collision Severity and Contributing factors

- **Distracted driving**, failure to yield the right way, and following too closely are the top contributing factors to collisions.
- Pedestrians account for a significant portion of fatalities, while motorists are more frequently injured.
- Most number of accidents are caused by **passenger vehicles** followed by **transport**.
- **0.544 deaths** occur per 1000 collisions and **104.334 injuries** occur per 1000 collisions

Recommendations

1

Road maintenance activities, such as salting and snow removal, should be increased during the winter months.

2

More stringent traffic regulations and fines should be implemented in Manhattan and Brooklyn.

3

Police patrols and sobriety checkpoints should be increased during peak accident hours like 5pm - 6pm.

4

Stricter penalties should be implemented for distracted driving, and technology solutions, such as speed cameras and mobile phone usage detectors should be introduced.

5

Infrastructure improvements, such as better signage and road designs should be done in the accident prone areas.

6

More visible traffic light signals, pedestrian barriers and crosswalk markings should be installed at major intersections.

7

Focus should be on providing a more robust and efficient public transit system to encourage usage by commuters

8

A task force can be established to review collision data regularly and recommend necessary changes.



THANK YOU

Kindly refer to the SQL files and Excel dashboard for more in-depth details. Also please read the “README” section of the github repository.