RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Bhavana Vemula
Layakishore Reddy Desireddy
Vikram Sriram Yasaswi Sagaram

NYC Motor Vehicle Collisions

INTRODUCTION AND OBJECTIVE

This project analyzes comprehensive data on motor vehicle collisions in New York City, aiming to enhance traffic safety and reduce fatalities. Utilizing the NYPD's Finest Online Records Management System (FORMS), which captures detailed crash data from the MV-104AN forms, we conducted in-depth statistical analyses to identify trends and risk factors associated with traffic incidents.

Data Retrieval

A Python script was developed to retrieve detailed traffic collision data from New York City's open data portal via the Socrata Open Data API (SODA). A specific query was crafted to select all reported traffic collisions since 2013, ordered by the date of occurrence in descending order. The query was URL-encoded to ensure compatibility with web requests, facilitating the efficient fetching and analysis of the most recent and relevant data for the traffic safety study.

Data Attributes

- crash_date
- crash_time
- borough
- zip_code
- latitude
- longitude
- location
- on_street_name
- off_street_name
- 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
- contributing_factor_vehicle_1
- contributing_factor_vehicle_2
- collision_id
- vehicle_type_code1
- vehicle_type_code2
- cross_street_name
- contributing_factor_vehicle_3
- contributing_factor_vehicle_4
- vehicle_type_code_3
- vehicle_type_code_4
- contributing_factor_vehicle_5
- vehicle_type_code_5
- date/time

Data Cleaning & Transformation

The dataset underwent a thorough cleaning process where date and time data were parsed and converted to a standardized datetime format. This transformation was critical for accurate time series analysis and reporting on traffic incidents. Extraneous records lacking essential geographical coordinates were removed, and all column names were standardized by converting to lowercase and replacing spaces with underscores, which streamlined subsequent data processing stages.
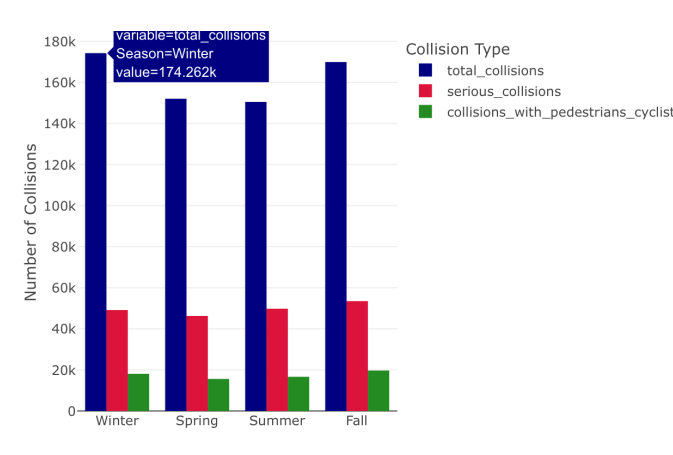
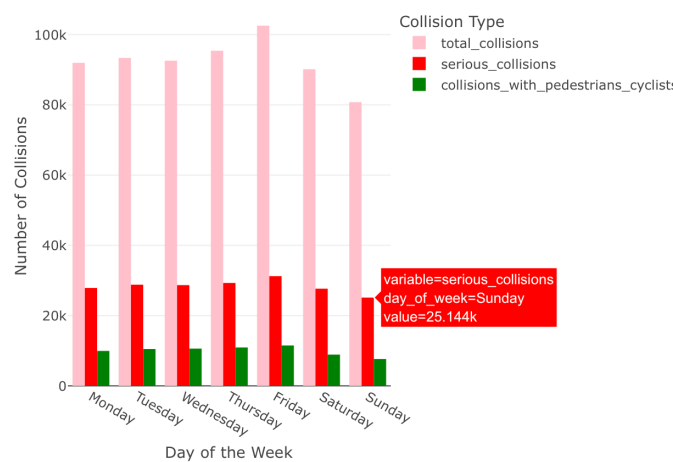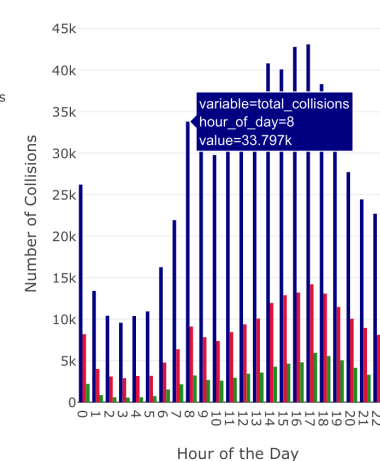## VISUALIZATIONS

Total Collisions - Yearly
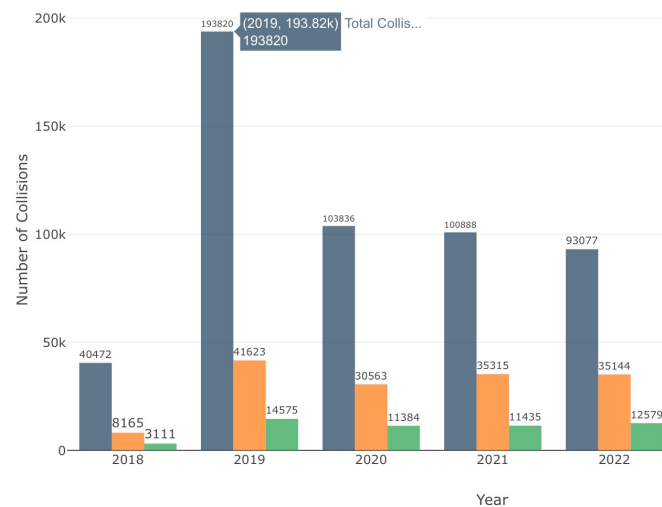
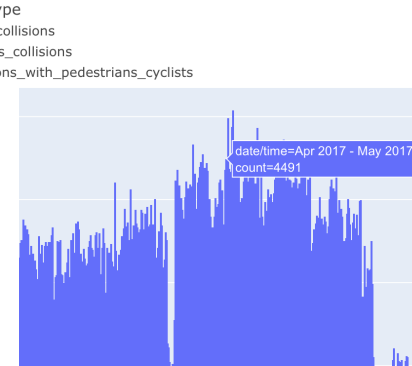**Monthly Collisions**
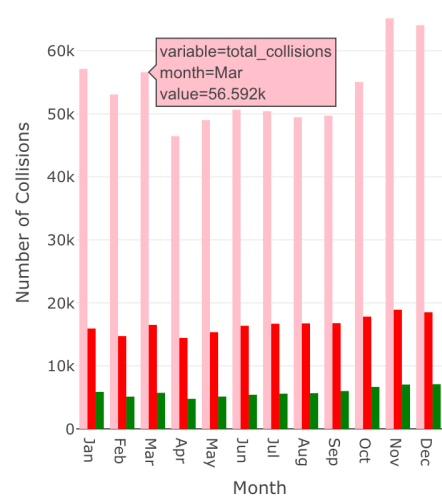
Seasonal Collisions
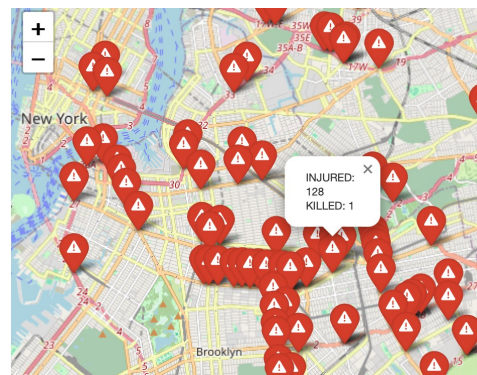

Weekly Collisions

Hourly Collisions

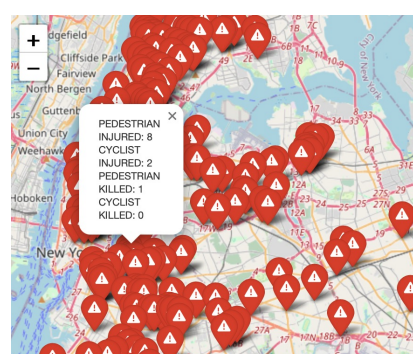**Incidents Distributed over time**
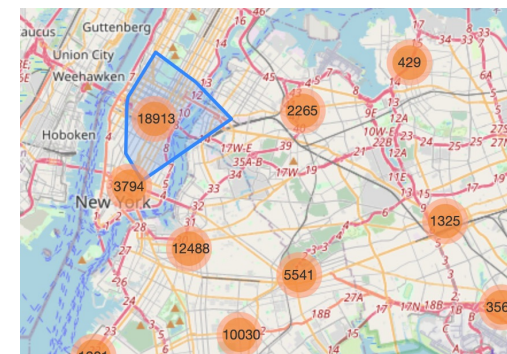

Dangerous Single points

Single points with a high number of injuries and deaths from vehicle collisions

Dangerous Single Points for Pedestrians and Cyclists
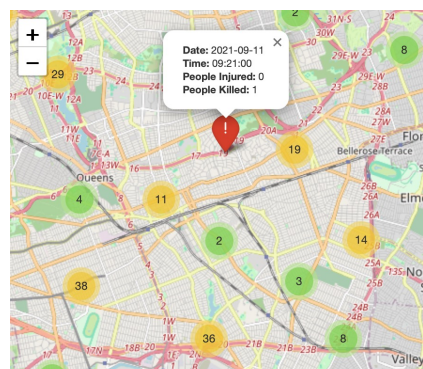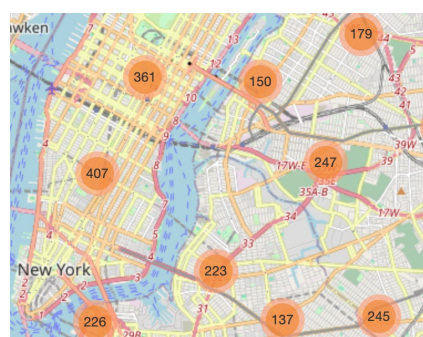

Single points with a high number of pedestrian / cyclist injuries and deaths from vehicle collisions.
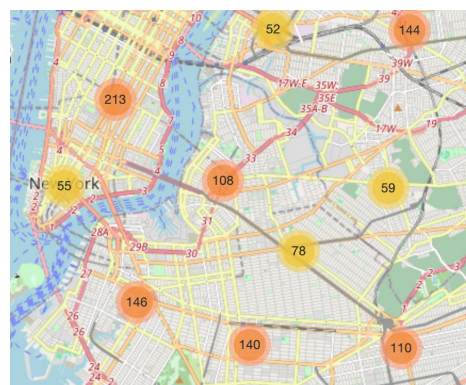
Dangerous Areas by Clustering

Dangerous areas are clusters of collisions with a high number of injuries and deaths from vehicle collisions. Clusters were identified by tuning a density-based clustering algorithm.
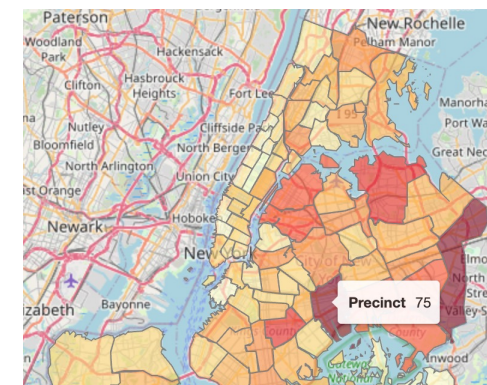
Serious Collisions in 2024(Latest)

Collisions with Pedestrians and Cyclists in 2024(Latest)

Serious Collisions by NYPD Precinct

Average annual serious collisions per precinct, and the number displayed is the precinct number.

Fatal Collisions

Collisions where at least one person is killed

Git Commit History

Graph	Description	Author	Commit
	Uncommitted Changes (1)		
	main origin origin/HEAD Created a dashboard	Layakishore Reddy	1171196e
	origin/gh-pages Built site for gh-pages	Layakishore Reddy	2e323589
	sample code for dashboard and added results	Layakishore Reddy	0cab46be
	added code required for viz and all viz are present	Layakishore Reddy	05f3e260
	added code for summary stat	Layakishore Reddy	49ae4c6e
	backup external service	Sriram Yasaswi	2b99b16e
	added pytest to test quality and format conversion	Sriram Yasaswi	47d7abcf
	'enrichment'	Sriram Yasaswi	0a0d4c48
	ml	Bhavana Vemula	99cccb79
	cleaning	Bhavana Vemula	5b6e67ee
	tabular	Bhavana Vemula	cea37424
	retrival	Bhavana Vemula	02cccc12
	introduction	Bhavana Vemula	441a0a1a
	Built site for gh-pages	Bhavana Vemula	6950b19b
	Update index.qmd	Bhavana-Vemula	795bf3c8
	Built site for gh-pages	Bhavana-Vemula	fc54b682
	Update _quarto.yml	Bhavana-Vemula	4db5f8f9
	Built site for gh-pages	Bhavana-Vemula	3c10feda
	Update README.md	Bhavana-Vemula	c3a1a595
	9,11 dan	Bhavana Vemula	d617760d
	Made Change For Question 4 and 5	Sriram Yasaswi	a77670b6
	made changes in Q2,Q3	Layakishore Reddy	874fcc04
	1st doc	Bhavana Vemula	85065225
	Setting up GitHub Classroom Feedback	github-classroom...	5797055c
	origin/feedback GitHub Classroom Feedback	github-classroom...	970576c8
	Initialize gh-pages	github-classroom...	30bb4e9c
	Initial commit	github-classroom...	9981970c

Inferences

• Serious collisions, including collisions where pedestrians or cyclists are injured or killed is fairly widespread throughout NYC.
• Clustering can be used to reveal multi-block areas with elevated injuries and deaths that would be missed by tracking high-collision intersections.
• The 105th and 109th precincts in Queens and the 75th precinct in Brooklyn lead the city in total number of collisions.
• Areas with higher numbers of collisions tend to have higher numbers injuries or deaths (motorist, pedestrian, and cyclist).
• However pedestrian and cyclist injuries and deaths are not as strongly linked to the number of vehicle collisions.
• Surprisingly, pedestrian and cyclist injuries and deaths do not seem to correspond to density, suggesting that other factors like street design, signage, and traffic enforcement are important to reducing injuries and deaths.
• North Queens, much of Brooklyn, and midtown Manhattan stand out as having high numbers of pedestrian and cyclist injuries and deaths.

Scan for more Interactive Visualizations