Reveal Crash Severity: A Comprehensive Analysis of Location, Weather, and Time Factors

Methods of Advanced Data Engineering

Author: Md Junayed Miah (23120467)

Introduction:

Motor vehicle crashes have significant impacts on public safety, infrastructure, and urban planning. This project seeks to identify critical factors contributing to motor vehicle crashes, understand their geographical pattern or trend and temporal distribution, and evaluate the severity of incidents based on contributing factors such as weather, road conditions, and driver behavior. The results aim to provide meaningful insights for stakeholders and policy makers in transportation safety and urban policy planning.

Used Data:

Data Source-1: Montgomery County Crash Data Source:

The dataset provides me with detailed information about motor vehicle crashes in Montgomery County, including crash date, route type, road conditions, weather, and driver at fault.

- **Url:** **Crash Reporting – Driver Data**.
- **Data Structure:**
    1. **Columns:** crash_date/time, route_type, road_name, collision_type, weather, surface_condition, light, latitude, longitude etc.
    2. **Data Types:** Date (datetime64), borough (strings), and number_of_persons_injured (integers).
- **License:** Open Data Common Public Domain Dedication and License (ODC PDDL).
    1. **Obligation:** Proper attribution and ensuring transparency in usage.
    2. **Link:** License details on **Montgomery County Open Data Portal**.

Data Source-2: New York City Vehicle Collision Data Source:

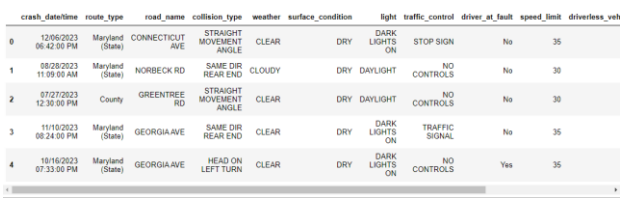
This dataset offers comprehensive data on motor vehicle collisions in New York City, including details about crash dates, times boroughs, injured or killed persons, and vehicle types involved.

- **Url:** **NYC Vehicle Collision Data**.
- **Data Structure:**
    1. **Columns:** crash_date, crash_time, borough, number_of_persons_injured, number_of_persons_killed, latitude, longitude etc.
    2. **Data Types:** Date (datetime64), borough (strings), and number_of_persons_injured (integers).

- **License:** Available through NYC Open Data, aligning with open-data principles, ensuring it can be used without legal constraints.
    1. **Obligations:** Use data responsibly and attribute the source.
    2. **Link:** NYC Open Data **terms and conditions**.

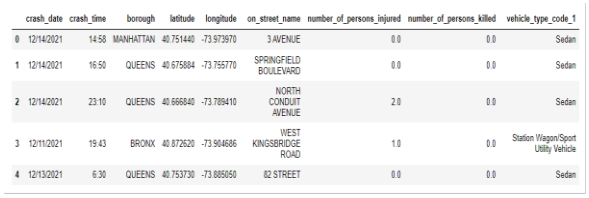
Analysis

The analysis used crash data from Montgomery County (figure-1) and New York City (figure-2) to examine patterns and correlations in vehicle accidents. The datasets were merged on latitude and longitude to integrate insights, with care taken to resolve inconsistencies. Exploratory analysis focused on trends over time, geospatial clustering, weather impacts, and severity metrics, providing a comprehensive understanding of factors influencing crash occurrences and outcomes.



	crash_date/time	route_type	road_name	collision_type	weather	surface_condition	light	traffic_control	driver_at_fault	speed_limit	driverless_veh
0	12/06/2023 06:42:00 PM	Maryland (State)	CONNECTICUT AVE	STRAIGHT MOVEMENT ANGLE	CLEAR	DRY	DARK LIGHTS ON	STOP SIGN	No	35	
1	08/28/2023 11:09:00 AM	Maryland (State)	NORBECK RD	SAME DIR REAR END	CLOUDY	DRY	DAYLIGHT	NO CONTROLS	No	30	
2	07/27/2023 12:30:00 PM	County	GREENTREE RD	STRAIGHT MOVEMENT ANGLE	CLEAR	DRY	DAYLIGHT	NO CONTROLS	No	30	
3	11/10/2023 08:24:00 PM	Maryland (State)	GEORGIA AVE	SAME DIR REAR END	CLEAR	DRY	DARK LIGHTS ON	TRAFFIC SIGNAL	No	35	
4	10/16/2023 07:33:00 PM	Maryland (State)	GEORGIA AVE	HEAD ON LEFT TURN	CLEAR	DRY	DARK LIGHTS ON	NO CONTROLS	Yes	35	

Figure-1: Pipeline Output of Crash-reporting



	crash_date	crash_time	borough	latitude	longitude	on_street_name	number_of_persons_injured	number_of_persons_killed	vehicle_type_code_1
0	12/14/2021	14:58	MANHATTAN	40.751440	-73.973970	3 AVENUE	0.0	0.0	Sedan
1	12/14/2021	16:50	QUEENS	40.675884	-73.755770	SPRINGFIELD BOULEVARD	0.0	0.0	Sedan
2	12/14/2021	23:10	QUEENS	40.666840	-73.789410	NORTH CONDUIT AVENUE	2.0	0.0	Sedan
3	12/11/2021	19:43	BRONX	40.872620	-73.904686	WEST KINGSBRIDGE ROAD	1.0	0.0	Station Wagon/Sport Utility Vehicle
4	12/13/2021	6:30	QUEENS	40.753730	-73.885050	82 STREET	0.0	0.0	Sedan

Figure-2: Pipeline Output of Vehicle Collision

Results and Interpretation

Crashes by Year:

Analyzing crash frequency over years in NYC revealed a downward trend, as shown in Figure 1. This trend suggests improving road safety measures over time, but fluctuations in specific years indicate further exploration of external factors (e.g. weather or policy changes).

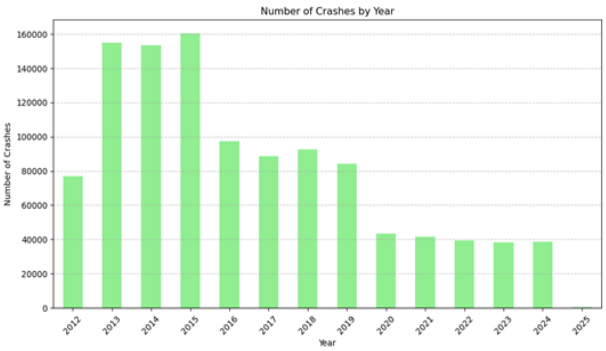


Figure-3: Crashes by Year

Injuries and Fatalities by Borough: NYC data Revealed that Manhattan reported the highest number of injuries, whereas Brooklyn had the most fatalities (figure-4). These findings call for targeted safety initiatives based on the specific challenges of each borough.

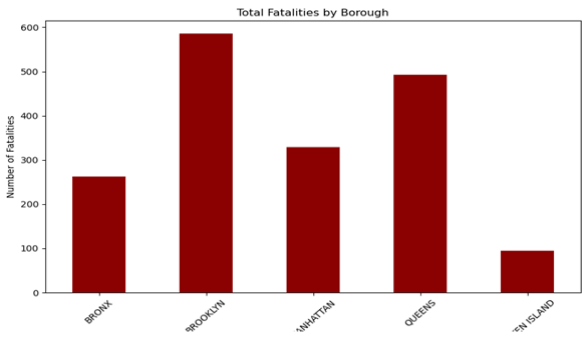


Figure-4: Injuries and Fatalities by Borough

Impact of Weather and Surface Conditions: Clear weather had the highest crash frequency, followed by rainy and foggy conditions (figure-5). Wet and icy roads showed a sharp increase in crashes (figure-6).

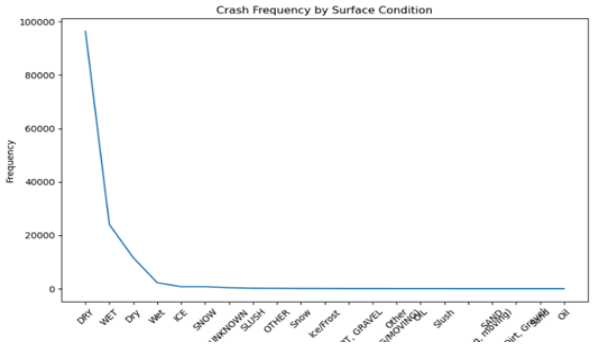


Figure-5: Crashes for surface

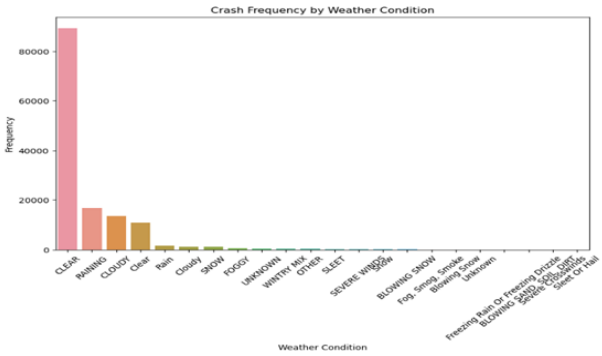


Figure-6: Crashes for weather

Speed Limit Impact:

Higher crash frequencies were observed at speed limits of 30-40 mph (figure-7). This suggests that areas with these limits, such as urban and suburban roads, require enhanced traffic calming measures.

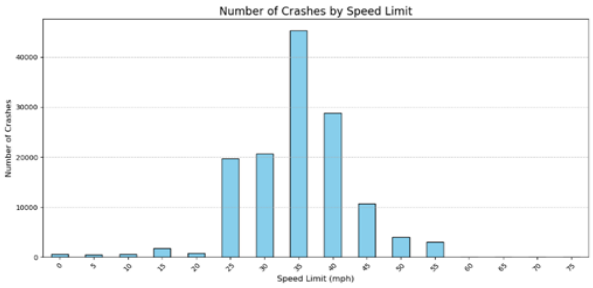


Figure-7: Impact of Speed Limit

Comparison of Crashes by Year:

Here we found distinct trends between the two datasets. While dataset 1 shows a stable pattern of crashes over the years, dataset 2 demonstrates a sharp rise until 2015, followed by a consistent decline, likely reflecting the impact of safety measures or reporting changes. The significantly higher crash counts in NYC also reflect its larger urban density and traffic volume.

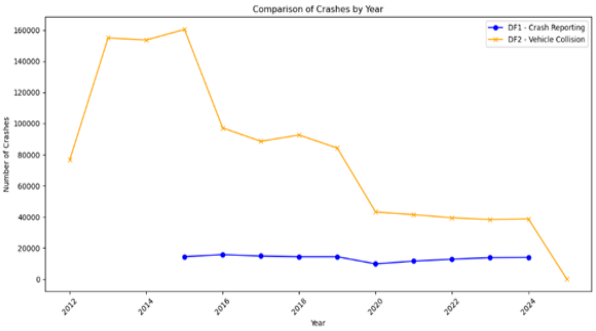


Figure-8: Crashes by Year (Comparison)

Correlation of numerical factors with severity: The heatmap (figure-9) visualizes the correlation between the number of injuries, fatalities, and crash severity. It shows that as the severity of a crash increases, both injuries and fatalities tend to rise. This highlights the direct relationship between crash severity and the outcomes of accidents, emphasizing the need for improved safety measures in high-severity areas.

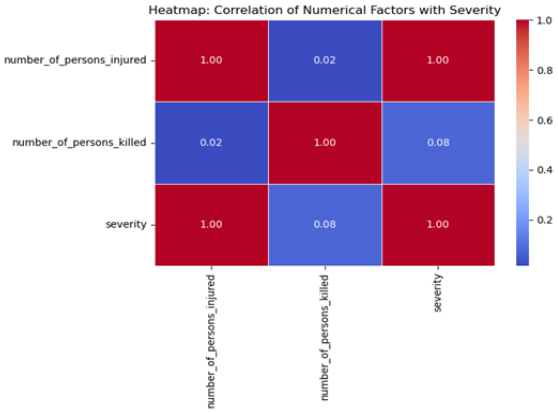


Figure-9: Correlation of severity

Conclusions

This analysis provides actionable insights into crash patterns and their contributing factors. Crashes exhibit temporal peaks, suggesting external factors like policy changes, weather anomalies, or population behavior. Adverse weather and road conditions significantly contribute to crashes. High-density urban areas are hotspots for crashes, indicating the need for localized interventions. Speed limits, borough characteristics, and driver behavior strongly influence crash severity. Incorporate additional datasets, such as traffic volume and road quality, for a holistic analysis. Develop predictive models to identify high-risk areas and conditions proactively.

Limitations

While the analysis provides valuable insights, it is limited by potential biases in data reporting and missing entries. For example:

- The datasets do not include information on unreported minor crashes.
- Factors like driver distractions (e.g., mobile phone use) were not directly analyzed.