### **ANALYSIS REPORT**

### **NewYork City Motor Vehicle Collision Data Analysis**

#### **Overview:**

This analysis document provides a comprehensive overview of the motor vehicle collision data for New York City, covering observations from July 1, 2012, to March 22, 2024. The dataset encompasses a wide range of variables that detail the circumstances, outcomes, and involved parties in motor vehicle collisions across the city.

#### **Dataset Statistics:**

Number of Observations: 2,075,427

Number of Variables: 29

Missing Cells: 17,761,579 (29.5%)

Duplicate Rows: 0

Total Size in Memory: 459.2 MiB

Variable Types:

DateTime: 2

Categorical: 6

Unsupported: 1

Numeric: 8

Text: 12

#### **Key Variables and Insights:**

* **Crash Date and Time**: Data spans from July 1, 2012, to March 22, 2024, with crash times recorded on March 26, 2024, indicating the dataset's recent update. This comprehensive timeframe allows for detailed temporal analysis of trends and patterns in motor vehicle collisions.
* **Location Data**: Boroughs and ZIP codes provide a geographic breakdown of incidents, though both have significant missing data (31.1%). Latitude and longitude data offer a precise mapping of collision sites, with notable gaps in recording.
* **Injuries and Fatalities**: The dataset details the number of persons injured and killed, including specific counts for pedestrians, cyclists, and motorists. A significant majority of collisions result in no fatalities, highlighting the non-lethal nature of most incidents.
* **Vehicle Information**: Extensive details on up to five vehicles involved per collision, including type codes and contributing factors, though there's a notable increase in missing data for variables related to additional vehicles involved.
* **Contributing Factors**: Categorized for up to five vehicles involved, the data identifies factors contributing to the collision, with a large proportion of entries unspecified, especially as the number of vehicles involved increases.

**Observations:**

* A significant portion of the dataset's entries is dedicated to the precise location (latitude and longitude) and timing (date and time) of collisions, emphasizing the importance of temporal and spatial analysis in understanding traffic incidents.
* Missing data, particularly in borough, ZIP code, and vehicle-specific variables, indicates challenges in data collection or reporting, potentially affecting analyses related to geographic trends or vehicle types.
* The data reveals the frequent occurrence of non-fatal injuries in collisions and provides a granular view of the impact on different road users, including pedestrians, cyclists, and motorists.
* The absence of duplicate rows underscores the dataset's integrity, ensuring a reliable basis for analysis.

**Conclusion:**

The New York City motor vehicle collision dataset presents a rich source of information for analyzing traffic incidents' spatial, temporal, and categorical dimensions. Despite challenges posed by missing data, the dataset's breadth and depth offer valuable insights into the city's road safety dynamics, contributing factors to collisions, and the differential impacts on various road users.

This analysis aims to provide stakeholders, including city planners, policymakers, and safety advocates, with a foundation for informed decision-making towards enhancing road safety in New York City

### **Austin Traffic Incident Analysis Report**

**Overview:**

This report presents an analysis of a comprehensive dataset detailing traffic incidents, capturing various aspects such as incident identifiers, date and time of crashes, location data, severity, and outcomes. With 147,750 observations and 54 variables, the dataset offers a rich source for understanding traffic patterns, incident severity, and safety concerns.

**Dataset Statistics:**

Number of Observations: 147,750

Number of Variables: 54

Missing Cells: 1,725,084 (21.6%)

Duplicate Rows: 0

Total Size in Memory: 60.9 MiB

**Variable Types:**

Numeric: 17

Boolean: 11

DateTime: 2

Text: 7

Unsupported: 2

Categorical: 15

**Key Insights:**

* **Incident Identifiers**: The crash\_id variable is unique for each incident, ensuring reliable identification.
* **Incident Severity**: The boolean variable crash\_fatal\_fl indicates the majority of crashes (146,884) were not fatal, with 866 incidents resulting in fatalities.
* **Location and Timing**: Geographic data (latitude and longitude) show a high percentage of missing values, challenging spatial analysis. The crash date spans from March 26, 2014, to March 11, 2024, indicating a decade of recorded incidents.
* **Severity and Outcomes**: The dataset includes detailed counts of various injury levels and fatalities, providing insight into the human impact of these incidents. It highlights the prevalence of non-fatal injuries and a relatively small number of deaths, emphasizing the potential for preventive measures.
* **Contributing Factors**: With a significant portion of data missing for contributing factors (contrib\_factr\_p1\_id and contrib\_factr\_p2\_id), there's a notable gap in understanding the causes behind these incidents.
* **Mode of Transport**: Specific flags for pedestrians, motor vehicles, motorcycles, bicycles, and other forms of transport illustrate the diverse participants in traffic incidents, offering a foundation for targeted safety interventions.

**Observations:**

The dataset's structure and variable types suggest a comprehensive approach to capturing traffic incident details, yet the significant missing data, especially in location coordinates and contributing factors, limit some analytical capabilities.

The diversity in incident severity, from non-injury incidents to fatalities, underscores the complex nature of traffic safety and the need for multifaceted preventive strategies.

The clear distinction between different modes of transport involved in incidents provides a valuable perspective for designing mode-specific safety measures.

**Conclusion:**

The analyzed dataset offers vital insights into traffic incidents' dynamics, outcomes, and potential safety interventions. Despite challenges posed by missing data, the detailed records of incident severity, involved parties, and temporal patterns present a valuable resource for traffic safety analysis. Enhancing data collection and reporting practices, particularly for location and contributing factors, could significantly improve the dataset's utility for preventing future incidents and shaping traffic safety policies.

### **Chicago Traffic Incident Analysis Report**

**Dataset Overview:**

The dataset comprises records of 817,723 traffic incidents in Chicago, encapsulating a broad spectrum of information across 48 variables. It covers a time span from March 3, 2013, to March 26, 2024, providing a comprehensive look at traffic dynamics over nearly a decade. The data is characterized by its diverse variable types, including text, boolean, datetime, numeric, and categorical, indicating a rich dataset conducive to in-depth analysis.

**Key Statistics and Findings:**

* **Observations and Variables**: The dataset contains 817,723 observations with 48 variables, occupying a total memory size of 299.5 MiB. Each record on average uses 384.0 B of memory, pointing towards a dataset of considerable detail and depth.
* **Missing Data**: Approximately 21.1% of the dataset's cells are missing, which signifies a substantial amount of incomplete records across various variables. The extent of missingness varies significantly, with some variables like CRASH\_DATE\_EST\_I missing up to 92.5% of its values.
* **Traffic Control and Device Conditions:** A significant number of incidents occurred in areas with "No Controls" or under "Traffic Signal" regulation, indicating potential areas for improving traffic signal visibility or compliance.
* **Weather, Lighting, and Road Conditions:** The majority of incidents occurred under "Clear" weather conditions and "Daylight" lighting conditions, suggesting that factors other than adverse weather or poor lighting primarily contribute to these incidents. The road surface condition was mostly "Dry" at the time of incidents, and most roads had "No Defects".
* **Common Crash Types**: The most frequent first crash types include "Parked Motor Vehicle" and "Rear End", pointing towards a high prevalence of stationary or slow-moving traffic incidents.
* **Injuries and Fatalities**: The dataset captures various injury levels, with most incidents resulting in "No Indication of Injury". Fatal injuries, while relatively rare, still present a critical concern.
* **Location Data**: The dataset provides detailed latitude and longitude information for each incident, although approximately 0.7% of these records are missing. This geographic detail is crucial for identifying hotspots of traffic incidents within the city.
* **Time of Incidents**: Data on the hour, day of the week, and month of crashes could inform temporal patterns in traffic incidents, potentially guiding targeted safety measures or enforcement efforts.

**Recommendations and Implications:**

* **Enhanced Traffic Controls**: Addressing high-incident locations with improved traffic signaling and control measures, especially at intersections or areas prone to "Parked Motor Vehicle" and "Rear End" incidents.
* **Road Safety Campaigns**: Focusing on periods of high traffic and daylight hours where the majority of incidents occur, through public safety campaigns or increased enforcement of traffic laws.
* **Data Quality Initiatives**: Efforts to reduce the rate of missing data, particularly in key variables that influence safety analysis and decision-making, would significantly enhance the utility of this dataset for policy formulation and evaluation.

**Conclusion:**

The analysis of the Chicago traffic incident dataset reveals a city with diverse and complex traffic safety challenges. While the data points towards a multitude of factors influencing traffic incidents, it also offers valuable insights for targeted interventions. Addressing these challenges through data-driven policy, infrastructure investment, and community engagement can significantly enhance road safety in Chicago.