Title: The Impact of Road Conditions and Traffic Patterns on Accident Severity

Github : [Link]

Domain:

(Powerful image of a car accident or congested traffic)

Road accidents are a significant global concern, with devastating consequences. Traffic congestion in cities is also on the rise, impacting everything from travel time to economic costs. This presentation explores these issues by analyzing two datasets.

The first dataset provides detailed information on car crashes, including factors like collision severity, weather conditions, and road types. The second dataset offers traffic volume data for various junctions at different times.

Through data visualisation and analysis, we will investigate how road conditions and traffic patterns might be linked to the severity of road accidents.

Dataset Type:

Tabular

Attributes:

- **Year (from Car Crash data):** Year of the accident (numerical)
- **Month (from Car Crash data):** Month of the accident (numerical)
- **Day (from Car Crash data):** Day of the accident (numerical)
- **Weekday? (from Car Crash data):** Weekend or Weekday indicator (categorical)
- **Hour (from Car Crash data):** Hour of the day (numerical)
- **Collision Type (from Car Crash data):** Type of collision (categorical: e.g., 2-Car, 1-Car)
- **Injury Type (from Car Crash data):** Severity of injuries sustained (categorical: e.g., No injury/unknown, Non-incapacitating)
- **Primary Factor (from Car Crash data):** Main contributing factor to the accident (categorical: e.g., FAILURE TO YIELD RIGHT OF WAY)

- **Reported_Location (from Car Crash data):** Text description of the accident location
- **Latitude (from Car Crash data):** Latitude coordinate (numerical)
- **Longitude (from Car Crash data):** Longitude coordinate (numerical)
- **DateTime (from Traffic Congestion data):** Date and time of traffic data collection (datetime)
- **Junction (from Traffic Congestion data):** Name of the junction where traffic data was collected (categorical)
- **Vehicles (from Traffic Congestion data):** Number of vehicles observed at that time (numerical)
- **ID (from Traffic Congestion data, optional):** Unique identifier for each traffic data point (may not be used in the analysis)

Number of Records:

The process of combining the datasets resulted This number depends on the chosen merging method:

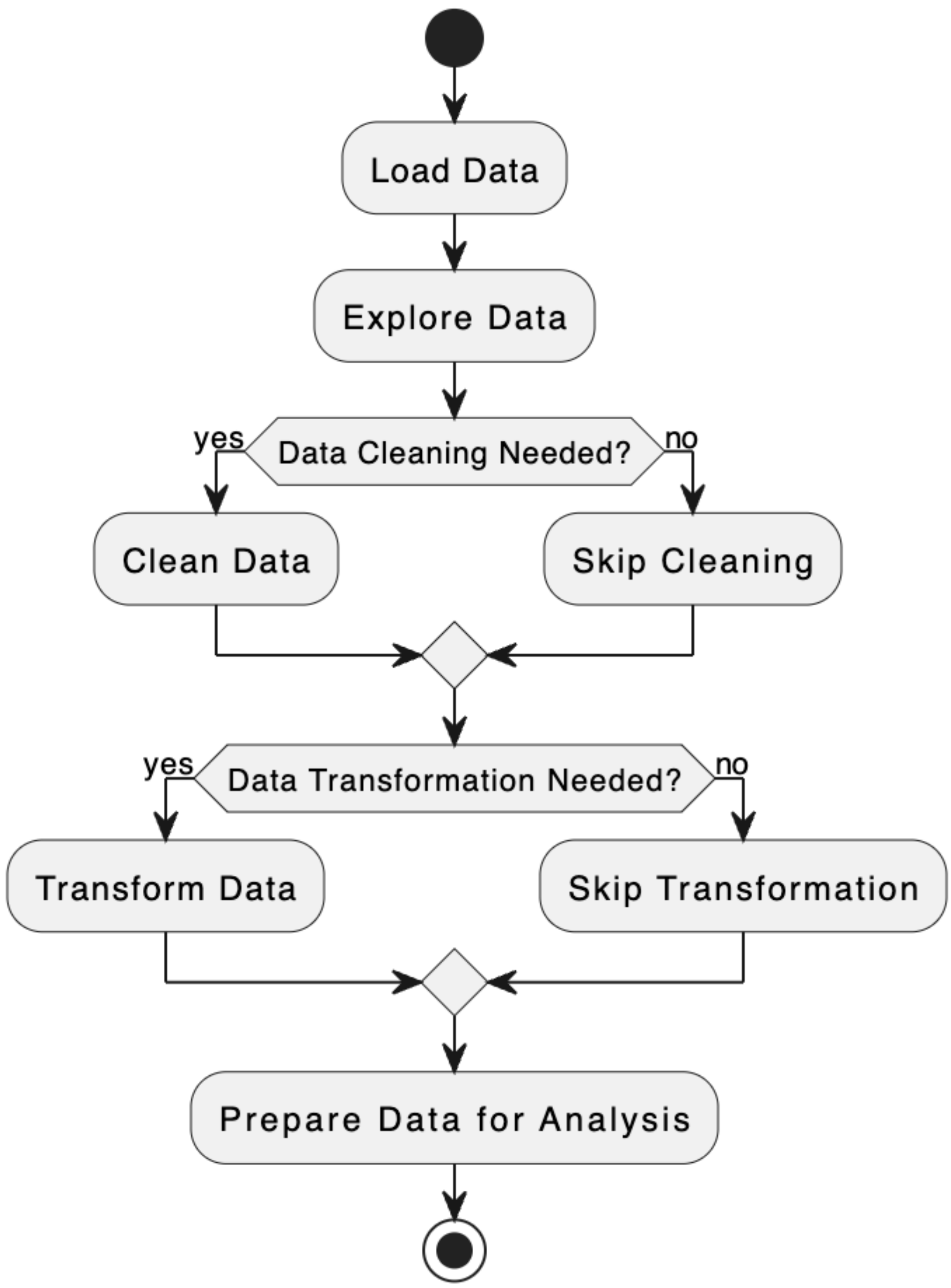
- **Matching by Location and Time:** If matching occurred based on location and time window, this might be a smaller dataset focusing on accidents potentially linked to specific traffic conditions.
- **Enriching Car Crash Data:** If traffic volume was added to the car crash data based on matching junctions and time windows, this approach might not have a perfect match for every accident, but it enriches the car crash data for further analysis.

Data Transformation :

The data underwent the following transformations to facilitate the merging process:

- **Extracting Features from DateTime:** Features like Year, Month, Day, and Hour were extracted from the DateTime attribute in the traffic congestion data to enable matching with the car crash data timestamps.
- **Matching Logic Definition:** A logic was defined to match entries based on location (reported location vs. junction names) and a chosen time window.
- **Handling Missing Traffic Data:** We addressed potential missing traffic data points for car crash locations/times (depending on the chosen method).
- **Adding Traffic Volume (if applicable):** If applicable, a new column was created in the car crash data to hold the matched traffic volume for each accident.

Traffic Congestion Analysis Workflow



Target: Create a combined dataset suitable for analyzing the relationship between car crashes and traffic congestion.

Actions:

1. **Data Loading (Action):**
   - Load the car crash data from its original format (e.g., CSV) into the analysis environment.
   - Load the traffic congestion data from its original format (e.g., CSV) into the analysis environment.
2. **Data Exploration (Action):**
   - Examine the structure of both datasets (attributes, data types).
   - Identify missing values or inconsistencies in each dataset.
   - Understand the distribution of key variables (e.g., time of day, location) in each dataset.
3. **Data Cleaning (Optional Action):**
   - Address missing values in either dataset (if necessary).
   - Handle formatting inconsistencies (e.g., date formats).
   - Address outliers in the data (if applicable).
4. **Data Transformation (Optional Action):**
   - Depending on the chosen merging approach, this might involve:
     - Extracting features from the "DateTime" attribute in the traffic congestion data (e.g., Year, Month, Day, Hour) for matching with car crash data timestamps.
     - Defining a matching logic based on location (reported location vs. junction names) and a chosen time window.
   - In case of enriching car crash data: creating a new column to hold the matched traffic volume for each accident (if applicable).
5. **Data Merging (Action):**
   - Choose a method to combine the datasets:
     - **Matching by Location and Time:** Match entries based on location (reported location vs. junction names) and a chosen time window. This might result in a smaller dataset focusing on accidents potentially linked to specific traffic conditions.
     - **Enriching Car Crash Data:** Keep the car crash data as the main dataset and add a new column indicating the traffic volume (number of vehicles) based on the closest matching junction and time window from the traffic congestion data. This approach might not provide a perfect match for every accident, but it can enrich the car crash data with traffic information for further analysis.
   - Merge the datasets based on the chosen method.

- Implementation using tools

Matplotlib (Python):

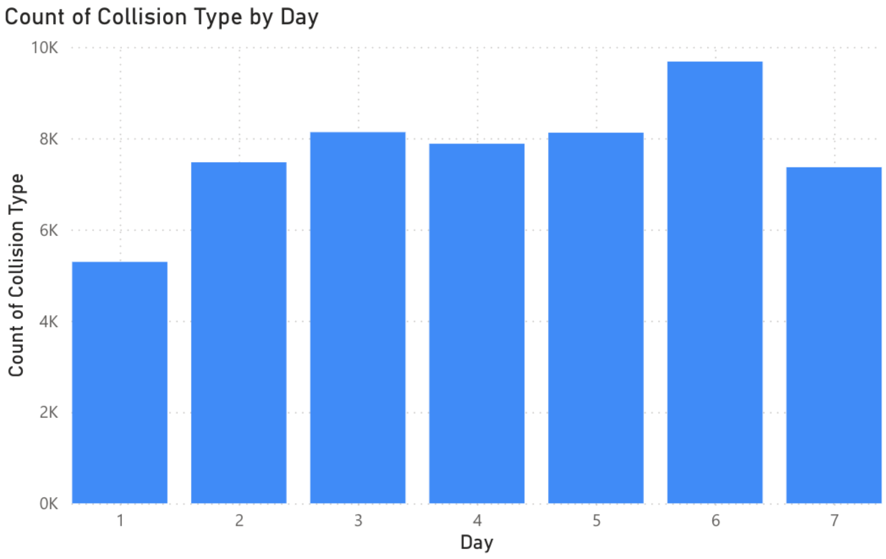
- **Suitable for:**
    - Basic exploratory visualizations like scatter plots, histograms, and bar charts.
    - Creating custom visualizations with more control over the visual elements.

Power BI Online:

- **Suitable for:**
    - Interactive dashboards and reports with various chart types.
    - Easy data exploration and filtering for identifying trends.

D3.js (JavaScript):

- **Suitable for:**
    - Highly interactive and customizable visualizations.
    - Creating complex visualizations not readily available in other tools.



We're looking at a bar chart, most likely. The horizontal axis (X-axis) represents days, and the vertical axis (Y-axis) represents the count of collision types. There might be multiple bars for each day, signifying different collision types (e.g., 2-Car, 1-Car, Hit-and-Run). The height of each bar indicates the number of collisions of that type that happened on that particular day.

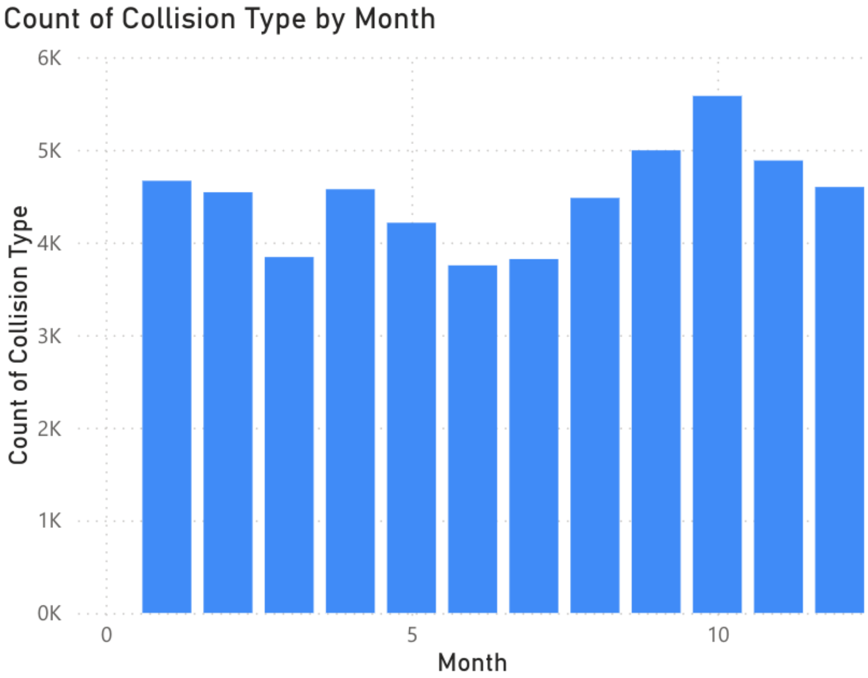
Key Elements:

- **Title:** The graph should have a clear title describing the information it presents, such as "Count of Collision Types by Day" or "Daily Breakdown of Car Crash Types."
- **Labels:** The X-axis should be labeled "Day," and the Y-axis should be labeled "Count" or "Number of Collisions."
- **Legend (Optional):** If there are multiple bars per day representing different collision types, a legend will be helpful. The legend will typically appear outside the plotting area and use colors, patterns, or symbols to represent each collision type.

Interpreting the Data:

By looking at the heights of the bars, we can see which days had the most and least collisions of each type. We might also be able to identify patterns in the data, such as:

- Are there specific days of the week with higher collision counts?
- Does a specific collision type (e.g., 2-Car) occur more frequently than others throughout the week?
- Are there any trends or outliers in the data that warrant further investigation?

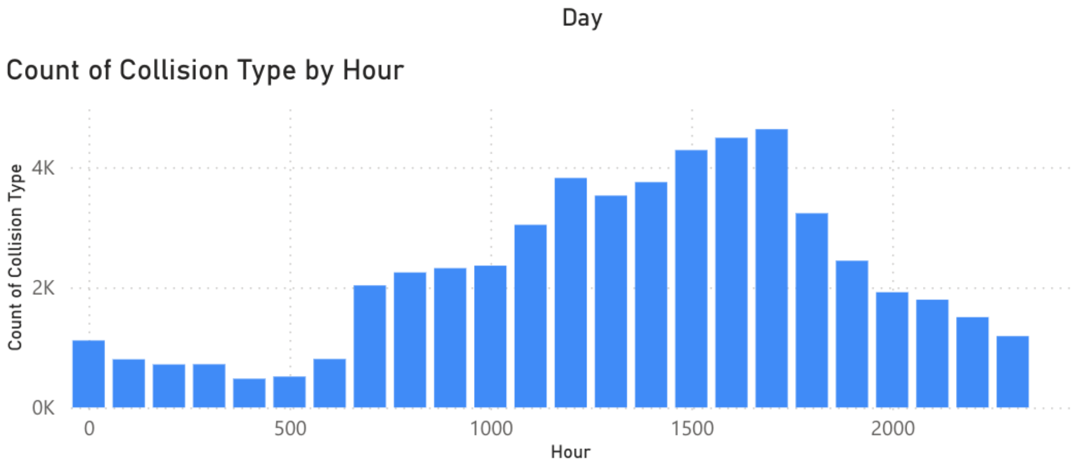


Data Visualization:

- The X-axis (horizontal) represents the month, likely abbreviated (e.g., Jan, Feb, Mar).
- The Y-axis (vertical) represents the count of collisions.
- There are multiple bars for each month, representing different collision types (likely indicated by the legend on the right). The legend uses colors (blue, orange, green) to differentiate between collision types.

Key Findings:

- We can see that some months have more collisions overall than others, based on the total height of all bars for that month.
- By looking at the height of each colored bar within a month, we can identify which collision types were most frequent in each month. For example, in the month labeled "Jul" (July), the blue colored bar seems to be the tallest, followed by the orange and green bars. This suggests that the collision type represented by blue was most common in July.

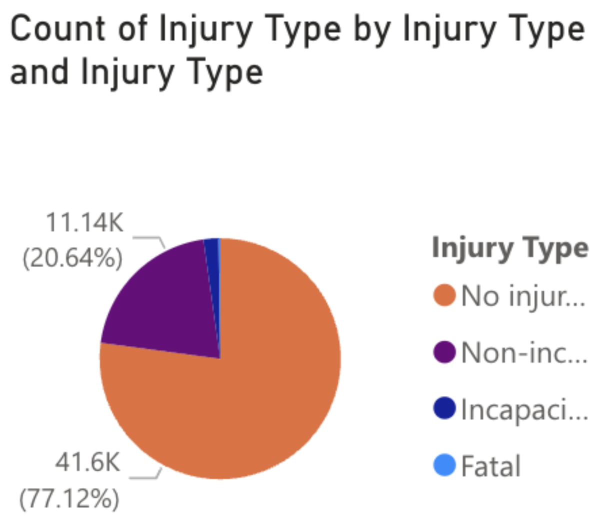


Data Visualization:

- The X-axis (horizontal) likely represents the hour of the day (0-23 or 12 AM - 11 PM).
- The Y-axis (vertical) represents the count of collisions.
- There are likely multiple bars for each hour, representing different collision types (indicated by the legend, colors, or patterns).

Key Findings:

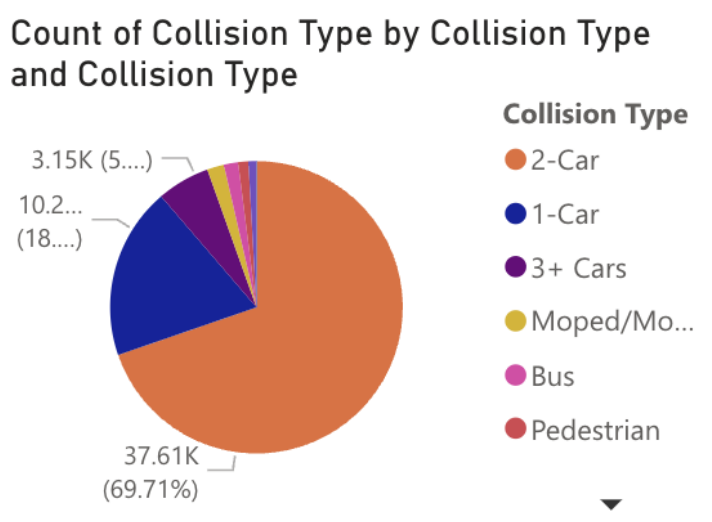
- We can see which hours of the day have the most and least collisions overall, based on the total height of all bars for that hour.
- By looking at the height of each colored bar within an hour, we can identify which collision types were most frequent at different times of the day. For example, the hour labeled "5" might have a tall green bar, indicating a high number of collisions of that type (e.g., rush hour fender benders) at 5 AM.



- The pie chart is likely divided into slices representing different injury types (e.g., No Injury, Minor Injury, Major Injury, Fatal).
- Each slice's size (angle) should correspond to the proportion of crashes resulting in that injury type out of the total number of crashes.

Possible Interpretation :

- The pie chart shows the distribution of injury severity levels in car crashes represented in the data set.
- The largest slice would indicate the most frequent injury type sustained in crashes. For example, a large "No Injury" slice might suggest a high number of minor fender-benders.

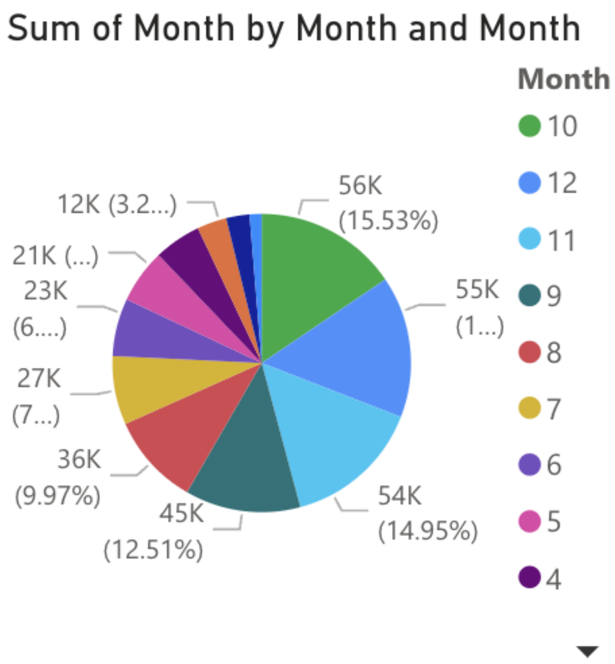


Key Elements:

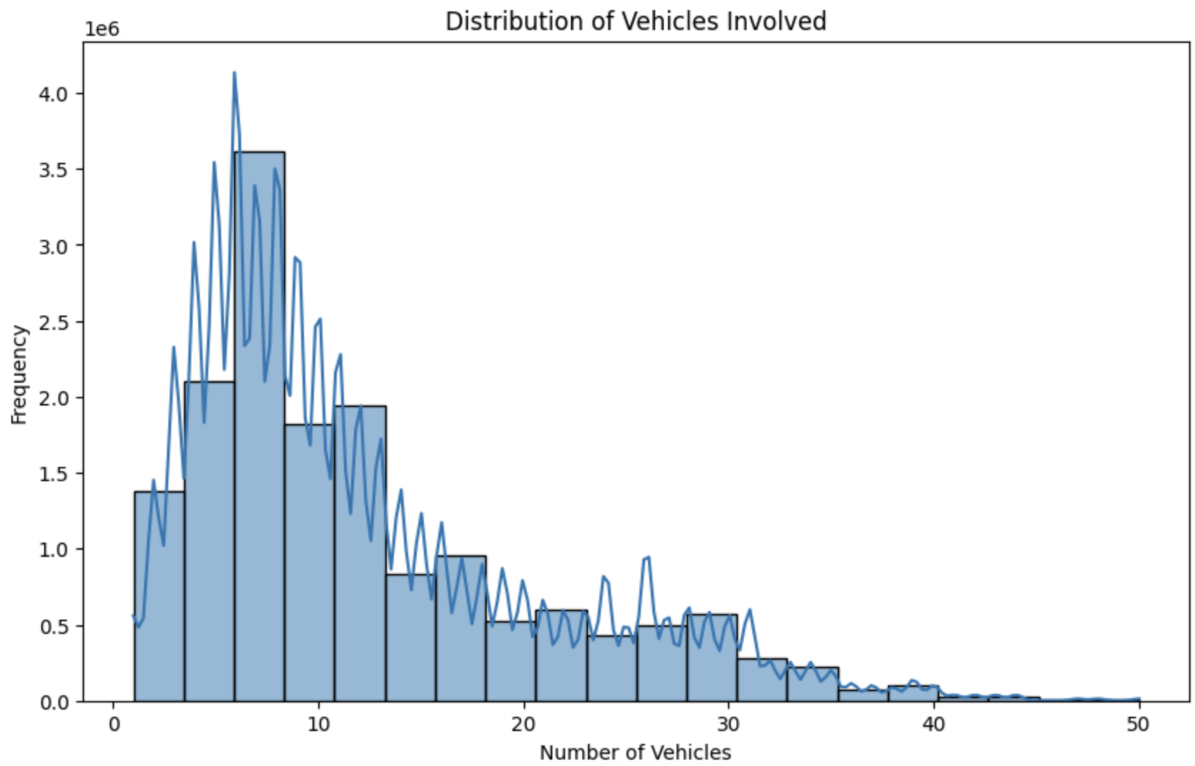
- The pie chart is likely divided into slices representing different collision types (e.g., 1-Car, 2-Car, 3+ Cars, Moped/Motorcycle, Pedestrian, Bus).
- Each slice's size (angle) should correspond to the count of crashes belonging to that specific collision type.

Possible Interpretation:

- The pie chart shows the distribution of crashes by collision type in the data set.
- The largest slice would indicate the most frequent type of collision. For example, a large "2-Car" slice might suggest that most crashes involved two vehicles.



- **Total Collisions by Month:** If the stacked bars represent the count of collisions categorized by a specific factor (e.g., collision type, severity, location), the total height of each stack would indicate the cumulative number of collisions for that month. In this case, we could potentially see if there are seasonal trends in collision occurrences.
- **Traffic Volume by Month:** If the stacked bars represent different vehicle classes (e.g., cars, trucks, buses) contributing to traffic volume, the total height of each stack would indicate the cumulative traffic volume for that month. This could reveal seasonal patterns in traffic depending on the location.



Data Visualization:

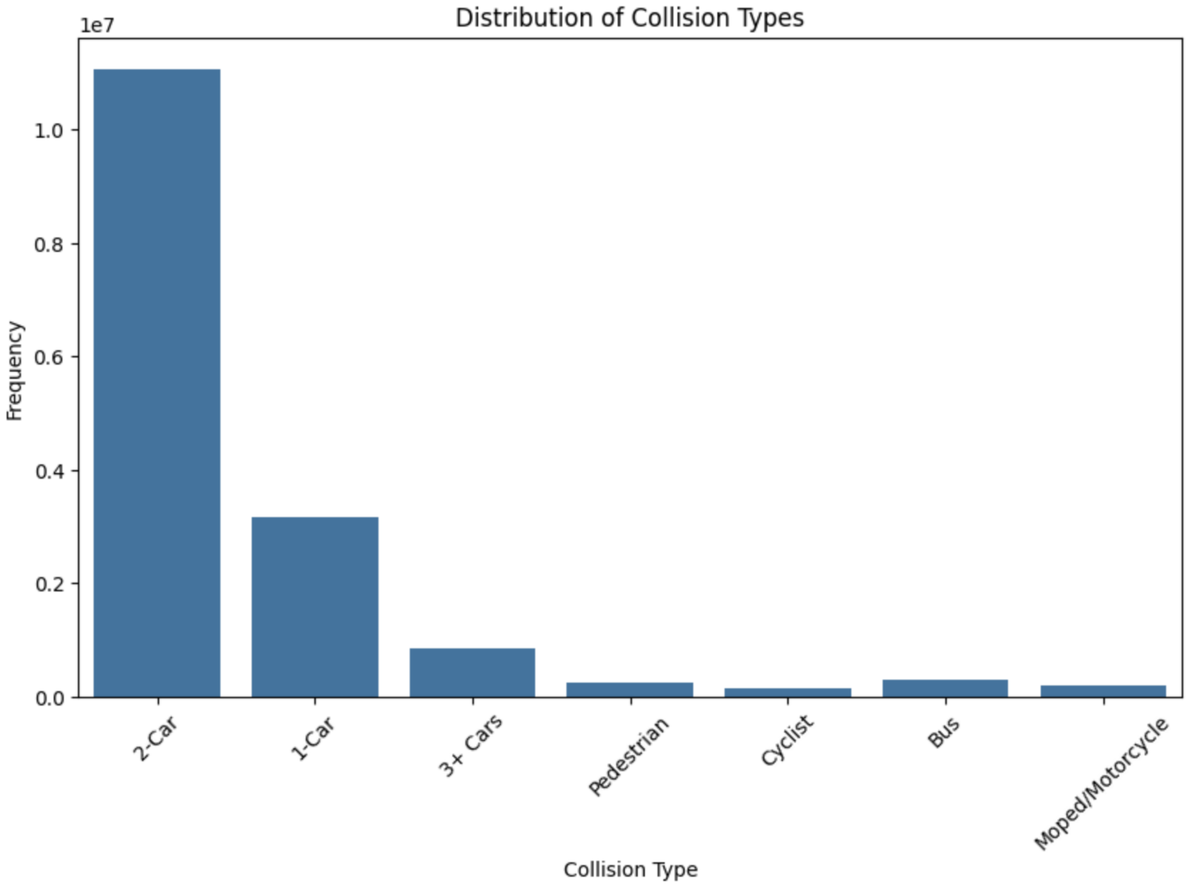
- The X-axis (horizontal) represents the number of vehicles involved in a car crash.
- The Y-axis (vertical) represents the frequency (count) of car crashes that had that number of vehicles involved.

Key Elements:

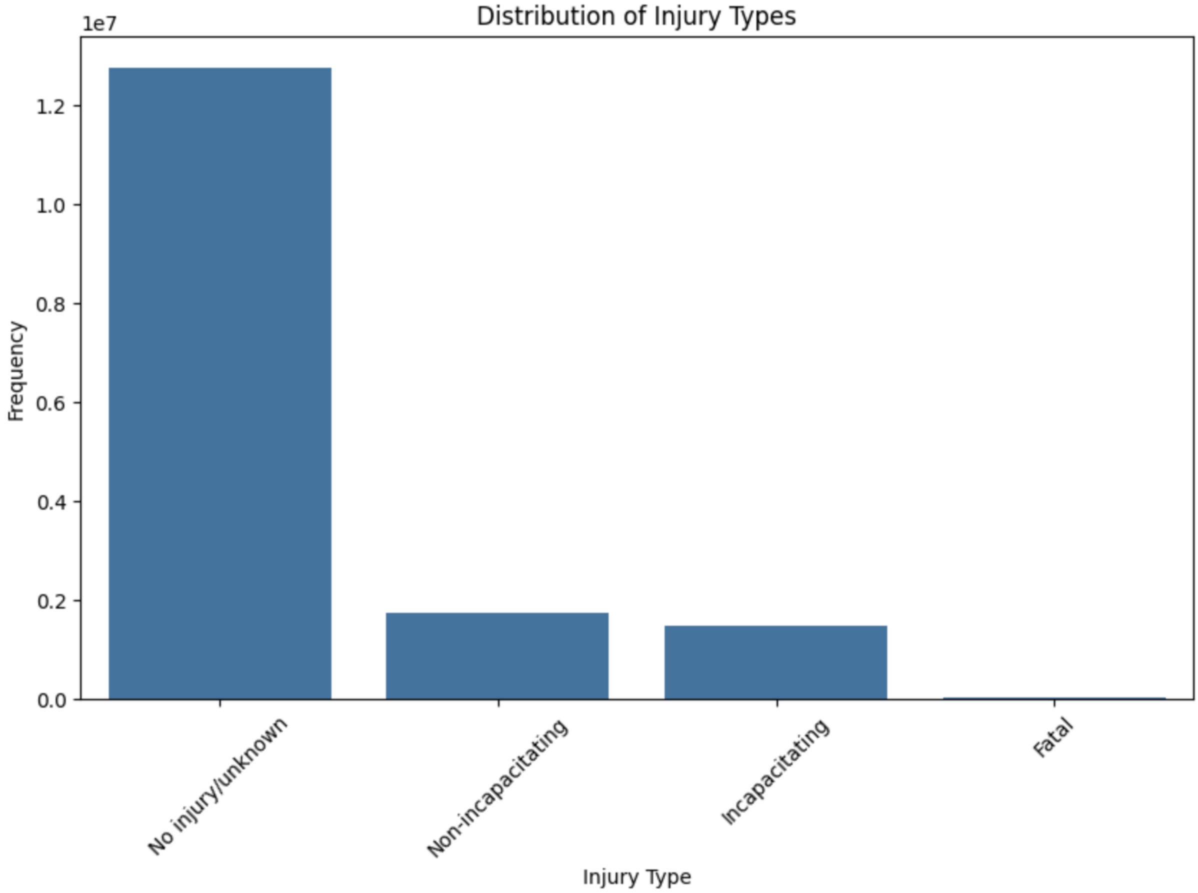
- The graph displays bars at various points on the X-axis. The height of each bar represents the number of crashes (frequency) where that specific number of vehicles were involved (e.g., if a bar is at X=2 and Y=10, it means 10 crashes involved 2 vehicles).

Interpreting the Distribution:

- By observing the heights of the bars, we can see how many crashes involved a certain number of vehicles. The most frequent number of vehicles involved will be represented by the tallest bar.
- The distribution of the bars indicates whether most crashes involve few vehicles (single car accidents), many vehicles (pileups), or fall somewhere in between.



- **Collision Counts by Month and Severity:** If the Y-axis represents the total number of collisions, and the stacked bars represent collision severity (e.g., minor, major, fatal), we could see how the total number of crashes and the breakdown by severity vary across months. This might reveal seasonal trends in crash occurrences and severity.
- **Collision Counts by Month and Vehicle Type:** If the Y-axis shows the total number of collisions, and the stacked bars represent the type of vehicle involved (e.g., car, truck, motorcycle), we could identify any patterns in the total number of crashes and the types of vehicles involved across months.



- The stacked bar chart shows the distribution of injury severity outcomes for different crash types. For each crash type (on the X-axis), the height of each colored bar segment represents the number of crashes that resulted in that specific injury severity level.
- By looking at the height of each color segment within a bar (crash type), we can identify the most common injury severities associated with each crash type. For example, a crash type with a tall red bar segment (potentially representing fatal injuries) might be a cause for concern.