**1. Introduction**

**Traffic accidents pose a significant threat to road safety, and analyzing accident data is crucial to identifying patterns that can help prevent future incidents. This project aims to analyze traffic accident data to uncover insights related to road conditions, weather, and the time of day. It also visualizes accident hotspots and contributing factors that lead to accidents.**

**2. Objective**

The primary objectives of this analysis are:

1. Identify patterns in traffic accidents based on:
   * Road conditions
   * Weather conditions
   * Time of day
2. Visualize accident-prone areas (hotspots) on a geographical map.
3. Understand contributing factors such as weather, road types, and other related features.

**3. Dataset Overview**

The dataset used for this analysis contains 7,728,394 records and 46 features that describe various aspects of traffic accidents in the United States. Some important features include:

* **ID**: Unique identifier for each accident.
* **Severity**: Indicator of accident severity on a scale of 1-4.
* **Start\_Time & End\_Time**: Timestamps for when the accident occurred and ended.
* **Weather\_Condition**: Weather conditions during the accident.
* **Temperature (F)**, **Wind\_Speed (mph)**, **Visibility (mi)**: Weather-related metrics.
* **Sunrise\_Sunset**: Whether the accident occurred during daylight or night.
* **Location Information**: Latitude and longitude coordinates for mapping accident locations.

**4. Data Preprocessing**

**4.1 Handling Missing Data**

* A few columns, such as weather conditions, had missing values. These missing values were handled using imputation techniques where appropriate, or rows were dropped if they contained critical missing data.

**4.2 Feature Selection**

* From the 46 available columns, key features related to time, weather, and geographic information were selected for deeper analysis. Columns such as **Start\_Time**, **Weather\_Condition**, **Temperature**, and **Location** were used for visualizing trends.

**5. Exploratory Data Analysis (EDA)**

**5.1 Accident Severity Distribution**

* The data shows the distribution of accidents across severity levels 1 to 4. Most accidents are of severity level 2, indicating they are moderate in nature.

**5.2 Time of Day Analysis**

* A time-based analysis reveals that most accidents happen during peak traffic hours (7-9 AM and 3-6 PM). This indicates higher accident risk during commuting hours.

**5.3 Weather Conditions and Accidents**

* **Clear weather**: Accounts for a majority of accidents, but this might be because clear weather is the most common condition.
* **Rainy or Foggy conditions**: Higher severity accidents occur in adverse weather conditions.
* **Low visibility**: Directly correlates with more severe accidents.

**5.4 Road Condition Impact**

* **Wet roads**: Show a higher frequency of accidents.
* **Paved roads**: See more frequent but less severe accidents, whereas unpaved roads have a lower accident count but higher severity.

**6. Hotspot Analysis**

**6.1 Geographic Mapping of Accident Locations**

* Using the latitude and longitude data, accident hotspots were identified. The most accident-prone areas are typically urban areas and highways with high traffic volumes.
* Geographic mapping tools like Plotly and GeoPandas were used to visualize these hotspots.

**6.2 Accident-Prone States**

* States with the highest number of accidents include **California**, **Texas**, and **Florida**.
* Heatmaps were generated to show clusters of accidents, especially around major cities like **Los Angeles**, **Houston**, and **Miami**.

**7. Contributing Factors**

**7.1 Weather Conditions**

* Adverse weather conditions (fog, rain, snow) contribute significantly to accidents. Data shows that accidents during such conditions tend to be more severe.

**7.2 Time of Day and Traffic Patterns**

* Accidents are highly concentrated during rush hours (morning and evening). Night-time accidents, though less frequent, tend to be more severe due to low visibility and potentially risky driving behaviors.

**7.3 Road Infrastructure**

* Certain road conditions like intersections, roundabouts, and junctions are common sites for accidents. Accidents at intersections tend to be more severe due to conflicting traffic flows.

**8. Visualization**

**8.1 Bar and Line Graphs**

* **Accident Trends**: Line graphs were used to show accident frequency over time, highlighting peaks during rush hours.
* **Severity vs. Weather**: Bar charts helped visualize the severity of accidents under different weather conditions.

**8.2 Geographic Visualizations**

* **Heatmaps**: Used to highlight accident hotspots based on latitude and longitude data.
* **Scatter Maps**: Depict accident-prone locations based on weather conditions, road types, and severity.

**9. Conclusion**

Through this analysis, it was possible to identify several patterns related to traffic accidents:

* Most accidents occur during peak traffic hours, and weather conditions like rain or fog lead to more severe accidents.
* Geographic mapping allowed the identification of accident hotspots, primarily in densely populated urban areas.
* Specific road conditions and time of day have a significant impact on accident frequency and severity.

**Key Recommendations:**

1. **Increase Safety Measures in Accident Hotspots**: Enhanced monitoring and road safety measures in identified accident hotspots can help reduce accidents.
2. **Improving Road Signage and Lighting**: Especially during night-time or in areas with frequent adverse weather conditions.
3. **Public Awareness Campaigns**: Focused on reducing accidents during peak hours by encouraging safer driving habits.

**10. Future Work**

Further analysis can include:

* Modeling accident prediction based on contributing factors.
* Investigating the impact of other potential causes such as driver behavior, vehicle type, or traffic flow data.

**Bobby Narang**  
*MCA Graduate, Data Science Intern*  
Date: 20.09.2024