US Accidents Data Analysis (2016-2023)

# Project Overview

This project analyzes US accidents from 2016 to 2023 using a dataset containing over 7 million rows, sourced from the internet. The dataset helps explore various factors contributing to accidents, such as weather conditions, traffic, and geography, to better understand the frequency and causes of accidents in different cities and states across the country.  
  
The key variables in the dataset include:  
- City and State  
- Weather Conditions, Temperature  
- Start Time, End Time, Severity  
- Wind Speed, Traffic, Visibility  
- Latitude, Longitude  
- Road Features such as bumps and curves

# Data Cleaning

The first step was to load the dataset and identify columns with a high percentage of missing data. Columns with excessive null values were removed to streamline the analysis. The cleaned dataset allowed for smoother data exploration and ensured the analysis focused on relevant and high-quality data.  
  
One notable observation was that data from New York City was missing. As one of the most populated cities, its absence prompted questions about the dataset's completeness and the possible reasons behind this missing data.

# Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted to examine accident trends across various factors:  
- Cities with Most Accidents: A key discovery was that a small percentage of cities account for the majority of accidents, while over 1000 cities reported only one accident.  
- Accidents by Day of the Week: The analysis showed that weekends typically had fewer accidents compared to weekdays.  
- Weather Conditions: Surprisingly, most accidents in major cities occurred during fair weather.  
- Accidents Over Time: We analyzed accidents by year and month to uncover trends. There was a clear decrease in accidents in some years, while others showed an increase, reflecting evolving traffic conditions or policy changes.

# Key Metrics and Questions Answered

Several questions were posed to gain deeper insights into the dataset:  
1. Are there more accidents in warmer or colder areas?  
2. Which five states have the highest number of accidents?  
 - We also examined accident rates per capita.  
3. Does New York appear in the data? It was found that New York City is missing, which is unexpected given its population size.  
4. Accidents by City: Among the top 100 cities with the most accidents, which states are represented?  
5. Accidents by Time of Day: The most common times for accidents are between 6 AM and 10 AM, followed by 3 PM to 6 PM.  
6. Accidents by Day of the Week: Weekdays see more accidents than weekends.  
7. Accidents by Month: Monthly analysis highlighted peak accident periods.  
8. Yearly Trends: The analysis revealed both increasing and decreasing accident trends year over year.  
9. Accidents per Traffic Unit: We explored when accidents per traffic volume were the highest.  
10. Accident Distribution by Hours: We compared hourly accident rates on weekdays versus weekends.  
11. Impact of Weather Conditions: We analyzed how different weather conditions influence accident frequency.  
12. Temperature Effects: Temperature's correlation with accident severity was also explored.

# Insights and Findings

- No Data for New York: A major insight was the absence of New York City data.  
- Accident Distribution: The number of accidents per city decreases exponentially. Less than 8% of cities experience more than 1000 yearly accidents.  
- Rare Accidents: Over 1000 cities reported only one accident, raising the need for further investigation.

# Data Visualization

Using Tableau, we visualized the insights from the dataset. Key visualizations include:  
- Heat Maps: Based on latitude and longitude data, we created heat maps to visualize accident density across the US.  
- Bar Charts: These display the number of accidents by state, city, and weather conditions.  
- Line Graphs: We used line graphs to depict accident trends over time, highlighting significant increases or decreases year by year.