

CSE564: Visualization

PROJECT PRELIM REPORT

VISUAL ANALYTICS ON TRAFFIC ACCIDENT TRENDS IN USA

Submitted By:

Group 27

Goutham Kalla (116385130)

Shreyas Reddy Gaddampally (116439954)

Introduction

Traffic accidents remain one of the most significant public safety concerns across the United States, claiming thousands of lives annually and resulting in substantial economic costs. Our project "Visual Analytics on Traffic Accident Trends" aims to transform raw accident data into actionable insights through comprehensive visual analysis. Initially focused on New York state patterns, we have expanded our scope to encompass all states, providing a national perspective on traffic safety challenges.

The nationwide analysis allows us to identify regional variations in accident patterns, seasonal trends, environmental impact factors, and infrastructure-related risks. By visualizing these complex relationships, we create a powerful decision-support tool for transportation planners, safety researchers, and policy makers to develop targeted interventions that can potentially save lives.

Our interactive dashboard presents geospatial distributions, temporal trends, severity classifications, and environmental correlations in an intuitive interface. This enables users to explore both macro-level national patterns and micro-level state-specific details. The visual analytics approach transforms millions of accident records into comprehensible patterns that reveal the underlying story of America's roadway safety challenges and opportunities for improvement.

Through this project, we aim to contribute to the ongoing efforts to enhance traffic safety by providing data-driven insights that can inform policy decisions, infrastructure improvements, and public awareness campaigns across the United States.

Data Source

The US Accidents dataset used in this project was compiled by Sobhan Moosavi and is available on Kaggle. This comprehensive collection contains approximately 7.7 million traffic accident records covering 49 states of the United States from February 2016 to March 2023. For our analysis, we initially focused on a representative sample of 1,000 entries from New York state but have expanded our scope to include nationwide data for more comprehensive insights.

The dataset was collected using multiple traffic APIs that stream traffic incident data from various sources, including departments of transportation, law enforcement agencies, traffic cameras, and sensors within the road networks. This data has been extensively validated, cleaned, and normalized to ensure consistency and reliability for analytical purposes.

The dataset features detailed information about each accident, including precise geolocation (latitude/longitude coordinates), temporal data (start/end times), environmental conditions present during the accident, and road infrastructure characteristics. This rich collection of attributes allows for multi-dimensional analysis of traffic safety patterns across the United States.

URL: <https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents>

This dataset is recognized as one of the most comprehensive publicly available repositories of traffic accident information in the United States and has been extensively used by transportation departments, insurance companies, urban planners, researchers, and government agencies for safety analysis and policy development.

Current Progress

Our team has made significant strides in developing the traffic accident analysis dashboard. Here's an overview of our current accomplishments:

Data Ingestion & Cleaning

- Successfully loaded the US Accidents dataset and filtered initially for New York state analysis
- Parsed date-time fields to extract temporal components (Year, Month, Day, Hour)
- Handled missing values using median imputation for environmental variables
- Normalized numerical fields to ensure consistent scaling across visualizations
- Sampled data strategically to optimize rendering performance while maintaining statistical validity

Exploratory Analysis

- Completed preliminary statistical analysis of accident severity distribution across states
- Identified key temporal patterns including peak accident hours and seasonal variations
- Analyzed correlations between weather conditions and accident frequency/severity
- Mapped geographic hotspots at both national and state levels
- Calculated summary statistics for top accident-prone counties in New York

Dashboard Implementation

- Designed responsive single-screen interface layout with no scrolling requirement
- Implemented the US accident map visualization showing nationwide distribution patterns
- Created state-specific accident location map (focused on New York)
- Developed trend analysis charts showing yearly accident counts (2016-2023)
- Implemented severity distribution visualization with filtering capabilities
- Added weather condition analysis component to assess environmental factors

Technical Integration

- Ensured smooth data flow between visualization components

- Implemented cross-filtering capabilities for interactive exploration
- Optimized rendering performance for handling large datasets
- Created consistent visual styling across all dashboard elements
- Added tooltips and interactive elements for enhanced user experience

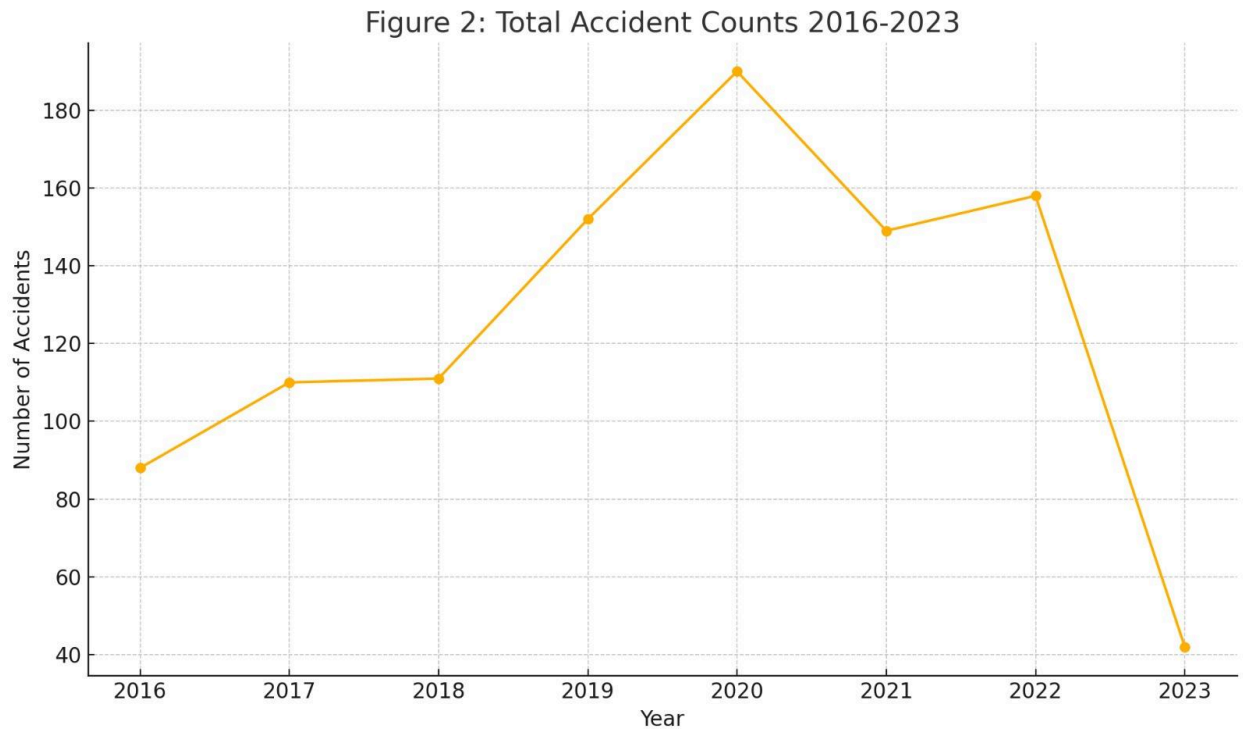
We are currently finalizing additional interactive elements and ensuring all visualizations accurately represent the expanded nationwide scope of our analysis.

Visualizations

To uncover trends and hidden patterns within New York's traffic accident data, we have implemented a diverse set of interactive and static visualizations using Plotly and Dash. These visual tools are essential for transforming raw data into digestible, actionable insights.

1. Line Chart

- Depicts the trend of total annual accident counts in New York from 2016 to 2023.
- Indicates overall fluctuations, with a noticeable drop post-2020, possibly due to pandemic-related changes in travel behavior.

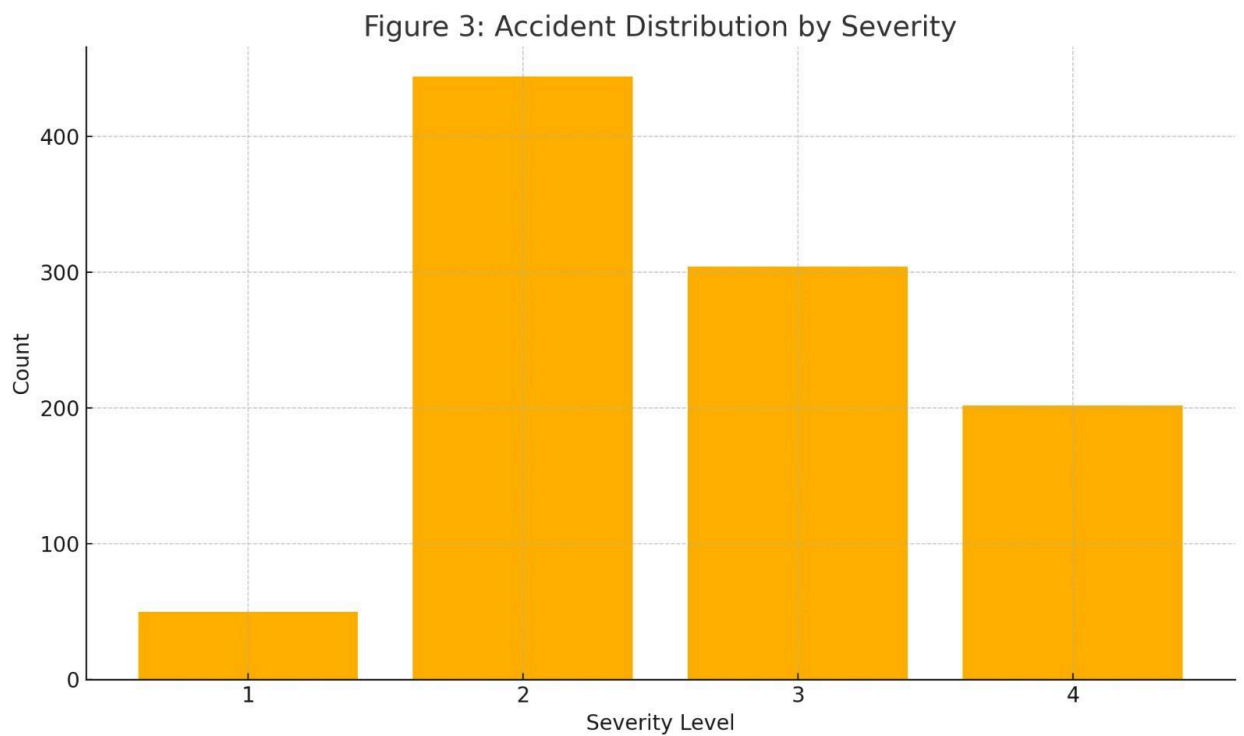


2. Choropleth Map

- Displays accident frequency by county in New York for the year 2023.
- Highlights county-level accident frequency in New York using a color gradient. Counties like Kings, Queens, and Albany show higher accident density.

3. Bar Chart

- Illustrates the distribution of traffic accidents by severity (Levels 1 to 4).
- Helps identify the dominant accident types and understand their implications.



4. US Map & State Selection

- Interactive point map showing accident locations across all 49 states for a selected year.

- Supports dropdown selection for states, dynamically updating visuals and insights.

Figure 1: Accident Locations in NY (2023)

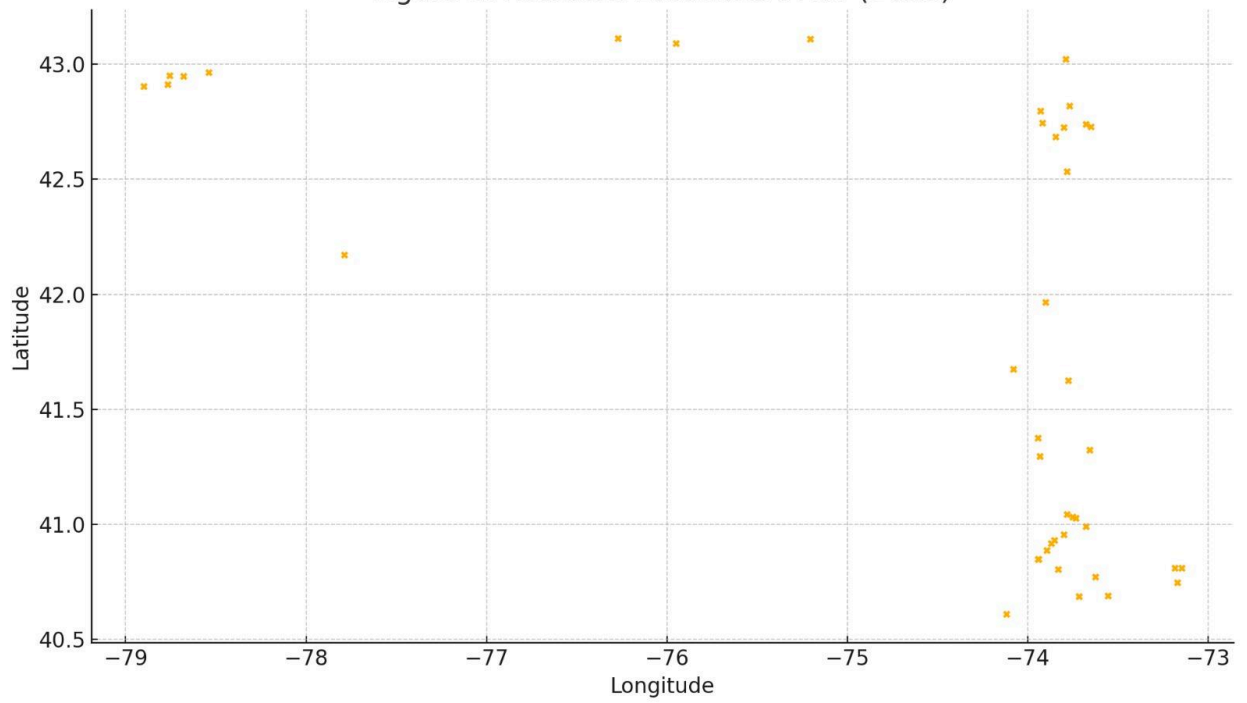
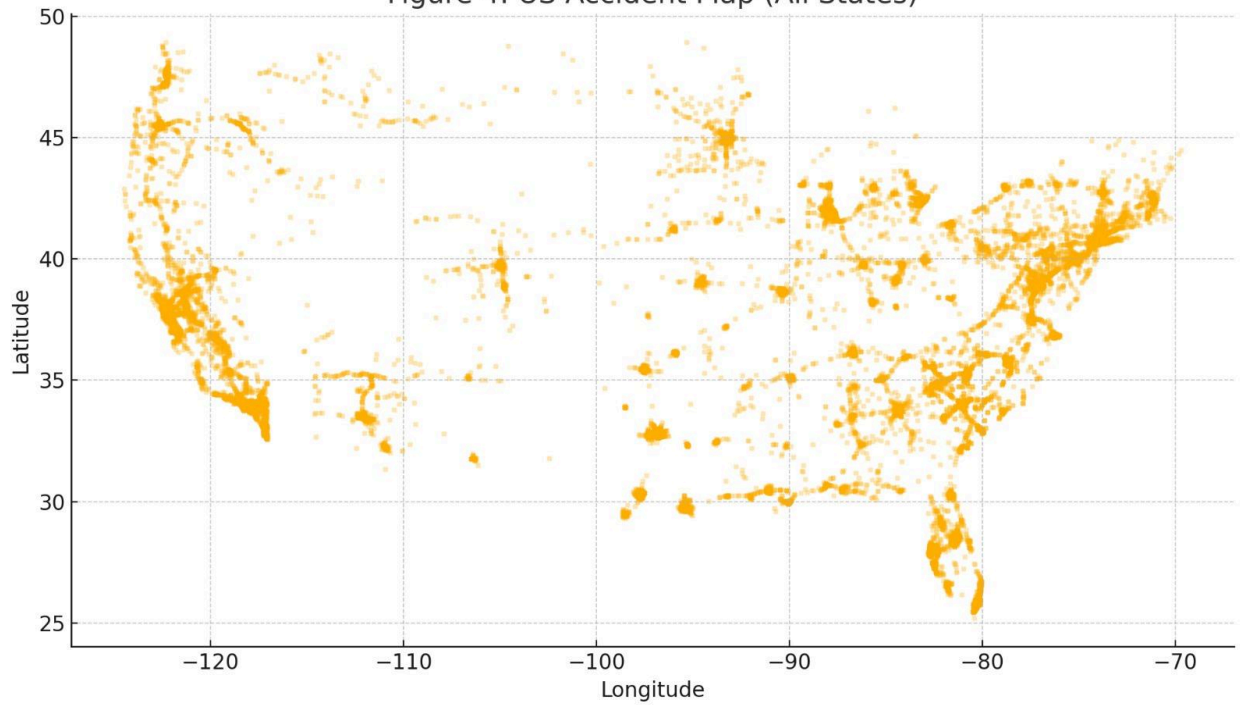
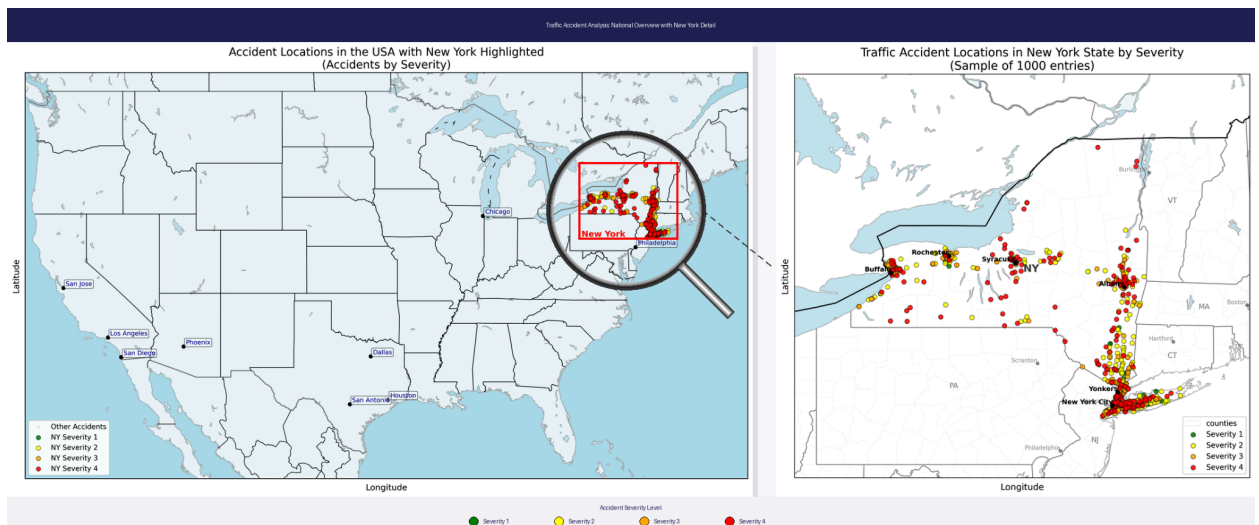


Figure 4: US Accident Map (All States)



5. State-wise Analysis (Dynamic)

- Once a state is selected:
 - **Line Chart** shows annual accident trends for that state (2016–2023).
 - **Bar Chart** reveals severity distribution specific to the selected state.
- Facilitates micro-level exploration while preserving national context.



Dashboard Overview

We have developed an interactive web dashboard using **Plotly Dash** to consolidate visual insights into a single interface. The dashboard enables users to explore traffic accident trends across the United States with a focus on **New York State**. It provides geospatial, temporal, and categorical breakdowns through intuitive visual components.

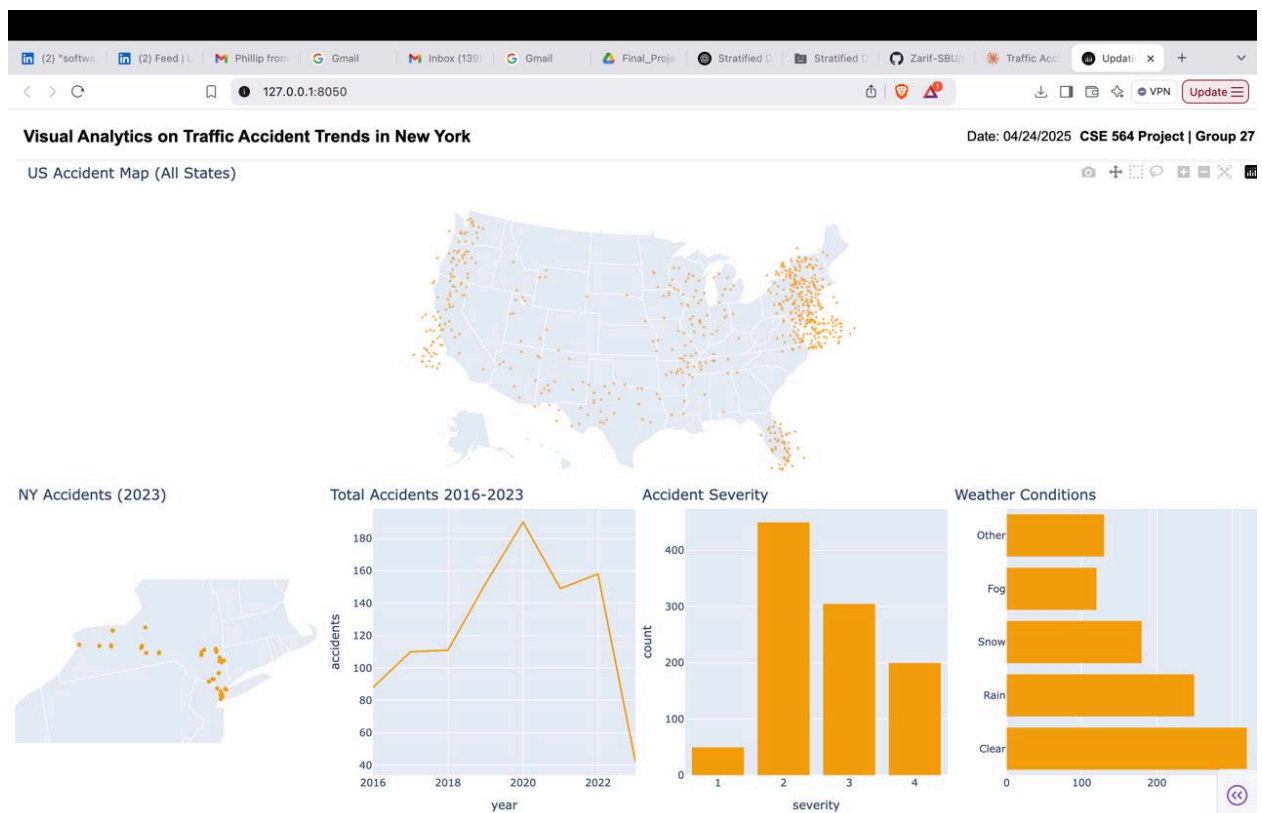
Key Features:

- **US Accident Map (All States):** A scatter plot displaying accident locations nationwide for macro-level analysis.
- **NY Accidents (2023):** A filtered geographic view showing all accidents reported in New York State for the year 2023.
- **Total Accidents (2016–2023):** A line chart illustrating yearly accident counts, highlighting significant fluctuations such as the pandemic dip in 2020.

- **Accident Severity Distribution:** A bar chart representing the count of accidents across severity levels (1–4).
- **Weather Conditions:** A horizontal bar chart showing the number of accidents under different weather types, with 'Clear' and 'Rain' being the most common.

The dashboard is fully responsive, with future enhancements planned to include:

- Dynamic dropdown for state selection.
- Cross-filtering for severity, time-of-day, and weather condition.
- Heatmaps and scatter plots for deeper correlation analysis.



FUTURE WORK:

- Complete remaining plots: Heatmap of accidents by hour/day, Scatter plots for Temperature vs. Frequency, Histogram of Visibility, Parallel Coordinates, Radar Chart comparing counties, Box Plots for Severity vs. Environmental factors.
- Integrate all elements into a single interactive Dash dashboard with linked brushing and filter controls (weather, time, severity).

- Refine layout for no scrolling; ensure all plots visible on one screen.
- Add interactive tooltips and custom filters for deeper exploration.