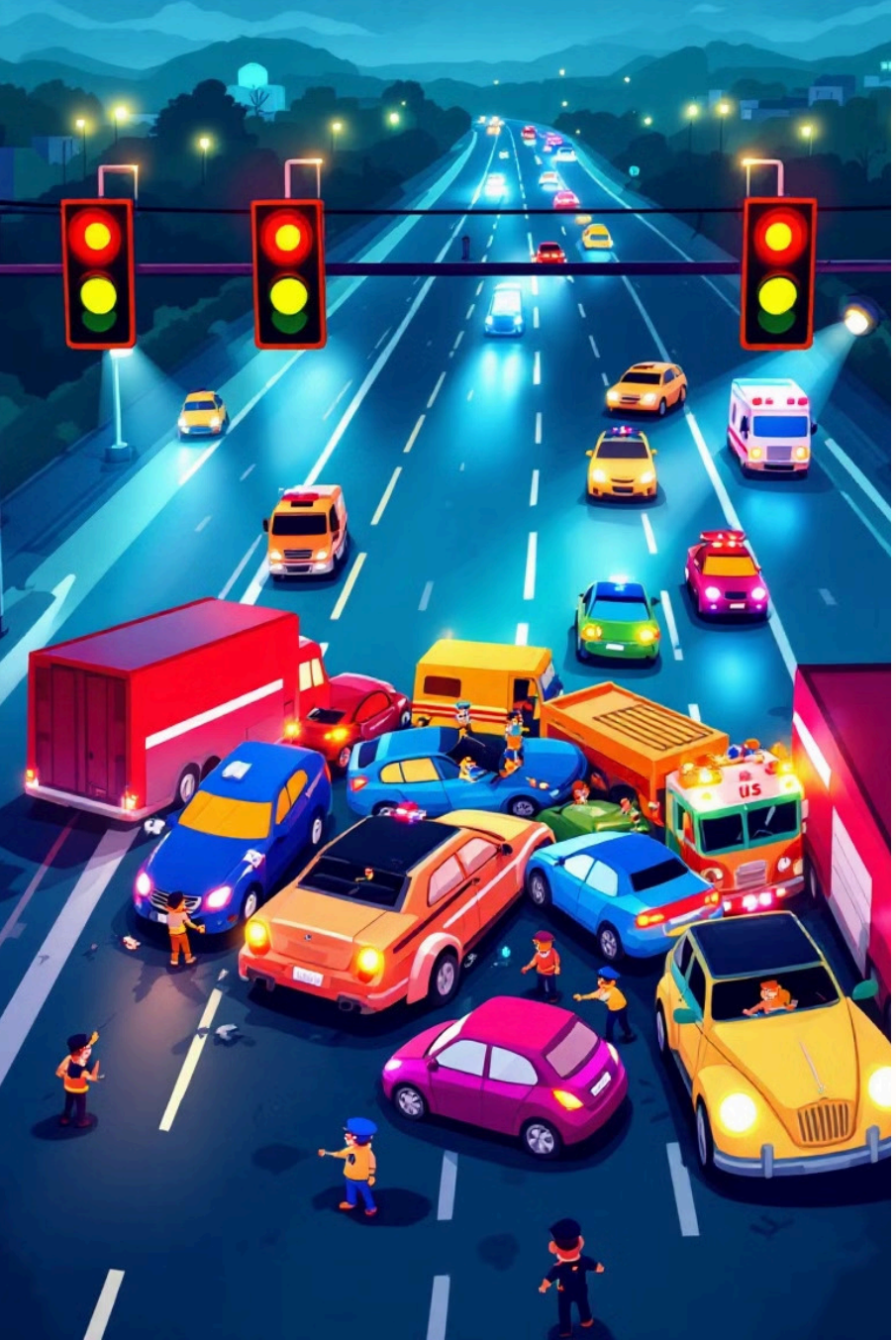


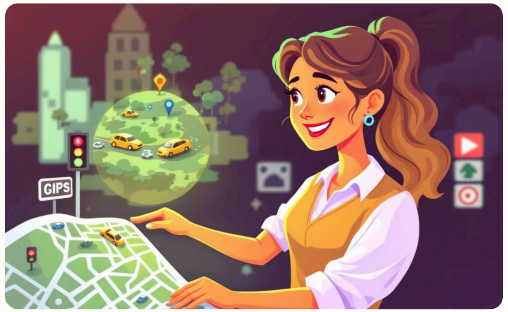
Traffic Accident Patterns in the United States

"Fasten your seatbelts, because we’re diving into the world of traffic accidents across the U.S.! From uncovering the busiest crash hotspots to decoding how weather and road design influence our safety, this presentation is your roadmap to understanding accident patterns. We'll explore how clever road features and smart planning can steer us toward fewer accidents. Join us as we navigate the twists and turns of the data to drive home safer solutions for everyone."

Meet the Project Team

Olga, Data Analyst

Collects, cleans and plots the traffic accident data



Nathaly, GIS Specialist

Visualizes the data geographically

Harrison, Traffic Engineering Analyst

Provides transportation network analysis

Diya, Project Manager

Oversees the project and coordinates the team

Who Benefits from This Analysis?

Our target audience includes transportation planners, insurance companies, public safety officials, and the general public.

1 **City Planners**

Identify high-risk areas needing infrastructure improvements.

2 **Transportation Agencies**

Optimize traffic flow and enhance safety measures.

3 **Policymakers**

Inform legislation aimed at accident reduction.

Understanding Our Data

Data Sources

US Accidents 2016-2023 (Kaggle)

Data Types

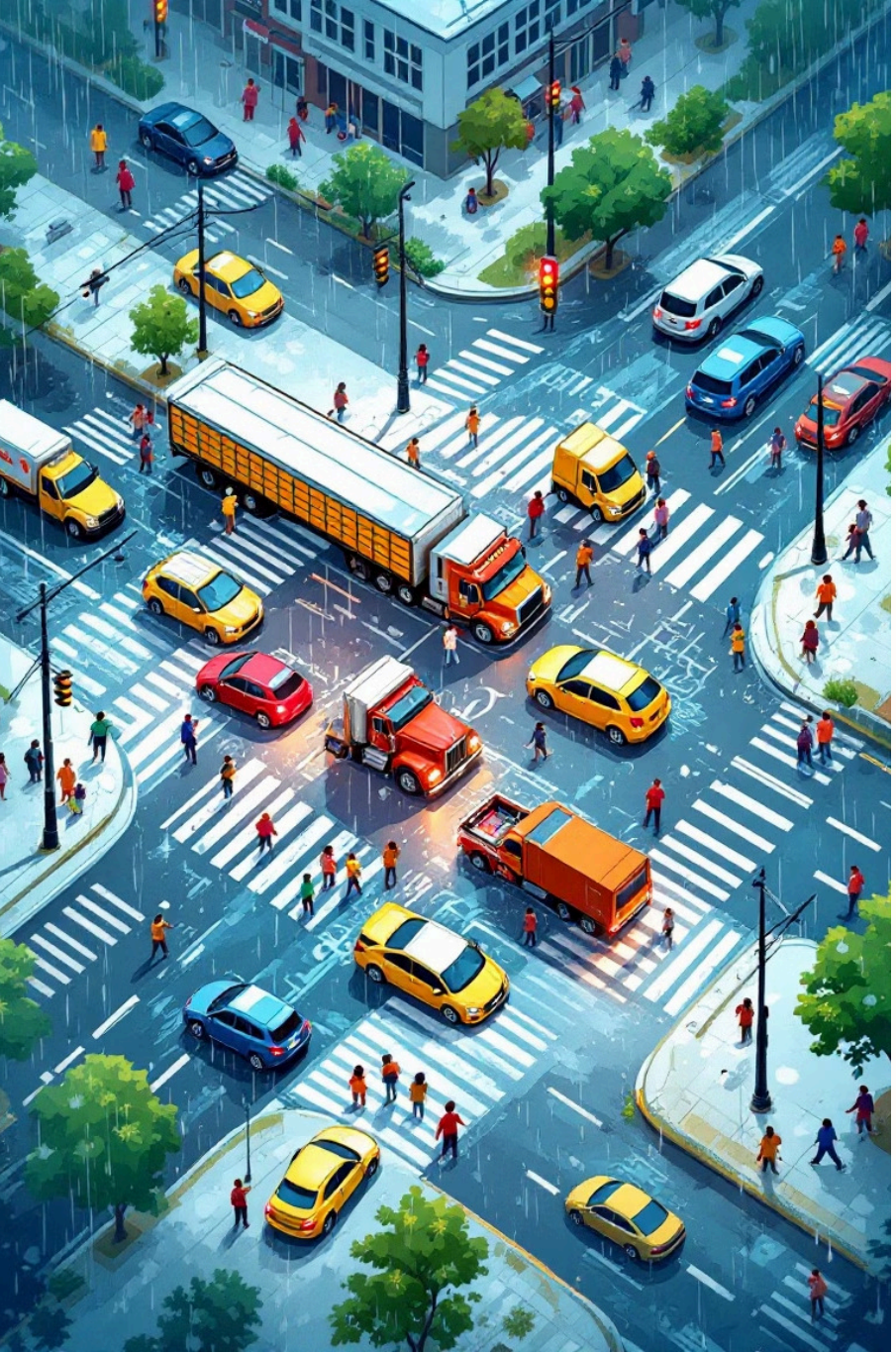
Location, time, weather, number of drivers, traffic mitigation road features

Data Cleaning

Standardize formats, handle missing values, random sampling

Aggregation

Group by time, location, and other factors



Examining Accident Factors

Locations

States

Seasonality

does it Increased traffic and impaired driving?

Weather

Rain, snow, fog impact visibility and control

Discovering the Traffic Accident Data (ETL)

Data Set: US Traffic Accidents. Size: ~2.8M rows of data, spanning 49 states. Subset used for this analysis: 1% sample (~68,000 records). Source: **Kaggle - US Accidents Dataset**

1

Extract:

CSV File
- Pandas `.read_csv()`
- Leaflet.js for mapping

2

Transform:

Data Cleaning Steps

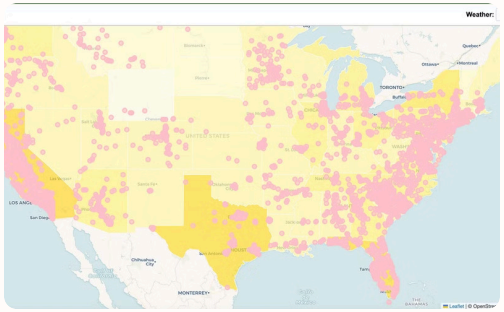
- Squeezed data to 1% from 3G to 30Mb
- Converted NaN values to 0 instead of deleting, to avoid deleting several cities.
- Converted `Start_Time` and End_Time to `datetime`
- Renamed and reformatted columns for consistency
- Mapped weather conditions to categories (rain, snow, clear, etc.)
- Retained essential columns (46 columns)

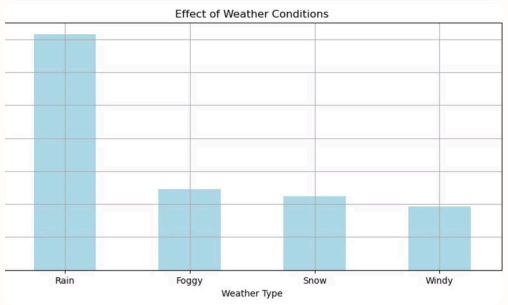
3

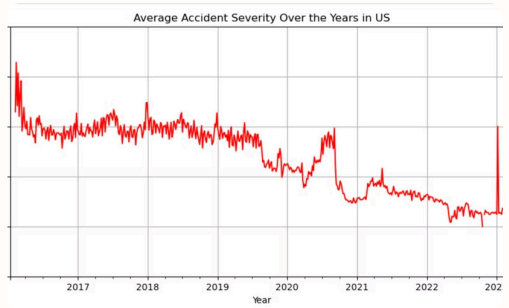
Load:

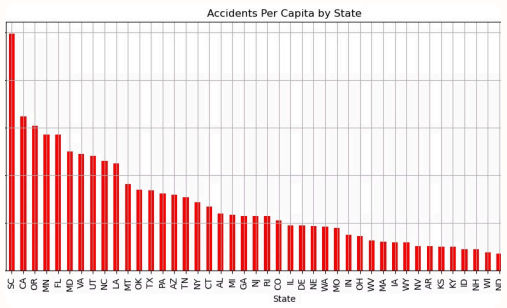
Used `to_sql()` for seamless database storage
- Data served to visualizations using JavaScript APIs

Identifying Accident Hotspots and Trends: README









Interactive Traffic Accident Map

Temporal Trends

Severity analysis

State analysis

Interactive Traffic Accident Map

This map visualizes U.S. traffic accidents by year and weather, highlighting hotspots like California, Texas, and Florida, and showing correlations with driver density.

Temporal Trends

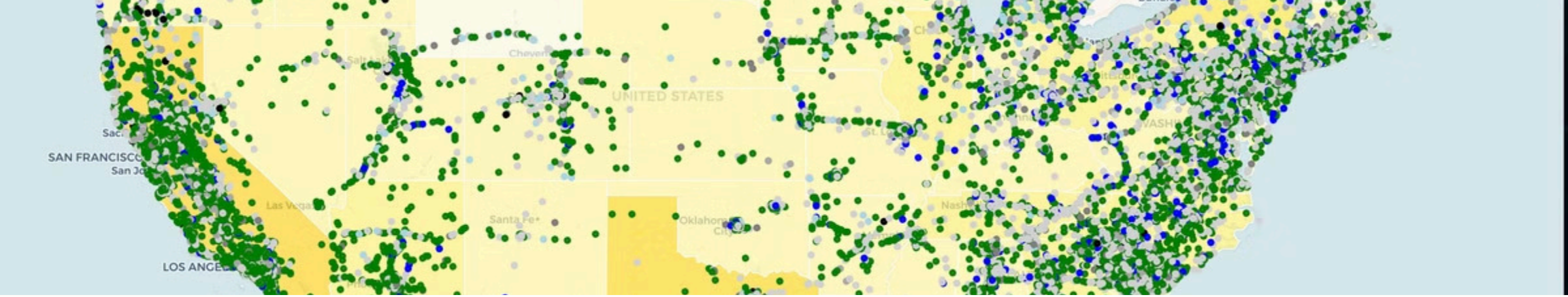
Spikes during holidays, weekends, or/and weather

Severity Analysis

Understand causes of the severity of the accident, the metric on our data showed a number between 1 and 4, where 1 indicates the least impact on traffic

State analysis

This analysis identifies the top number of accidents by state



Visualizing the Data in a Web Application

1

Geographic Heatmap

Identify accident hotspots

2

Time Series Charts

Analyze trends by hour, day, month

3

Interactive Dashboards

Allow users to explore data filters

Key Challenges and Recommendations

Data Size

Managing our large dataset required efficient cleaning and processing to maintain performance and ensure timely analysis.

Data Accuracy

Deleting NaN values resulted in the loss of valuable data, so we chose to convert NaN values to 0 instead. Defining precipitation was also challenging due to the variety of terms used to describe rain and snow.

Random Data Sampling

We created a program to randomly sample 1% of the data. This technique, commonly used in machine learning for handling very large datasets, allowed us to share the data, maintain the integrity of the results, and keep the data random and unbiased.

Data bias

While charting accidents, we observed an unusual data bias. The majority of accidents occurred in temperatures between 62-72°F with perfect visibility and clear skies. This suggests that the dataset has a higher number of entries under these conditions, likely because people tend to drive more in favorable weather. To make more accurate predictions, we need data on accidents during every type of drive, comparing how people drove in clear conditions versus adverse conditions like rain, wind, or snow.

And above all, remember to drive safe—don’t drink, don’t text, and stay focused on the road. Thank you!