

Exploratory Data Analysis of U.S. Accidents



CSIT 553 – Project 2

Student Names: Joana Dervishaj
Vinayak Vemula
Vishwanath Bhupathi
Lam Nguyen

Project Description

- This project analyzes fatal motor vehicle crashes in the U.S. using 2022 data from the Fatality Analysis Reporting System (FARS). The dataset includes detailed information on over 39,000 incidents.
- Each record includes details like time, location, number of people and vehicles, and conditions such as weather and lighting—helping identify factors behind fatal crashes.
- The goal is to uncover trends and risks through visual tools like maps and interactive charts, supporting safety efforts and informed decision-making.

Project Objectives

- Analyze the temporal and spatial distribution of fatal traffic accidents.
- Create geographic maps and aggregated charts to visualize crash data.
- Use interactive plots to explore the data in depth.
- Identify correlations between time, weather, location, and crash severity.

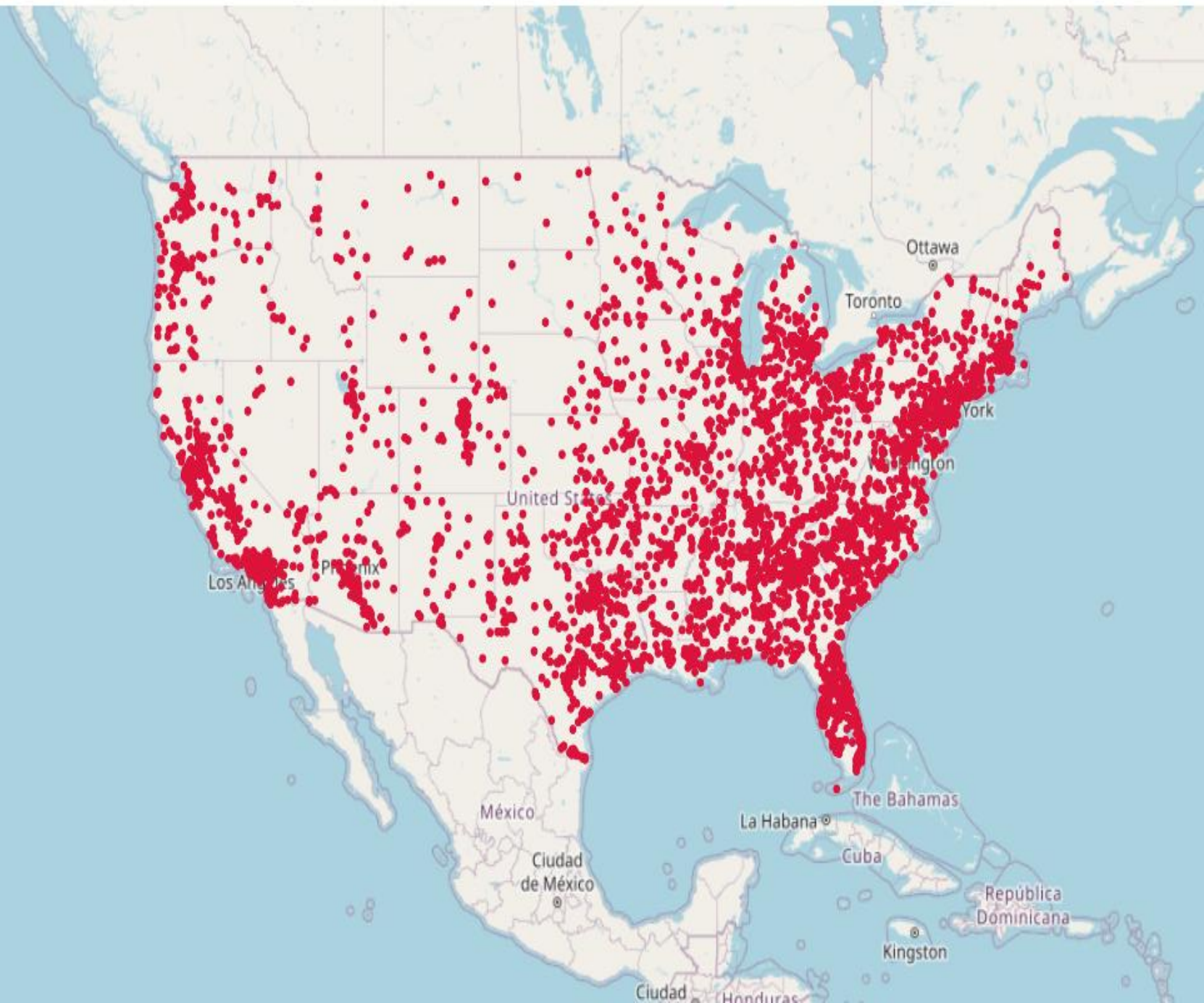
Dataset Overview

- Data comes from the FARS 2022 report on fatal traffic crashes in the U.S.
- Contains over 39,000 records with details on time, location, and conditions.
- Features include: time, location, vehicle types, weather, lighting, and number of fatalities.
- Cleaned to remove missing or incorrect coordinates for accurate mapping.

Bubble Map

- Shows 5,000 fatal crashes across the U.S. using location data.
- Each bubble's position is based on GPS, with size or color reflecting severity or weather.
- Hover details show state, hour, day of week, and weather condition.
- Map created using Plotly with Mapbox and styled as an open street map.

Bubble Map for Fatal Crashes Across the U.S. (2022) - Sample of 5,000

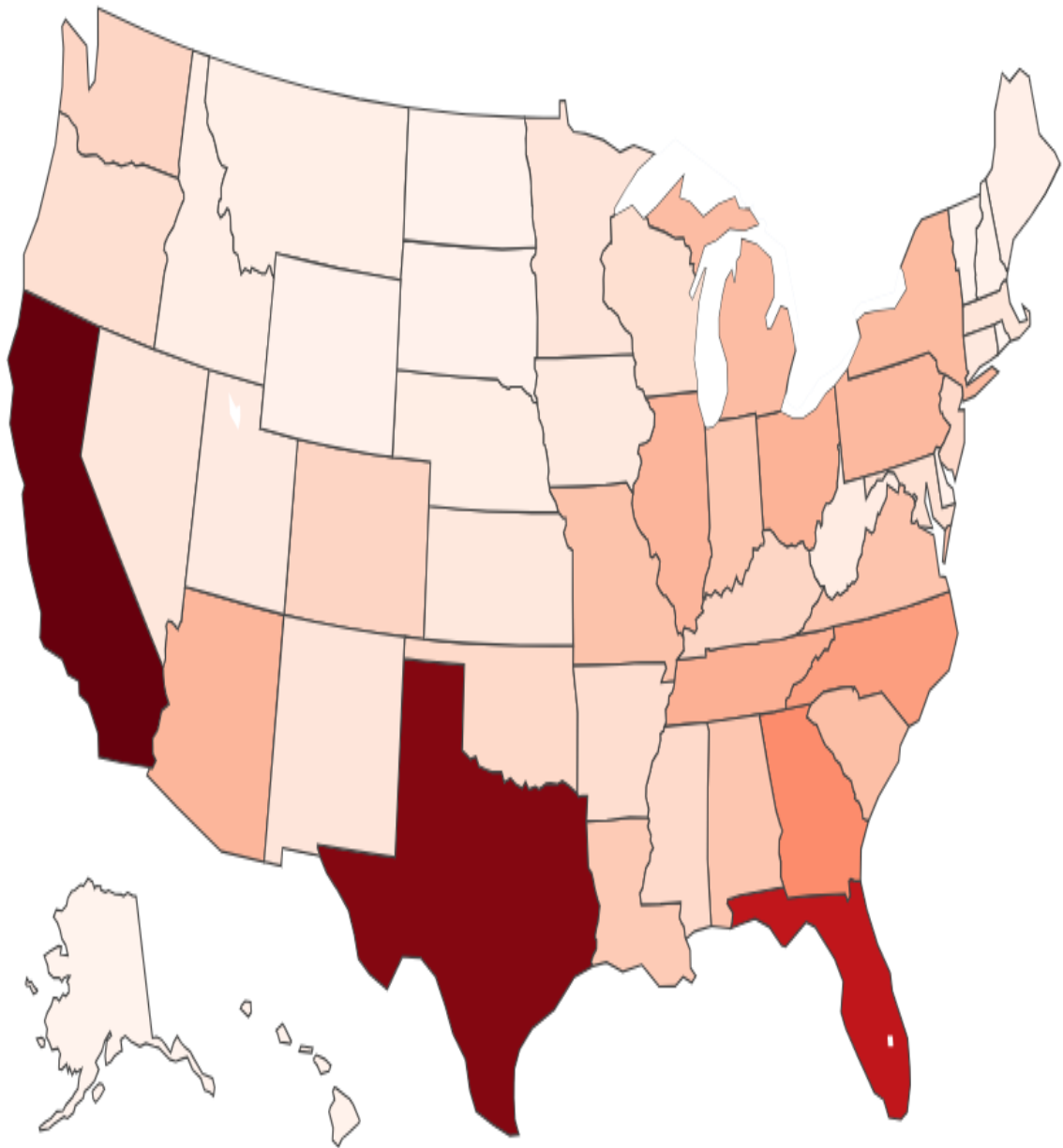


Choropleth Maps:

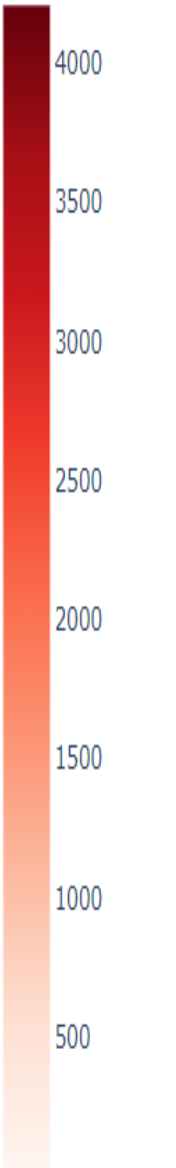
1. Fatal Crashes by State

- Visualizes the number of fatal crashes across U.S. states using color intensity.
- Darker shades represent states with higher numbers of fatal crashes.
- Helps identify regions with the highest crash frequency.

Number of Fatal Crashes by U.S. State (2022)



NUM_CRASHES



2. Crashes & Fatalities per 100K

- Two maps:
 - (1) Crashes per 100K people
 - (2) Fatalities per 100K people.
- Adjusting for population highlights risk levels more fairly.
- Useful for comparing states regardless of size.

Fatal Crashes per 100K Population by U.S. State (2022)



Total Fatalities per 100K Population by U.S. State (2022)

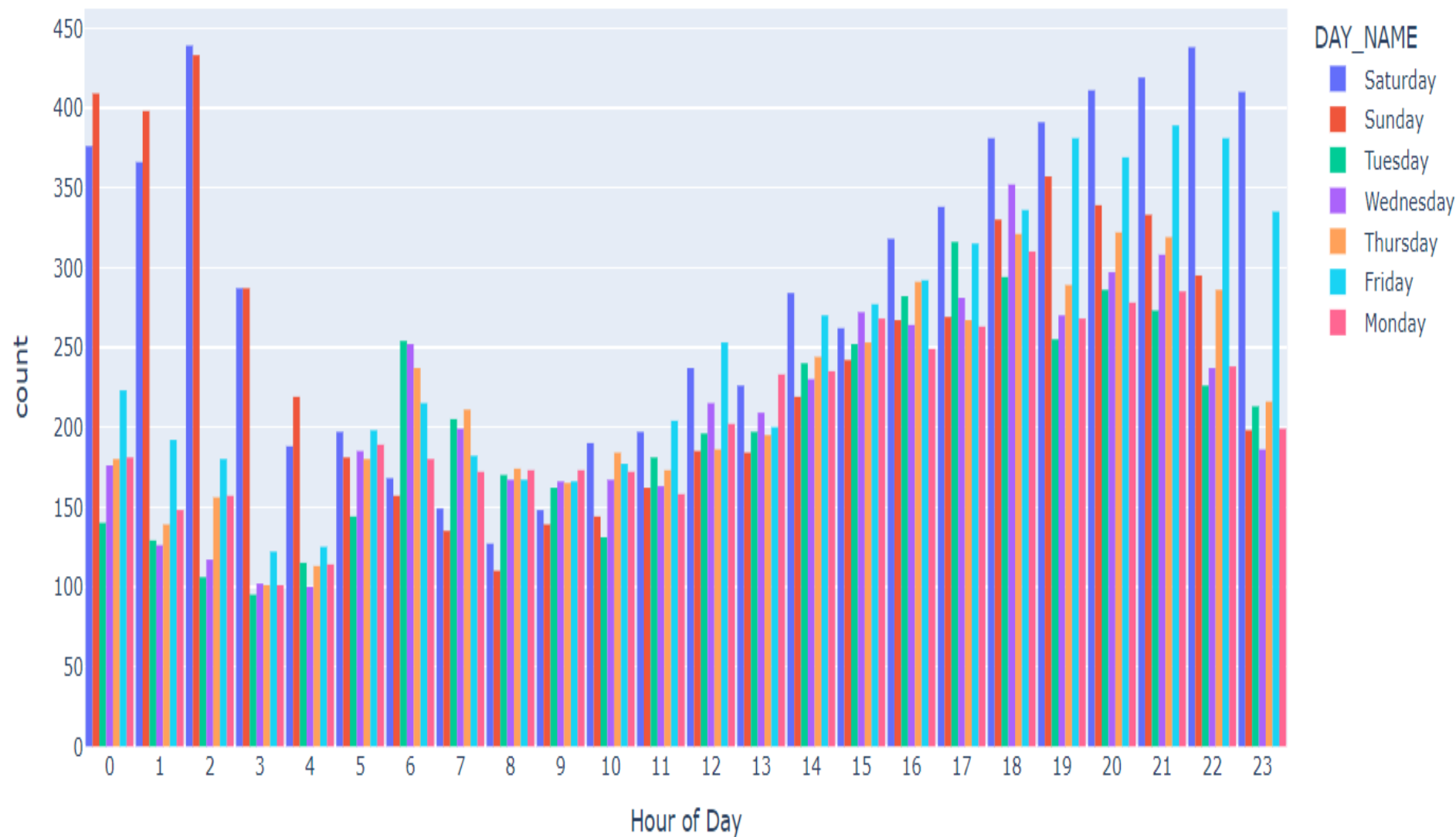


Histogram Maps:

1. Hourly Crashes by Day

- Shows number of crashes by hour, broken down by day of the week.
- Highlights peak hours for fatal accidents, like evening rush.
- Color-coded bars make weekday patterns easy to compare.

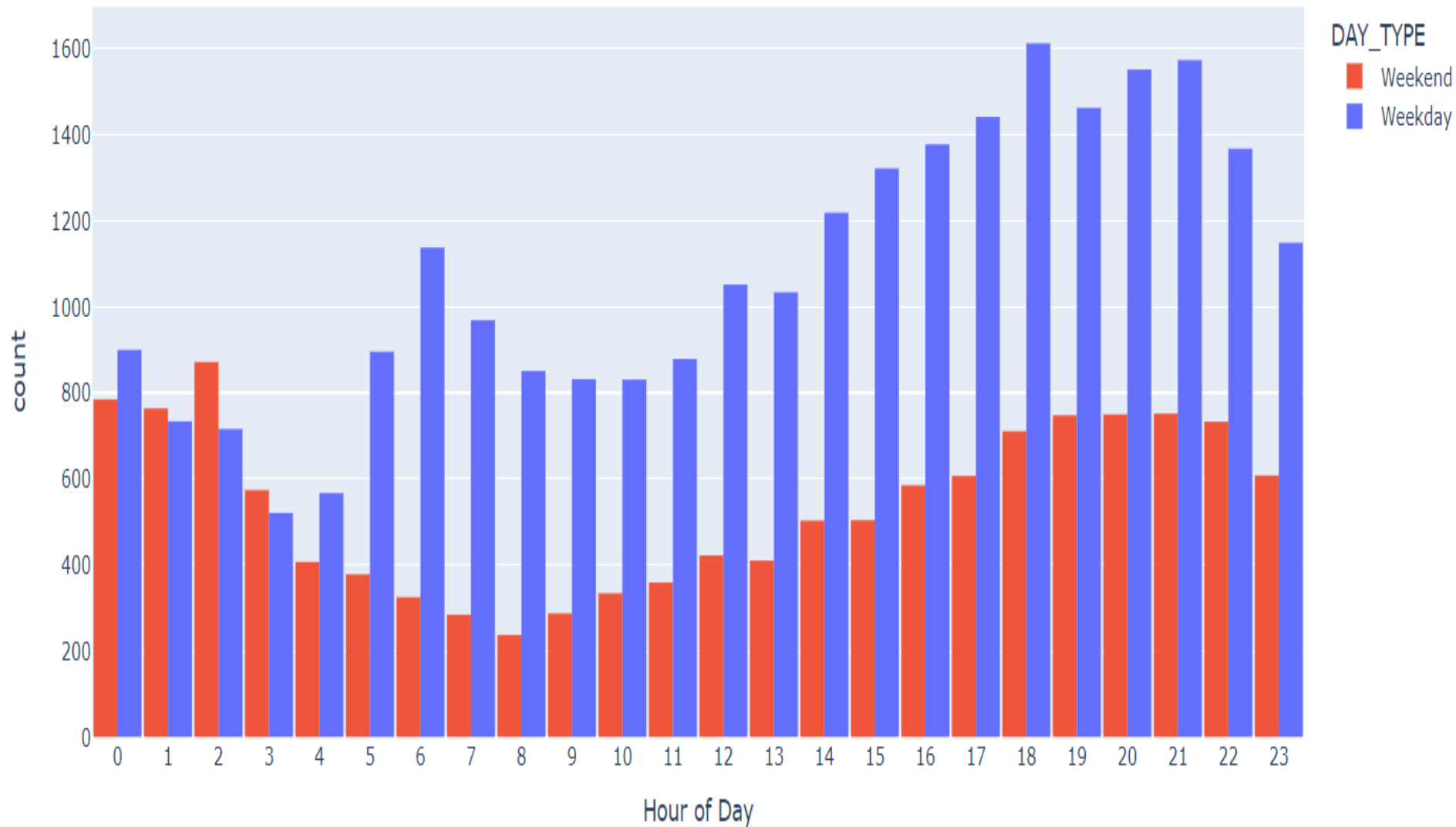
Hourly Distribution of Fatal Crashes by Day of Week (Valid Hours Only)



2. Weekday vs Weekend Crashes

- Bar chart separating crash counts by hour for weekdays and weekends.
- Shows different patterns between workdays and leisure days.
- Clear grouping helps identify when roads are most dangerous.

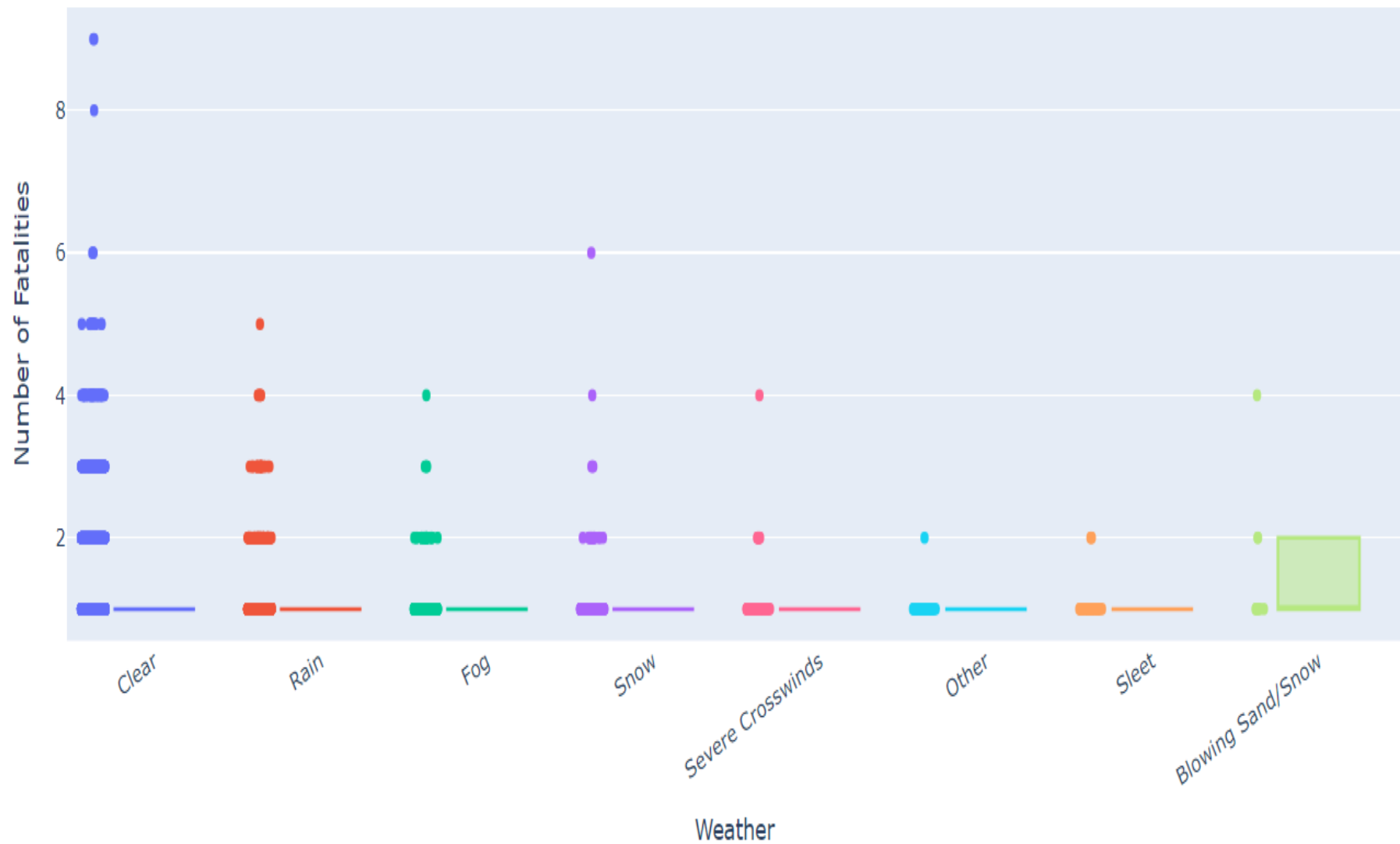
Fatal Crashes by Hour: Weekday vs Weekend



Box Plot

- Displays the distribution of fatalities per crash under different weather conditions.
- Useful to understand how adverse weather impacts crash severity.
- Box plots show distribution, with outliers and medians.

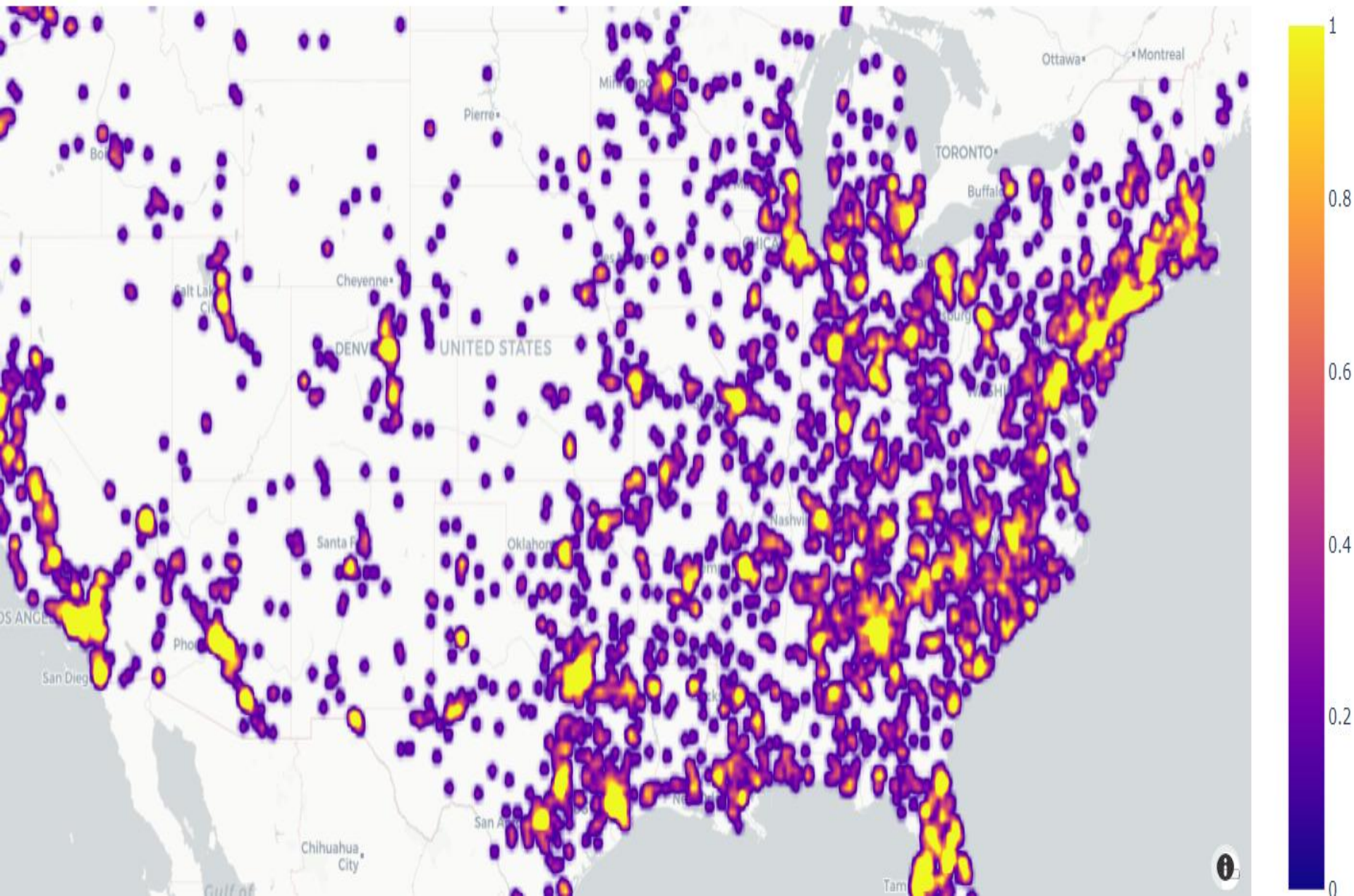
Box Plot for Distribution of Fatalities per Crash by Weather Condition



Heat Map

- Heatmap showing concentration of fatal crashes in the U.S.
- Identifies high-risk areas where crashes are concentrated.
- Brighter areas indicate higher crash density.
- Useful for spotting high-risk regions visually.

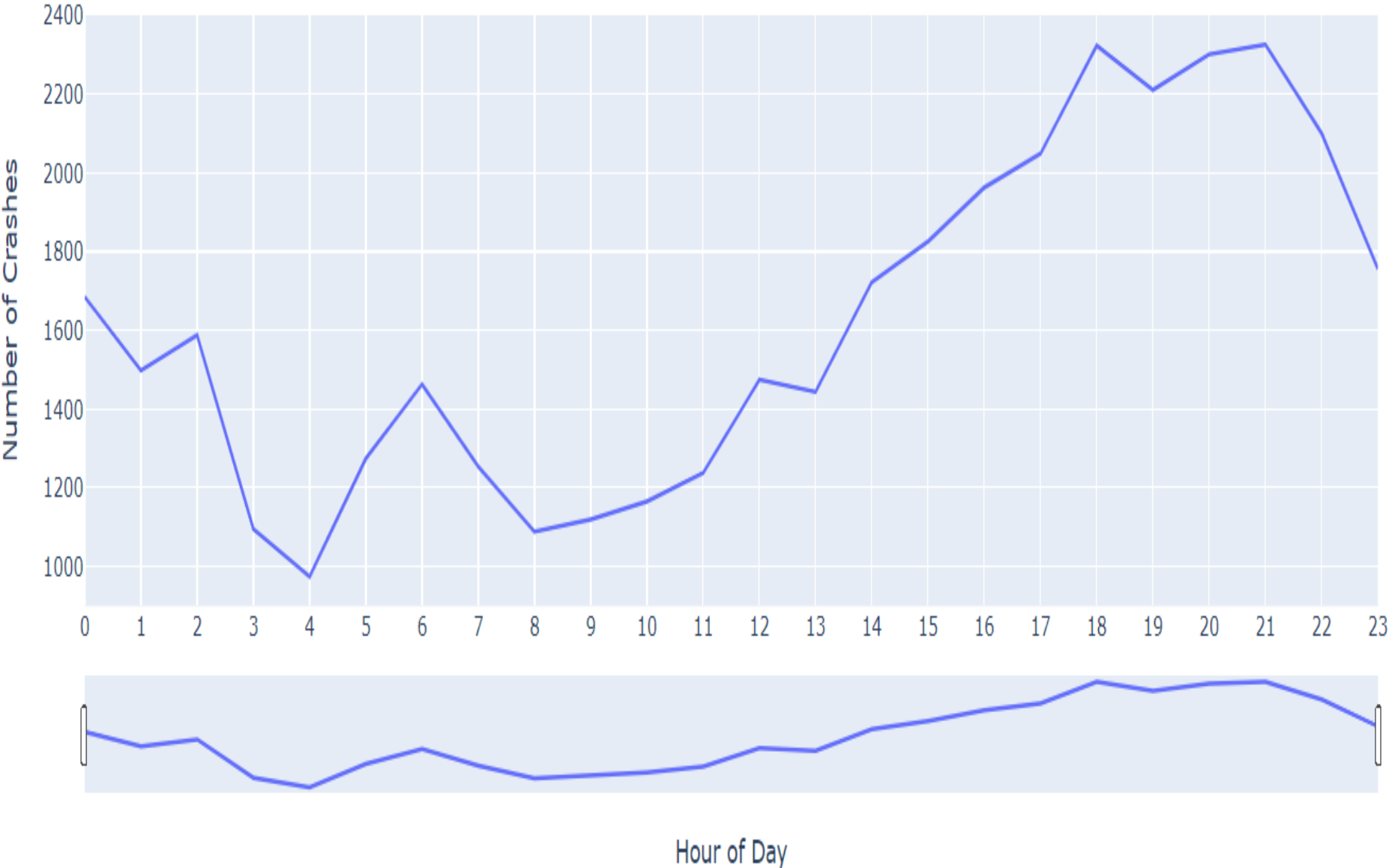
Refined Heatmap of Fatal Crash Hotspots in the U.S.



Time-Series

- Line plot showing number of crashes by each hour of the day.
- Interactive with scrubbing and zooming enabled.
- Reveals overall crash trends over the day.

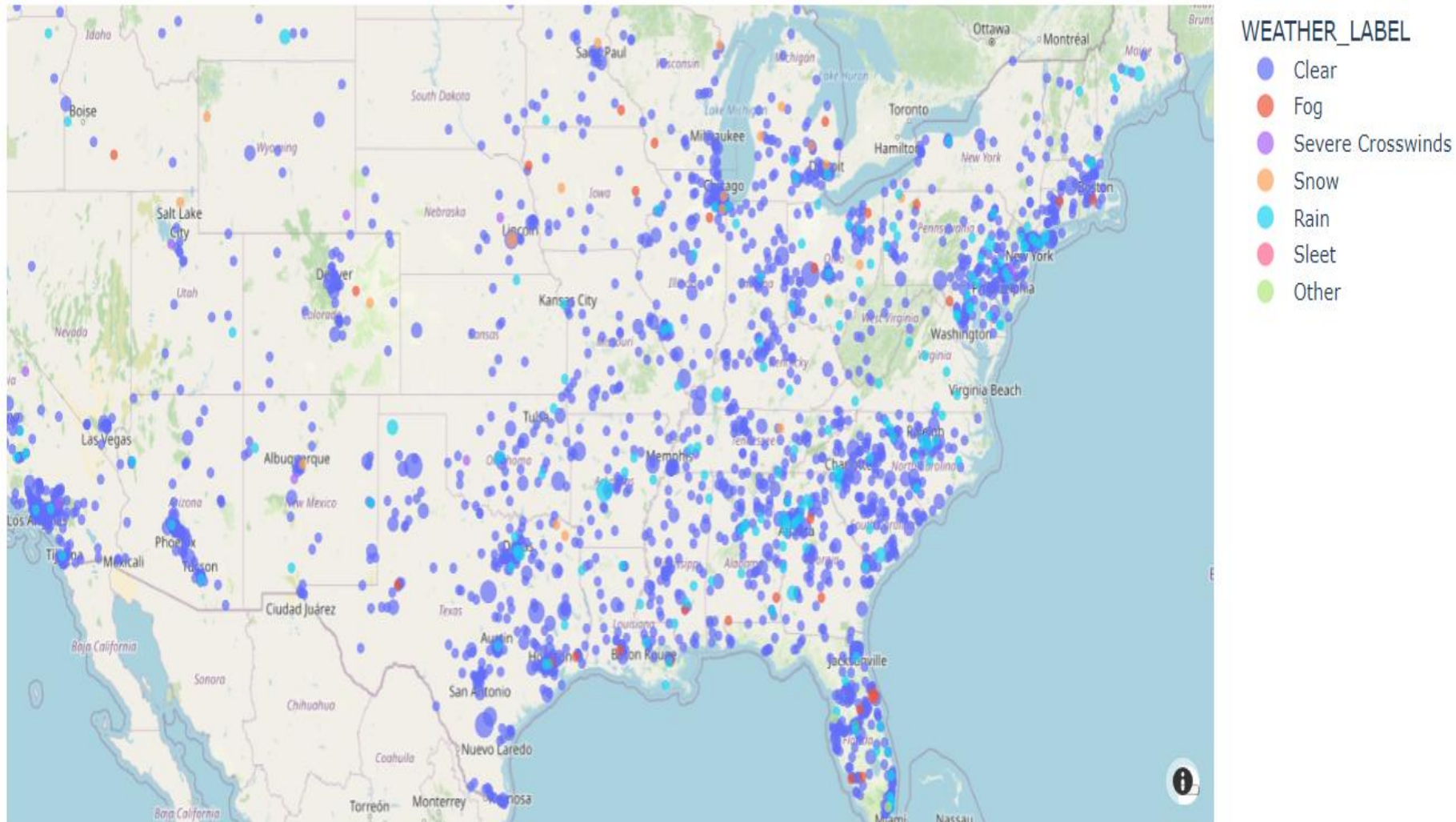
Time-Series of Fatal Crashes by Hour of Day



Bubble Map

- Bubble map with crashes color-coded by weather conditions.
- Bubble size represents number of fatalities at that location.

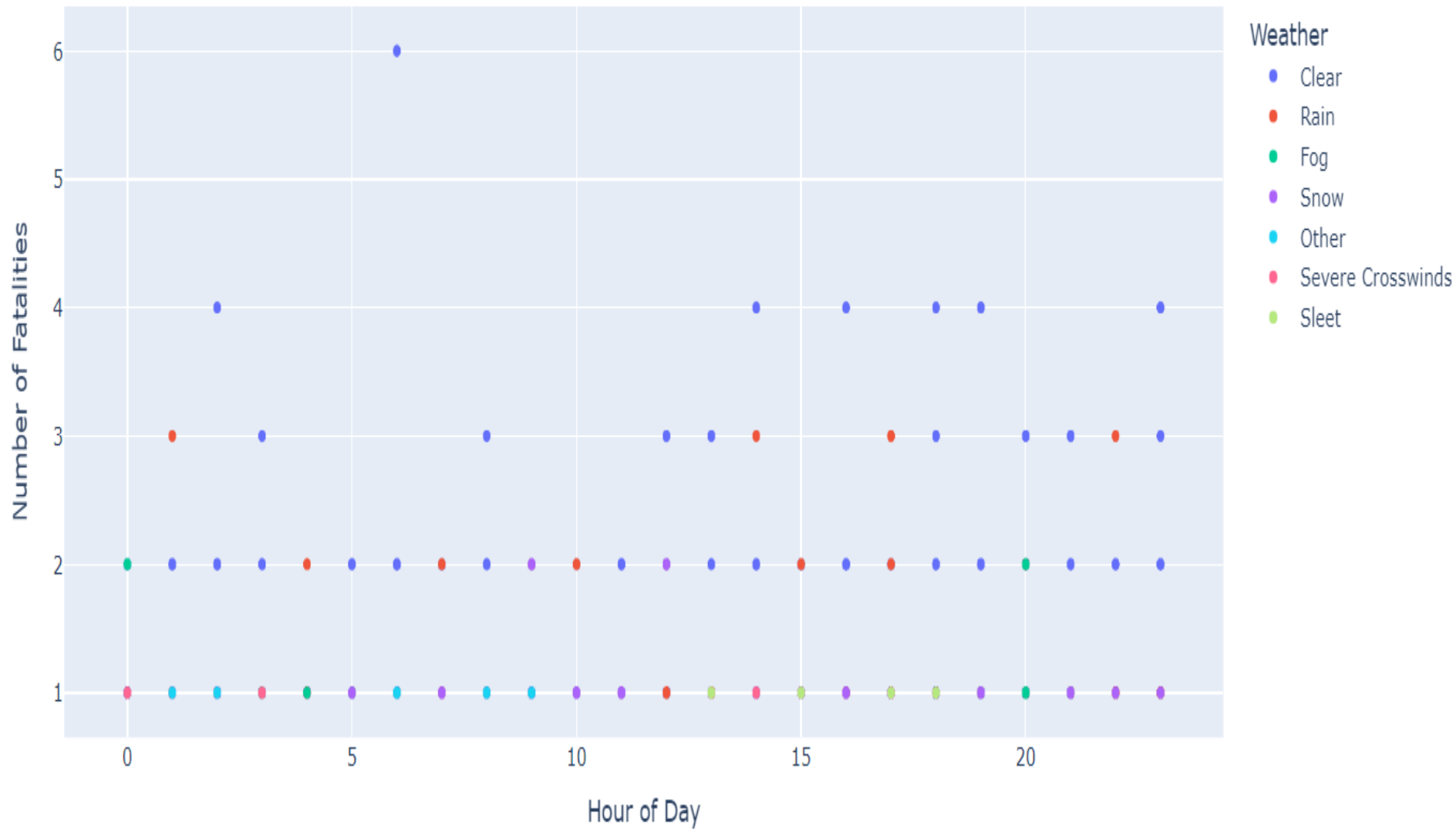
Interactive Map of Fatal Crashes by Weather Condition (Sample of 3,000)



Scatter Plot

- Scatterplot showing fatalities at different hours, colored by weather.
- Shows how weather conditions may affect severity over time.
- Helps compare patterns between clear and adverse conditions.

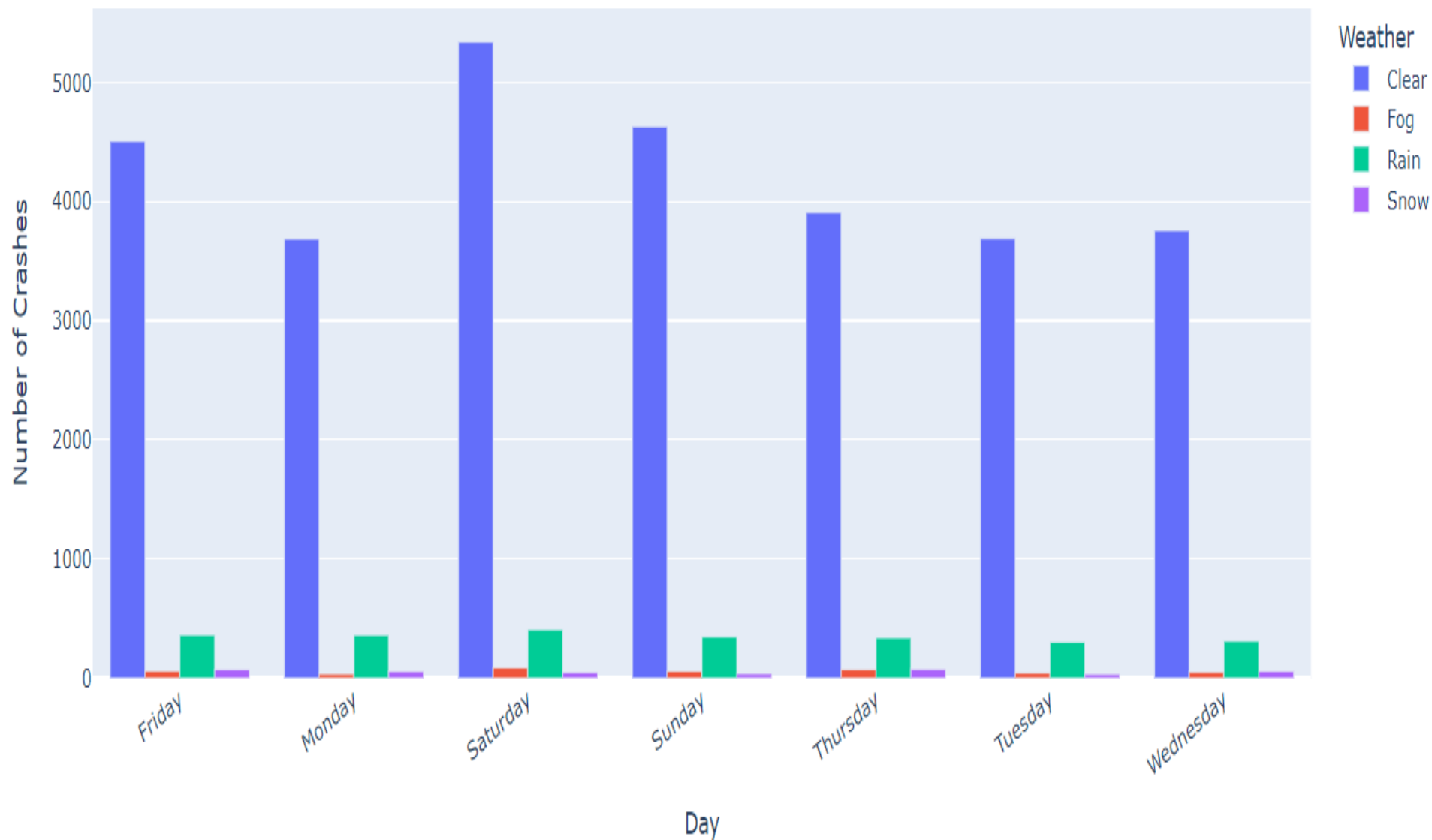
Fatalities by Hour Colored by Weather Condition



Bar Chart

- Grouped bar chart displaying crash count per day under top 4 weather conditions.
- Reveals interaction between weekday and weather effects.
- Makes it easy to compare which days are riskier in bad weather.
- Highlights the interaction between daily trends and conditions.

Fatal Crashes by Day of Week (Top 4 Weather Conditions)



Demo Overview

- All code implemented in a Jupyter Notebook using Python.
- Libraries: pandas, plotly, seaborn, mapbox.
- Dataset cleaned and visualizations generated live.
- Interactive maps and plots demonstrate real-time exploration of crash patterns.
- Python Code Link:
<https://shorturl.at/jqPms>

Conclusion

- Evening hours (5–8 PM) are the most dangerous.
- Poor lighting and adverse weather conditions increase fatality severity.
- States like Texas and California show highest raw crash numbers.
- Visualizations provide actionable insights for traffic safety planning.

Q & A



Thank you for your
attention!



We welcome your
questions and feedback.