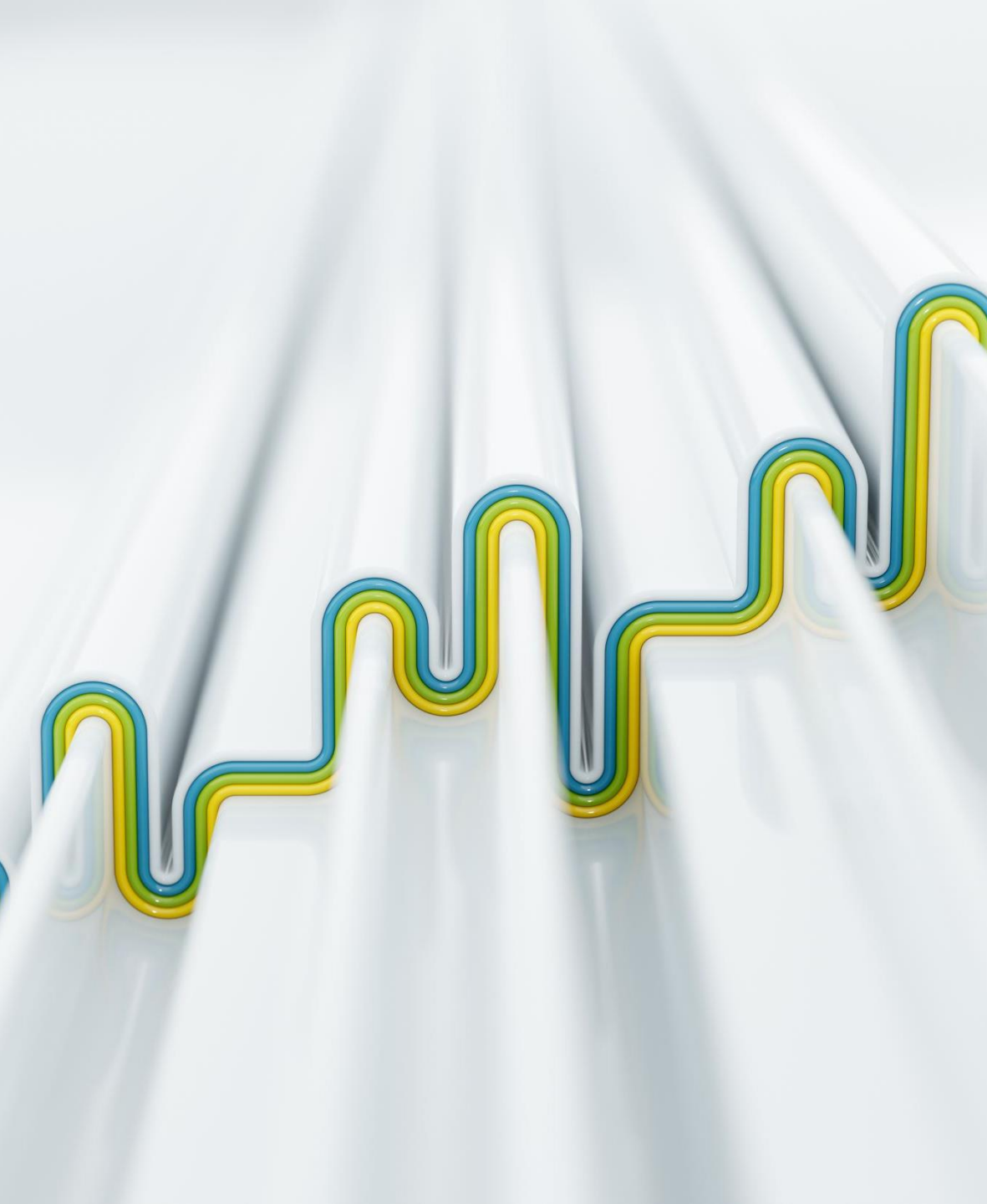# Analysis of Factors Affecting Road Accidents in the US

An In-depth Exploration of Accident Data
Presented by: Charles Paulinus

Introduction

This analysis explores the factors affecting road accidents in the US. The objective is to understand the key elements influencing accidents and the length of road impacted. Data Source: US Accidents dataset.

Data Preparation and Preprocessing

- - Handling missing values.
- - Dropping irrelevant or duplicate records.
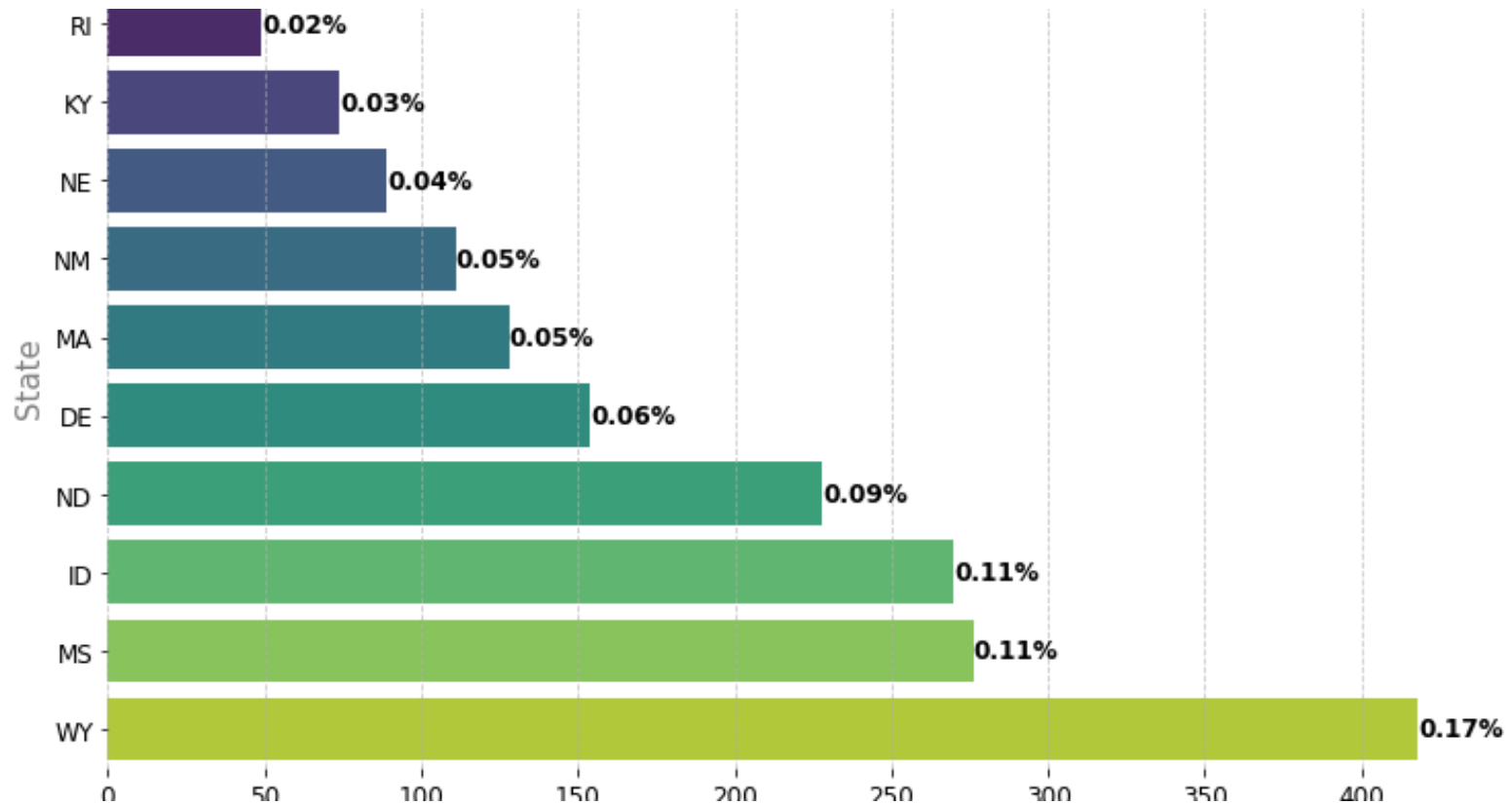- - Converting data types for analysis.
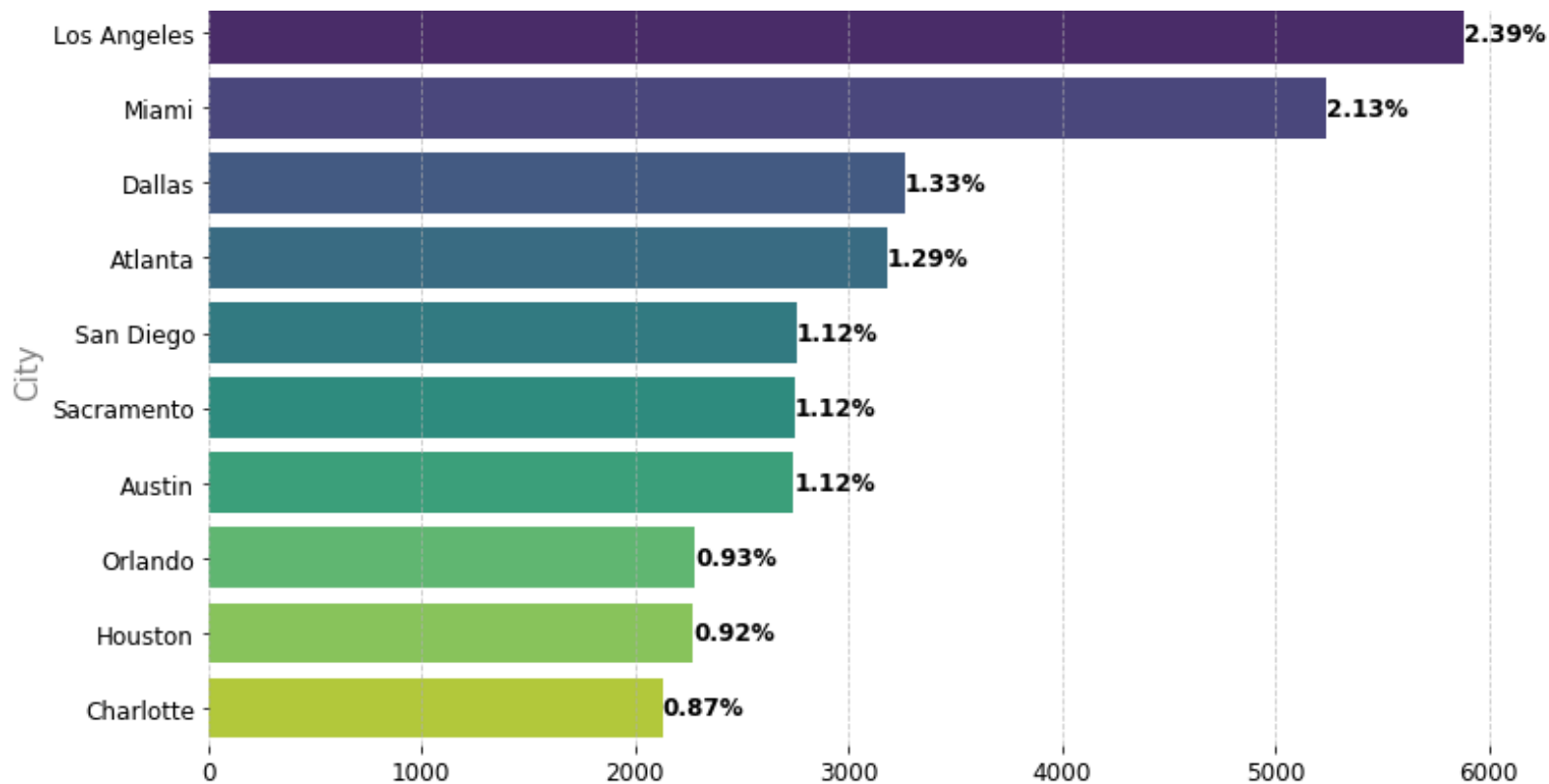
Exploratory Data Analysis (EDA)

Cities with most accidents:

Los Angeles (2.39%)
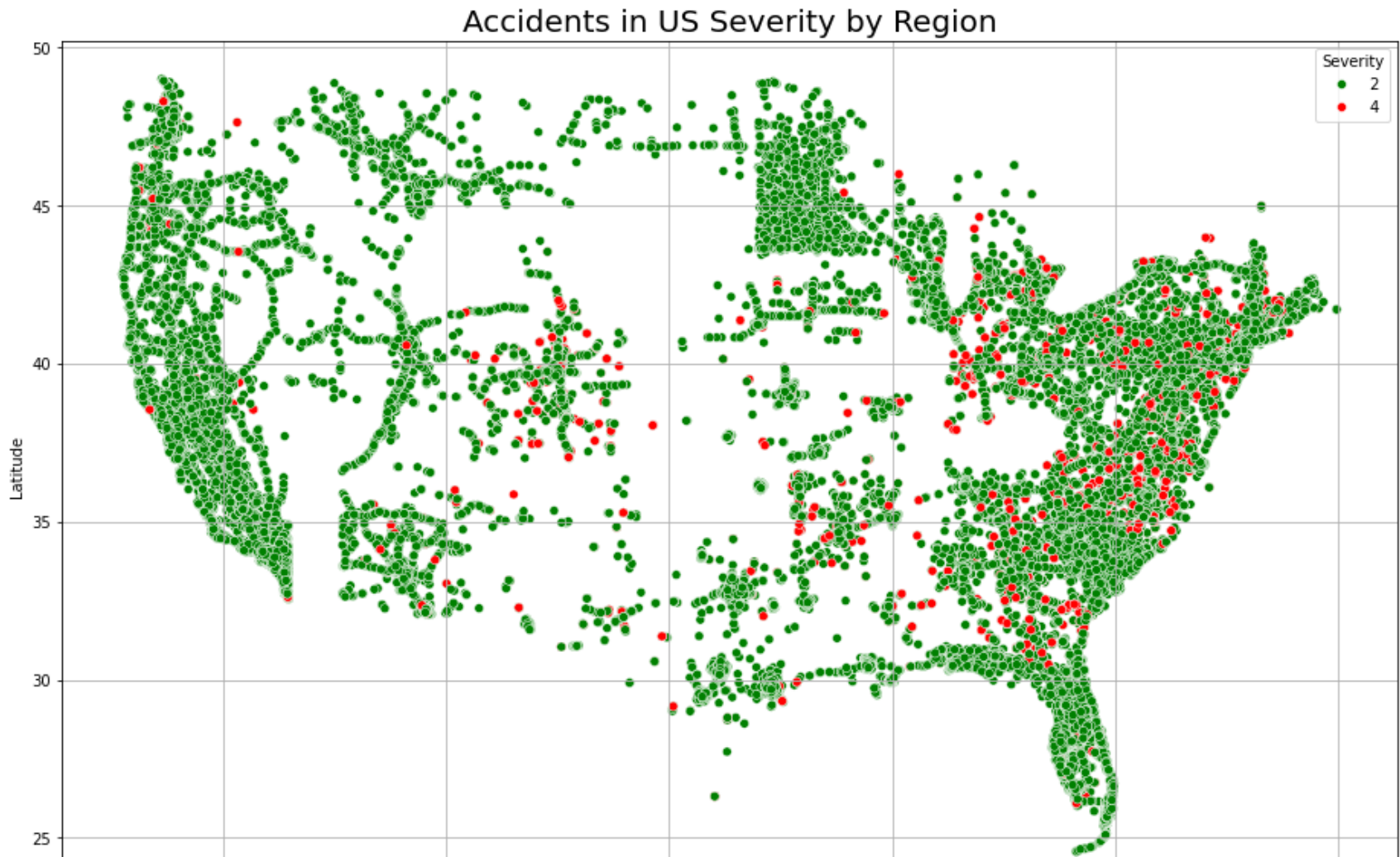
Miami (2.13%)

Exploratory Data Analysis (EDA)
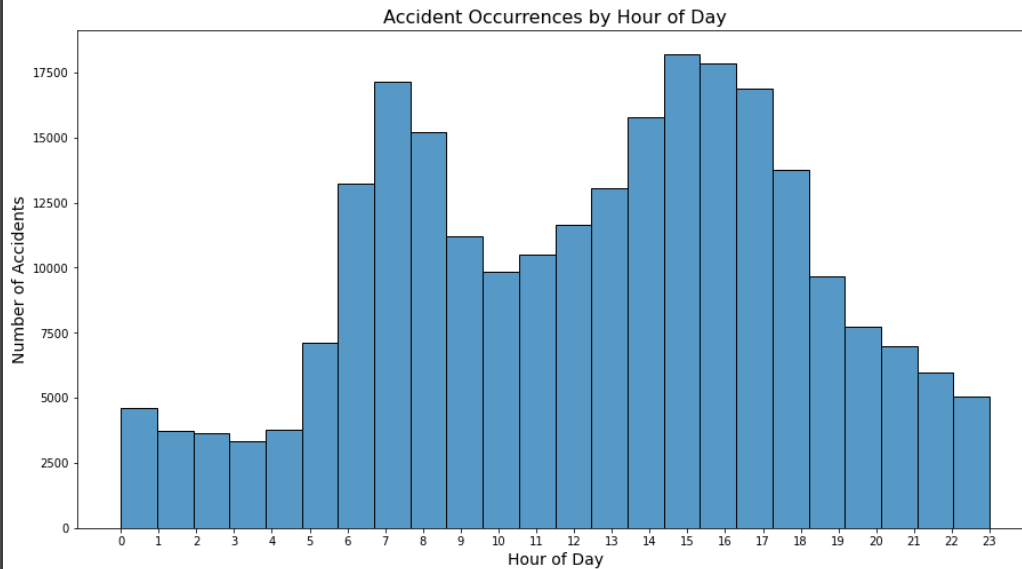
Cities with Least Accidents:

Rhode Island (0.02%)

Kentucky (0.03%)

Spatial Analysis

- Key Insight: The eastern part of the US is more prone to severe road accidents compared to the western part.
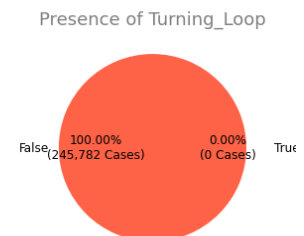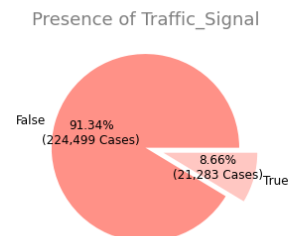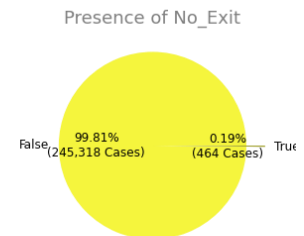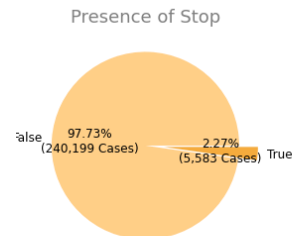

Accidents in US Severity by Region

Accident Occurrences by Hour of Day

Temporal Analysis

Key Insight: Accidents show a bimodal distribution, peaking during morning and evening rush hours.

Accidents by Road Condition

The presence of traffic signals, crossings, and give-ways are highly associated with road accidents.
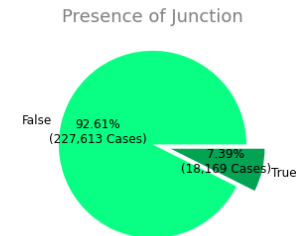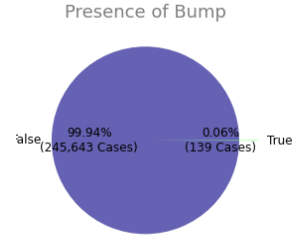
Severity Count for each Weather Condition
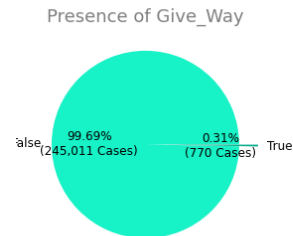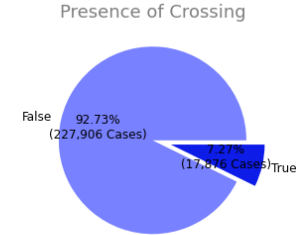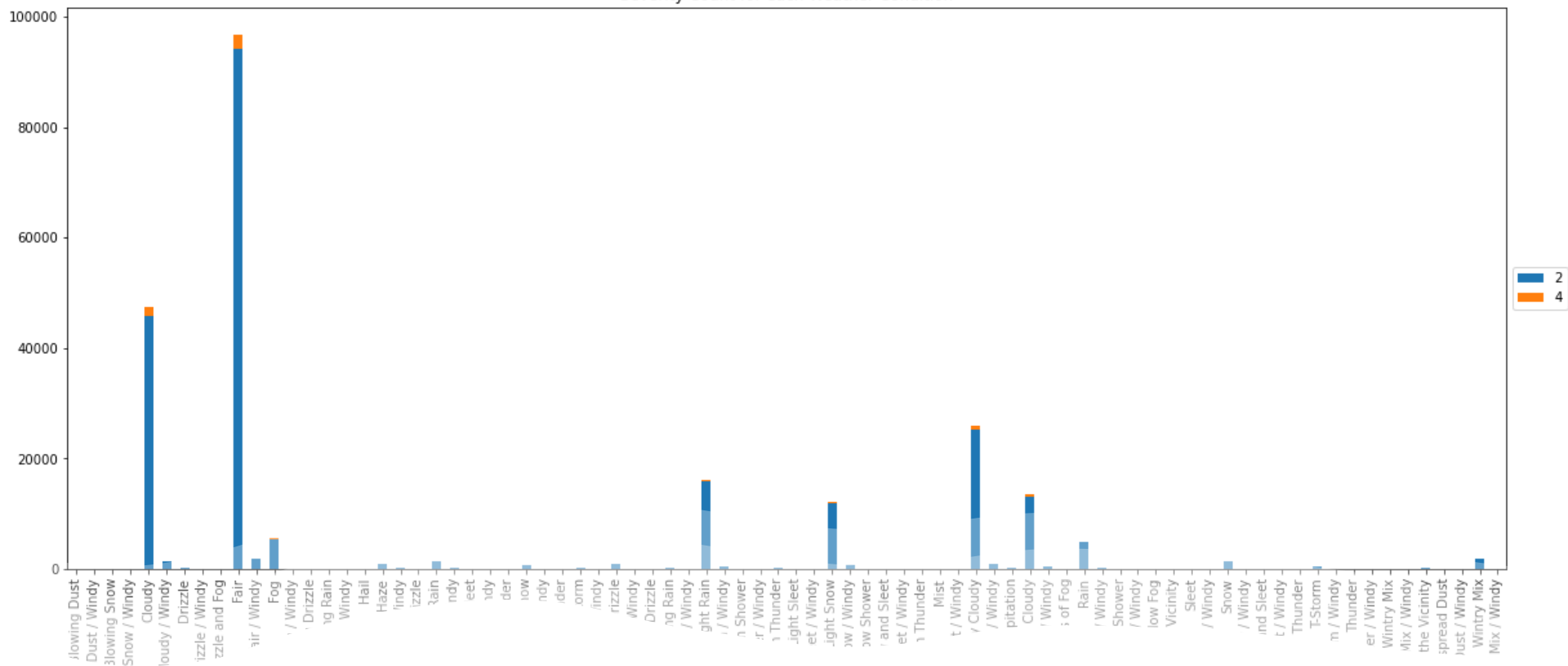
Accidents by Weather Condition

- Most accidents occur in fair weather conditions, suggesting either that most driving occurs during fair weather conditions or potential mislabeling or further investigation needed.

Timezone Analysis

Percentage of Accident Cases for different Timezone in US
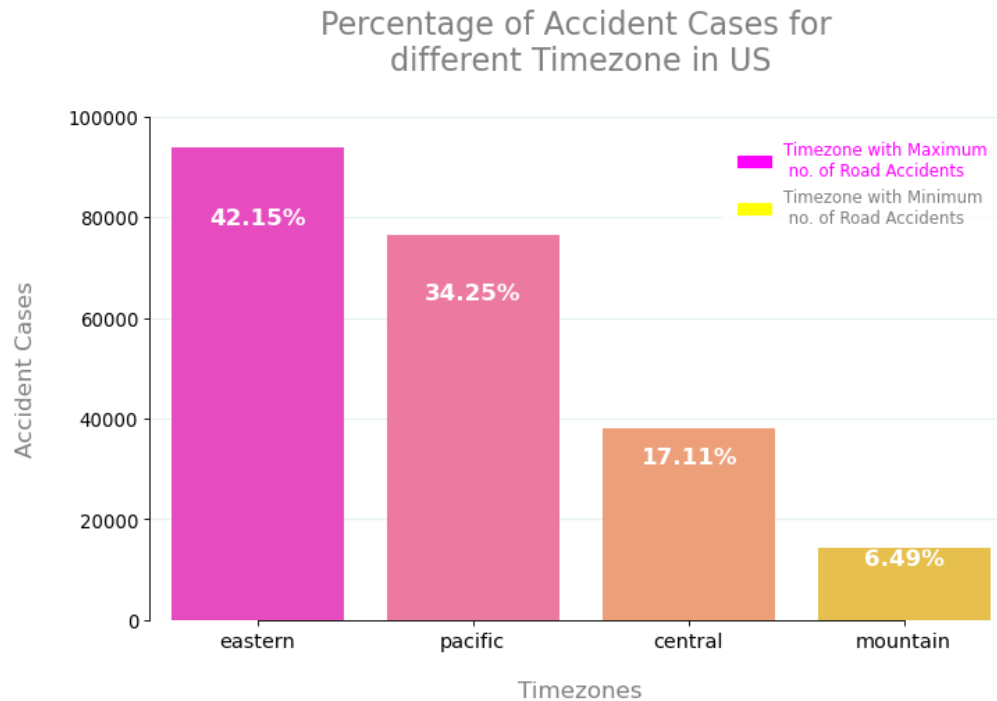
- The Eastern timezone has the highest percentage of accident cases, while the Mountain timezone has the least.

Feature Selection and Model Building

- Process Overview:
- Selection of relevant features for the model.
- Removal of highly correlated features.

- Model: Random Forest Regression used to predict the length of the road affected by accidents.
- Feature Engineering: Inclusion of engineered features like median temperature in Celsius, UTC offset, etc.
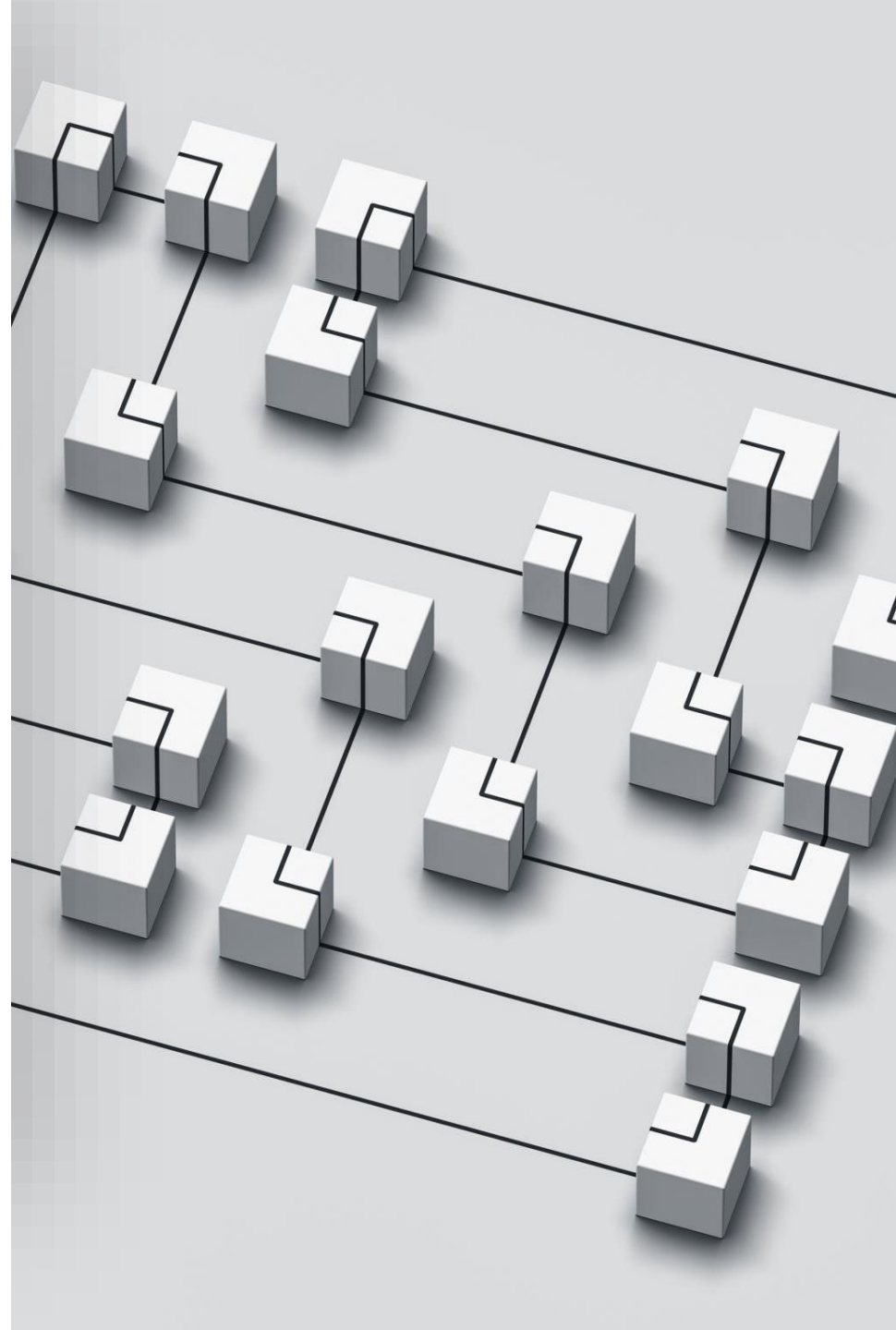
Model Evaluation

Model Performance:

- - Mean Absolute Error (MAE)
- - Mean Squared Error (MSE)
- - Root Mean Squared Error (RMSE)
- - R-squared (R2)

Key Factors Affecting Road Length Impacted by Accidents

- Air Pressure in inches
- Visibility in miles
- Starting location of accident
- Average temperature per location
- The presence of snow
- Wind blowing in the NW direction

Conclusion and Recommendations

- Focus on improving road conditions in high-risk areas.

- Implement better traffic management during peak hours.

- Conduct further investigations into weather condition labeling.

Q&A

Questions and further discussions.