RoadSafe Analytics

US Road Accident Data Analysis & Interactive Dashboard

1. Introduction

Road accidents are a major public safety concern worldwide. In the United States, millions of accidents occur every year due to traffic congestion, poor weather, low visibility, and human error. Traditional static reports are insufficient to analyze such massive datasets. RoadSafe Analytics addresses this challenge by using data analytics and interactive visualization to uncover hidden patterns and insights from large-scale accident data.

2. Problem Statement

Despite the availability of large accident datasets, decision-makers struggle to extract meaningful insights due to data volume, lack of interactivity, and complex relationships between variables. This project aims to analyze US road accident data and present insights through an interactive dashboard.

3. Objectives

- Analyze large-scale accident data
- Identify temporal accident patterns
- Study weather and visibility impact
- Detect accident-prone regions
- Validate accident-related hypotheses
- Build an interactive Streamlit dashboard

4. Dataset Description

Dataset: US Accidents Dataset (Kaggle)

Period: 2016–2023

Records: Over 4 million

Features include severity, time, location, weather, visibility, and traffic signals.

5. Technology Stack

Python – Core programming language

Pandas & NumPy – Data processing

Matplotlib & Seaborn – Visualization

Streamlit – Interactive dashboard

6. Methodology

- Data Loading & Cleaning
- Feature Engineering (Hour, Weekday, Month)
- Univariate, Bivariate & Multivariate Analysis
- Geospatial Analysis
- Hypothesis Testing
- Dashboard Development

7. Dashboard Features

- Dynamic filters (State, Weather, Severity, Day/Night)
- Multiple analysis tabs
- Interactive plots and KPIs
- Real-time updates based on filters

Milestone-wise Project Implementation

Milestone 1: Data Acquisition & Understanding

In the first milestone, the US Accidents dataset was collected from Kaggle. The primary goal was to understand the structure, size, and attributes of the dataset. Initial exploration involved checking the number of records, identifying important columns, and understanding accident severity levels, time attributes, location data, and weather-related features. This stage helped in defining the analysis scope and identifying relevant variables.

Milestone 2: Data Cleaning & Preprocessing

The second milestone focused on preparing the data for analysis. Missing values were handled, incorrect data types were corrected, and unnecessary columns were removed. New features such as Hour, Weekday, and Month were extracted from the accident start time. Outliers in visibility and distance were treated to improve data quality. The cleaned dataset was saved for efficient reuse.

Milestone 3: Exploratory Data Analysis (EDA)

In this milestone, extensive exploratory data analysis was performed. Univariate analysis helped understand individual variables such as accident severity, time distribution, weather conditions, and traffic signals. Bivariate and multivariate analysis explored relationships between severity and factors like visibility, weather, time of day, and weekdays. Geospatial analysis identified accident hotspots, high-risk states, and cities.

Milestone 4: Visualization, Insights & Dashboard Development

The final milestone involved transforming analysis results into an interactive Streamlit dashboard. Dynamic filters for state, weather, severity, time, and visibility were implemented. All plots from previous milestones were integrated into structured tabs. Insights and hypothesis testing were presented visually, enabling data-driven conclusions. This milestone completed the end-to-end analytics workflow.

8. Key Insights

- Accident peaks during rush hours
- Poor visibility increases severity
- Rain and fog lead to severe accidents
- Certain states and cities are consistently high-risk

9. Challenges & Solutions

Large dataset size – Solved using caching and sampling

Data inconsistency – Solved through preprocessing

Performance issues – Optimized rendering

10. Real-World Applications

Traffic safety planning

Emergency response optimization

Policy-making and insurance analysis

11. Future Enhancements

Machine learning prediction models

Real-time data integration

Mobile app support

12. Conclusion

RoadSafe Analytics successfully transforms complex accident data into actionable insights using data analytics and interactive visualization, supporting informed decision-making for road safety.