US Accidents Insights

IA640: Information Visualization

Project -2 (Report) -Shiny

Instructor - Thevasha Sathiyakumar

Rahul Tammalla – 1010517, Clarkson University

Project Overview:

Road safety is a critical issue in the United States, with millions of accidents occurring annually, causing loss of life, injuries, and economic damage. This report presents an analysis of US traffic accident data spanning from 2016 to 2023 using an interactive Shiny application, US Accident Insights. The objective is to uncover patterns and trends in accidents, enabling better policy formulation and resource allocation to improve road safety.

Shiny App URL:

https://clarksonuniv.shinyapps.io/US-Accident-Insights-Rahul/

Dataset:

The dataset contains detailed 42768 records of traffic accidents across the United States, sourced from Kaggle.

Link for dataset: https://www.kaggle.com/datasets/pateluttam/usaccidents

Pre-processing data:

1. **Library Importation**
   - The preprocessing utilized several R libraries, including:
     - dplyr for data manipulation.
     - lubridate for date and time operations.
     - shiny, leaflet, ggplot2, and plotly for building the interactive Shiny application and creating visualizations.

2. **Dataset Loading:**
   - The dataset is loaded using read.csv with stringsAsFactors = FALSE to avoid automatic factor conversion

3. **Column Selection:**

- There are a total of 47 columns in the file, but I have retained relevant columns to streamline the analysis. Selected columns:

  - Accident Details: ID, Severity, Start_Time, End_Time.

  - Location Information: Start_Lat, Start_Lng, State, City.

  - Weather and Environmental Factors: Weather_Condition, Temperature.F., Humidity..., Wind_Speed.mph., Visibility.mi.

  - Traffic Data: Traffic_Signal, Distance.mi.

4. Handling Missing Values

- Rows missing critical fields, such as Severity, Start_Time, End_Time, and State, were removed.

- Weather_Condition: Missing values were replaced with "Unknown."
- Numeric Columns: Columns such as Temperature.F., Humidity..., Wind_Speed.mph., and Visibility.mi. had missing values replaced with the mean of their respective columns.

5. Date and Time Processing

- The Start_Time column was converted to a POSIXct format for accurate temporal analysis.

- A new column, Year, was derived using the lubridate::year() function to facilitate year-wise analysis.

6. Filtering and Finalization

- Data was filtered to retain only rows with complete critical fields (Start_Time, End_Time, and State).

- The processed dataset was stored in a new data frame (accidents_filtered) for use in the interactive Shiny application.

Positive Feedback:

1. The project emphasizes actionable outcomes, such as identifying high-risk locations and times. This practical orientation makes the project not just a data exercise but a meaningful contribution towards enhancing traffic safety strategies.
2. I have utilized f libraries like dplyr, lubridate, and ggplot2 for preprocessing and visualizations, as well as shiny and leaflet for interactivity. Which helped me learn and implement my learnings.
3. I have an excellent dataset with a variety of fields which enabled me to create a geospatial representation of the data.
4. The Shiny app's interactive features, such as filters and dynamic visualizations, help various stakeholders, from researchers to policymakers.

Challenges:

1. The primary challenge was to decide relevant columns, I have spent significant time understanding the dataset and picked the relevant columns which suits my idea of interactive shiny app.
2. I have tried multiple interactive shiny features narrowing it down to interactive histogram and maps which suit perfectly to represent the data.
3. I have incrementally developed the application adding new features without effecting the existing functionality was a challenge.

Conclusion:

The US Accident Insights Shiny app has immense potential for future applications, such as integrating real-time data for traffic monitoring and emergency response, aiding policymakers in prioritizing high-risk areas, and guiding infrastructure development. Additionally, incorporating predictive analytics could enable proactive accident prevention, while its insights can inform educational campaigns to raise awareness about road safety. This app serves as a foundation for expanding data-driven solutions to enhance traffic management and reduce accidents.