Car Accident Severity Analysis

Sahil Rajendrakumar Raut
Rutgers University
Piscataway, NJ, USA
sahil.raut@rutgers.edu

Kaushik Vakadkar
Rutgers University
Piscataway, NJ, USA
kaushik.v@rutgers.edu

Sohailabbas Saiyed
Rutgers University
Piscataway, NJ, USA
ss3723@scarletmail.rutgers.edu

***Abstract*— Reducing traffic accidents is an essential public safety challenge all over the world; therefore, accident analysis has been a subject of much research in recent decades. The objective of the project is to analyze the US accident data from 50 states to inform the US government agencies and the public on trends and possible causes of traffic accidents and what could be done to reduce them. We established a multilevel interactive map to demonstrate over a few million-accident data inside the United States for 49 states. Some of the visuals include the number of accidents by year, number of accidents by state, the best time to travel by month, day and hour, an accident-prone area in each state, factors responsible for the accidents like weather, wind flow, temperature, location, etc. The dashboard plots will be built using D3.js and Plotly, while the front-end will be developed in HTML/CSS and will be deployed through Flask.**

***Keywords*— Traffic accidents, data visualization, data interaction, exploratory analysis, statistical analysis, dashboard, D3, Plotly, Flask, HTML, CSS**

I. Project Description

The economic and social impact of road accidents is affecting U.S. citizens. hundreds of billions of dollars each year. And much of the loss is due to several serious accidents. Road crashes cost the U.S. $230.6 billion per year or an average of $820 per person [1]. Reducing road accidents, especially the worst accidents, however, remains an important challenge. The acceleration method, one of the two main ways to deal with road safety problems, focuses on preventing unsafe road conditions from happening in the first place. For effective use of this method, risk forecasting and difficulty forecasting are essential. If we can identify patterns of how these bad accidents occur and the key factors, we can take informed actions and better share financial and human resources. The primary objective of this project is to recognize key factors contributing to road accidents, without any detailed information about itself, like driver attributes or vehicle type. This visualization is supposed to be able to find the pattern of the accident in the United States. The aim of this project is to establish an interactive dashboard to demonstrate information about accidents and discover patterns with respect to weather conditions, location, and time of the year.

A. Stage 1 - Requirements Gathering

Our dataset [2] is based on data collected from "A Countrywide Traffic Accident Dataset" by Moosavi et. al. [3] with a total size of about a few GB. This dataset has been collected in real-time, using multiple Traffic APIs. Currently, it contains accident data that are collected from February 2016 to Dec 2020 for the Contiguous United States. In this project, we will focus on 2 sets of columns i.e., Accident spatial position and the corresponding weather conditions. We performed exploratory data analysis with Python on Jupyter Notebook using several types of plots such as heat map, time-series, scatterplot, and histogram. The dataset was available in .csv format on the Kaggle.

- Expected Target Users: The main recommendations from the visualization focus on Infrastructure, Policy, Administrative, and Human behaviour-related changes that can be implemented by the state and the federal government. Government agencies and the public can leverage these insights and take preventive measures which can reduce road accidents in the US. Apart from this, the visualization can help state transportation agencies and logistics companies to select the safest and optimal route, weather, and time of the year to reduce losses.
- Project Timeline:

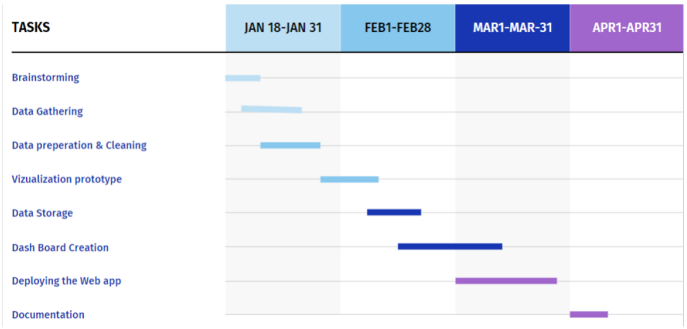

Fig. 1. Gantt chart depicting tasks and estimated timeline

- Division of Labor:
  - Sahil Raut: UI, exploratory data analysis, prototype
  - Kaushik Vakadkar: UI, statistical analysis and data processing
  - Sohailabbas Saiyed: UI, documentation and deployment

B. Stage 2 - Design

- Short Textual Project Description: Our web application majorly focuses on providing interactive statistical graphs which give detailed statistics of our data set built upon two major groups of the data, Accident spatial position and the corresponding weather conditions.

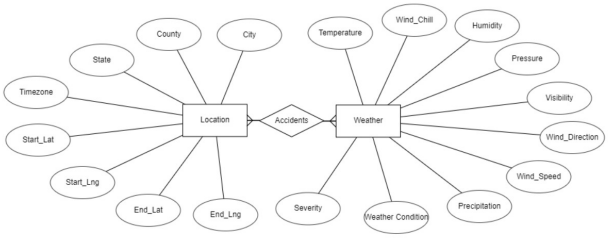

Fig. 2. The ER diagram.

- System Flow Diagram:

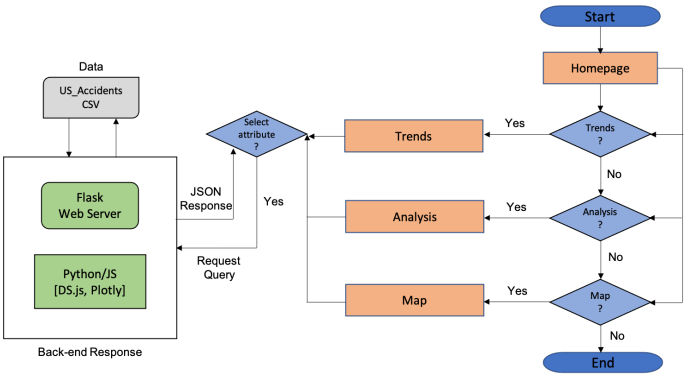

Fig. 3. System flow diagram

C. Stage 3 - Implementation

The components forming the front-end and back-end along with the flow of data and query between them can be seen in Fig 4. The data is read from the CSV flat file. D3.js and Plotly are used to generate all the plots and visualizations. In order to create a menu-driven, interactive dashboard HTML/CSS, Bootstrap, and JavaScript will be used.

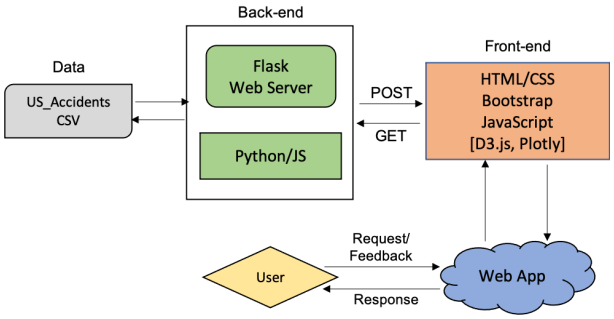

Fig. 4. Component integration diagram.

- Language(s) used:
  - Backend: Python
  - Frontend: HTML 5, CSS, JavaScript
- Software platform (Back-end + front-end)
  - Data: CSV flat file
  - Web server: Flask [4]
  - Front-end framework: Bootstrap
- Software libraries
  - Plotting and visualization libraries: D3.js [5], Plotly [6]

D. Stage 4 - User Interface

The user interface will be developed using HTML, CSS, and JavaScript. The application will focus on three main parts - Trends, analysis and mapping. The user can select from these three options through a side menu bar as shown in Fig. 5. The interactive statistical graphs and plots will be generated using D3.js and Plotly. The plots showcased from Fig. 6 onward have been rendered using Tableau to thoroughly analyse and better understand the plots which will then be used as a basis for developing the main dashboard further on. We examined the hierarchical clustering and partitioning algorithm on our dataset with Python on Jupyter Notebook, and we determined interesting statistics we should display. Several essential libraries were used as listed below.

- Plotly
- NumPy
- Pandas
- Matplotlib

Our data set is divided into two major parts, accident spatial position and the corresponding weather conditions. Therefore, we use attributes from both these parts to plot charts for the following statistics.

- Severity
- Weather Condition
- State
- Total number of accidents

A tree map is used to analyse the accident data and its severity on various states and streets.

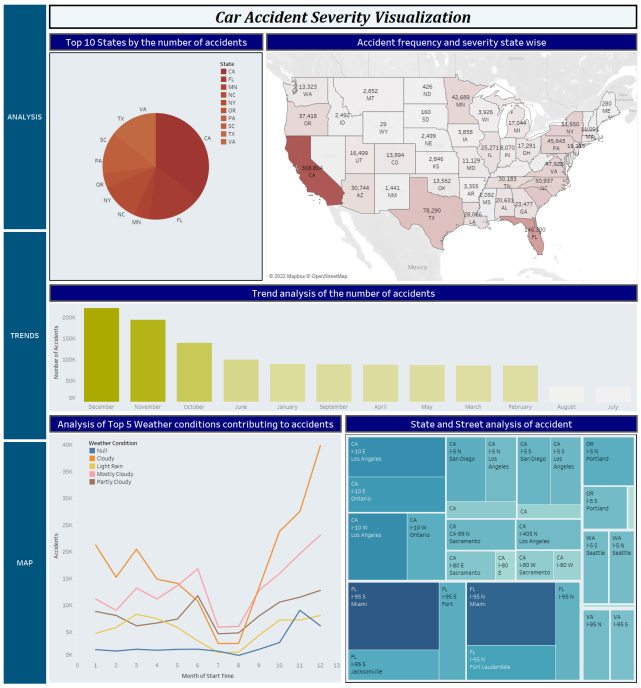

Fig. 5. Home page.

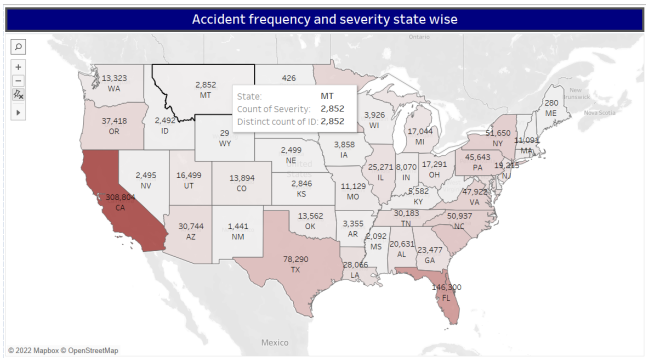

Fig. 6. State wise accident frequency and severity.

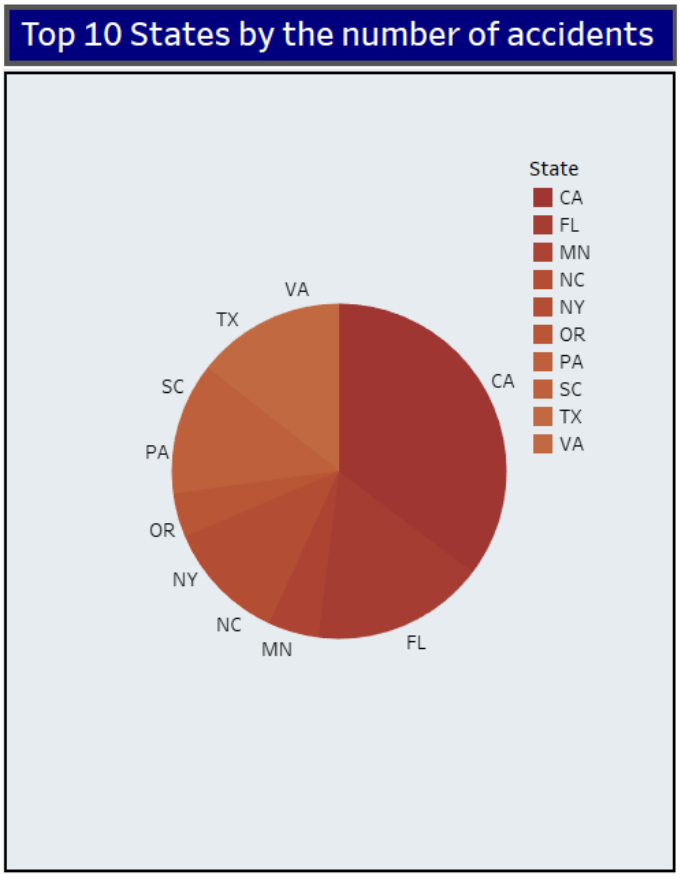

Fig. 7. Top 10 States by the number of accidents.

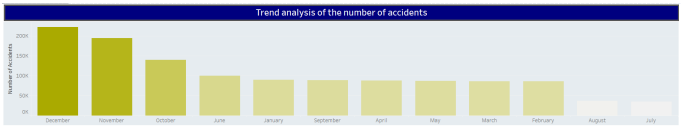

Fig. 8. Trend analysis of number of accidents.

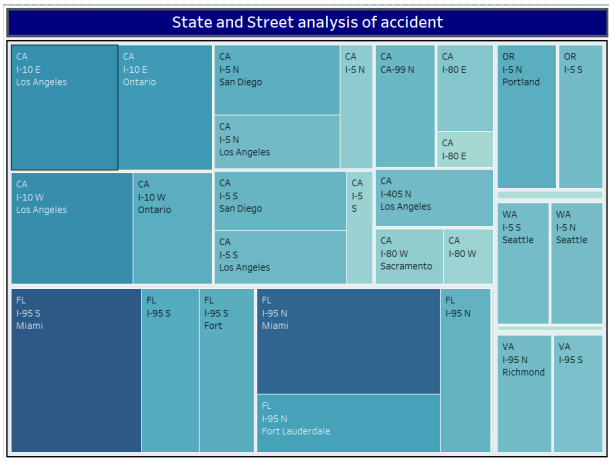

Fig. 9. State and Street Analysis of accidents.

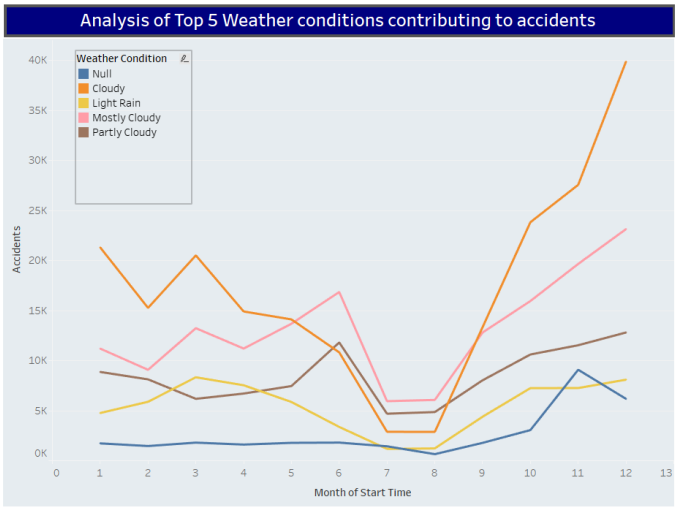

Fig. 10. Analysis of Top 5 weather conditions contributing to accidents.

II. Project Highlights

Our system aims to uncover some essential insights from the data. Some of the most important questions to be answered include:

1) What key factors affect accident severity?
2) What's the impact of different weather conditions or environmental stimuli on severity of accidents in USA?
3) Are there any spatial patterns in terms of area size?
4) What are the accident hotspot locations?

III. Future work

The focus of our current system is on visualizing patterns found in the data and facilitating exploratory analysis. If time permits, we would like to incorporate the following features to further augment the system's functionality as well as usability.

- Prediction: Employ techniques like linear regression, neural networks or time series analysis on the data to help make predictions on accident events in the future.
- User Query Interface: With a user query interface, researchers from different disciplines can apply their own analysis methods to explore the data. The data can be accessed by querying some attributes like the weather condition, location, day of week, etc.
- Feedback: Through a feedback interface we can gather valuable information from different types of users, which can help us to constantly improve user ex- perience, interface layout, and the findings they are interested in.

References

[1] (2022) Road safety facts. [Online]. Available: https://www.asirt.org/safe-travel/road-safety-facts/
[2] Us accidents (updated) - a countrywide traffic accident dataset. [Online]. Available: https://www.kaggle.com/sobhanmoosavi/us-accidents
[3] S. P. Sobhan Moosavi, Mohammad Hossein Samavatian and R. Ramnath. (2019) A countrywide traffic accident dataset. [Online]. Available: https://arxiv.org/pdf/1906.05409.pdf
[4] Flask. [Online]. Available: https://flask.palletsprojects.com/en/2.0.x/
[5] V. O. M. Bostock and J. Heer. (2011) D3 data-driven documents. [Online]. Available: https://doi.org/10. 1109/TVCG.2011.185
[6] Plotly python open source graphing library. [Online]. Available: https://plotly.com/python/