Predictive & Visual Analytics of Car Crashes for Road Safety & Route Planning

Group 29: Aishwarya Shenoy, Ayushi Goyal, Joel Jude, Shiven Barbare, Tyler Burns
Georgia Institute of Technology

1. Introduction and Overview of Methodology

While recent developments in urban transportation, such as Google Maps, have prioritized reducing travel time to estimate route duration across various modes of transportation, they continue to overlook potential safety hazards and accident-prone areas within the city, raising concerns among commuters and policymakers alike.

Our project aims to create a safer transportation system by assessing the severity of car accidents across roads in Georgia, identifying the weather patterns that contribute to this severity, and providing users with the safest route from an origin point to a destination of their choice.

The main components of this project are:
➢ Basic exploration of the data (EDA)
➢ Cluster accident hotspots using based on the latitude and longitude of all roads in Georgia.
➢ Each road is assigned a severity score reflecting the danger level of the road, based on past accident data.
➢ An optimization model using Google Maps API and the severity score data to provide the route with lowest severity score, depending on the user's choice of origin and destination
➢ Visualize both results on a map of Georgia, done on both Tableau and Python.

2. Data Sourcing

➢ All accidents in the United States from 2016-2023, their location, datetime(temporal data), and weather-related metrics taken from a Kaggle dataset that is 3.06 GBs and 7.7 million rows long.
➢ All zip codes with their corresponding latitude and longitude data. This was taken to be a superset of all roads in Georgia, as the accident data does not cover the areas that have had 0 accidents in the past. 4.4 MB, 42,735 rows
➢ Google Maps API to provide routes from point A to B, expressed as latitude + longitude points. Size dependent on start and end points, live route data, size dependent on route choice

3. Exploratory Data Analysis and Severity Score Calculation

Initial attempts at getting a safety score involved logistic regression and other machine-learning techniques. However, this was determined to be the wrong approach. We decided to move forward with a much simpler approach. We got the average severity of the accident on a scale of 1-4 as well as a count of accidents by grouping by each city, street, zip code, weather, weekday, and day or night. Then, we created a severity score that is calculated by multiplying count by the average accident severity and dividing that number by 84, the number of months our dataset covers.

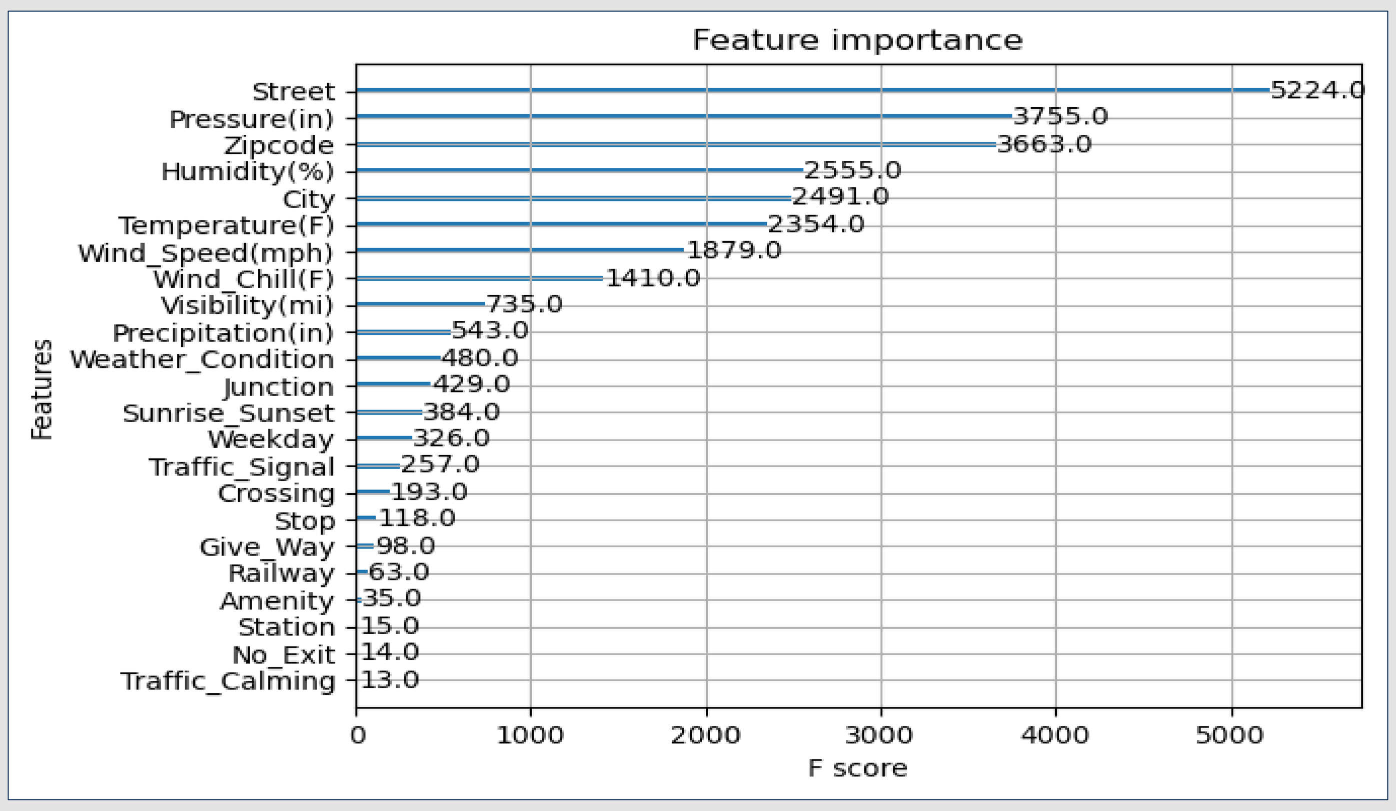
For the visualization in section 5, the severity score is categorized into four levels: ➢
➢ Very Safe (Score <= 2)
➢ Moderately Safe (2 < Score <= 5)
➢ Slightly Dangerous (5 < Score <=50)
➢ Very Dangerous (Score > 50+).

The severity score was leveraged for initial data analysis, as illustrated in the bar charts within Section 5. These charts demonstrate noticeable differences across different weather patterns and times of the day, offering viewers an understanding of how weather conditions may influence accident severity.

4. Predictive Modeling and Experiments

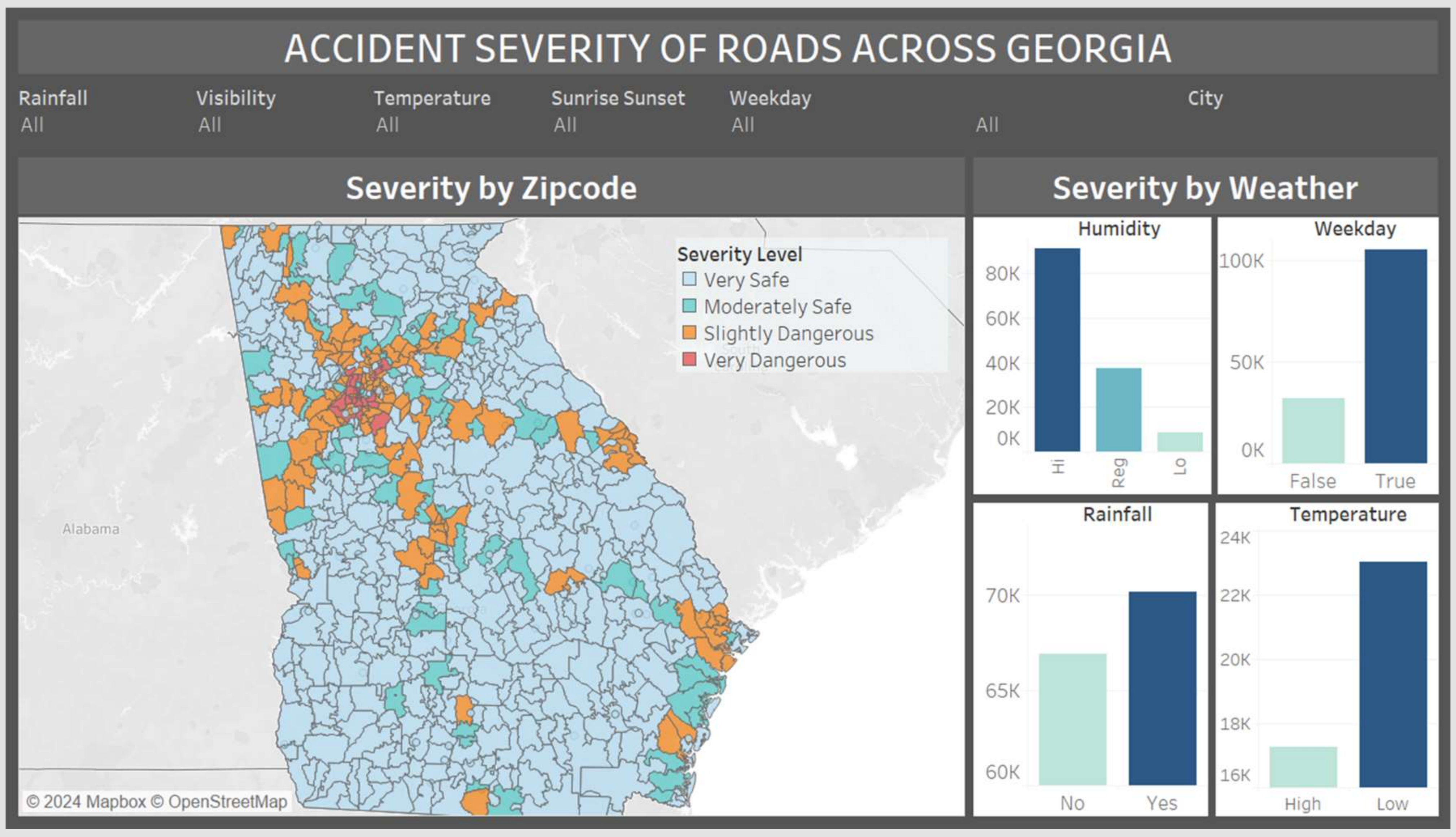
Factor based predictive models were used to predictive severity scores based on location and weather-related factors. Three models were considered: XGBoost, Multi Layer Perceptron, and Random Forests. All models were trained using a standard 80-20 data split and tuned with GridSearch CV.

Experiments involved comparing the tune models based on their performance metrics like Mean Squared Error, and XGBoost was found to be the best model with parameters: {'learning_rate': 0.1, 'max_depth': 6, 'n_estimators': 500}. The relative feature importance was then analyzed based on F score. The results are presented in the graph below:



5. Accident Severity across Zip codes and Weather Patterns

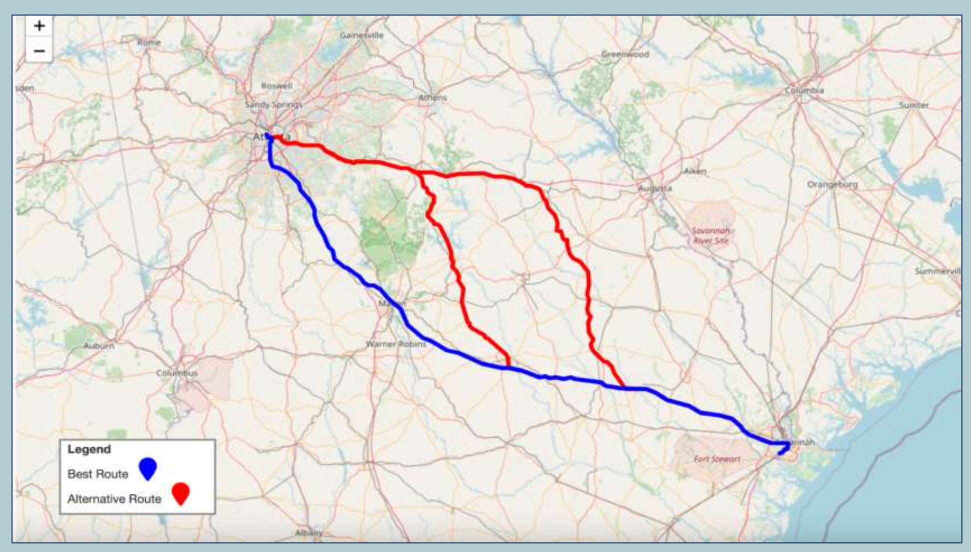
The following interactive visualization was built to analyze the accident "hotspots", by taking the mean severity score per zipcode. The landing page below demonstrates the areas having the most severe accidents, and the user can select varying degrees of temperature, time of day, rainfall, and humidity to see the impact it causes across all areas in Georgia.
There is also a macro-level depiction of weather patterns and time of day/week to see the increase or decrease in severity based on weather patterns.



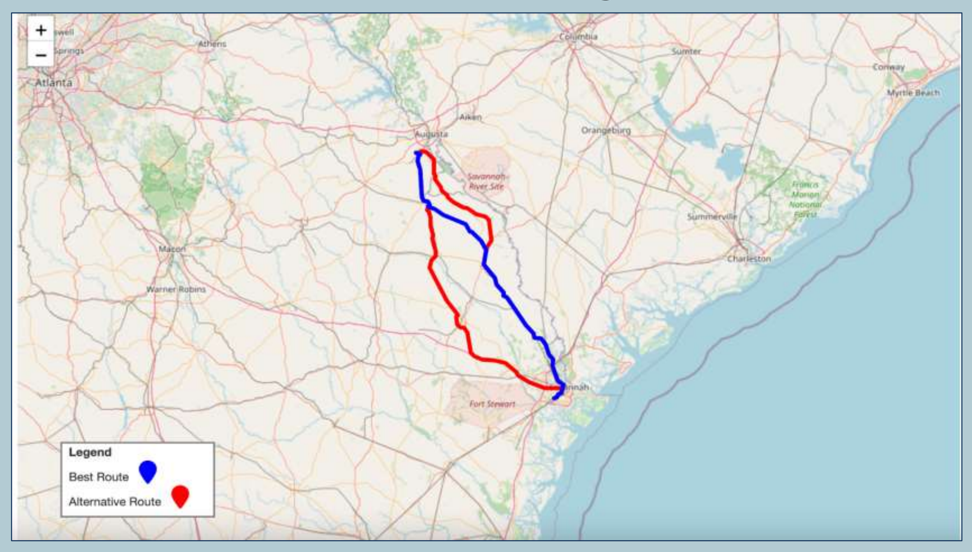
6. Route Planning Optimization Model

This model generates all possible routes given any latitude and longitude and chooses the best route using our desired metric of safety (and nothing else).
➢ Using Google Maps API, we allow the user to input any valid latitude and longitude points in Georgia for generating multiple routes and selecting the safest option.
➢ All alternative routes are computed and displayed on the map City names and severity scores data matrix with latitude and longitude are enhanced using Geopy package.
➢ Severity score is calculated for each step by finding the closest city and multiplying corresponding severity score with step distance.
➢ Used polyline & folium package to map polylines based on cumulative danger calculations.
➢ The safest route is marked in BLUE and less safe ones are marked in RED

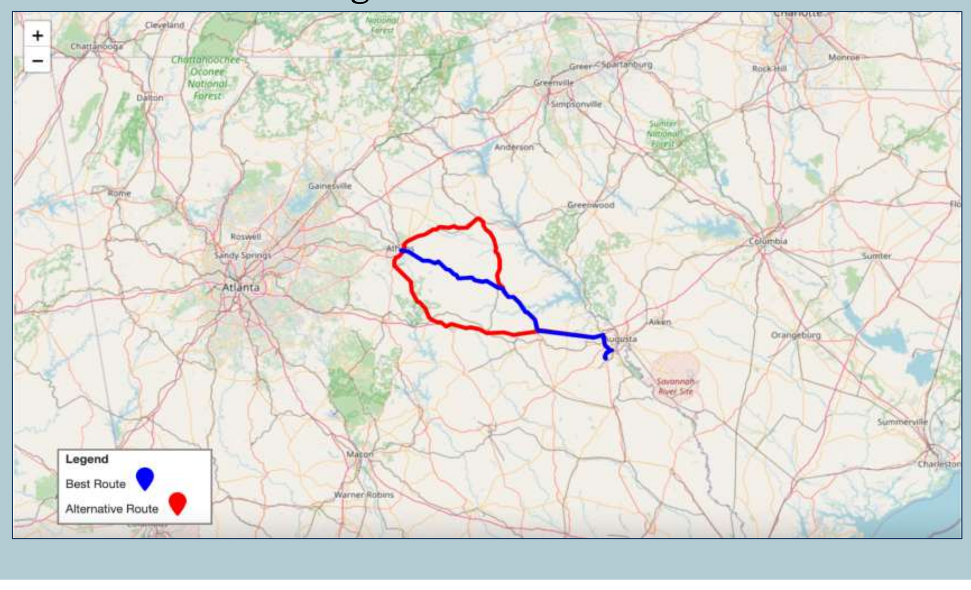
Atlanta to Savannah



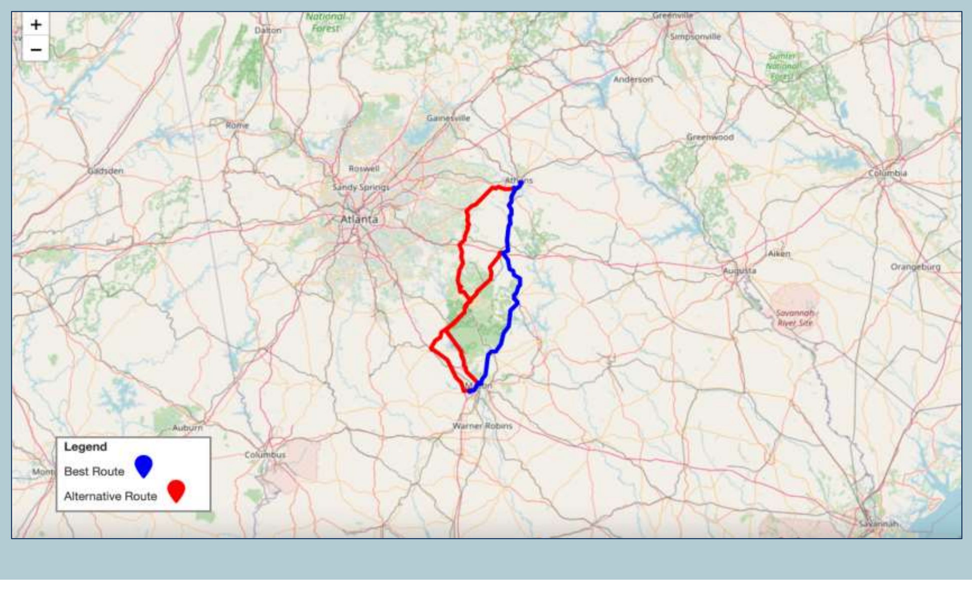
Savannah to Augusta



Augusta to Athens



Athens to Macon-Bibb



7. Results and Conclusions

The results from the Tableau dashboard and the route optimization revealed some interesting insights into road safety that are critical for route navigation:
➢ Routes with high accident severity are concentrated in majorly two areas: (a) Along major interstates where the speed limits are high, and (b) Around metropolitan regions like Atlanta, with high population density
➢ Almost all major accidents occur on weekdays, which is logical since there are much fewer vehicles on road on weekends.
➢ The impact of factors like low temperature and rainfall on crash severity is significant, making them important factors for safe route estimation.
➢ The route optimization model reveals that safest routes are often the fastest ones, demonstrating that more time spent on roads increases the likelihood of a crash.

In conclusion, our dashboard provides a very useful visual analytics tool to study crash severity and its spatio-temporal distribution. The optimization model provides a safer alternative to traditional navigation systems thus improving road safety.