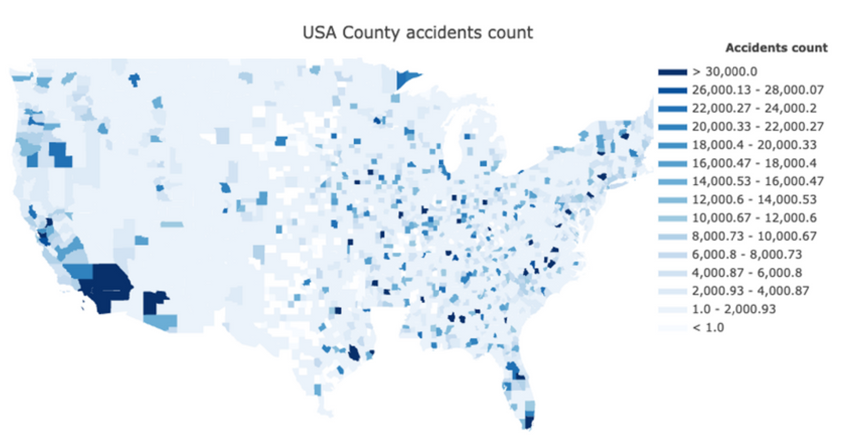
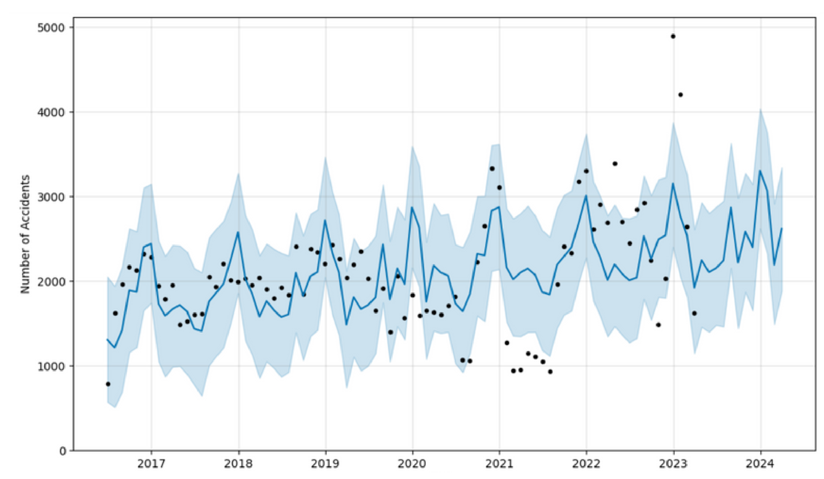
Advancing Road Safety: Innovations in Real-time Monitoring, Predictive Planning, and Community Engagement using Car Accident Data

Team 046: Gabriel Gros, Camille Migozzi, Aubin Rey, Tien Lu, Ting-Yang Kao
Georgia Institute of Technology

Motivation





As the number of car accidents continues to increase in the United States, there's a wealth of data available that could be used to better understand and improve safety on the roads.

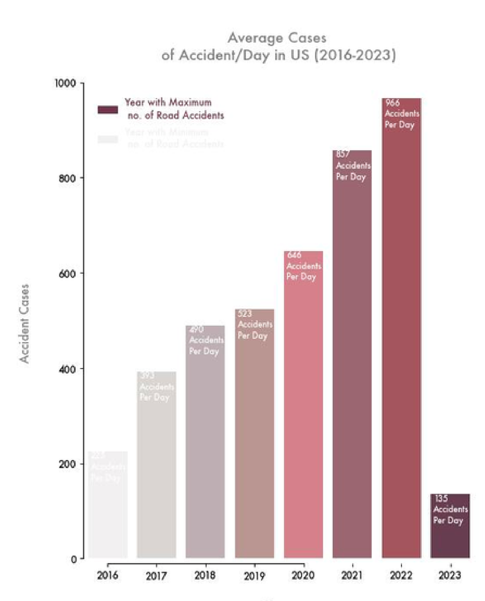
We've been working on a **Real-time Safety Index** in the state of Georgia that combines accident data, weather conditions, time of day, and more to provide up-to-the-minute safety assessments for different routes. This could **help users make informed decisions based on current road conditions**, reducing the risk of accidents. We've been also developing **Predictive Route Planning** using historical data to predict the likelihood of accidents on specific routes. This allows users to choose safer paths, enhancing overall road safety. Additionally, we've developped lots of **dynamic data visualizations** to help users explore accident patterns and safety indices, fostering a deeper understanding of road safety.

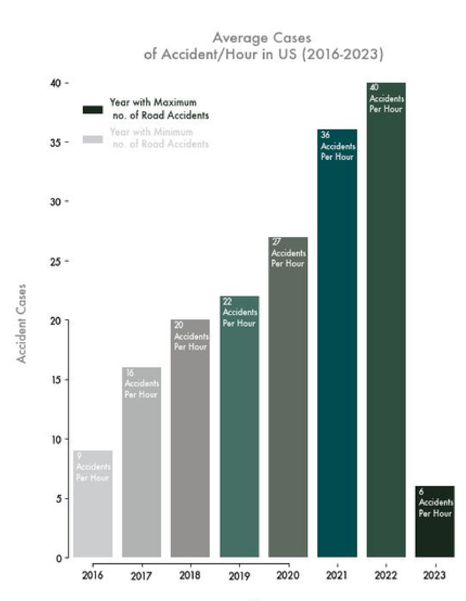
Caring about this project is crucial because it directly addresses the escalating issue of road safety. Our tools empower individuals to make informed decisions based on current conditions, decide safer routes, and contribute to a collective effort in reporting real-time information. **Investing in this project means investing in technologies that have the potential to prevent accidents, save lives, and improve overall road safety.**

The Data

📊For this project we used the **Kaggle dataset of US accidents from 2016 to 2023**. This is a countrywide car accident dataset that covers 49 states of the USA. The dataset contains and detailing geographical, meteorogical and dynamical information about 7.7 million accident records.

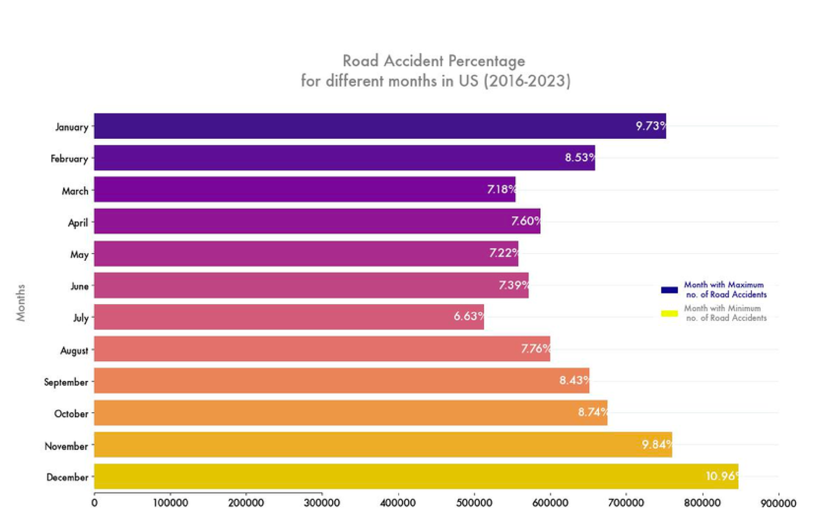
To dive deeper into the analysis of the accident data, there are some static and interactive graph being shown on the webpage. Throughout the graph and plot, we get better understanding on the pattern of the data. Below are some examples.

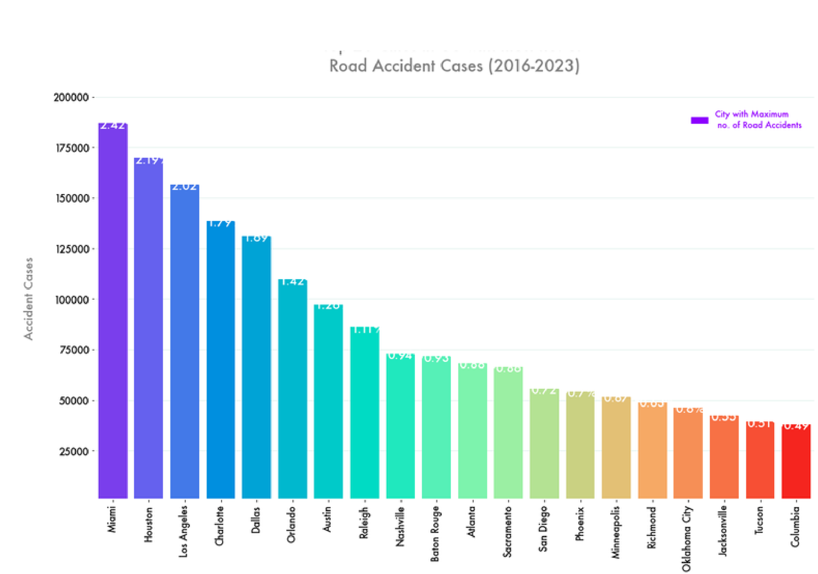




As shown in the left pictures, the total number of accidents happened is obviously increasing. In 2022, the **daily average accident count** is 966 per day. (Note that the data is only updated till March 2023).

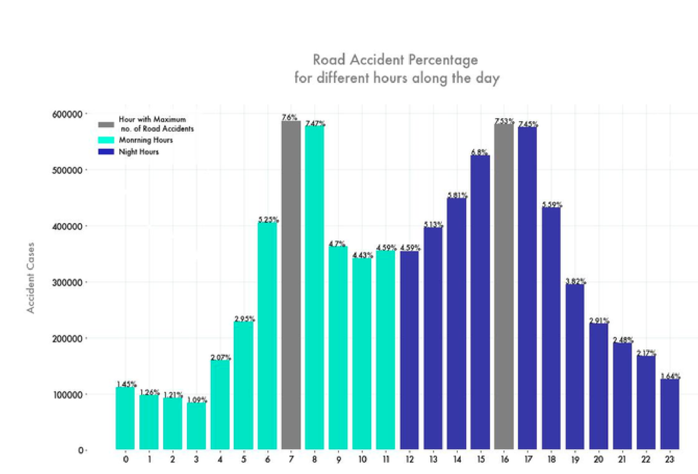
We all hear the that Atlanta is notorious for its transportation, but is it really true? In the graph below, we can see the total count of the accidents on a city level, **Atlanta ranks at no. 11 among the US**. Even take the population into account, it's still not rank in the top 10.

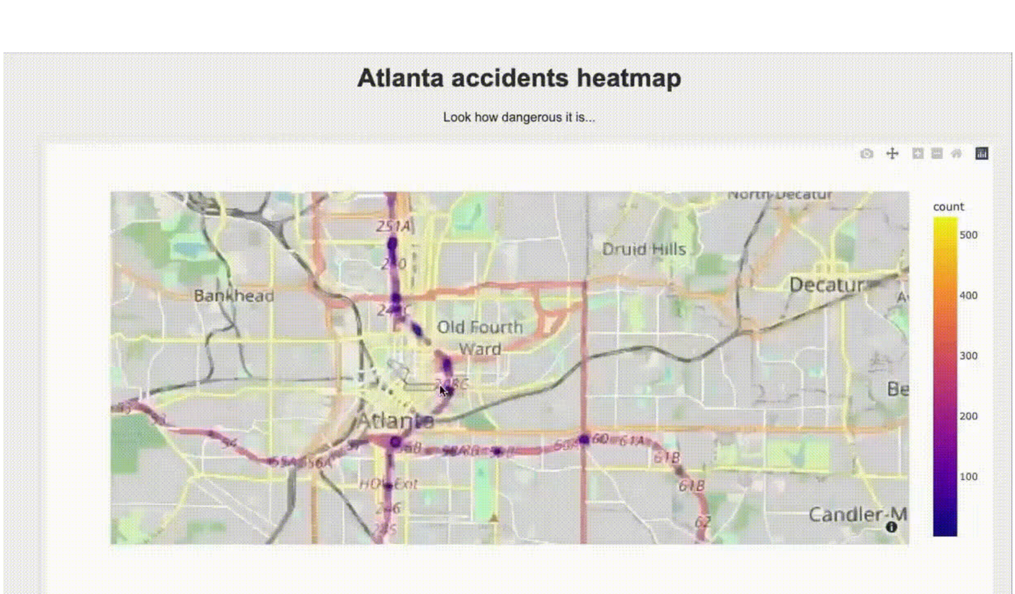




As we are more interested in the time series pattern of the car accidents data, here is a perspective showing the **relationship between the car accidents count and the month** when the acccident happen in the upper-left graph. As shown in the graph, accidents are more likely to happen in winter than other season.

Another perspective of the time series pattern is the **time during the day when the accidents happen**, as shown in the graph, the tint bar shows the accidents happen in the morning (AM), while the blue bar indicates the accidents average in the evening (PM). The grey bar indicates the highest (most) accidents happened during the day, which coinicide with the rush hours (7AM to 8 AM, 4PM to 5 PM).

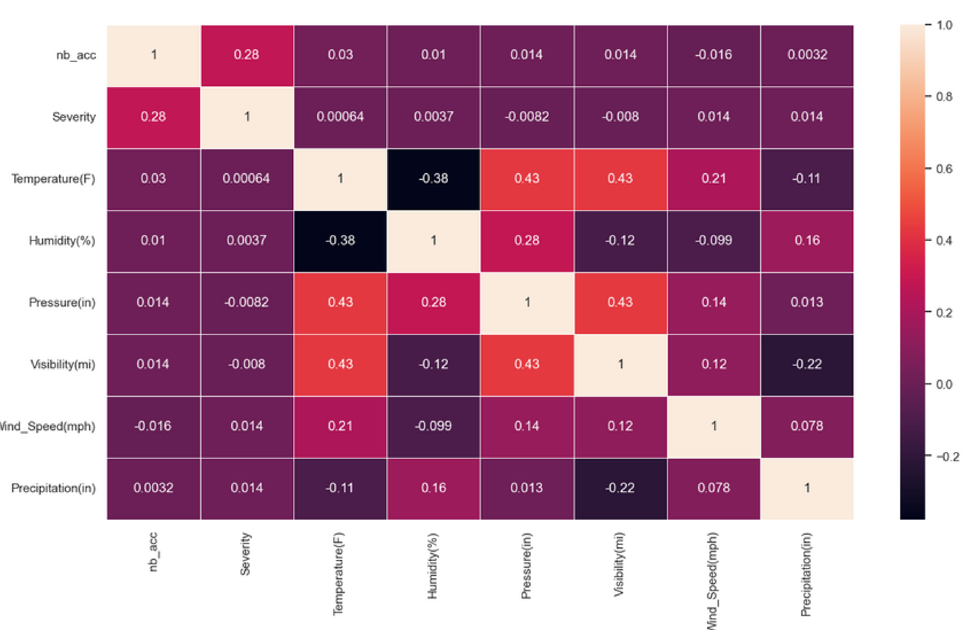




On the left is the interactive graph, this is the demo of the **heatmap of accidents within Atlanta region**. Although static on this document, it is possible to zoom in and zoom out to see the accidents in different region. As we zoom in, we can find that most of the accidents happened on highways. One thing worth noticing is that I-75 is in fact on the top-10 list of the most accidents highways.

Our approach

We first had to create a graph **G = {V,E}** of intersections **V** and road segments **E**, assigning accident locations to the least-squared closest roads. The resulting graph captures the number of accidents per road segment in the state of Georgia. By defining a path planning algorithm, it is enough to modify the weights of the graph (the real distance between each intersection) to obtain a graph taking into account the distance and the dangerousness of this intersection.



To do so, we have developed two approaches:

- A **static hour-based method** to see the influence of daytime on the number of accidents
- A **predictive method** to evaluate the upcoming month number of accidents; it is a robust method that takes into account changes in the number of accidents

For the **static hour-based method**, we referred to the **PASSAGE system** and managed to obtain 24 indexes aggregated on each road for each hour of the day depicted by **x**. Then we normalized the indexes by a customized **sigmoid function** which is depicted in the equation on the right, with its coefficients α, β, and γ set to 2, 0.21, and -0.5.

$$SI_{history} = f(x) = \alpha\left(\frac{1}{1+e^{-\beta x}}+\gamma\right)$$

For the **predictive method** chose to use **Facebook Prophet**, a very popular time series forecasting algorithm. It additively expresses a time series as shown on the right with trend **g(t)**, seasonnality **s(t)**, noise component **ε(t)** and holiday effect **h(t)**.

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t)$$

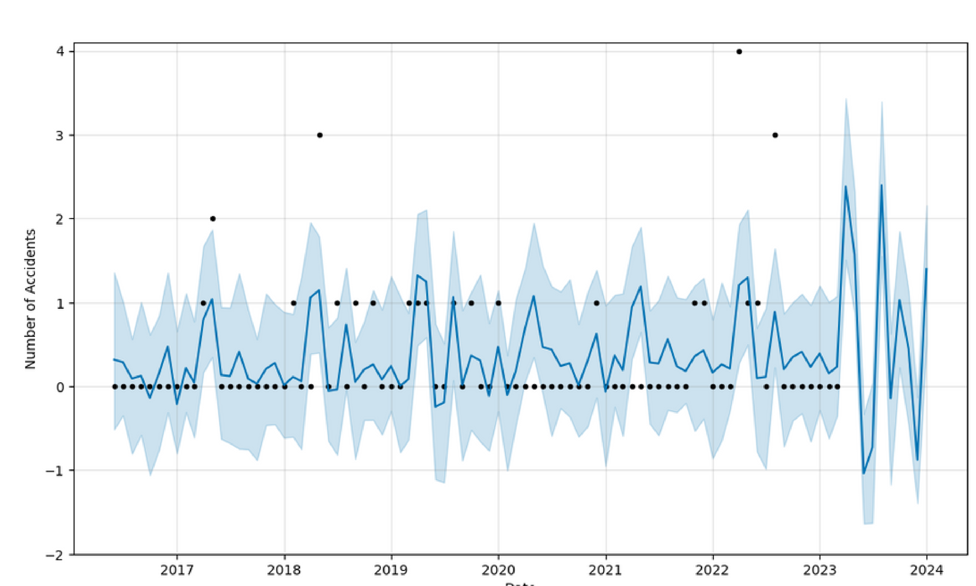
To prevent unnecessary complexity in the analytical framework, we computed a **correlation matrix**. It showed the limited correlation of accidents number with weather variables such as Temperature, Humidity or Pressure. Thus, we prioritized the accident count as the primary feature for both statistical and predictive safety index modeling.

The expected number of accidents for the currrent month is then computed for each road segment **e** and the new weight of each edge can be expressed as follows:

$$f'(\omega_e) = (1 + \mathbb{E}_e[\hat{y}|X])^p\,\omega_e$$

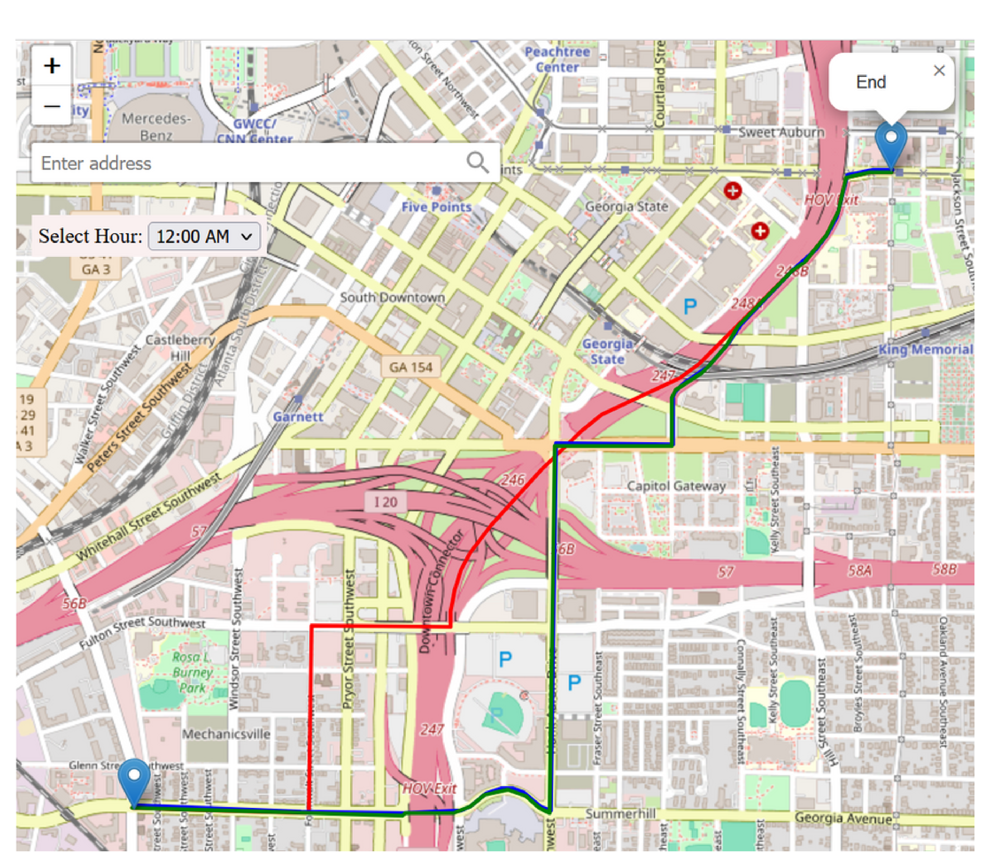
We conducted a comprehensive study to discern notable but not excessive changes during the update step. We opted for **p = 2**. Finally, we opted for **Dijkstra's algorithm** to find the shortest path between two points in the updated graph.

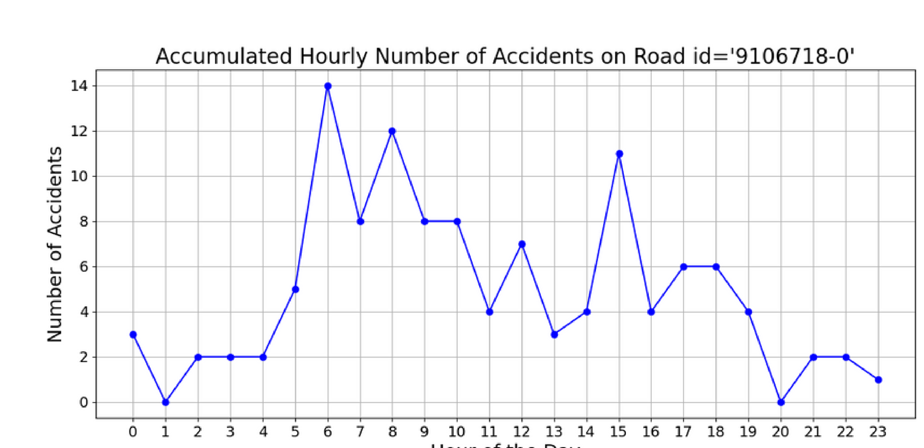
Experiments and results

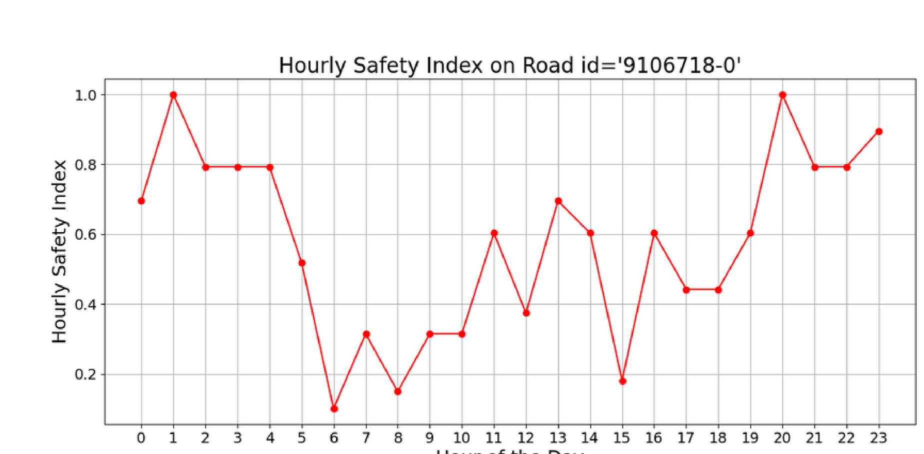


Using FBProphet, we can see on the figure on the left a **prediction** over the next month on a given route, roving its powerful time series analysis capabilities.

The two line charts below show an example of the safety index generated by the **static hour-based method**. The blue one shows an example of the accident counts on a road in Atlanta, and the red one shows the corresponding hourly safety indexes on the same road.







By utilizing our web application, a traveler has the flexibility to designate a **departure point** and an **arrival point** either by clicking directly on the map or by employing the search bar for a specific address. Subsequently, three distinct itineraries are generated using **Dijkstra's algorithm**, each based on a different cost function:

- **Fastest Itinerary** (red) : This itinerary prioritizes the fastest travel time. The travel time of a given edge is computed by multiplying the edge's length with the speed limit corresponding to its road type.
- **Safest Itinerary using the Static Hour based Method** (blue) : This itinerary minimizes accidents within a one-hour time frame. The normalized accident occurrence, ranging between 0 and 1, is multiplied by the edge's travel time, rewarding routes with lower danger.
- **The Safest Itinerary using Future Accident Predictions** (green) : This itinerary incorporates a second safety index, predicting accidents for the next month. The index corresponds to an adjusted distance, which is then multiplied by the speed limit of the road.

Results reveal that the fastest itinerary prioritizes motorways for expedited travel, while both safe itineraries strategically seek alternatives to mitigate the higher accident rates associated with such roads. Another notable observation is that the **Safest Itinerary using the Static Hour based Method** consistently produces the same route as the **Safest Itinerary using Future Accident Predictions**. This pattern remains unchanged irrespective of the selected hour of travel. One possible explanation could be a limitation in the available data, where the number of accidents for each edge at a given hour might be relatively low. To enhance precision, a more accurate metric, beyond solely the number of accidents per road, would involve evaluating the frequency of accidents per person-hour specific to each road.