CSE 6242 Final Report: Traffic Route Safety Visualization

Team 28: Christian Lee, Charles (Charlie) Liu, My Duyen (Tina) Nguyen, Borun (Patrick) Song

1. Introduction and Problem Definition

Current routing systems prioritize travel time without adequately considering safety metrics such as historical collision data and traffic patterns, limiting users' ability to make informed, risk-aware route choices. This gap in explainability often leads to increased frustration for commuters, and prevents traffic planners from addressing critical safety concerns effectively. The objective of this project is to develop an interactive application that integrates real-time and historical data on traffic, collisions, and routing patterns to model, analyze, and visualize safety and speed statistics for routes in Georgia. We aim to calculate how traffic patterns are affected by various types of accidents and how it affects commute, and visualize concrete safety and speed statistics of different routes from origin to destination.

2. Background and Related Work

People often blindly follow their navigation app's route, even if they know the route by heart. This is because the route can take into account real-time location and crash data. However, time is not the only attribute to consider when determining the optimal, safest routes during commute. You may also want to consider the accident history of certain paths on your route, the severity and frequency, and can learn to adjust accordingly to risks. Pi et. al. stressed the importance of identifying the root causes of traffic congestion. They developed a system that utilized data visualization to showcase the traffic signal system, its bottlenecks, and road capacity. In combination, by examining past documentation, the researchers determined the root causes of traffic congestion and accidents in specific streets at specific times. Our project shares the same goal, to showcase the history of each route to tell the story to not only city planners, but the general public. Building these charts, highlighting specific routes, and attaching an interactive graph of Georgia routes would be useful to show how often traffic congestion happens, the timing, and the area on the road where most accidents happen.

Several methodologies for predicting traffic flow have been explored in the past. Abadi et. al. used dynamic traffic simulation and autoregressive models to predict traffic flow in San Francisco, while Lv et. al. explore applying deep learning methodologies in the form of stacked autoencoders for the same purpose. As we work to predict traffic patterns in Georgia, these papers offer key insights regarding handling issues with limited traffic data and learning latent traffic features. Li et. al. go a step further and integrate auxiliary data sources on weather and time with historical traffic data to better capture the impact of external factors on traffic. Likewise, we plan to include historical crash data and location information to enrich our traffic datasets.

We are also interested in assessing the risk of a road to better understand how accidents affect traffic. Retallack and Ostendorf outlined previous efforts on the correlation between traffic volume and accidents and concluded that more recent research shows a concave relationship. Milton takes a step further by isolating other influencing factors for accidents and finds negative binomial distribution to be the best model. To better adjust such models, Xu et. al used clustering technique and concludes that there are three types of traffic state, each of them showing a different correlation. We plan to start our investigation on finding a good model that could better predict the risk of accidents.

Overall, applications such as Google Maps and Apple Maps are often a black box when it comes to how optimal routes are being selected. Considering many applications do not account for the route safety and traffic volume, we plan to bridge this gap for Georgia drivers to bring awareness and promote more conscious driving habits. Researchers have also used this data combination to get an in-depth understanding of a city's traffic history. The research completed from Shaadan et. al. and Sakib et. al. showcased figures describing the relationship between accident cases by injury, road conditions, time of day, vehicle type and more, that would be informative for future urban planning. The researchers used multiple figures to map out the information, but adding one interactive graph blending all the information could be useful to describe the story. For our own interactive visualization, we plan to select similar relationships to provide a historical overview of traffic conditions for all major routes in Georgia.

Incorporating more insights on each of the possible routes and what makes one route better than the rest also improves user trust and reliability of routing applications.

An annotated bibliography can be found in the Appendix section below.

3. Innovations

The application combines vital information from both the collision dataset from Georgia Department of Transportation (GDOT) with TomTom's routing patterns. Using Javascript and D3.js, we will retrieve the top 3 or 5 routes between any two locations from the interactive map of metro Atlanta. Many similar applications usually only focus on the top route at a time.

Statistics will be generated and historical trends on traffic flow, collisions, and overall safety will be visualized for each of the routes, providing a more holistic approach for commuters to determine the best route. Manap et. al has shown that using buffer analysis we can locate accidents and accident hotspots that happen along a route. In our implementation, each route will have a 24-hour distribution of the safety of that route and the speed of that route. Rakha & Van Aerde and Borucka et. al shows that there are significant differences in accident frequency between both time-of-day and day of the week. Stretch goals are to give advanced safety statistics of these routes. Crash visualizations will be proportional to traffic volume to normalize roads.

This can have real impact - state and federal departments of transportation could use our statistics to put real-time specific warning signs along the part of the road that was identified as "dangerous" for drivers' safety. It could also be used to find specific pockets of time where certain roads are performing poorly, helping design roads better. Of course, this can also be used by drivers themselves, to understand their commute better, reducing frustration and increasing driver awareness.

4. Methods

4.1 Data Pipeline and Preprocessing

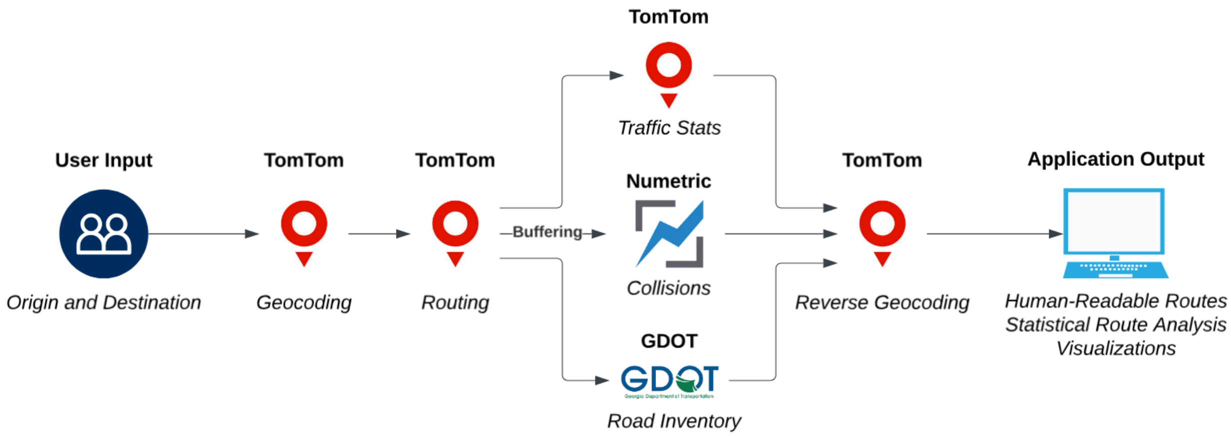
The end-to-end flow of data in our project is as follows: first, the user passes in an origin and destination location as human-readable addresses. Using the TomTom Geocoding API, the addresses are transformed into latitude and longitude coordinates. These coordinates are used as inputs for the TomTom Routing API, which takes in origin and destination coordinates and calculates time-optimized routes between the two locations. The routes are returned in the form of an array of coordinates. The routes are joined with traffic flow data, accident data, and roadway infrastructure data using the latitude and longitude coordinate pairs as keys.

Traffic flow data is acquired through the TomTom Traffic Stats API's route analysis functionality. This API provides historical traffic data by hourly granularity. It takes in coordinate-defined routes as input for generating route statistics, and is therefore easily compatible with the Routing API's output.

Numetric vehicle collision data is provided by AASHTOWare, and contains latitude and longitude coordinates for most collisions. The dataset consists of 3.8 million collisions in Georgia from 2014 to 2023. The dataset must be cleaned and validated for records that have non-null latitude and longitude values, and are spatially and temporally relevant to the scope of our application. In order to reconcile collision data with route information, we must develop a method to query for collision data based on the generated routes. To achieve this, we apply buffer analysis to retrieve all collisions within a specified latitude-longitude radius of the routing coordinates. All collisions within these radii are considered relevant to the route.

Lastly, the Georgia Department of Transportation provides road inventory and annualized traffic data in the form of geodatabases. Each record is identified by a specific road segment, with latitude and longitude values. A similar buffering algorithm can be applied once again to connect routes with GDOT data.

While latitude and longitude coordinates are useful keys to use for reconciling and associating various data sources, they are not user-friendly to interface with due to their lack of readability. To resolve this, we call on the TomTom Reverse Geocoding API to transform coordinates into addresses and road names. This allows us to better format and visualize the final output of the project. The figure below outlines the architecture of the data pipeline as discussed:



4.2 Calls for Route Finding and Traffic Flow of Route

From the form on the FrontEnd, we received the origin location and destination location in the form of an address. We use TomTom's API to convert the address to coordinates. Using TomTom's Routing API and Javascript, we get the top 5 routes recommended to be used as a starting pool. To get the accident information of each segment, we query for accident information based on segments within a route. The route coordinates are then fed into the HERE Map API to display on the map. The route is represented as an array of coordinates and we will query safety history based on the location passed along the route.

We also made request calls for Flow Segment Data to obtain real-time information about the flow of speeds and travel times within specified coordinates. Combined with annual average daily traffic information from GDOT from previous years, we get a more holistic view of the area's traffic conditions to help assess route recommendations based on the users' prioritization of travel time and safety.

4.3 Accidents on Route

For GDOT's collision data and AADT dataset, we populated PostGIS. PostGIS allows for easy storage and efficient query when we run safety analysis of a given route. We populated our database based on the accident history. With route information as input, we performed buffer analysis to find any accidents that are within the route's area. Buffer analysis allows for us to detect accidents that occurred along the area of the coordinates of the route.

Populating PostGIS was tricky, with manual input of 21 fields and guessing the correct data types and with inconsistent column header formatting (white space, capitals, spaces at the end of columns but not others, etc.). This ended up being the bulk of our data cleaning.

Fortunately, the AADT dataset from GDOT includes detailed geometries for road segments monitored by traffic meters. Each traffic meter corresponds to an individual road segment, complete with specific attributes and geometric data for visualization. For instance, a road segment like Spring Street NW between 14th Street and 10th Street is represented as a single unit, making it straightforward to break down long roads into smaller, more manageable segments for analysis. This segmentation is crucial because accidents occurring far from the designated route may not provide meaningful insights. By focusing on localized segments, we ensure the relevance and precision of our safety analysis.

Once accidents were able to be found, it was also then crucial to write functions grouping this by hour, by day, and by specifically road segments. These functions were then used in both the map visualization, route recommendation, and hourly accident chart.

4.4 Statistical Analysis and Recommendation

In order to determine the most optimal route, we designed a route index based on metrics for traffic and safety. However, we first needed to determine how to prioritize advantages in traffic and speed against improvements in safety. We conducted a survey among drivers on their priorities when it comes to their commutes. 39 out of 245 respondents (~16%) indicated that they would be strongly willing to increase their travel times if it meant a safer commute. Thus, to reflect the population's preferences in the prioritization of travel time and safety, we compute the route index using a weighted average of indices for traffic and safety, with coefficients of 0.8 and 0.2, respectively.

The traffic index is computed based on two terms, the first being travel time. In order to compare relative route efficiency, we normalize this metric by dividing the fastest travel time by each route's travel time. Thus, the fastest route would have a "travel time" of 1, while a route that takes twice as long would have a "travel time" of 0.5. This metric would have a linear correlation with the traffic index. The next factor is the relative traffic, defined by the percent increase in travel time caused by traffic congestion. Since delays of a few minutes are very likely along longer routes and often negligible, low relative traffic only slightly punishes the traffic index. However, we want high relative traffic times to have a much more significant impact on the traffic index, in order to minimize driver frustration. Thus, we use a logarithmic term to capture this relationship, $1 + log_{10}(1 - (\beta \cdot Relative\,Traffic))$, where β controls the steepness of the traffic penalty. β is dynamically adjusted based on the fraction of the length of the route that is affected by traffic. The final traffic index is computed by averaging the linear travel time term with the logarithmic traffic term:

$$I_{Traffic} = AVG(\frac{Fastest\,Travel\,Time}{Route\,Travel\,Time},\ 1 + log_{10}(1 - \beta \cdot Relative\,Traffic)),$$
$$\text{where } \beta = 1 + \frac{Length\,of\,Traffic}{Length\,of\,Route}$$

We assess the safety of certain roads by aggregating the number of accidents that took place around a given location. Since the accident includes the exact coordinate of the incident, we only consider accidents close to the coordinates that we are interested in. In other words, instead of counting the number of accidents along the entire road, we segment it so that only accidents that are close enough to be relevant are taken into consideration. From there, the safety index is computed based on three metrics: the likelihood of a collision occurring, the likelihood of a collision being fatal, and whether the current time of day is more or less dangerous for the route. First, the likelihood of a collision occurring is calculated based on the number of crashes divided by the total annual average daily traffic (AADT) through the route. The average likelihood across the whole collision dataset is also calculated, and the percent difference between the two is mapped to a value between -1 and 1 using a tanh function as follows: $tanh(\frac{Collision\,Likelihood - Route\,Likelihood}{Collision\,Likelihood})$. This term is negative when the likelihood of a collision on the route is higher than the average, and vice versa. The likelihood of a crash being fatal is calculated and compared in the same way against the National Highway Traffic Safety Administration's statistic of 0.7% of collisions being fatal. Finally, the number of collisions that have occurred on the route in the current hour is compared to the expected number of collisions on the route per hour, again with the same methods. These three terms, each ranging from -1 to 1, are averaged to construct the safety index:

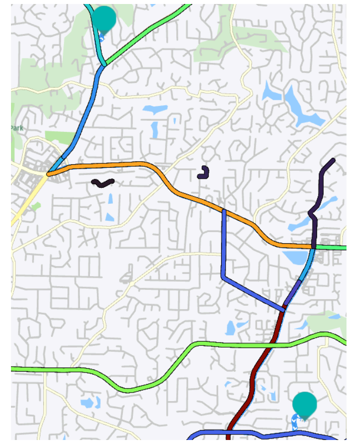
$$I_{Safety} = AVG(tanh(\frac{Collision\,Likelihood - Route\,Likelihood}{Collision\,Likelihood},\ tanh(\frac{0.7 - Fatal\,Percentage}{0.7}),\ tanh(\frac{E - Hourly\,Collisions}{E})),$$
$$\text{where } E = \frac{Total\,Route\,Collisions}{24}$$

Thus, the final equation for our route index is as follows:

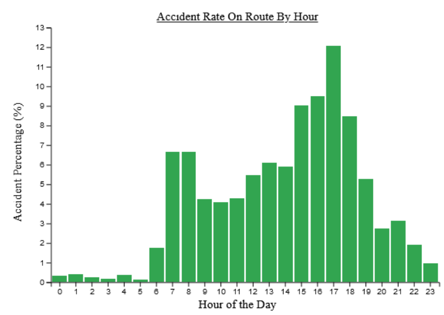
$$Route\,Index = 0.8 * I_{Traffic} + 0.2 * I_{Safety}$$

4.5 Visualization

Our main visualization involves a map of the best route between points A and B. This best route is calculated with numerous factors like speed, safety, and general flow speed. Overlaid on this route are the accident road segments as defined by the Georgia Department of Transportation (GDOT). These road segments are colored based on accident density, showing a driver where along their route where they should watch out for. This also can help city planners very quickly pinpoint where there are accident-prone road designs, and how to alleviate these areas.



Additionally, we created a visualization of the distribution of accidents in general for a road over the course of 24 hours. This allows users to understand at which times during the day it's relatively safer to take a route than others. For example, if one were to look at the best time to take their commute to work, this could aid in understanding at what hours between 7 - 10 they should head in. While there is a strong correlation between traffic volume and accident rate, this still gives a bit more specific information that could be insightful.



5. Impact Measurement, Experiments, and Evaluation

If successful, this project can educate people on traffic patterns in their area to make more informed decisions on their daily commute, and provide transparency for those who are frustrated on their daily commutes by helping people understand the potential root causes of traffic, accidents, and road abnormalities. Success can be measured by analyzing real-world use cases, and comparing routes found from those examples by various applications. Progress will be measured through deliverables at midterm and final checkpoints. By the midterm, a route-finding algorithm should be developed to identify crashes along optimal routes. These results will be evaluated against current industry-standard routing algorithms such as Google Maps and Waze. A minimum viable product for the end of the semester would involve displaying visualizations of routes, and visualizations of 24-hour safety and speed distributions.

For the project evaluation, we invited classmates currently enrolled in this class to test our product. We reached out to 32 different peers. Each person was introduced to the project and were provided with an explanation on how to use the system. Each person was then asked to use the application for 5 minutes. We asked users to navigate routes that they are familiar with, verbally explaining their thought process, opinions, and ideas along the way. After testing the system, users were given a survey asking how close our routing result was compared to their impression about that route if they are familiar with it. The survey also asked about how the subject usually navigates around Georgia, what tools are used, and their overall experience with navigating apps. If applicable, we also asked them to compare the system to the application they currently use.

A majority of the users put in routes to their hometowns or Georgia Tech as those were the addresses they knew at the top of their heads. The results showed that 28 subjects explicitly mentioned their surprise at the accident visualization as some of the routes that the users took in their own neighborhoods were more dangerous in comparison to the average. For many, this was an easy fix as they knew of multiple other routes from point A to point B. Some subjects said that the visualization also confirmed their suspicions about which roads were dangerous. For example, one road segment from a subject's house to their grandparent's house was more dangerous because of a yielding left turn. The subject sometimes took that route because it saved them an extra minute, but after seeing the visualization, they now know that they can take another route to avoid that left turn.

The survey results showed that the users valued the information for our project and believed it would be useful to use in conjunction with the other routing systems they normally use. The most popular ones were Waze, Google Maps, and Apple Maps. It was found useful for those with driving anxiety. Ten subjects reported having driving anxiety, especially in small cities and new areas. While they all valued efficient routing, just having the information of how safe or unsafe some roads may be made them feel more at ease. Two subjects noted that it would be very useful when trying to help their younger family members learn how to drive. The map allowed them to evaluate what familiar roads were safer for inexperienced drivers and plan a quick route in their heads accordingly. The current applications lack this visual information as all focus on getting a person most efficiently from point A to point B, avoiding traffic and road construction. Our project, however, provides an informative visual that promotes more mindful driving. Rather than just following the map, users are able to get a more holistic idea of the safety of their driving routes. The con for our map was that it can only be used for pre-planning and is not as effective as the other applications when the users do not have time to digest the visualization information.

GDOT Crash Data has charts visualizing crash reports for different Georgia counties, but it is difficult to plan a course of action with the information.

6. Conclusions and Discussion

SafeMaps is a superset of current navigation apps like Google Maps. While Maps has historically been the go-to app with more information about speed, SafeMaps also includes safety in our recommendations. Users are able to be more mindful of their safety while driving, and cities are able to visualize and make key decisions in infrastructure with our application. Additionally, accident frequency is highly correlated with increased traffic time in general. Understanding the relationship between accident frequency and travel speed of a given road, we believe, should lower driver frustration. For example, understanding that taking the highway is both a high-speed form of travel but also prone to accident backups by looking at our map visualization can create the correct expectation for a route.

At the same time, our current solution is far from perfect. We could create a more interactive experience that allows users to dive deeper or zoom out to get micro or macro details. The visualizations could still be cleaned up a bit, especially in how we query for road segments for a given route. Many road segments, which just slightly graze the route are included. In the future, we continue to hone our query to be more specific to the route and also only include the bits of the road segments that touch the route. We also want to better visualize our recommendation route and the thought process behind it. If there was a better way to display our statistical analysis of a route other than text, that would be a great step forward.

A common way of saying something isn't dangerous is saying, "Oh, you're scared of doing X? Hey, driving is more dangerous! You're not complaining about that, are you?" The premise here is that driving isn't thought of as dangerous, but it is, more than one thinks. And we believe our application is one step in making society more aware about what driving is really like.

Contribution Statement: *All team members have contributed a similar amount of effort.*

References

A. Abadi, T. Rajabioun and P. A. Ioannou, "Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 653-662, April 2015, doi: 10.1109/TITS.2014.2337238.

Borucka, A., Kozłowski, E., Oleszczuk, P. & Świderski, A. (2021). Predictive analysis of the impact of the time of day on road accidents in Poland. *Open Engineering*, *11*(1), 142-150. https://doi.org/10.1515/eng-2021-0017

Li B, Xiong J, Wan F, Wang C, Wang D. Incorporating Multivariate Auxiliary Information for Traffic Prediction on Highways. *Sensors*. 2023; 23(7):3631. https://doi.org/10.3390/s23073631 (http://dx.doi.org/10.3390/s23073631)

Manap N, Borhan MN, Yazid MRM, Hambali MKA, Rohan A. Identification of Hotspot Segments with a Risk of Heavy-Vehicle Accidents Based on Spatial Analysis at Controlled-Access Highway. *Sustainability*. 2021; 13(3):1487. https://doi.org/10.3390/su13031487

Milton, J., Mannering, F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* 25, 395–413 (1998). https://doi.org/10.1023/A:1005095725001

M. Pi, H. Yeon, H. Son and Y. Jang, "Visual Cause Analytics for Traffic Congestion," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2186-2201, 1 March 2021, doi: 10.1109/TVCG.2019.2940580.

Sakib, A., Ismail, S.A., Sarkan, H., Azmi, A., Mohd Yusop, O. (2019). Analyzing Traffic Accident and Casualty Trend Using Data Visualization. In: Saeed, F., Gazem, N., Mohammed, F., Busalim, A. (eds) Recent Trends in Data Science and Soft Computing. IRICT 2018. Advances in Intelligent Systems and Computing, vol 843. Springer, Cham. https://doi.org/10.1007/978-3-319-99007-1_9

Shaadan, N., Azhar Suhaimi, M. I., Hazmir, M. I., & Hamzah, E. N. (2021). Road accidents analytics with data visualization: A case study in Shah Alam Malaysia. *Journal of Physics: Conference Series*, *1988*(1), 012043. https://doi.org/10.1088/1742-6596/1988/1/012043 (https://iopscience.iop.org/article/10.1088/1742-6596/1988/1/012043/pdf)

Rakha, H., & Van Aerde, M. (1995). Statistical analysis of day-to-day variations in real-time traffic flow data. *Transportation research record*, 26-34.

Retallack, A. E., & Ostendorf, B. (2019). Current Understanding of the Effects of Congestion on Traffic Accidents. International journal of environmental research and public health, 16(18), 3400. https://doi.org/10.3390/ijerph16183400

Xu, C., Liu, P., Wang, W., & Li, Z. (2012). Evaluation of the impacts of traffic states on crash risks on freeways. *Accident Analysis & Prevention*, *47*, 162–171. https://doi.org/10.1016/j.aap.2012.01.020

Y. Lv, Y. Duan, W. Kang, Z. Li and F. -Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865-873, April 2015, doi: 10.1109/TITS.2014.2345663. (http://dx.doi.org/10.1109/TITS.2014.2345663)

Appendix

A. Abadi, T. Rajabioun and P. A. Ioannou, "Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 653-662, April 2015, doi: 10.1109/TITS.2014.2337238.

The paper tackles challenges in traffic flow prediction models, such as a lack of data due to limited sensors, and the prevalence of unforeseeable events such as crashes and road closures. To address these issues, a dynamic traffic simulator was used to connect sparse traffic data, and an autoregressive prediction model was trained to predict traffic flow in San Francisco with relatively high accuracy. We can apply similar models and methodologies in building a traffic prediction tool for the streets of Atlanta, and improve on the model infrastructure by bringing in auxiliary data sources to further remediate limited data availability and gauge the likelihood of formerly unpredictable events.

Y. Lv, Y. Duan, W. Kang, Z. Li and F. -Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865-873, April 2015, doi: 10.1109/TITS.2014.2345663. (http://dx.doi.org/10.1109/TITS.2014.2345663)

This paper is an early exploration in the potential of applying deep learning models to big transportation datasets, for the purposes of predicting traffic flow. The researchers trained a stacked autoencoder model on two months of traffic data from the Caltrans Performance Measurement System database. We could potentially apply similar deep learning methodologies to capture latent relationships between traffic data features that might otherwise be unrepresented.

Li B, Xiong J, Wan F, Wang C, Wang D. Incorporating Multivariate Auxiliary Information for Traffic Prediction on Highways. *Sensors*. 2023; 23(7):3631. https://doi.org/10.3390/s23073631 (http://dx.doi.org/10.3390/s23073631)

Most traffic flow prediction models are built on historical traffic data alone. However, this paper expands the scope of traffic prediction by incorporating supplemental data sources (e.g. weather data) to better predict their impact on traffic events such as crashes and road closures. In order to combine data sources, the researchers used a multi-horizon bidirectional long short-term memory model. Given our project will also involve multiple data sources, we can consider similar techniques as displayed in this paper, while further exploring options for auxiliary data sources.

Shaadan, N., Azhar Suhaimi, M. I., Hazmir, M. I., & Hamzah, E. N. (2021). Road accidents analytics with data visualization: A case study in Shah Alam Malaysia. *Journal of Physics: Conference Series*, *1988*(1), 012043. https://doi.org/10.1088/1742-6596/1988/1/012043
(https://iopscience.iop.org/article/10.1088/1742-6596/1988/1/012043/pdf)

The paper identifies the root causes of the accidents to further Malaysia's future development of their road design and policies. To generate an effective overview of the scene, the researchers split their visualizations into inferential analytics and descriptive analytics. The visualizations provide a guideline and ideas into the type of information that would be useful to depict for the average Atlanta driver. Specifically, creating an interactive visualization describing road accident timing patterns may be useful to showcase to Atlanta drivers the times where most accidents are recorded. The researchers used multiple figures to map out the information, but adding one interactive graph blending all the information could be useful to describe the story.

Sakib, A., Ismail, S.A., Sarkan, H., Azmi, A., Mohd Yusop, O. (2019). Analyzing Traffic Accident and Casualty Trend Using Data Visualization. In: Saeed, F., Gazem, N., Mohammed, F., Busalim, A. (eds) Recent Trends in Data Science and Soft Computing. IRICT 2018. Advances in Intelligent Systems and Computing, vol 843. Springer, Cham. https://doi.org/10.1007/978-3-319-99007-1_9

The paper provides an analysis behind the popular types of U.K. traffic accidents by examining previous works on this topic. The researchers also visualized the accident trends to describe the information found. The paper's methods on pre-processing, filtering, and linking the dataset into a data model acts as a baseline into how to properly organize the dataset to be used for interactive visualizations. The results from the visualization revealed that the majority of the accidents happened in dry, urban areas under daylight and in low speed limits which defies intuition and thus inspires further research to identify the root causes.

M. Pi, H. Yeon, H. Son and Y. Jang, "Visual Cause Analytics for Traffic Congestion," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2186-2201, 1 March 2021, doi: 10.1109/TVCG.2019.2940580.

The paper suggests a novel method that aims to identify the causes of traffic congestion based on traffic flow theory and develops a visual analytics system. The researchers identified that traffic congestion analysis can utilize not only average speed, but entropy from the information theory to classify flow changes in congestion zones. The paper tested the system on traffic data on four cases and determined, through visualizations, that traffic signal system, bottlenecks, and reaching road capacity caused traffic congestion. For each case, the visualizations highlighted the routes and traffic direction in various cities around the world that cause traffic congestion and mapped the times of congestion with the number of cars using a line chart. Building these charts into an interactive graph of Atlanta routes would be useful to show how often traffic congestion happens, the timing, and the area on the road where most accidents happen.

Borucka, A., Kozłowski, E., Oleszczuk, P. & Świderski, A. (2021). Predictive analysis of the impact of the time of day on road accidents in Poland. *Open Engineering*, *11*(1), 142-150. https://doi.org/10.1515/eng-2021-0017

Through statistical methods like Kruskal-Wallis and Kolmogorov-Smirnov tests, this paper was able to establish a strong correlation between accidents and time of day and seasonality. However, it did note that not accounting for other factors negatively affected their models, so there are still other critical factors to what causes an accident.

Rakha, H., & Van Aerde, M. (1995). Statistical analysis of day-to-day variations in real-time traffic flow data. *Transportation research record*, 26-34.

This paper shows traffic patterns stay relatively regular during the weekdays, but have increased deviations on the weekends. Major incidents, like accidents and road hazards or big events in town, lead to drastically different times where real-time data could be needed. Although we can't predict real-time deviations, we want to maximize a user's ability to understand the deviations of their route on weekdays and on weekends.

Manap N, Borhan MN, Yazid MRM, Hambali MKA, Rohan A. Identification of Hotspot Segments with a Risk of Heavy-Vehicle Accidents Based on Spatial Analysis at Controlled-Access Highway. *Sustainability*. 2021; 13(3):1487. https://doi.org/10.3390/su13031487

This paper identified accident prone areas in highways with buffer analysis technique. Using the same technique, we could identify where there are accident hotspots along routes.

Retallack, A. E., & Ostendorf, B. (2019). Current Understanding of the Effects of Congestion on Traffic Accidents. International journal of environmental research and public health, 16(18), 3400. https://doi.org/10.3390/ijerph16183400

This paper outlines previous efforts on understanding the correlation between traffic volume and accidents. It is interesting to note that different independent research in different years and locations shows slightly different results. For example, several earlier surveys show a concave relationship while several more recent ones show a convex trend. The location, time of day, and layout of highway affect the result, and thus we attempt to visualize accident risk at different locations and time of day to better reflect the reality.

Xu, C., Liu, P., Wang, W., & Li, Z. (2012). Evaluation of the impacts of traffic states on crash risks on freeways. *Accident Analysis & Prevention*, *47*, 162–171. https://doi.org/10.1016/j.aap.2012.01.020

This paper suggests that the state of traffic affects the risk as well. The authors used a clustering approach to separate the traffic flow pattern into three different states to assess the risk under that specific pattern and they obtained three different correlations between traffic speed and accident.

Milton, J., Mannering, F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* 25, 395–413 (1998). https://doi.org/10.1023/A:1005095725001

The authors identify negative binomial distribution to be the best model to describe the correlation between traffic and accidents while isolating other factors such as geometry of the highway. Their approach could be used as the base model for general cases, and we could adjust it to better reflect edge cases.