REINFORCEMENT LEARNING BASED ROUTE AND STOP PLANNING FOR AN AUTONOMOUS VEHICLE SHUTTLE SERVICE IN AN URBAN CITY

&

ACCIDENT RISK LEVEL PREDICTION FOR INDIVIDUALS

A Technical Report Submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia – Charlottesville, Virginia

In the Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Vismita Uppalli
Fall, 2020

Technical Project Team Member

Shohaib Mahmud
University of Virginia

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: ______________________________ Date 12/1/2020
Vismita Uppalli

Approved: ______________________________ Date 12/1/2020
Haiying Shen, Department of Engineering and Society

Introduction

This technical paper consists of the work of two main problems completed over the course of two semesters. The main idea of the first problem, *Reinforcement Learning Based Route and Stop Planning for Autonomous Vehicle Shuttle Service in an Urban City*, is to determine an efficient route and stops for an autonomous vehicle (AV) shuttle service in an urban city. Public transportation is quite prevalent in major cities such as New York City or San Francisco, as it can be expensive for the city residents to buy and maintain their own private car. One study shows that there are many factors in these cities, such as traffic, the distance from their residence to other areas, and economic factors that influence a lower number of car ownership and higher amount of mobility needs in these areas, (Potoglou and Kanaroglou, 2008). In such scenarios, it can be useful to provide public transportation to residents to travel from one place to another. Furthermore, an autonomous vehicle shuttle service can be used to take the residents along the most efficient route based on human mobility patterns of the city.

The adoption of autonomous vehicles to a city can bring many advantages. Autonomous vehicles, also known as driverless vehicles, use a fully automated driving system that results in the vehicle responding to external conditions the way human drivers do. One great advantage of this technology includes a potential increase in safety for drivers, passengers, and pedestrians. The software used in these vehicles is less likely to make as many errors as humans would on the road, resulting in a decrease of casualties caused by vehicle crashes. Another advantage of autonomous vehicles includes the ability for many kinds of people to drive, including those that may be disabled or elderly. Since the vehicles take control over the drive, it's easier for anybody to take the wheel of an autonomous vehicle. Lastly, automated cars may eliminate driver fatigue and allow some to sleep during overnight journeys, ("What is an Autonomous Vehicle?", 2020).

With these advantages, it is easy to see that the arrival of automated cars can affect a society drastically. To investigate the issue of determining an efficient route for an autonomous vehicle service, I have used a human mobility dataset that collected data from users' cellphone GPS sensors in Richmond. Using this data, I have observed the various routes the users traveled, the speed in which they moved around, a variety of other factors to strategically plan the best places the AV shuttle should stop and take residents around the urban city.

The second problem I worked on, *Accident Risk Level Prediction for Individuals*, includes observing the patterns of accidents in Florida and determining the factors that cause these accidents. This can later be used to group users and situations into various risk categories in relation to the decisions they make when driving. Research shows that there were 3,133 traffic fatalities in 2,915 crashes in Florida 2018, (Car Accident Statistics in Florida for 2018 & 2019, 2020), a surprisingly significant number for just one year. Additionally, Florida has one of the highest traffic death rates at 14.7 per 100,000 residents, (Car Accident Statistics in Florida for 2018 & 2019, 2020), presenting the need for further analysis on the causation of these accidents and the amount of risk residents are taking when traveling on the road. These fatalities are not only harmful for the drivers and passengers, but also pedestrians, where 22% of the mentioned fatalities were suffered by walkers, (Car Accident Statistics in Florida for 2018 & 2019, 2020). This problem was investigated by analyzing a US Accidents dataset which consisted of accidents in the entire country of the first half of the year 2020. Various factors such as the outside temperature, visibility, weather, and wind speed were observed. Other factors such as the place of the accident were analyzed as well.

<u>Part 1: Autonomous vehicle shuttle service route and stop planning</u>

As previously stated, public transportation is widely used in urban cities. Using an autonomous vehicle as a shuttle service can prove advantageous for many reasons, such as improving safety and eliminating driver fatigue. Additionally, it can alleviate public transportation drivers of a potentially tedious job of traveling along the same routes every single day. For an AV shuttle service to prove affective in a city, the route it goes along should be well suited to the mobility patterns of the area's residents. Factors such as the usual routes of the population, parking deficiencies present in the city, the number potential public transportation (PT) users vs. car-owners, points of interests present in the area, locations of user residences, and crowd positions were all analyzed to determine the best places for an AV shuttle to stop in a route across the city.

How do autonomous vehicles work?

To effectively plan an autonomous vehicle shuttle service route for an urban city, the inner mechanisms of how they work must be thoroughly explored. First, a driver sets a destination and the car's software calculates a route. Autonomous vehicles contain a rotating, roof-mounted LIDAR sensor that monitors a 60-meter range around the car, (Ondruš, Kolla, Vertaľ, & Šarić, 2020). Consequently, this creates a dynamic 3-dimensional map of the car's current environment. The vehicles also contain sensors that detect its place relative to the 3D map and calculate distances between obstacles, (Ondruš, Kolla, Vertaľ, & Šarić, 2020). Additionally, there is an override function available for humans to take control of the vehicle.

A significant technology that autonomous cars use is Machine Learning (ML) and Deep Neural Networks (DNNs). A key component of an autonomous vehicle, as mentioned before, is a perception module which senses its surroundings. This module is controlled by an underlying

DNN which takes inputs from different sensors to maneuver the car safely under the current conditions. These sensors include radar, laser light, GPS, odometers, and computer vision to interpret sensory information, identify appropriate navigation paths, and calculate any obstacles. LIDAR, Light Detection and Ranging, is a remote sensing technology that measures distance by illuminating a target with a light beam and analyzing reflected light, (Ondruš, Kolla, Vertaľ, & Šarić, 2020). Radar, Radio Detection and Ranging, can estimate the velocity of objects using electro-magnetic waves. Ultrasonic sensors and video cameras are used to detect objects near the mirror on all sides of the vehicles. Lastly, GPS, Global Positioning System, is a space-based satellite navigation system that provides current location and time information from an unobstructed line of sight from satellites, (Ondruš, Kolla, Vertaľ, & Šarić, 2020). A typical DNN is composed of multiple layers stacked together to extract various representations of the input from the information provided by the sensors (Tian, Pei, Jana, & Ray, 2018). For example, the first few layers of an autonomous car DNN can extract basic information such as stop signs or other cars, whereas the final layer calculates decisions such as the direction to steer, (Tian, Pei, Jana, & Ray, 2018). Each layer of a DNN consists of computing units called, neurons, which are connected together to form each layer and send outputs to one another, (Tian, Pei, Jana, & Ray, 2018).

It is important to note that despite tremendous progress with DNNs and autonomous cars, the software can demonstrate incorrect or unexpected corner-case behaviors that can lead to dangerous consequences, (Tian, Pei, Jana, & Ray, 2018). There have even been several real-world cases where crashes under rare previously seen cases occur. For example, a fatal Tesla crash resulted from a failure to detect a white truck against the bright sky, (Tian, Pei, Jana, & Ray, 2018). These erroneous behaviors can be fixed through bug detection and a patching cycle;

however, this is a very challenging problem. Large companies like Google and Tesla have indicated so as well, despite having already deployed machine learning techniques in production environments, (Tian, Pei, Jana, & Ray, 2018).

AV Shuttle Route and Stop Planning

To determine an AV shuttle service route and stops, the mobility of an urban city must be thoroughly observed. For this research, a human mobility dataset in Richmond, Virginia from January 2019 to March 2019 was analyzed. This dataset contains GPS location data tracked from the phones of approximately 36,000 unique users when using certain applications. The most significant information of the dataset consists of a User ID, their location, the speed at which they are traveling at, and the time when the information was extracted. The remainder of this section highlights the ways in which this dataset was analyzed to create some key observations.

Determining user's routes

As mentioned before, the mobility dataset consists of user locations, and times they visited the particular location. When the locations for each user are sorted by the time, they visited them on a particular day, the route that the user traveled throughout that particular day can be observed. This data can be used to find the popular routes in which several users travel and determine the places users start out their day. Furthermore, these routes can be mapped to find overlaps and popular routes the residents take throughout the city. Therefore, AV shuttle routes and stops can be modeled off of the popular routes and places users visit throughout their day.

Vehicle speeds

City residents can be categorized into two groups: car owners and public transportation (PT) users. To determine if a user owns a car or not, the speed at which they travel from their residence can be calculated. If the user is traveling over 20 mph from their home, they are most likely driving a car. If not, they may be walking or biking to a nearby PT stop. Figure 1 shows a 3D plot of car speeds, and Figure 2 shows a 3D plot of bus speeds throughout Richmond.

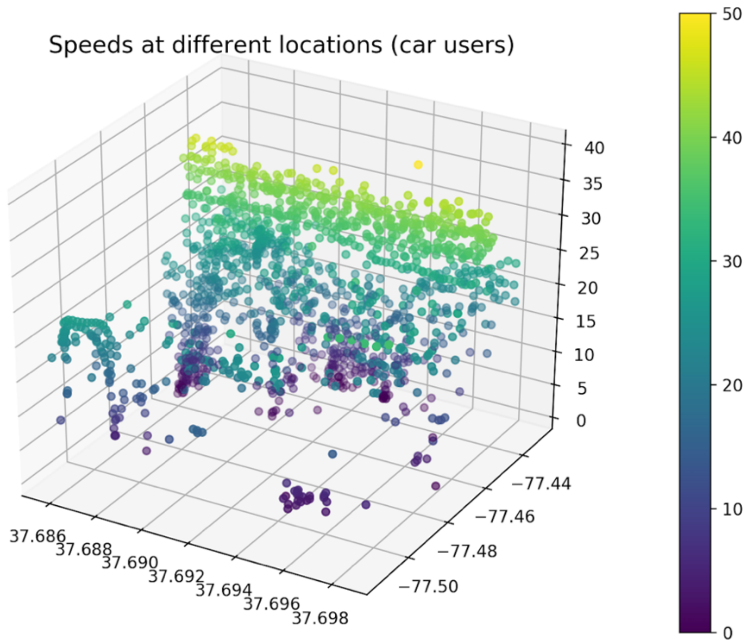

Figure 1: Speed of car users throughout Richmond.

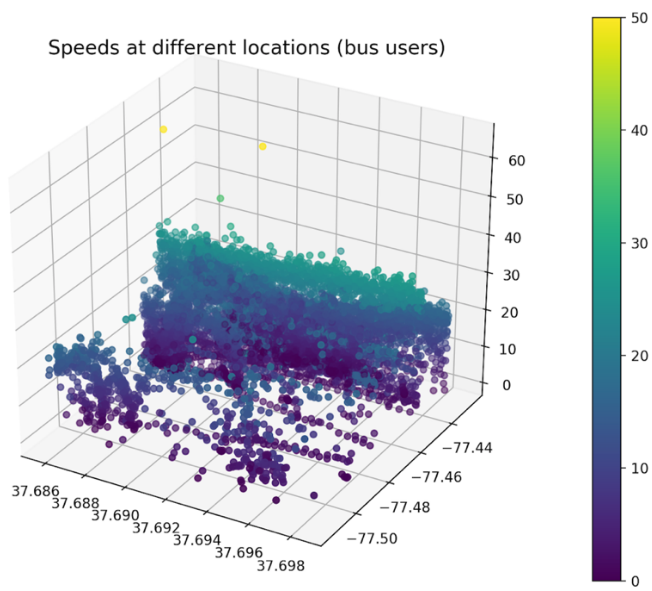

Figure 2: Speed of bus users throughout Richmond.

When considering the routes of an AV shuttle service, PT users should be specifically targeted as they will be the ones to use the service. Additionally, observing vehicle speeds throughout the city of Richmond can provide an idea of how fast the AV shuttle should travel, and a rough outlook of the speed at which users can get from one place to another.

User residences

When analyzing user locations between 12 AM to 7 AM for several days, their place of residence can be extracted. To view this data easily, the area of Richmond was divided into 100 grids. After this, the place of residence for each user was categorized into one of the grids. Figure 3 shows a graph of the population per grid, when Richmond is divided into 100 grids.

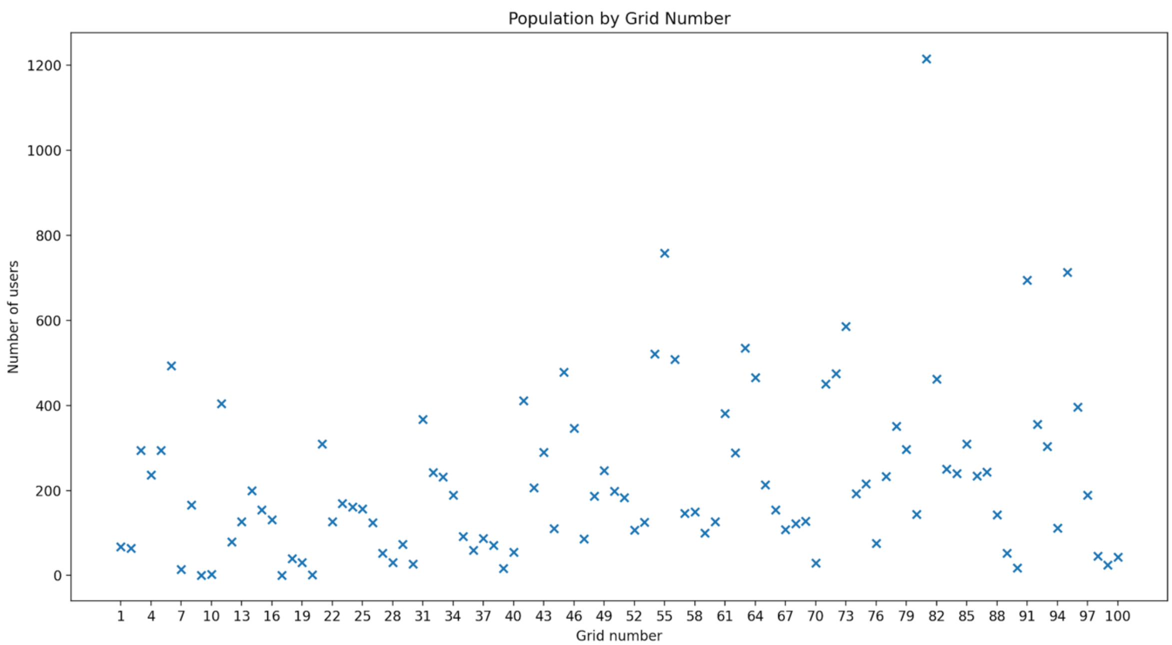

Figure 3: The population of Richmond per grid number when divided into 100 grids.

It can be beneficial to determine residences of users for multiple reasons, the main one being to target certain individual’s homes for AV shuttle stops. Keeping the shuttle service stops near the residences of users who use public transportation, can help them get to their destination faster and easier.

Point of interests

Various points of interests (POI's) throughout Richmond were extracted and observed from external data. The mobility dataset was used to create an intermediate dataset for each user who has visited a POI on a certain day. Analyzing the points of interests (POI's) throughout Richmond can provide a good idea of the most popular places city residents visit throughout the day. An AV shuttle service's route should consist of the most popular POI's and arrive at these places during the times when passenger frequency would be high.

Crowds

The various crowds present throughout Richmond can be extracted by tracking users who are near each other during the same time. If the number of users exceeds 10, they can be considered a crowd. Figure 4 shows a heatmap of the area of Richmond where darker spots represent a greater number of crowds for that area.

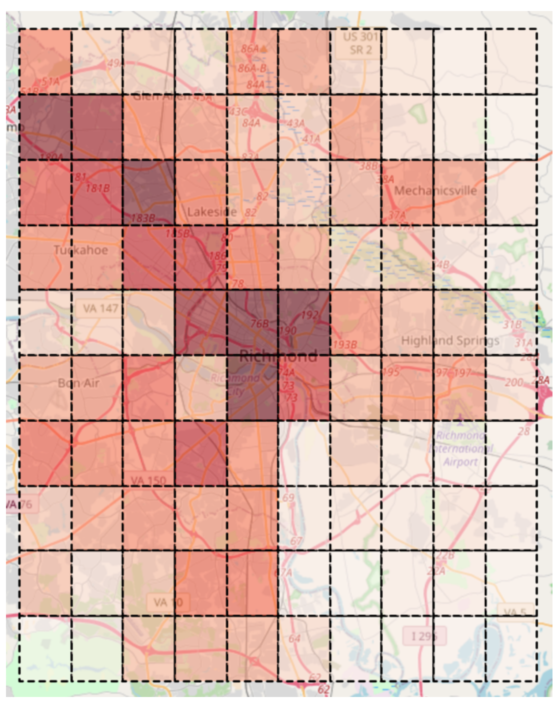

Figure 4: A heatmap of the number of crowds present during the day throughout Richmond.

By analyzing the various crowds present throughout the city, the areas with most people present throughout the day can be pinpointed. The AV shuttle can then travel in areas with high crowds to ensure a high passenger frequency.

Parking deficiency

Another goal of the autonomous vehicle shuttle service is to lessen the parking burdens of city residents. To analyze the areas of the city in which residents struggle to find parking, first, the number of parking spots around Richmond was determined using a website (https://www.parkme.com). Figure 5 shows the number of parking spots available for each grid in Richmond, when divided into 100 grids.

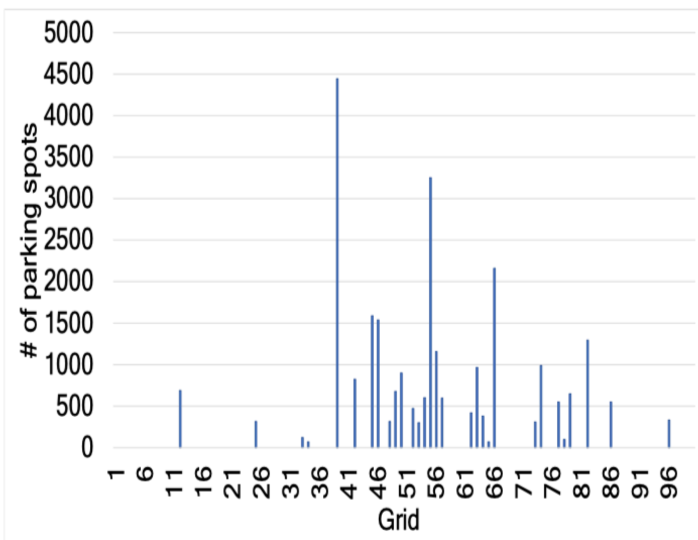

Figure 5: The number of parking spots available for each area in Richmond when divided into 100 grids.

Then, the number parking spots residents needed throughout Richmond was calculated by counting the number of car owners who stayed in an area other than their home for at least a couple of hours. Lastly, the number of parking spots needed was subtracted from the ones available to determine the parking deficiency in each area of Richmond. Figure 6 shows the parking deficiency in Richmond, when divided into 100 grids.

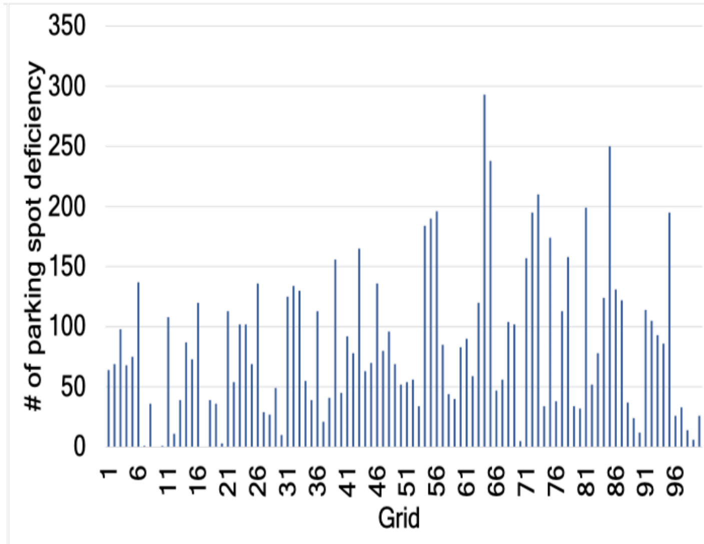

Figure 6: The number of parking spot deficiencies per each area in Richmond when divided into 100 grids.

The figure above shows that there is a significant amount of parking deficiencies throughout Richmond and some specific areas that lack a lot more parking spots than others. Therefore, the autonomous vehicle shuttle should target these areas, potentially lessening the parking woes of some city residents.

Part 2: Accident risk level prediction

It's well known that far too many traffic accidents occur throughout the United States every year. The US is one of the busiest countries when it comes to road traffic with almost 280 million vehicles and more than 225 million drivers holding a valid driver's license, (Wagner, 2020). There are many factors that influence the accident distribution in the United States, including external influences such as weather and internal reasons such as an individual's thought process. This research investigates all of these factors to categorize individuals into various accident risk levels. It can be beneficial to have a method to accurately predict these risk levels so that the main factors that influence these traffic accidents can be identified and avoided. To perform this research, a US accident dataset from June 2020 was analyzed. This dataset includes information such as the day/time, location, and severity of the accident.

First, the accident distribution of Florida was closely analyzed. To do this, a plot comparing the percentage of locations to the number of accidents was generated. Figure 7 shows this plot with the number of accidents capped at 50.

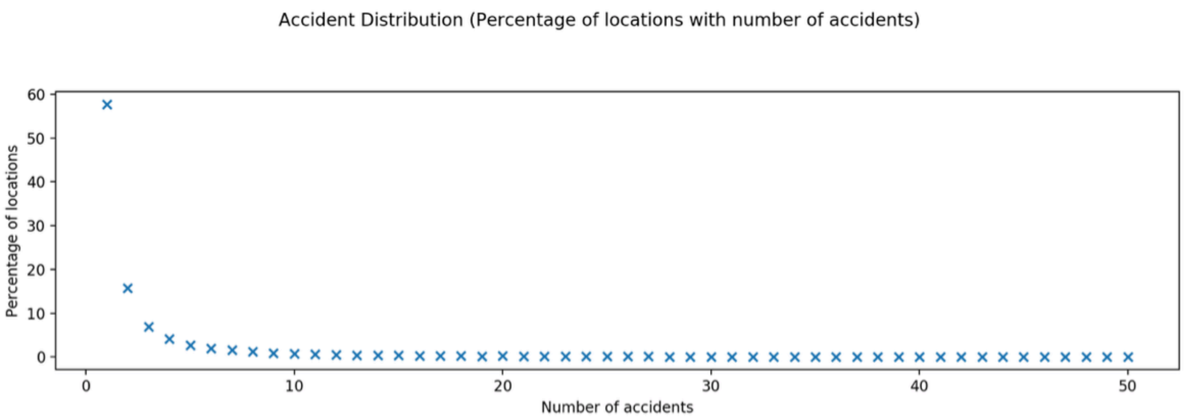

Figure 7: The percentage of locations vs. the number of accidents (capped at 50) in Florida.

The plot shows that most locations (almost 60%) have one accident, and almost 20% have 2 accidents. There is only 1% of locations at most with 10 or more accidents. Therefore, graph shows that the distribution of accidents is quite scattered around Florida, where most locations only have 3 or less accidents.

External factors

Several external factors such as the climate and weather of the area during the accidents in Florida were analyzed. To do so, plots of various variables were graphed to view the correlations.

Weather

A graph of the number of accidents vs. the temperature of the area in Florida was plotted to observe any correlation between the two variables. A plot of this comparison can be seen in Figure 8.

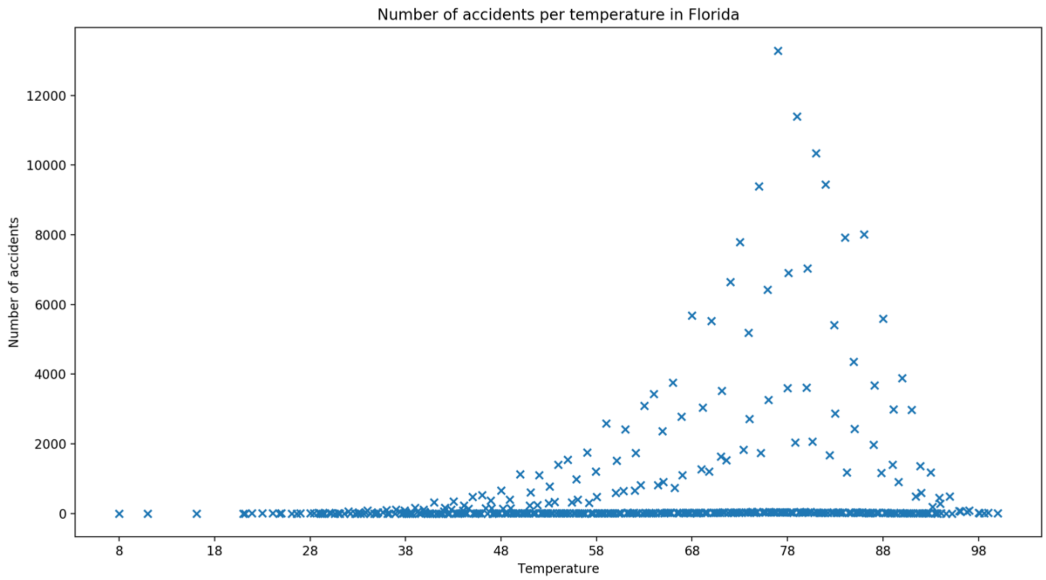

Figure 8: The number of accidents vs. the temperature in Florida.

The graph shows that the most accidents occurred when the temperature was around 80°F. Other temperatures below 60°F or above 90°F show 2 or less accidents. This plot suggests that people tend to go out during warmer temperatures in Florida, which can result in a higher number of accidents. In other words, there are not many drivers out during temperatures that are very low or very high, resulting in a lower number of accidents.

Visibility

A graph of the number of accidents vs. the visibility of the area in Florida was plotted to observe any correlation between the two variables. A plot of this comparison can be seen in Figure 9.

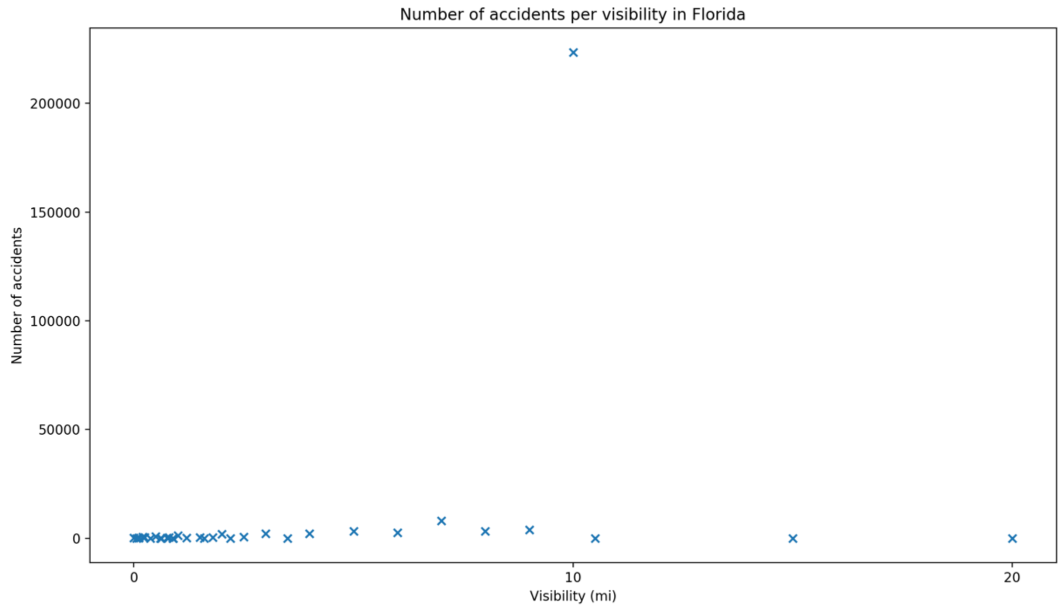

Figure 9: The number of accidents vs. the visibility (miles) in Florida.

The graph shows that the most accidents occurred when the visibility was around 10 miles. Therefore, it can be suggested that there are not many drivers out when the visibility is too low, resulting in a lower number of accidents. In other words, drivers tend to go out when there is a good visibility present that is ideal for driving, possibly resulting in a higher number of accidents.

Weather

A graph of the number of accidents vs. the weather of the area in Florida was plotted to observe any correlation between the two variables. A plot of this comparison can be seen in Figure 10.

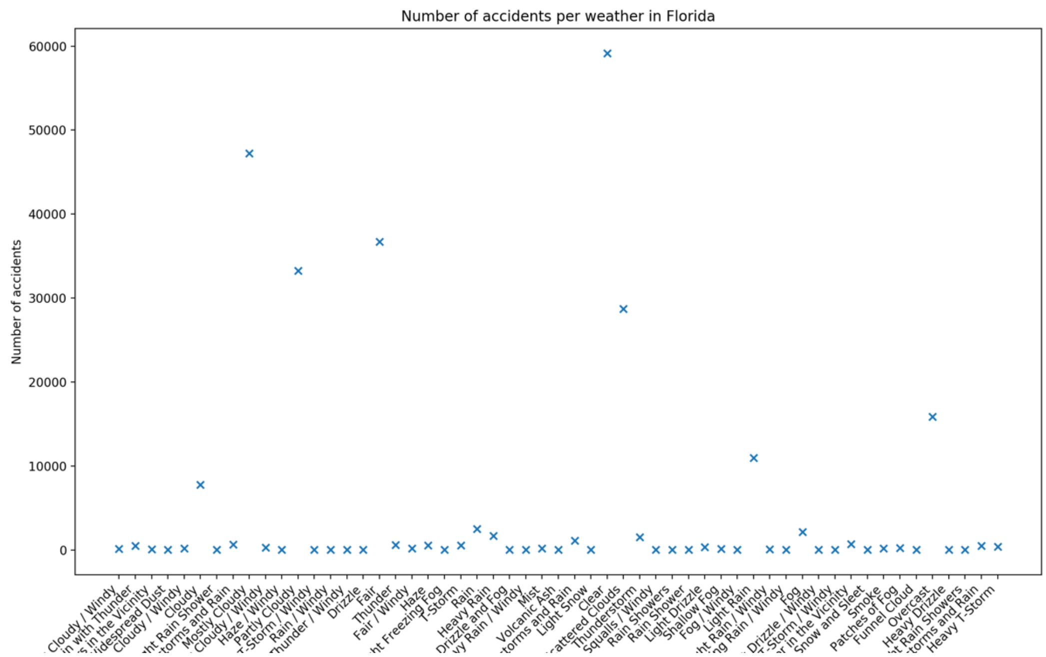

Figure 10: The number of accidents vs. the weather in Florida.

The graph shows that the most accidents (over 30,000) occurred when the weather was ideal for driving: mostly cloudy, partly cloudy, fair, and clear. When the weather was cloudy, rainy, foggy, or overcast, there were at least 5,000 accidents. This type of weather is common but not ideal for driving, which results in less people going out, but still enough to cause a lot of accidents with the weather being a large hindrance during travel. The fact that there are not many accidents when the weather is very severe, suggests that many drivers do not go out when there are unfavorable conditions. Therefore, it can be proposed that there are more accidents when the weather is ideal for traveling because there are more cars out on the road.

Wind speed

A graph of the number of accidents vs. the wind speed of the area in Florida was plotted to observe any correlation between the two variables. A plot of this comparison can be seen in Figure 11.

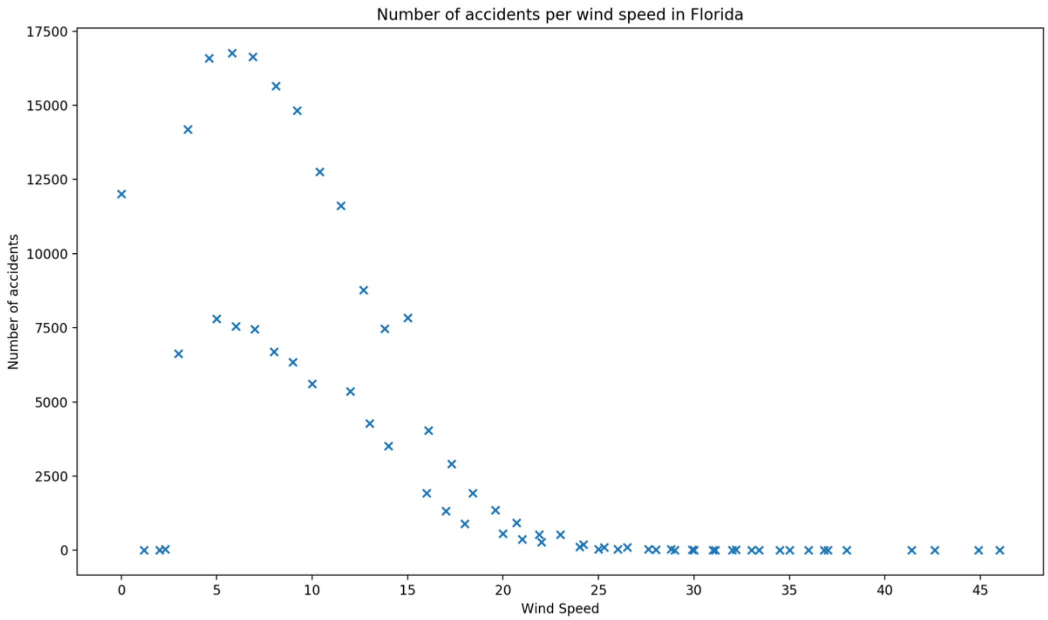

Figure 11: The number of accidents vs. the wind speed (MPH) in Florida.

The graph shows that the most accidents occurred when the wind speed was around 5 to 10 MPH, with almost all accidents occurring when the wind speed is less than 15 MPH. This indicates that are not many drivers who go out on windy days in Florida, resulting in a lower number of accidents when compared to days with a low wind speed. As drivers tend to travel more when the wind speed is less and favorable for driving, more road accidents can occur.

As seen from the graphs, many external factors can affect the accident distribution in an area. Almost all of the analyses indicate that there are more accidents when the conditions are more favorable to driving. Therefore, it can be suggested that more people out on the road result in a greater number of accidents, rather than the external factors themselves affecting people's driving. Additionally, the more favorable conditions can influence a driver to be more confident and less careful than if the conditions were severe and not ideal.

Intersections

Lastly, the accident distribution near intersections was specifically observed. To do this, a plot comparing the percentage of locations at intersections to the number of accidents was generated. Figure 12 shows this plot.

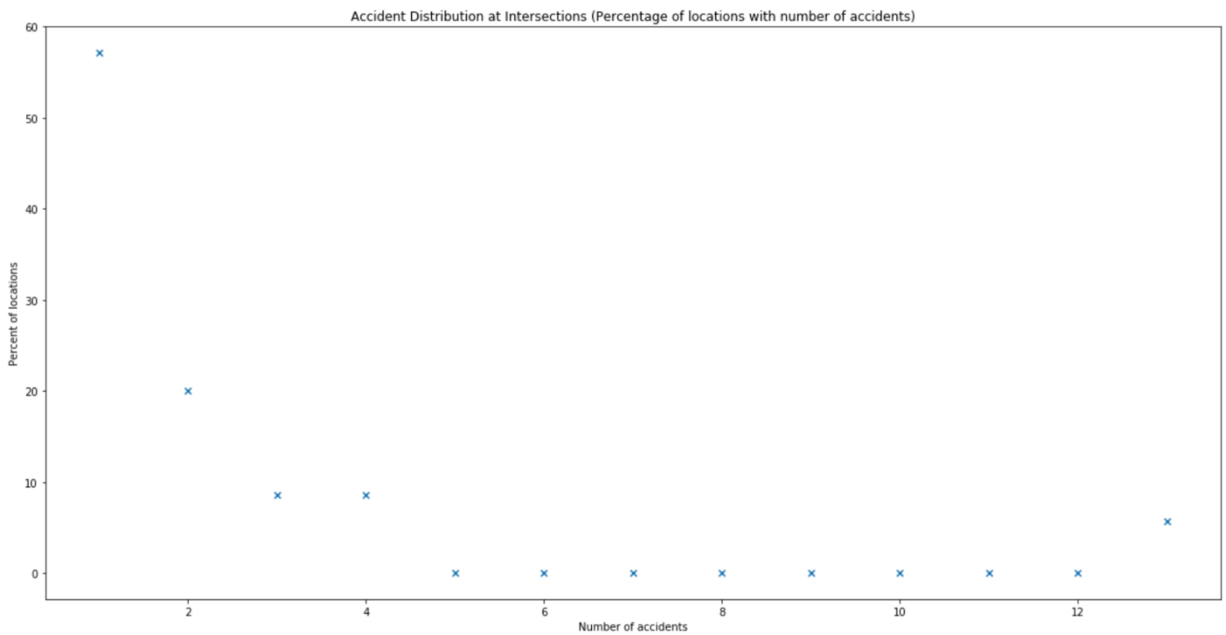

Figure 1: The percentage of locations at intersections vs. the number of accidents in Florida.

The graph shows that most intersections (almost 60%) have only 1 accident, and around 20% of intersections have 2. This is very similar to the accident distribution of all locations in Florida, where many accidents are scattered around different locations.

Conclusion

The first research discussed, route and stop planning for an autonomous vehicle shuttle service, looked into various aspects of the city and mobility patterns to plan the route and stops for a shuttle service. Characteristics such as the city residents' routes, vehicle speeds, residences, points of interests, locations of crowds, and parking deficiencies were all investigated. In general, the autonomous vehicle should include stops of users who are more likely to use a public transportation, and in places where the visitation is high in the city. Additionally, the autonomous vehicle should aim at lessening the burdens of city residents such as parking woes or difficulties in reaching the nearest public transportation stop. Observing a city's mobility pattern closely can result in the most effective route for an autonomous vehicle shuttle service.

The second research discussed, accident risk level prediction, closely observed the accident distribution in Florida and several factors in which accidents occur. The data generated suggests that accidents throughout Florida are scattered, with almost 80% of locations having only a couple of accidents or less. Additionally, more accidents occur when the conditions for driving are more favorable, suggesting that the number of accidents is proportional to the number drivers on the road. Furthermore, this data can imply that drivers are more confident or not as cautious when the environments are ideal in comparison to conditions that are more severe and harmful.

References

Car Accident Statistics in Florida for 2018 & 2019. (2020). Retrieved November 5, 2020, from https://www.toddminerlaw.com/blog/2020/august/car-accident-statistics-in-florida-for-2018-2019/.

Ondruš, J., Kolla, E., Vertaľ, P., & Šarić, Ž. (2020). How Do Autonomous Cars Work? *Transportation Research Procedia,44*, 226-233. doi:10.1016/j.trpro.2020.02.049

Potoglou, Dimitris, and Pavlos S. Kanaroglou. (2008). "Modelling Car Ownership in Urban Areas: a Case Study of Hamilton, Canada." *Journal of Transport Geography*, vol. 16, no. 1, 2008, pp. 42–54., doi:10.1016/j.jtrangeo.2007.01.006.

Tian, Y., Pei, K., Jana, S., & Ray, B. (2018). DeepTest: Automated Testing of Deep-Neural-Network-driven Autonomous Cars. *Proceedings of the 40th International Conference on Software Engineering*. doi:10.1145/3180155.3180220

Wagner, I. "Topic: Road Accidents in the United States." (2020). Retrieved November 18, 2020, from www.statista.com/topics/3708/road-accidents-in-the-us/.

What is an Autonomous Vehicle? (2020). Retrieved October 19, 2020, from https://www.twi-global.com/technical-knowledge/faqs/what-is-an-autonomous-vehicle