Predicting Hourly Shared E-scooter Use in Chicago:

A Machine Learning Approach

by

Lauren Fietz

A THESIS

Presented to the Department of Business Administration
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2020

An Abstract of the Thesis of

Lauren Fietz for the degree of Bachelor of Science
in the Department of Business Administration to be taken June 2020

Title: Predicting Hourly Shared E-scooter Use in Chicago: A Machine Learning Approach

Approved: ___________*Ryan Cabinte*___________________
Primary Thesis Advisor

Shared e-scooter programs were first implemented in 2017 to solve problems with the current transportation landscape. Combining ideas from mobility as a service, micromobility, and multimodal transportation, shared e-scooter systems and other forms shared of transportation programs have the potential to reduce or eliminate the need for unsustainable personal vehicles. However, shared e-scooters can create more problems than they solve. Some problems e-scooters can create include vandalism, lack of accessibility, hazards for the rider and pedestrians, and added pollution to the environment. With proper management, these problems can be mitigated.

Using frameworks from optimizing bike sharing programs, a predictive algorithm for shared e-scooters to predict hourly trips for e-scooter pilots was created. Features that help predict hourly e-scooter trips include time of day, number of days since its inception, rainfall, wind speed, and more. Machine learning models with the best accuracy at predicting e-scooter trips includes K nearest neighbors, decision tree, and random forest. Shared e-scooter system managers can use these models for optimal allocation of e-scooters to maximize ridership.

Acknowledgements

I would first like to thank all my advisors with their help and patience in this process. My advisors include primary advisor Professor Ryan Cabinte Program Manager for the Center of Sustainable Business Practices, CHC representative Doctor Trond Jacobsen Director of Forensics and Instructor of Information Science, and second reader Doctor Pradeep Pendem, Assistant Professor of Operations and Business Analytics. Without their guidance and support, I would have not been able to complete my research. Each of these professors gave me valuable insights into my topic, new contribution research, and how to write a thesis.

I would also like to thank my family and friends for supporting me during this time. It is stressful writing a thesis, especially during a pandemic, but with their support and guidance I was able to stay focused on my research. I would like to thank my mom Jill Fietz specifically because she offered to help with data entry. Although her help was not needed, I appreciate the sentiment.

Table of Contents

List of Figures

List of Tables

Introduction

In its current state, the urban mobility system is socially and environmentally unsustainable. As cities are growing rapidly and becoming denser, the current urban transportation systems will not be able to have effective modes of transportation. Vehicle combustion accounts for about 20% of the world's carbon dioxide emissions and average vehicle speed in cities can be as low as 20 kilometers per hour (Burns 2013). Furthermore, personal vehicles are generally expensive to purchase and hard to park in cities. Low-income families can be hurt by the transportation system, often further increasing social inequality (Barreto 2018). The United States dependency on gas dependent vehicles in urban cities causes pollution and a barrier to social and economic equality. Solutions to solve urban mobility issues have been suggested, such as ride sharing programs or autonomous electric vehicles.

Alternatives to the Current Transportation System

Mobility as a Service, or MaaS, could solve some of the problems caused by transportation. Mobility as a Service would eliminate the need for private vehicles, making transportation public or shared. In New York, it is predicted that the implementation of ridesharing services could reduce the number of taxis on the road by 75% without impacting travel time, reducing fuel consumptiob and vehicle miles travelled (Helveston 2017). MaaS business model can be completely sustainable if, "1) the modes that are offered are sustainable 2) the quality of the service is high enough so that users develop positive attitudes towards sustainable transport modes 3) the users shift to sustainable modes and 4) the financial model allows the MaaS companies to earn a profit by offering sustainable transport (Tol 2017).” There are different vehicles that are used in MaaS, ranging from bike sharing to public transit to ride sharing. Users of MaaS benefit from multimodal or intermodal travel, choosing which service or services they would most benefit from, providing more flexibility in cities than the traditional personally owned transportation.

Multimobility, using several forms of transportation to get to your destination, improves flexibility within transportation and can decrease dependence on personal vehicles. Forms of multimodal transportation include public transportation, bike sharing, or e-scooter sharing. Shared vehicles, such as bikes or e-scooters, can help to increase mobility, reduce greenhouse gas emissions, and improve economic development (Shaheen 2019). Shared e-scooter systems have been implemented recently and have been met with some criticism. However, with proper management, shared e-scooters have the potential to be an innovation that changes the current transportation landscape for good.

Benefits of Shared E-scooters

Starting in September of 2017, the company Bird started a revolution in transportation with their implementation of a shared, dockless electric scooter system in Santa Monica, California. These electric scooters were touted as a solution to the “last mile” problem, describing a problem that occurs in cities when a person’s destination is too far to walk, but too close to drive. The “last mile” problem frequently occurs in cities often due to a lack of parking or not being able to easily access a destination through public transportation, encouraging the use of cars in cities. Micro-vehicles or types of micro-mobility, such as scooters, skateboards or bikes, are viable alternatives for trips that are less than five miles, which makes up 60% of all trips according to a survey completed in the United States in 2017. Furthermore, this survey showed that 76% of trips under five miles are taken by personal vehicles making short trips extremely inefficient (US Dept. of Transportation 2018). Shared vehicles like e-scooters and bikes can solve the “last mile” problem by creating a more efficient form of transportation by improving trip times, increases public transportation access, reducing congestion in parking spaces and the road, and decreasing the need for car ownership in urban cities.

By using multi-modal transportation systems that incorporate e-scooters, people would be able to get to a destination quicker in a more sustainable way due to the flexible nature and efficiency of dockless e-scooters. E-scooters provide a flexible form of transportation for tourists who wish to get around efficiently in cities. People who consistently use shared e-scooters cite the service’s speed and convenience as motivators (Fitt & Curl 2019). For people without a personal vehicle, e-scooters might be a cheaper option than buying a car. Companies like Bird and Lime offer discounts to low-income households, attempting to solve some of the inequalities in the current transportation system. E-scooters have the potential to bring transportation justice to individuals who have unequal access to employment, healthcare, shopping, and recreation (Deka 2004). Additionally, e-scooters can help relieve problems that motorized vehicles cause such as traffic, noise pollution, and lack of parking. Except for walking and an advanced transportation system, a shared bike or scooter system seems to be the most sustainable alternative for short distance transportation in dense cities (Chang 2019).

Drawbacks of Shared E-scooters

However, several critics suggest that e-scooters do not to solve the last mile problem and can create new problems for cities. Incidents of vandalism have occurred in nearly every city that e-scooters are in. People throw dockless e-scooters in lakes and on top of trees to protest e-scooter systems, creating more waste, a potential hazard, and an eye sore as well. Other concerns about e-scooters include blocking pedestrian pathways, especially impacting citizens who are disabled. A shared e-scooter service is not accommodating to people with physical disabilities due improper parking on sidewalks, blocking ADA parking spots, and hard to balance on scooters. In a focus group completed by the Portland Bureau of Transportation about concerns regarding accessibility when an e-scooter service is implemented, one participant, who is blind, noted that they bump into e-scooters often and it is more challenging to walk after the implementation of an e-scooter service (PBOT 2019).

Shared e-scooters are not as financially accessible as they seem, with most users being millennial, wealthy, white, and male. In Portland's 2018 e-scooter pilot, only a few people with low income signed up for discounts. In another focus group conducted by PBOT, one individual of color says, "it is not in our culture to pick up something off the street, ride it, and leave it for the next person." Their report explains their comments further, writing, "for some focus group participants, the overall threat of an escalating incident outweighed the desire to try e-scooters." Although proponents of shared e-scooter services like to say that shared e-scooters are socially sustainable, e-scooter services are not equally accessible.

Furthermore, e-scooters are not as sustainable as e-scooter companies may suggest. Shared, dockless e-scooters short life cycles create more e-waste and release more CO2 than half of the alternative transportation methods in cities. Although e-scooters are much more efficient than personal automobiles, the short life cycles of shared e-scooters and the emissions associated with picking up and dropping off e-scooters make e-scooter sharing have more of an environmental impact than the modes it displaced. According to Joseph Hollingsworth's paper "Are e-scooters polluters? The environmental impacts of shared dockless e-scooters," using a Monte Carlo simulation, Hollingsworth found that shared e-scooters current operations have a 65% chance to be worse for the environment than their displaced trips. If more efficient operations were put in place, like efficient pick up and drop off vehicles or low collection distance, shared e-scooters would have less emissions than 35-50% of transportation it displaced. If shared e-scooter life span extended to two years, dockless shared e-scooters would emit less than 4% of the trips it displaced (Hollingsworth 2019). Without drastic operational improvements, they are more likely to be worse than better for the environment. These problems can be mitigated through changes in public policy and e-scooter operations.

Lastly, shared e-scooter services are not safer than other transportation modes. E-scooter riders ride in unsafe ways, such as riding on the sidewalk, with multiple people on one scooter, while intoxicated, and without a helmet. Unsafe riding can create accidents that could harm not only the scooter rider, but other pedestrians or passengers in other vehicles. Accidents occur more frequently when e-scooters are used more frequently. In a study completed in Brisbane, Australia, helmets were only worn in 46%

of accidents and the rider had been drinking alcohol in 27% of accidents. Helmet use is a large concern, with the severity of injuries increasing when the rider does not wear a helmet (Mitchell 2019). On UCLA's campus, it was observed that 94.3% of riders were not wearing a helmet (Quitana 2019). This statistic is concerning due to the severity of head injuries, something that public officials must try to prevent. One example of using regulation to increase helmet use was that e-scooter companies participating in the Portland 2018 e-scooter pilot were required to give away helmets.

Improving E-scooter Service through Policy, Strategy and Data Analytics

With comprehensive operational and public planning, the concerns surrounding shared e-scooters such as safety, pollution, accessibility, and equality can be improved. Literature that provides a comprehensive guide for e-scooter regulation includes Mason Herrman's report, "A comprehensive guide to electric scooter regulation practices", and Susan Shaheen's paper, "Shared Micromoblity Policy Toolkit: Docked and Dockless Bike and Scooter Sharing". These documents review e-scooter regulation across the United States and give suggestions for public officials to regulate their own e-scooter service. Herrman's report compiles information about cities e-scooter services in financial, legal, and operational categories. Operational components that public officials should consider include fleet regulation, parking, equipment, education, communication, and data collection. Important data for cities to collect include trip starts and ends, crashes, trip distance, map of route, vehicle counts, location of towed vehicles, number of riders per time period, demographics, low-income users numbers, active customers, injuries, theft, vandalism and losses, parking compliance, weather, helmet use, maintenance reports, battery level, customer complaints, community outreach, and rider surveys.

Operational data was broken down into four different categories: usage, vehicle, user and survey data. Usage data helps city officials increase or decrease fleet sizes based on the number of riders per day, week, and month. Usage data can also help cities alter scooter deployment locations. User data can help show if e-scooters are being used by certain demographics or are causing issues for the public. Vehicle data shows e-scooter movement, distribution, and vehicle compliance while in operation. Lastly, survey data shows public officials what their constituents think should be changed about the program and give insights that numerical data cannot provide (Herrman 2019). Collecting, analyzing, and sharing data is vital for e-scooter systems to improve cities' mobility ecosystems. Through improved operations from data analysis, shared e-scooters have a chance to be a completely sustainable service (Tol 2017).

Bike Sharing Insights, Connections and Differences

Insights into improved shared e-scooter operations and strategy can be learned from bike sharing. The first bicycle sharing system was started in 1965 in the Netherlands. Dubbed as ‘Witte Fietsen’ (or ‘ White Bikes’), several bicycles were painted white and were put out for public use. The program did not go as planned as the public threw bicycles into the canals or were stolen. ‘Witte Fietsen’ collapsed after a few days. Bike sharing was not widely accepted until the beginning of the 21st century (DeMaio 2009). Bicycle sharing became more popular in the 2000’s due to improvement of bicycle sharing operations, concern for the environment, and to improve first/last mile connections, similar to some of the reasons for the rise of e-scooter sharing.

Differences between shared e-scooters and shared bicycles are important to consider for city planners and micro-mobility companies. Shared bicycles tend to be docked instead of dockless, creating less flexibility for shared bicycle riders. However, shared bicycles are more reliable, causing them to be used at different amounts of time, certain hours of the day and distance. Dockless e-scooters were ridden at a much lower average time and average distance than the docked bike system (McKenzie 2019). After studying docked bike share and dockless scooter share usage patterns in Washington D.C., McKenzie concluded that bike sharing services were primarily used for commuting to work, especially for people who were members of the bike sharing service (McKenzie 2019). A shared, dockless e-scooter service supports leisure, recreation, and tourism activities more so that commuting.

Although scooter and bike sharing generally service different types of trips, bike sharing is the closest transportation system to a shared e-scooter service. Since e-scooter sharing is relatively new, research on bicycle sharing can provide valuable insights into how the service should be operated. One main aspect of bike sharing research is placement of stationary bike sharing racks and distribution of bikes among these racks. By using rider data like station activity, customer behavior, location factors, and more, bike sharing providers can ensure high bike availability for customers. Satisfactory shared bike allocation is a difficult task due to the high cost of redistributing bikes and highly dynamic movements of shared bike users (Vogel, P., Greiser, T., & Mattfeld, D. C. 2011).

Several studies using algorithms such as supervised machine learning and mixed integer linear programming have been conducted to solve the demand allocation problem. For example, Arnab Kumar Datta’s Master’s Thesis on “Predicting bike-share usage patterns with machine learning” and Giot & Cherrier’s paper on “Predicting bikeshare system usage up to one day ahead” both use machine learning techniques to predict demand allocation or general demand. By accurately estimating bike share usage, bikes can be optimally allocated throughout the day. Properly allocated bikes lowers the probability of failure for riders not being able to secure a bike and reduces the chance for overpopulated bike racks. An optimally managed bike sharing system can reduce reliance on cars. Lessons from managing bike sharing can be used to manage e-scooter sharing.

Purpose of Study and Research Questions

This paper will be focused on the operational component of implementing a shared e-scooter system, specifically focusing on improving fleet regulation through collected and analyzed data. Although there are several bike sharing studies that use predictive algorithms for bike allocation, there have not been many studies completed on demand allocation for e-scooters. The lack of research in this area is due to the new nature of e-scooters and the complexities of a dockless system. However, a predictive algorithm for e-scooter sharing demand can provide benefits for the mobility company, public officials, customers, and the public.

An analysis of hourly and daily e-scooter demand can cut costs and optimize operations for the mobility company. E-scooter managers can place e-scooters on the sidewalk or remove them depending on hourly predictions. An optimized e-scooter demand can eliminate unnecessary e-scooters on the street, decreasing the chance for an e-scooter to be damaged or block pedestrians' path. Furthermore, finding daily and hourly optimal demand for e-scooters can ensure that there will be enough e-scooters available to ride. Depending on the predicted number of e-scooter demand and utilization rates per e-scooter preferred by the company or city, shared e-scooter operators can accurately predict the number of e-scooter they want on the street. This analysis was dependent on what data is accurate and publicly available. The City of Chicago is one of the few cities that has publicly available data from their e-scooter pilot. To determine hourly trips, data was collected from Chicago's open data portal and open use weather data from National Centers for Environmental Information. This analysis will only include data from the Chicago area and will primarily determine e-scooter pilot predictions. With more data this model could predict general e-scooter trips rather than pilot trips. This analysis is not meant to be fully comprehensive or completely accurate predicting e-scooter trips but is meant as a starting point for this type of analysis. Additionally, this analysis will show what factors are most important in predicting hourly e-scooter trips.

Research Questions

- What features/predictors are important in predicting hourly e-scooter usage?
- Can a predictive algorithm accurately depict hourly and daily trends of e-scooter usage in Chicago?
- What are the limitations of this analysis? What conclusions can be drawn?

Data Collection and Description

Machine learning takes independent or input variables ($X_{1,} X_2 \ldots X_n$) to predict a dependent or output variable (Y). When predicting hourly e-scooter usage rates (Y_1), it is important to determine what independent variables ($X_{1,} X_2 \ldots X_n$) influence the number of e-scooter trips. Some factors that might influence shared e-scooter trips include weather, other modes of transportation, time of day, what day it is, and more. For shared e-scooter programs that are just beginning, days since e-scooters were placed on the streets might be another predictor. For Chicago's e-scooter pilot, a significant portion of rides were in the earlier months of the e-scooter pilot. Not only are e-scooters used in warmer weather like during the summer, e-scooter trips tend to be higher at the start of the pilot since people were encouraged to try the vehicles for the first time. The city of Chicago reported that 49% of customers used an e-scooter only once, while only 15% took more than five trips(E-Scooter Pilot Evaluation 2020).

Data was collected from Chicago's open data portal and National Centers for Environmental Information. Limitations of the data set included missing e-scooter trip data for the month of October and lack of other features that could have predicted e-scooter trips, like large events that would increase the need for flexible transportation. The number of observations in the dataset is 2374, which is a smaller dataset to use for machine learning. The small number of observations and high number of features will contribute to some machine learning algorithms having low accuracy.

Predictor Variables

- time: Hour of the day grouped into 6 different periods. Categorical.

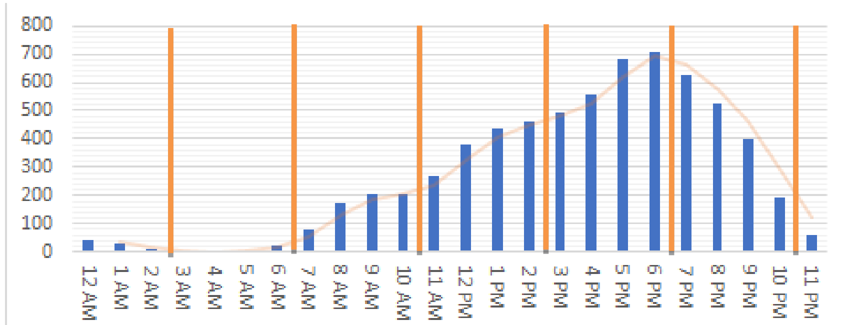
- days_since: Days since the start of the pilot. Numeric.
- day_type: 0 for Sunday through Thursday and 1 for Friday or Saturday. Binary.
- holiday: 0 for typical day and 1 for federal holidays. Binary.
- temp_f: Temperature at the start of the hour in Fahrenheit. Numeric.
- rain: Rain measured in inches for the hour. Numeric.
- wind_mph: Speed of wind at the start of the hour in miles per hour. Numeric.
- avg_transportation: Average metro and bus use for the day. Numeric.

Table 1.0 gives an example of the first several observations. Tables 1.1 and 1.2 depict a data summary and correlation for each variable.

Time

Arguably the most important variable when determining e-scooter trips is the time of day. Time is cyclical and cannot be a numeric value as it is. Instead of using time as a number, time was split up into 6 four-hour sections. Categories include earlymorning (3am-6am), morning (7am-10am), midday (11am-2pm), afternoon (3pm-6pm), night (7pm-10pm), and latenight (11pm–2am). Time started at 0 for 12:00am to 12:59am and ended at the 23rd hour for 11:00 to 11:59pm. If a ride was started at 4:45pm, then it would fall under the afternoon category. There are 108 days from the pilot that are recorded in the dataset. Some categories, such as midday, each hour included from the 108 days included. Early morning and late night have the fewest entries as people are rarely riding during this time. The only times with a few entries are hours 3am and 4am.

Figure 1.0: Average Trips per Hour



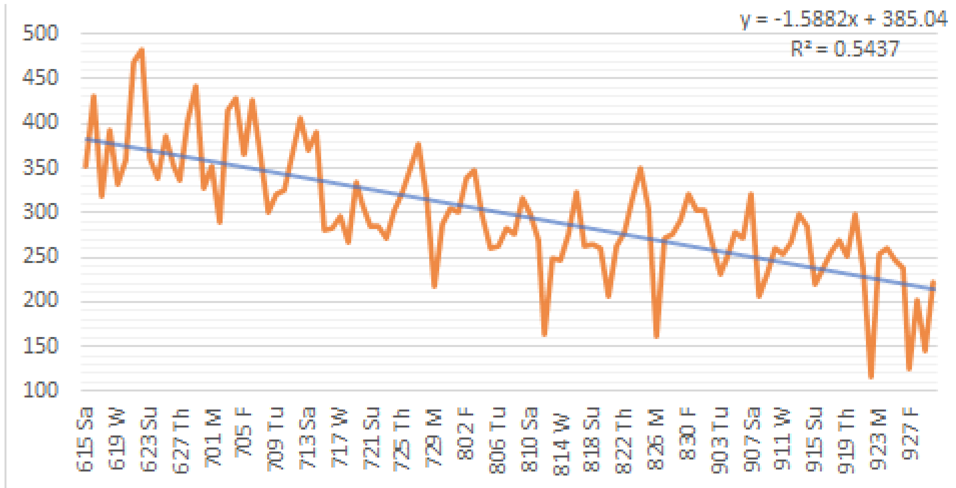
The chart above depicts average trips per hour throughout the 108 days of the e-scooter pilot. The trendline is the moving average of hourly data. Average trips is on the y axis.

Figure 1.0 shows a clear relationship between number of trips and time of day. Managers can also receive insights hourly from this chart alone, as times have clear patterns can tell a manager what shared e-scooter use will look like for most days. The peak of average trips is from 5pm to 7pm, with the highest number of average trips starting from 6:00pm to 6:59pm. There is a gradual incline from 5am to 4pm and a steeper drop from 6pm to 11pm. From 12am to 5am, there are hardly any trips. When predicting hourly e-scooter trips, this trend will be the important in the predictive algorithm.

Days Since Pilot Start

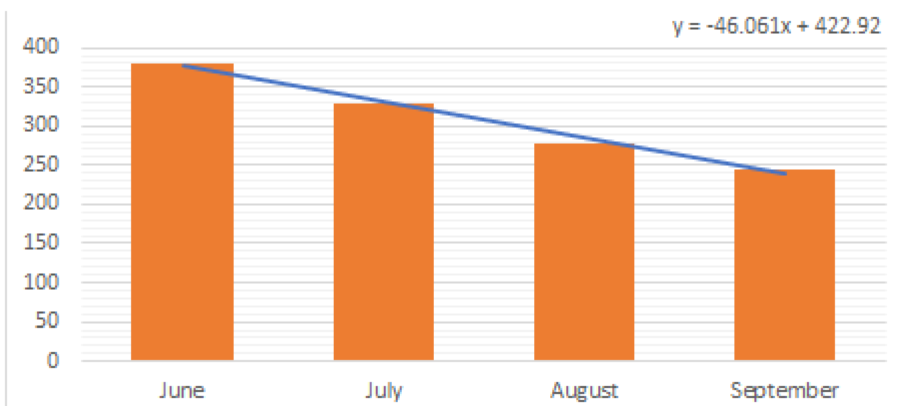
The number of days since the pilot started is important to consider because ridership tends to be higher at the beginning of the pilot, shown in Figures 1.1 and 1.2.

Figure 1.1: Average Trips per Day



The chart above is a line chart that depicts the relationship between average trips per day over time. The day is on the x axis and average trips is on the y axis. The linear trendline describes the relationship between variables. “615 Sa” on the x axis can be interpreted as June 15th, which occurs on a Saturday in 2019.

Figure 1.2: Average Trips per Day by Month



The chart above depicts average number of trips per day by month with a linear trendline. Although October is not included in the dataset, it follows the same pattern as shown in Chicago’s e-scooter pilot report. Average trips is on the y axis.

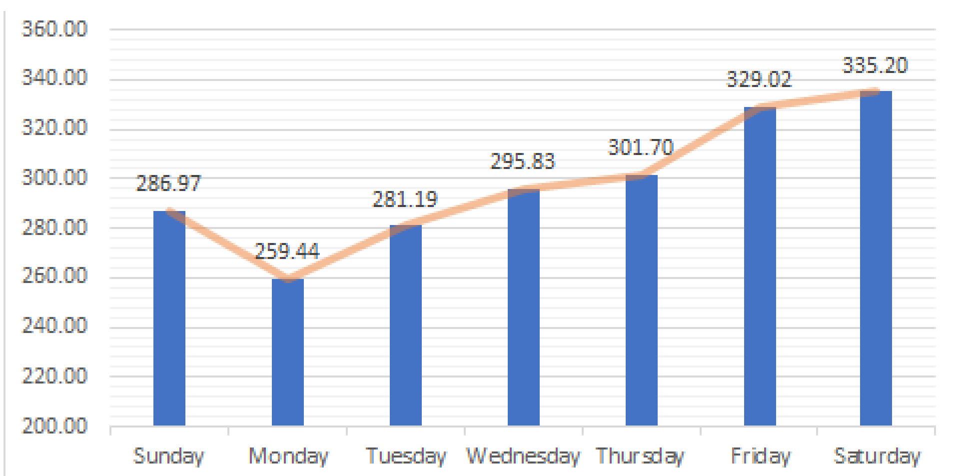
Figures 1.1 and 1.2 shows a negative correlation between days since the start of the pilot and average trips taken per day. Figure 1.1 describes a relationship where for each day that goes by, average e-scooter trips decreases by 1.5. The R^2 number for

Figure 1.1 is 0.54, meaning that number of days since the start of the pilot explains the average e-scooter trips well. Compared to the other variables with a linear relationship and R^2 number, days since pilot start explains the variation in average e-scooter trips the best. For each month in Figure 1.2, average e-scooter trips are expected to decrease by 46. The clear relationship between trips and number of days since the start of the pilot should be included in the machine learning algorithm as it will help predict hourly number of trips.

Day Type

The day of the week is another important factor that should be considered when predicting e-scooter trips. Since e-scooter trips tend to support leisure and tourism, it is likely that e-scooter trips will increase on days of the week when these types of activities are more prominent.

Figure 1.3: Average Trips by Day of the Week



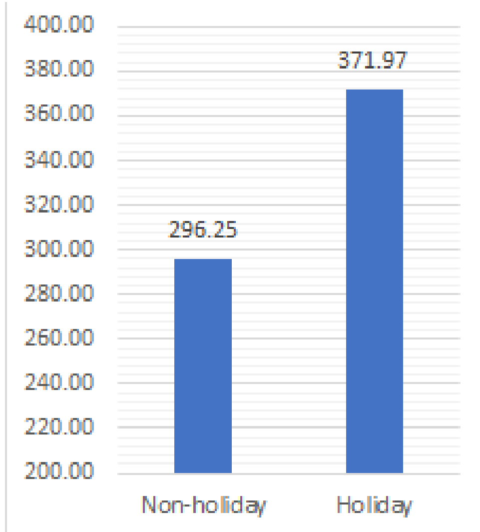
The figure above depicts average trips for each day of the week. A trendline is included to show day to day changes more prominently. Average trips is on the y axis and day of the week is on the x axis.

Figure 1.3 shows a pattern for average trips per day of the week. Sunday through Thursday have a lower number of average trips than Friday and Saturday. Since machine learning algorithms with many features and a small number of observations tend to be inaccurate, only one predictor column or feature for day of the week will be used. Friday and Saturday have the highest average of trips and tend to support leisure or tourism activities more so than any other day of the week, so the predictor column for day of the week will describe whether the day is Friday or Saturday. A feature or predictor column for each day of the week would cause more inaccuracy in the model unless there were substantially more observations.

Holidays

Holidays tend to change people's typical behavior, especially in transportation. On a holiday, people will either travel less or more depending on the holiday. During the e-scooter pilot, there were only 3 large federal holidays in the United States: Father's Day, the 4th of July, and Labor Day. To implement holidays as a predictor of e-scooter trips, a binary column describing whether the observation lands on a holiday will be included in the machine learning algorithm.

Figure 1.4: Average Trips for Holidays



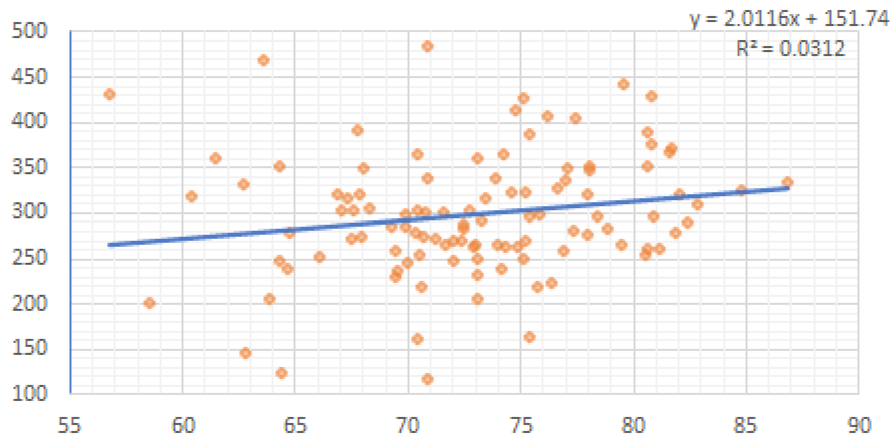
The chart above depicts average trips for each holiday. There were only three days that are averaged for the holiday column.

As Figure 1.4 shows, there is about a 77 increase in average trips when a holiday occurs. The flexible nature of e-scooters supports holiday or unusual travel, likely describing why holidays have a higher than average number of trips. However, there are only 3 days that are labeled as holiday, so there cannot be many assumptions based on this data. Although the holiday feature can be helpful to predict deviations in trip data, it will be a less important of a predictor than other variables like time or days since pilot start.

Temperature

Temperature at the start of the hour was recorded at the Chicago O'Hare International Airport and was collected from National Centers for Environmental Information. Temperature is another important feature in predicting hourly e-scooter trips. The colder it is, the less likely people will be willing to ride a scooter outdoors. A personal vehicle or public transportation is a warmer option. In hotter weather, a shared e-scooter might be a more comfortable option than a crowded bus or metro. Figure 1.5 confirms assumptions about trips relationships with weather.

Figure 1.5: Daily Average Trips and Daily Average Temperature



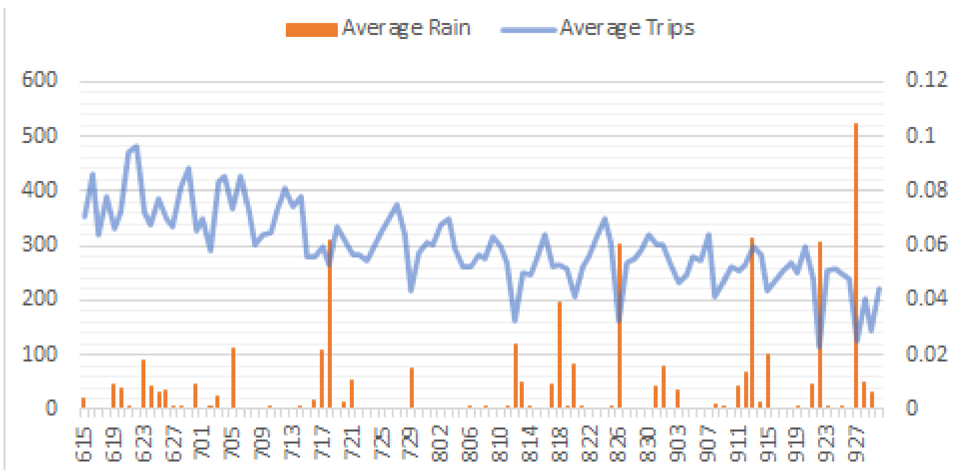
Daily average trips is on the y axis and daily average temperature in Fahrenheit is on the x axis. Each point on the plot represents a day. A linear trendline is added to describe the relationship between the variables. Variables such as temperature that fluctuate due to the hour of the day were averaged for the day to optimize the visualization of the relationship between trips and temperature. In other words, time of day needed to be eliminated from the visualization.

A relationship is confirmed between trips and temperature. As temperature increases by one, the daily trip average will increase by 2. The R^2 number is 0.03, meaning that temperature is bad at explaining the variation in the number of trips. Although this factor is not as significant as time of day in explaining variation in the number of hourly trips, the relationship between temperature and trips can help predict some of the nuance in hourly e-scooter trips and should be included in a machine learning algorithm.

Rainfall

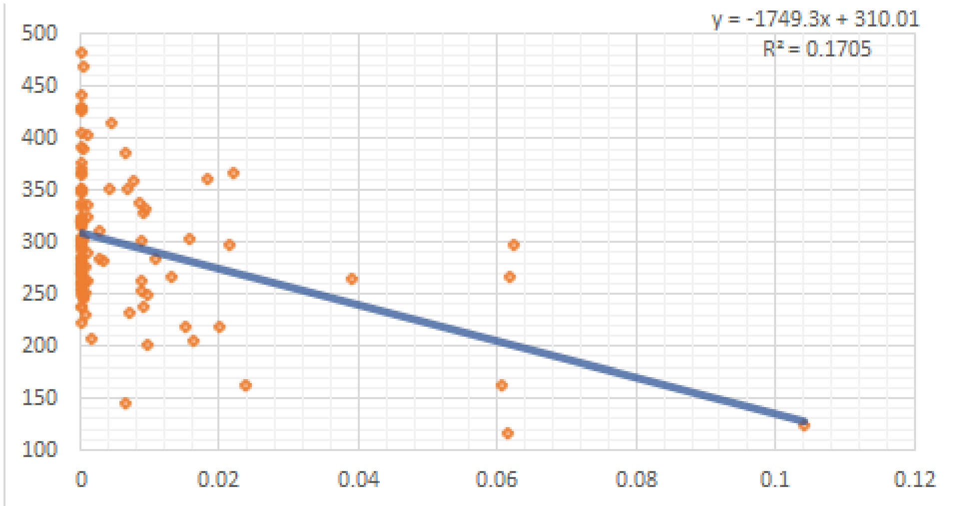
Data on rain was recorded from the Chicago O'Hare airport and describes how many inches of water was collected during the hour. Trace amounts of rain were filled in as 0.005, the average between zero and the smallest amount of rain able to be measured, 0.01. Rain is an important predictor to include when predicting e-scooter trips per hour. The more it rains, the less likely people will be to ride shared e-scooters outdoors. People do not want to be wet, cold, or feel unsafe on an electric scooter.

Figure 1.6: Average Daily Trips and Average Daily Rainfall



The chart above is a combo chart depicting average daily trips in a linear format and average daily rain in a bar chart format. Rain is in inches. Variables such as rainfall that fluctuate due to the hour of the day were averaged for the day to optimize the visualization of the relationship between trips and rainfall.

Figure 1.7: Average Daily Trips and Average Daily Rainfall



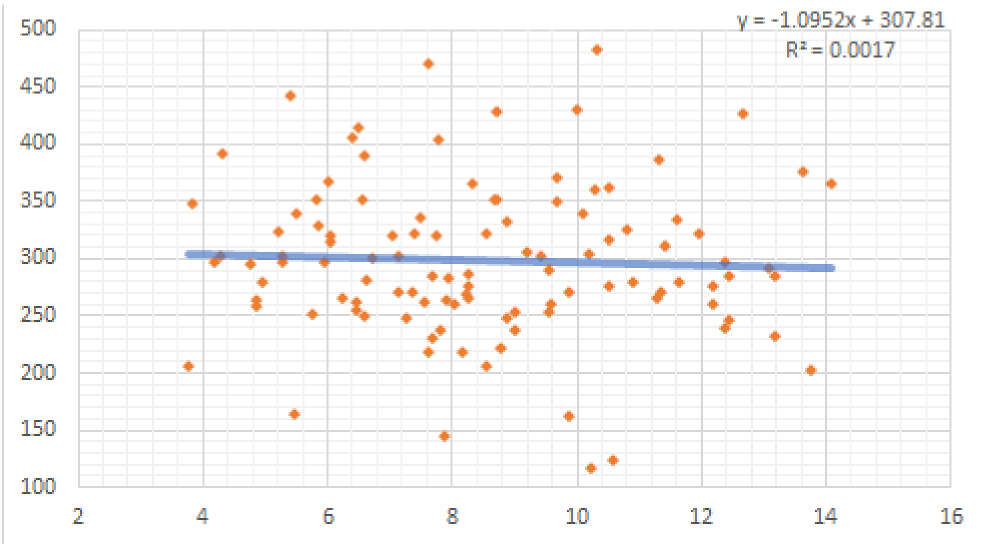
The chart above is a scatter plot with average trips per day on the y axis and average rain per day on the x axis. There is a linear trendline to describe the relationship between the variables. Rain is in inches. Variables such as rainfall that fluctuate due to the hour of the day were averaged for the day to optimize the visualization of the relationship between trips and rainfall.

Figures 1.6 and 1.7 describe a negative linear relationship between trips and rainfall. In Figure 1.6, except for one day, when the average rainfall is above 0.06 there is a sharp drop in average trips. In Figure 1.7, there is a negative trend line. For each 0.01-inch increase in average rainfall per day, there will be a decrease in average trips for the day by about 18. The R^2 number is 0.17, meaning that rainfall is bad at explaining the variation in average daily trips. However, rainfall has the largest R^2 number and explains the variation in average trips per day better than other weather feature. Rainfall should be included as a feature to help predict variation in hourly e-scooter trips.

Wind Speed

Like temperature, the wind measurement was taken at the beginning of the hour at the Chicago O'Hare airport. The relationship between wind and number of trips might be a factor in predicting trips taken per hour, especially when predicting e-scooter rides in "the Windy City". Wind can discourage e-scooter trips as the rider might be colder or feel unbalanced.

Figure 1.8: Average Daily Trips and Average Daily Wind Speed



The chart above is a scatter plot depicting the relationship between average wind speed in miles per hour and average trips per day. Average trips is on the y axis and average wind speed is on the x axis. A linear trendline is added to help explain the variables relationship. Variables such as wind speed that fluctuate due to the hour of the day were averaged for the day to optimize the visualization of the relationship between trips and wind speed.

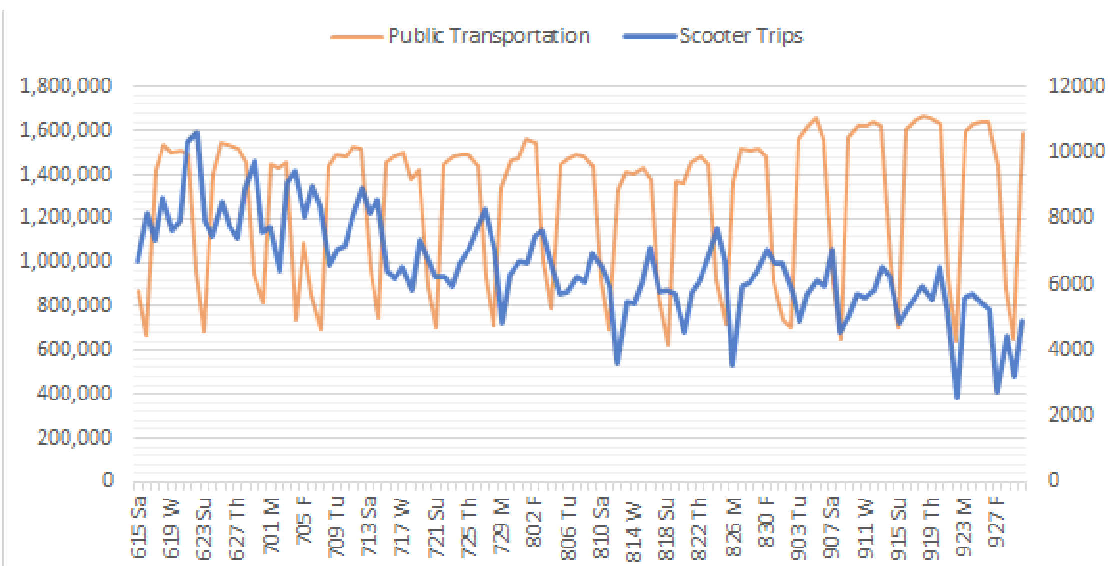
There is a negative relationship between average trips and average wind speed per day, but it is not strong. For each one mile per hour increase in the average wind speed per day, there is one less average trip per day. The R^2 number is 0.0017, meaning that average wind speed per day does not explain the variation in average daily trips.

Although wind speed has a poor linear relationship and R^2 number, it might help explain variation in hourly trips with a more complex algorithm.

Average Transportation

Transportation data was collected from the City of Chicago's open data page. However, only daily transportation data was available, not hourly. Transportation data combines the number of riders from both the bus and metro. Transportation could be an indicator of e-scooter trips because of multimodality. Shared e-scooters are meant to be used to solve the 'last mile problem', often connecting people to and from bus stops and metro stations. An increase in public transportation could mean that more people are using shared e-scooters and are less reliant on cars.

Figure 1.9: Daily Public Transportation and E-scooter Trips



The chart about depicts two lines, the orange line visualizing public transportation and the blue line visualizing total e-scooter trips per day. Numbers for public transportation is on the left y axis and numbers for e-scooter trips are on the right y axis. The x axis lists the day. For example, 709 Tu stands for July 9th which is on a Tuesday.

There is not a clear relationship between public transportation and e-scooter trips based on Figure 1.9. Public transportation tends to have its peaks on Monday through Friday, with lower travel on Saturday and Sunday. Shared e-scooters have their peaks on Thursday through Saturday, with a lower number of trips on Sunday through Wednesday. Even in a scatter plot depicting the relationship between public transportation and e-scooters, there were two clusters of data points with no clear relationship.

Additional Variables

Other variables I wish to analyze are whether there was a large event that could dramatically increase e-scooter use. For example, events such as the music festival Lollapalooza could be an important predictor for e-scooter use, especially since e-scooters support leisure activities more so than bike sharing. When comparing average e-scooter usage to the weekend of Lollapalooza, there was a higher number of trips taken than expected. Variables that are correlated with leisure or tourism such as number of people travelling from outside of Chicago or number of purchases at retail stores could also help predict shared e-scooter trips. Other forms of transportation should be considered when predicting shared e-scooter usage. Other variables like days_since could be changed since they are better for predicting shared e-scooter pilots rather than general shared e-scooter use. Instead of using days since launch as a predictor, the month would be a better predictor when the e-scooter program is no longer new.

Analysis

Machine Learning Algorithms

For this analysis, decision tree, random forest, and k-nearest neighbors are used. Other machine learning algorithms, like linear regression and support vector machine, were used to predict hourly trips but had a higher mean absolute percentage error (MAPE) and lower test accuracy). This analysis is a regression problem since the number of shared e-scooter trips are numeric and are not classified into a category. All analysis was conducted in Python using the scikit-learn module. The data was collected and cleaned in Excel, so there were no strings or missing data values.

Machine learning works by splitting data into independent variables or features ($X_{1,}$ $X_2 \ldots X_n$) and dependent variables or targets (Y). Both X and Y variables are then split into train and test datasets, typically with 75% of X and Y variables in train and 25% in test. After training the X and Y train dataset, it is then used to predict a Y variable using X test. Finally, the predicted Y is compared with Y test to get a score. Training and test scores can be produced, showing the accuracy of the trained model and test data. The optimal machine learning model finds the lowest error for both test and training data by finding the balance between bias and variance. A too complex model has high variance and overtrains or overgeneralizes the data. A too simple model has high bias and undertrains or undergeneralizes the data. Changing what features are included, using cross validation, and changing parameters can help to improve train and test accuracies. A validation curve can be used to find optimal hyperparameters by getting an accuracy score for several different hyperparameters.

Decision Tree

A decision tree splits a complex problem into several simpler problems through hierarchal organization. By using a process of splitting a root node into internal decision nodes into lead nodes, an easy to interpret, trained model can be created (Ahmad, M., Reynolds, J., & Rezgui, Y. 2018). A model using hierarchal, logical decisions can be used to predict hourly shared e-scooter trips. After the model was created and tested, a validation curve, where testing different hyperparameters on the training dataset, was used to prune the tree. Pruning the decision tree with hyperparameters such as max_depth or min_samples_split ensures that the model is optimally fit.

After finding the model with the best hyperparameters with a max_depth of 6 and min_samples_split of 60, the decision trees final training accuracy is 81.80% and the test accuracy is 78.66%. The feature importance for the decision tree algorithm can be calculated based on feature usefulness. Time is the most important variable at 85% after calculating the sum of each time category. The most important time categories include afternoon with an importance of 39.07%, midday with 21.99% and night with 19.10%. Following these three categories is days since the pilot started, with a feature importance of 8.29%. Morning has a feature importance of 4.00%. Next, average transportation has a feature importance of about 2.97%. In order of feature importance, wind speed, rain, temperature, and day type all rounded to 1%. Finally, holiday, late night, and early morning categories has 0.00% importance, hardly contributing to predicting the target variable. A decision tree can be visualized to help understand feature importance and how it makes decisions. However, given the decision tree's large length and span, it is not easy to interpret. The whole decision tree can be seen in Figure 2.0. A section of the regression tree can be seen in Figure 2.1.

Random Forest

The random forest method is like the decision tree method but addresses some of its shortcomings by using multiple decision trees or an ensemble. Using bagging and feature randomness, multiple uncorrelated trees are created to be more accurate than an individual tree. The final prediction is the average of all the individual decision tree predictions (Ahmad, M., Reynolds, J., & Rezgui, Y. 2018). The optimal fit with the hyperparameter n_estimators of 300 and max_depth of 8 for this model was found using a validation curve. After training and testing the model, the training accuracy for the random forest model is 86.29% and the test accuracy is 80.14%. The feature importance for the random forest model is like the decision tree feature importance with a few changes. Time is similarly the most important variables. Afternoon has a feature importance of 36.99%, midday has a feature importance of 19.61% and night has a feature importance of 17.15%. Days since start of the pilot is the next most important feature, with 9.17%. Average transporation has a feature importance of 3.31%, temperature at 3.01%, wind at 2.1%, and rainfall at 1.4%. Day type, late night, and early morning all have a feature importance that rounds to 1% and are most important according to their order. Holiday hardly predicts anything at all, with an importance of .01%. Unlike the decision tree model, a random forest model cannot be mapped easily.

K-Nearest Neighbors

The k-Nearest Neighbors or k-NN algorithm uses similar observations to help predict new observations. By averaging k closest observations, a new prediction can be made. K is a hyperparameter that can be changed depending on how many of the closest neighbors the model chooses to average. However, k-NN is a simple algorithm and does not have great accuracy for complex predictions. When the k-NN model included every feature that is included in the decision tree or random forest model, the test accuracy was 74% with a k of 1. To get a more accurate test score, features were reduced to get a simpler model. Additionally, a validation curve was used to find the best k number of neighbors for the model. After removing nearly all features except for the most important, time categories and days since start of the pilot, the best k-NN test accuracy of 70.0% was achieved. The training accuracy for the simplified model is 84.2%.

Comparing Algorithms

Out of the three machine learning models, Random Forest proved to be the most accurate model on training and testing scores. To compare the different models in depth, a random sample of 1.5% of samples from the data set was taken from the main dataset shown in Table 2.0. Each model then predicted the number of trips from each of the 24 random observations. After a prediction for each model was made, it was compared to the actual number of trips to generate a number for the absolute percentage error (APE) shown in Table 2.1. A mean APE was calculated for each model. For a specific random sample, the random forest model had the lowest mean APE of 62.68%. The next most accurate model was the decision tree with an MAPE of 69.49%. Finally, k-NN had the largest MAPE of 84.91%.

Limitations and Improvements

There are several limitations with this predictive algorithm, most of which can be solved with dataset manipulations. The main problem with this dataset is that there is a small amount of observations. Future e-scooter trips would be difficult to predict. For example, the only e-scooter trips recorded for October for the whole day was on October 15. Using the random forest model, the average absolute percentage error is 181% for actual vs predicted trips. The percentage error for October 15 is much higher than any random sample of the full dataset. Table 2.2 depicts the predicted number of trips, actual number of trips, and the APE for October 15. The large error for October 15 is likely because there were not enough observations to be able to predict farther into the future. This model is better for shorter term predictions. Furthermore, the dataset from the Chicago's e-scooter pilot might have not been complete. An incomplete dataset could explain some of the errors, especially the large error in the October 15 predictions that were too high.

A larger dataset be beneficial to the prediction in multiple ways. A larger dataset would likely improve feature importance. There were not enough observations to understand the nuances of e-scooter trips. Although hour of the day would still be the most important predictor, I expect other variables such as rainfall and temperature to be more important. E-scooter trips tend to drop dramatically during the colder or rainier months due to the exposure to the cold or rain when riding e-scooters. Only having data during the summer would not be good at predicting e-scooter use in January. Adding new observations daily would improve the model and predictions, helping e-scooter operators and public officials plan the next day better. Additionally, the time variable could be manipulated to be numeric using sine or cosine as the models were more accurate when predicting time numerically. After changing the time variable from numeric to categorical, the MAPE doubled for decision tree and random forest. However, time cannot be represented numerically as it ignores the cyclical trends of time.

Data could be improved to provide more accurate predictions. For example, weather data was taken from the Chicago O'Hare airport which does not necessarily reflect the weather that e-scooter riders experience. Weather should be recorded in Downtown or central Chicago to reflect what riders experience more accurately. However, I was limited to what data was openly available online. Furthermore, mean temperature and wind speed for the hour rather than the number recorded at the beginning of the hour would be a better representation of weather. Average transportation can be broken down by hour to provide more accurate representation of transportation. However, the correlation between variables such as transportation or temperature with number of trips might cause issues for an accurate model. Models using average rainfall, wind speed, temperature, and transportation should be tested against hourly rainfall, wind speed, transportation, and temperature to determine the best possible model. Additional features that were not included such as events or other leisure activities could be added to improve the model.

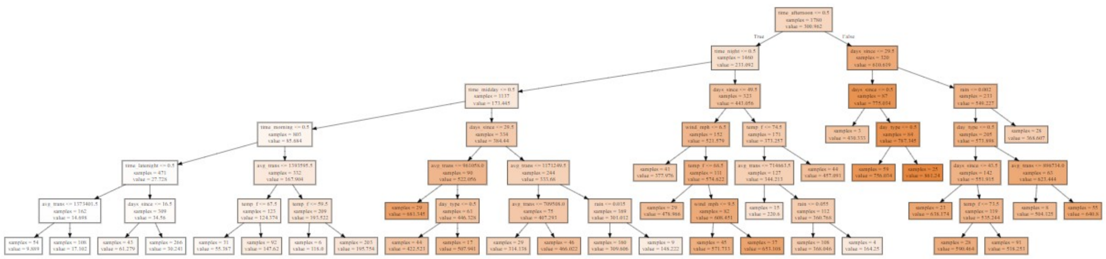
Conclusion

This research is meant to be a starting point for using machine learning models to improve e-scooter strategy and operations. By optimizing the number of e-scooters on the street, businesses and governments can correctly manage e-scooter systems to reduce personal vehicle dependence. Hourly changes to shared e-scooters trips fluctuate quickly, so a predictive algorithm that can highlight fluctuations before they happen in real time would be beneficial for optimal allocation of e-scooters. For example, if it is predicted that the temperature will drop and rainfall increases in a few hours, the shared e-scooter operators can remove e-scooters that will not likely be used under harsh weather conditions. If there are too many e-scooters on the street, it may encourage e-scooter hate and block pedestrians' path. A poorly managed e-scooter system will lead to more vandalism, create an eye sore, disrupt public space, and decrease the lifetime of shared e-scooters. The small lifetime of e-scooters is the main contributor to pollution from shared e-scooters and can be costly for the company. Furthermore, placing too few shared e-scooters will lead to a high rate of failure for customers trying to find an e-scooter. Finding the optimal number of e-scooters for the day or hour is necessary for shared e-scooter systems success and to keep the public happy.

The optimal number of e-scooters placed can be found from the average ratio of e-scooter trips to e-scooters. The average number of trips per e-scooter could also be used to find an optimal Another solution would be to change the predictor variable from number of trips to e-scooters placed. However, I did not have access to data regarding how many e-scooters were placed on the street for the day. With this information, the model could be improved to find the most accurate prediction for how many e-scooters should be placed that day. This research is meant to describe the most important features in shared e-scooter management and start a basis for e-scooter predictive models. The same model could also be used to predict daily e-scooter trips, daily number of e-scooters placed, or even utilization rates. By using predictive models, shared e-scooter fleet management can be improved and increase the likelihood of shared e-scooter becoming a permanent solution to current transportation problems.

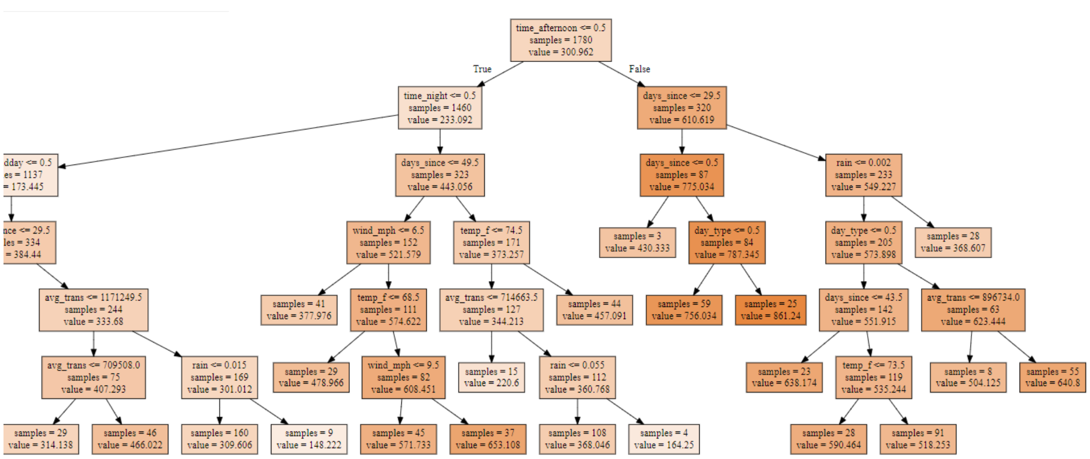
Figures

Figure 2.0: Regression Tree



Decision tree with a max_depth of 5 and min_samples_split of 50. This a small section of the larger decision tree.

Figure 2.1: Section of Decision Tree



Simplified decision tree. Hyperparameters are max_depth of 5 and min_samples_split of 50. This a small section of the larger decision tree so individual decisions can be seen.

Tables

Table 1.0: First Several Observations in Excel

id	num_trips	time	days_since
61504	1	earlymorn	0
61506	10	earlymorn	0
61507	38	morning	0

The table above contains the first 40 rows of data.

Table 1.1: Data Summary

	num_trips	days_since	day_type	holiday	temp_f	rain	wind_mph	avg_trans	ti
count	2374.000000	2374.000000	2374.000000	2374.000000	2374.000000	2374.000000	2374.000000	2.374000e+03	
mean	298.262005	53.566133	0.285594	0.026537	72.970514	0.006594	8.510531	1.281428e+06	
std	256.738616	31.129429	0.451792	0.160761	8.029915	0.044293	4.402775	3.370921e+05	
min	1.000000	0.000000	0.000000	0.000000	7.000000	0.000000	0.000000	6.207380e+05	
25%	50.000000	27.000000	0.000000	0.000000	67.000000	0.000000	6.000000	9.274770e+05	
50%	242.000000	54.000000	0.000000	0.000000	73.000000	0.000000	8.000000	1.444651e+06	
75%	491.750000	81.000000	1.000000	0.000000	78.000000	0.000000	11.000000	1.516955e+06	
max	1134.000000	107.000000	1.000000	1.000000	95.000000	0.960000	24.000000	1.666985e+06	

This data summary was generated in Python using the .describe() function. A time summary is not included in this image.

Table 1.2: Feature Correlation

	num_trips	days_since	day_type	holiday	temp_f	rain	wind_mph	avg_trans	time_afternoon
num_trips	1.000000	-0.191883	0.083573	0.047411	0.441032	-0.075867	0.326838	-0.027591	0.573052
days_since	-0.191883	1.000000	0.008724	-0.101021	-0.196155	0.064606	0.042397	0.097109	-0.001002
day_type	0.083573	0.008724	1.000000	-0.104393	0.007201	0.014826	0.027935	-0.181043	0.001507
holiday	0.047411	-0.101021	-0.104393	1.000000	-0.044443	-0.024586	-0.010219	-0.282177	0.003640
temp_f	0.441032	-0.196155	0.007201	-0.044443	1.000000	-0.037261	0.191654	0.007755	0.269229
rain	-0.075867	0.064606	0.014826	-0.024586	-0.037261	1.000000	0.015024	-0.033836	-0.016857
wind_mph	0.326838	0.042397	0.027935	-0.010219	0.191654	0.015024	1.000000	-0.051036	0.247891
avg_trans	-0.027591	0.097109	-0.181043	-0.282177	0.007755	-0.033836	-0.051036	1.000000	-0.001781
time_afternoon	0.573052	-0.001002	0.001507	0.003640	0.269229	-0.016857	0.247891	-0.001781	1.000000
time_earlymorning	-0.351743	-0.008194	-0.005887	0.001462	-0.272053	0.063286	-0.164781	-0.001305	-0.150732
time_latenight	-0.480759	0.010223	-0.001611	-0.015743	-0.252337	-0.033693	-0.261474	0.008148	-0.220560
time_midday	0.161249	-0.001002	0.001507	0.003640	0.281604	-0.005270	0.208206	-0.001781	-0.222451
time_morning	-0.243447	-0.001002	0.001507	0.003640	-0.046137	0.011618	-0.014274	-0.001781	-0.222451
time_night	0.251650	-0.001002	0.001507	0.003640	-0.049265	-0.003544	-0.057926	-0.001781	-0.222451

The table above depicts feature correlation generated in Python using the .corr() function. Correlation among different times of the day are not included because it is not insightful information.

Table 2.0: Sample for Predictions

	num_trips	days_since	day_type	holiday	temp_f	rain	wind_mph	avg_trans	aft	em	lan	mid	mrn	ni
1231	35	56	1	0	68	0.000	3	918556	0	0	1	0	0	0
1848	10	84	1	0	59	0.000	3	906925	0	0	1	0	0	0
772	12	35	1	0	78	0.000	10	886543	0	1	0	0	0	0
48	178	2	0	0	65	0.000	8	1424323	0	0	0	1	0	0
1144	8	52	0	0	69	0.005	5	1475853	0	0	1	0	0	0
610	726	27	1	0	80	0.000	9	1515902	0	0	0	0	0	1
202	156	9	0	0	72	0.050	11	1408425	0	0	0	0	1	0
840	183	38	0	0	71	0.000	5	1486495	0	0	0	0	1	0
1646	194	74	0	0	65	0.000	7	1509399	0	0	0	0	0	1
447	548	20	1	0	89	0.000	10	1085402	0	0	0	1	0	0
1452	16	66	0	0	73	0.000	0	1359419	0	0	1	0	0	0
829	662	37	0	0	70	0.000	10	1449490	0	0	0	0	0	1
233	842	10	0	0	81	0.000	10	1546968	0	0	0	0	0	1
1789	189	81	0	0	67	0.000	8	1624154	0	0	0	1	0	0
2170	499	98	1	0	75	0.020	11	929619	1	0	0	0	0	0
2352	20	107	0	0	63	0.000	0	1588492	0	0	1	0	0	0
998	355	45	0	0	77	0.000	9	1465901	0	0	0	1	0	0
1208	55	55	1	0	68	0.000	6	1436165	0	0	1	0	0	0
681	23	31	0	0	76	0.000	8	1483030	0	0	1	0	0	0
1669	69	75	0	0	71	0.000	11	1516786	0	0	1	0	0	0
486	18	22	0	0	67	0.000	14	693347	0	0	0	0	1	0
1765	98	80	0	0	71	0.000	16	1560073	0	0	0	0	1	0
2322	442	105	1	0	57	0.000	18	877979	1	0	0	0	0	0

A random sample of 1.5% of the total entries. The only columns included are features. This table was created in Python.

Table 2.1: APE for Sample

	dt	rf	knn	num_trips	dt_APE	rf_APE	knn_APE	time	days_since	day_type	holiday	temp_f	rain	wind_mph	avg_trans
0	30	34	41	35	14.29%	2.86%	17.14%	latenight	56	1	0	68	0.000	3	918556
1	30	26	24	10	200.0%	160.0%	140.0%	latenight	84	1	0	59	0.000	3	906925
2	10	14	6	12	16.67%	16.67%	50.0%	earlymorning	35	1	0	78	0.000	10	886543
3	423	386	463	178	137.64%	116.85%	160.11%	midday	2	0	0	65	0.000	8	1424323
4	30	32	33	8	275.0%	300.0%	312.5%	latenight	52	0	0	69	0.005	5	1475853
5	572	620	690	726	21.21%	14.6%	4.96%	night	27	1	0	80	0.000	9	1515902
6	196	148	202	156	25.64%	5.13%	29.49%	morning	9	0	0	72	0.050	11	1408425
7	196	190	183	183	7.1%	3.83%	0.0%	morning	38	0	0	71	0.000	5	1486495
8	368	351	505	194	89.69%	80.93%	160.31%	night	74	0	0	65	0.000	7	1509399
9	508	537	538	548	7.3%	2.01%	1.82%	midday	20	1	0	89	0.000	10	1085402
10	30	33	57	16	87.5%	106.25%	256.25%	latenight	66	0	0	73	0.000	0	1359419
11	653	599	490	662	1.36%	9.52%	25.98%	night	37	0	0	70	0.000	10	1449490
12	653	660	454	842	22.45%	21.62%	46.08%	night	10	0	0	81	0.000	10	1546968
13	310	275	267	189	64.02%	45.5%	41.27%	midday	81	0	0	67	0.000	8	1624154
14	369	380	506	499	26.05%	23.85%	1.4%	afternoon	98	1	0	75	0.020	11	929619
15	30	16	19	20	50.0%	20.0%	5.0%	latenight	107	0	0	63	0.000	0	1588492
16	310	313	250	355	12.68%	11.83%	29.58%	midday	45	0	0	77	0.000	9	1465901
17	30	33	48	55	45.45%	40.0%	12.73%	latenight	55	1	0	68	0.000	6	1436165
18	30	36	28	23	30.43%	56.52%	21.74%	latenight	31	0	0	76	0.000	8	1483030
19	30	33	37	69	56.52%	52.17%	46.38%	latenight	75	0	0	71	0.000	11	1516786
20	55	63	144	18	205.56%	250.0%	700.0%	morning	22	0	0	67	0.000	14	693347
21	196	188	78	98	100.0%	91.84%	20.41%	morning	80	0	0	71	0.000	16	1560073
22	504	496	429	442	14.03%	12.22%	2.94%	afternoon	105	1	0	57	0.000	18	877979
23	196	197	205	255	23.14%	22.75%	19.61%	morning	67	0	0	75	0.000	6	1457018
24	423	384	463	420	0.71%	8.57%	10.24%	midday	2	0	0	66	0.000	8	1424323
25	148	216	120	219	32.42%	1.37%	45.21%	morning	43	0	0	84	0.000	9	713285

Predictions for each sample observation. An absolute percentage error was calculated for each model. The order of the predictions is the same as the order of the sample data. This table was created in Python.

Table 2.2: Observations, Predictions, and APE for October 15

	rfoct	num_trips	rfoct_APE	days_since	day_type	holiday	temp_f
0	15.436477	1	1443.65	122	0	0	43
1	16.002157	3	433.41	122	0	0	43
2	16.157540	15	7.72	122	0	0	44
3	168.758541	81	108.34	122	0	0	46
4	169.038664	175	3.41	122	0	0	50
5	169.316635	168	0.78	122	0	0	53
6	169.038664	96	76.08	122	0	0	55
7	273.631677	99	176.40	122	0	0	55
8	273.631677	165	65.84	122	0	0	57
9	273.333738	175	56.19	122	0	0	60
10	278.616904	148	88.25	122	0	0	62
11	391.205245	198	97.58	122	0	0	62
12	557.974789	244	128.68	122	0	0	61
13	558.813400	286	95.39	122	0	0	62
14	384.054642	304	26.33	122	0	0	60
15	321.125225	194	65.53	122	0	0	53
16	311.496154	167	86.52	122	0	0	52
17	312.586154	103	203.48	122	0	0	50
18	299.279990	54	454.22	122	0	0	49
19	15.529179	17	8.65	122	0	0	48

The table above is created with pandas in Python. A random forest model is used to predict since it has the best accuracy of all machine learning methods. MAPE for this day of predictions 181.32%.

Bibliography

Ahmad, M., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production, 203, 810-821.

Barreto, L., Amaral, A., & Baltazar, S. (2018). *Urban mobility digitalization: Towards mobility as a service (maas).* Paper presented at the 2018 International Conference on Intelligent Systems (IS).

Chang, A., L. Miranda-Moreno, R. Clewlow, and L. Sun. (2019). *Trend or Fad? Deciphering the Enablers of Micromobility in the U.S.*. A Report of SAE International.

Clark, J., & Curl, A. (2016). Bicycle and car share schemes as inclusive modes of travel? A socio-spatial analysis in Glasgow, UK. *Social Inclusion, 4,* 83+.

Clewlow, R. R. (2019). *The Micro-Mobility Revolution: The Introduction and Adoption of Electric Scooters in the United States* (No. 19-03991).

Deka, D. (2004). Social and environmental justice issues in urban transportation.

DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of public transportation*, *12*(4), 3.

Fitt, H., & Curl, A. (2019). *E-scooter use in new zealand: Insights around some frequently asked questions.* University of Canterbury.

Helveston, J. P. (2017). Navigating an uncertain future for us roads. *Issues in Science and Technology, 34,* 21+.

Mason, H. (2019). *A comprehensive guide to electric scooter regulation practices.* Kansas State University.

McKenzie, G. (2019). Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C. *Journal of Transport Geography, 78,* 19-28. doi: 10.1016/j.jtrangeo.2019.05.007

Mitchell, G., Tsao, H., Randell, T., Marks, J., & Mackay, P. (2019). Impact of electric scooters to a tertiary emergency department: 8   week review after implementation of a scooter share scheme. *Emergency medicine Australasia*.

Portland Bureau of Transportation. (2019). 2018 e-scooter findings report (pp. 36). Portland, OR.

Quintana, C. (2019). Electric scooters hit speed bumps on campuses. *The Chronicle of Higher Education*, *65*, A49.

Shaheen, S., & Cohen, A. (2019). *Shared micromobility policy toolkit: Docked and dockless bike and scooter sharing*. University of California Berkeley.

Tol, L. (2017). *Mobility as a sustainable service? An exploration of the sustainable impacts of maas using a business model perspective.* Utrecht University.

U.S. Department of Transportation. (2018). *2017 National Household Travel Survey*.

Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences*, 20, 514-523.