Analyzing the Usage Patterns of Electric Bicycles

by

Ivan Rios

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2016

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Electric bicycles (e-bikes) are growing as an energy efficient alternative to public transportation, and cars. However, their adoption in North America has not increased at the same rate as in other countries like Germany, and China. Additionally, very few studies have been conducted to collect data about the patterns of usage of electric bikes. For these reasons, we distributed 31 e-bikes to staff, faculty members and students at the University of Waterloo in order to identify the main barriers and misconceptions about e-bikes, and obtain information about the riding habits of their users.

Each e-bike has a sensor kit that collects data about its location, battery state, acceleration, charging and discharging current, among other information. This data is constantly collected, therefore an algorithm had to be developed to analyze and identify trips. This algorithm is able to successfully identify 98% of the trips without relying on GPS data and is an alternative for field trials that have problems with the geolocation data. Additionally, we processed charging data of each battery in order to analyze the charging patterns of the participants in our study.

The results obtained provided insights about the differences between segments of the population, range anxiety of participants, riding and charging patterns. We identified weather as one of the main factors that influence the riding habits of participants. Additionally, we determined that e-bikes are mostly used for commuting. Similarly, when analyzing charging events, we found that most of participants generally start charging their batteries when the state of charge (SOC) is above 30% and in most of the cases, they prefer to charge until it is full. The results obtained in this thesis describe the main characteristics of riding and charging patterns of electric bikes, and can be used as an input for bicycle manufactures, policy makers, and utility companies to encourage the use of electric bikes in North America and prepare the infrastructure required for this transition.

Acknowledgements

Besides my advisors, I would like to thank the rest of my thesis readers: Prof. Keshav, and Prof. Hancock, for their insightful feedback, and also for the advice which encouraged me to widen my research from various perspectives.

Dedication

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents, Martin and Nary Rios whose words of encouragement helped me through this journey. Also to my brothers Martin, Renato and Eduardo who have never left my side.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

Electric bicycles (e-bikes) are bikes with a battery and an electric motor installed on them to provide assistance to the user. The level of assistance is controlled from a panel where the rider can select different configurations. Currently, electric bikes are being adopted in several parts of the world because of the convenience they offer to travel long distances through different types of topographies while avoiding traffic, keeping transportation expenses low, generating a small carbon footprint, and health benefits [13].

Even though this mode of transportation has many advantages over cars and traditional bikes, the adoption of electric bicycles in North America is still low, since they only represented 0.8% of the total bike sales in 2013, and 1.3% in 2014, in the United States [15]. These values are small compared to data obtained from Europe where in Germany, for example, 12% of the total bike sales are electric bikes [10]. For this reason, it is important to identify the barriers for the adoption of this technology, and also understand the potential behavior electric bike riders have, in particular the North American population.

Considering these facts, and the intention of North American nations to encourage people to use this mode of transportation in the next years [2], an eBike Field Trial was designed in order to collect data about the habits of people who have electric bicycles and be able to model their usage based on real data.

As of October of 2015, we have collected over 100 GB of data. This raw data required processing since it is collected even when the bike is not in use. For this purpose, we needed to develop algorithms that would allow us to extract meaningful information. The first analysis involved understanding how people use their bikes and the characteristics of each trip taken; therefore, we needed an algorithm that can accurately identify start and end times of trips. This algorithm is of great importance since the main purpose of this project is to analyze bike usage and the influence that factors like weather, gender, and occupation have on the riding habits of people.

This thesis describes the process followed to develop the trip detection algorithm along with the results obtained from the analysis of the data available until October of 2015.

1.2 Motivation

The number of electric bicycles sold around the world has increased during the past few years. They have become a great alternative for personal transportation in continents like Asia, and Europe. In contrast, North American sales of electric bikes are still very low. Additionally, there are very few studies that have collected data about e-bikes around the world. For these reasons, we decided to collect and analyze data from electric bikes.

For this study, we identified three main stakeholders who would benefit from our analysis:

1. Electric bikes manufacturers: Information about how electric bikes are used by different segments of the population could help to efficiently target different segments of the market with more inclination to acquire e-bikes.

2. Policy makers: There are three main topics of interest for this group. First, they need to understand the misconceptions that people have about e-bikes in order to create policies to encourage their acquisition. Second, from the point of view of road design, and safety, it is important for them to know the habits and characteristics of electric bike users. Finally, along with utility companies, they may need to create policies to help minimize potential negative effects of higher electricity consumption from charging the batteries.

3. Utility companies: It is important for them to understand what the patterns are that people follow to charge their batteries. This can be used to estimate the effects of electric bikes, and electric vehicles on the electrical grid.

In the particular case of this thesis, our focus is on usage and charging patterns that e-bike owners have. This will allow the stakeholders of this study to understand the habits of e-bikes riders and take actions to increase their use in North America. It is important to start the transition towards energy efficient modes of transportation, and e-bikes are a great option for this purpose.

1.3 Objectives

This section describes the main objectives for this thesis and gives a general description of future projects that will be developed using the results from this work.

- Develop algorithms to extract information from the raw data collected from e-bikes: for the work in this thesis, we focused on analyzing the usage of electric bikes which required us to identify trips taken by participants.
- Analyze the riding habits of participants in our project, and identify the main factors that influence their behavior: we want to understand what external factors affect the decision of using a electric bike on a regular basis.
- Identify the main differences between sections of the population: we want to determine the difference between habits of men compared to women, students vs staff and faculty members, in order to understand how demographic features affect the riding, and charging behavior of a participant.
- Analyze charging habits of participants: similarly as before, we want to identify the main factors that influence when participants charge their batteries, and also understand how demographic factors affect behaviors.
- Determine the relationship between the answers given in a survey with the actual usage of electric bikes: a survey was conducted before selecting participants for the study in order to determine the correlation between their riding habits and the answers collected.
- Obtain data to model the impact of electric bicycles at a larger scale: we want to determine the impact of the adoption of eBikes by creating models that would scale the effects generated in different areas of research such as transportation, electricity consumption, etc.

Along with these objectives, the work presented in this thesis is expected to facilitate opportunities to study the following topics in the future:

- Create models for range prediction: predicting models will be developed in order to estimate the distance remaining with different states of charge (SOC) or battery levels, considering external factors such as participant's aggressiveness, speed, etc.
- Compare electric bikes with electric vehicles (EV): the main objective of this project is to extrapolate the knowledge generated from electric bikes to EV's. With this, it would be easier, and cheaper to perform studies with a greater number of electric bikes in order to predict the habits of people who own EV's, and estimate their impact on different areas of research.

1.4 Technical Challenges

Since the conception of this project, several technical challenges have arisen. Hardware design, and the development of software for data collection were developed prior to the start of work presented in this thesis. However, by introducing data processing and analysis, other challenges were identified. Just as previous work served as the foundation for this thesis, it is similarly expected that the challenges described and solved in this thesis will serve as the base for future research. The following list presents the main challenges faced during the development of this thesis:

- As described in the next chapter, data is collected in cellphones that are located in the sensor kit that is installed on each e-bike. This data is transmitted, as plain text files, into a server that has a public IP address. Every hour this data is parsed, and put into a database; however, this recurring script had to be modified several times in order to handle problems with data.
- As the base for most of the analysis presented in this thesis, trip detection of e-bikes was a challenge considering the particular characteristics of this project. Even though there is a considerable amount of previous work in this area, factors such as GPS fidelity, data collection frequency, and accuracy of data collected forced us to develop an algorithm to accurately identify trips from our data.
- During the development of the data collection software, the phone that collects and stores data presented some problems. For this reason, the frequency of data collection is not evenly distributed in time; currently, four data points are collected during the last 10 seconds of every minute. This had to be considered when analyzing the data since it produced 50 seconds gaps of no data along with only 10 seconds of data (sampled 4 times) every minute.

- For similar reasons, the frequency of data collection had to be modified which became a challenge to face when developing the scripts that process raw data. Initially, the phones were collecting data every 0.5 seconds during the last 10 seconds of every minute (20 data points every minute), but this value went down to 4, as described in the previous point. These types of changes had to be considered in order to guarantee that our algorithm can perform efficiently with a different number of data points collected every minute.

- After starting to analyze the data, several problems were found specifically with GPS location data. Since bikes are generally stored indoors, GPS takes at least one minute to find satellites to determine the location of the phone. Additionally, several phones had sporadic problems when collecting this data which ended up causing a considerable amount of missing GPS points. Finally, the inherent noise of GPS data is a factor to consider when developing algorithms that depend on this information.

- When participants received their bikes, they agreed to provide information about their vacation times and other events that could potentially affect their riding frequency. Similarly, some bikes broke or had parts stolen, which affected the data collected from them. This information had to be considered when analyzing the data since it could have introduced a bias into the final results. All these factors were taken into account during the analysis in order to minimize the effects of this type of event on our conclusions.

- Bikes were given to each participant expecting to have only this rider using them throughout the duration of the project. However, some bikes changed owners since the original participants did not comply with the conditions of the project. Our analysis had to consider transference of ownership in order to correctly keep track of the users who took each trip on their assigned bikes.

- In order to obtain feedback and allow participants to access their own data, a web site was previously designed and implemented where several graphs and general information were displayed. This web site started presenting high latency due to the amount of data that had to be processed with each request from the users. For this reason, preprocessing routines were implemented in order to offer faster access to data in this portal, and encourage participants to give feedback about the results presented.

1.5 Main Contributions

The work presented in this thesis generated 3 main contributions. The first contribution is the trip detection algorithm which is the foundation for this thesis and other future projects, while the other contributions are related to the conclusions obtained from the data. Each one of these contributions is described in detail in the following sections.

1.5.1 Trip Detection Algorithm

Bike field trial projects generally rely on GPS data in order to identify trips taken by participants, or require their riders to keep a journal with all the trips taken on their bikes along with some additional information. Initially, we developed an algorithm that uses GPS to detect trips, however, it did not perform well on our data set. For this reason, we developed an algorithm that do not depend on GPS data, and can accurately determine the start and end times of trips from raw data. Our algorithm only depends on data from the phone's gyroscope sensor or similar movement sensors, and the charging current of the battery.

1.5.2 Riding Habits Stereotypes

There are stereotypes regarding gender and occupation, and how these demographic factors influence the behavior and habits of bike riders. After analyzing the data generated in this study, several differences were found between these groups of population.

Gender showed influence on the riding habits of people. As described in the following chapters, several differences were identified between male and female participants. Even though men and women tend to use their bikes (potentially for commuting) at the same time in the morning, men have a tendency to take their trips at later times during the afternoon and night, suggesting that they tend to work until later times of the day. Additionally, women tend to take shorter trips than men do. Finally, when analyzing the average speed of trips, men generally presented greater values.

Similarly, when analyzing and comparing students with staff and faculty members, similar differences were found. Participants with a full time job have a more defined schedule, mainly matching regular working hours of the day. Conversely, students take bike rides at more scattered times during the day, even showing a considerably higher amount of trips late in the night. In addition, students reach higher average speeds during their rides in comparison to their counterparts. Finally, staff and faculty members tend to take shorter trips compared to students, which is likely to happen because they commonly have access to more convenient modes of transportation such as cars.

1.5.3 Surveys for prediction of electric bikes usage

Several studies that aim to understand the effects of adoption of new technologies consider data obtained from surveys as their main source of information. Similarly, detailed surveys and interviews have been conducted aiming to predict the usage of electric bikes. For this reason, in order to analyze the correlation between the riding habits of participants and the results obtained from surveys, we asked participants to answer a questionnaire which had one section dedicated to describing the expected frequency of use of their bikes during winter and summer seasons.

This information was expected to be used as an input to predict the way people use their bikes, however, as shown in the following chapters, there was no relationship found between the answers and the data generated by each participant. These results suggest that surveys, with questions that ask participants to estimate usage of their bikes, should not be considered a reliable source of information. The e-bike field trial showed that the expected usage of the bikes, estimated by the participants, did not reflect the reality of the their actual riding habits.

Chapter 2

Project Overview

2.1 Description of the project

The e-bike project was conceived at the Information Systems and Science for Energy Research Group at the University of Waterloo. This project started in 2014 by designing and building the sensor kit, and writing the code for data collection. The sensor kit was designed considering all the information required to analyze the riding, and charging habits of participants.

The electric bikes were obtained from a supplier in British Columbia, Canada. After the bikes arrived, the sensor kit was installed on each battery. This sensor kit contains a phone that collects information and also works as the central computer of the system where all the data is stored. Every time the battery is on campus, the phone connects to the wireless network and transmits all the stored data to our servers.

After installing and testing the bikes, they were given to a group of people selected from within the university. The participants of this study were selected considering that they should at least use the bikes once a week in order to consistently obtain data to be analyzed. Additionally, all the participants are people who are“”’ expected to work or study at the university for at least three years; additionally, each participant was asked to bring the bike to school at least once a week to guarantee that the data collected is regularly transmitted to our servers. After these 3 years, participants will officially have ownership of the bikes and no more data will be collected from their bikes. The decision of giving the bikes away was made to encourage people to take care of the physical state of the bikes, and also use them regularly.

Also, in order to make the study more friendly for participants, no additional information is asked from them. Similar studies in the past have asked participants to keep a log of their trips and some general information that would be relevant to the study. However, this data does not always reflect reality, and can be overwhelming for participants to constantly keep track of it, therefore, we decided to develop algorithms that would automatically extract the information required for our study.

The field trial started in May of 2014 when the first bikes were given to participants. Even though data collection started at this time, some problems were detected that impacted the quality of the initial information collected. After identifying and fixing these problems, reliable data started to be collected during the month of September in the same year.

The e-bike project is expected to be the initial step of a set of projects that will eventually consist of a bigger fleet of bikes which will help obtain more accurate results. As mentioned before, these results are expected to provide information that could be related to Electric Vehicles. Currently, there are very few projects that have EV's on field trials because of the high costs that these studies incur into. Moreover, this data is generally not available to the public which makes it complicated to analyze the real habits of people who own EV's. E-bikes are expected to generate meaningful results that will later be used to better understand how EV's will be used, and the impact of their adoption.

2.2 Related Work

2.2.1 E-bikes Field Trials

Dozza and Fernandez [10] describe a platform developed to obtain data from traditional bikes. The objective of this project is to model the behavior of bike riders and study their usage. In a similar way as in our project, the bikes used for this study have sensors installed on them in order to obtain information about braking pressure, location, inertia, and images during participants' trips. The data obtained in this study helped the authors compare the behavior of users of traditional, and electric bikes. This study has similar objectives as our work, however, the authors focused on analyzing longitudinal, lateral and vertical accelerations to obtain information about braking patterns, reactions of participants during collisions, and behavioral analysis during trips.

In another paper, Dozza et al. [11] present a study in which they focus on determining the behavior of electric bikes users in order to understand their impact in traffic, and road safety. For this experiment, they collected data from 12 electric bikes during 2 weeks of use. Using this data, they focused on analyzing critical events where there was a risk of collision, or a crash was registered. This publication then compares the results obtained with data collected in a similar study that used regular bikes. The study determined that electric bikes are ridden faster, and the interactions of their riders with the surroundings are different from a regular bike user, which suggests that specific policies should be defined for e-bikes. This research obtained results mostly aiming to identify the causes for accidents involving electric bikes; they also obtained general statistics about trips taken during the study but did not perform an analysis on the riding habits of participants.

Similarly, this research group developed another study [22] where they compared several types of electric bikes with traditional bikes. In this case, they had 90 participants, representing three age groups, using the bikes during four weeks with additional sensors such as speedometers, and cameras. The data for this study was collected during four weeks. The results showed that electric bikes with motors capable to reach higher speeds presented higher average speeds. Additionally, they mention that this characteristic makes it more likely to have e-bikes riders involved in more severe accidents compared to riders who use conventional bikes. In contrast to our work, [22] mostly focuses on analyzing accidents and potentially threatening events for e-bikes riders, and they do not perform any analysis of riding habits.

To the best of our knowledge, there are no studies focusing on analyzing the riding habits of participants; most of the related work has focused on analyzing behaviors of participants in specific situations, and safety related interactions with surroundings during trips. Additionally, these publications used data collected during short periods of time (up to four weeks), with bikes that were temporarily distributed to their participants; for this reason, the data collected in these studies cannot be used to analyze patterns of usage of electric bikes which makes our data set unique and representative for this purpose.

2.2.2 E-bikes Surveys

Cherry and Cervero [8] performed a survey in two large cities of China where electric bikes are commonly used. This data was used to determine a generic profile of e-bikes users, their frequency of use, and the main factors that influence when using electric bikes. With the survey information, this study concludes that people with electric bikes generally ride longer distances, and they use their bikes as an alternative to public transportation (generally these users cannot afford a car). Particularly for China, they conclude that even though electric bikes are very popular, they have not been able to displace cars and more policies are needed to encourage their use. The results obtained in this study are different from the ones obtained in our study, most likely due to the differences between the analyzed populations. In contrast to the publication, we have participants who most likely do not belong to low income segments of the population and generally have access to private transportation, and public transportation. Additionally, the main factor identified for people to use their e-bikes in China was due to reduction of commuting times, which is not a limitation for our participants since staff, and students generally live close to campus.

Similarly, An et al. [1] conducted a survey in Shangai, China in order to identify the main user group of electric bikes. The authors collected data from 470 respondents and determined that most of the electric bike owners are in the middle and low income brackets. Additionally, they identified commuting as the main purpose for their bikes (47.2 %) but only in cases when the total time goes up to 40 minutes. Finally, they analyzed transitions between modes of transportation and determined that 56% of the e-bike owners used traditional bikes before, while 33% originally switched from public transportation. Similarly, the results obtained on this study are different to our results due to the different populations being analyzed and the different characteristics of the cities where the studies were performed.

Gordon et al. [21] describe the results obtained from interviews with 27 electric bike owners in Sacramento, United States. This study focused on identifying the reasons for choosing electric bikes, along with their usage patterns. The results showed that e-bikes owners tend to go longer distances for longer times compared to their previous experiences with traditional bikes; this was the main advantage identified by interviewees, while they mentioned safety and range anxiety as the main disadvantages. Regarding e-bikes usage, the results presented by the authors are estimations limited to ranges of values.

2.3 Participant Selection

In order to select the group of people who would receive an electric bike, a survey was conducted. Faculty, staff, and students from several departments in the University of Waterloo were invited to respond the questionnaire. The survey had four main sections:

- Prior knowledge: this section contains questions that determine how much participants know, and are interested in electric bikes. The intention of this section is to analyze the perceptions that people have about this mode of transportation.

- Transportation related opinions: this section contained questions about the opinion of each participant concerning different modes of transportation, and their perception about each one of them.
- Estimated usage: this section asked participants to estimate how often they will use the electric bikes, number of trips, and total distance expected per week.
- Background information: this section contains questions about the background of each participant, and general demographic information.

25 participants were selected by an external collaborator of the project in order to prevent selection bias, and the remaining 6 bikes were distributed among members of our research group. The group of participants was evenly divided between men and women, and faculty/staff members and students, in order to have similar amount of data from all these populations. In summary, the sample selected has 18 male and 13 females; also, out of all the participants 16 hold a faculty or staff position while the rest are students.

2.4 Technology Overview

The following sections describe the hardware and software components used and developed for this project.

2.4.1 Electric bikes

Electric bikes are similar to traditional bicycles but they have additional electronic components to provide assistance for propulsion [4]. The main components are an electric motor, and a battery that provides the energy required for the motor to function. This mode of transportation has become popular due to the diverse advantages that they offer.

First, due to the use of a battery, e-bikes are a gasoline-free way of transportation and are considered to run on clean energy. Currently, a main factor when traveling short distances is the emissions generated by the mode of transportation. For this reason, countries such as Germany [10], and China are encouraging the use efficient technologies that can reduce the overall emissions.

Second, due to the portability of the battery, e-bikes owners can ride their bikes for longer distances and duration without much additional effort. The e-bikes used for this study can assist for a distance of up to 45 km [12], and electric bikes in general go up to 50 km. Additionally, heavier weights can be carried which makes these bikes ideal for shopping, and similar uses. This advantage is particularly attractive to people who do not have access to cars since they can use their bikes for a wider range of applications.

Finally, even though electric bikes are more expensive than regular bikes, they are considerably more affordable than cars. As mentioned in Section 2.2, electric bikes are ideal for people who want a versatile, and convenient mode of transportation without having to spend large amounts of money.

2.4.2 Batteries

There are two main types of batteries commonly used for electric bikes, and vehicles. First, Lithium-ion (LiON) which is an energy dense battery commonly used in cases when it is the only source of energy [19]. The second type is Nickel Metal Hydride (NiMH) which is generally used in cases when the battery is part of a hybrid system like a Toyota Prius (combination of a gas engine and an electric motor).

A typical LiON battery contains copper foil coated with black graphite, the main material of the electrode. Additionally, there is a white electrolyte-soaked separator between the cathode and anode, which works as the path for ions to travel between electrodes while blocking electrons [5]. The battery stores energy by collecting electrons on the foil and it transfers energy using chemical reactions to send electrons and ions from one electrode to another [20]. Additionally, these types of batteries have a fuse to prevent them from catching fire. With all these components, LiON batteries can have different configurations.

The type of battery used for the e-bikes in this project is LiON, in a cylindrical configuration. Considering that space, and weight of components must be minimized for e-bikes, LiON batteries are ideal since they are lightweight and compact [5] with a considerable energy storage capacity that allow users to ride long distances.

2.5 Hardware and Software Details

2.5.1 E-bikes

The e-bike selected for this project is the 2014 eProdigy Whistler produced by the company eProdigy [12] shown in Figure 2.1. The bike has a motor maximum speed of 32 km/h which is limited by a controller, but can reach higher speeds when the rider pedals. The user has the option to choose one of five levels of assistance which will allow him or her to reach different speeds without necessarily having to pedal.

The retail price of this e-bike is 2,499 USD, which is approximately 10 times higher than the average price of a regular bike. The price difference is generally a factor for users to choose whether to buy a regular or an electric bike, however, by providing the bikes we can safely remove the bias introduced by their price.

The bike comes with a 37 Volts Lithium rechargeable battery [12]. This battery can be detached from the bike's body in order to be charged by plugging it into a regular outlet using an adapter. The battery takes approximately 5 hours to charge completely and is expected to have a life of 1000+ charge cycles.

Considering all the components, the bike has a total weight of 21 kg, including the 2.5 kg from the battery; this value is greater than an average bike. For this reason, and considering the advantage offered by the electric motor, it is expected to collect data during all the trips taken by the participants since the sensor kit was installed on top of the battery.

Figure 2.1: eProdigy Whistler bike

2.5.2 Sensors

The sensor kit, shown in Figure 2.2, contains a cellphone as the main component, which is a Samsung Galaxy III that runs Android Operating System and contains the data collection software installed as an application. The following fields are measured and recorded by the phone:

1. Time: the time stamp when the data was collected.
2. GPS: the geographical position coordinates calculated using information provided by satellites. Also, additional GPS data was collected using the cellphone network to estimate the position.
3. Angular speed: instantaneous angular speed measured using the Android API. Each measurement has 3 values corresponding to the axes of the phone. This data is obtained using the gyroscope from the phone.
4. Acceleration: instantaneous acceleration of the phone also considering gravity. Each measurement has 3 values corresponding to the axes of the phone.
5. Linear Acceleration: instantaneous acceleration of the phone without considering gravity. Each measurement has 3 values corresponding to the axes of the phone.

Additionally, the kit contains the following sensors:

1. Phidget Voltage Sensor to measure the battery voltage.
2. Phidget Current Transducer to measure the charging rate.
3. Digikey Current Transducer to measure the discharging rate.
4. Digikey Temperature Sensor to record the temperature inside the box.

As shown in Figure 2.2, all the circuitry and sensors are enclosed in an aluminum box. This box was hermetically sealed with plastic covers on the top, and bottom to protect the electric components while also allowing the phone to connect to the network and obtain GPS data.

Figure 2.2: Sensor Kit for Data Collection

2.5.3 Data Collection

The data is collected using the phone which runs the application developed for data collection. This software uses the standard Android API in order to access the sensors of the phone. Additionally, this code accesses the values recorded by the external sensors.

Initially, it was intended to collect data between fixed intervals of time. However, due to some complications found using the Android API, the phone was initially set up to collect one data point every half a second during the last ten seconds of every minute. During the execution of the project, some changes were made to this the software, and currently, the phone collects four data points during the last ten seconds of every minute. Each data point contains one reading from all the aforementioned sensors.

The phone saves this data into its internal memory. The data is stored in individual Comma Separated Values (CSV) files that contain one hour worth of data, and are indexed by date and time. The size of these files is very small compared to the total memory of the phone, therefore, it is possible to buffer data for several weeks. When the battery is at any location with Wi-Fi coverage in the University of Waterloo, the data files are transmitted to a server and deleted from the phone.

With this set up, the project is easily adaptable to the inclusion of new sensors. When new data is required, sensors can be installed in the box and the data collection software has to be slightly modified in order make this new information available in the data files.

This advantage was used before the beginning of the summer of 2015 in order to collect data about the discharge current of the battery.

Finally, in order to implement changes on the code of the data collection software without having to recall all the bikes, the application that contains this code can be upgraded automatically via Wi-Fi when there is a new version available on the server. This feature not only upgrades the main application, but it is also able to install a new application on the phone. This capability gives the opportunity to easily make changes and/or add new software to the phone.

Once the data is on the server, Python and Bash scripts are used to process the files. Once every hour, the scripts are run to parse the files and put the data into a MySQL database to allow further analysis. Additionally, other routines are executed hourly in order to identify the trips taken on each bike since the last time the phone transmitted data to the server. These scripts also confirm that data is being collected properly, and the users are consistently using their bikes.

2.6 Data Management

This project highly depends on the quality of the data collected from the bikes. For this reason, we wanted participants to take active part of our work by providing feedback of our analysis. In order to achieve this objective, we developed a web portal for participants to be able to access their data. The following sections describe the main parts implemented to achieve these objectives:

2.6.1 WeBike website

We developed a website in order for participants to see a summary of statistics about the data collected from their bikes. The first version was launched in 2014, and gave users the ability to generate plots from their data. This first version had high latency since each one of the requests was being processed in real time, and put a high load on the server due to the amount of data to analyze.

In order to offer a better user experience for participants, the website was redesigned and launched again by the end of September, 2015. The current version, shown in Figure 2.3 offers the ability to plot the total distance biked between a range of dates, the state of charge (SOC) of the battery over time on a specific date, and a map with the trips taken on a specific date.

Mapping trips was a challenging task considering that the GPS measurements were not reliable. For this reason, a GPS data cleaning algorithm had to be implemented so that a smooth, and more accurate trajectory could be displayed to the users. The algorithm implementation used is described by Katsikouli, et al. which only relies on the original signal (GPS in this case) to clean potential noise and determine a subset of data points that represent the actual trajectory [17]. This algorithm was suitable for our project since bikes do not necessarily follow routes defined in public databases, and therefore cannot simply be approximated to mapped roads.

In order to keep improving the website, other options are being developed for people to obtain information about the health benefits of biking, such as the calories burned, to encourage their participation and use of of the website.

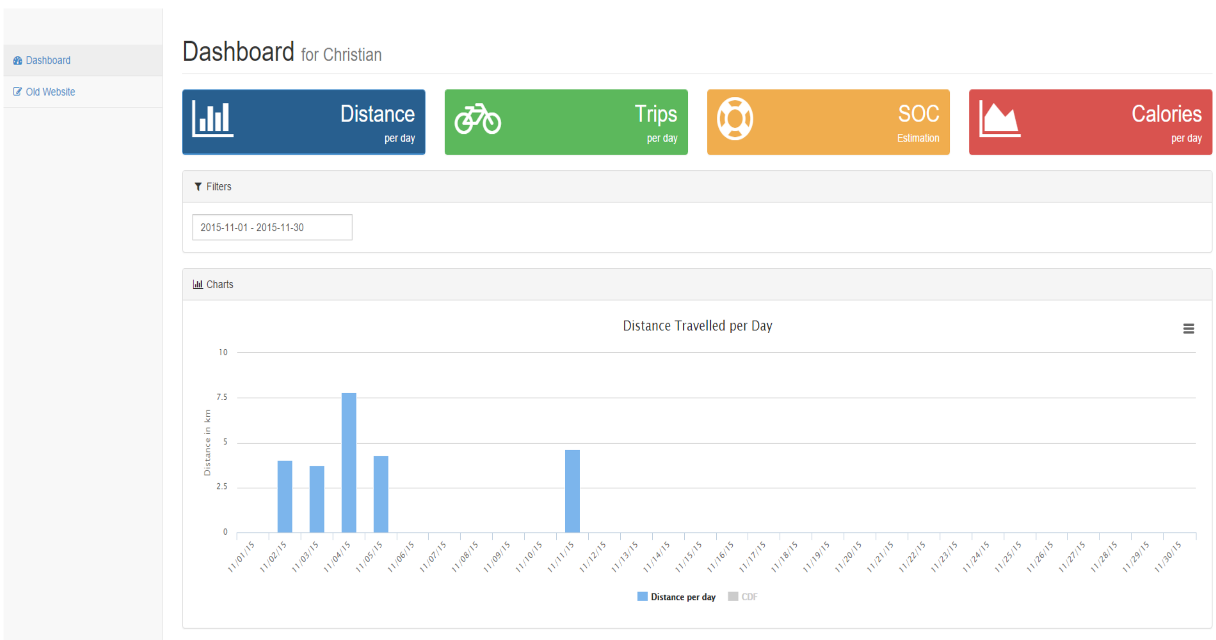
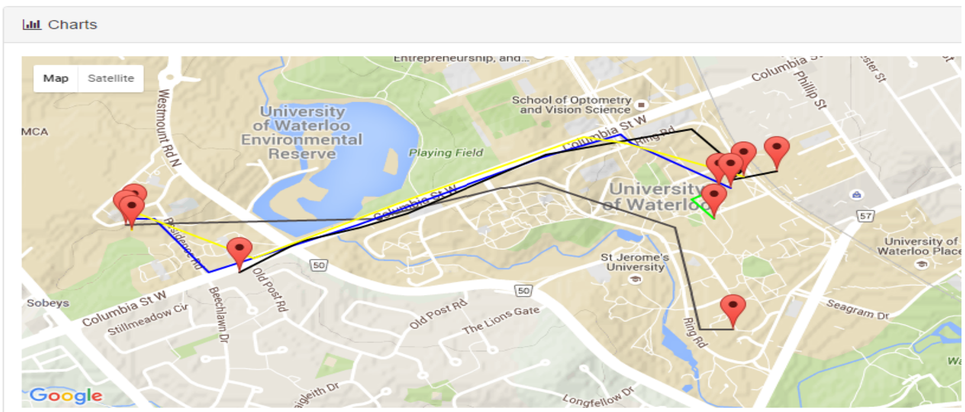


Figure 2.3: WeBike website

2.6.2 Feedback Collection

Once the website was launched, we started developing algorithms to analyze our data. In order to determine the accuracy of these algorithms, we decided to obtain ground truth from researchers involved in the project who also had an e-bike. Although this helped temporarily, we wanted to obtain more information from participants in a consistent manner. For this reason, we developed an option to obtain feedback from the users. Until this point, the main algorithm used to analyze data is the one for trip detection, therefore, the feedback module was implemented in the tab where users can plot their trips.

This implementation gives the option to confirm if trips were detected correctly, and also provide comments about each trip. Figure 2.4 shows an example of how participants can provide feedback about their trips; the table in the figure presents general information about each trip, such as start time, end time, total time, and distance while comments can be inputted on the last two columns.



Data

Show 10 entries

Trip #	Start time	End time	Time Taken (mins)	Distance (km)	Accurate?	Comments	
302	08:17	08:31	14.88	2.19	☐		Save
303	09:59	10:05	5.98	0.21	☐		Save
304	14:32	14:43	10.33	2.06	☐		Save
305	17:59	18:09	10.00	1.96	☐		Save
306	20:37	20:49	12.00	2.09	☐		Save

Figure 2.4: Feedback Collection

Chapter 3

Trip Detection

This chapter presents a description of the process followed to develop the trip detection algorithm. This work set the bases for the analysis presented in the following chapters, along with future research work. The first section describes the motivations for developing this algorithm, and the second section lists the main challenges faced while determining an optimal approach to solve this problem. The following sections show previous algorithms presented in related work with their shortcomings, and finally, a general description of the algorithm.

3.1 Problem Statement

Considering that data is permanently being collected from the e-bikes, even when users are not riding them, it was required to develop an algorithm that will take this information and determine specific periods of time when the bike was being used. Once those time periods were determined, we were able to develop analysis about how people use e-bikes, frequency, and other behaviors targeted in this project.

For this reason, and also as an essential part of future work to be developed for the analysis of this data, we had to determine a process to efficiently identify trips. This algorithm is an important milestone for our study and its development was expected to allow us to obtain accurate conclusions from the data while guaranteeing robustness against changes in the number of data points collected per minute, and GPS availability. Considering these facts, the following sections describe the approach selected in order to fulfill the described requirements.

3.2 Challenges

The algorithm presented several challenges due to the original requirements, and the specific characteristics of the data. It was expected to be robust and able to adapt to potential changes in the data collection process. Additionally, it was expected to not only rely on GPS data, but also be able to accurately identify the start, and end times of trips. The following list details the main challenges found:

- The inherent advantages of bikes hinder the process of analyzing the collected data: considering that bikes can be taken into different types of roads (not always mapped), can be stored both indoors and outdoors, can be carried on a car, can reach high speeds, among other features, it is difficult to define specific parameters or thresholds to characterize a standard behavior of this mode of transportation. For this reason, the trip detection algorithm has to be robust enough to consider these situations, and properly determine when a trip actually happened.

- Usage of GPS data: since this data was not accurate, contained several missing points, and was not available immediately after a trip started, its use was limited only for one parameter required by the algorithm. Specifically, GPS helped to determine the average speed during a trip. Moreover, due to the amount of noise present in this data, we also implemented another algorithm to calculate the total distance in a more accurate way.

- Correct identification of the mode of transportation: since bikes can be carried in buses, or cars, it was necessary to determine a methodology to identify these situations in order to properly filter them out.

- Identification of the approach to process the data: initially, we processed our data point by point and kept states which represented if a trip was active or not, however, this approach presented several disadvantages. After experimenting with different variations of the original algorithm, we decided to use a sliding window approach that generated higher accuracy. This approach relies on saving a fixed number of the last values analyzed and obtain conclusions based on the calculation of aggregated statistics such as average, number of points above average, and standard deviation; this is similar to calculating the moving average of a data set.

- Selection of variables to be used to efficiently identify trips: initially, we considered GPS data as the most suitable information to determine trips; however, this changed once we saw the low reliability of those measurements. After an analysis of the behavior of each measurement during trips, we determined that the Gyroscope and Linear Acceleration data are more suitable for this purpose.

- Definition of thresholds for each variable used in the algorithm: our algorithm uses the values from different sensors in order to make decisions. When the algorithm determines if a trip should start or finish, it calculates the number of values above a threshold for each variable (sensor) in order to decide if the behavior is actually a change of state or simply noise. Many trips had to be analyzed in order to determine the right value of the threshold of each variable. Even then, once we did an analysis of the entire database we were able to identify additional problems. Modifying these values has been challenging considering the different frequency at which data is collected, amount of data collected, variability of data, and different versions of the data collection software.

- Analysis of data collected during the last 10 seconds of every minute: This was a challenge for all the stages of the analysis since, in time series analysis, it is more common to have data points collected between fixed periods of time. Additionally, we started collecting 20 data points per minute, and eventually went down to collect 4 because of technical reasons.

3.3 Related work

This section describes the approaches used by other authors in order to define an effective algorithm to detect trips. Several publications were found in this area; the suggestions provided were used to develop, and improve our algorithm.

Chung et al. [9] describe a software tool developed to process GPS data from a personal GPS tracker (wearable), and identify routes and modes of transportation. The algorithm developed has a preprocessing stage of GPS data, where the system uses the number of satellites available in order to filter out invalid points. Additionally, the system uses Horizontal Dilution of Precision (HDOP), an index that describes how satellites are arranged in the sky at the time of the record, to calculate the quality of GPS data (lower angles between satellites cause lower precision). They use Map Matching to determine what route (streets and roads) was followed by the user. This technique consists in identifying the closest street to the observed GPS location using an algorithm proposed by Kim [18]. Similarly, trips are identified using parameters such as distance, and time between subsequent points. If these values are within a certain range, they will be identified as potential trips. They determine modes of transportation by using some characterization of each (bicycle, bus, car, walk) and additional features such as directionality of roads, transit network, waiting behavior, moving patterns (eg. speed and distance), transit vehicle stop pattern, etc.

The paper by Gong et al. [14] describes an algorithm that is able to reach 80% of accuracy when identifying the combination of different modes of transportation using GPS data in complex urban environments such as New York City. Their algorithm is able to identify any kind of combination of transportation modes. They mention that generally other algorithms used standard combinations for their analysis, for example bike-bus-bike; however, they did not use this approach. Again, the number of satellites available and HDOP are used as a primary tool to filter non-accurate data. Trips start, and end when a cluster of points stays inside or moves outside of a radius of 50 meters for over 200 seconds. Moreover, modes are detected using parameters such as speed, average speed, and distance. With their experiments, they estimated the warm start time of a GPS to be between 39s to 106 s. This is the time that the GPS takes to start calculating its location after being in places with no satellites access such as the subway.

Kasemsuppakor and Karimi's paper [16] describes an algorithm used to process GPS data in order to automatically identify walking routes and map them. This data is used to merge routes and create a pedestrian network in areas where this information is not available.The authors make a distinction between 7 different types of path segments: sidewalk, cross- walk, pedestrian walkway or footpath, accessible entrance, pedestrian bridge, pedestrian tunnel, and trail. Similarly as before, HDOP and the number of satellites available is used to filter inaccurate data. The authors use the bearing change in order to identify significant points (points with a high probability of determining the geometry of the walking path). Bearing change is the absolute value obtained from subtracting successive bearings using the great circle navigation formula [25]. This value is calculated between subsequent points, and the destination of the trip, and then it filters out the angles under a certain threshold. After filtering the significant points of a trajectory, they apply a clustering algorithm called Partitioning Around Medoids (PAM) to isolate outliers and clean the data.

Finally, the paper by Schuessler, Nadine and Axhausen [23] describes an algorithm that only requires raw GPS data as input, and is able to identify trips and modes of transportation. The study was done using data from approximately 5000 participants for an average of 6.65 days. For the data cleaning stage, they use a variation of HDOP which is called PDOP (Position DOP) that considers the position in a 3-d space. Additionally, they take into account the number of satellites available as a filtering tool. Besides those approaches, they also use altitude to estimate possible miscalculations of the GPS receiver; they suggest that errors in GPS data are generally accompanied by strong jumps in altitude.

Finally, trips are detected by analyzing the data that contains speed close to 0 and stay within a circle with diameter of 30 meters. This value is equivalent to 3 times the standard deviation of the measurement accuracy. They mention that most of the trips in literature are detected by detecting periods of inactivity (without considerable movement) between 45 and 300 seconds. Generally, the standard is to use 120 seconds. For mode detection purposes, the authors recommend using fuzzy logic because of the overlapping features of the different modes.

3.4 Background

In order to analyze the patterns of usage of electric bikes, we need to precisely determine the trips taken by participants. Several algorithms have been developed for trip detection, however, particular characteristics of this project forced us to develop a different algorithm for our data. First, as presented in the previous section, several algorithms consider that all trips are taken in mapped roads which is not the case of e-bikes. Also, the frequency of data collection is another difference with other data sets. Finally, due to the frequency of data collection there were several problems with the GPS since the phone was not able to start collection information immediately after a trip started. For these reasons, and as an essential first step to understand the usage of electric bikes, we developed the algorithm described in the following sections.

3.5 Initial versions of the algorithm

Before the final version, our research team worked on two other versions of algorithms for trip detection. At first, an algorithm was implemented using the approach suggested in similar projects described in our literature review. This approach used GPS data as the primary source of information to detect movement of the bike. However, GPS proved to not be a reliable source of information since it not only takes at least one minute to obtain data from satellites (which generates problems when detecting the precise starting time of a trip), but also takes additional time to stabilize the final location when the bike stops moving. This means that if the bike is stored indoors (which is generally the case), the phone will take 1-2 minutes to locate the satellites before generating the initial GPS points. Also, when the bike stops, and hence the trip is over, the location keeps changing (due measurement errors), making it complicated to precisely identify end time of trips. Finally, we detected a considerable amount of missing readings on our geolocation data.

These characteristics inherent to GPS data not only generate errors when identifying the times of a real trip, but also generate a high number of false positives. Moreover, real trips with missing GPS measurements would not be detected since the movement was not recorded. These two problems generated both a high number of trips detected that did not happen, and also a high number of trips that could not be detected.

The second approach we took was to combine GPS data with another sensor (magnetic field on the z axis) to determine when the bike was being ridden. Even though this approach improved the rate of detected trips, there were still several problems with the algorithm. In this case, the algorithm relied on GPS measurements again, and the magnetic field data was highly dependent on the direction of the bike when it was moving. Even though this second version took advantage of other available data, it was still producing errors due to GPS measurements. With this analysis we determined that it would be better to not use GPS data and define a new approach that will increase the accuracy, by adapting our solution to the specific data available in our study.

Finally, we concluded that analyzing data point by point caused the algorithms to be highly sensitive to measurement errors and noise, therefore, we decided to use a different approach. Similarly, the initial algorithms were sampling the data due to the amount of time that it took to analyze every measurement, however, this method generated loss of valuable information. To sum up, similar problems were detected in both algorithms which generated negative results; therefore, we concluded that a different solution should be developed considering all the information available in our data set.

3.5.1 Variable Selection

Since the sensor kit contains various sensors, we had different types of data to use when developing our algorithm. In order to determine which ones would have higher accuracy when predicting trips, we collected ground truth during 2 months from the bikes used by members of our research group in order to analyze the behavior of each one of the sensors while a trip happens. With this information, we analyzed the variations of the measurements from the sensors and confirmed which ones were more sensitive to detect trips.

The first step was to calculate the resultants of the sensors that had readings for axes X, Y and Z of the phone. Figure 3.1 shows how the resultant of the linear acceleration measurements varies over time; the graph shows the specific blocks of time when trips happened with red rectangles marking them. Similar results are presented in Figure 3.2 but for the resultant of the gyroscope measurements. These examples introduce the two variables used in the algorithm due to their sensibility when detecting trips. As shown in the plots, the values obtained over time clearly present a change of their average when the bike is being ridden while also showing very few outliers. Conversely, Figure 3.3 shows how the resultant of the magnetic field varies over time with the real trips marked using red rectangles again. In this case, the data collected from this sensor was not considered suitable to detect trips due to the lack of stability over time and the amount of noise detected.

Both Figure 3.1 and Figure 3.2 show some outliers. They may be generated when participants took the battery off from their bike to charge it, or due to measurement errors by the sensors. However, using the sliding window approach appropriately handles this problem since those types of outliers, detected for short durations, are automatically filtered out.

After this analysis, we concluded that there are three essential characteristics to consider when selecting the sensors that will provide data for the algorithm. First, the data should be stable, which means that it should be quick to present changes in its values when there is a trip happening, and it should also go back to the stable state as soon as a trip is over. This feature will allow the algorithm to accurately identify start and end times. Second, the data from the sensor should have clear thresholds that can be defined in the algorithm in order to identify when the bike is being used. These thresholds will be used in the algorithm to determine when movement is detected from the bike. Finally, the sensors should have minimal noise in their measurements so that outliers can be easily detected and filtered out.

We analyzed the data from all the sensors that provide relevant information for trip detection. This data considered a period of time of three weeks, and was obtained from the bikes used by our research team since we had access to the information about all the trips taken using them. This was used to identify the sensors which have constant (stable) values registered during times without trips, have clearly defined thresholds to identify movement, and have the lowest average number of outliers registered per hour. Stability was determined using visual inspection, and it was classified into three categories: high, medium, and low, where high stability means the values registered during times with no trips do not vary from the stable state. Each threshold's definition was also determined using visual inspection, where high threshold definition means there is a clear increase in the measurements of the sensor when a trip starts, and also the measurements immediately decrease to the stable state after a trip is over. Finally, the average number of outliers per hour was calculated by identifying the data points that presented high values during times when no trips actually happened. Table 3.1 presents the results obtained.

Sensor	Stability	Threshold definition	Average number of outliers
Gyroscope	High	High	4
Linear Acceleration	High	High	7
Acceleration	Medium	Low	4
Magnetic field	Low	Low	50

Table 3.1: Analysis of sensor data for trip detection algorithm

From Table 3.1 we can see that the most stable sensors are gyroscope and Liner acceleration sensor. Also, these two sensors have highly defined thresholds and low number of outliers. For these reasons, we concluded that the two sensors with higher accuracy are Gyroscope, and Linear Acceleration sensor.

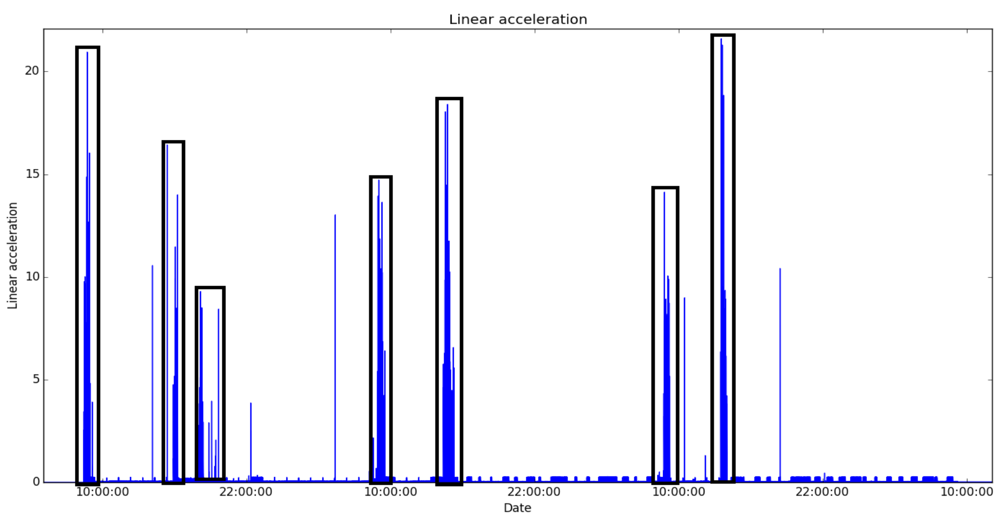


Figure 3.1: Linear acceleration resultant during trips

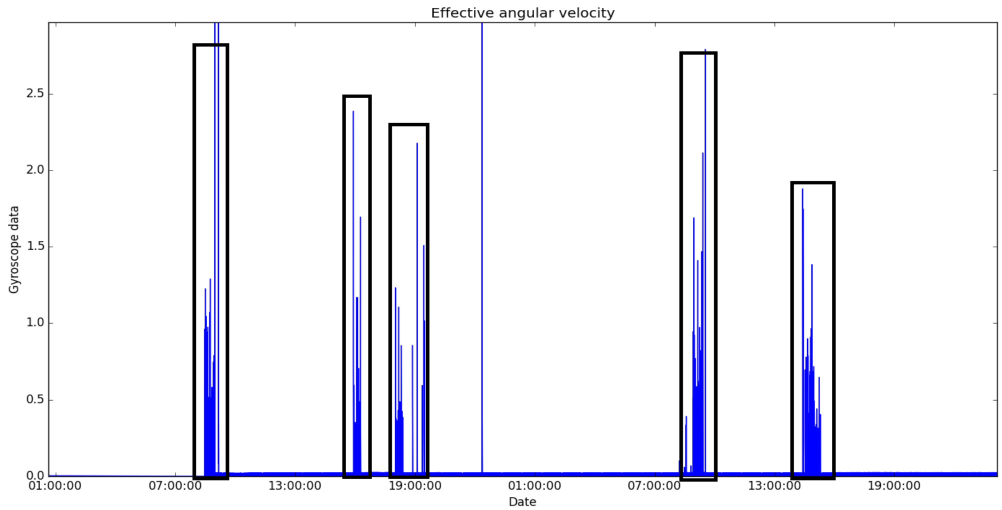


Figure 3.2: Gyroscope resultant during trips

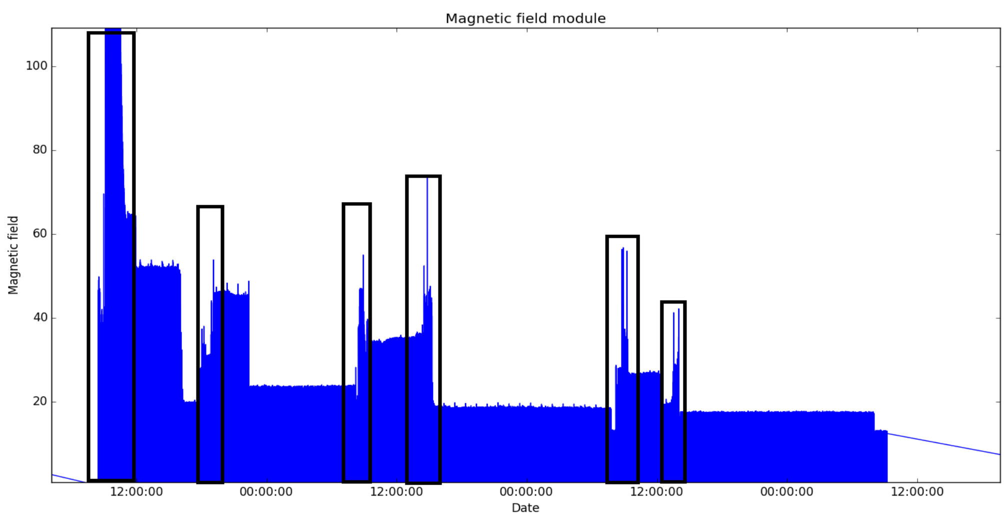


Figure 3.3: Magnetic field resultant during trips

3.6 Algorithm description

In order to minimize the effects of noise and outliers in our algorithm, we used a sliding window approach that keeps track of the average, high values, low values, standard deviation, among other features, of the last 12 values processed from each sensor. The width of the sliding window was selected considering that less than 8 values makes the algorithm sensitive to noisy data which leads to lower accuracy. Similarly, widths greater than 18 would not allow the algorithm to properly identify trips that lasted 5 minutes. Therefore, we tested the algorithm using a width between 8 to 18 values (equivalent to the time range of 2 to 4.5 minutes); we determined that the accuracy was slightly better with a value of 12. The use of a proper width helped us to detect random variations, such as the outliers detected in Figure 3.1, from real changes in the behavior of the variables analyzed.

Additionally, simplify the calculation of the average of values in the sliding window, and other statistics mentioned before, we created an object that handles these tasks, and also keeps track of the number of values above the defined thresholds. In the pseudo code, these objects are represented with the class name "variable".

Each one of these objects has a threshold defined, which is the minimum value for each sensor to consider the bike to be moving. These thresholds were selected by determining the middle point between the average of the measurements of each sensor during times with no trips, and the average during trips. With this approach, the threshold is equally distant to both the stable state of measurements when the bike is moving and when it is static.

Similarly, a limit for the number of measurements above the threshold of the variables was defined by experimenting with values within the range of 2 to 10. The same methodology as before was used by choosing the value that returned higher accuracy. In a similar way as before, low values will cause the algorithm to have problems when the variable has many outliers, while high values (close to 12) would decrease the sensitivity of the algorithm to detect trips. After evaluating the results with each number in the range, the value selected as the limit was 4.

Additionally, we used the charging current data, since no trips can happen during charging events (the battery must be plugged in to an electricity outlet to start charging). The algorithm uses this information to constantly verify that there is no charging events happening. With all this information, we were able to not depend on a specific variable, but instead, the algorithm uses all the data available to make the best decision. Using these objects that represent the data from specific sensors, we detect all the trips that happened during a defined period of time.

After identifying all the potential trips, we merge consecutive trips that have less than 5 minutes of time between them. Finally, using some filters defined from suggestions found in our literature review, a final list of trips is returned which contains all the trips that had a duration longer than 5 minutes, a total distance over 300 meters, and average speed of under 25 km/h.

The following section presents the pseudo code used to develop this algorithm; this code exemplifies the algorithm using only two variables: gyroscope, and linear acceleration variables:

```
# Definition of the class variable:
class Variable:
        # Array that contains the last processed data points
        values = list()
        # Variable that contains the number of values greater than threshold
        number_of_values_above_threshold
        # Variable that contains the threshold defined for this object
        threshold
        # variable that contains the width of the sliding window
        width

        # Method to determine if the battery is being charged
        def is_charging:
                if average(values) > 20:
                        return true
                return false

        # Method to determine if the bike is moving by counting the number of data points in the sliding window that are above the threshold and comparing the average of those values with the threshold
        def is_moving:
                if number_of_values_above_threshold > MIN_NUMBER_VALUES_OVER_THE_LIMIT and average(values) > threshold:
                        return true
                else:
                        return false

        # Method to obtain the earliest time when movement was detected from the data stored in the variable
        def get_start_time:
                for record in variables do:
                        if record.value > threshold:
                                return record.time_stamp
                done
```

```

# Function to merge subsequent trips
def merge_trips(i):
        # replace the end time of the first trip for the end time of the next trip
        end_times[i] =  end_times[i + 1]
    # add the distances of both trips
    distances[i] = distances[i] + distances[i + 1]
    # remove the second trip
    del start_times[i + 1]
    del end_times[i + 1]
    del distances[i + 1]

# Definition of variables:
CHARGING_CURRENT_LIMIT = 25 # Threshold for charging current
GYRO_LIMIT = 0.09 # Threshold for resultant of 3 gyro components
LIN_ACCEL_LIMIT = 1 # Threshold for resultant of 3 linear acceleration components
MIN_NUMBER_VALUES_OVER_THE_LIMIT = 4 # Variable that defines the minimum number of values over the threshold to consider that there is movement.
WINDOW_WIDTH = 12 # Number of values sliding window will keep
MIN_TRIP_LENGTH = 300 # Minimum number of seconds to be considered a trip
MAX_TIME_BETWEEN_TRIP = 280 # Maximum time with no movement within a trip
MAX_AVG_SPEED = 20 # Maximum average speed in a trip

trip_started = False # Variable that keeps track if a trip has been started
# Variable that contains the charging current data
charging_currents = Variable(width=WINDOW_WIDTH, threshold=CHARGING_CURRENT_LIMIT, number_of_values_above_threshold=MIN_NUMBER_OF_VALUES_OVER_THE_LIMIT)

# Variable that contains the gyroscope data
gyroscopes = Variable(width=WINDOW_WIDTH, threshold=GYRO_LIMIT, number_of_values_above_threshold=MIN_NUMBER_OF_VALUES_OVER_THE_LIMIT)

# Variable that contains the linear acceleration data
lin_accels = Variable(width=WINDOW_WIDTH, threshold=LIN_ACCEL_LIMIT, number_of_values_above_threshold=MIN_NUMBER_OF_VALUES_OVER_THE_LIMIT)

start_times = [] # List of starting times
end_times = [] # List of ending times
distances = [] # List of distances

raw_data = obtain_data_from_db() # Calls DB to obtain data
### PHASE 1 ###
```

```
# Obtaining all trips in the data
for each record in raw_data do:
        # Adding data to the corresponding variables
        charging_currents.push(record.charging_current)
        gyroscopes.push(record.gyro)
        lin_accels.push(record.lin_accel)

        # Case when trip has not started
        if not trip_started:
                #Check no charging event and movement from gyro/linear acceleration
                if not charging_currents.is_charging() and (gyroscopes.is_moving() or lin_accels.is_moving()):
                        trip_started = true
                        # Obtain earliest time when movement was detected
                        tmp_start_time = get_start_time()
        # Case when trip has started:
        else:
                trip_started = False
                # Save start and end times
                start_times.append(tmp_start_time)
                end_times.append(get_end_time())
done # end of for loop

# Calculate the distances of all the trips detected
distances = calculate_distances()

### PHASE 2 ###
# Process the detected trips and merge trips if the end of the first is less than 280 seconds before the beginning of the next trip
i = 0
while i < length(start_times) do:
        # Obtain seconds between end of trip and start of next trip
        if (start_times[i+1] - end[i]).total_seconds() > MAX_TIME_BETWEEN_TRIP:
        # merge trips and add total distances
                merge_trips(i)
    i = i + 1
done # End of for loop
# Apply filters of average speed, total distance, and duration
validate_all_trips()
# Return results
return start_times, end_times, distances
```

3.6.1 Trajectory Cleaning

As shown in the algorithm, one of the final filters for trips uses the total distance of the trip in order to calculate the average speed. This value is compared with a the threshold defined of 25 km/h; all the trips that have greater average speeds are deleted. When there is GPS data available, this is used to recognize the mode of transportation since bikes can be carried on buses or inside cars, and average speeds greater than 25 km/h are not characteristic of a bike ride. However, as mentioned before, GPS data is not completely accurate which generally ended up inflating the total distance computed by summing all the distances between points.

For this reason, we implemented an algorithm for trajectory cleaning. The algorithm is described by Katsikouli et al. [17] and its purpose is to reduce large volumes of GPS data by identifying the most important points that can be used to describe the trajectory. Figure 3.4 shows the results obtained after applying the trajectory cleaning algorithm on four different trips drawn on the map.

This algorithm approximately reduces the total number of data points collected during a trip by 90%. The examples shown in Figure 3.4 originally have more than 150 data points collected, but the algorithm determines up to 15 points to characterize each trajectory. Additionally, in cases where few GPS points were collected during a trip, the algorithm does not decrease the total number of points; this is an important feature since it will guarantee to use all measurements when data is scarce. With this solution, we were able to determine the distances more accurately, and plot trips on the website where participants can access the data collected from their bikes.

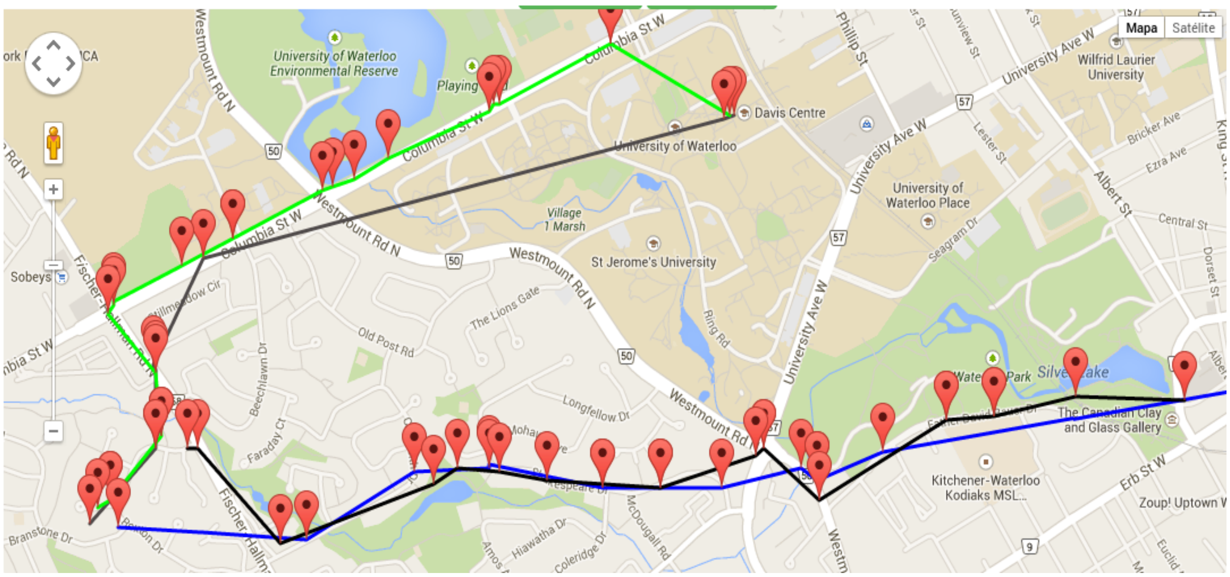


Figure 3.4: Resulting routes after applying trajectory simplification to four sample trips

3.7 Evaluation

In order to evaluate the final version of our algorithm for trip detection, we compared the results with the two previous algorithms detailed in Section 3.5. The first algorithm, Algorithm 1 in Table 3.2, is a generic implementation described in our literature review which uses GPS as the main feature. The second algorithm, Algorithm 2, is a modified version of the first algorithm with an additional feature: magnetic field measurements.

Table 3.2 presents the results obtained from this comparison. Here, Duration Accuracy refers to the total duration of trips correctly detected divided by the sum of durations of real trips. Precision is the total number of correctly detected trips divided by the total number of trips identified. Finally, false positives is the total number of trips detected that were not actual trips, and false negatives is the number of trips that were not detected by the algorithm.

Using these values, we can estimate the accuracy of the results of each algorithm in three main areas of interest. First, we want to guarantee that all (or most) of the trips are detected. Moreover, these trips should not only be detected, but should also have accurate start, and end times. Lastly, an important aspect we were concerned about is to prevent the identification of false trips detected due to noise. These three aspects are highly related to the variable selection section presented before.

The following results were obtained using 225 trips taken during May, and June of 2015.

Algorithm	Features	Duration accuracy	Precision	False positives	False negatives
Algorithm 1	GPS	65%	89%	18	79
Algorithm 2	GPS, Magnetic Field	69%	94%	11	47
Proposed algorithm	All (hierarchized)	95%	99%	2	5

Table 3.2: Evaluation of trip detection algorithms

After analyzing Table 3.2 and the data collected during the aforementioned dates, we determined that the false positives and false negatives obtained by applying our algorithm were caused due to random noise, and trips with durations close to five minutes. Our approach aimed to minimize the effects of noise, however, there are sporadic outliers that could be causing our algorithm to have false positives. Similarly, trips that have a short duration can be filtered out by mistake because of the frequency at which we collect data; if a trip lasted 5 minutes, and the phone did not start collecting data until the end of the first minute of the trip there is a chance it could be filtered out in the final validation of trips in our algorithm.

As shown in Table 3.2, our algorithm performed better than the initial versions in all the evaluation metrics. Additionally, we are constantly testing it with new data, and so far it has presented similar results. We recommend using our algorithm in cases when GPS data has low fidelity and it contains a high percentage of missing measurements. This algorithm presents a different way to identify trips using data from other types of sensors, and a different approach which decreases the effects of outliers and noise in the data.

Chapter 4

Data analysis

This chapter presents the analysis performed on trips, charging events, and survey data. Trips are the events identified with the algorithm described in Chapter 3. Charging events are periods of time when the batteries were charged; more details about this are presented in Section 4.3.

Three main stakeholders were identified for this study: Electric Bikes Manufacturers, Policy Makers, and Utility Companies. The results shown in this chapter can be used by these stakeholders for different purposes; each section presents the potential application of our conclusions.

For the sections in this chapter, we compared the distributions of the different segments of the population. For continuous distributions, we performed t-tests, since the standard deviation of the population is unknown and both sample were random, in order to determine if the means of both sets of data are significantly different. In this case, the null hypothesis is that both distributions are equal and the alternative hypothesis is that they are different. The value selected for the significance level is 5%.

4.1 Background

After defining a trip detection algorithm, the next step of our project was to analyze the main usage patterns of e-bikes such as average speed during trips, duration of trips, among others. This information, along with similar data obtained from charging events, is used to describe the main characteristics of e-bike users and their habits. Considering that, as described in the Related Work about e-bikes section, there is no information available about e-bike patterns of usage, we expect that the results presented in this chapter will be the starting point to develop models about e-bike usage. This data can be used by researchers to study the specific effects of e-bikes adoption in different fields (e.g. mobility planning), and can further be utilized by the aforementioned stakeholders in order to successfully increment the number of energy efficient vehicles used in North America.

4.2 Trips

4.2.1 General Statistics

Table 4.1 presents the summary information about the trips detected, it shows that a total of 4668 trips were detected as of the end of October 2015. Also, we can see some differences between the populations analyzed. On average, there is an 18% difference between the number of trips taken by students and staff/faculty; similarly, female participants took 16% more trips than their male counterparts. Regarding total time of trips, there is no difference between male, and female participants; however, students accumulated 15% more minutes, on average, than staff/faculty participants.

Description	Total	Female	Male	Students	Staff/Fac.
Total Trips	4668	2124	2544	2470	2124
Average trips per participant	156	163	141	164	137
Total duration [min]	78426	33712	44714	40778	37648
Average duration per participant [min]	2529	2593	2484	2718	2353
Average duration per trip [min]	16.8	15.8	17.6	16.5	17.7

Table 4.1: General statistics of trips

These initial observations lead to have an idea of the potential differences between the populations part of the study. To confirm these findings, a more detailed analysis is presented in the following sections focusing on specific parts of the data. These sections use the frequency of events to obtain conclusions from the data; frequency is presented in the y-axis of the plots and it represents the normalized number of events detected for the ranges shown in the x-axis.

4.2.2 Distribution of Trip Start Time of the Day

Initially, we analyze the hour of the day when each trip started. This data will help us to estimate the time when trips are generally taken and identify the differences between some segments of the population. This information can be used by policy makers to predict times of the day when there will be more congestion of electric bikes, and create plans that focus on mobility, and road safety during peak hours. Additionally, bike manufactures can offer extra accessories for their target customers based on their patterns of usage; for example additional lightning can be offered to riders who use their bikes during evening times.

Figure 4.1 shows the distribution obtained from all the data; here, we can see that there are two peaks happening at 8 am, and 4 pm which is consistent with the start and end times of a regular business day. Additionally, we can see that participants take trips between 7 am and 8 pm, and there are a minimal number of trips happening during the remaining hours of the day. This graph suggests that electric bikes are mostly used for commuting purposes, and they are not considered as an alternative for transportation during night hours.

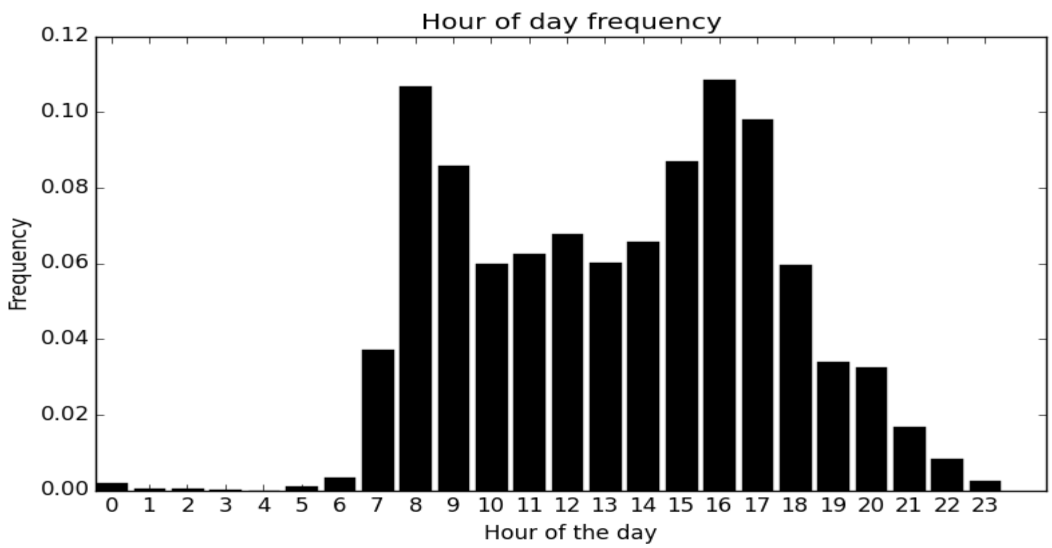


Figure 4.1: Distribution of hour of the day when trips started

Figure 4.2 shows the distribution of starting times of trips for men and women separately. Initially, when this analysis was first performed, clear differences were identified between these two groups. However, after we collected more data during the summer of 2015, the differences became less clear. This plot shows no visible difference between the populations besides a reduced number of trips taken by women after 5 pm compared to male participants.

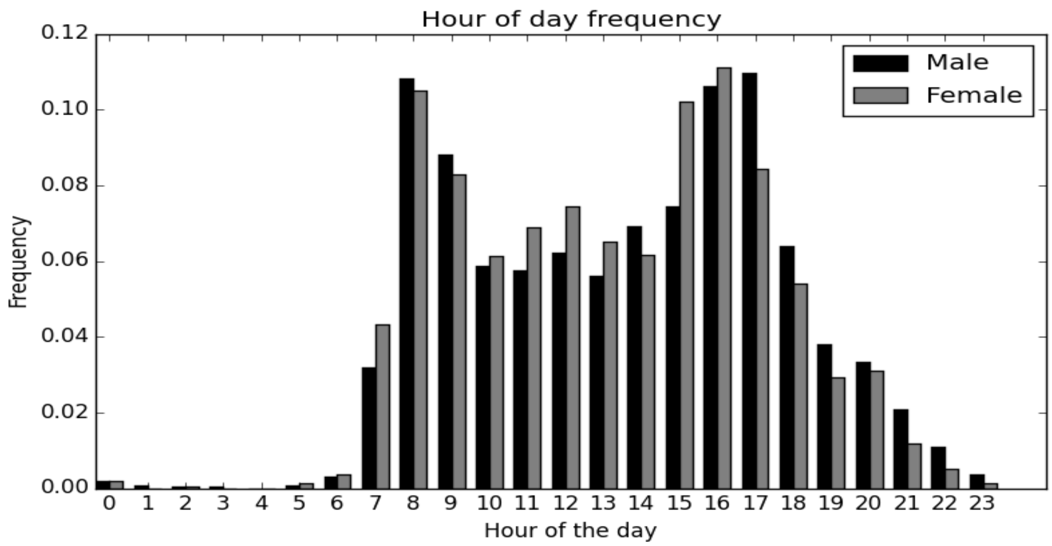


Figure 4.2: Distribution of hour of the day when trips started divided by gender

Finally, Figure 4.3 presents some differences between students and staff/faculty members. This plot shows that staff and faculty members use their bikes mostly at the beginning and end of business days, while students have a flatter plot with less defined peaks. Additionally, we can see that students have a higher number of trips happening after 5 pm compared to staff/faculty members. This is consistent with a more defined schedule that people with a full time job follow compared to students who generally have changing schedules and tend to work until later times in the night.

As identified in [8], [1] and [21], electric bikes are generally used for commuting to work. This is consistent with the results obtained, since the highest frequencies of trips match with the regular start and end of business hours.

4.2.3 Distribution of Trip Duration

This section analyzes the frequency of the duration of each trip divided into 5 minutes intervals. Using this information, bike manufactures can estimate the proper battery size

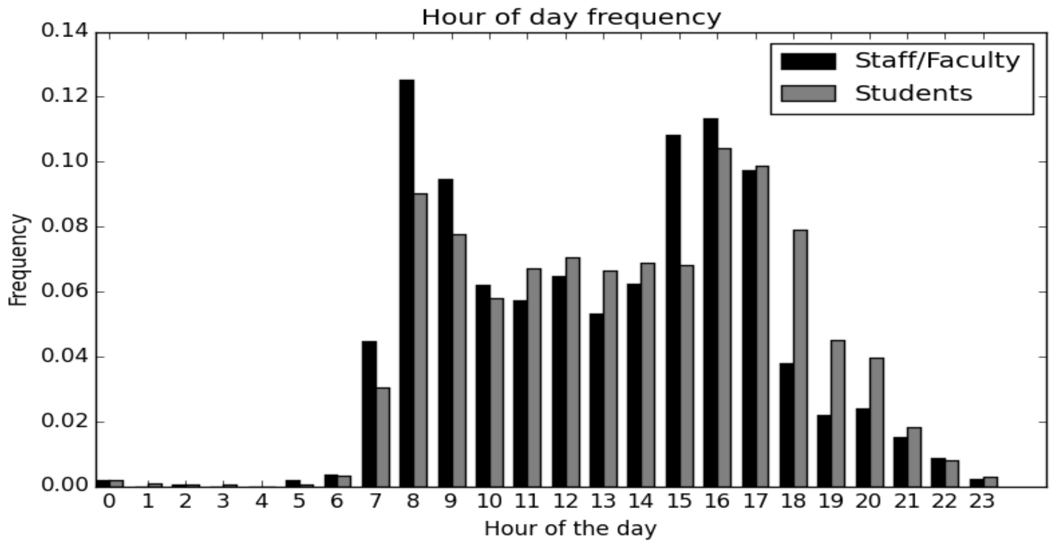


Figure 4.3: Distribution of hour of the day when trips started divided by occupation

for specific segments of the population, and adjust the prices accordingly. This can help manufactures to adapt to the actual necessities of users while offering a more diverse range of e-bikes prices to choose from. Additionally, policy makers can estimate the amount of bike lanes required for users to reach popular destinations in cities, by considering the maximum duration they are generally willing to bike.

Figure 4.4 shows the distribution of frequency in black, and the cumulative distribution in gray; this plot was obtained using all the trip data available. This graph shows that 70% of all the trips last less than 20 minutes, while 90% last less than 25 minutes. Additionally, we can see that most of the trips last between 5 to 20 minutes. These results might be due to the fact that students and staff/faculty members generally live close to campus, and therefore, most commuting trips are short. Additionally, it is important to consider that Waterloo is not a large city, and traffic is not commonly a problem which might influence the decision of choosing other modes of transportation when the trips are longer.

Similarly, Figure 4.5 presents the distribution of duration divided by the gender of participants. In this case, we can see that female participants have more trips that last less than 20 minutes compared to their male counterparts. These differences suggest that men take longer trips than women, although this conclusion may need further analysis.

Figure 4.6 was obtained by dividing the data based on the occupation of participants (staff/faculty and students). This plot shows that there is a slight difference between these

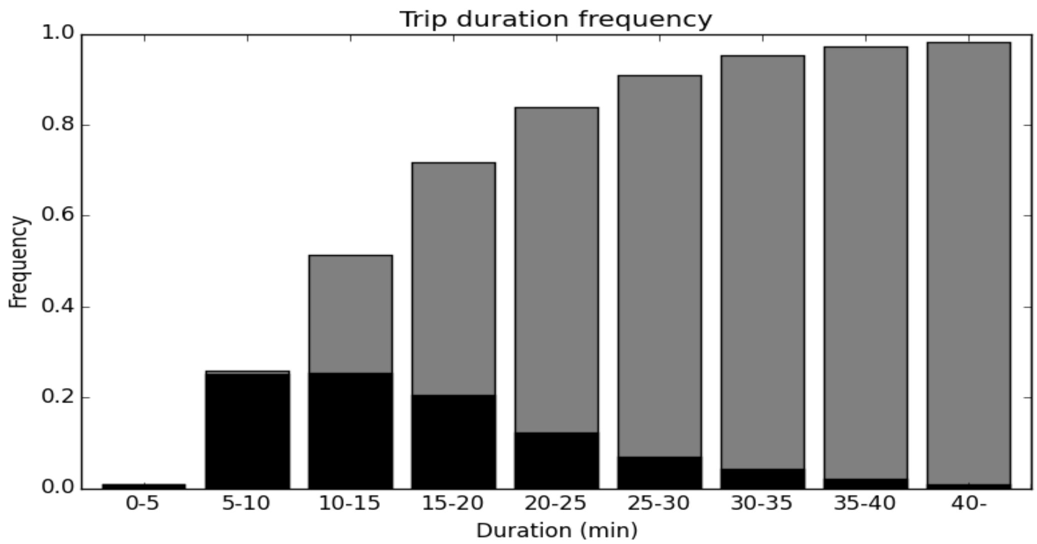


Figure 4.4: Distribution of trip duration

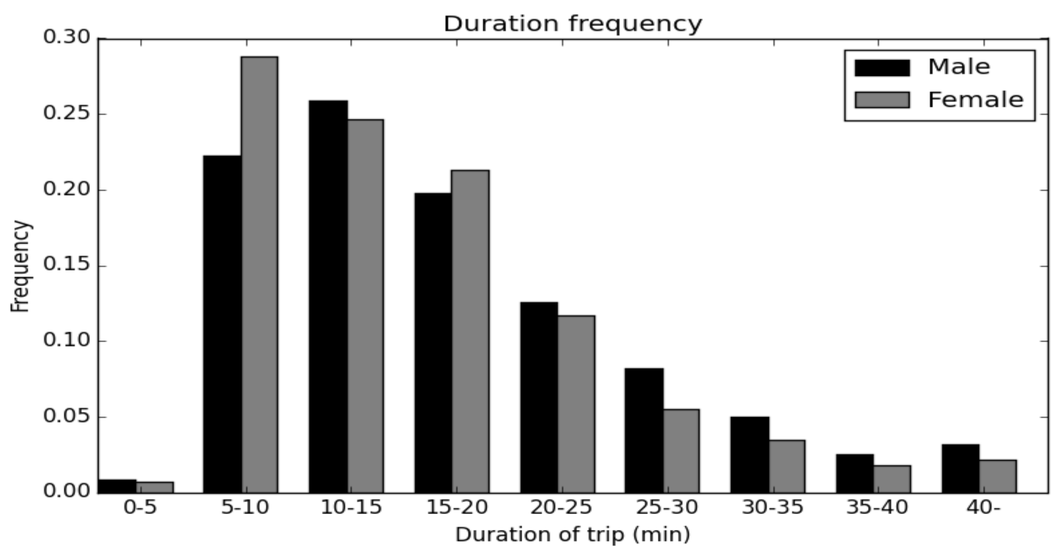


Figure 4.5: Distribution of trip duration divided by gender

two populations where students seem to take shorter trips than staff and faculty members. This conclusion needs to be verified with a larger population of participants, since students in Waterloo generally live closer to school which might be causing the mentioned results. Our findings were confirmed by the t-test which rejected the null hypothesis with a p-value of 0.0001.

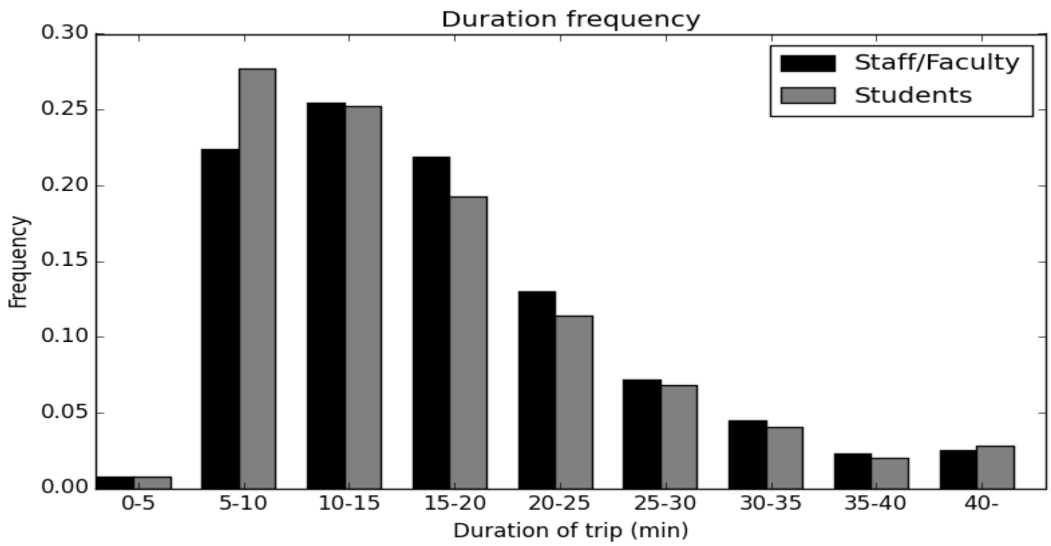


Figure 4.6: Distribution of trip duration divided by occupation

The distribution of duration of trips obtained in this section is similar to the results detailed in [11] where they mention that the average duration detected is 14 minutes. However, our average is lower than the results obtained by [8] which estimated an average of 20-25 minutes. These differences may be due to lower distances in commuting trips, and also to different characteristics of the populations who have electric bikes in China and the group of participants considered for our project. Additionally, the location might influence the results since Shanghai and Kumming, cities where the authors collected data from, are considerably larger than Waterloo-Kitchener, where University of Waterloo is located.

The t-test in this case confirmed that the mean of both populations is statistically different with a p-value of 0.038.

4.2.4 Distribution of Trips per Month

This section analyzes the distribution of trips per month. In this case, it is important to consider that participants were encouraged to keep using their bikes during the winter months. The conclusions obtained from this analysis can be used by manufactures to estimate their sales of bikes throughout the year. Additionally, manufactures can also consider selling seasonal accessories for bikes depending on the pattern of usage identified during each month. For policy makers, this information can be used to design different mobility plans for specific times of the year along with the implementation of safety measures to decrease accidents during times when the weather makes it riskier to use e-bikes.

Figure 4.7 shows the results obtained from the aggregated data when plotting the frequency of trips per month. This graph shows that most of the trips were taken during the months of May and October. This is an expected result since the weather is generally known as one of the main factors that influence usage of bikes. However, it is important to notice that approximately 15% of trips happened during the winter months.

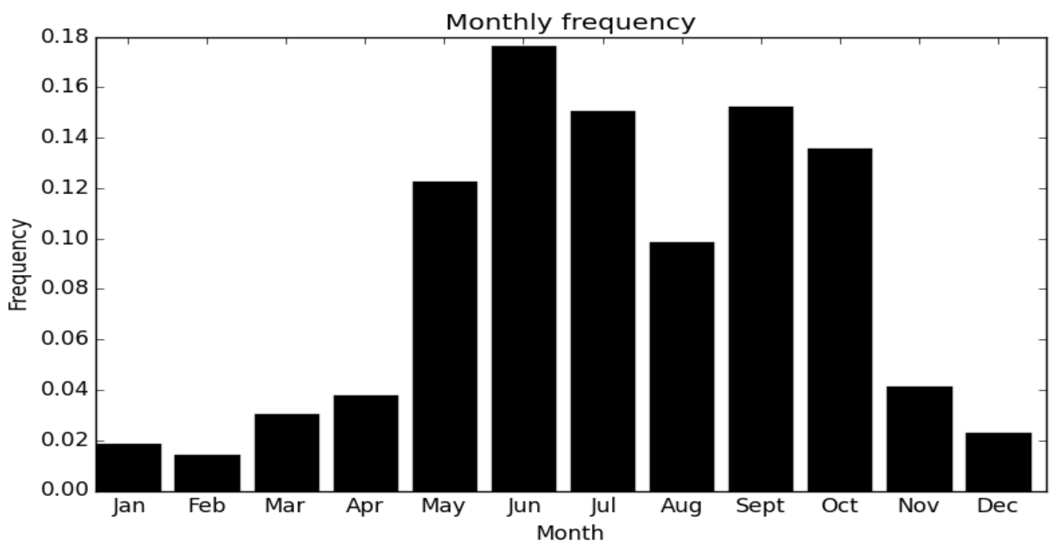


Figure 4.7: Distribution of trips per month

Figure 4.8 shows that female participants mostly bike during the months of June and July, while male users present peaks during the months of September and October. Additionally, we can see that the number of trips by women reaches a peak in June, and then it consistently decreases until January. On the other hand, trips taken by male participants have similar values during the high season of the year.

Finally, Figure 4.9 presents the frequency of trips per month divided between staff/faculty members and students. In this case, no major differences are noticeable but the fact

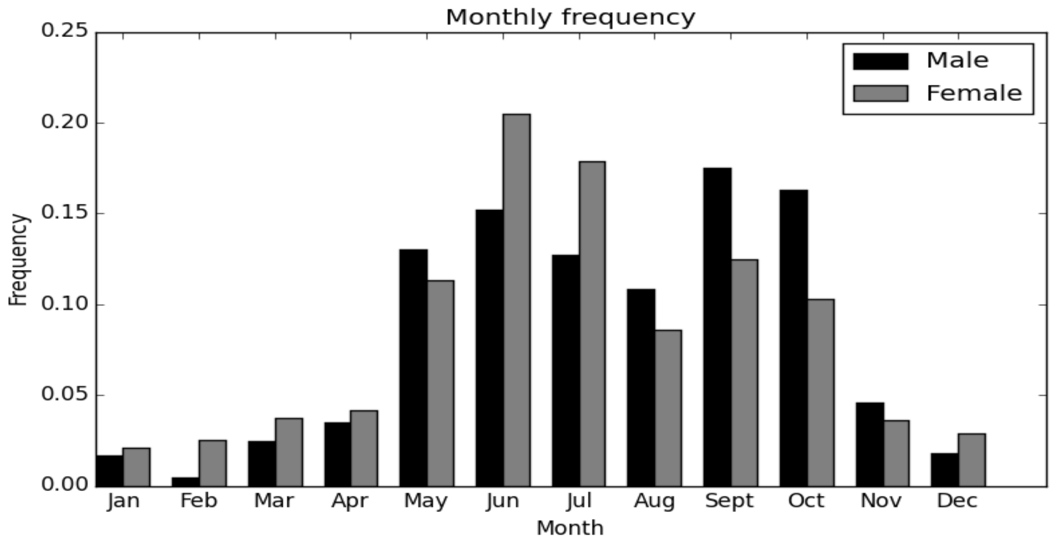


Figure 4.8: Distribution of trips per month divided by gender

that staff and faculty show a higher number of trips during two of the coldest months of the year: December and January.

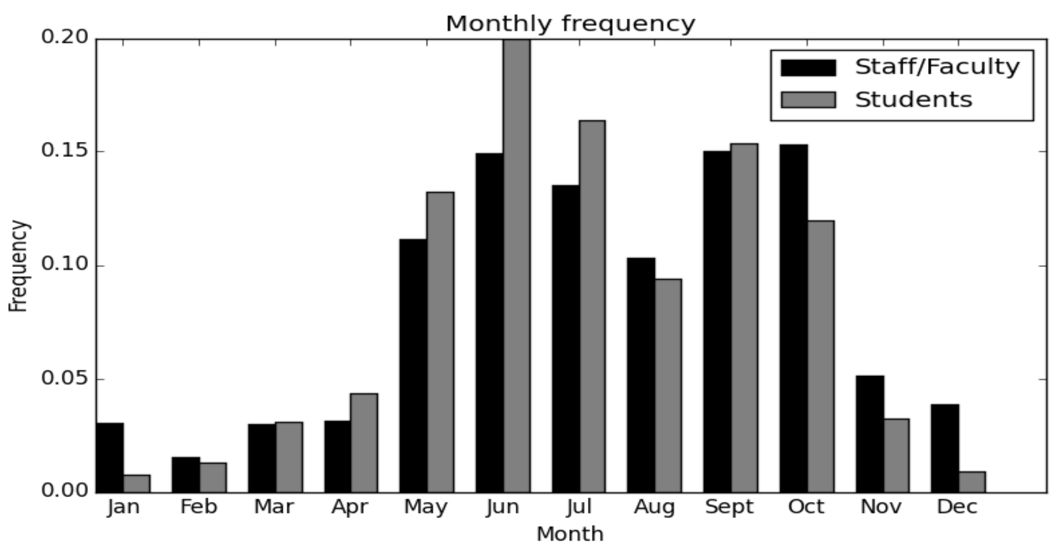


Figure 4.9: Distribution of trips per month divided by occupation

4.2.5 Distribution of Average Speed

This section analyzes the average speed calculated per trip. This data can be used by manufactures to adapt the components of electric bikes based on the common speeds reached by users. Additionally, policy makers can use this information to redefine speed limits, or consider adding laws specific for electric bikes in order to guarantee the safety of all types of vehicles that share the roads.

This analysis could not be performed on the whole data set since it requires GPS data to be available during trips to properly calculate the distance of each trip. In our data set, 60% of the trips contain enough geolocation data to calculate the trajectory followed by the bikes. The results presented in Figure 4.10 show the frequency of average speeds of trips for each speed defined in the horizontal axis. This graph shows that most of the trips have an average speed under 13 km/h. After this mark, the number of trips reduces linearly as the speed increases.

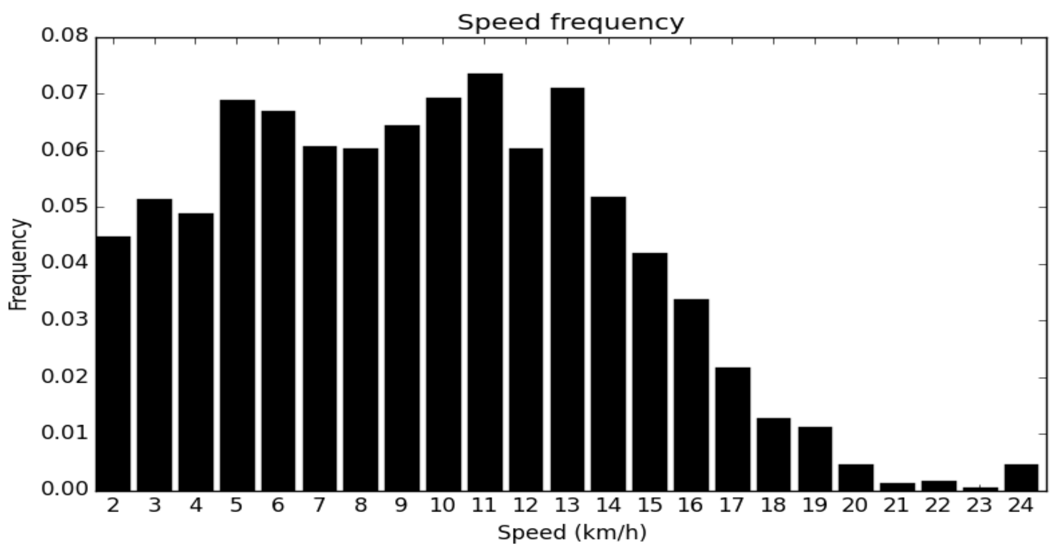


Figure 4.10: Distribution of average trip speed

Figure 4.11 presents the results divided by gender. This graph shows that on average, men tend to reach higher speeds since they have most peaks after the 10 km/h point. On the other hand, women have their two highest peaks at 5 and 6 km/h. Besides one exception, men have a greater number of trips with an average speed of at least 10 km/h than women do. These results support the idea that men tend to ride faster than women and are confirmed by the test which rejected the null hypothesis with a p-value of 0.01.

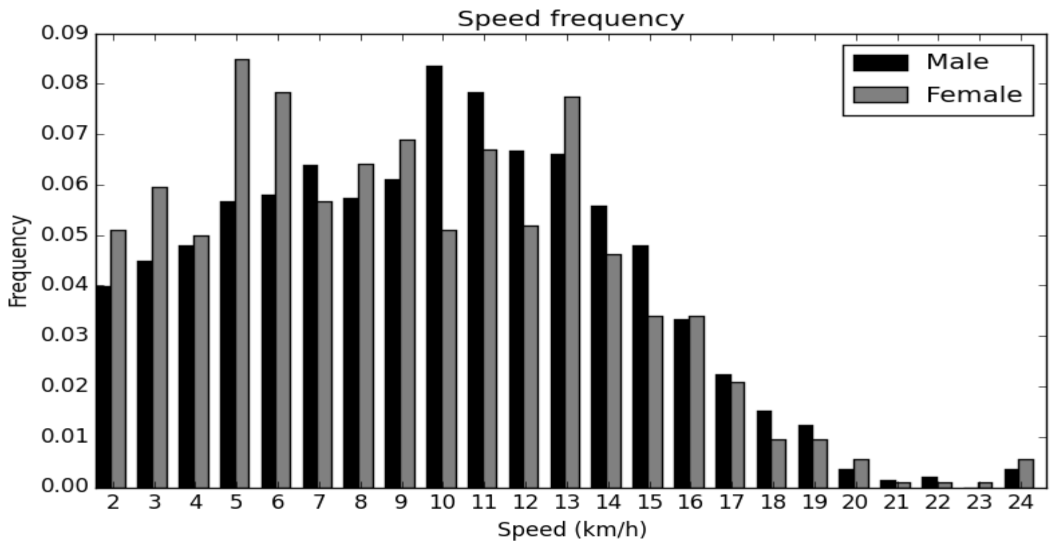


Figure 4.11: Distribution of average trip speed divided by gender

The last plot of this section analyzes the speeds divided by occupation. In this case, Figure 4.12 shows that the highest peaks for students happen at the speeds of 5 and 13 km/h. Similarly, staff and faculty have peaks in the same range, however, they have a slightly higher number of trips with an average speed above 13 km/h. The t-test for this data showed that the means of these distributions are statistically different with a p-value of 0.018.

The results obtained in this section have a similar a distribution as the ones described by Dozza and Fernandez [10]. However, our results have a lower average speed.

4.2.6 Distribution of Average Temperature of Trips

This section presents the frequency of trips with the average temperature defined on the horizontal axis. The range of temperatures goes from -10 to 35 degrees Celsius, however, most of the trips were registered within a shorter range. The plot shows that most of the trips happened with a temperature of 23 °C. These results reaffirm the idea that weather, and temperature in particular, play an important role for people when they decide whether or not to use their electric bikes. Finally, the plot shows that the range of temperatures between 20 to 30 degrees Celsius contains most of the trips detected.

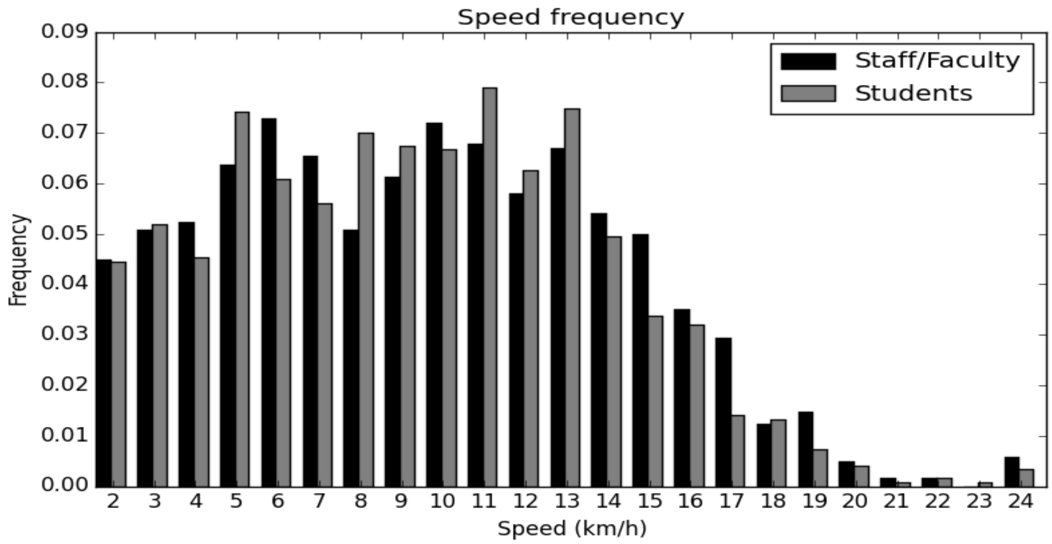


Figure 4.12: Distribution of average trip speed divided by occupation

The conclusions obtained in this section will help manufactures to determine the typical temperatures when their users prefer to ride their bikes. Additionally, this information can be used by policy makers to estimate the amount of electric bikes on the roads based on weather conditions. This could also help them to implement safety measures during different seasons of the year based on the expected temperatures.

Our conclusions from this section are aligned with the results presented in [21] since most of participants in that survey mentioned that weather is one of the main factors that influences their usage patterns of e-bikes.

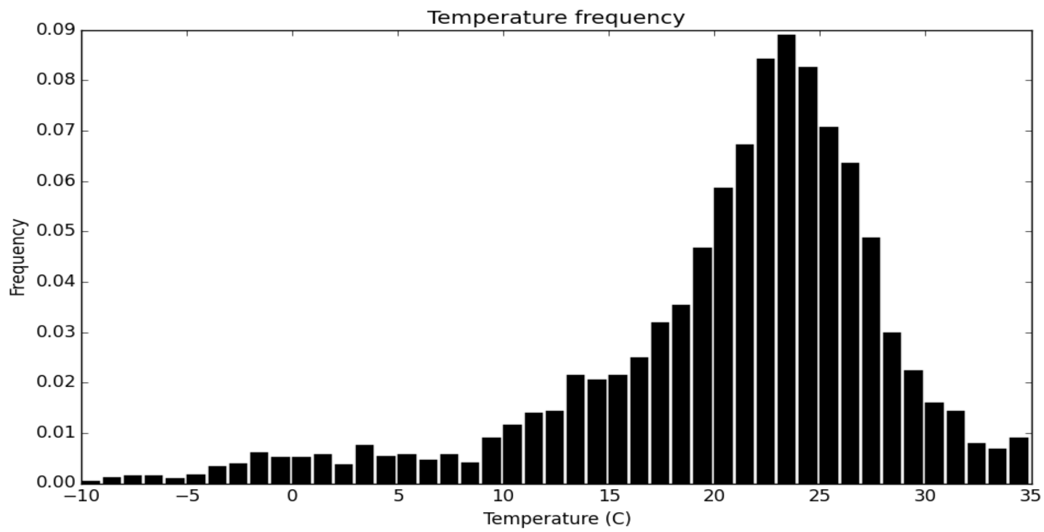


Figure 4.13: Distribution average trip temperature

4.3 Charging Events

4.3.1 General Statistics

This section focuses on the analysis of charging events which are essential to identify the potential effects of electric bicycles on the grid. These events were identified by detecting spikes in the charging current; the starting time of a charging event was obtained by identifying the first spike in the charging current. Similarly, the end time of a charging event was obtained by identifying the moment when the charging current decreased to zero. More details about identifying charging events can be found in Carpenter's work [6].

Table 4.2 presents an aggregated summary of the identified charging events. We were able to detect 2007 charging events which results in an average of 2.3 trips per charging event. Additionally, we can see that, on average, female participants take more trips between charging events, which is consistent with the fact that they generally take shorter trips, allowing them to not deplete their batteries that fast. Similarly, students have a greater average number of trips between charging events, which suggest that they generally take shorter trips than staff and faculty members.

The following sections analyze different aspects of charging events which can be used by stakeholders of this study.

Description	Total	Female	Male	Students	Staff/Fac.
Charging events	2007	791	1216	927	1080
Average charging events per participant	67	61	68	62	68
Average trips per charge	2.3	2.7	2.1	2.7	2.0

Table 4.2: General statistics of Charging events

4.3.2 Distribution of Charging Start Time

This section analyzes the time of the day at which charging events started. This can be used by utility companies in order to estimate the impact of this extra load on the grid. Additionally, new policies can be created by policy makers in order to reduce the impact of battery charging during the peaks of energy consumption throughout the day.

Figure 4.14 presents the aggregated results. This plot shows that most of the charging events start between 8 am and 6 pm. Also, we can see that there is a similar number of charging events within this range which suggest that the price variations of energy during the day do not influence the charging habits of participants. Additionally, when comparing these results with Figure 4.14, we can see that charging events times do not show peaks as high as the ones in trip times, however, there is an increase in the number of charging events at similar times as the peaks in trips (8 am and 4 pm). This suggest that participants prefer to charge their batteries after riding their bikes.

Figure 4.15 shows that female participants have 3 peaks of times when they start charging their batteries, while male participants have similar number of events during the day, and night. Additionally, the graph shows that male participants have more charging events starting at unusual times of the day such as late in the night and after midnight.

Finally, Figure 4.16 shows that staff and faculty have a more defined peak around noon, while students have several peaks around 8 am, noon and 5 pm. Again, this confirms that students have changing schedules which is consistent with the analysis of their trip's start times in the Section 4.2.2.

4.3.3 Distribution of Charging Duration

This section studies the duration of each charging event measured in hours. For this analysis it is important to consider that the battery takes up to 5 hours to charge. This data should also be considered by utility companies in order to estimate the impact of

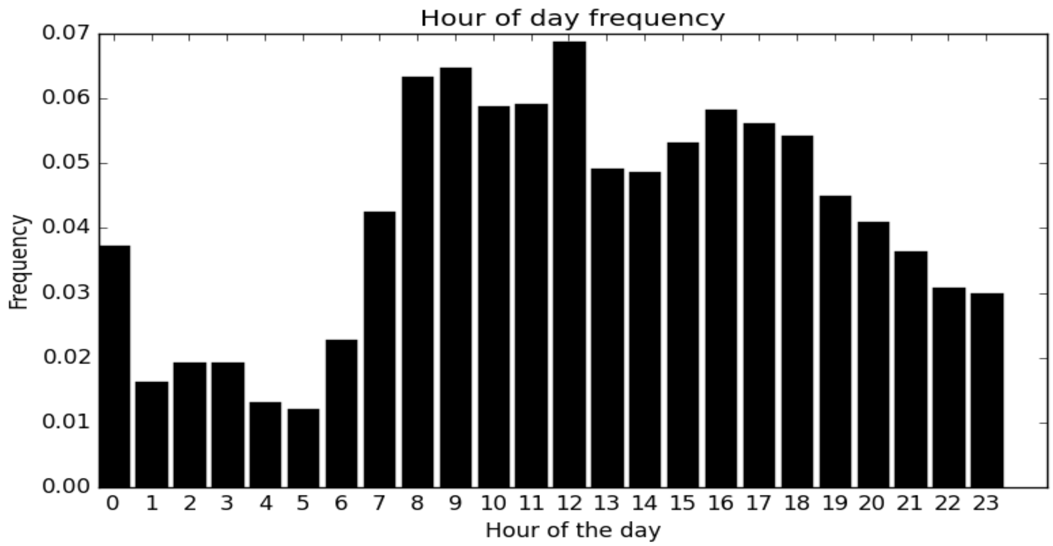


Figure 4.14: Distribution of hour of the day when charging events started

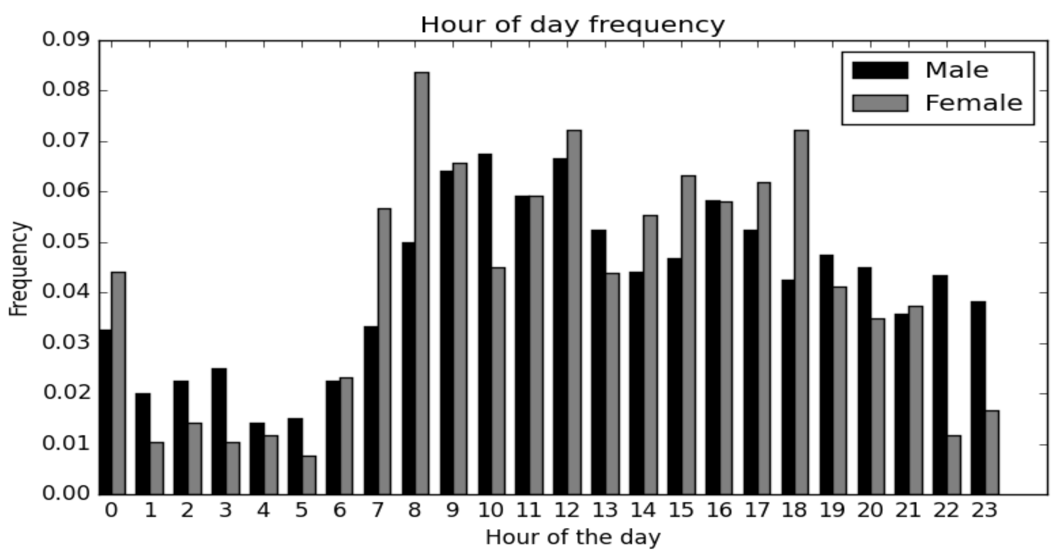


Figure 4.15: Distribution of hour of the day when charging events started divided by gender

battery charging events on the grid, since the duration of these events is needed to estimate the total energy consumption.

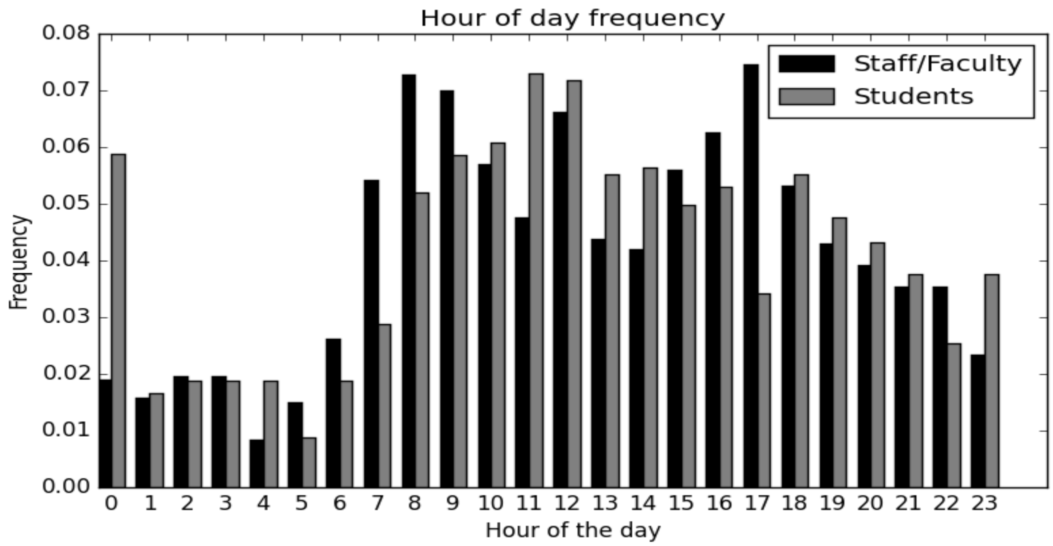


Figure 4.16: Distribution of hour of the day when charging events started divided by occupation

Initially, we plotted the aggregated data which is shown in Figure 4.17; the plot shows the frequency distribution in black, and the cumulative distribution in gray. In this graph, we can see that approximately 35% of the charging events takes 5 hours or more. Additionally, 27% of the total charging events lasted less than an hour.

Figure 4.18 presents the results divided by the gender of participants. This plot shows that it is more common for male participants to have charging events with shorter durations; 31% of their total charging events lasted less than an hour. In contrast, female participants generally tend to have longer charging events, specifically, they have greater percentages of charging events with durations of over three hours. However, with the t-test performed comparing these populations, we obtained a p-value of 0.1 which means that the null hypothesis was not rejected and hence both means do not present an statistical difference.

Lastly, Figure 4.19 shows the results divided by occupation. In this case, we can see that 43% of the charging events from students lasted over 5 hours. Additionally, the plot shows that it is more common for faculty and staff members to have charging events with a duration of under four hours. The t-test performed with these sets of data determined that their means are statistically different; the p-value obtained in this case is 0,0001.

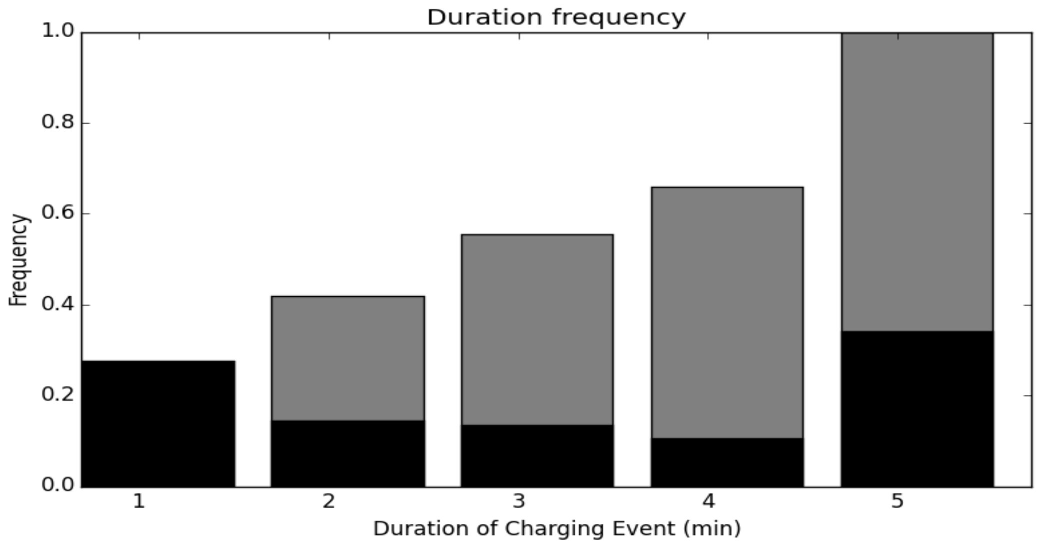


Figure 4.17: Distribution of charging duration

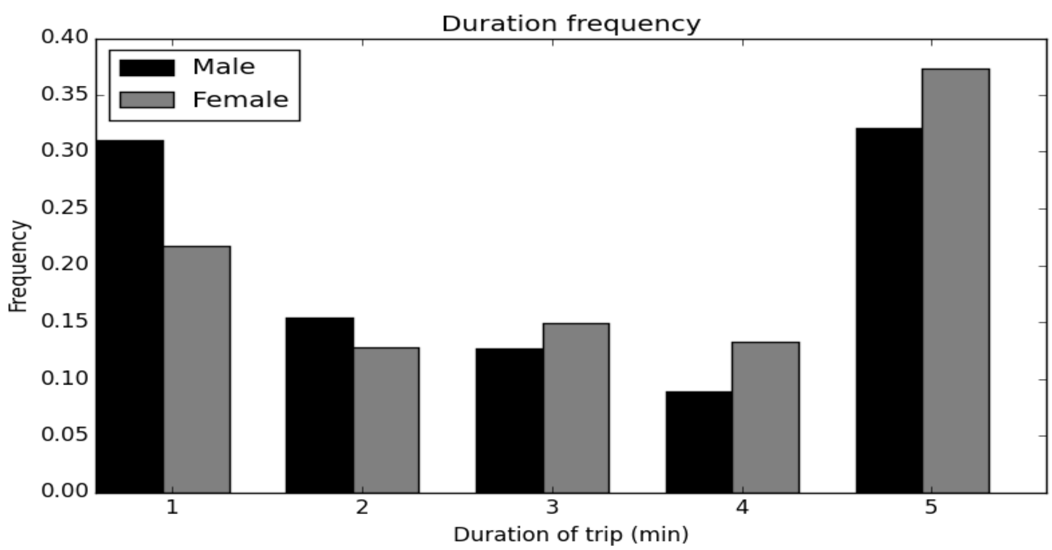


Figure 4.18: Distribution of charging duration divided by gender

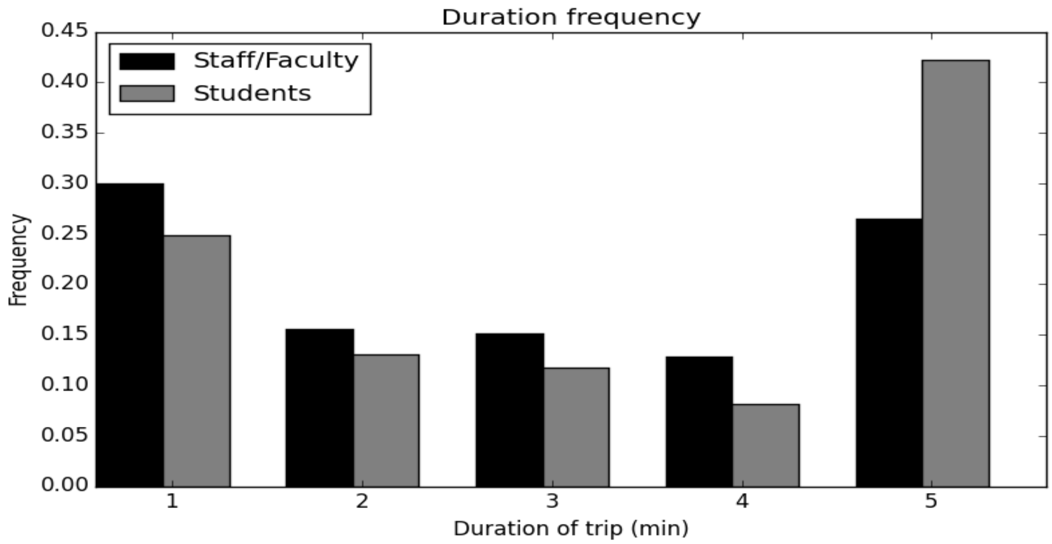


Figure 4.19: Distribution of charging duration divided by occupation

4.3.4 Charging Event Start Times vs State of Charge

The last part of this section is dedicated to analyzing the number of charging events in relation to the state of charge (SOC) of the battery. SOC is the percentage of remaining energy left in the battery. The conclusions obtained in this section can help understand how comfortable people feel with different levels SOC, which is translated into range anxiety; the interpretation is that participants willing to deplete their batteries to lower percentages generally have less range anxiety.

Electric bikes manufactures can use this information to provide tools to reduce range anxiety. A common example is presenting range prediction to the user, which is an estimation of the total distance remaining with battery assistance at different SOC levels. Additionally, this data can be used to estimate the battery capacity for different segments of the population. Similarly, policy makers can use this information to offer options to reduce range anxiety, for instance, by building charging stations in popular destinations for e-bike riders.

Figure 4.20 shows the frequency of events that started within the ranges of SOC defined in the x-axis in black, with the cumulative distribution in color gray. We can see that 10% of the charging events started when the battery had a SOC of less than 10%. Additionally, note that most of the events start when the battery has over 30% SOC. Finally, we can see that 25% of the charging events were started when the SOC was above 90% which still suggests that range anxiety is common among participants.

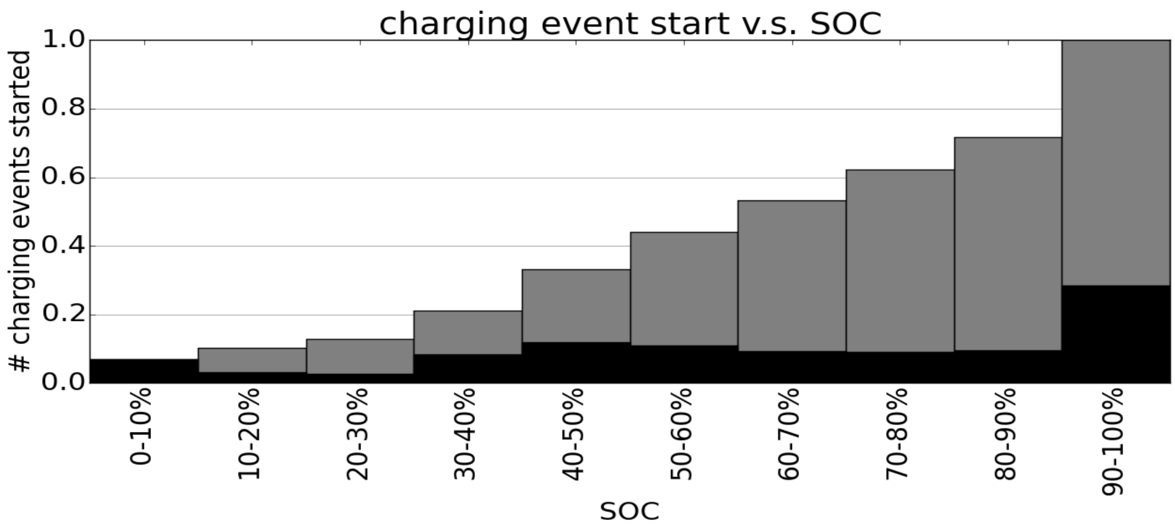


Figure 4.20: Charging events start vs SOC level

Figure 4.21 presents the results obtained after dividing the data by the gender of participants. In this plot, we can see that most of the charging events start with SOC levels greater than 40% and there is a peak after 90%-100%. This peak confirms that participants prefer to keep their batteries fully charged. Additionally, male participants show a higher frequency of events (35%) in the range of the aforementioned peak; this suggests a more preventive attitude since they start charging their batteries even when the SOC is greater than 90%.

Finally, when comparing the populations of students with staff/faculty members, Figure 4.22 presents similar results between both groups. However, students have a higher peak in the range of 40-50% which suggest that they feel more confident to deplete their batteries to lower SOC levels. This conclusion is also confirmed by the fact that faculty and staff have a higher percentage of charging events starting with an SOC in the range of 90-100% which shows that they are more conservative and prefer to keep their batteries fully charged.

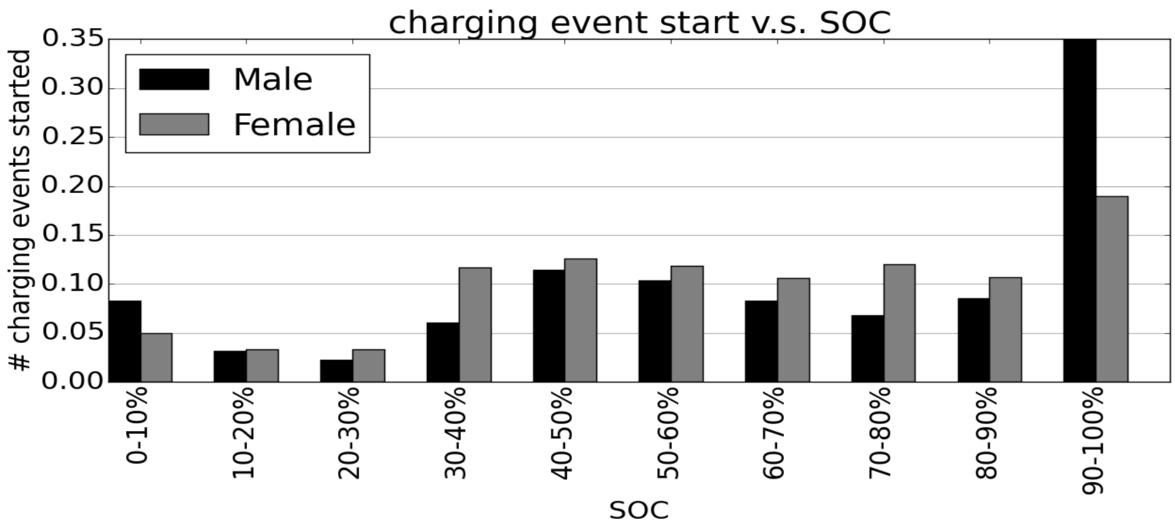


Figure 4.21: Charging events start vs SOC level divided by gender

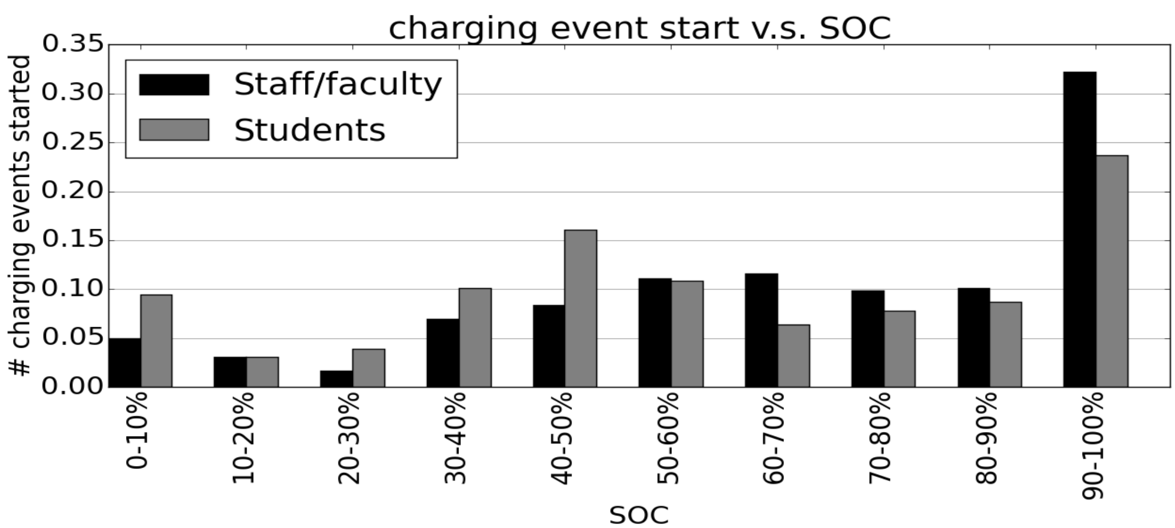


Figure 4.22: Charging events start vs SOC level divided by occupation

4.3.5 Charging Event End Times vs State of Charge

In this section, we analyze the SOC levels measured at the end of charging events. Figure 4.23 presents the distribution of frequency of events in black, and the cumulative distribution of events in gray. This plot shows that almost 80% of all the charging events end with a battery fully charged (90% or more). This means that even in cases when the charging time was less than an hour, most of this events ended up with a fully charged battery.

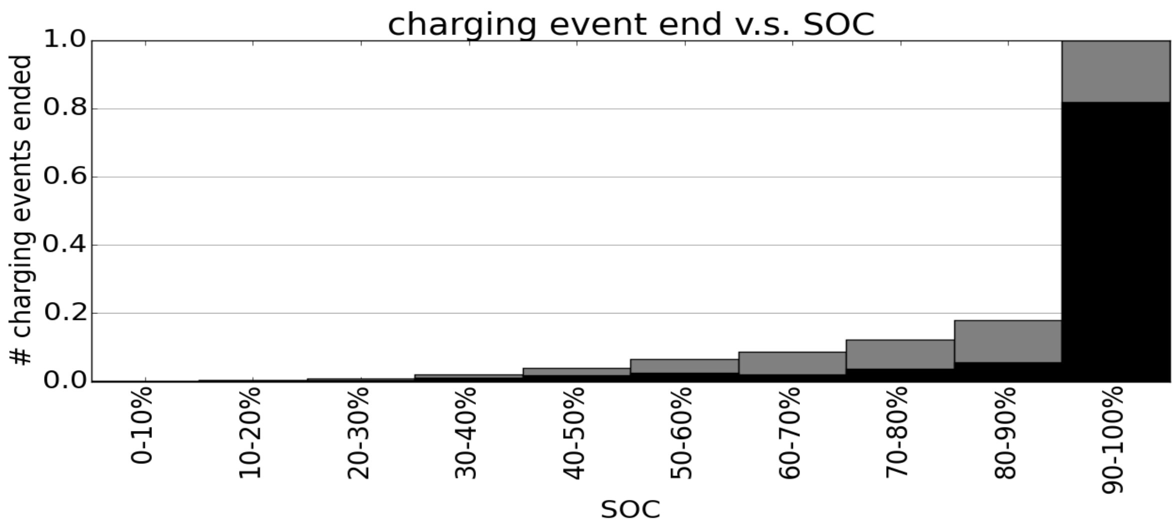


Figure 4.23: Charging events end vs SOC level

Figure 4.24 and Figure 4.25 present the plots obtained when dividing the data by gender and occupation, respectively. In this case we can see that there is no difference between these segments of the population since there is a common pattern to charge their batteries until the SOC reaches ranges between 90-100%.

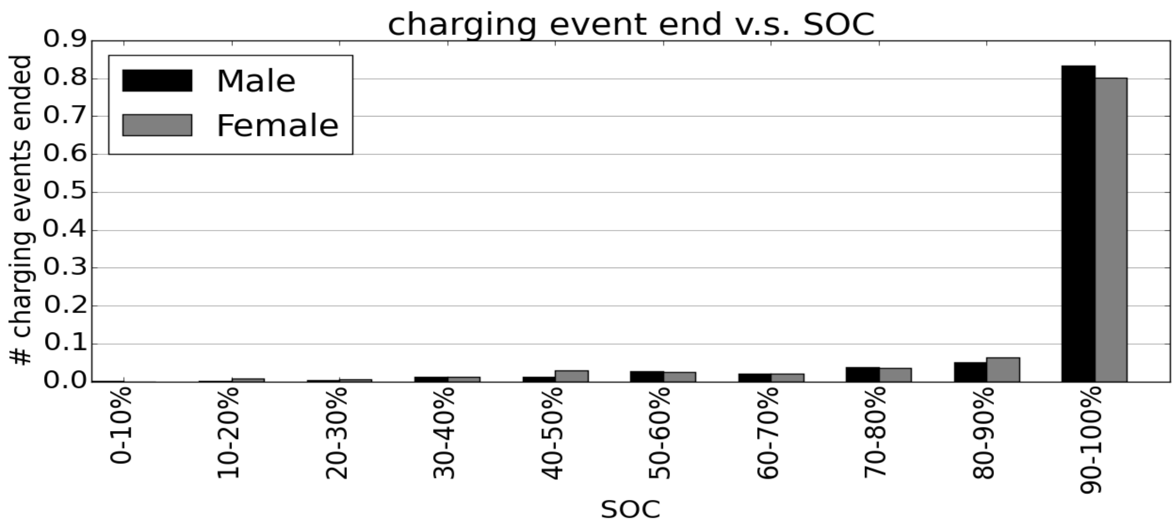


Figure 4.24: Charging events end vs SOC level divided by gender

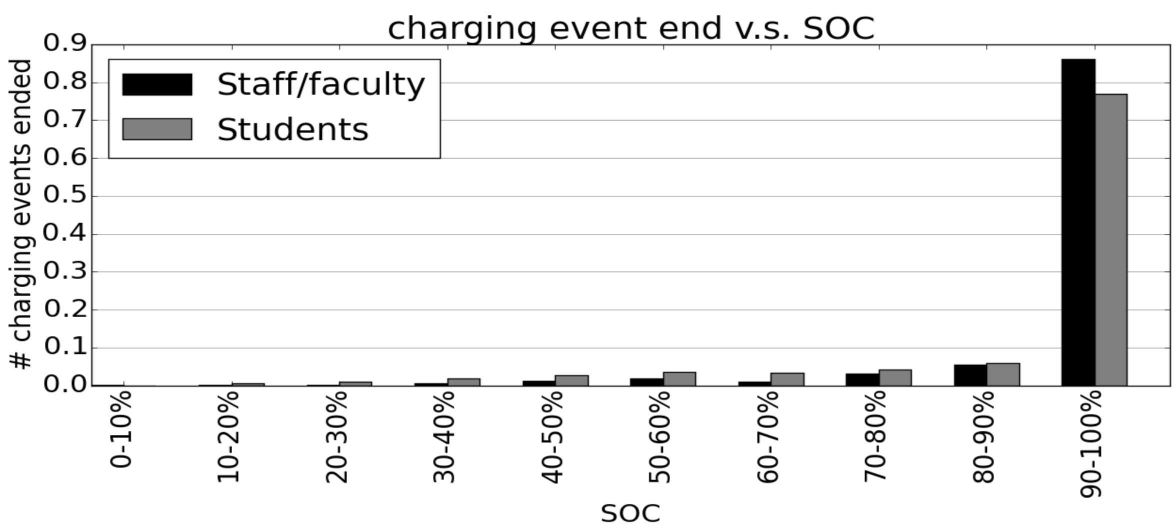


Figure 4.25: Charging events end vs SOC level divided by occupation

4.4 Survey Analysis

4.4.1 Correlation of Answers vs Riding Behavior

In this section, we present the comparison performed between the responses given to the survey and the actual riding habits of each participant. As mentioned before, we consider important to determine if it is possible to predict the behavior of participants based on their answers to the survey. This questionnaire had two sections where participants were asked to estimate the usage of their e-bikes during the summer and winter. The following sections describe the correlation analysis of these answers with the riding habits of participants:

Estimated distance per week during the summer

The first question asked participants to estimate the number of km's they will ride every week during the summer. Using their responses, we did a correlation analysis between this information and the average minutes ridden weekly during the two months with more trips from each participant. We decided to use the two months with more trips in order to reduce the effects of weather, vacation times, and other factors that modify the riding habits of participants. Additionally, we decided to compare the duration of trips since this data is highly correlated with the distance ridden and, in our case, these measurements were more accurate than the calculated distances.

Figure 4.26 shows 25 points, one for each participant, where the x-axis represents the expected average number of kilometers per week, and the y-axis describes the average number of minutes ridden weekly during the two months with more trips. As shown in the plot, there is no clear relationship between these two variables. In order to confirm this result, we calculated the Pearson coefficient of correlation, and the value obtained was 0.18 which can be interpreted as no correlation between the variables. This approach was also modified by using different number of months worth of data but all the results confirmed that there is no clear correlation between these variables.

Similarly, we tested for correlation between the expected number of km's responded in the survey, and the average number of trips per week during the two months with more trips for each participant. The results are presented in Figure 4.27 where the y-axis represents the average number of trips calculated for each participant. This plot presented no correlation between the variables again. As shown in the plot, there is no clear relationship between the variables, and the Pearson coefficient confirms this fact since the calculated value was -0.03. With this analysis, we confirmed that the answers to the

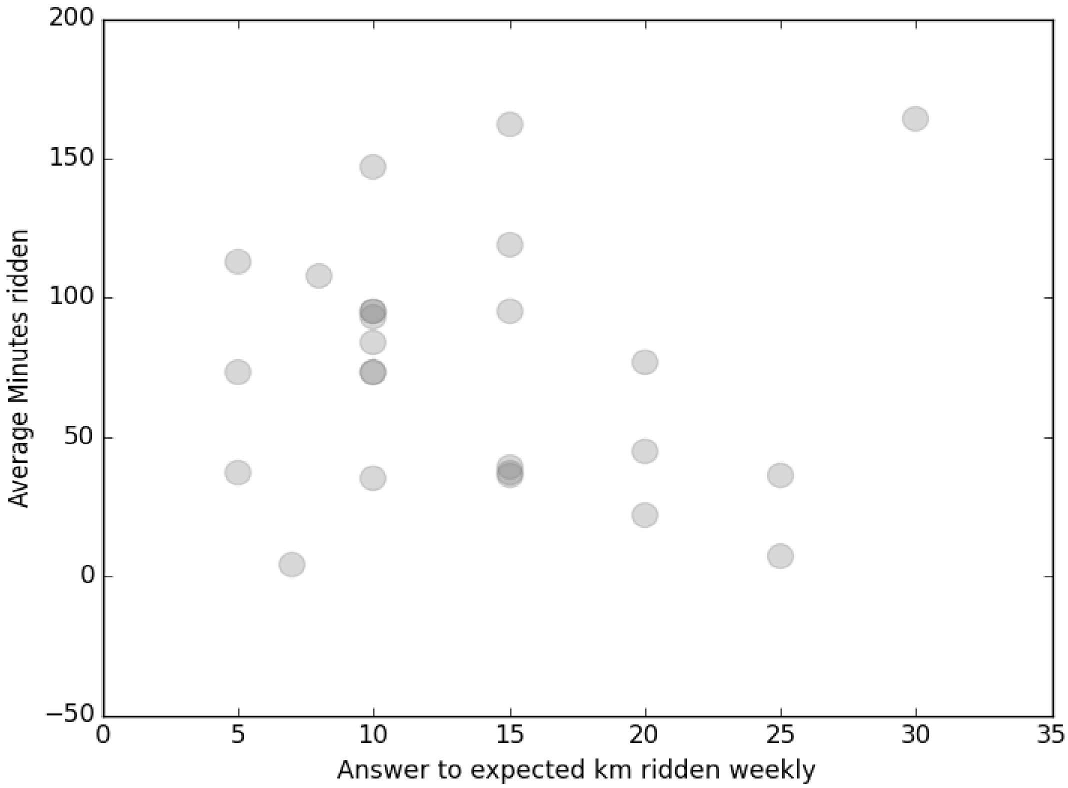


Figure 4.26: Correlation between expected distance per week (survey) and average minutes ridden per week of each participant

survey did not reflect the actual riding behavior of each participant during the summer months.

Estimated riding during winter months

The same analysis from the previous sections was performed using the question where participants were asked to estimate the usage of their electric bikes during the winter. Figure 4.28 shows the correlation between the estimated usage of the electric bikes and the average number of minutes ridden during the winter by each participant. Similarly as before, no clear correlation is detected between the variables. As we can see on the plot, there are several participants who said they will use their bikes on a weekly basis, however,

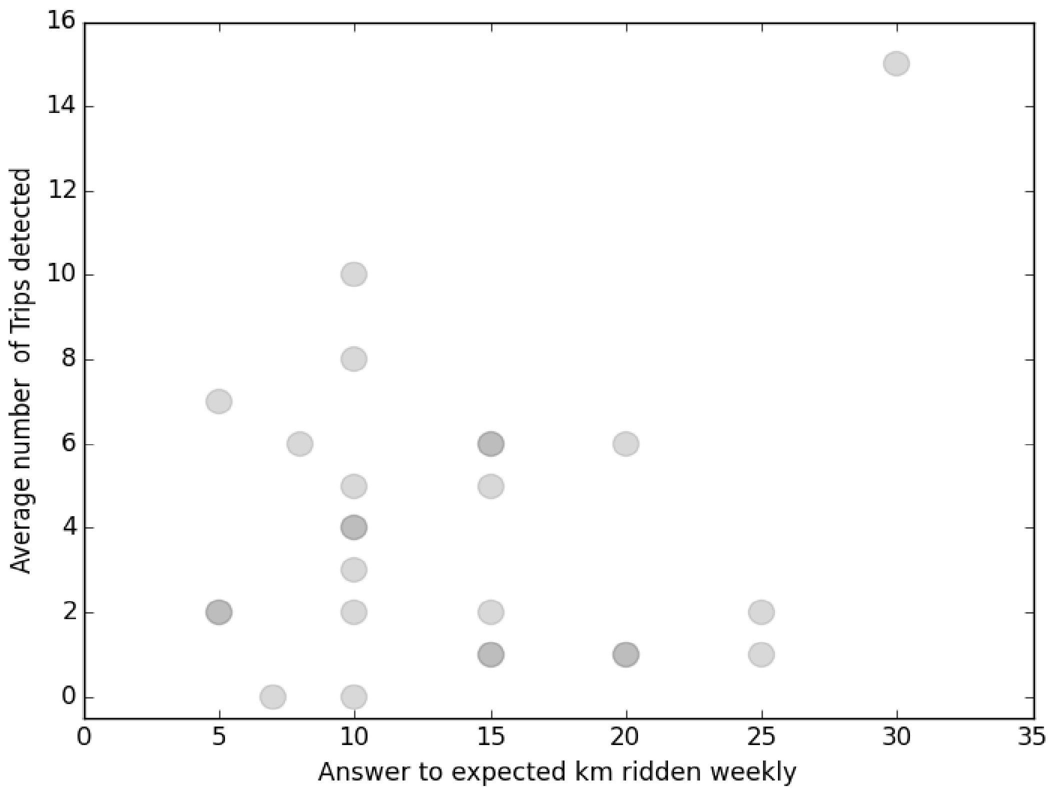


Figure 4.27: Correlation between expected distance per week (survey) and average trips per week of each participant

the total number of minutes accumulated per month is not greater than 60, which is also similar to values obtained from participants who gave a different response.

Finally, comparing the answers given with the average number of trips detected per month over the winter we obtained the results presented in Figure 4.29. The graph shows that, again, there is no correlation between the variables. With these two correlation analysis we can see that the answers given in the surveys do not seem to provide information about the real usage of the bikes during the winter.

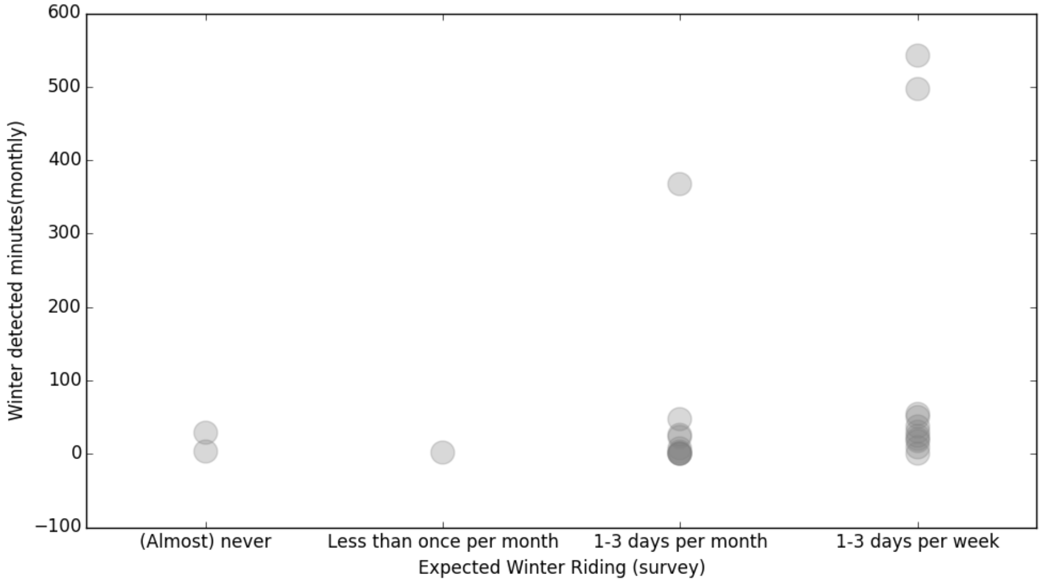


Figure 4.28: Correlation between usage (survey) and average minutes ridden per month of each participant

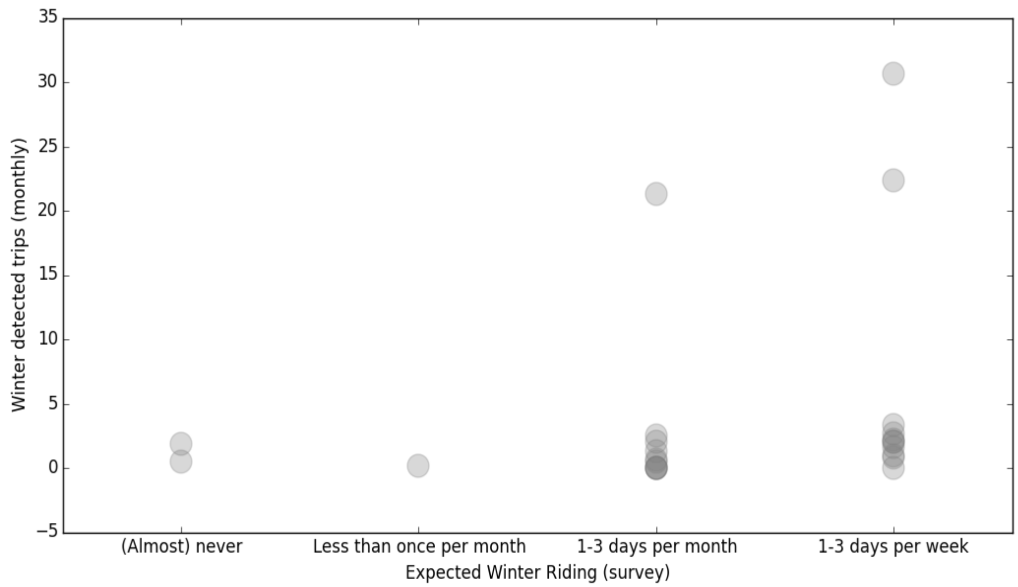


Figure 4.29: Correlation between usage (survey) and average trips ridden per month of each participant

Chapter 5

Conclusions and Future Work

The main conclusions obtained from this thesis are described in the following points:

- As expected, we confirmed that weather is one of the main factors of influence on the usage patterns of electric bikes. Winter months presented a drastic decrease of usage while the months with more trips are June, July and September (summer months). Similarly, the temperature range in which most of the trips happened was between 20 and 27 degrees Celsius.
- Electric bikes are most commonly used for commuting purposes, and their usage has noticeable peaks at the beginning and end of regular business days.
- Participants did not commonly use their electric bikes after sunset probably because of higher risk of accidents due to lower visibility.
- Regarding riding behavior, female participants presented lower average speed, and shorter trips compared to male participants.
- Concerning charging events, male participants presented more scattered times when they start charging their batteries. Additionally, men tend to charge their batteries more often which causes shorter charging times.
- When comparing the riding habits of students and faculty/staff members, we found that students start trips at more scattered times of the day, and it is more common for them to use their bikes during evening hours. In contrast, faculty and staff members showed a higher number of trips taken during winter months of the year, and they consistently reached higher average speeds compared to students.

- Regarding charging events, students showed scattered times of the day when they charge their batteries. Additionally, staff and faculty members presented shorter but more frequent charging events.

- When we analyzed the SOC levels at which charging events start and end, we were able to see that participants still prefer to not deplete their batteries under 30%. Also, participants generally choose to charge their batteries until they are full. These results suggest that range anxiety is common among participants, and they generally prefer to keep their batteries with high SOC levels.

- We determined that the means of trip duration, and average speed of men and women, students and staff/faculty are statistically different.

- By analyzing the initial surveys taken by all the participants, we were able to see that there seems to be no correlation between the expected usage, initially defined by the users, and the data collected from each one of them. This may be caused due to a change of behavior as participants became more familiar with electric bikes. This result may discard surveys as valid information to estimate or model the usage of electric bikes, however, we expect to confirm this results in the future with a bigger sample of the population.

As mentioned before, the work presented in this thesis will be used in future research projects. In the long term, there are several other studies to be developed, the following list presents the main ones:

- Comparison with other modes of transportation: we will compare our results to similar studies that have focused on usage, and behavior analysis of other modes of transportation. With this, it will be possible to understand the particular characteristics of how people use electric bikes in order to be able to extrapolate our results to other applications, in particular, electric vehicles usage.

- Carbon footprint of electric bicycles: using the trip detection algorithm, one further analysis is to develop criteria to identify leisure, and health related trips from commuting trips. With this information, we will be able to identify the carbon footprint generated by the bikes when they are used as a replacement to other modes of transportation for commuting.

- Range prediction: using the trips detected and combining it with information from additional sensors, we will develop models to accurately estimate the range of distance that can be covered using the remaining energy left in the battery. These results can be used as feedback for e-bike users in order to help them reduce the energy consumption of their batteries during trips.

Finally, similar analysis will be performed once more data is collected in order to confirm the results presented in this thesis. Additionally, we are expecting to increase the size of our fleet of electric bikes in order to have a sample that represents the population in a better way. Hence the present work will be used as a point of reference for comparison, and it will help to study the evolution of the behavior of participants over the life of this project.

References

[1] Kang An, Xiaohong Chen, Feifei Xin, Bin Lin, and Longyu Wei. Travel characteristics of e- bike users: Survey and analysis in shanghai. *Procedia - Social and Behavioral Sciences*, 96:1828–1838, 2013.

[2] Anonymous. The new wave of electric bicycles. *Institute of Transportation Engineers.ITE Journal*, 83(11):27, 2013.

[3] Brandon Baker. Electric bike sales soar worldwide, 2013. Accessed November 1st 2015. http://ecowatch.com/2013/11/06/electric-bike-sales-soar-worldwide.

[4] Hui Zhi Zak Beh, Grant A. Covic, and John T. Boys. Wireless fleet charging system for electric bicycles. *Emerging and Selected Topics in Power Electronics, IEEE Journal of*, 3(1):75–86, 2015.

[5] Kevin Bullis. Lithium-ion battery, 2012. Technology Review, Inc. Jul/Aug 2012; Last updated - 2012-10-17. Accessed November 1st 2015. http://search.proquest.com.proxy.lib.uwaterloo.ca/docview/1112262444?accountid=14906.

[6] Tommy Carpenter. Measuring and mitigating electric vehicle adoption barriers, 2015. Thesis (Ph.D)–University of Waterloo, 2015.

[7] C. Chen, Dq Zhang, P. S. Castro, N. Li, L. Sun, S. J. Li, and Zh Wang. iboat: Isolation- based online anomalous trajectory detection. *Ieee Transactions On Intelligent Transportation Systems*, 14(2):806–818, 2013.

[8] Christopher Cherry and Robert Cervero. Use characteristics and mode choice behavior of electric bike users in china. *Transport Policy*, 14(3):247–257, 2007.

[9] Eh Chung and A. Shalaby. A trip reconstruction tool for gps- based personal travel surveys. *Transportation Planning and Technology*, 28(5):381–401, 2005.

[10] Marco Dozza and Andre Fernandez. Understanding bicycle dynamics and cyclist behavior from naturalistic field data (november 2012). *Intelligent Transportation Systems, IEEE Transactions on*, 15(1):376–384, 2014.

[11] Marco Dozza, Giulio Francesco Bianchi Piccinini, and Julia Werneke. Using naturalistic data to assess e- cyclist behavior. *Transportation Research Part F: Psychology and Behaviour*, pages 199–212, 2015.

[12] eProdigy Bikes. eprodigy whistler, 2013. Accessed November 1st 2015. http://www.eprodigybikes.com/whistler.aspx.

[13] Boris Gojanovic, Joris Welker, Katia Iglesias, Chantal Daucourt, and Grald Gremion. Electric bicycles as a new active transportation modality to promote health. *Medicine and Science in Sports and Exercise*, 43(11):2204–2210, 2011.

[14] H. M. Gong, C. Chen, E. Bialostozky, and C. T. Lawson. A gps/ gis method for travel mode detection in new york city. *Computers Environment And Urban Systems*, 36(2):131–139, 2012.

[15] Statista Inc. Estimated sales of electric bicycles in the united states, 2014. Accessed November 1st 2015. http://www.statista.com/statistics/328871/sales-of-electric-bicycles-in-the-united-states/.

[16] P. Kasemsuppakorn and Ha Karimi. A pedestrian network construction algorithm based on multiple gps traces. *Transportation Research Part C-Emerging Technologies; Transp.Res.Pt.C-Emerg.Technol.*, 26:285–300, 2013.

[17] Panagiota Katsikouli, Rik Sarkar, and Jie Gao. Persistence based online signal and trajectory simplification for mobile devices. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 371–380. ACM, 2014.

[18] J. Kim. Node based map matching algorithm for car navigation system. *International Symposium on Automotive Technology and Automation*, pages 121–126, 1996.

[19] Le Yi Wang Kwo Young, Caisheng Wang and Kai Strunz. *Electric Vehicle Integration into Modern Power Networks*. Springer, New York, USA, 2013.

[20] Sergio Manzetti and Florin Mariasiu. Electric vehicle battery technologies: From present state to future systems. *Renewable and Sustainable Energy Reviews*, 51(Complete):1004–1012, 2015.

[21] Natalie Popovich, Elizabeth Gordon, Zhenying Shao, Yan Xing, Yunshi Wang, and Susan Handy. Experiences of electric bicycle users in the sacramento, california area. *Travel Behaviour and Society*, 1(2):37–44, 5 2014.

[22] K. Schleinitz and P. The german naturalistic cycling study - comparing cycling speed of riders of different e-bikes and conventional bicycles. *Safety Science*, 2015.

[23] Nadine Schuessler and Kay Axhausen. Processing raw data from global positioning systems without additional information. *Transportation Research Record*, (2105):28–36, 2009.

[24] Vladimir Usyukov. Development of a cyclists' route-choice model : an ontario case study, 2013. Thesis (Ph.D)–University of Waterloo, 2012.

[25] E. Williams. Aviation formulary, 2010. Accessed November 25th 2015. http://williams.best.vwh.net/avform.html.

[26] J. Wolf, S. Schonfelder, U. Samaga, M. Oliveira, and Kw Axhausen. Eighty weeks of global positioning system traces - approaches to enriching trip information. *Data And Information Technology*, (1870):46–54, 2004.