Weekly report: August 5th, 2015

Ivan Rios S.

August 6, 2015

1 Goals for the week

- Analyze the challenges faced during the algorithm development.
- Analyze the results from the algorithm.
- Obtain insights from the data analysis

2 Activities

- Automate the process to update the trip tables.
- Document the structure of the algorithm
- Identify the challenges and the intrisically hard problems for developing the algorithm
- Obtain insights from the initial analysis
- Develop further analysis of the trip data

3 Summary

From the initial analysis of the survey, the following insights can be obtained. Participants with less interest and experience with ebikes seem to have their number of trips above the average, and they tend to take longer trips; this could be caused by an increased interest in ebikes, and also due to positive experiences (it would be interesting to analyze the changes in their perception after a couple years of using ebikes). This could be a proof that additional factors, such as price, influence people's interest and willingness to use and learn about bikes.

Also, people with more interest and with more experience with ebikes have a lower than average number of trips, and lower total time spent on their ebikes. Finally, the factors which seems to directly influence in the total time spent riding their ebike is the participant's interest in replacing their bike, or car with an ebike; this seems to be a motivation to go for longer distances and time;

however, these two factors barely influenced in how often people use their bikes since the number of trips was similar for all the groups. As a general conclusion, this initial analysis showed a change in the perception of ebikes (assuming it is correlated to participant's use) in a positive manner after the participants got the chance to own and regularly use them.

4 What I learnt

Algorithm description:

The algorithm has two main phases, the first one is trip identification, which basically identifies all the trips in a specific period of time. These trips are identified using a sliding window approach in which a threshold of values has to be reached in order to consider real movement in the bike. Then, the second phase joins trips that could be potentially be considered one based on some defined rules, and finally it filters out trips that do not meet the minimum distance, and/or duration.

The following snippet presents the pseudo-code used in the algorithm:

```
1  #Definition of variables:
2  CHARGING_CURRENT_LIMIT = 4.8 #Threshold for charging current
3  GYRO_LIMIT = 0.09 #Threshold for resultant of 3 gyro components
4  LIN_ACCEL_LIMIT = 1 #Threshold for resultant of 3 accel components
5  MIN_NUMBER_VALUES_OVER_THE_LIMIT = 4 # Threshold for values over
       the defined limit of each variable
6  WINDOW_WIDTH = 12 # Number of values slidding window can hold
7  MIN_TRIP_LENGTH = 300 # Minimum number of seconds
8  MAX_TIME_WITHIN_TRIP = 280 # Maximum time with no movement within a
       trip
9  MAX_AVG_SPEED = 20 # Maximum average speed in a trip
10
11 trip_started = False # Variable that keeps track if a trip has been
       started
12 charging_currents = Parameter(WINDOW_WIDTH, CHARGING_CURRENT_LIMIT,
       MIN_NUMBER_OF_VALUES_OVER_THE_LIMIT) #Variable holding charging
       current data
13 gyroscopes = Parameter(WINDOW_WIDTH, GYRO_LIMIT,
       MIN_NUMBER_OF_VALUES_OVER_THE_LIMIT) #Variable holding charging
       current data
14 lin_accels = Parameter(WINDOW_WIDTH, LIN_ACCEL_LIMIT,
       MIN_NUMBER_OF_VALUES_OVER_THE_LIMIT) #Variable holding linear
       acceleration data
15 start_times = [] # List of starting times
16 end_times = [] # List of ending times
17 distances = [] # List of distances
18
19 raw_data = obtain_data_from_db() #Calls DB to obtain data
20 ### PHASE 1 ###
21 for record in raw_data do:
22         # Adding data to variables
23         charging_currents.push(record.charging_current)
24         gyroscopes.push(record.gyro)
25         lin_accels.push(record.lin_accel)
26
```

```
27          # Identying trips
28          # Case when trip has not started
29          if not trip_started:
30                  #Check no charging event and movement from gyro/
                        linear acceleration
31                  if not charging_currents.is_charging() and (
                        gyroscopes.is_moving() or lin_accels.is_moving
                        ()):
32                          trip_started = true
33                          # Obtain earliest time when movement was
                                detected
34                          tmp_start_time = get_start_time()
35          # Case when trip has started:
36          else:
37                  #Check for charging event or no more movement
                        detected
38                  if not charging_currents.is_charging() or
                        gyroscopes.is_moving() or lin_accels.is_moving
                        ():
39                          trip_started = False
40                          # Save start and end times
41                          start_times.append(tmp_start_time)
42                          end_times.append(get_end_time())
43  done # End of for loop
44  # save the distances of all the trips
45  distances = calculate_distances()
46  ### PHASE 2 ###
47  # Merging trips:
48  for start, end in start_times, end_times do:
49          # Obtain seconds between end of trip and start of next trip
50          if (start + 1)-end.total_seconds() > MAX_TIME_WITHIN_TRIP:
51                  merge_trips()
52  done # End of for loop
53  # Apply filters of average speed, and duration
54  validate_all_trips()
55  # Return results
56  return start_times, end_times, distances
```

Challenges when developing the algorithm:

- Identify the right approach to process the data: initially, we analyzed the data point by point but this approach was not effective. Once we moved to a slidding windows approach, the results were considerably better.

- Definition of thresholds for each variable used in the algorithm: 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 whole database we were able to identify some problems that were not possible to identify before. Tweaking these values has been challenging considering the different frequency in which data is collected, amount of data collected, variability of data, and different versions of the data collection software.

- Dealing with different number of records saved every minute and the frequency when they were collected: This was a challenge for all the stages of the analysis since it is more ideal for data analysis to have values collected

at a certain frequency and not one after the other for 10 seconds every minute. Since the values collected were very close in time to each other (within the 10 seconds that are collected), the data does not have enough time to show significant changes.

- Identify the data that should be considered for identifying 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 behaviour of each measurement during trips it was easy to determine that the Gyroscope and Linear Acceleration data was more suitable for this purpose.

- Definition of parameters related to the sliding window such as width and number of values over the limit: the frequency and the changing number of values collected presented difficulties since a slidding window generally needs to know all this information in advance and it is not necessarily easy to adapt to changes. However, after testing different values and confirming that the accuracy was not affected, the used values were determined in order to adapt to both 20 and 4 measurements per minute.

- GPS data: since this data was so unrealiable, an algorithm for cleaning it was required.

The next analysis was done using the survey answers from the users:
For each one of the answers analyzed, I obtained the number of trips and the total time spent on average to see how people have used the bike compared to their answers.

PREVIOUS EXPERIENCE OF USERS WITH EBIKES

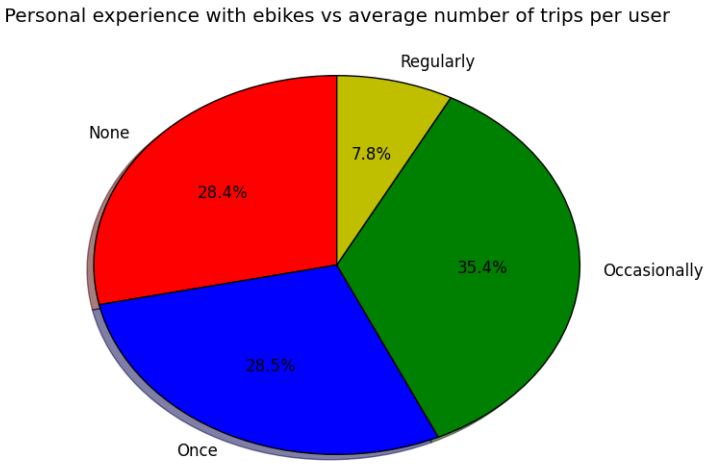

Figure 1: Average percentage of trips per user divided into their previous experience with ebikes
This graph shows the percentage of the total trips by users who answered that they have used an ebike "None" (never), "Once", "Ocassionally", and "Regularly". The total number of trips for each category was divided by the number of users on each category to estimate an average of their use. Here, we can see that people who have used ebikes before, also have the lowest number of trips; this could be caused since people who are not familiar with ebikes are curious about them and once they get them, they feel more interested in using them. Also, we can see that people who have occasionally used an ebike before are the ones who have had more trips on them, but there does not seem to be a considerable difference with the rest of the groups.

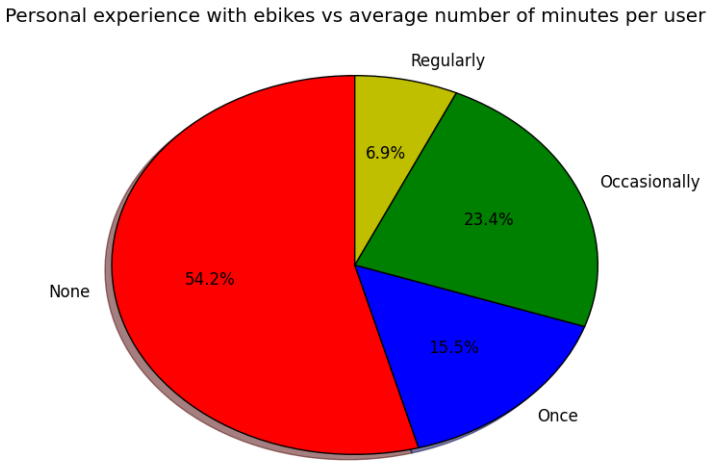

Figure 2: Average percentage of time per user divided into their previous experience with ebikes

This graph was obtained in a similar fashion as the previous graph but considering the time spent on trips, instead of the number of trips. Here, it is easier to note that people who had no previous experience with ebikes have used them for longer times. That means that these people are willing to take longer trips to farther places using the bike. Also, similarly as before, people with occasional interactions with ebikes are one of the main groups who spent longer times riding the ebikes.

INTEREST OF USERS ABOUT EBIKES

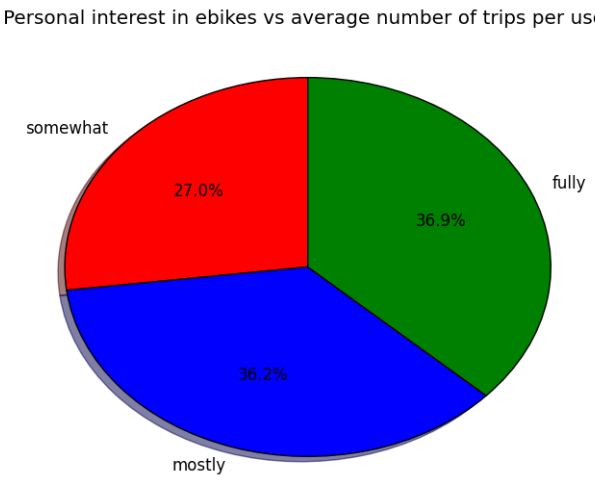

Figure 3: Average percentage of trips per user divided into their interest about ebikes

The previous graph presents three groups based on the answer of the users to identify themselves with the sentence "I am interested in ebikes". Their answers varied from "Somewhat applies to me", "Mostly applies to me", and "Fully applies to me". Similarly as before, this was analyzed by comparing the average percentage of trips taken by each one of the groups. Here, we can see that the interest in ebikes did not make a difference in the number of trips each participant took; all the values are very similar with a slightly lower value for people who presented less interest in ebikes.

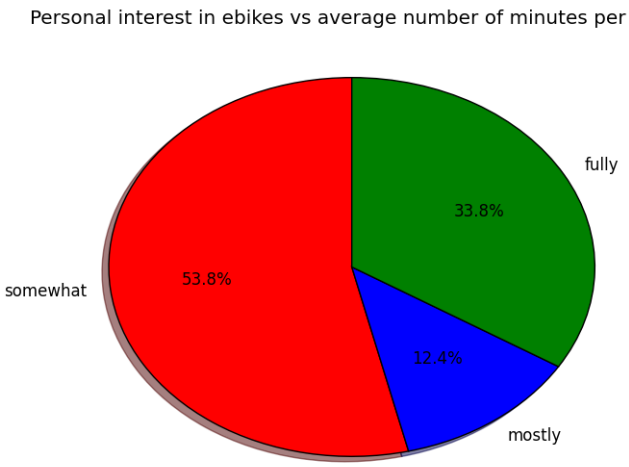

Figure 4: Average percentage of time per user divided into their interest about ebikes

This graph replicates the analysis from the previous one, but using the total riding time instead of the total number of trips. Here, we can see that even though people who showed less interest in ebikes had fewer trips, they tend to have longer ones convering over 50% of the total time of trips. This could probably mean a change in perception about ebikes and an increase of interest after their experience with them.

USERS PLANNING TO BUY AN EBIKE

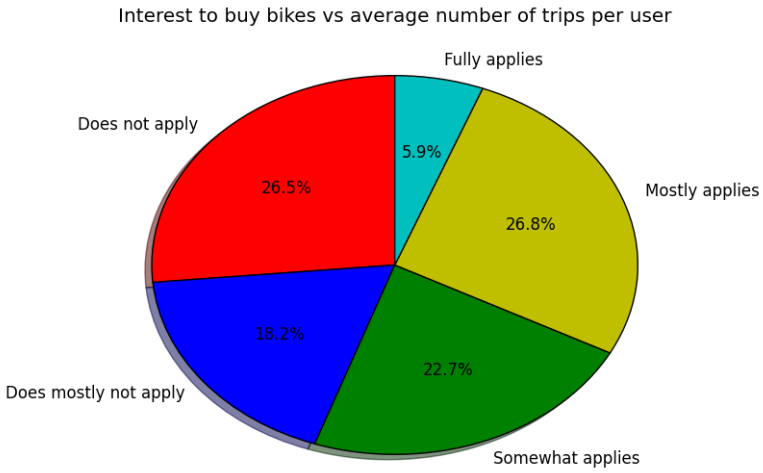

Figure 5: Average percentage of trips per user divided into their intention to buy an ebike

The previous graph analyzes the average number of trips divided into the 5 groups of answers to people identifying with the following statement: "I plan to buy an ebike within the next couple years". Here, we can see that people with intentions to buy an ebike have taken fewer trips and the rest have very similar number of trips with the lowest being people who responded that the statement "Mostly does not apply" to them.

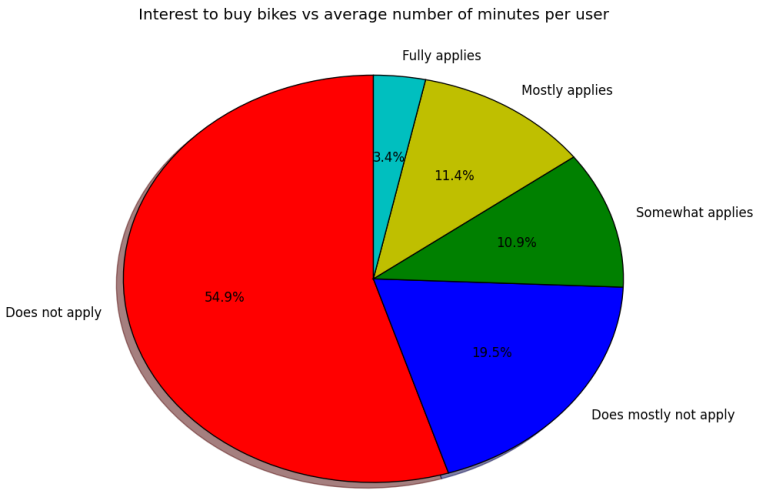

Figure 6: Average percentage of time per user divided into their intention to buy an ebike
Again, this graph is similar to the previous one but instead of using total number of trips, it uses total time riding the bike. This pie shows that people who are not planning to buy an ebike have been using theirs for a longer time. This could mean that people are actually interested and willing to use an ebike but other factors influence their decision to actually buy one (eg. price).

USERS CONSIDERING SUBSTITUTING THEIR REGULAR BIKES FOR EBIKES

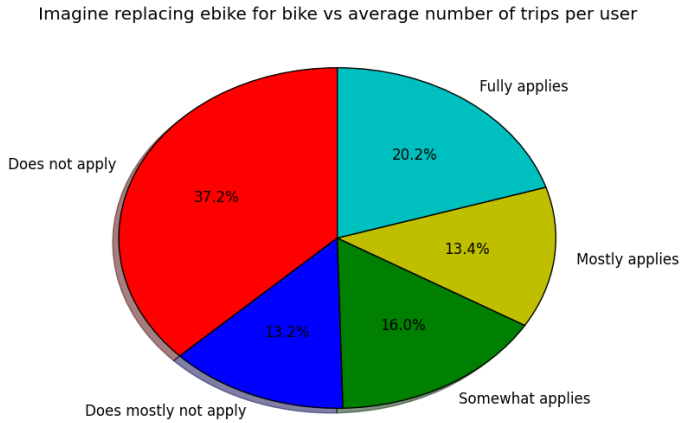

Figure 7: Average percentage of trips per user divided into their intention to replace their regular bikes with an ebike

The previous graph presents the total number of trips divided into the answers of users when asked to identify themselves with the statement: "I can imagine substituting my current bicycle with an e-bike". The plot shows that again, people with less interest in replacing their bikes with an ebike have been the ones who have more trips. Similarly as before, this could be caused by an increasing interest on ebikes after having a consistent experience with them.

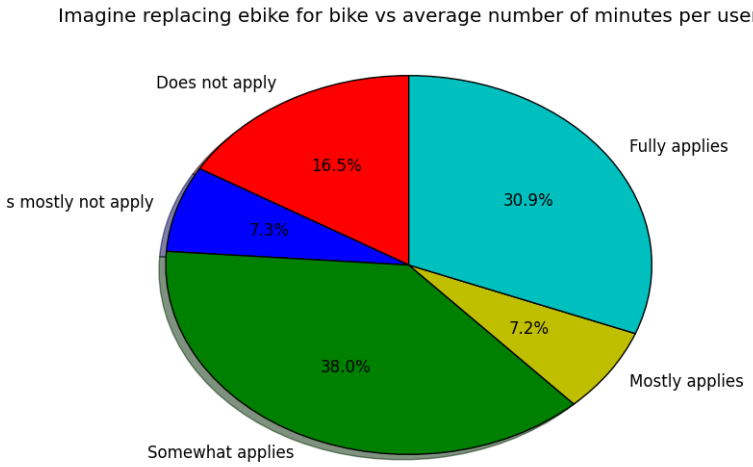

Figure 8: Average percentage of time per user divided into their intention to replace their regular bikes with an ebike

Similarly, this graph uses the same data as the previous one but replacing the number of trips with the total amount of time of their trips. This graph shows that people with more interest in replacing their bikes with an ebike have a tendency to use it for longer times. This could be caused because they try to use their ebike more often and for different purposes due to the advantage of having a battery and being able to reach longer distances with less effort than using a regular bike. Following these two groups, people who did not plan to replace their bikes with ebikes are the next group that rides their bikes the most.

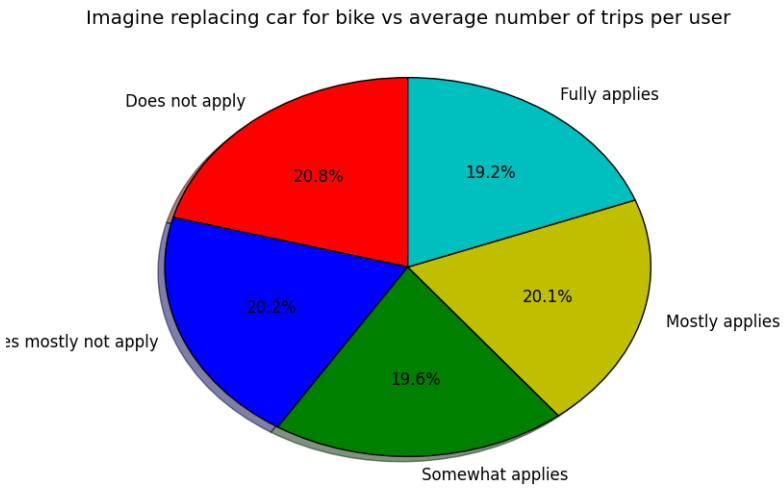

Figure 9: Average percentage of trips per user divided into their intention to replace their cars with an ebike
The plot presents the total number of trips divided into the answer of participants when asked about the following statement: "I can imagine substituting my car with an e-bike". Here we can see no difference between people who responded any of the possible 5 options. This means that the number of trips a person takes on the ebike is not influenced by their opinion about replacing their cars with ebikes.

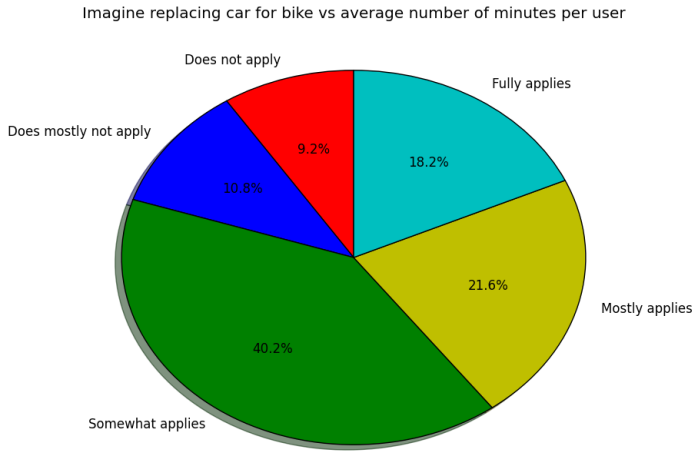

Figure 10: Average percentage of time per user divided into their intention to replace their cars with an ebike
This final graph shows that people with some/full interest in replacing their cars with bikes tend to go for longer rides than people who are not willing to replace their cars with ebikes. This means that distance and duration of trips is a factor that people consistently consider when they have to choose between using their car or ebikes. People with less interest in getting rid of their car use ebikes for shorter distances, and most likely use their car for longer trips.