Weekly report: June 5-June 12, 2015

Ivan Rios S.

June 12, 2015

1 Goals for the week

- Redesign of trip detection algorithm

2 Activities

- Read literature about self tuning algorithms
- Identify reference trips with precise data that can be used for the analysis
- Define the variables that will be used by the algorithm to identify trips
- Determine the accuracy of each variable.
- Determine the order in which the variables will be analyzed ("committee of experts").

3 What I learnt

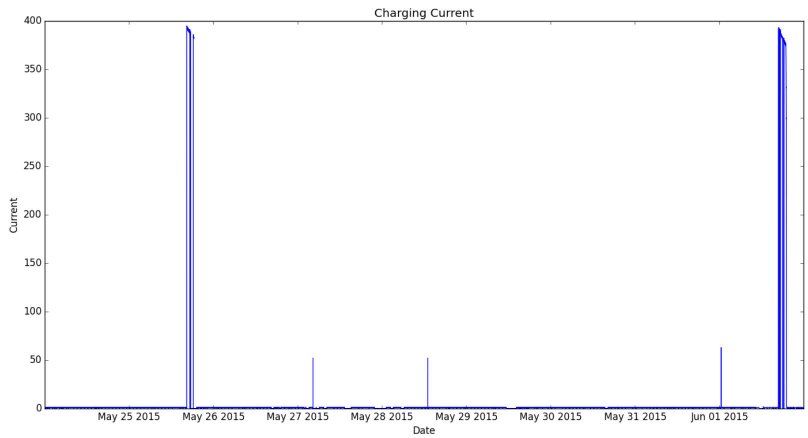
Once looking at the data, I realized that there are some considerations to take in count:

- 20 to 25 points of data are being collected every minute with a time of approximately 50 seconds with no data being collected.
- Some data is highly influenced by the way we are collecting data. Especially, the two most affected are distance between points and instant speed (based on the GPS data). This decreases the accuracy of the trip detection algorithm since noisy GPS data generally result in confusing conclusions.

Since I am making a complete change of the original algorithm, I started by analyzing each of the data that could be used for the algorithm. The following is a list of each of the records that can be used for the analysis:

1. CHARGING CURRENT:

The following plot shows the behavior of charging current over time:



The previous plot shows two very clear charging events which match with the real events.

Observations:

- During times when the battery is not being charged, there is some variation between 0 and 5.
- All the charging events show an increase in the value of the current. However, this does not stay constant at the new level reached but it has random values that go down to 0.
- There are random high values but they they do not happen consecutively.

Potential use:

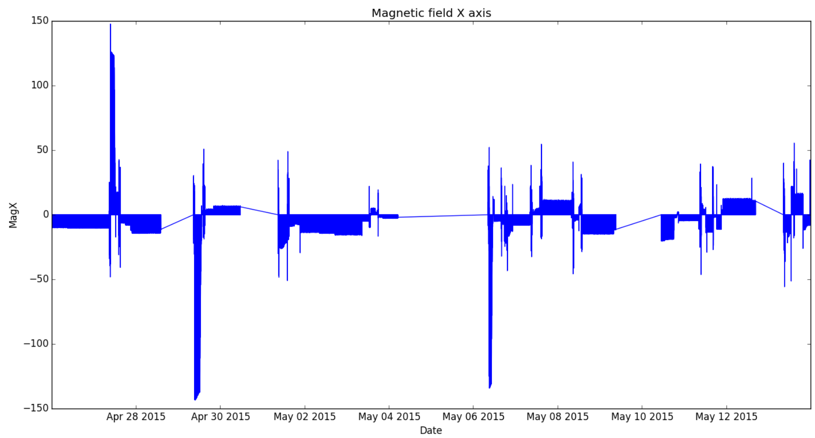
- If a charging event is detected no trips could be happening.
- If a charging event is not detected, more information is required to make a decision.

Potential way to identify if it is being charged:

- A considerable increase in the average of the last points would mean the battery is being charged.

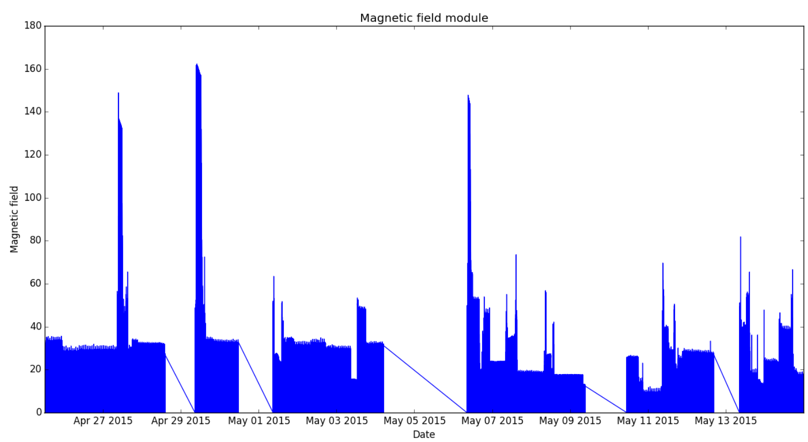
2. EFFECTIVE MAGNETIC FIELD

In the previous algorithm, the data used was the magnetic field measurement in the X axis. Looking at the graph of this data we get the following:

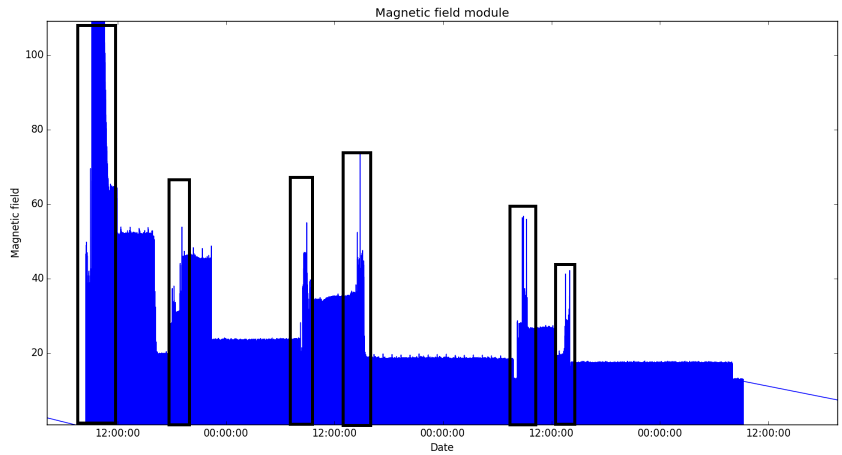


As shown here, variation (positive and negative) make it difficult to be able to identify trips. There generally are peaks (positive or negative) that define the existence of trips; however, once the trip is over, the values do not go get stable at a specific point which makes it a challenge to properly identify those events. After reading some literature, I realized that considering the movements that the bike goes through, the most efficient way to use this data is to obtain the effective magnetic field (module obtained form the 3 components).

With this change, the following plot was obtained:



Which shows a more organized and clear plot. However, it is still challenging to identify trips since a stable state is not immediately reached by the system after each trip. Again, the trips show peaks but it is easier to identify since all of them are positive. Here is a closer look at a couple dates that contain some trips:



Observations:

- The graph shows that trips (black rectangles) are reflected in the data as positive peaks. However, these peaks do not go back down when the trip is over which means it will be challenging to use this information.
- Even though I obtained similar results analyzing my data and Professor Golab's data, Professor Keshav's data shows a different pattern since the lower average is much higher.
- The data raises immediately when a trip starts.
- There is little or no noise. Generally, all the values stay in a stable value and they don't have random high or low values.

Potential use:

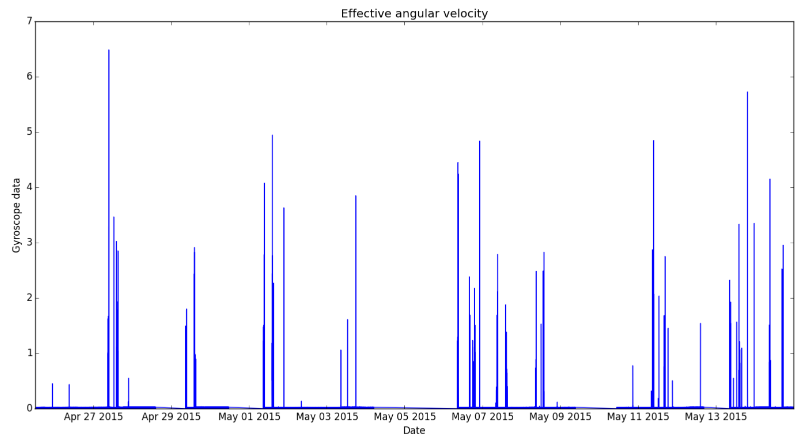
- It could be used a way to confirm that a trips has started. It does not seem useful to look at this data when determining if an active trip is over.

Potential way to identify if there is an active trip:

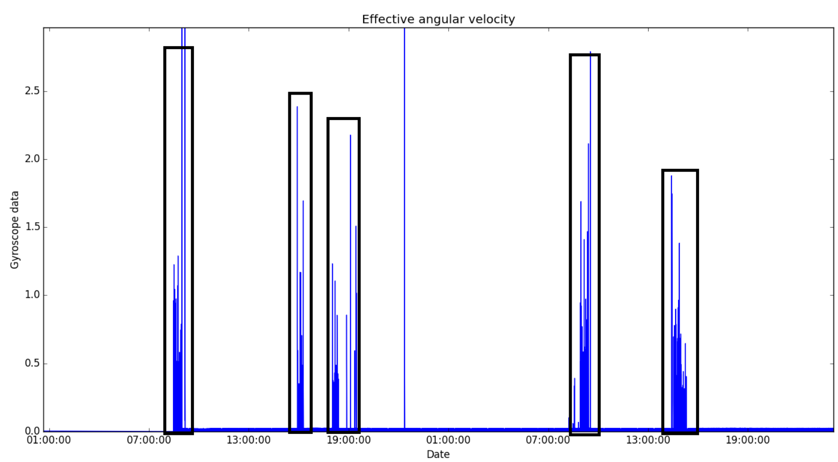
- A considerable increase in the average of the last points would mean a trip has started.

3. GYROSCOPE:

The following graph was obtained when plotting the effective gyroscope data over time:



This graph shows very clean stable data which could potentially help detect trips. The gyroscope data is very sensitive and is able to detect movement with high accuracy. In the same way, it does not contain that much noise. With these insights I plotted the data for specific trips:



Observations:

- The data helps to easily identify all the trips (black rectangles) and they match with the actual data.
- The graph shows very few noisy values, and they are never consecutive so it would be possible to identify them properly.
- The trips are defined properly since the sensor is able to pick up data immediately and it does not depend on external factors such as GPS signal.
- The data values collected when there is no movement/trips, do not go

over 0,25. This is a very positive feature since the values obtained during a trip go over 1,0.

- There are randomly low values during trips but generally they do not go lower than 0,40.
- The third trip is detected when the bike is in a car. For this reason, the data should be cleaned using other variables to confirm it was a bike trip.

Potential use:

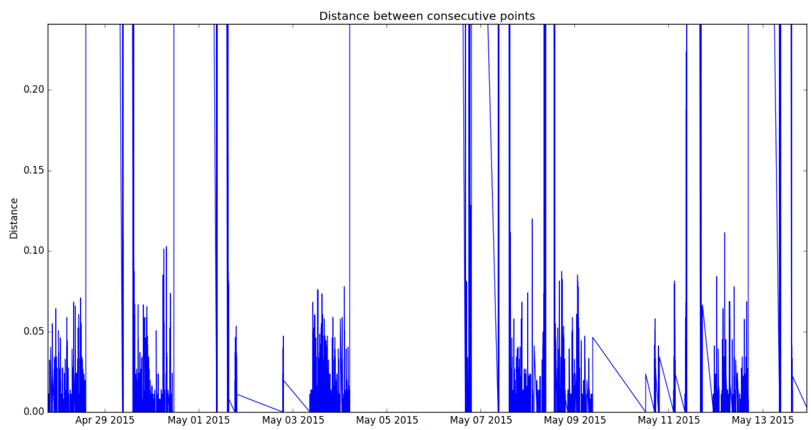
- This variable can be used to identify both start and end times of trips. The data is very accurate and can be the leading variable to identify trips.

Potential way to identify if there is an active trip:

- A considerable increase in the average of the last points would mean a trip has started.

4. DISTANCE BETWEEN CONSECUTIVE POINTS:

The following plot was obtained by looking at the distance between consecutive points:



Observations:

- Great amount of GPS values are 0 which means the bike is not on movement and is probably stored in a place where there is not GPS signal.
- There is a considerable amount of peaks in the data which is caused by the noisy GPS data. This is also influenced by the way data is collected; in this case it is possible to see that there are many low distances while the phone is collecting data but there are considerable changes during the idle time.
- It might be better to have a representative distance between each time the phone is in idle state in order to clean the noise in the GPS data.

Potential use:

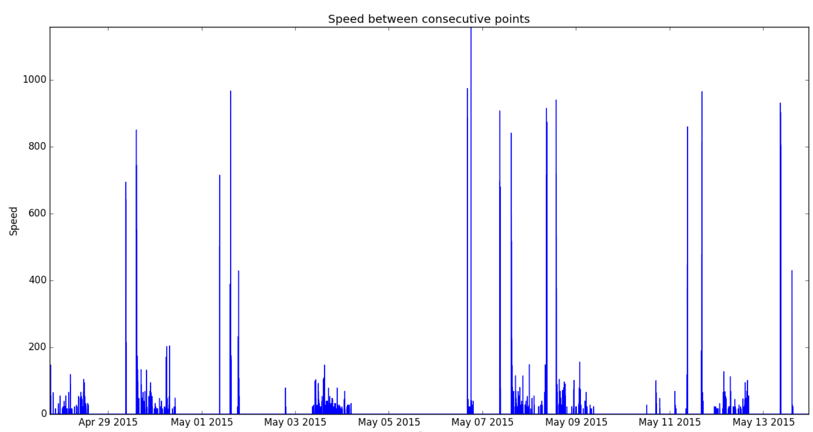
- When a certain number of points are null or zero, it is possible to assume that there is no trip going on.
- This data could be used a input for other variable to determine if there is trip; for example, confirming the distance between points we could clean the gyroscope data and make sure that we are just identifying bike trips.
- Even though it can give us useful information about a trip (such as the total distance), this data is not able to identify trips on its own but it should be used as a tool for other variables to increase their accuracy in identifying trips.

Potential way to identify if there is an active trip:

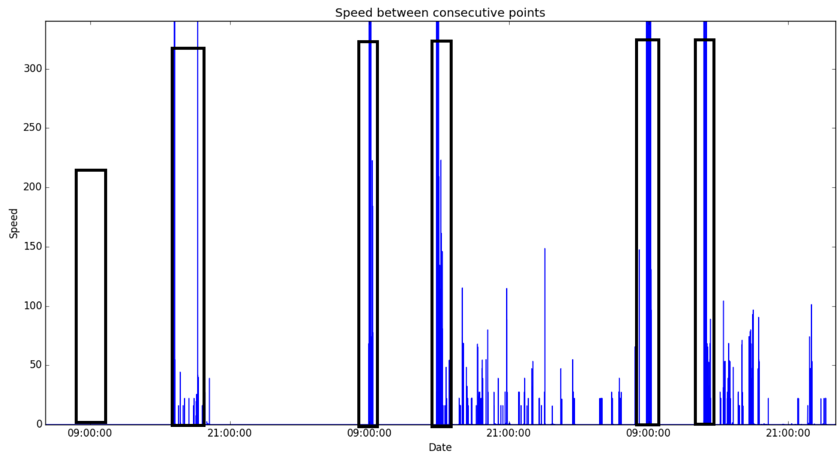
- A considerable increase in the average of the last points would mean there is potential movement of the bike and it might be in a trip.

5. SPEED BETWEEN CONSECUTIVE POINTS:

Calculating the speed between consecutive points based on the GPS data showed these results:



As shown in the previous plot, the data is not very noisy and actually seems to identify trips with acceptable accuracy. However, as we can see on the axis, the speed recorded goes regularly to 800 m/s which is not a realistic value. This happens because of the GPS problems. Next, we can see the result obtained when trying the identify trips with this data:



Observations:

- The data is not very stable and shows constant jumps in the values. However, the higher peaks work properly to identify trips (black rectangles).
- One trip is missing which is probably due to missing data since the rest are properly identified by the high peaks.
- The lower average is different between data from different users. This is obviously influence by the biking habits of the bike owners and also the characteristics of paths they generally use.
- As we can see, this data is useful as a tool to identify start and end of trips but not to actually estimate the speed of the bike in a certain moment.
- Even when the bike is not on a trip but has access to GPS signal, the data shows considerably high values calculated.

Potential use:

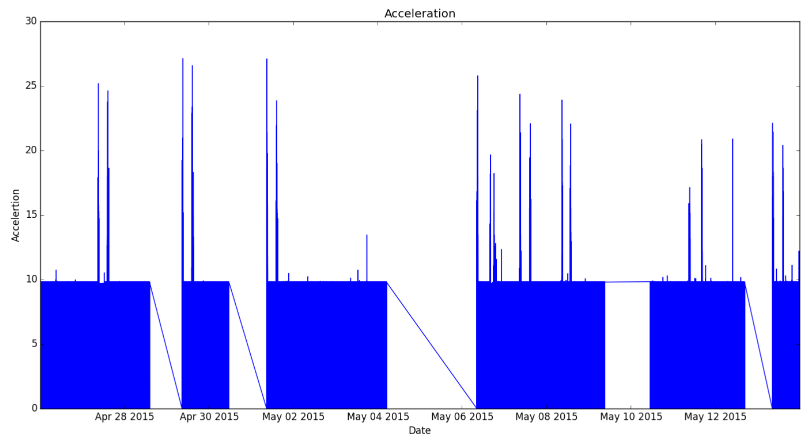
- This data could be used as a confirmation that a trip is active or has ended; however, it is not completely reliable to identify perfectly start and end times.

Potential way to identify if there is an active trip:

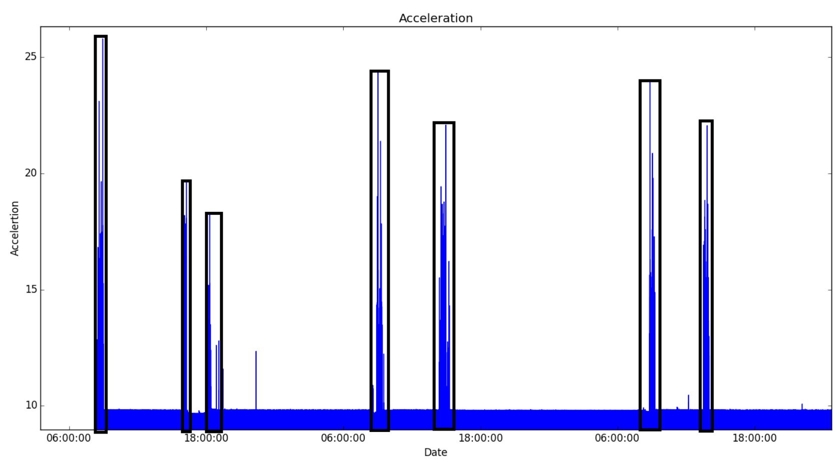
- A considerable increase in the average of the last points would mean there is potential movement of the bike and it might be in a trip.

6. ACCELERATION

This data considers gravity as part of its components so the acceleration could be affected by how it is positioned. For this reason, I calculated the effective acceleration (module of the three components), and obtained the following results:



The previous plot shows that acceleration has a constant value of approximately 10 (the value of gravity) with some variations. Here, it is possible to see that trips seem to be easy to identify; here are the results for obtained when looking for trips in specific dates:



Observations:

- The data is very stable and shows very few random variations over and under the values of 10 m/s2.
- The few random high points do not happen consecutively.
- There are some random low values in the middle of trips but the overall average still increases during this time.

Potential use:

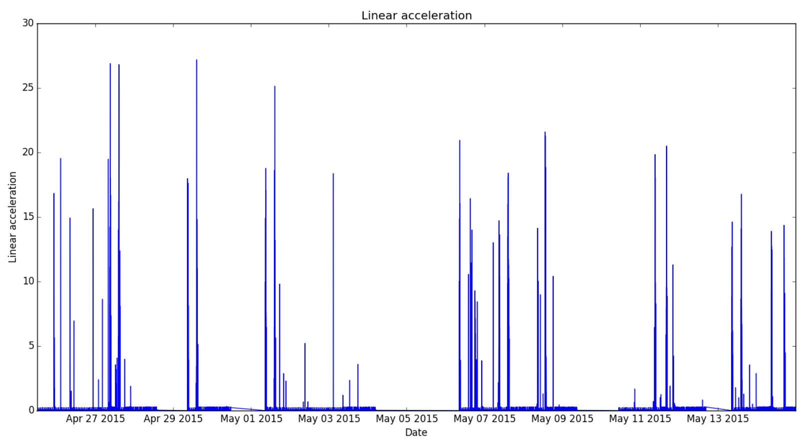
- This variable can be used to identify both start and end times of trips. The data is very accurate and can be the leading variable to identify trips.

Potential way to identify if there is an active trip:

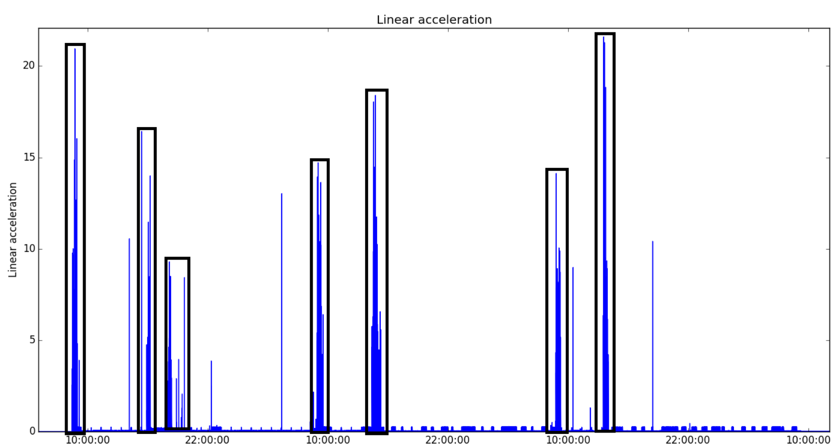
- A considerable increase in the average of the last points would mean there is potential movement of the bike and it might be in a trip.

7. LINEAR ACCELERATION:

Finally, similarly, I analyzed the data of linear acceleration over time; here are the results I obtained:



Again, similar results are obtained here as before but with the difference that gravity is not a component so the overall average is close to zero. Here are the results when identifying trips with this data:



Observations:

- The values are very stable between 0,25 and 1,7 when there are no trips.
- Real trips generally have values that go above 2
- There are very few random high noisy values which are not generally consecutive.
- Very accurate for trip detection since the changes in the average increase considerably.

Potential use:

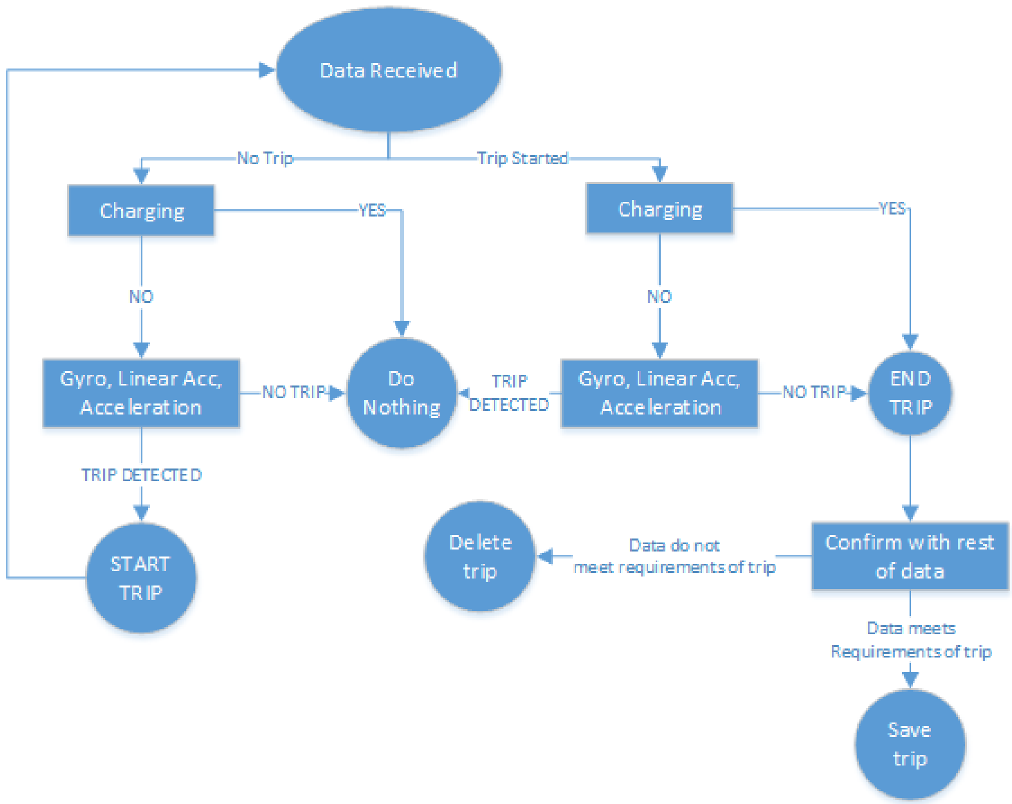
- This variable can be used to identify both start and end times of trips. The data is very accurate and can be the leading variable to identify trips.

Potential way to identify if there is an active trip:

- A considerable increase in the average of the last points would mean there is potential movement of the bike and it might be in a trip.

After this analysis of the data available, the following general conclusions can be obtained:

- The variables with more accuracy for trip detection are gyroscope, linear and regular acceleration. They could be use as primary variables to make decisions about the data because of their accuracy. They could be used for identifying both beginning and end of trips with very high accuracy.
- The main variables chosen for trip detection need support variables to make them more accurate, and to be able to identify other information about the trip, for example, if the movement fits the profile of a bike trip.
- Charging events could be identified with high precision so we could use this variable to help make a decision about the bike being charged and hence not being on a trip.
- The rest of the variables could be used as support for when the variables cannot offer an accurate decision. Also, they will be used as a filter to identify potential car trips and other causes that could lead to wrong conclusions.
- The rest of the data will need noise cleaning in order to obtain accurate conclusions. However, it is important to consider that the new algorithm should only use this data for confirmation purposes.
- The algorithm will handle decisions in the following way:



4 Proposed goals for next week

- Finish working on the algorithm. I have already started implementing it variable by variable in order to see what accuracy can be reached and when do we actually need additional information to make a decision.
- Make the new algorithm available in the webike webpage