בס"ד

# Otonomo Challenge - Fuel Theft

Submitted by Nomi Hadar

1. **A fuel theft:** I assumed that a theft can occur both when vehicle is on or off, therefore I distinguish between two cases:
   a. **A fuel theft when engine is on:** a significant increase in the decrease-rate of the fuel-level. Within a time-window, the decrease-rate is defined as the difference size in the amount of fuel, divided by the window size (liters / time-unit).
   This definition makes the reasonable assumption that thefts happen within a relatively short time (say, up to quarter of an hour), therefore during theft, the decrease rate of the fuel would be much higher than in the ordinary usage.

   I believe this is a proper definition because theft is defined relative to the fuel consumption of the individual vehicle. This means that a "significant" increase is relative to the rate at which the vehicle usually loses fuel, and not an arbitrary value.
   The disadvantage of this definition is that it requires to define a quantile threshold. More important, it will always find cases, resulting in false positives. To solve this, we can require a threshold of minimal stolen amount (threshold can be constant or relative to tank size threshold).

   b. **A fuel theft when engine is off:** Theft occurs when the vehicle is off and the owner is probably not around. A theft is when the fuel-level measured before turning off the vehicle, is higher compared to the level measured on the next turning on. A difference that is a greater than the minimal stolen amount threshold is considered a theft.

2. **Method for finding a fuel theft:**
   a. Data processing:
      i. Sort data by vehicle ID and time.
      ii. Smooth fuel-level data using the averaging method.
      iii. Calculate the difference in time and fuel-level of each observation, compared to the previous observation. This will be referred as "delta time" and "delta fuel-level", relatively.
   b. Calculating the difference rate of the fuel-level:
   Using a rolling window of minimal specified time (I used 10 min), divide the sum of delta time values in the window, by the sum of delta fuel-level values.
   Rolling window is necessary since time intervals are very small (usually 1,800 milliseconds). More important, since fuel-level values are still noisy, the window is necessary to discover a true trend of increase / decrease in fuel-level.
   c. **Find thefts when engine on:**
      i. Choose negative rates only, which mean decreases in fuel level.
      ii. Extract the highest rates values. A high rate (in absolute value) means a sharp decrease in fuel level. This is done by taking the values under the specified quantile (under because values are negative).

iii. Filter cases where delta fuel-level is smaller than the specified minimal stolen amount.
iv. The resulted observations are considered as thefts.

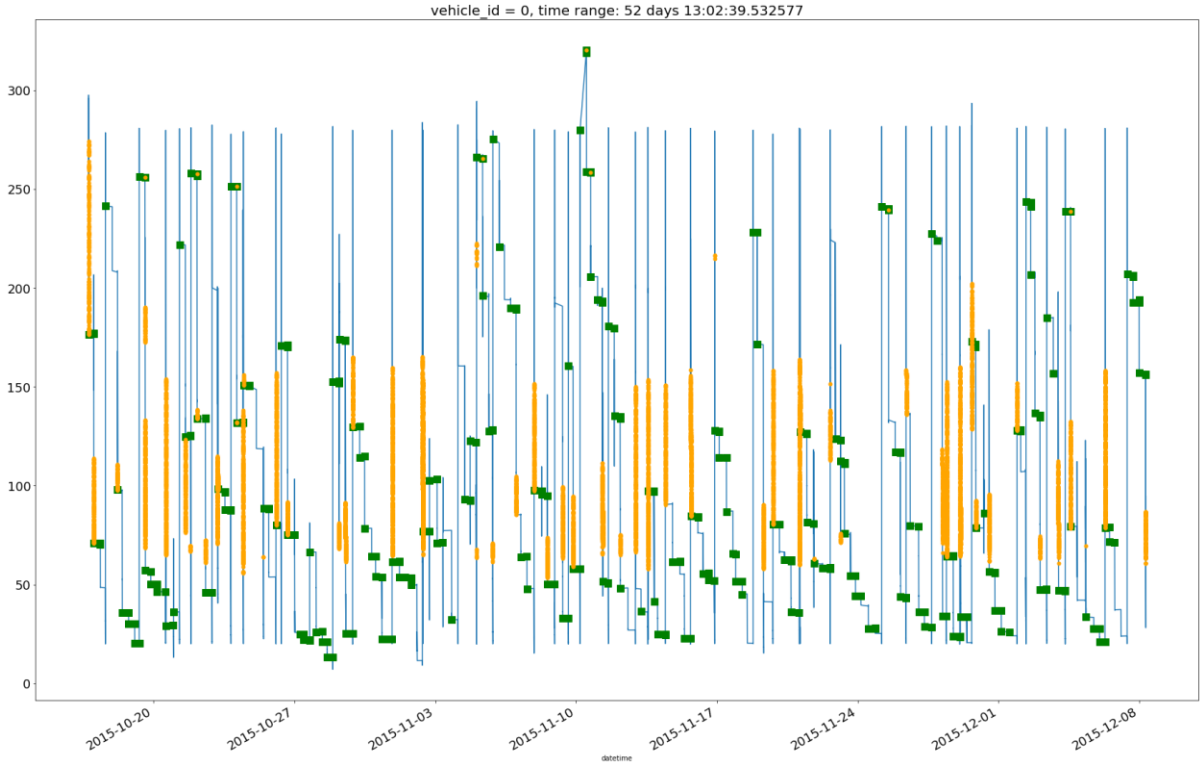
**d. Find thefts when engine off:**
For each observation where engine is off, calculate the average fuel-level observed during the few minutes (I used 5 minutes) before turning off, and the average fuel-level observed during the few minutes after turning on. If the difference between the first quantity to the second is greater than the specified minimal stolen amount, consider it as a theft.
Another way to do it (not implemented): For each observation where engine is off, sum the differences in fuel level observed a few moments after turning on. If sum is greater than the specified minimal stolen amount, considered as a theft.

**e.** Reduce results overlapping in time (not sure how to do it).

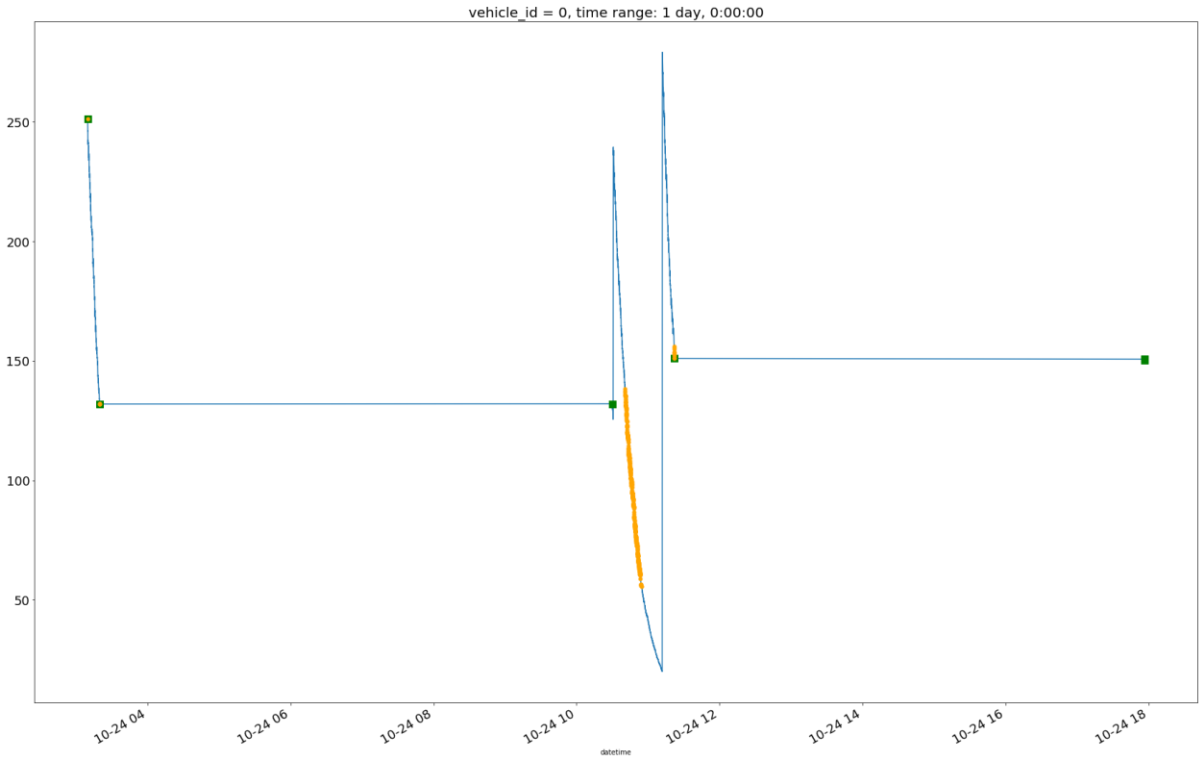
**f.** Find the thief and hang him.

3. File named "fuel_thefts.csv" is enclosed.

**5. Show a fuel theft.**
For example, the following graph shows the fuel-level of vehicle 0 against time series. In orange: suspicions of theft. In green: engine is off. We can clearly see that there are too many… I suspect that rate of false-positive is high.




Zoom in to one of the thefts:




We can see a sharp decrease from fuel-level of ~240L to fuel-level <50L, within less than two hours (engine is on). Disadvantage: only part of the descending line is marked as a theft.

Things I would do better:

- I wanted to demonstrate a theft when engine is off, but a little bit having trouble with it. I think that I have a bug. In retrospect, I would change the method to find thefts when engine is off. I would first divide the data into sections, where each section is "a drive" – from switching on to switching off, and work with these sections similarly (i.e., find difference before and after the vehicle "rest"). Probably this way is less prone to troubles.
- I would use plots with range slider (like offered in *plotly* – I only knew it now) and then zoom-in more comfortably.
- I would annotate the green points with "switch on" and "switch off" because currently it is not always easy to distinguish.
- Also, I would reduce noise, for example by filtering single cases of thefts.

7. Insights on the findings: Avg stolen fuel amount: 82.1 liter.
   I calculated it by resampling data time every 30Min, and if was a theft during this time interval, sum the difference in fuel level, and compute the mean.
   Other statistics that can be deduced are the avg. of duration of a theft, percentage cases where engine was on, and distribution of stolen amounts (not just the mean).

9. Additional pieces of information would help:

- Distance, velocity and average fuel consumption of vehicles of same type: these data (or part of) can help estimate if the observed loss rate of fuel is consistent with the usage of the vehicle, or not. It can be used to calculate quite accurately the amount of fuel that is expected to decrease in a specific usage (for example, driving in traffic, highway, etc.). In such case we don't have to use the quantile method, but to calculate for each time window if the decreased amount is what we expected to see. This can help better predict thefts when engine-on.
- Location: probably there are some locations that are more prone to thefts than others. This can help determining a theft when vehicle is off (so this supposed to be an easier case).
- Labels of whether or not a case is a theft. Known cases can help find thefts patterns that didn’t occurred to my mind.