Otonomo Challenge - Fuel Theft

You are provided with a CSV file names fuel_level.csv, this file represents fuel level reading through time taken from vehicles. Each row contains the following columns:

- fuel_level - The amount of fuel observed.
- time - The time this sample was received.
- vehicle_id - The ID of the vehicle.
- engine_on - Was the engine on during the observation?
- provider_id - The ID of the provider from which we received this sample

You can assume the following:

- Samples from the same vehicle will contain the same vehicle and provider IDs.
- Fuel level is given in liters.
- Time is given in millisecond since epoch.
- Fuel level can be noisy (i.e., some readings are incorrect), all other fields have the correct value in them

Please answer the following questions:

1. How would you define a fuel theft? Why do you believe this is a proper definition? You can base your answer on the provided data.
2. Given your definition, describe a method for finding fuel theft.
3. Using your suggested method, find all (vehicle_id, time) pairs where you believe a fuel theft has occurred.
5. Show that the pairs you have found in the previous question do indeed represent a fuel theft (i.e., via a figure), if they aren't please try and explain what could have gone “wrong”.
7. What types of insights can you provide from your findings (e.g., the avg. stolen fuel amount is...).
9. What additional pieces of information would help you better your prediction? How would this affect your answers to Q1 and Q2?

Expected results:

- We expect you to submit a jupyter notebook containing all code, figures, and relevant documentations so that we could understand your thought process.
- If needed, you can submit an additional README file containing the more “textual” answers.

You can reach out to us with any questions.

Good luck!
Otonomo DS Team