**Step 1**

* Remove entries of cars that have only gone through 1 siteid such as c1 below



* After wards we will have to deal with the more complicated case which exist for time differences such as c3 below which also needs removal
* 

**Step 2**

* Group the data by cars and then sort by time in order to get the routes for the cars
* 
* Using the time, calculate the time in range between consecutive siteid sensors for each car.
* 
* However, we should now create an id to keep track of which start-end trip corresponds to which entry such as like in below:
* 
* 
* This will be important later on when we want to remove some “trips” that are found to be outliers and also to mark the end and start of trips.

**Step 3**

* When we repeat this process again with a lot of car routes we should be able to build a distribution of actual trips made and the consecutive differences of time in range between all the possible two siteid routes that have been made.
* 
* Thus in this instance we were able to detect one of the trips as an outlier
* In theory we could consider all possible combinations/links of siteid1 to siteid2 but will be very tedious and we most likely won’t have the time for all the scenarios. So it may be assumed that since we have enough data, all the cars should go through the sites within some logical order that eliminates a lot of routes that don’t make sense practically.

**Step 4**

* Once we have been able to identify all outlier trips, we are then ready identify when trips begin and end. And it follows the following rule:
  + If a tripid is an outlier then it is marked as a trip point. This trip point can either later be the start or end of a trip depending on how many trips are after it for the same car.
  + If a car has no outliers, then we simply denote the first and last entry as the start and end of the trip.
  + 

**Step 5**

* Then we may preprocess the data into however form we need to do other stuff. But one thing I am not sure is if there is like a data structure that we could use to do some of the stuff effectively.