* What information can be extracted from the ? Any informative visualization that can improve the understanding?
* How would you go about building an interactive view of this data and its possibilities?
* Are you missing some vital information in the data? What should be added to make your life as a data scientist easier, and what new possibilities could this generate?
* A hot topic is understanding how goods flow in the European transport market. Can this data be used to estimate transport flows on the roads?
* Are there clear symptoms indicating whether a vehicle has quality issues or not?
* Some vehicles have unknown issue status (file “\_cx.csv” below). If we treat them as a test data set, could we already now, based on the patterns from c1 and c2, indicate if they are likely to have issues or not?
* What recommendations would you give our stakeholders for further analysis and next steps, given the information you have found in the data?

# Information on provided data

Three files have been extracted by our data engineers. The data is divided into two groups depending on if a vehicle had issues or not. Vehicles for which the issue status is still unclear have been separated into a third group.

* trackdata\_c1.csv – first group of vehicles
* trackdata\_c2.csv – second group of vehicles
* trackdata\_cx.csv – vehicles with unknown issue status

The files contain GPS traces of the vehicles, with each vehicle associated with one trip (representing its typical usage pattern). The format is similar in all files and the columns are:

* uid – unique vehicle identifier
* row\_number – counter for the data points that make up a trip (i.e. a “time coordinate” for the trip)
* lat – latitude position in WGS84
* lon – longitude position in WGS84
* ele – elevation above sea level in meters