**7. car sales.csv**

# Context

This dataset was collected by me from car sale advertisements for study/practice purposes in 2016. Though there is couple well known car features datasets they seems quite simple and outdated. Car topic is really interesting. But I wanted to practice with real raw data which has all inconvenient moments (as NA’s for example).

This dataset contains data for more than 9.5K cars sale in Ukraine. Most of them are used cars so it opens the possibility to analyze features related to car operation. At the end of the day I look at this data as a subset from all Ukrainian car fleet.

# Content

Dataset contains 9576 rows and 10 variables with essential meanings:

* car: manufacturer brand
* price: seller’s price in advertisement (in USD)
* body: car body type
* mileage: as mentioned in advertisement (‘000 Km)
* engV: rounded engine volume (‘000 cubic cm)
* engType: type of fuel (“Other” in this case should be treated as NA)
* registration: whether car registered in Ukraine or not
* year: year of production
* model: specific model name
* drive: drive type

Data has gaps, so be careful and check for NA’s. I tried to check and drop repeated offers, but theoretically duplications are possible.