“data1\_3.17.xlsx” records 27077 pieces of data from 78 drivers, 43 of them have demographic information. And 63 drivers have records of co2 emission (actually, a zero co2 emission means the driver did not drive or just pulled out the OBD that day). The data was collected from June 2019 to September 2020, and we are still collecting data right now.

**group:** 4 groups in total, and drivers in different groups have access to different functions of the App. Specifically, drivers in group 01 have access to both the ranking and alert functions; group 02—ranking and no alert; group 03—alert and no ranking; group 04—no ranking and no alert.

**ranking** : whether to show a ranking of driving performance to the driver in the experiment

**alert** : whether the driver can be alerted when detected risky driving behavior

**license plate number: (**literal meaning) “1.22E+16” is the data error.

**age:** a driver’s age

**gender:** a driver’s gender

**accident:** whether the driver had any accidents before

**driving experience:** the driving months of the driver

**gasoline consumption**: fuel consumption/day

**CO2 emission**: the amount of co2 emission/day, unit kg.

**Date**: **(**literal meaning)

**Mileage:** mileage/day

**Total mileage**: a driver’s accumulated total mileage

**Time:** total hours of driving/day

**Speed\_KMH** : average driving speed/day

**rapid\_acceleration:** number of rapid acceleration detected/day

**rapid\_deacceleration**: number of rapid deacceleration detected/day

**sharp\_turn**: number of sharp turn detected/day

**fatigued driving**:times of fatigued driving/day

**night driving**: hours of night driving

**driving score** : driving behavior scores (a higher score means better performance)

**app\_usage** : whether the driver have used the App that day

**log\_*variable*:** Some variables were remarkably skewed and I transformed them with logarithmic transformations. The variables named like “log\_xxxx” are new ones after transformation.

**log\_*variable*\_1**: We used the mean of certain “log\_xxxx” of each driver to fill the missing values. The variables named like “log\_xxxx\_1” are variables after filling the missing values.