**Rules of the Game**:

1. MANDATORY: Please do not make either your solution or this problem statement publicly available by, for example, using github or bitbucket or by posting this problem to a blog or forum.

**Problem Statement:**

We have a lot of driver partners on our platform. Whenever they are online, we get pings from their mobile phone every 15 seconds indicating that they are available.

You have been provided with 3 weeks (1 Jun 2017 to 21 Jun 2017) of training data and 1 week (22 Jun 2017 to 28 Jun 2017) of test data.

Training data contains driver id, gender, age, number of kids the driver has and all the pings that have been received (during the training data period).

We want to predict how many hours the driver will be online on a given day. So the test data contains driver id, and date (during the test data period). The test data also contains the actual online hours, which is what your model should predict.

We will be looking at Root Mean Squared Error or RMSE for short (lower the better) to see how good your model is. We have a separate held out test dataset against which we will evaluate your model for overfitting, underfitting, etc.