

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.
2. MANDATORY: Your submission should be in a standard archival format (.zip, .tar, .tar.gz, etc). Please **do not** use any proprietary archival formats.
3. We would really like to see how your solution has evolved. So please use Git for version control. We expect you to send us a standard zip or tarball of your solution when you're done that includes Git metadata (the .git folder) in the tarball.
4. MANDATORY: Please do not include any executable files, class files, pyc files, jars or other libraries or output from the build process.

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.

Expected Submission:

We expect you to submit all your code / scripts and a brief write up about your solution. You can even do it as an IPython (or any other) notebook.

Code:

- Please include all the code / scripts that you have written, even if it is a simple script to just explore the data.

- Include a README file with appropriate information.

Write Up:

- Include a short write up describing your approach to the solution.
- If you have tried different models, please include relevant details (along with their RMSE score) and reason for moving on to the latest model.
- Include any visualizations that you may have created.
- Include any interesting insights that you may have found in the data.
- Include any other information that you feel is relevant.
- Your write up should be preferably in pdf, odf, markdown or txt format.