**Dataset Description**

You are presented real-world user data from one of our games. The dataset includes user metadata and ingame behavior for each user. Ingame behavior for each user contains their first 15-day ingame data (e.g: The levels they played in their first 15 days, or the revenue they generated in their first 15 days).

The dataset also includes the target variable, the total revenue each user generated in their first 90 days. So, the task here is to predict day-90 total revenue of each user using their 15 days of data. The prediction should be done at user-level, that is, you should predict a value for each user, and optimize your predictions accordingly.

The training dataset is organized as three separate files. users\_train.csv includes user-metadata for each user, user\_features\_train.csv includes first 15 days of user behavior, and targets\_train.csv includes target values for each user. These files share a common ID column that you can use to match with each other.

The test dataset is organized as two files: users\_test.csv and user\_features\_test.csv. They have the same shapes and columns as the training dataset, except the target variables. Your submission should be a CSV file containing only two columns: ID column which should be the same as the ID column of the test dataset; and TARGET column, which should be the predicted value (the predicted day-90 revenue) for each row of the test dataset. An example submission can be found in sample\_submission.csv file.

As per evaluation of your submission, RMSE (Root mean square error) metric will be used. Lower RMSE values will result in higher scores.

Below you can find descriptions for each column in the whole dataset:

**ID**: Unique ID for every installation of the game (User ID)

**first\_open\_date**: Date of the first launch of the game

**first\_open\_timestamp**: Timestamp of the first launch of the game, in UTC timezone, Unix time in microseconds

**local\_first\_open\_timestamp**: First open timestamp in local timezone of the user

**country**: Country of the user

**platform**: Platform of the user; Android or iOS

**device\_category**: Category of the device; mobile or tablet

**device\_brand**: Brand of the device

**device\_model**: Model of the device

**has\_ios\_att\_permission**: Whether the iOS user has given ATT permission (true or false), false for Android users

**ad\_network**: Ad network the user has come from, null for organic users

**first\_prediction**: Initial predicted value of the user (in USD)

**RetentionD{i}**: Whether the user launched the game at i’th day (true or false)

**LevelAdvancedCountD{i}**: Number of levels the user completed at i’th day

**Level\_{i}\_Duration**: The time it takes for the user to complete i’th level (null if the user hasn’t completed i’th level)

**AdRevenueD{i}**: Amount of ad revenue (in USD) the user generated at i’th day

**IAPRevenueD{i}**: Amount of IAP (in-app purchase) revenue the user generated at i’th day

**TARGET**: Total amount of revenue the user generated in their first 90 days, this is the target value that you should predict