Take-Home Challenge: Relax Inc

By Muluemebet Ayalew
June, 2020

Question: Defining an 'adopted user' as a user who has logged into the product on three separate days in at least one seven day period, identify which factors predict future user adoption

Answer:

Data: Two data sets are given, takehome_user_engagement.csv and takehome_users.csv.

A user table ('takehome_users') contains 12,000 users who signed up for the product in the last two years. A usage summary table ('takehome_user_engagement') that has a row for each day that a user logged into the product.

Solution:

First, the data 'takehome_user_engagement' has to be grouped by user id and weekly frequency to compute the number of visits for each week. This helps us to count the number of visits per week and then to determine if a user logged in three separate days in a seven day period. To use this approach, the maximum number of daily logins has to be one. So that we can be sure that weekly visits/logins/ are coming from separate days. To confirm this, the data was grouped on a daily basis by user id and the number of visits are computed. The maximum number of visits on a daily basis was one.

After the data was grouped on a weekly basis and user id, the number of visits are counted. Then, the id of users whose visit was greater than or equal to three were extracted. Those users are called ‘adopted users’.

On the other table(takehome_users.csv), a new column “adopted_user” with label true or false was created depending on whether the column ‘object_id’ belongs to the list of adopted users id or not(see Table-1 for its frequency).

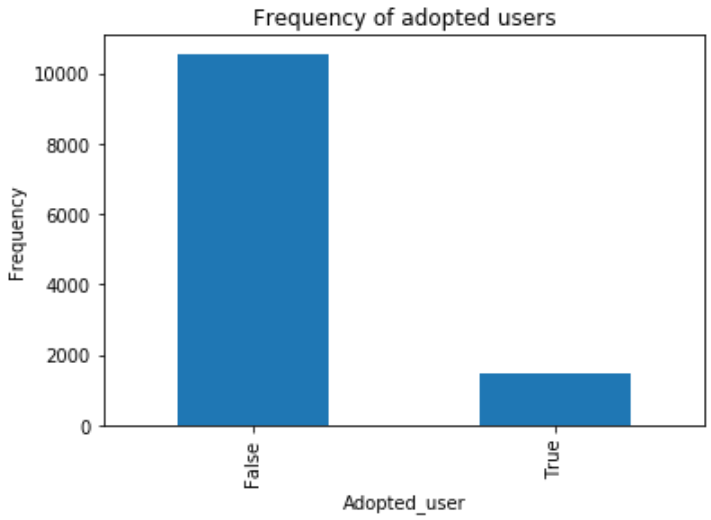

Table-1

Now, the data is ready for further analysis. We can use the categorical column ‘adopted_user’ as a dependent variable and other variables from the table(takehome_users.csv) as independent variables and make a predictive model. Classification methods such as logistic regression and random forest can be used to make prediction as well as identifying important predictor variables.