Relax, Inc Take-Home Challenge

# By Michael Ward

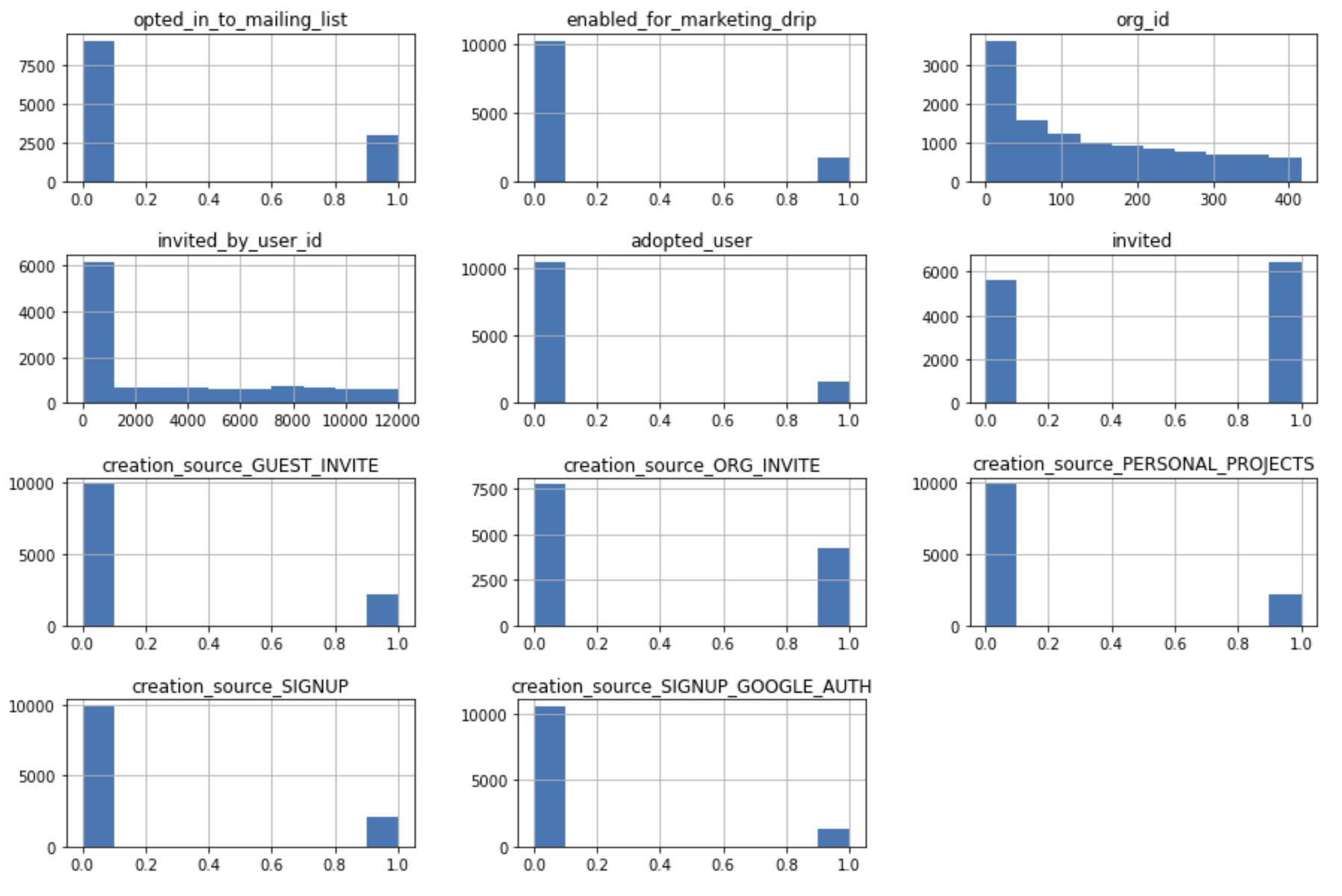
In order to help Relax Inc to retain users, I decided to create a machine learning model that could determine the most important features related to user adoption.

My first step was to use a rolling window to calculate the users that logged into the system for at least 3 days during a rolling 7 day period as this was the criteria for defining an adopted user. I then updated the users data to flag each one as an adopted user or not.

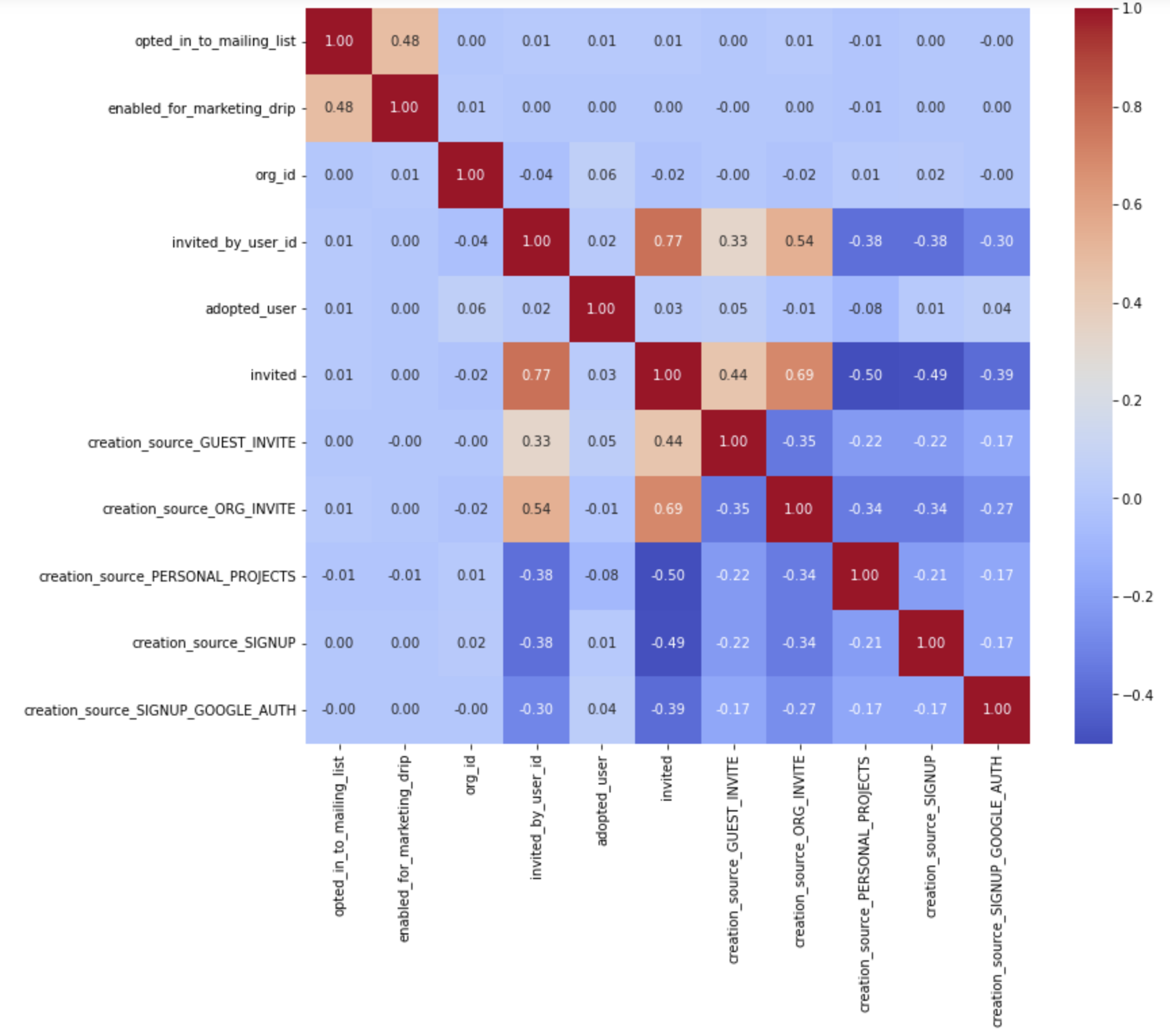
The creation source seemed like it could be an important feature for determining adopted users. I used one-hot encoding to create Boolean features for each creation source.

I also created a Boolean feature to indicate whether the user was invited to join or not based on whether the invited\_by\_user\_id field was populated.

I then dropped the fields that were unique to the user and were not likely to influence whether the user became an adopted user (user id, creation time, name', email, and last session creation time). I created the following histograms to visualize the distribution of each of the fields to be modeled.



I also created the following heatmap to visualize the Pearson Correlation Coefficients. Our target feature is the adopted\_user which has relatively low correlation coefficients to the other features.



I create an 80/20 split of the data for training and testing a machine learning model. I decided to use an XGBoost Classifier in order to model the data. I used RandomSearchCV in order to tune the hyperparameters for the model with the following values.

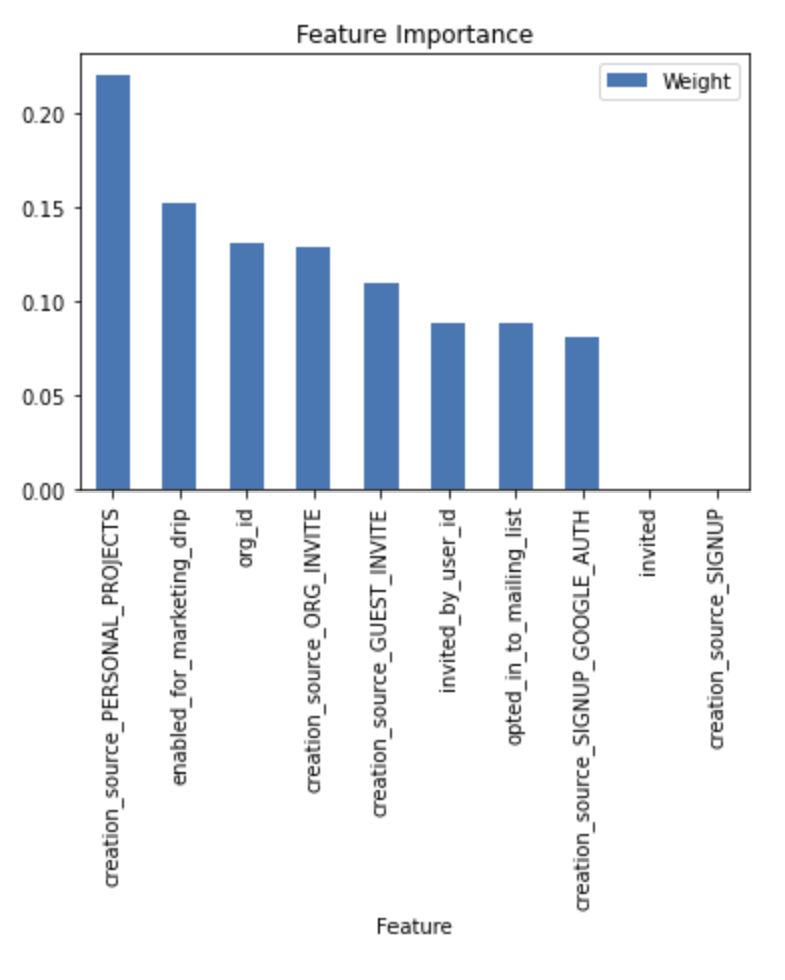
Model: XGB

Tuned Hyper Parameters: {'n\_estimators': 50, 'min\_child\_weight': 4, 'max\_depth': 3, 'learning\_rate': 0.15}

Best score is 0.8670833333333334

I then used the model to predict the adopted users based on the test data and it resulted in a 86.42% accuracy.

The following features were identified as the most important predictors of adopted users by the model.



The most important predictors were as follows:

1. The user was invited to join another user’s personal workspace in order to contribute to the other user’s personal project.
2. The user’s profile is enabled for the marketing drip emails.
3. The user is part of an organization and was invited by the organization to join.
4. The user is NOT part of an organization, but was invited by the organization to join as a guest.
5. The user opted in to the mailing list
6. The user signed up using Google Authentication.