Example Take-Home Challenge: Relax Inc.

I first changed the “object_id” column in the user dataset to “user_id” to match the user engagement dataset because I knew I would eventually want to merge the two together on the “user_id” column. I grouped the user engagement dataset by the sum of logins per week, then subset the dataset to just adopted users by essentially removing users that did not log in 3 or more times in a week. I then added a boolean column to the adopted user dataset equal to True in order to distinguish adopted users from the rest. I merged the adopted user dataset with the total user dataset and filled the NaN values in the “adopted” with False because there would only be a NaN value in the “adopted” column after the merge if the user was not adopted. In total, there were 1,445 adopted users out of 12,000, or about 12%.

I ran a GridSearchCV to determine that the best parameters were 100, while the average precision was about 93%. I used a Random Forest Classifier for feature importance, and found the most important feature by far for predicting future user adoption was the time of the last session, while the organization a user belonged to also played a small role.