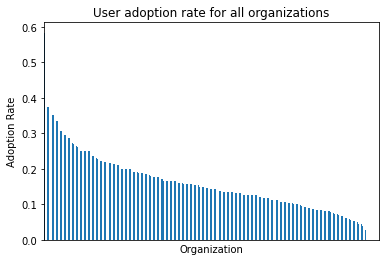
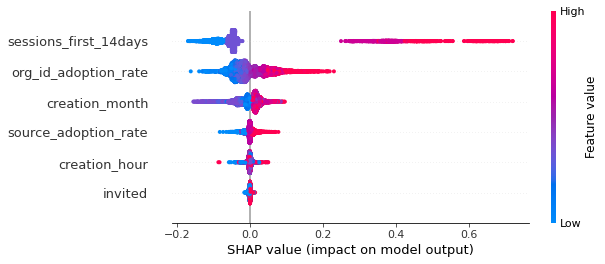
Relax Inc

## Subject: User Adoption Prediction

## From: Blaine Murphy

To determine which factors impact users adoption for Relax, I first imported the data and created the target feature from the usage summary table. Users are defined as adopted if they logged in for a session on three occasions during a 7-day period. I created a target column by grouping the usage data for each user and checking to see if there was at least one session with a previous and a following session all within a 7-day timespan.

Next, I began profiling all of the columns in the user’s data, creating new features when appropriate and testing each features association with user adoption. As an example, I calculated the adoption rate for all unique organizations, and added those values as a column to the data. The plot to the right shows the adoption rate varies significantly among different organizations. A chi-squared test with the target returned a chi2 value of 541 and a p-value of 0, indicating that this feature has a large and statistically significant impact on user adoption. After inspecting all of the user columns I created a final column using the usage data that contained the number of sessions that user had in the first 14 days since the account was created.

After analyzing all of the user data and building a Random Forest classifier, I utilized the *shap* library to visualize feature importance, as seen below. The top 3 most influential factors for predicting user adoption are the number of sessions in their first 14 days, the user’s organization, and the month the account was created in descending order of importance. The hour of the day the user created the account and the whether the user was invited to join by another user both had statistically significant association to user adoption but less impact on the prediction. The Random Forest classifier had an ROC AUC score of 0.88 on a blind test set, demonstrating the features predictive power for user adoption.