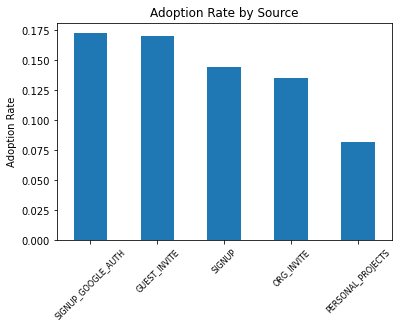
Relax Inc

## 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 the 5 different ways a user created an account. The plot to the right shows the adoption rate varies among the different creation sources. A chi-squared test with the target returned a p-value of 0.0, indicating that this feature has a statistically significant impact on user adoption. I one-hot encoded this feature to create modelling inputs. Finally, I created 3 more columns that calculated the number of user sessions in their first 7, 14, and 30 days.

After analyzing all of the user data I calculated the information value for each feature and dropped those with a value of less than .01. Next, I further reduced the data by removing collinear features. To verify which features are most important and to confirm their predictive power I created a Random Forest classifier to predict user adoption. I utilized the *shap* library to visualize feature importance, as seen below. By far the most influential feature for predicting user adoption are the number of sessions in their first 30 days. The account creation month and the source of the users account are also useful for prediction. The Random Forest classifier had an ROC AUC score of 0.92 on a blind test set demonstrating the features predictive power for user adoption.

