TL;DR: Number of days active is by far the most predictive variable of adopted users and can predict whether a user will adopt with ~98% accuracy.

Data Wrangling

To identify adopted users, I selected users who had >=3 logins within any week. I then merged this classification with the user data to identify adopted and not adopted users and prepared the data for modeling by converting categorical data into boolean columns (1 or 0; columns include creation source and organization ID). I added a column for invited v. not invited (1 or 0). I further calculated active days using last login date and creation date.

Modeling

I used a Logistic Regression to model how all features predict user adoption. I tuned the C parameter and used cross-validation to validate the accuracy metrics for this model. This model was able to predict user adoption with ~97.7% accuracy, about a 10.5% increase from randomly assigning all users to not adopting (the data are imbalanced; most users do not adopt). I validated this accuracy using the AUC, which was rather high for this model at 93%, indicating the model has a 93% chance of distinguishing between the two classes.

Feature Importance

To identify factors that could make a user more or less likely to adopt, I looked at the Logistic Regression coefficients (log odds). I converted these into odds and probabilities - e.g., all else being equal, belonging to organization 118 has 1.19:1 odds (54% chance) of *not* adopting - most of these features did not have easy to interpret coefficients (close to 50% chance). I did cycle through a few features in smaller groups to evaluate the model accuracy and AUC; of note, **active days alone** was able to replicate the model with 97.8% accuracy and an AUC of 93%, indicating this is by far the most useful feature for determining adopted users. This is intuitive in that users who are active for longer are more likely to be more active each week but could be useful nonetheless.