Task: Identify factors that predict future user adoption of the software. An "adopted user" is identified as a user who has logged into the product on three separate days in at least one seven-day period.

Data: user information csv file and user login time stamps csv file

After computing the number of logins per user over 7 day rolling periods, the users were classified as 'adopted' or not based on the 3 logins in separate days criteria. Subsequently, the following factors possibly impacting user adoption were explored:

1) Invited by another users or not

2) Creation source (guest invite, organization invite, etc.)

3) Whether user opted in to mailing list or enabled for marketing drip (showing stronger interests)

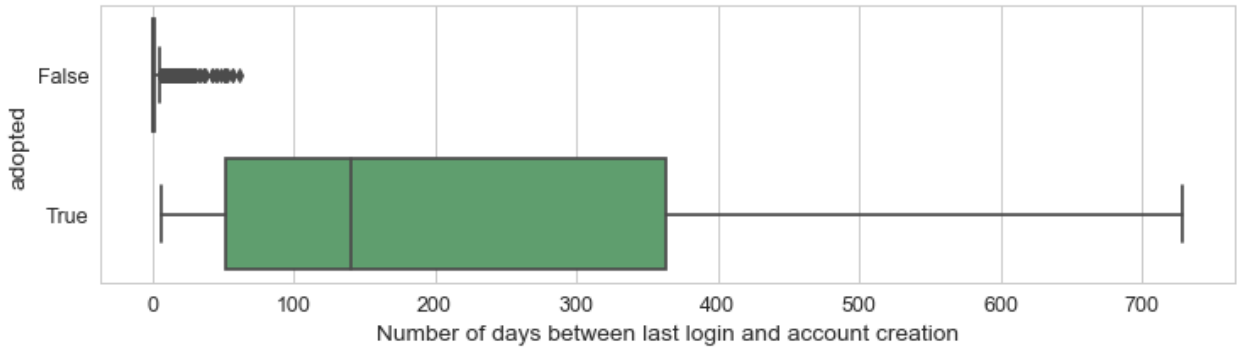
4) Number of users in the organization the user works (computed summing number of users per org_id)

5) Email domain type (some email domains in the database were disposable email addresses (gustr, cuvox and jourrapide) and other domains were typical domains (Hotmail, gmail, yahoo))

6) Total time active (computed subtracting date when the user created their account from their last known login time)

The features were one hot encoded for categorical variables and normalized. A logistic regression model was generated, from which, coefficients and p-values were extracted and the odds ratio was computed to evaluate feature importance.

The most predictive feature was the total active time (time elapsed between when the user created the account and their last login), allowing a nearly perfect split of the data by using a simple cutoff of 16 days. To a much smaller degree, creation time, type of email, and creation source (organization invite) were also found to have predictive power. Illustrating the predictive power of active time we can see the clear distinction in the box plots for adopted users and non adopted users:



Further research:

1) Creation source can be further explored by doing a one vs all approach to determine which sources are statistically significant different from others.
2) More organization features could be added to this data such as type of industry or size of organization and joined using organization id