

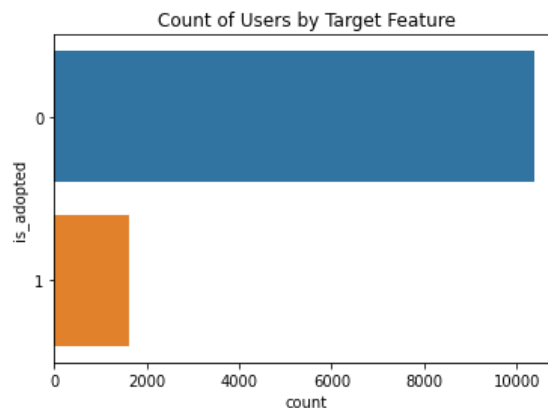
## Relax Challenge

The goal of this challenge was to “identify which factors predict future user adoption” based on the provided datasets of users and logins (“engagement”). I accomplished this through exploratory data analysis and statistical tests after performing some data wrangling and feature generation. For code, please see this notebook.

### Data Wrangling

First, I investigated the null values in the users table. I confirmed that null values for `invited_by_user_id` were only present for users who had not been invited to sign up. Nulls for `last_session_creation_time` are only for users who have no records in the engagement table, meaning that they never logged in. Because neither of these could be explanatory variables I did not impute null values.

The target feature had to be created by identifying users that had more than 3 logins during a 7-day period as recorded in the engagement table. I discovered that based on this definition, 13.35% of users have "adopted" the service

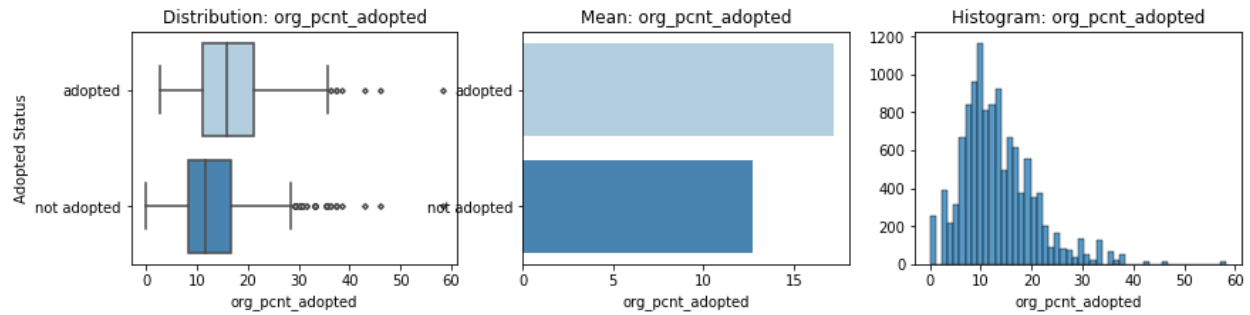


Because there were not very many potential explanatory features in the tables as is, I generated a few additional features for each user:

- Signup day of the year and day of the month
- Number of users in organization user is part of
- Number of active users in organization user is part of
- Proportion of users who are active in org user is part of
- Binary flag if user has invited others

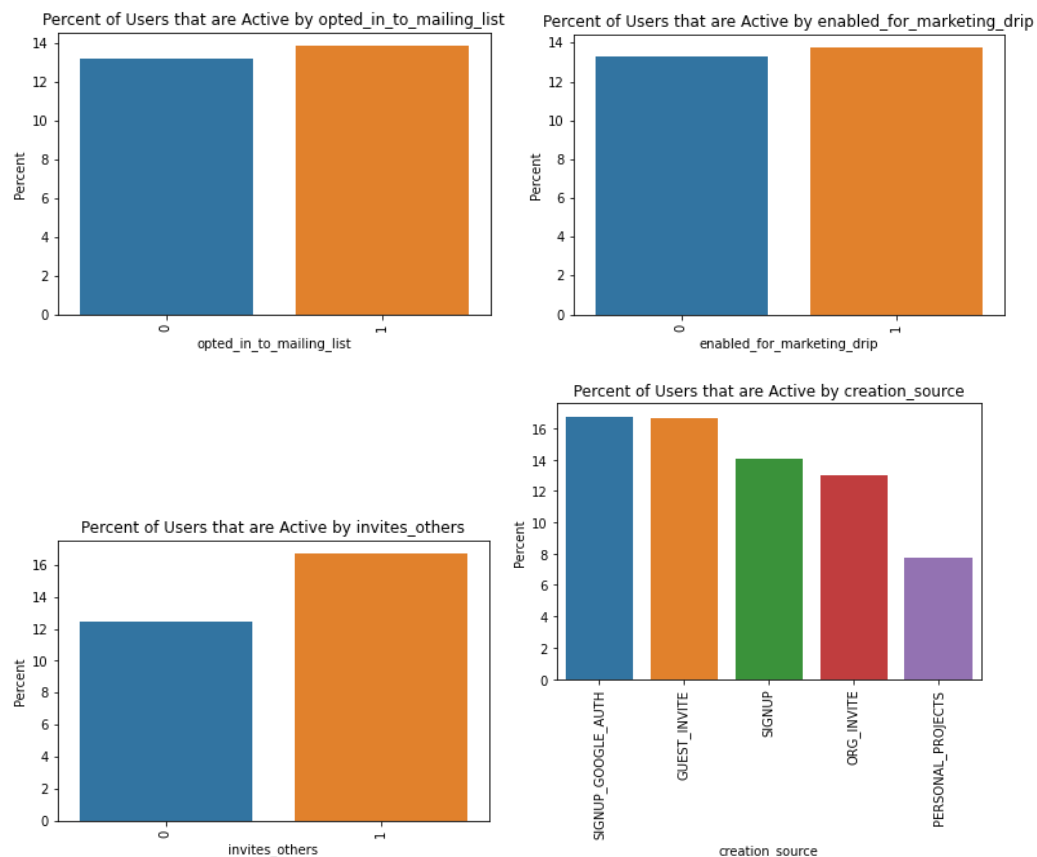
### Exploratory Data Analysis

For continuous features I plotted data visualizations comparing values for that feature to adopted/non adopted users including boxplots, barplot of means, and histograms (see example figure below). I also performed two-tailed T-tests to test for statistical significance (see table below). See conclusions for interpretation.



	T-Score	P-Value
creation_dayofmonth	-2.0602	0.0395
creation_dayofyear	4.1350	0.0000
org_num_users	-12.0972	0.0000
org_num_active_users	0.6549	0.5126
org_pcnt_adopted	20.5042	0.0000

For continuous features, I plotted frequencies of each category vs the target variable and then performed chi2 tests to test statistical significance (see charts and table below).



	opted_in_to_mailing_list	enabled_for_marketing_drip	invites_others
<b>chi2</b>	0.591	0.221	24.766
<b>pvalues</b>	0.442	0.638	0.000

### Conclusions

From the analysis described above I was able to identify which factors predict user adoption, and which do not. Strong predictors with statistically significant relationships included:

- Percent of users who have adopted service in org user is part of (positive,  $p < 0.001$ ). This relationship suggests that users are more likely to adopted the service if they are part of an org where a higher proportion of other members have also adopted the service.,
- Number of users who of have signed up for service in org users is part of (negative,  $p < 0.001$ ). Although a bit counter intuitive, this suggests that users who are part of organizations that have large numbers of users signed up to the service are actually less likely to adopt the service. This is likely because these users are being invited en mass and have little interest in the service.
- If users invite others ( $p < 0.001$ ) they are much more likely to have adopted the service.

These factors do not appear to have a statistically significant relationship with whether a user adopted the service or not: number of active users in org, opted in to mailing list and enable for marketing drip. This suggests that marketing efforts are not accomplishing much and should be revisited.