Take Home Challenge II: Relax Inc.

By Aisling Casey - July 16th 2021

See accompanying notebook in Github folder: Take Home Challenge II Notebook.ipynb

Question

Which factors predict future user adoption?

Adoption Prevalence

Of 8,823 users, 1,602 users adopted the platform while 7,221 did not, yielding a rate of 18.16%. User adoption was defined as, "a user who has logged into the product on three separate days in at least one seven-day period."

Missing Data

Of the original 12,000 users in the data set, only 8,823 of them had user engagement data, which was needed to determine if they had adopted the platform or not. Users without engagement data were not included in further analysis; therefore, *the following analysis assumes that the users without engagement data had no systematic differences in feature distributions compared to those with engagement data.*

Methods

To extract insights from the data, chi-square statistical tests and visualizations were used. Additionally, the following features were engineered:

- A new column, 'invited_other_user, indicates whether a user invited others or not.
- The email domain was extracted from each email address; those emails domains appearing less than 5 times in the data set were marked as 'other'.
- Organizations were stratified into size categories based on how many users belonged to them.
- Users were grouped by the user that recommended them.

Results

The chi-square test showed that 'invited_other_user', 'creation_source' and 'email 'were statistically significant in distinguishing users that adopted the platform vs those that did not. Of particular interest:

- Those who invited another to the platform adopted it 23% of the time.
- Those who created their accounts due to a guest invite or personal project were 22.6% and 21.4% likely to adopt the platform respectively, while the other sources had likelihoods below the overall prevalence.

As seen in figure 1, the size of a group invited by a single individual proved modestly informative in distinguishing the two groups. Those invited in groups of 1-4 people had a 19.4% adoption rate, compared to those not invited having just a 17% adoption rate; these two groups encompassed 3,665 and 4047 users, respectively.

As seen in figure 2, organizations with 10-24 users had a higher adoption prevalence of 20.6%, while representing the lion's share of overall users (over 50%). Meanwhile, organizations with over 100 users had a low adoption rate at 10.4%, and those not belonging to an organization had an adoption rate of just 4.8%

Future Work

From here, it would be worth building a machine learning model to better understand the data. Particularly to understand latent relationships between the variables – e.g. while on its own 'opted_into_mailing_list' did not yield any insight in user adoption, perhaps when paired with other features it could. In building a predictive model, it would be critical to know when in the user lifecycle the client is looking to predict user adoption. For example, using the user engagement data, features based on frequency & time between sign-up and first use could be explored; but some of this information would not be available until weeks or months after user sign up.

group_size	adoption_prevalence	group_total
1-4 Invites	0.193724	3665
5-10 Invites	0.182718	1111
Not Invited	0.170250	4047

Figure 1 – Adoption prevalence and number of members for various invitee group sizes

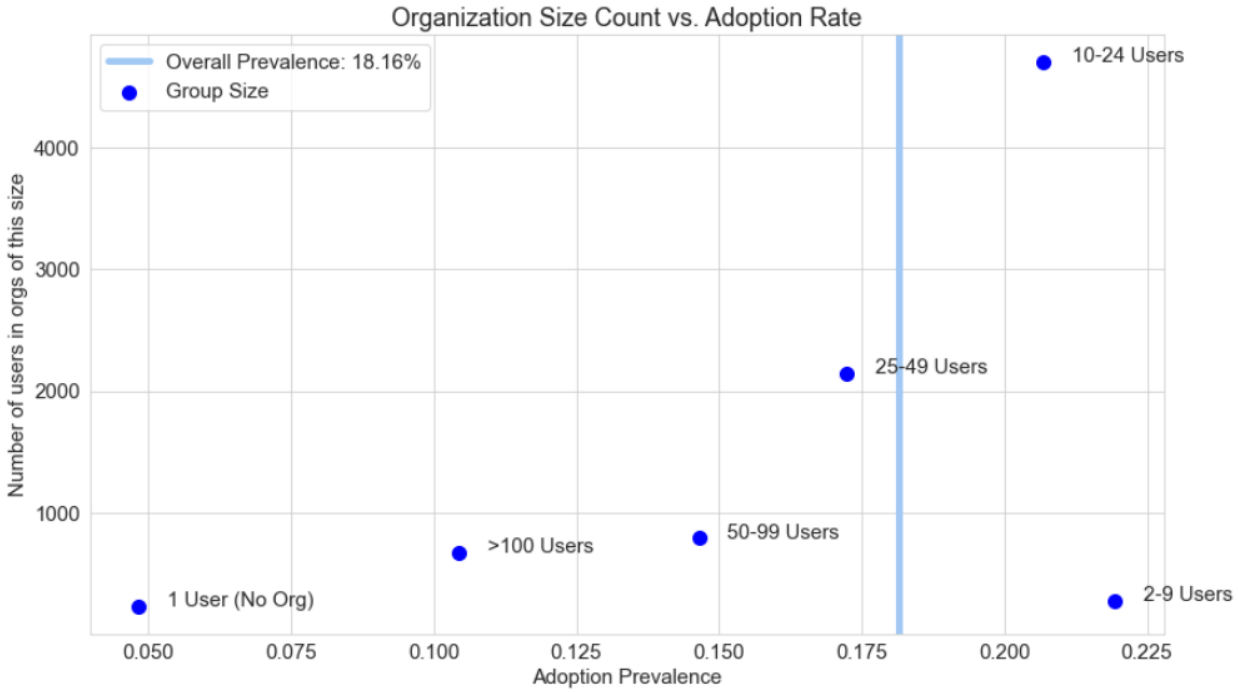

Figure 2 – In this graph, the adoption rate is plotted against number of people in an organization of a certain size.