

## PREDICTING USER ADOPTION

*Prepared by Alexander Olden for Relax Inc.*

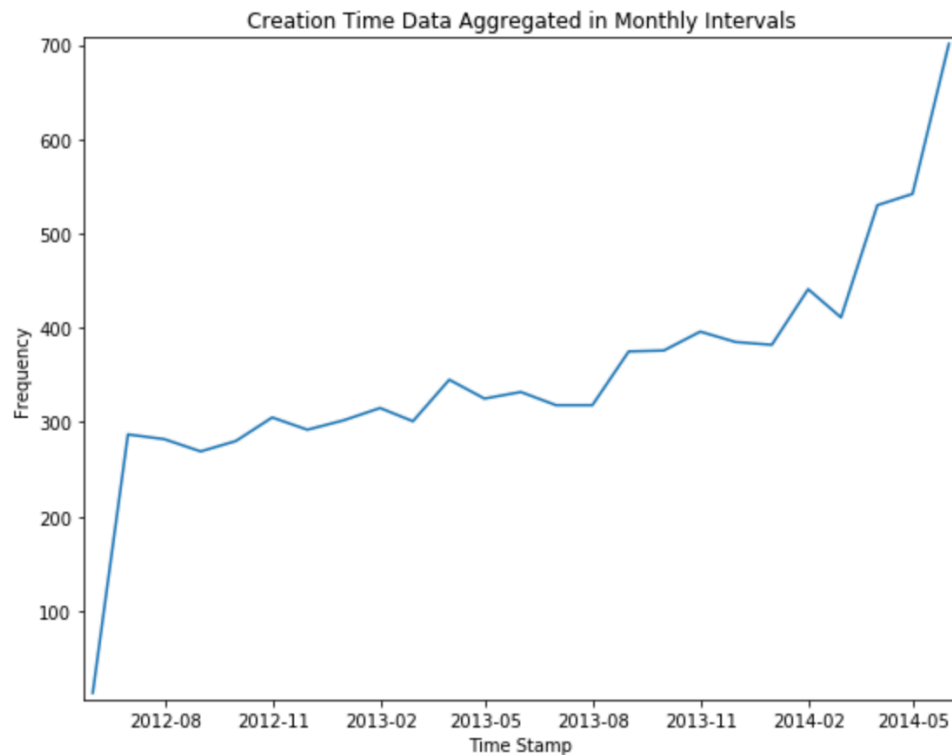
In my analysis of user adoption (any user who has logged in at least 3 times within a seven-day time period) at Relax Inc., I found that the months during which a user is active and organization IDs are important predictors of adoption. In particular, users who logged in most recently during the spring of 2014 are more likely to stay involved. This may seem a bit obvious, since those dates are more recent in the context of the dataset, but it also tells us that older users are not necessarily staying active and that the company needs to take steps to maintain relationships with users over the long term. There were also a few organization IDs (82, 119, and 148) that proved important. Without knowing more about the organizations per se, I cannot offer actionable advice beyond looking for what distinguishes those three organizations from other and seeking to replicate those traits. The ten most important features and their coefficients appear in a table on the next page.

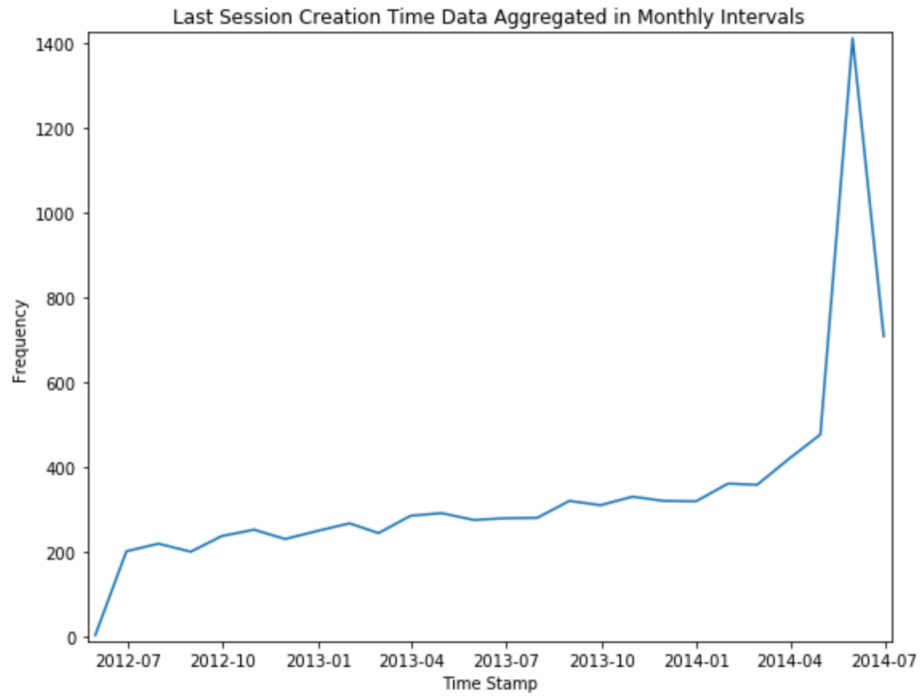
It is import to note that this model would likely change markedly with a few enhancements to the dataset. After converting categorical features to binaries, I had a very high percentage predictors that were either dates or organization IDs. Ideally, I would have more information about IDs so that I could bin them. With about 8,000 observations in the dataset, reducing hundreds of organization IDs to a few grouped ones would make a big difference. Removing group IDs with relatively few observations could help as well, but I would want to know more about the groups before removing representation for certain ones. Depending on the cycles of the Relax business, grouping dates by season instead of month could help reduce dimensionality, too.

### Coefficient-feature Pairings from Logistic Regression

coefficient	feature
14.542737	last_session_creation_time_2014-06
12.969305	last_session_creation_time_2014-05
10.321495	creation_time_2012-06
9.147000	last_session_creation_time_2014-04
9.097680	org_id_82
8.651756	org_id_119
8.217541	org_id_248
7.472719	creation_time_2012-10
7.261015	creation_time_2012-08
6.866192	last_session_creation_time_2014-02

The next two graphs capture the frequencies of the two date variables, binned by month.





Finally, as the bar graph on the next page shows, there is a relatively small number of groups that represent a large portion of the users. Some form of binning or dimensionality reduction would help, but I would like to have more business information before taking those steps. (I realize the group IDs are not legible here, but they can be viewed in the interactive graph in my Python notebook.)

Bar Chart for Organization IDs

