

Relax Challenge

Tamara Monge

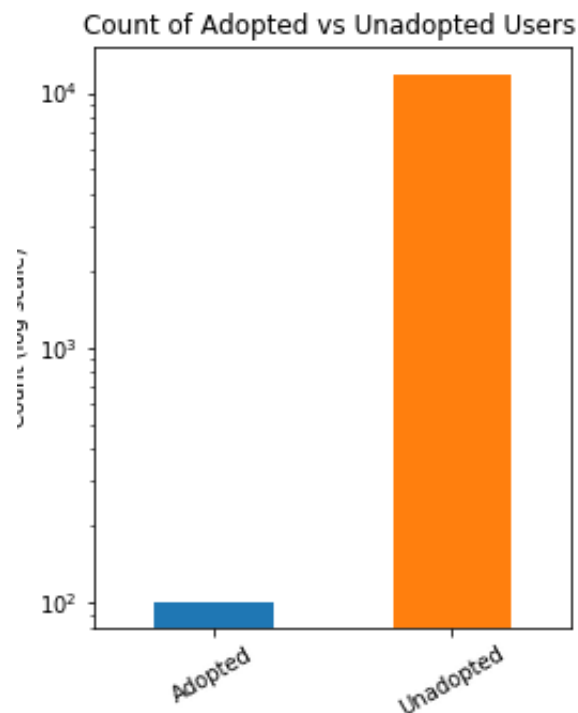
Objective: The objective of this challenge is to identify which factors predict future user adoption.

Client: The client is the productivity and project management software company, Relax.

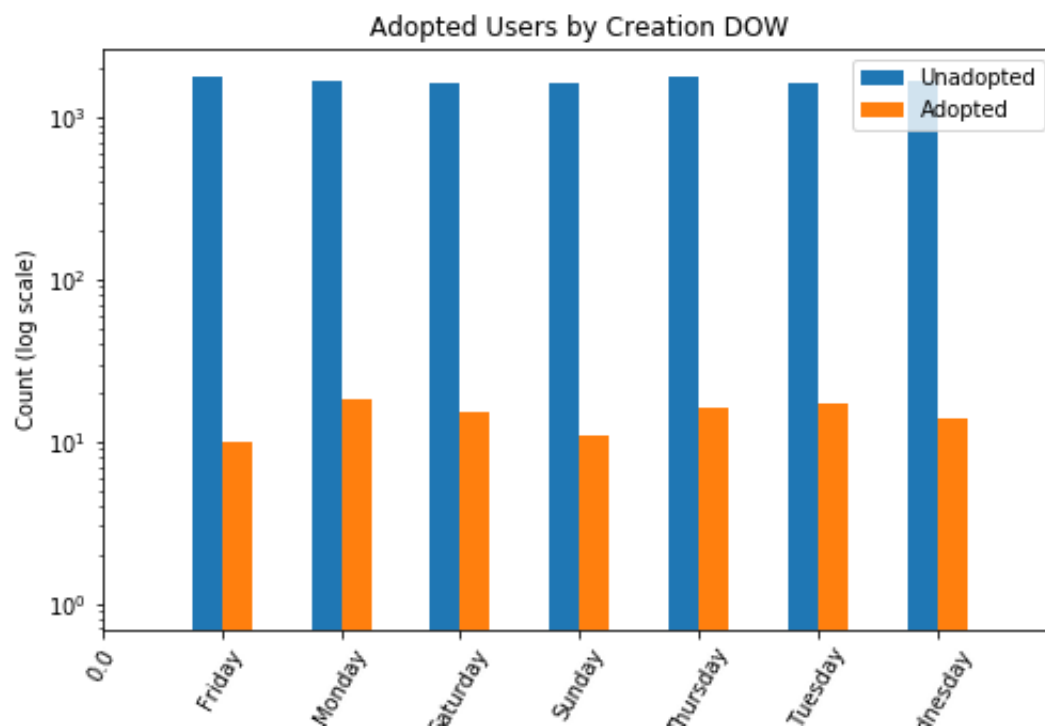
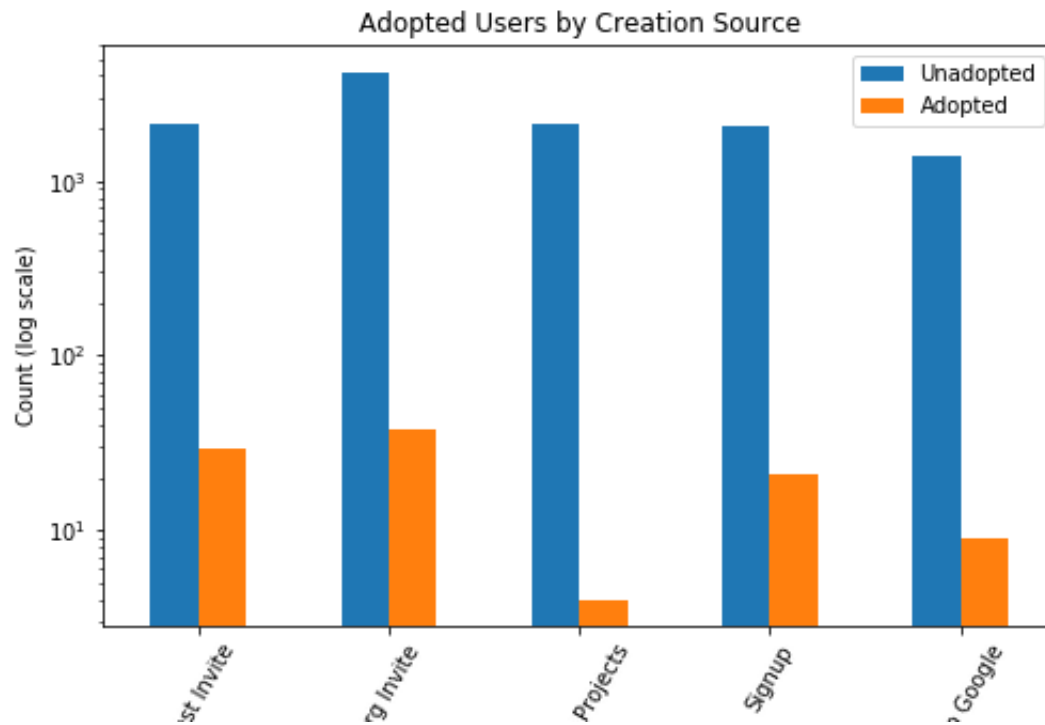
Data Wrangling steps: The data came as CSV files which allowed them to be imported as pandas dataframes. The two dataframes were joined. Missing values were set to 0. Datetime variables were set to datetime objects. The index of the dataframe was set as the user id.

Feature Engineering: A feature `adopted` was created to indicate whether or not a user had adopted. Another feature `creation_dow` was created to indicate the day of the week the account was created.

Exploratory Data Analysis: Approximately 0.84% of users adopted (figure below).



Visual inspection of the explanatory variables revealed that `creation_source` and `creation_dow` may be a meaningful predictors of whether a user adopts (figures below).



Supervised Classification: I treated this as a supervised classification problem with `adopted` as the target variable and the other features as the predictive variables.

I decided to go with a tree-based classifier for two reasons: 1) they do well with imbalanced classes, such as our dataset, and 2) they do well with many categorical features, such as our dataset. I selected the random forest because it is less prone to overfitting than the decision tree and feature importances can still be extracted.

I split the data for training and testing using a 70%-30% random split.

The model was fit on the training set. It was then tested on the test set to evaluate its performance on unseen data. On unseen data, the random forest model predicted which users will adopt with 99% accuracy.

Results: The 10 factors that best predict future user adoption are shown below:

Feature	Importance
creation_source_ORG_INVITE	0.037
opted_in_to_mailing_list	0.036
creation_dow_WEDNESDAY	0.034
creation_source_SIGNUP	0.032
creation_dow_MONDAY	0.032
enabled_for_marketing_drip	0.28
creation_dow_TUESDAY	0.26
creation_dow_THURSDAY	0.23
creation_source_SIGNUP_GOOGLE_AUTH	0.22
creation_dow_FRIDAY	0.21

Client recommendation:

- Users most likely to adopt are those who: create their account via `ORG_INVITE` or `SIGNUP`, create their account on a Monday, Tuesday, Wednesday, or Thursday, enable marketing drip, and/or opt in to the mailing list.
- **To increase adoption, focus more marketing attention on users who: create their account via `SIGNUP_GOOGLE_AUTH`, create their account on a Friday, disable marketing drip, and/or opt out of the mailing list.**

Copyright 2018 Tamara Monge