

Relax Data Science Challenge

https://github.com/bjnugent/Springboard/blob/main/relax_challenge/relax_challenge_code.ipynb

Through grouping by user_id and weeks, I determined that 16.38% of the users were adopted for all users who have logged in. However, 26.47% of all users have not logged in at all, which should be investigated as they have not even tried the product. I explored the creation source, opting for the mailing list, enabled for marketing drip, and whether the person's organization was in the top 10 organizations with the most users.

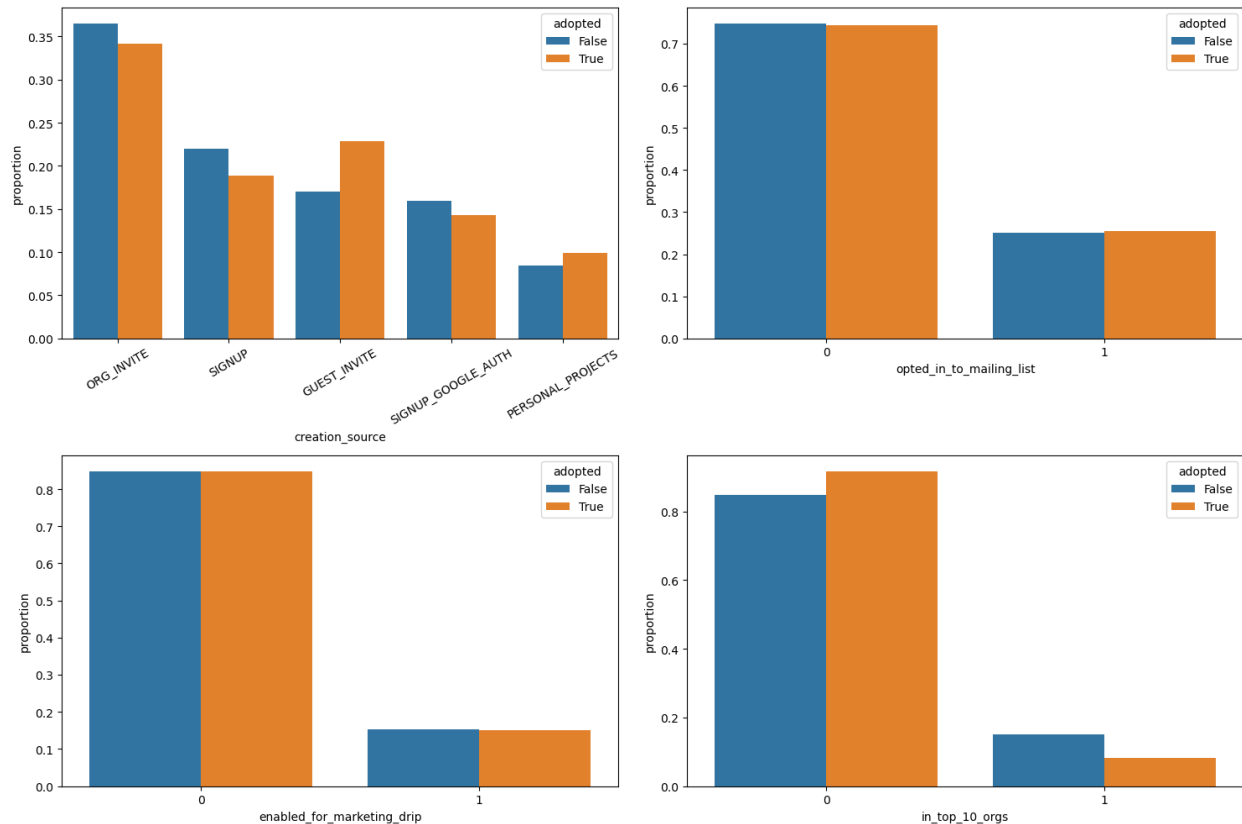


Figure 1. Features of adopted and non-adopted users

The mailing list and marketing drip both yielded almost identical percentages for adopted and non-adopted users. Users who signed up using a guest invite were more likely to become adopted users while all other categories, including organization invites, are more associated with non-adopted users. Additionally, users in the top 10 organizations were less likely to become adopted users.

I continued my analysis through a Chi-squared test to quantify the correlation in these categories.

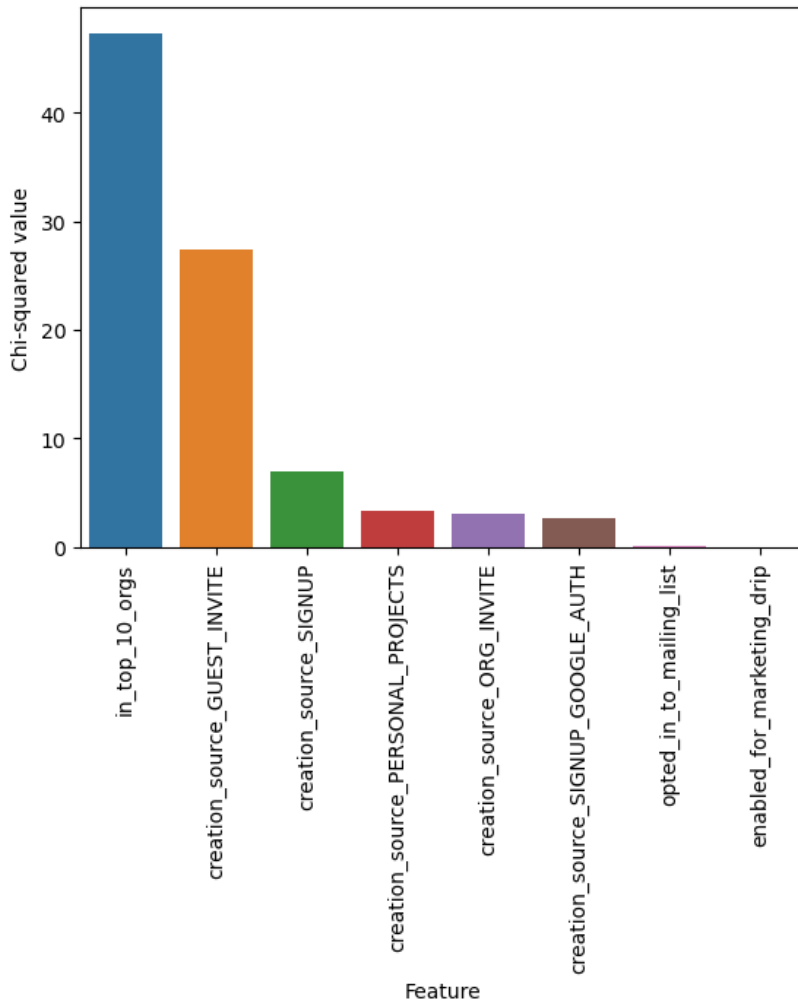


Figure 2. Chi-squared analysis of features

The Chi-squared analysis supported my initial findings, with being in the top 10 organizations and being invited by a guest being the only two statistically significant correlations with user adoption ($p = 1.3 \times 10^{-9}$ and $p = 1.6 \times 10^{-5}$ respectively).

I also built decision tree, random forest, and logistic regression models to view the feature weights and importance that are optimal for determining user adoption. While all model's feature importances revealed the same two top predicting features, all models failed to predict adopted users due to the weak correlation between these features and adoption, revealing a need for more data on adopted users and new features.