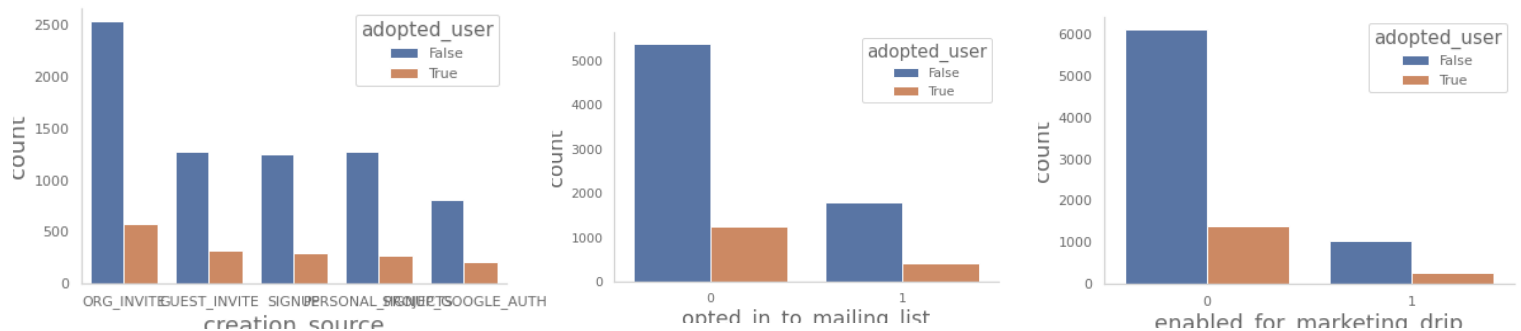


Relax Inc. Report

This is the report for the take-home exercises on Relax Inc.

The goal of the exercise is to identify which factors effecting the adoption rate which is defined as logging in 3 days in a 7 days period. After some cleaning and simple EDA, we can see that many creation sources, opted in mailing and marketing drip have very little effect on the adoption rate. As seen below the percentage of True is almost the same for every category.



Thus, feature engineer is needed. The time delta between the last session and the creation date seems to be a good sign of an active and adopted user.

Finally, I run the random forest classifier with standard hyperparameters (this can be improved by GridsearchCV) and found the most important features as followed:

- | | |
|-------------------------------|----------|
| 1. opted_in_to_mailing_list | 0.008866 |
| 2. enabled_for_marketing_drip | 0.007462 |
| 3. org_id | 0.725748 |
| 4. time_create_last_session | 0.237597 |
| 5. source__ORG_INVITE | 0.006394 |
| 6. source__PERSONAL_PROJECTS | 0.004450 |
| 7. source__SIGNUP | 0.004886 |
| 8. source__SIGNUP_GOOGLE_AUTH | 0.004597 |

Org_id or the organization that a user belongs to has a big effect on whether the user will become an adopted user or not. This makes sense since being a part of an organization who adopt the feature will incentivizes users to do the same.

Secondly, if user is still active (time between creation date and last session is large), they are more than likely an adopted user. But this might be reversal causation.

Many improvements can be made to this short report such as using CV with 3 partitions, engineer more features using current time...

