The first step I took in analyzing this data was creating the target variable. This involved resampling the time-series login data and calculating whether each individual user had at least 3 logins within any 7-day period. Next, I turned to cleaning the user data. Since the goal of this analysis is to uncover patterns that predict future adoption, there are a few variables that can be ignored on the basis that they will not be useful for this type of analysis. These variables are object\_id, creation\_time, name, email, and last\_session\_created\_time. The variables that I chose to explore further were creation\_source, opted\_in\_to\_mailing\_list, enabled\_for\_marketing\_drip, org\_id, and invited\_by\_user\_id. After cleaning these variables, some basic investigation and visualizations revealed the following relationships:

* Incidence of adopted users for the PERSONAL\_PROJECTS category of creation\_source variable is significantly lower than the other categories.
* Enabled\_for\_marketing\_drip, opted\_in\_to\_mailing\_list, and referred variables seem to have little to no impact on the likelihood of a user being adopted.
* Org\_id variable seems to have a major impact on the likelihood of a user being adopted. This variable is difficult to model as it is difficult to treat as a categorical variable since it has 417 unique values, but it also can’t be treated as a numerical variable since it doesn’t follow any clear linear or non-linear relationships. The adoption rate for some of the different organizations can be seen in the table on the following page, which should help illustrate the issues with this variable.

With these issues in mind, I then attempted to create two models. I chose two different tree-based models as they typically perform well with little tuning. However, in this case both models performed poorly, and I was not able to achieve results that would be worth reporting to the client. My final recommendations to the client are as follows:

1. Reach out to organizations with high user adoption rate and try to identify elements of their programs and communications that might lead to higher adoption rates if applied to lower adoption rate organizations.
2. Explore revamping the marketing drip and mailing list communications, as they currently seem to have little impact on user adoption.
3. Explore ways to increase user adoption rates among users who are invited to join personal projects.

Graphical user interface, table

Description automatically generated