Relax Inc. Challenge

Problem statement: How can Relax Inc. increase its adopted user base by 30% over the next year.

Methodology for data wrangling:

I explored the users data using ydata-profiling and I discovered that the last_creation_time series was missing more than 1/4 of the data, and invited_by_user_id was missing nearly 1/2 of the data. I dropped the invited_by_user_id series from my data frame and left the last_creation_time untouched.

Feature engineering:

I created a target variable called "adopted" by assigning a 1 to each record that has logged into the product on three separate days in at least one seven-day period, and 0 otherwise using data from the takehome_users csv. After merging with the engagement data, I I compared the adopted user count to each account creation source type and I got the following results:

Creation Source	Adoption Percentage	Total Adopted User Count
GUEST_INVITE	23	2163
ORG_INVITE	18	4254
PERSONAL_PROJECTS	11	2111
SIGNUP	20	2087
SIGNUP_GOOGLE_AUTH	22	1385

The highest adopted user conversion happens when a user is invited to an organization as a guest, and the lowest conversion happens when a user is invited to join another user's personal workspace.

In addition, I thought the count of visits per user could be useful so I created another variable called 'total_visits' using data from the takehome_users csv. After observing average visit count by adoption status, the obvious result indicates that non adopted users have very few sign ins.

Preprocessing and Modeling:

To prepare for modeling, I created dummy variables from the create_source categorical variable, and I also extracted the year, month, day, and hour from the creation_time and last_session_creation_time variables. I used a Histogram-based Gradient Boosting Classification Tree model the results were very good supported by an accuracy score of 99.6%. After using permutation importance to rank the features, my engineered total_visits variable ranked first with 30%.

Recommendation for stakeholders:

To increase the adopted user base, I recommend an incentive for users to sign in more often. The user creation source for personal_projects has the lowest adoption percentage and plenty of room for improvement considering the remaining four hover around 20% adoption.