

Relax Inc Challenge

Nolan Long

After importing the data frames I first used this line of code to find which users were considered adopted .

```
# creates a list of users who have logged in on three separate days in at least one seven-day period
adopted_users_list = []
for user in engagement_df['user_id'].unique():
    user_df = engagement_df[engagement_df['user_id'] == user]
    user_df = user_df.set_index('time_stamp')
    user_df = user_df.resample('D').count()
    user_df = user_df[user_df['visited'] > 0]
    user_df = user_df.rolling(window=7).sum()
    if user_df['visited'].max() >= 3:
        adopted_users_list.append(user)
```

I then took a look at the feature correlations to user adoption. I found that the last session time was highly correlated. Which makes intuitive sense because users who continually use your product are more likely to use it lately.

I then created a random forest classifier to look at which features were important to the model. It had a 97% accuracy score which means we can assess these importances as accurate indicators. This also showed last session time as the most important, but also creation time as a valuable indicator.

adopted_user	1.000000
last_session_creation_time	0.248590
org_id	0.063737
GUEST_INVITE	0.043657
SIGNUP_GOOGLE_AUTH	0.034821
invited_by_user_id	0.021602
opted_in_to_mailing_list	0.006780
SIGNUP	0.006635
enabled_for_marketing_drip	0.005074
ORG_INVITE	-0.003146
PERSONAL_PROJECTS	-0.075949
creation_time	-0.088143

	importance
last_session_creation_time	0.681473
creation_time	0.214105
org_id	0.052932
invited_by_user_id	0.031788
opted_in_to_mailing_list	0.004325
enabled_for_marketing_drip	0.003989
PERSONAL_PROJECTS	0.003604
GUEST_INVITE	0.002188
ORG_INVITE	0.002156
SIGNUP_GOOGLE_AUTH	0.001868
SIGNUP	0.001571
0.9680555555555556	