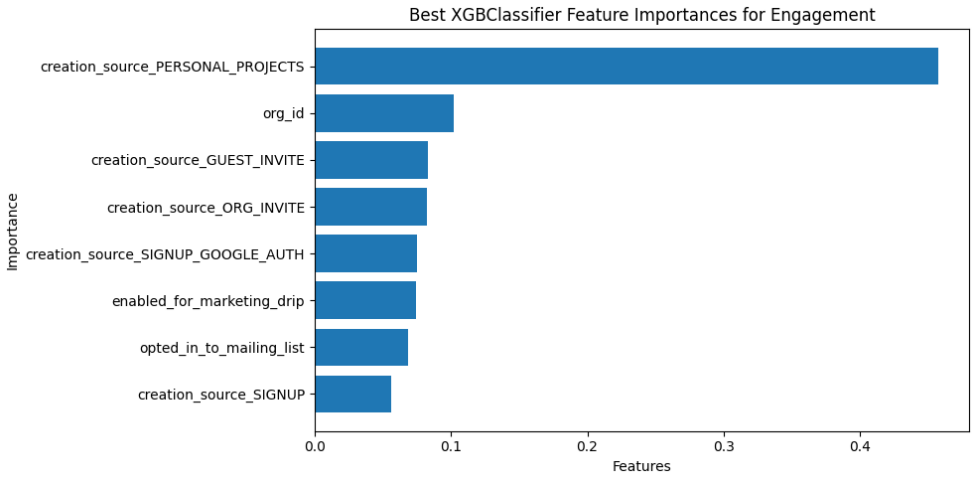
Relax Challenge
Andrew McLaughlin

The first thing I did with the dataset was group the logins by week and then find the user id of everyone that logged in three times in a one week period. I then added a column called engaged to the original data set, with a value of one if the user is considered engaged and a value of zero if the user is considered not engaged. I considered the columns that held the user's id, name, and email unlikely to hold any relation to the user's engagement so I disregarded those columns. Account creation time, timestamp of last login, and user invitations are also unlikely to hold a relation to the engaged status. That leaves behind only 'creation source', 'opted into mailing list', 'enabled marketing drip', and the 'organizational id' column to analyze for relationship to engagement. After making dummy variables for the creation source column I did a correlation between every columns and engaged the results of which can be seen below:

```
CS Personal Projects : -0.07209265102695622
CS Guest Invite : 0.04365728885273945
CS Org Invite : -0.011376030281460423
CS Signup : 0.016003947764581968
CS Signup Google Authentication : 0.03143208592027584
Opted Into Mailing List : 0.010339067805264338
Enabled For Marketing Drip : 0.0059006291320358025
Org ID : 0.060682576648155344
```

From this analysis, it appears that the correlation is weak between all columns but is strongest from users that were invited to join another user's personal workspace. Another method we can try to find column importance is to try making a model. So, I created an XGBoost classifier model on this data set and then found the feature importance in creating that model, which can be seen below:



Once again, personal project invitation seems to be most important. Thus Relax, Inc. could get more engagement by encouraging collaboration between individuals on projects on it's platforms.