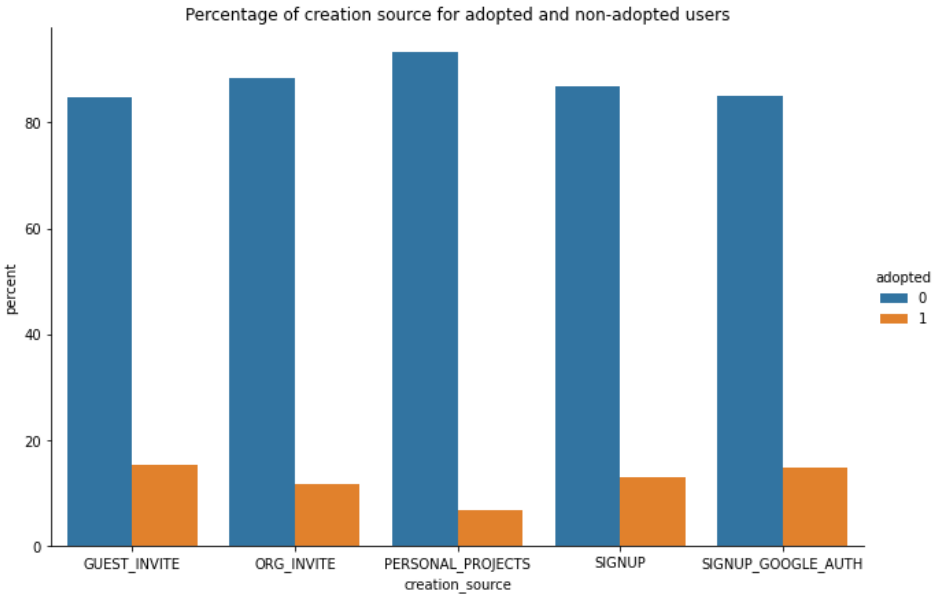
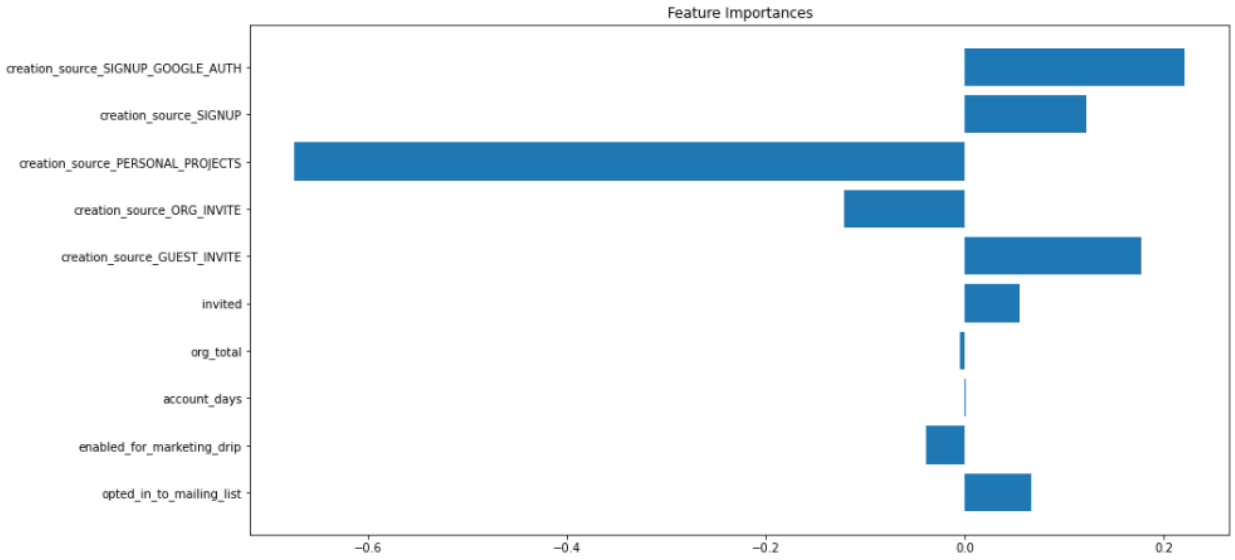
Relax Take Home Challenge Approach and Findings

The provided .csv files were loaded into pandas DataFrames in a jupyter notebook. Adopted users were identified by grouping and manipulating the data in the user_engagement table. 1445 of 12000 total users were identified as adopted. A new feature for adopted users was created, 1 for adopted and 0 for non-adopted. The bar plot below shows percentages of adopted users for each creation source type. Users who signed up via invite to another user's personal workspace "PERSONAL_PROJECTS" had the lowest rate of adoption. Those who were invited to an organization as a guest "SIGNUP_GOOGLE_AUTH" had the highest rate of adoption.



Features were created for days since user account was created, total number of users in an organization, and a binary feature for whether a user was invited by another user or not. A logistic regression classification model was used for prediction of adopted users. Although precision of adopted users was not high, precision of non-adopted users was good. The model may be of limited use for class prediction. Feature importances were determined from the model coefficients. The most important positively influencing features are creation sources SIGNUP_GOOGLE_AUTH, GUEST_INVITE, and SIGNUP. The feature that most negatively influences user adoption is PERSONAL_PROJECTS. It is clear from this analysis that the creation_source of a user account is important to whether the user will become an adopted user. Feature importances are shown in the bar plot below.


Feature Importances
creation_source_SIGNUP_GOOGLE_AUTH
creation_source_SIGNUP
creation_source_PERSONAL_PROJECTS
creation_source_ORG_INVITE
creation_source_GUEST_INVITE
invited
org_total
account_days
enabled_for_marketing_drip
opted_in_to_mailing_list
−0.6
−0.4
−0.2
0.0
0.2