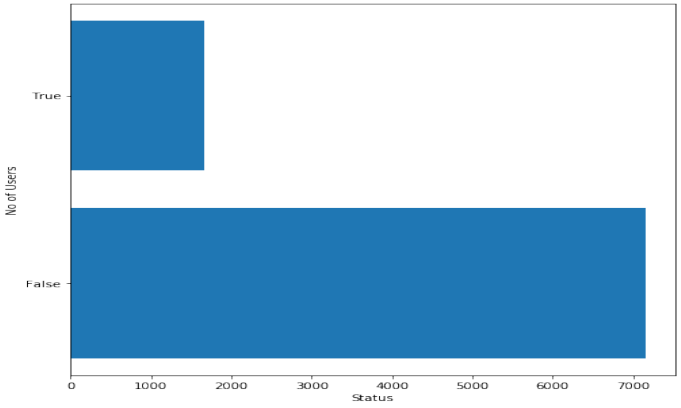
Relax User Engagement Report

The adopted users was created by identifying who has logged into the product on three separate days in at least one seven-day period as suggested in the brief. This information was then merged to the takehome_users dataset. Out of the 12,000 user records in the takehome_users dataset, only 8823 users have records in the summary table that could be classified as either adopted or not adopted as shown below, with about 19% adopted users and 81% not adopted.



Distribution of users

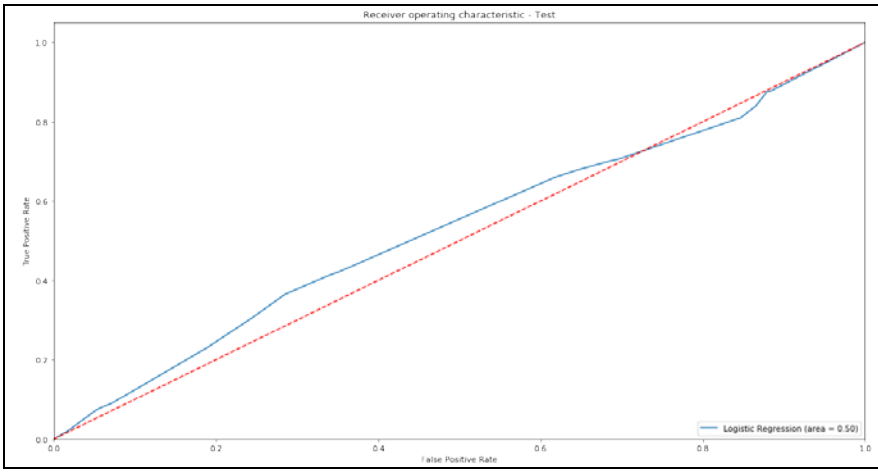
Feature Engineering

After identifying the adopted users, the characteristics of the adopted and non-adopted users were explored to look for features that can be used to determine who will be adopted or not. The first step was to remove variable that will not add value to the analysis. These include the name, email address and organization the belong to. The organization that the user belong to was dropped because we do not have further information for binning or creating a group. But, this variable will be useful if we can determine the different categories represented in the future.

Furthermore, creation time and last session creation time was used to calculate how long the user has been on the system as a new variable called total number of days. The original variables were then excluded. The variable invited_by_user_id variable was also converted into a binary variable to indicate who go invited by someone who is already a user or those not invited at all. Finally, the dummy variables were created from creation source in order for it to be used in a machine learning algorithm.

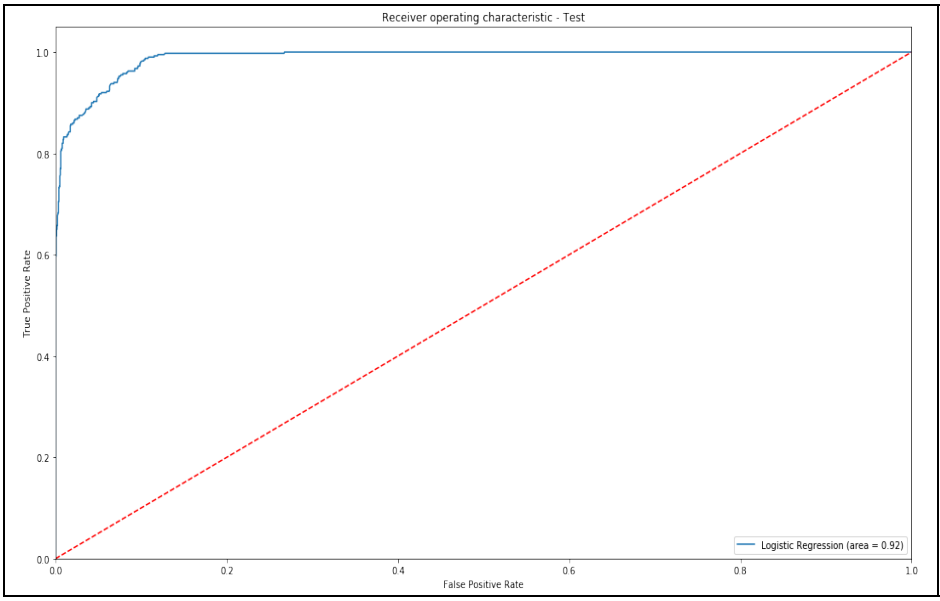
Logistic Regression

Logistic regression was selected because of the need for interpretability of the result. Two experiments were carried out to identify factors that predict user adoption. The first experiments used the variables **creation_source (1-4), opted_in_to_mailing_list, enabled_for_marketing_drip and invitedby (binary).** This produced a very weak result as shown below.



AUC = 0.50
Accuracy = 81%
True Positive = 0
False Negative = 0
True Negative = 1806
Fsl

This result indicated a poor predictor and it was the reason why the **total number of day** on the system was considered. Below is the result of the second experiment.



AUC = 0.92
Accuracy = 96.6%
True Positive = 339
False Negative = 31
True Negative = 1775
False Positive = 61

Based on the above results, the total number of days was the most influential predictor of adoption. This is followed by if the user opted to receive mailing list and the if enabled for marketing drip. The next influential predictor of adoption is if the user was invited to an organization (as a full member) and then those that sign up using google authentication.

Future Work

In the future, it will be interesting to see the influence of how long a user is registered on the system before being categorized as adopted and not just how long the user has stayed on the system altogether.