

1. In “takehome_user_engagement” file, there are only 8823 unique users, and therefore we were only able to determine the 8823 user’s adoption of the product. Among the 8823 users, about 84.8% users did not adopt the product and only about 15.2% users did adopt the product.

There is no missing values in the “takehome_user” file for the 8823 users.

Features such as 'object_id', 'name', 'email', 'org_id', 'invited_by_user_id' were dropped before further data exploration and only features such as 'creation_source', 'opted_in_to_mailing_list', 'enabled_for_marketing_drip', 'creation_time', 'last_login_creation_time' were entered for classification. At EDA step, I examined bi-vari

2. ate correlation between each feature with the target variable ‘adoption’. As can be seen from the barplots presented in EDA section, there is no clear correlation between the target feature with 'enabled_for_marketing_drip' or 'last_login_creation_time'. Distribution of “creation_source” showed very little difference between users who adopted the product and users who did not. Visually examination of the boxplots of Creation_time and last_login_creation_time showed very distinct distributions between two groups.
3. I applied logistic regression for classification, at 80/20 train/test split. The model performs quite well. The following is a summary of performance of the modeling method.

Adpotion	precis ion	recall	f1-score
Yes	0.97	0.99	0.98
No	0.95	0.81	0.88
accuracy		0.96	
macro avg	0.96	0.90	0.93
weighted avg	0.96	0.96	0.96

4. The most important factor in predicting whether the user will adopt the product is how long ago the user last logged into the product, the more recent the user's last login is, the more likely the user will adopt the product. The second most important factor in predicting adoption of the product by user is the elapsed time from the user first created login. The longer history the user has with the product, the more likely the user will adopt the product. Factors such as how the user's account was created, whether the user has opted into receiving marketing emails, and whether the user is on the regular marketing email drip seem to have very little impact on predicting adoption of the product by the user, which is consistent with our observation in EDA step. Feature importance are ranked and plotted in Feature Importance Plot in Section Train Model/Feature Importance.

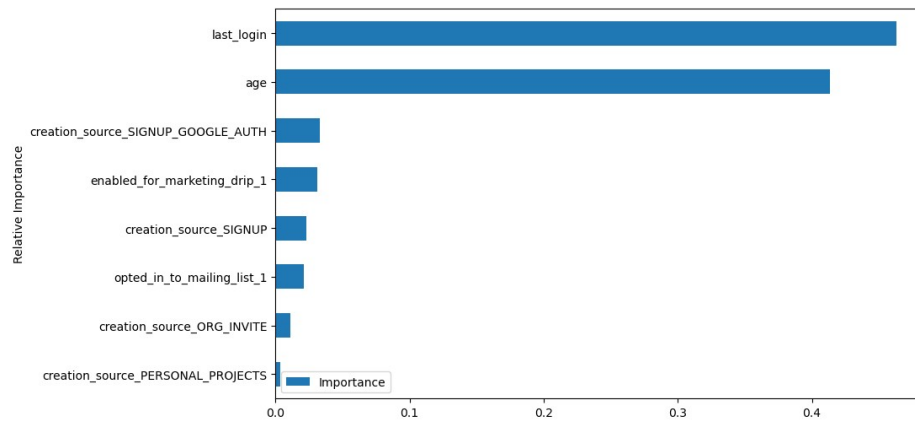


Figure: Feature Importance Plot

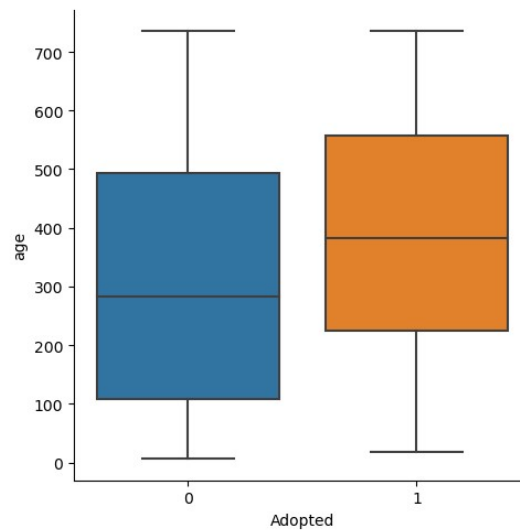


Figure: Creation_Time by Adoption

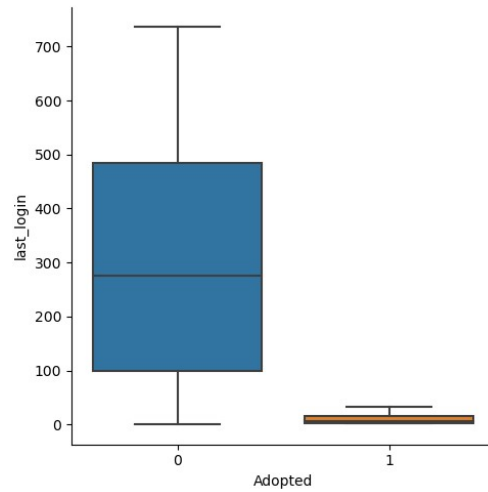


Figure: Last_login_creation_Time by Adoption

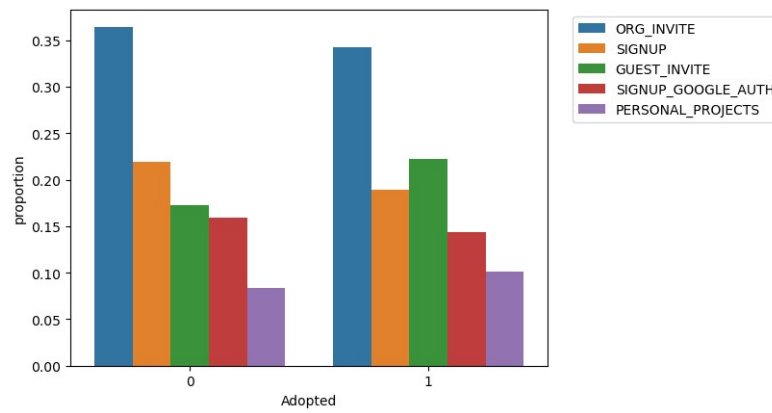


Figure: Creation_Source by Adoption

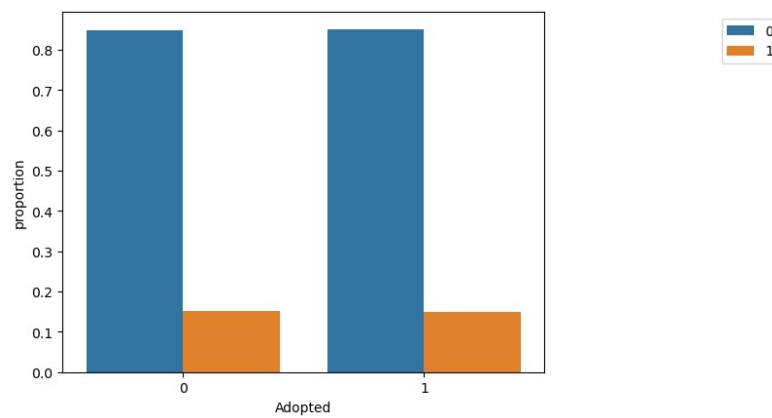


Figure: enabled_for_marketing_drip by Adoption

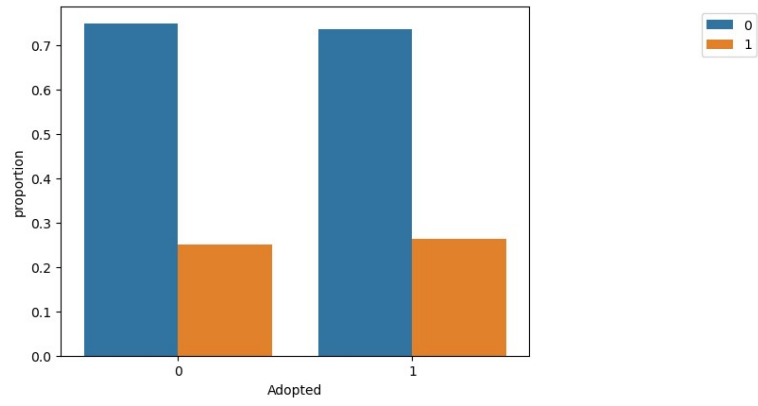


Figure: opted_in_to_mailing_list by Adoption