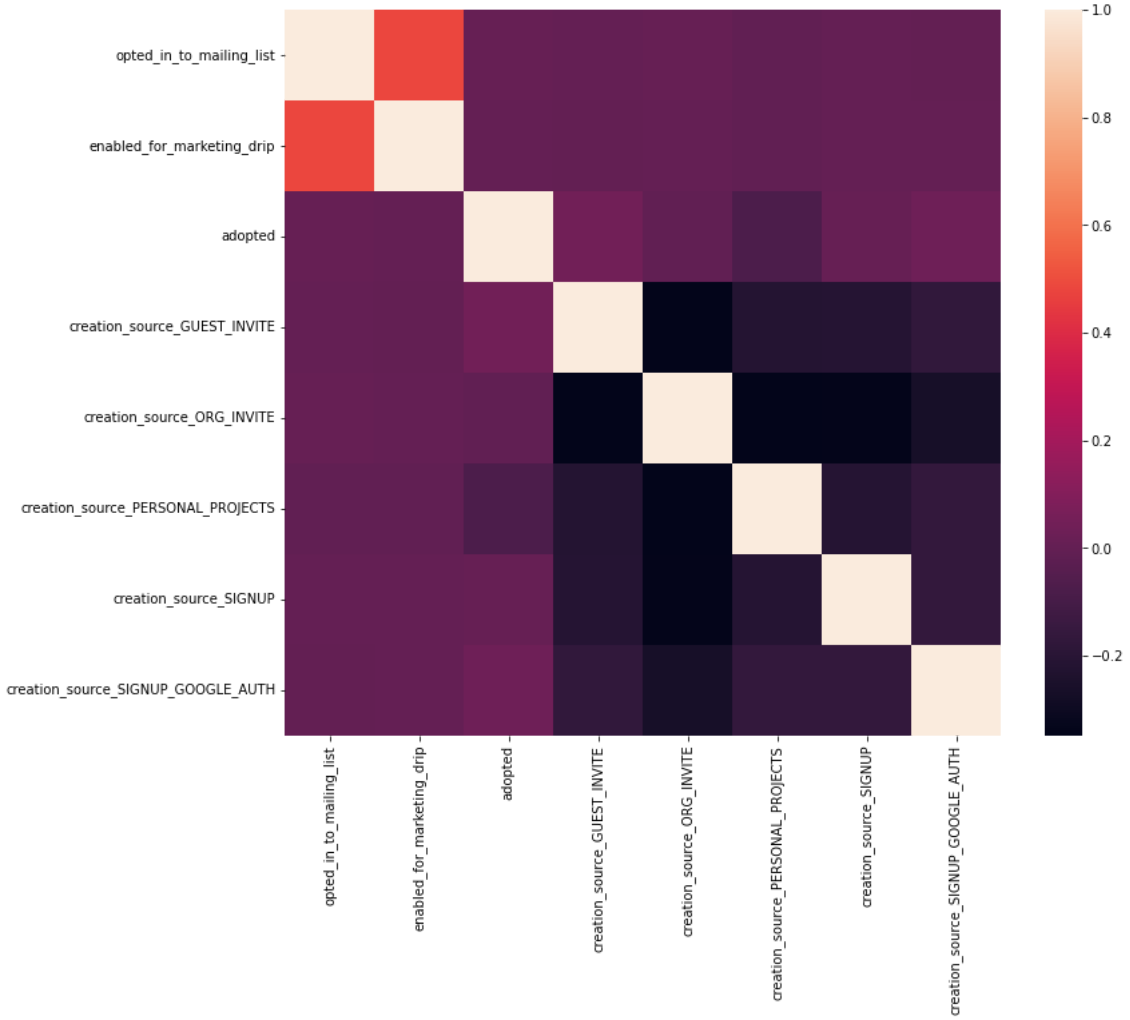
Initial Analysis and Cleaning

The given data had information for 12,000 users 1,602 of which are considered adopted users based on the criteria provided (at least 3 logins within a 7 day period). The first step was to create the target feature, which I did by using a rolling window of 7 days to count how many times each user logged in within a given 7 day period.

Other than a missing target feature, the data was mostly complete. Only two columns had any missing values, both of which could be explained and filled in very easily. The first column with missing values was one for the most recent time a user had logged in, the null values in this column were for users who had never logged in after creating their account, so filling these values with account creation date made the most sense. The second column with missing values was one indicating which existing user invited the new user to join. The null values in this column were for users who were not invited to join by existing users.



Logit Regression Results

Dep. Variable:	adopted	No. Observations:	12000
Model:	Logit	Df Residuals:	11993
Method:	MLE	Df Model:	6
Date:	Fri, 29 Apr 2022	Pseudo R-squ.:	0.01055
Time:	15:13:48	Log-Likelihood:	-4666.1
converged:	True	LL-Null:	-4715.8
Covariance Type:	nonrobust	LLR p-value:	3.170e-19

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.8257	0.065	-28.021	0.000	-1.953	-1.698
opted_in_to_mailing_list	0.0506	0.071	0.717	0.473	-0.088	0.189
enabled_for_marketing_drip	0.0051	0.086	0.060	0.952	-0.163	0.173
creation_source_GUEST_INVITE	0.2010	0.085	2.352	0.019	0.034	0.369
creation_source_ORG_INVITE	-0.0891	0.078	-1.145	0.252	-0.242	0.063
creation_source_PERSONAL_PROJECTS	-0.6617	0.103	-6.432	0.000	-0.863	-0.460
creation_source_SIGNUP_GOOGLE_AUTH	0.2091	0.096	2.186	0.029	0.022	0.397

I performed a logistic regression on the relevant features and examined the coefficients to determine which features have the greatest impact. The results indicated that the most impactful feature is the creaton source for each account. Surprisingly neither of the coefficients for the mailing list or marketing drip are significant, indicating that we cannot say they have a meaningful impact on adoption. Similarly a creation source of Org Invite signup is not significant. The creation source of Personal Projects, Guest Invite, and Google signup are all significant at 95%. Based on these results, we can say the most impactful feature in the given data is the method of signup. This makes a lot of sense, since the method of signup will give some indication of what the account is being created for.