Overall, I decided to use Logistic Regression for ease of interpretability. For preprocessing I encoded creation time as seconds since 2012-01-01. I dropped the name, org_id, and invited by user_id, as their cardinality was too high and were unlikely to be informative. For email, I extracted out the email domain (yahoo, hotmail, etc.) as a categorical variable. Almost all the emails belonged to the top 5 domains, I encoded the rest as "other". Creation Source was dummy encoded with one category removed to remain linearly independent. Every other feature was left more or less as is. Creation time and last session creation time were very large as compared to the rest of the features which were essentially dummy variables, so they were rescaled and centered to the unit Gaussian.

I used L1 regularization in order to attempt to gain a simpler model, the results of the regression was as followed:

Feature	Coefficient
Last_session_creation_time	3.553
opted_in_to_mailing_list	-0.022
enabled_for_marketing_drip	0
creation_time_seconds	-1.951
creation_source_GUEST_INVITE	-0.006
creation_source_ORG_INVITE	-0.173
creation_source_PERSONAL_PROJECTS	-1.094
creation_source_SIGNUP	-0.033
email_group_gmail	0.215
email_group_gustr	-0.104
email_group_hotmail	0.224
email_group_jourrapide	-0.047
email_group_yahoo	-0.145

Here we see how recent they last started a session to be the most significant factor by far. The second most significant factor was how early they started their account. Creation via Personal Projects was the 3rd most significant factor but it had a negative effect on being an adopted user. Overall this can probably be improved if we understand the type of account this is and then see if we can find relevant data based on that.