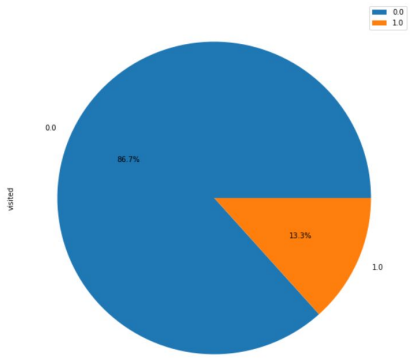
Relax Take Home Report

Jorge Londono

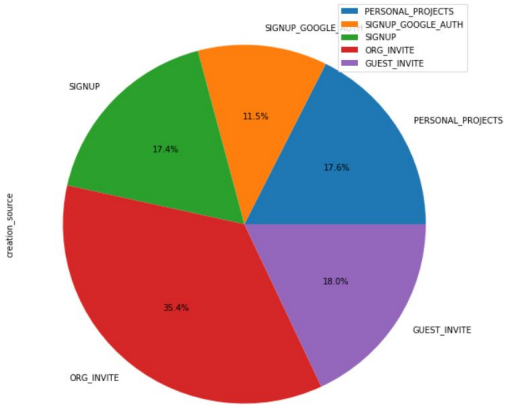
User Adoption

The chart below shows the user adoption rate. ‘1’ corresponds to adoption while ‘0’ corresponds to non-adoption.



Creation Source

The chart below shows the creation source for all users.



Findings

Using a Logistic Regression classification model with 5-fold cross validation I was able to create a model with 0.87 accuracy that would find the most important features of someone becoming an adopted user. The table below shows the highest coefficients of the linear regression model.

	Feature	Coefficient
5	creation_source_SIGNUP_GOOGLE_AUTH	0.114362
381	org_id_62	0.050962
325	org_id_387	0.048282
40	org_id_13	0.046160
138	org_id_218	0.043341
216	org_id_289	0.041092
403	org_id_82	0.040737
171	org_id_248	0.036711
75	org_id_161	0.036048
208	org_id_281	0.035900
92	org_id_177	0.035809
26	org_id_117	0.033388
273	org_id_34	0.032390
284	org_id_35	0.032055
209	org_id_282	0.031155
236	org_id_306	0.030441
199	org_id_273	0.030081
331	org_id_392	0.029698
211	org_id_284	0.029540
196	org_id_270	0.028629
0	opted_in_to_mailing_list	0.028035

Signing up with Google authorization is the most important feature followed by a numerous amount of specific organizations, then followed by users who opted into the mailing list.

Further Research & Comments

From the findings above, we can ask some questions for follow-up research. First, why is Google authorization so important to adopted users? Is the company doing enough to incentivize people to join through collaboration? Additionally, if we know more about the types of organizations that perform the best, there might be a relationship there

worth looking into. Additionally, since the mailing list is very successful in predicting adopted users, it might be worth it to advertise it more to users.

I thought about using a feature that would be the last login time minus the account creation time. This would be the total amount of time an account has existed. The problem with this method is that I would have to get rid of approximately ⅓ of the data, and this is something I view this as too prohibitive.