Summary

To identify factors that predict user adoption, I first conducted hypothesis tests. For numerical features, I conducted two-sample t tests, and for categorical features, I conducted chi-squared tests.

After identifying the factors that predict user adoption, I used them to create a predictive model to evaluate their predictive power. I chose to implement a gradient boosting model with XGBoost because its performance is not negatively impacted by correlated features.

Through this approach, I created a predictive model with log-loss score of 0.07 and accuracy score of 0.97. These scores are very encouraging, and even better performance could be achieved by tuning model hyperparameters.

Hypothesis Testing Results

Before conducting hypothesis tests, I created new features. The new features I created are:

- creation_quarter: Quarter (i.e. 2012Q1) when user created account
- days_from_creation: Days since account creation date
- last_session_quarter: Quarter (i.e. 2012Q1) when user last logged in
- days_from_last_session: Days since last user login
- logged_in: Whether user has logged into his or her account at least once
- invited: Whether user was invited by another user to create an account
-

The table below displays the hypothesis test results. Conclusions were based on a significance level of 0.05.

Factor	Type	P-Value	Conclusion
creation_quarter	Categorical	< 0.001	Significant
days_from_creation	Numerical	0.023	Significant
creation_source	Categorical	< 0.001	Significant
last_session_quarter	Categorical	< 0.001	Significant
days_from_last_session	Numerical	< 0.001	Significant
logged_in	Categorical	< 0.001	Significant
opted_in_to_mailing_list	Categorical	0.394	Insignificant
enabled_for_marketing_drip	Categorical	0.637	Insignificant
org_id	Categorical	< 0.001	Significant
invited_by_user_id	Categorical	0.217	Insignificant
invited	Categorical	0.003	Significant

Neither the mailing list nor the marketing drip were effective in changing user behavior. It was also interested that being invited by another user was predictive and it didn't matter which user sent the invite.