The goal of this exercise was to **identify factors that predict future user adoption**.

The target variable ('is_adopted_user') was derived from the user_engagement table by counting login days over rolling 7 day periods, and identifying users with total counts exceeding 3 at least once. Of the 12,000 users in the dataset, 1602 were flagged as adopted, for an overall adoption rate of 0.13.

Graphs, tables and statistical tests were employed to identify user characteristics that may differentiate between adopted and non-adopted users. Several features provided by the raw dataset were evaluated, including: 'creation_source', 'opted_in_to_mailing_list', and 'enabled_for_marketing_drip'. Newly engineered features that were also evaluated included: 'domain' (email domain), 'invited' (whether user was invited), and 'invited_by_adopted' (whether user was invited by another adopted user). The parameters and evaluation results are summarized in the following table:

Feature	**Type**	**Adoption Rates**	**chi-square p < 0.01**
creation_source	Category (5 distinct)	0.08 - 0.17	Yes
opted_in_to_mailing_list	True/False	True: 0.14 False: 0.12	No
enabled_for_marketing_drip	True/False	True: 0.14 False: 0.12	No
domain	Category (7 distinct)	0.11 - 0.17	Yes
invited	True/False	True: 0.14 False: 0.12	Yes
invited_by_adopted	True/False	True: 0.21 False: 0.13	Yes

Although several features were found to display statistically significant differences between adopted and non-adopted users, the overall effects of these features were weak. The greatest increase in adoption rates (7%) was observed for users who were invited by other adopted users, however the number of users in this class was low (<10%). Slightly higher adoption rates were associated with users who created accounts using guest invitations or using Google authentication, or who signed up with either hotmail or gmail email domains. Lower adoption rates were associated with users who created accounts from personal project invitations, or who used yahoo or cuvox email addresses. The mailing list and marketing drip did not appear to have any relation to adoption.

Another approach considered for this problem was to include organizational factors, such as the number of registered and/or adopted users in a user's company at the time each user account was created. The distributions of these two factors did not appear very different between the adopted and non-adopted user groups, so this approach was dropped.

Given these generally weak results, predictive modeling of user adoption with these features was not pursued. Additional research and data appear necessary to identify factors that relate to user adoption. My recommendations for this additional work include:

- Explore alternative metrics for user adoption that distinguish between users who have short periods of engagement, and those who have more sustained engagement.
- Integrate more information about users and their organizations, such as department types and industries, whether user accounts in an organization were created all at once or sporadically, and pricing associated with individual or organizational accounts.

- Incorporate more metrics around the user sessions, such as session length and usage of particular features.
- Revisit existing marketing strategies, since neither the mailing list nor marketing drip appears to have any impact on engagement.