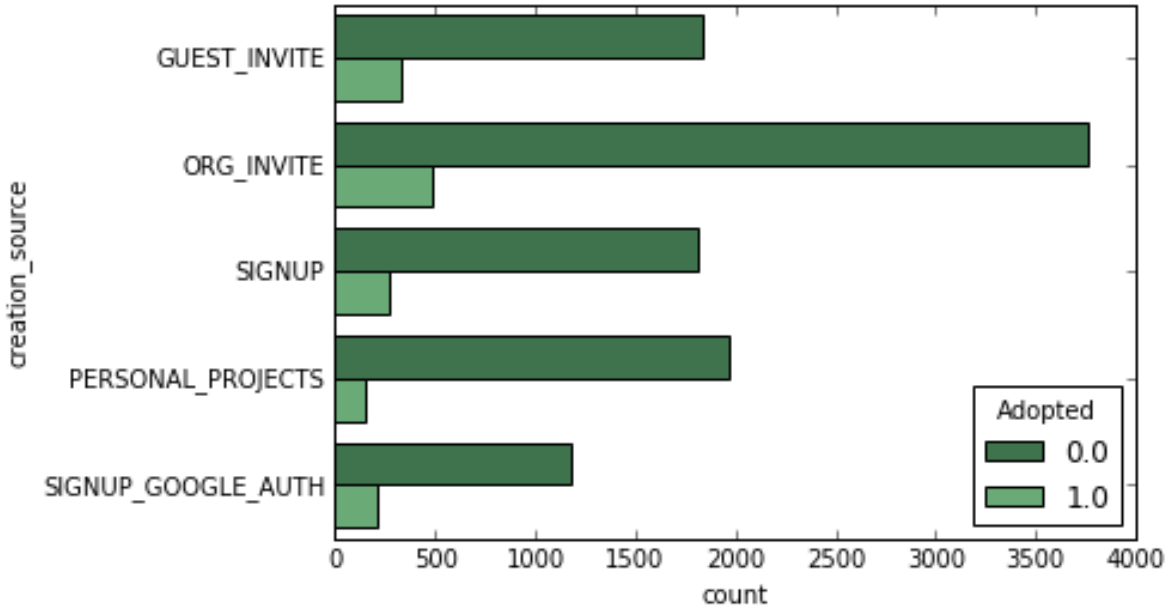
Relax Inc. Data Challenge

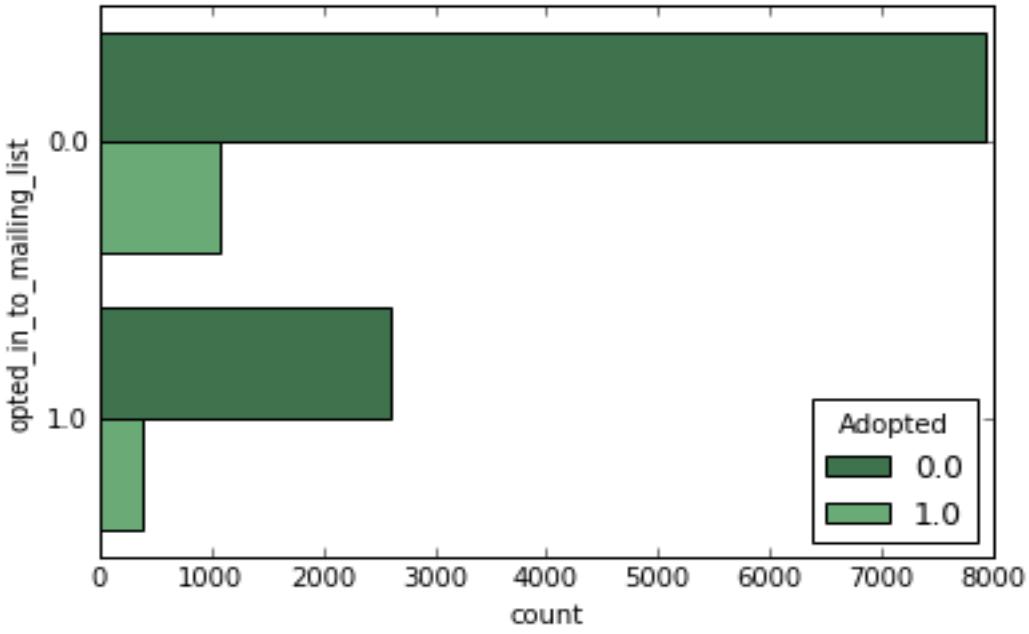
The problem has 2 datasets takehome_user_engagement.csv and takehome_users.csvcontaining subscribed user and user engagement data. A user is considered adopted if he/she has used the service for over 3 times during any week. The first task I performed was to outer join the two tables on object id and user id to get a view of the users who were active for more than 3 days in a week. Performed some data engineering to convert dates to weekday number. The weekdays were stored in the Weekday number column. Then the user_engagement dataframe was iterated over to see if a user is active for more than 3 times in a week, i.e if a user has contains a unique week number more than 3 times.

If it is present, the user is considered adopted and if not, the user is considered no adopted. The adopted data is stored in the 'Adopted' column as 1's and 0's. 1 if Adopted and 0 otherwise. After plotting some graphs, the effects of certain columns can be observed on the Adoption rate of the users.

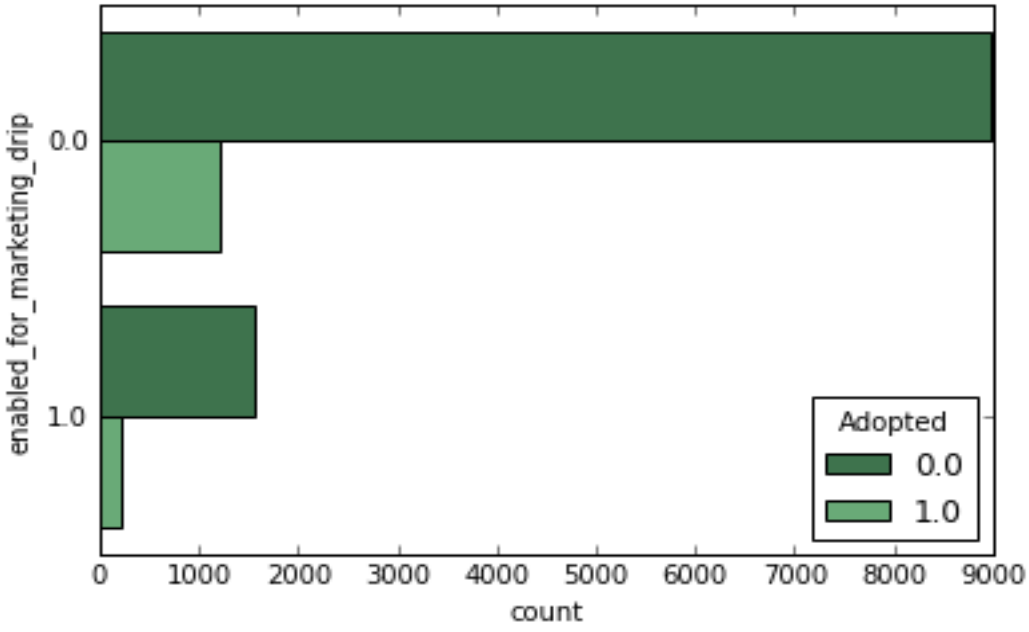
Creation source and Adoption rate: Users subscribed by org_invite seem to have a very high no adoption rate, but at the same time, it also has a higher adoption rate than other categories. Users invited through personal projects have the lowest adoption rate and users invited by guest_invite and sign_up have a relatively higher adoption rate. It can thus be inferred that users invited via guest invite and org invite have a higher adoption rate.



Opted to mailing list:
Users who opted to mailing list relatively have a higher rate of adoption than the users who did not opt to join the mailing list.



Enabled for marketing drip:
Similar to the previous one, users who have enabled marketing drip relatively have a higher adoption rate when compared to the users who have not enabled marketing drip.



The notebook for generating the adopted users column can be found in relax_challenge.ipynb.