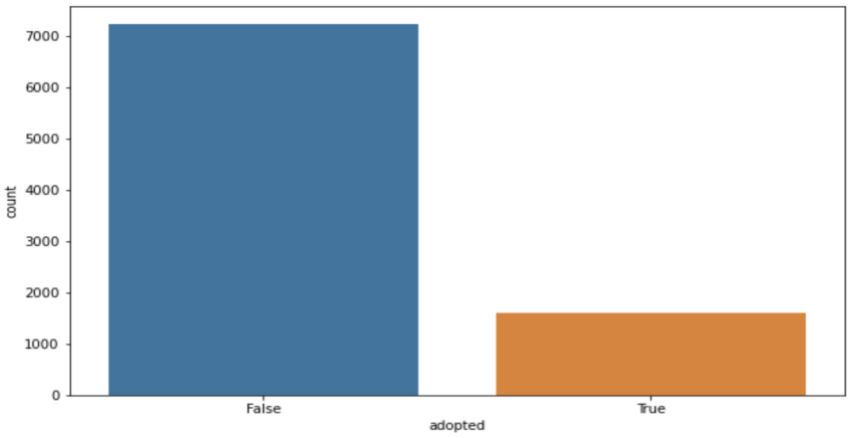
Relax Data Science Challenge

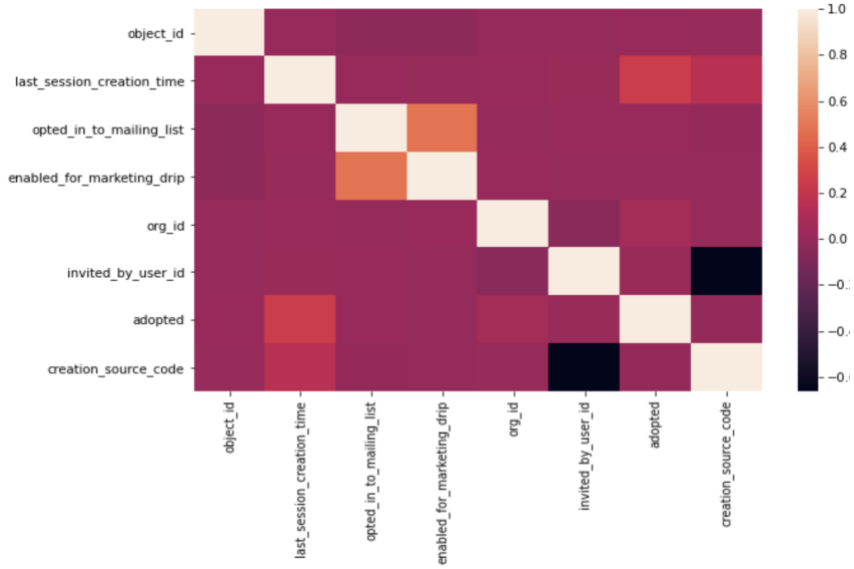
1. **Define an "adopted user" as a user who has logged into the product on three separate days in at least one seven-day period**

This is a crucial part of this challenge. “adopted user” is the dependent variable for our model. Following method was used to derive this value.

After reading user_engagement.csv, we group by “user_id” and resample “time_stamp” to ‘1D’. This returns all the days in between the user logins as well. *rolling()* function is used to move by 7 days across the “time_stamp” and the sum of number of logins in each window is calculated. “adopted” is set to 1 for all those users who have >= 3 logins in 7 days windows.
We have a very low user adoption,



“user_id” and “adopted” column is merged with users DataFrame from takehome_users.csv. After filling null values, changing categorical values to numeric values and dropping unwanted columns, we proceed with the correlation matrix,



2. **Identify which factors predict future user adoption**

“last_session_creation_time” has the most influence on user adoption. This is further confirmed by Random Forest model’s feature_importance_ attribute. In order to improve user adoption frequent email notifications can be sent to users who don't login for couple of weeks.