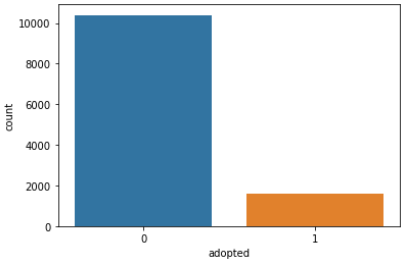
Relax Inc Report by Ketaki Nagarkar

PreProcessing:

From the two data tables provided, the common column was identified as user_id/object_id. In the user engagement table, the user logins by each day were identified and the maximum number of logins for each user in a 7 day period were obtained. From all the users created in the user table, only a subset of users had login sessions created in the user_engament table. To preserve all user id's a Left join was done, with the user table on the left and engagement table on the right. I then created a column 'adopted' indicating only those users that had logged in 3 times or more in a 7 day period(1: adopted users, 0: non-adopted users)

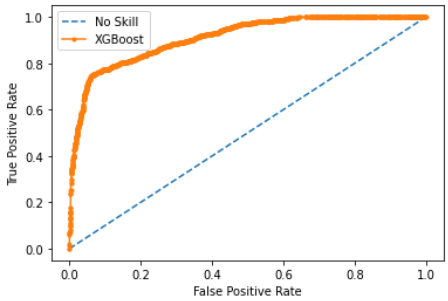
EDA:

Figure 1 shows that a smaller number of users were ***adopted*** compared to total created users.
Fraction of adopted users is 9%



Modeling and discussion:

Last session created date was an important feature for predicting if a user was adopted or not.



I used a XG boost classifier for predicting the binary class for the ***adopted user.*** Due to the class imbalance with the positive class in the minority, f1 score was used as a metric to determine model performance. At best threshold the f1 score for the positive class was 0.71, with an overall accuracy of 92%.

Further Study:

If we can get more data from a user survey during session login, it may help us to obtain user insight regarding repeat/future login behaviours.