Rane Najera

Relax Challenge

This take home challenge was a test to predict user engagement. The datasets given to me were 2 separate csv's file, one being a user table ("takehome_users") with data on 12,000 users who signed up for the product in the last two years, and the other dataset (takehome_user_engagement) contained info on each day a user logged onto the product. I investigated which factors were most important for getting a user to be an "adopted_user", which was defined as a user who logged on 3 times within a 7 day period.

After exploring the data I found out that out of 8823 total users, there were a total of 1656 adopted users, or 18.7% of all users. After figuring this out, I created 2 separate variables, "profile_history" which was the time in between first login and most recent login, as well as "profile_age", which was the time between first login and when the data was collected.

Next step I took was simple encoding my variables and training my data, to try to predict whether a user would become an adopted user. I decided to use a random forest classifier to train the data because the data was originally in both categorical and numerical form. The Random Forest Classifier was fit with the training data and was able to predict whether a user would become an active user with an accuracy of >95%.

```
Accuracy score:  0.9524079320113314
              precision    recall  f1-score   support

           0       0.96      0.98      0.97      1409
           1       0.92      0.84      0.88       356

    accuracy                           0.95      1765
   macro avg       0.94      0.91      0.92      1765
weighted avg       0.95      0.95      0.95      1765
```

Here were the features with the most importance:

profile_history	0.845239
profile_age	0.054218
org_id	0.042518
invited_by_user_id	0.040841
creation_source	0.010890
opted_in_to_mailing_list	0.003449
enabled_for_marketing_drip	0.002846

Based on these numbers it is clear that profile history, or time between first login and most recent login, was by far the most important factor for predicting an adopted user. With this information I believe the best course of action would be to try to encourage or incentivize users who have not logged-in in a certain period of time to log back in. Another strategy that may be beneficial would be to encourage already active users to invite their friends or family.