Amanda Ahn

<u>Relax Inc. Take-Home Challenge</u>

In order to identify which factors predict future user adoption, I filtered through the given datasets to find the adopted users based on their definition and created supervised machine learning models to find the most predictive feature. After loading in and cleaning the data by removing unnecessary columns, dropping null values, one-hot encoding features, and changing data types to the proper ones, we analyze our problem to determine the next steps. Because we are trying to predict which feature will lead to more future user adoptions, the first step is to identify and label which users have logged into the product on three separate occasions in at least a seven-day period. This was achieved by finding all the users who logged in on 3

separate occasions first and dropping the rest from the dataframe. Then, I used the following code, on the right, to find which of these kept users logged in on 3 separate occasions in a seven-day period. By creating a new list of adopted users, I was able to create a new column to the dataframe called 'adopted_user' with values 1 and 0 stating if the user is defined as an adopted user or not.

```python
def within_seven_days(timestamp1, timestamp2):
    return abs((timestamp1 - timestamp2).days) <= 6

adopted_users = []

for user_id in engage['user_id'].unique():
    user_df = engage[engage['user_id'] == user_id]

    for i in range(len(user_df)):
        timestamp = user_df.iloc[i]['time_stamp']
        count = 1

        for j in range(i+1, len(user_df)):
            if within_seven_days(timestamp, user_df.iloc[j]['time_stamp']):
                count +=1
            if count>=3:
                adopted_users.append(user_id)
                break

adopted_users = list(set(adopted_users))
```

Then, I was able to split the data into testing and training subsets based on our desired feature of being an adopted user or not and the rest of the features. The first model I used was a logistic regression model from sklearn, which returned an accuracy score of 0.81. However, I noticed the precision and recall scores for the adopted users were 0. Therefore, I created a random forest model, which produced far better precision and recall scores for both groups and an accuracy score of 0.83. Based on my random forest model, I was able to find the most predictive feature in predicting future user adoption was the 'enabled_for_marketing_drip' feature with a very high importance score of 0.99.
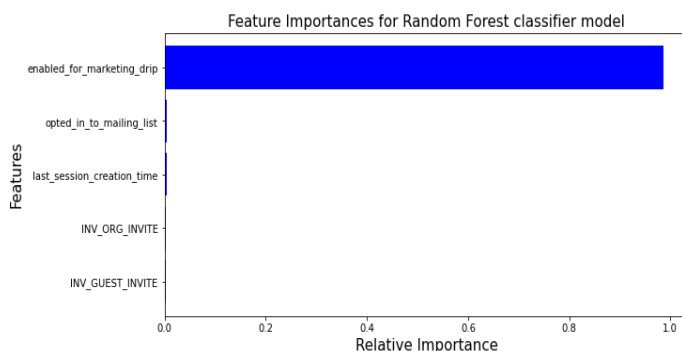
Therefore, our analysis strongly supports that in order to increase future user adoption, it is desired for the user to be on the regular marketing email drip. Before reaching this conclusion, I had hypothesized that whether the user had a guest invite or organization invite would highly influence them being an adopted user. However, finishing my analysis clearly displayed that was not the case.

I believe it would be very valuable to investigate what about being on the regular marketing email drip helps create adopted users through the use of surveys or other data. After determining this factor, Relax Inc. would be able to not only promote users to sign up for the regular marketing email drip, but also create similar marketing factors like this feature, to further increase future adopted users.