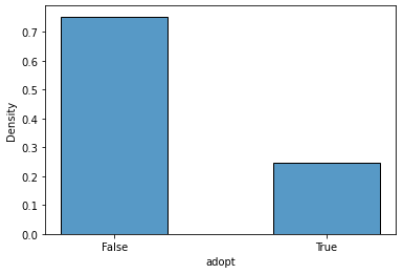
YANLING MENG
SPRINGBOARD TECH CHALLENGE

Relax_Challenge:

Step1: using the "usage summary table' to find out who are the adopted users:

Get the datetime type of index and output a "date" column just to be clear. Calculate how many times the user visited in 7 days. Output a new dataframe "adopted" with two columns: user_id(object_id) and adopt (True or False)
non_adopted/adopted (False/True) = 6639/2184



Step2: Left join adopted and the "takehome_users" dataframe: excluded users without usage information.

2.1. Drop one row that the id was not in the adopted dataframe, and recessive "id"s, "name" and "email" from the joined dataframe.

2.2. Fill NAs and convert some values.

- Fill the "last_session_creation_time" with median
- Convert "creation_time" to days from the last day in the column
- Convert "invited_by_user_id" to True/False categorical with "True" indicating the user is invited by another user.

Step3. Make predictions using RandomForestClassifier

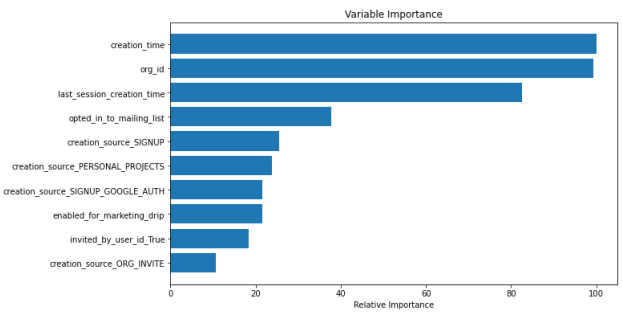
- Use "adopt" as the target variable
- Use others as predictors, use "get_dummies" for the categorical columns
- Use SMOTE() to balance the True/False values
- Split the whole data set into "train" and "test"
- Use MinMaxScaler to transform X matrices
- Fit a RandomForestClassifier.
- Classification report showed a 0.78 average accuracy.

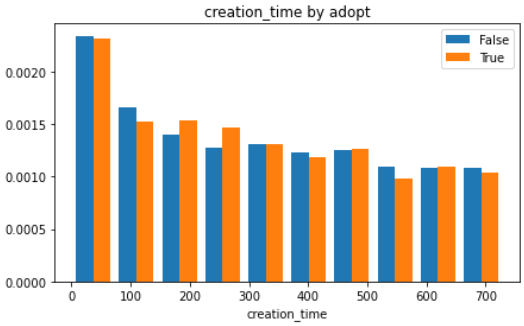
	precision	recall	f1-score	support
False	0.75	0.85	0.80	1355
True	0.82	0.71	0.76	1301
accuracy			0.78	2656
macro avg	0.79	0.78	0.78	2656
weighted avg	0.79	0.78	0.78	2656

Step4. Variable importance:

- Check the model features and print out the importance of variables for predictions.
- The most important seemed to be "creation_time": the user_ids created in the first 3 months of 2014 have relatively higher adopt rate.

Other information may help with the prediction:
Information about organizations "org_id" could be useful: org_id could be group as different types of industries and have a clearer understanding rather than just numbers.
Getting more data for both tables after 2014 would definitely help.





YANLING MENG
SPRINGBOARD TECH CHALLENGE