I was asked to investigate two (2) excel data frames as described below:

1. takehome_users
    a. There were a number of variables herein & many null values on two (2) of the variables; the null values of one (1) of the two (2) variables was addressed
2. takehome_user_engagement
    a. Consisted of 207,916 non-null observations with two (2) variables; notably
        i. user_id : the ID of the user
        ii. visited : contained 1's to confirm the user visited
        iii. time_stamp : the index but wasn't presented as a time_stamp dtype but was converted to one

The goal for the project was to **identify the factors predict future user adoption**. **User adoption was defined as a user who has logged into the product on three separate days in at least one seven-day period**.

I setup a variable called:

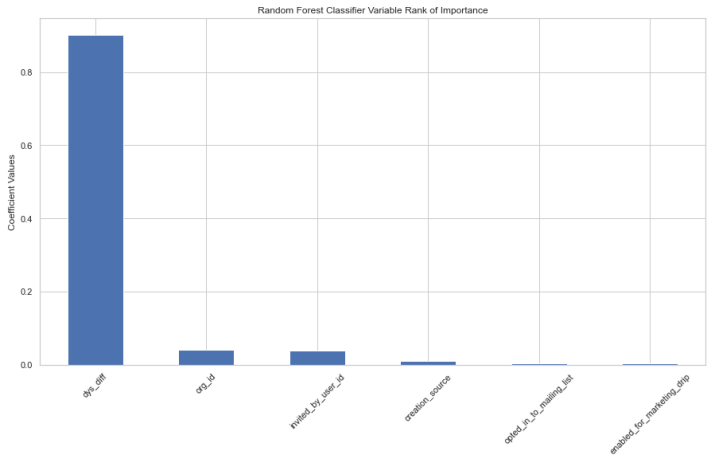
1. dys_diff
    a. The total number of days (days only) between when `last_session_creation_time` & `creation_time`

I used the Scikit-learn family throughout the modeling section. I first turned to Label Encoder to encode target values. I then identified there were a great number of null values in one of the variables ('invited_by_user_id'); I then replaced the nulls with a unique number as dropping the nulls & others would have been substantial.

Staying with the Scikit-learn family, I then built out a Random Forest model on a split. This yield accuracy of 96.22% on the test set. I were then inclined to turn to a Confusion Matrix which yielded very high results (below).

	precision	recall	f1-score	support
0	0.97	0.98	0.98	2133
1	0.91	0.89	0.90	514
accuracy			0.96	2647
macro avg	0.94	0.94	0.94	2647
weighted avg	0.96	0.96	0.96	2647

I ended on determining a Rand Forest Feature Importance scale which presented the variable which I built to be have the controlling interest in.



`org_id` & `invited_by_user_id` helped but given their relative size of uniqueness, they may not have had as significant of an influence on an `adopted user`. The controlling interest in `dys_diff` is arguably due to how it's structured. Adopted users are likely to be more active since commencement.