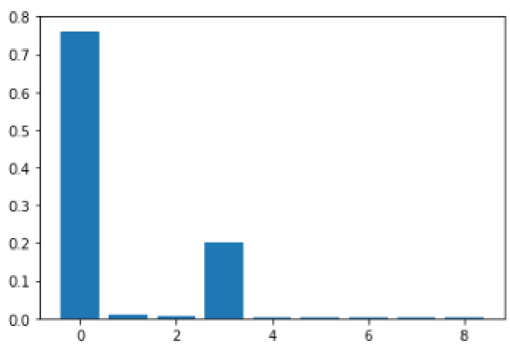
First, I aggregated the user engagement data and added together the counts for each login attempt made by a user during a seven-day period. I was able to identify there were 10555 not adopted users and 1445 adopted users within the dataset. An adopted user is a user who has logged into the product on three separate days in at least one seven-day period.

After cleaning the data, I preprocessed the data by creating dummy variables for the 'creation source' feature and standardized the 'org_id' feature because it had a large range of values.

_GUEST_INVITE	_ORG_INVITE	_PERSONAL_PROJECTS	_SIGNUP	_SIGNUP_GOOGLE_AUTH
1	0	0	0	0
0	1	0	0	0
0	1	0	0	0
1	0	0	0	0
1	0	0	0	0

Now that the data was ready for modeling, I create a dataframe of features that could be used in the random forest classifer. The features I selected were: 'last_session_creation_time', 'opted_in_to_mailing_list', 'enabled_for_marketing_drip', 'org_id', '_GUEST_INVITE', '_ORG_INVITE', '_PERSONAL_PROJECTS', '_SIGNUP', '_SIGNUP_GOOGLE_AUTH'. I also created a dataframe for the target variable. I fit the features and target variable to the random forest model and then called the feature_importances_ attribute to get the importance of the features and plot them.



From the plot above it can be seen that 'last_session_creation_time' and 'org_id' have the highest feature importance. Below is a plot of feature importance after I removed 'last_session_creation_time' and 'org_id' from the features dataframe and refit the data to the model. Now it can be seen that 'personal_projects' and 'guest_invite' have the 1st and 2nd highest feature importance respectively.