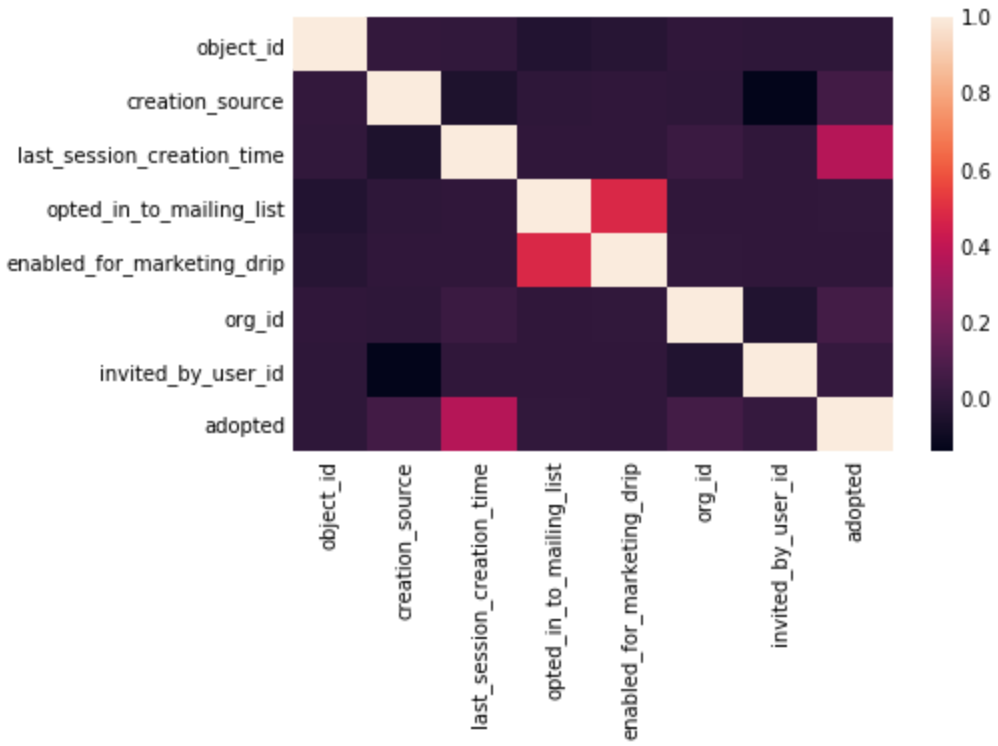
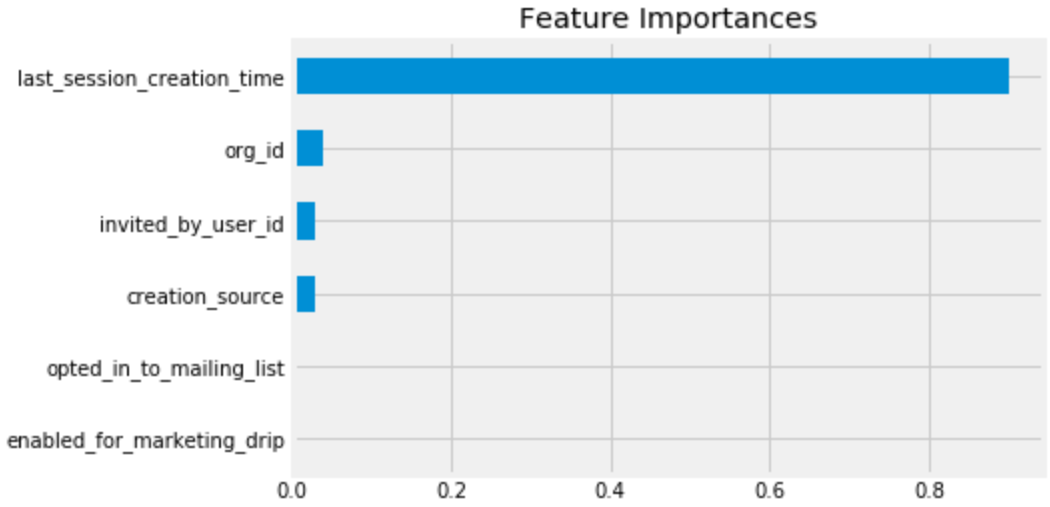
A function was written to count the number of logins within 7 day intervals and the results are mapped to a feature that describes whether a user is adopted or not. After imputing null values in the dataset, plotting a correlation matrix of the features shows none of the features are highly correlated. The only feature that has medium correlation with user adoption is last_session_creation_time. The null values in last_session_creation_time were filled with the median of the existing values.



An over sampling strategy was used to balance the classes prior to modeling since only about 10% of the observed users are adopted users. Gradient Boosting model predicted on test set with an accuracy of 90%.



The gradient boosting model shows last_session_creation_time as the single most important predictor of adoption. Decision tree model predicted with similar accuracy. Since there is only one feature that is somewhat reliable for predicting adoption, it would be expected that any model for classification would show comparable performance.