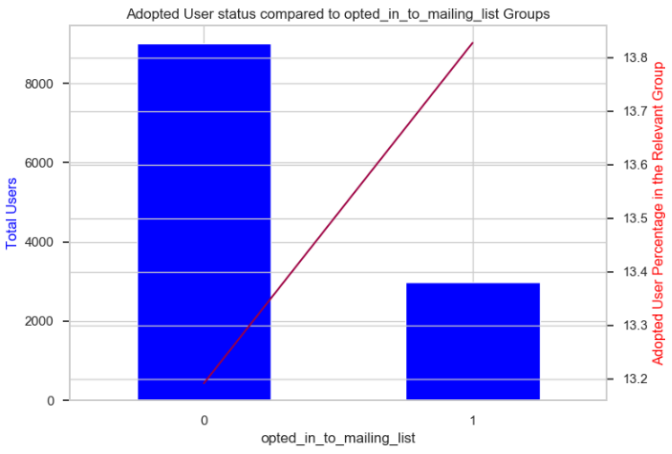
RELAX KEYNOTES

There are 12000 signed up for the application and only 8823 of them users used it. There are 3177 inactive users in the system. Total percentage of the adopted users is 13.35 %. I have done some research to identify which factors predict future user adoption as it might be seen below.

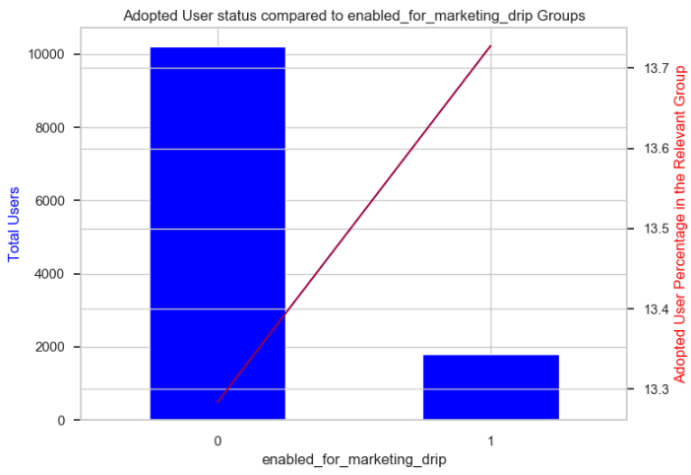
Analyze opted_in_to_mailing_list with adopted users

There is a negligible percentage difference between the two classes. This implies that customers who opted in to mailing list has very little to do with if they ended up being an adopted user.



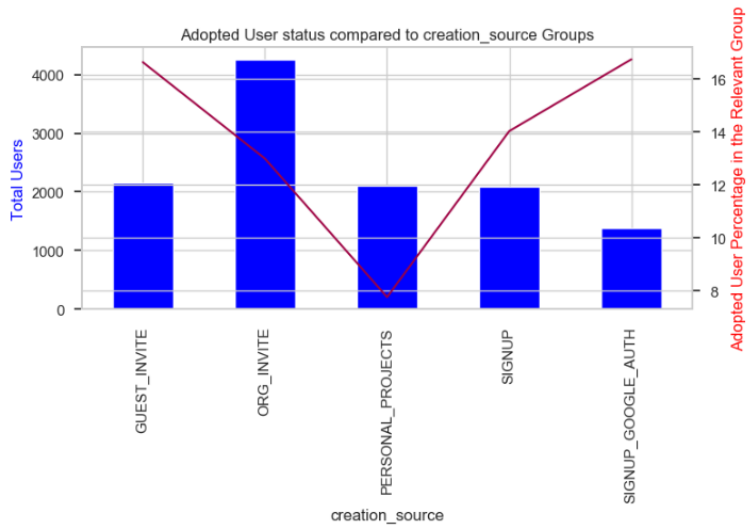
Analyze enabled_for_marketing_drip with adopted users

We obtain extremely similar results as above. Unfortunately, the Marketing Drip feature also isn't too helpful in identifying adopted users.



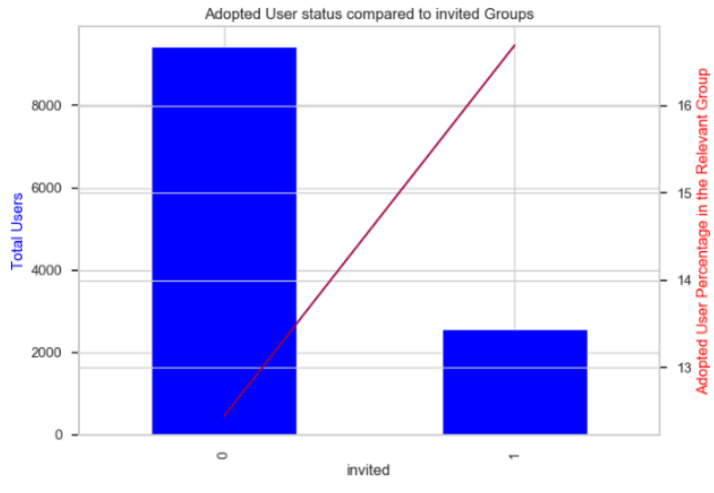
Analyze creation_source with adopted users

People with Personal Projects are extremely less likely to be an adopted user as compared to other creation source methods. But since this does not help us in identifying adopted users, this too, isn't a very useful feature.



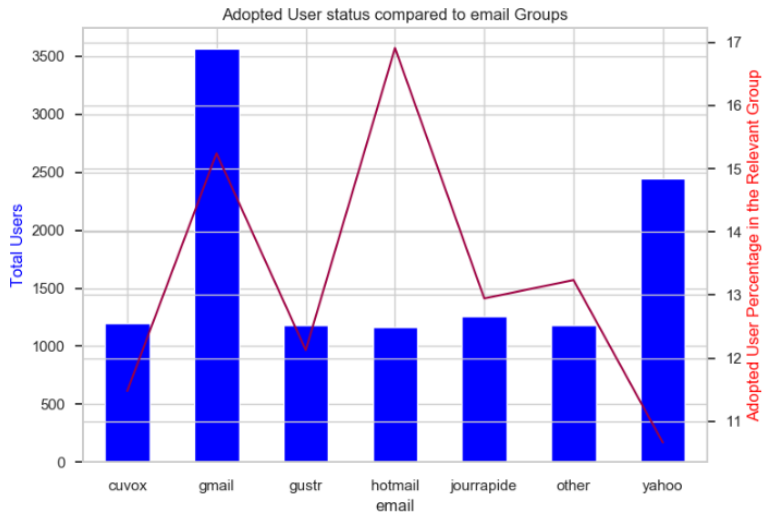
Analyze invited_users with adopted users

People who have invited other users have a slightly larger percentage chance of being an adopted user.



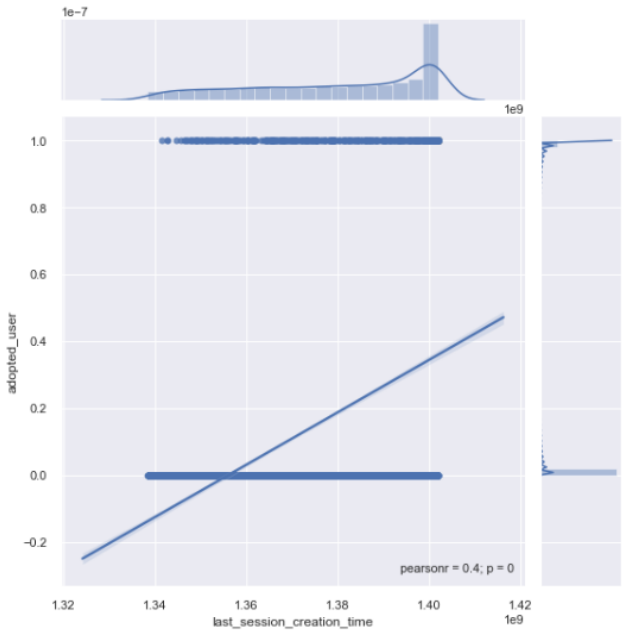
Analyze email domain with adopted users

People who have signed up using hotmail are most likely to become adopted users whereas people with yahoo IDs are the least likely. However, since the difference is so small, it might be entirely possible that these observations are due to noise.



Analyze last_session_creation_time and adopted users

There seems to be a medium correlation between the last session creation time and the target variable. This may be the most important feature to predict our adopted users as all the other variables have been extremely weak in explaining the target variable.

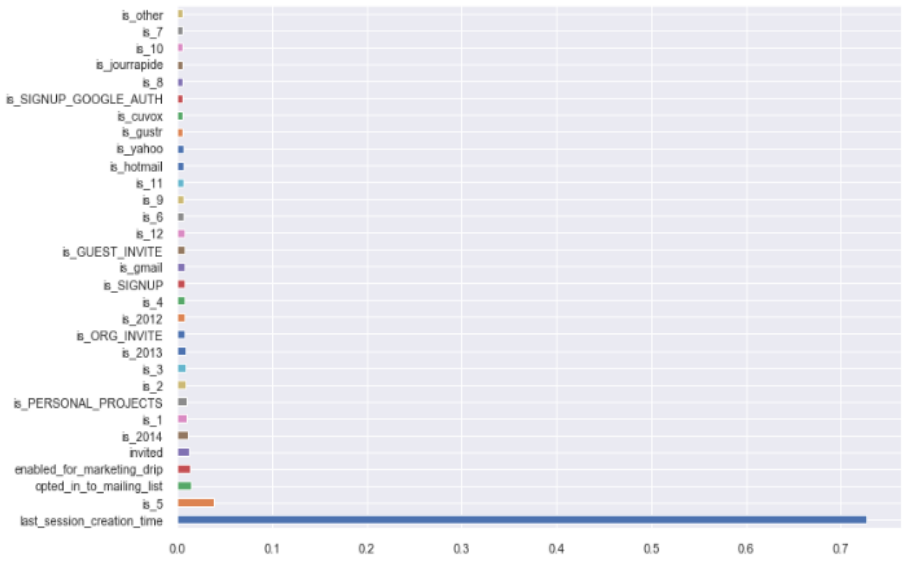


Modelling

Since we have a small feature set and sample size, Kneighbors and RandomForestClassifier are both good choices for this data set. However, we have a mix of categorical and numerical data so RandomForestClassifier will be easiest to implement. I will use the metric for this classification problem ROC_AUC.

Feature Importance

Whatever coefficients the classifier has assigned to the features will indicate which ones it found more useful in making the correct classifications.



The most important feature to predict user adoption is last_session_creation_time.

Conclusion

From the analysis and the predictive modeling that we have performed above, the results can be summarized as follows:

86% of our userbase are non-adopted users. This clearly indicates that there is a lot that can be done to decrease this proportion relative to the number of adopted users.

last_session_creation_time is the most important feature when it comes to predicting adopted users. It accounts for more than 72% of the importance in our Random Forest Classifier.

All the other features have close to zero importance and do not explain the target variable well.