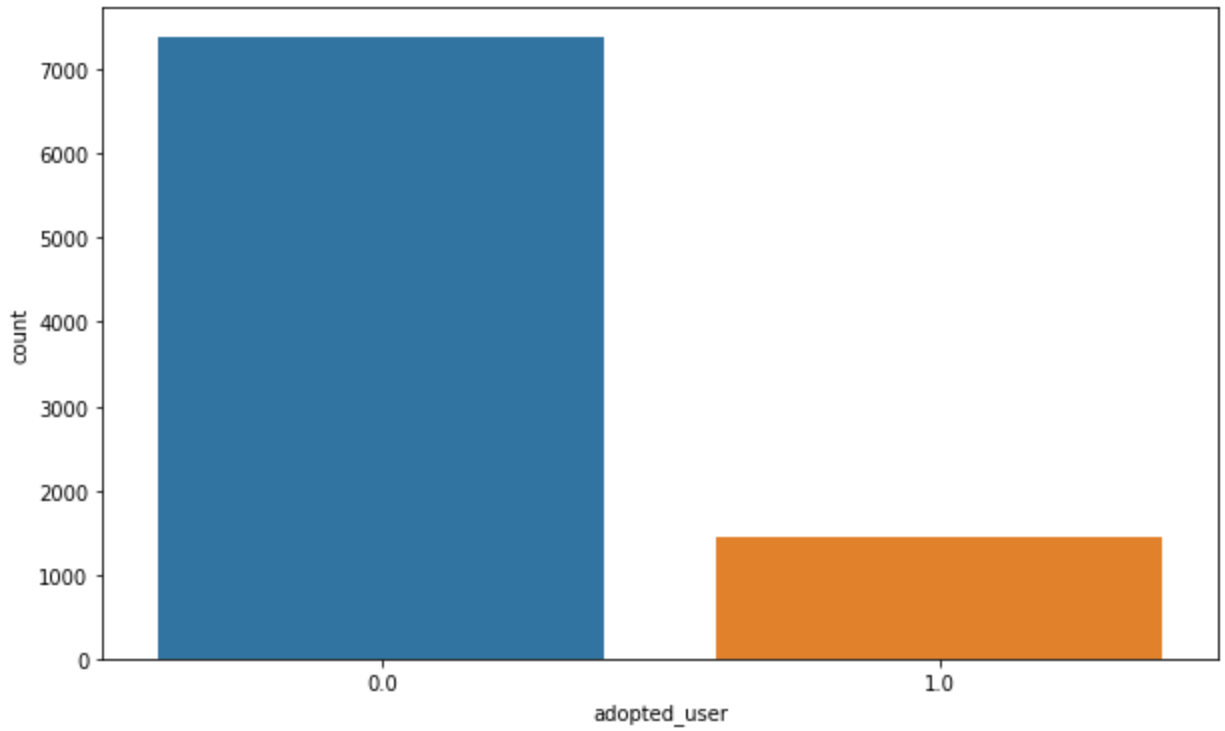
Relax Data Science Challenge

1. **Defining an "adopted user" as a user who has logged into the product on three separate days in at least one seven-day period.**

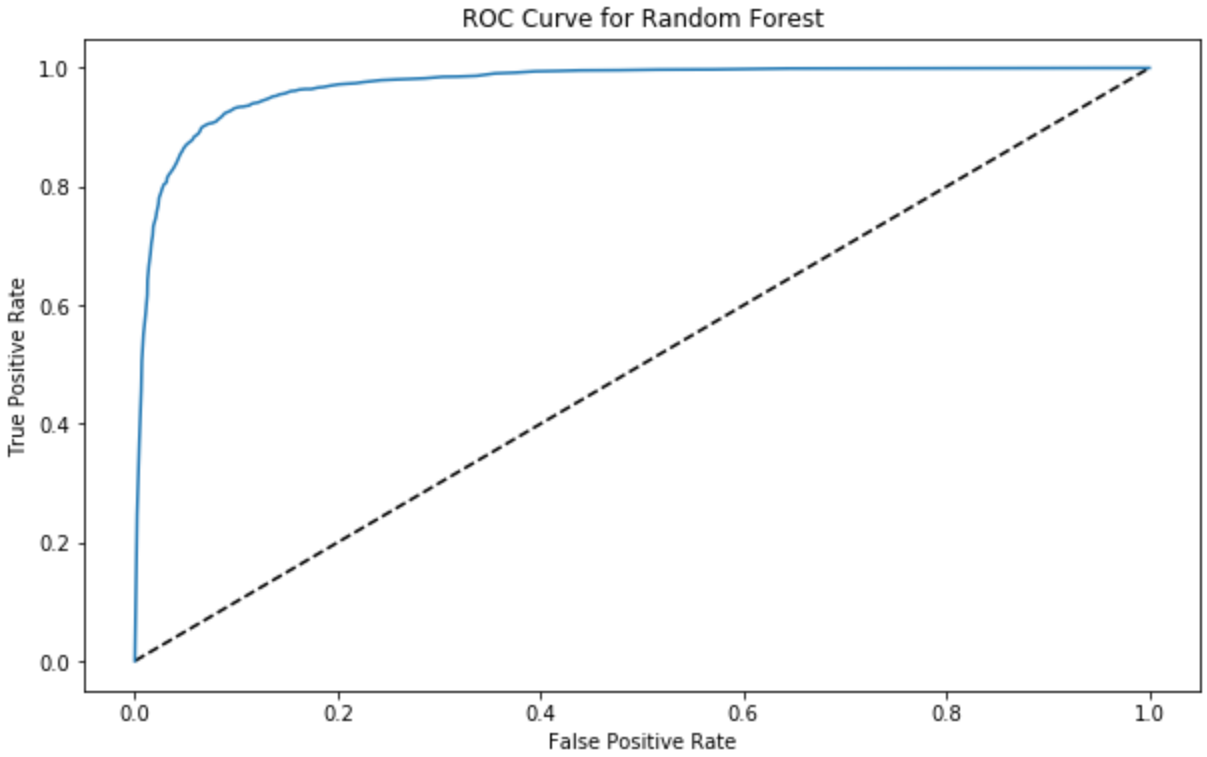
In this problem the “adopted user” is the target variable. I used the following methods to derive “adopted user” values.
First, I load the ‘user’ and ‘user_engagement’ csv files to pandas dataframe and then I grouped the user visit to weekly and sum the number of visits using groupby() and pandas Grouper() methods. Then made a new column “adopted_user” for users who had in a week visited 3 days separately with the value as 1 and others as 0. Kept only ‘user_id’ and ‘adopted_user’ columns and removed the rest column from the ‘user_engagement’ DataFrame. and then merged with users dataframe on ‘object_id’ column.
The percentage of users who are adopted users is 12.04 %.



2. **Identify which factors predict future user adoption.**

Before Machine Learning, I removed unwanted columns, changed categorical values to numerical values and filled the null values. Compared the Logistic Regression with Random Forest classifier and the Random Forest classifier had higher ROC-AUC score of 0.97. 'last_session_creation_time' has the most influence on user adoption so, to increase the user

engagement we can send email notifications periodically. The org_id, the organization (group of users) they belong to has also second most influence on user adoption.



	importance
last_session_creation_time	0.636937
org_id	0.114062
object_id	0.092636
invited_by_user_id	0.081916
creation_source_code	0.040766
opted_in_to_mailing_list	0.021379
enabled_for_marketing_drip	0.012303