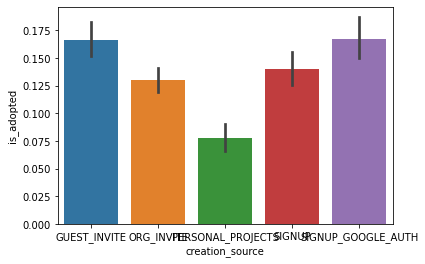
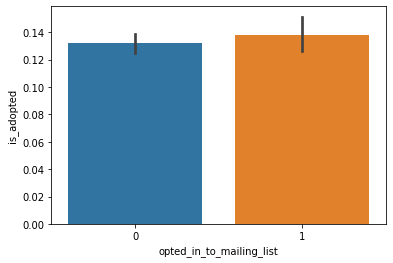
**Data Preparation:** There are two datasets, one logs the user summary (takehome\_users.csv) and the other (takehome\_user\_engagement.csv) details the regular user activities. The main challenge was to properly prepare the dataset for a classification model.

An adopted user is defined by a person who logs at least thrice within a week. I have used the second dataframe to first calculate the number of user login within a week and then took the max of that value to check whether that person logged in for more than twice or not. Next another column “is\_adopted” was created to classify that user. Then this modified dataset was joined properly with the first dataframe to create the final dataframe (df\_merged).

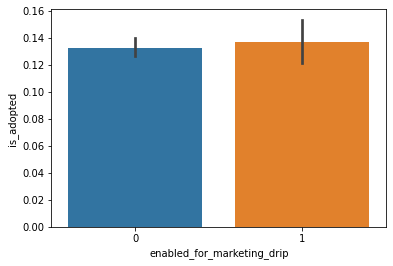
There are 12000 data points in my final dataset. The details of each step of creating this dataset have been mentioned in the colab notebook.

**Dataset Understanding:** The dataset is highly imbalanced as there are only 13.65% data from the adopted class.

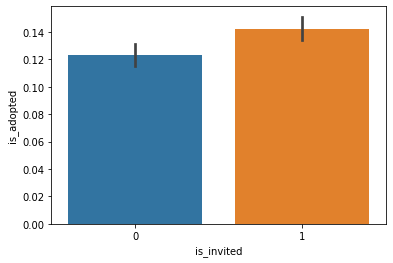
Those who have joined through either of the guest invitation or google authorization are more adopted than others.



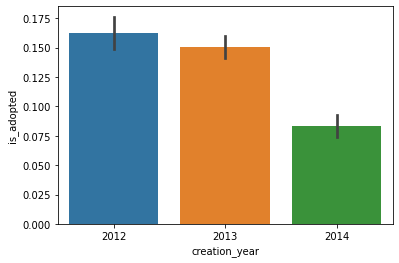
This result is a little surprising to me. Contrary to my assumption that the people who opted in to the mailing list are more adopted; I didn't find any significant difference between the two groups.

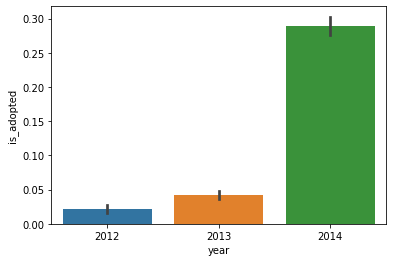


Again there is no significant effect of enabling marketing drip.

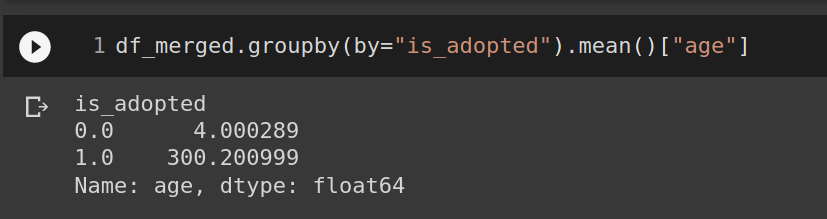


Those who are invited (is\_invited is a derived feature) are usually more adopted.

Those who created their account in 2014 (creation\_year is a derived feature) are significantly less adopted than those who created in 2012 or 2013.



Most adopted users have logged in the year (year is a derived feature) 2014



Age is defined by the difference between the last login date and the date of creation. The average ages of not adopted and adopted users are 4 days and 300 days respectively. So, those who have used the product for long times are usually more adopted.

Please check my collab notebook for the model building and the feature importance part.