Relax Take Home Challenge

Goal: To identify 'adopted users' and identify factors that could predict future adopted users. An adopted user is a user who has logged into the service at least 3 times within a 7-day period.

Datasets:
- A user table ("takehome_users") with data on 12,000 users who signed up for the product in the last two years
- A usage summary table ("takehome_user_engagement") that has a row for each day that a user logged into the product

Finding 'adopted users': The user engagement dataset was used to determine the adopted users. Here is the procedure:
- Convert the index to the 'time_stamp' variable so that we can perform timeseries analysis on the dataset.
- Group the dataset by 'user_id' and then resample the data by 7 day intervals and aggregating by counts. This provides us with the number of logins per 7-day period for each user. See snippet of resulting dataframe below.

```
user_id  time_stamp
1        2014-04-22    1
2        2013-11-15    1
         2013-11-22    0
         2013-11-29    1
         2013-12-06    1
         2013-12-13    0
         2013-12-20    1
         2013-12-27    1
         2014-01-03    1
         2014-01-10    0
         2014-01-17    0
         2014-01-24    0
         2014-01-31    1
         2014-02-07    3
         2014-02-14    1
         2014-02-21    0
         2014-02-28    0
         2014-03-07    2
         2014-03-14    0
         2014-03-21    0
         2014-03-28    1
3        2013-03-19    1
4        2013-05-22    1
5        2013-01-22    1
6        2013-12-19    1
7        2012-12-20    1
10       2013-01-16    2
         2013-01-23    0
         2013-01-30    2
         2013-02-06    1
         2013-02-13    3
         2013-02-20    1
```

- Then, only the adopted users were selected by filtering only those users who had at least one value of 3 or more in the table above.

user_id	time_stamp	count
2	2014-02-07	3.0
10	2013-02-13	3.0
	2013-02-27	3.0
	2013-03-20	3.0
	2013-04-10	3.0
...	...	...
11988	2014-03-31	3.0
	2014-04-28	4.0
	2014-05-12	3.0

- These user IDs were then collected in a list.
- A new variable in the main dataset was created which was created by comparing the user IDs in that dataset to the filtered user IDs who were the adopted users.

Exploratory data analysis

A few variables were analyzed against the 'adopted_user' variable to see whether they explained some of the variability in that feature. Below are some of the plots for that analysis. Out of the 4 variables analyzed, the month of account creation seemed to be the most promising.

From the month of creation data, it appears that people who tend to signup during May (5) don't tend to be adopted users. In fact, May has the lowest frequency for adopted users. Also, there is a steady decline in signups for adopted users from Jan to May whereas for other users, there is an increase.