Relax Challenge Analysis

First thing I did with the data is to create a ‘Week’ column by using a datetime method for year and week. A group by function on user_id and week, and applying a size() aggregate gave me the count of logins for each week, by beach unique user ID. With this I could easily create a new dataframe to get the users with a count of 3 or greater, and give me the unique ID’s of each.

Using pandas.isin method applied a boolean of True or False for object ID’s, I created a column for this called ‘adopted’.

1. By the means of how the user signed up show a significant trend; users who were invited by a guest are among the largest group of retain users at 15%, and with users who signed up through Gmail as their identification right behind. Users who were invited to work on a personal project were among the lowest percent of retained users at 6.5%. This makes sense as their use for the software was probably temporary, the original inviter is likely a retained user.

2. Hotmail users are the highest retained users at 15%, and Yahoo users are at the lowest at 9% This is likely arbitrary, but it's possible theres a reason why.

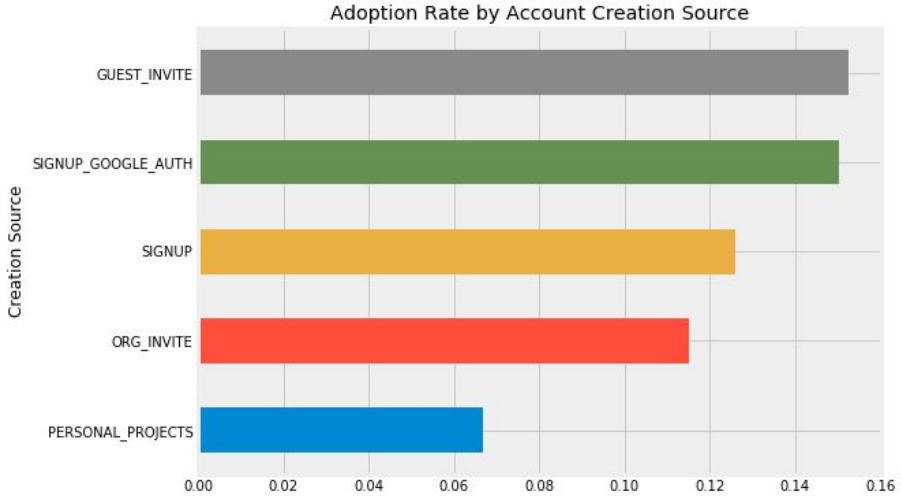
3. Can see a downward trend that the greater the organization size, the lower the retention. A Pearson correlation of -.49 confirms this trend.

4. Of all the users that created a session time in the latest month of June 2014, about 90% were adopted users, and of the month before that in May, 40% of users were adopted users.

```
plt.figure(figsize=(8, 6))


users.groupby('creation_source')['adopted'].mean().sort_values().plot(kind='barh')

plt.title('Adoption Rate by Account Creation Source')
plt.xlabel('Adoption Rate')
plt.ylabel('Creation Source')
plt.show()
```




```
users.groupby(['domain'])['adopted'].mean().sort_values().plot(kind='barh')
plt.xlabel('Adoption Rate')
plt.ylabel('Email Domain')
plt.title('Adoption Rate by Email Domain')
plt.show()
```