# ANALYZING CUSTOMER SIGNUP AND RETENTION FOR AN E-COMMERCE COMPANY

**Abstract:** This work illustrates trends for customer signup and retention for a US based e-commerce company. The data consists of two files: first, the signups.csv file and second, the visits.csv file. The signups.csv file contains information about unique customers, listed by their customer id's. This file provides information about the date on which each customer signed up on the website. Also provided is the information about the type of device they signed up from; listed as "device" and the mode of signing up; listed as "auth_type". The other file, named visits.csv consists of information about the date that each customer, identified by their customer id visited the website after signing up. A customer may have visited the website multiple times after signing up, hence the rows in this file are not unique. The date information in both files is in the format of YYYY-MM-DD. The overall task of visualizing the signups and visits to the website by customers is divided into two tasks: first, identifying the online method and the device type customers use to sign up, and second, investigating proportion of users and the trends for customer visits after they have initially signed up on the website. The rest of the document will go into detail of analysing and visualizing each of the above mentioned aspects.

Keywords: signup, visits, retention

## 1. Frequency distribution of users by device and auth_type

The head of the signups table (signups.csv) appears as follows:

|   | Unnamed: 0 | uid | signup_dt | auth_type | device |
|---|---|---|---|---|---|
| 0 | 1 | 21639492 | 2016-06-01 | C | 1 |
| 1 | 2 | 21640161 | 2016-06-01 | A | 6 |
| 2 | 3 | 21640773 | 2016-06-01 | C | 5 |
| 3 | 4 | 21641605 | 2016-06-01 | A | 1 |
| 4 | 5 | 21643130 | 2016-06-01 | C | 6 |
| 5 | 6 | 21644509 | 2016-06-01 | C | 1 |
| 6 | 7 | 21645033 | 2016-06-01 | A | 2 |
| 7 | 8 | 21645932 | 2016-06-01 | C | 2 |
| 8 | 9 | 21645974 | 2016-06-01 | C | 1 |
| 9 | 10 | 21646765 | 2016-06-01 | A | 1 |

The table only shows the first 10 rows of the signups dataframe, as printed using the

PANDAS library in Python. Using this data, the frequency distribution of the users based on the type of device they used for sign up on the website is illustrated in Figure 1. The data only provides the device type information as an integer and not by name. However, Figure 1 clearly shows that device type – 1 is the most popular amongst the customers to sign up on the company website. The overall distribution resembles a double delta distribution with device type – 6 also contributing to ~ 15000 user sign ups.

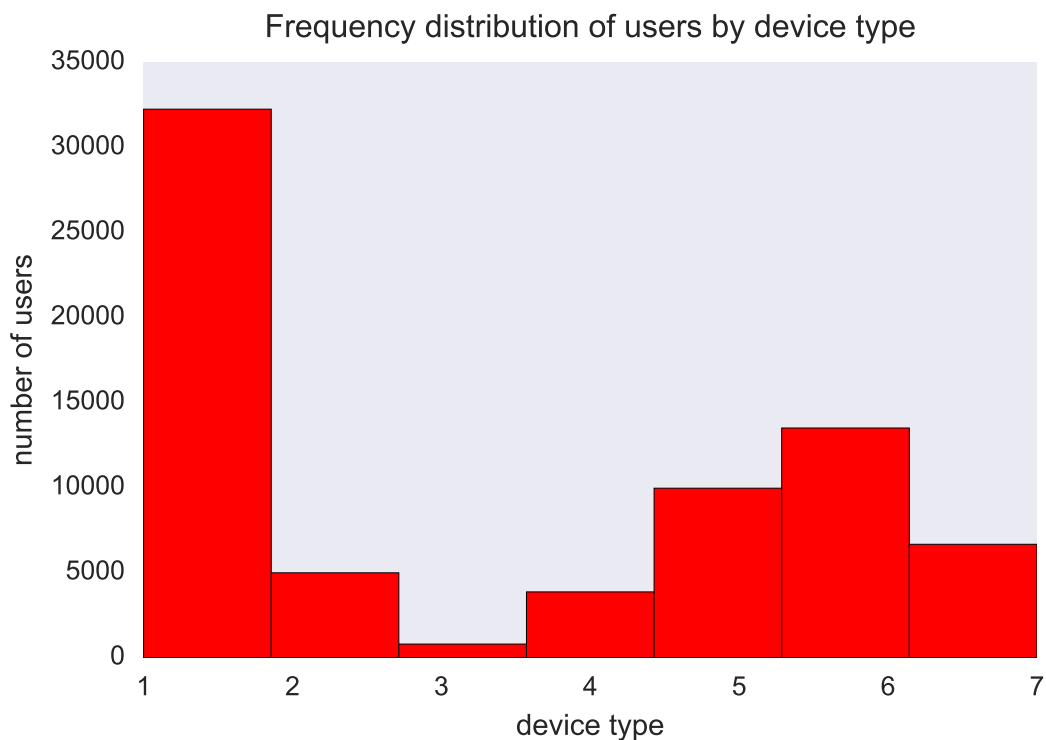Frequency distribution of users by device type



Figure 1 – Frequency distribution of all users by device type

Similar to Figure 1, the frequency distribution of users by device type and segmented by the mode of signing up on the website, given as -  A, B or C is shown in Figure 2. Again, it can be seen that device type 1 and 6 are the most popular among customers, irrespective of the method of signing up online. Customers who sign up using auth_type 1 and 3 also tend to use device type 5, which on the distribution plot comes in at a close 3rd behind device type 6.

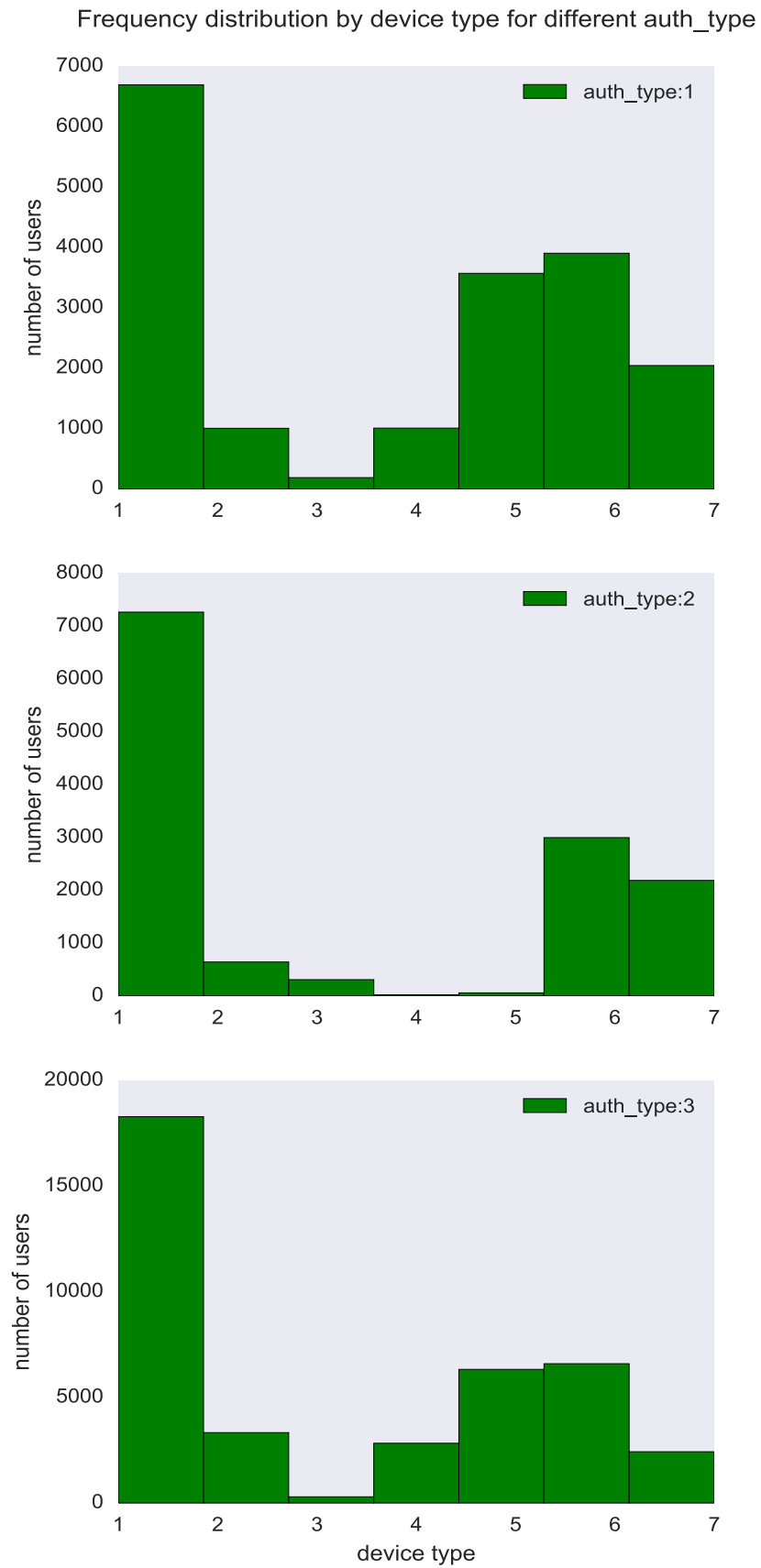Frequency distribution by device type for different auth_type



Figure 2 – Frequency distribution of users by device type, segmented by auth_type

Figure 3 is a bar graph that illustrates the most popular online methods of signing up on the websites which customers use, based on different device types. Although the auth_types are listed as variables – A, B and C, we can think of them as similar to signing up using a Facebook account, Google account etc. The figure shows that method "C" is the most popular and method "A" is the least popular overall, irrespective of the type of device used to sign up. For customers who use device type 4 and 5, only 1% of them use auth_type – A to sign up online, whereas customers who use device type 3 tend to sign up equally likely using auth_type B and C.
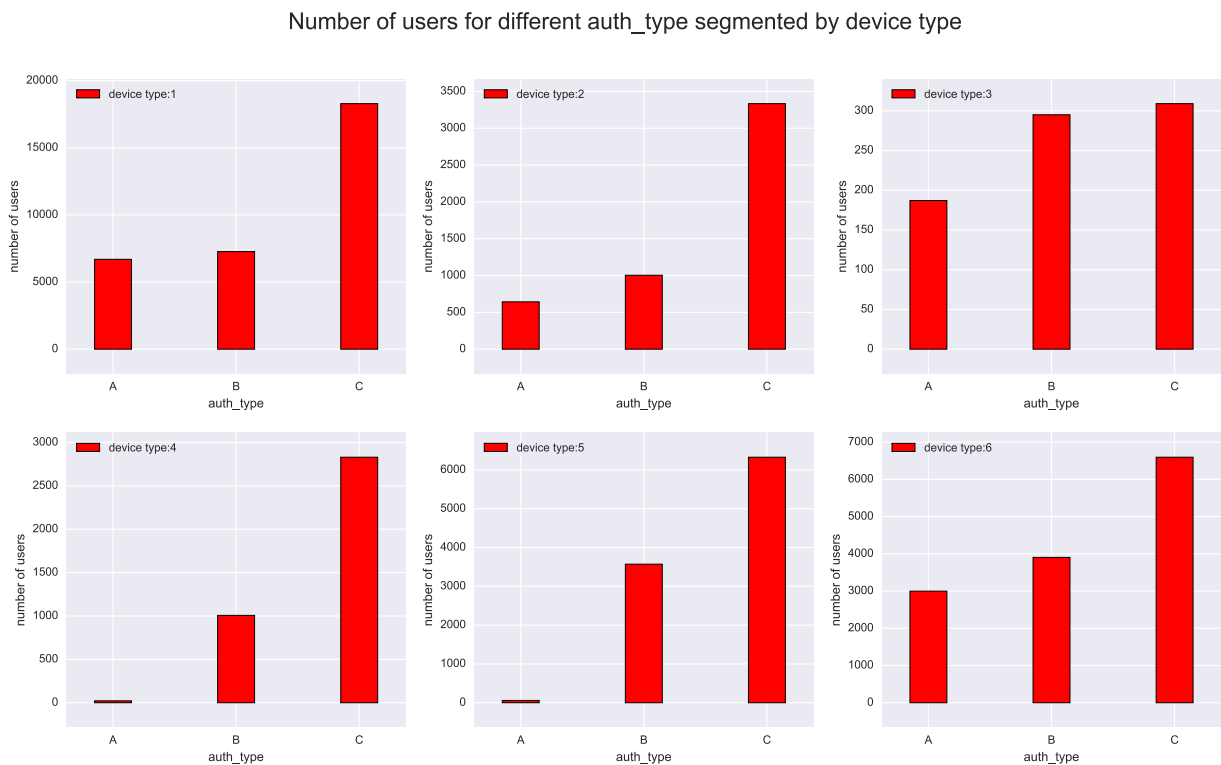
Number of users for different auth_type segmented by device type



Figure 3 – Bar graph showing the mode of signing up (auth_type) for different device types

## 2. Retention of customers after signing up on website

A key question which is analysed to investigate the retention of customers is that for users who signed up on any date starting from 06/01/16 until 10/30/16, what is the proportion of users who came back to visit the website in the $1^{st}$ week after signup, in the $2^{nd}$ week

after sign up, 3$^{rd}$ week after sign up and so on. For example, for users who signed up on 06/01/16, what is the proportion who came back between 06/02 – 06/08, then 06/09 – 06/15 and so on, all the way up to 24 weeks in the future. The head of the visits table (visits.csv) after splitting the date column (dt) into separate columns for year, month and day is given as follows:

| | Unnamed: 0 | Unnamed: 0.1 | uid | dt | year | month | day |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 21655745 | 2016-11-11 | 2016 | 11 | 11 |
| 1 | 1 | 2 | 22635182 | 2016-11-02 | 2016 | 11 | 2 |
| 2 | 2 | 3 | 22361195 | 2016-07-24 | 2016 | 7 | 24 |
| 3 | 3 | 4 | 22132291 | 2017-03-19 | 2017 | 3 | 19 |
| 4 | 4 | 5 | 22639754 | 2017-02-15 | 2017 | 2 | 15 |
| 5 | 5 | 6 | 22998116 | 2016-09-06 | 2016 | 9 | 6 |
| 6 | 6 | 7 | 23344314 | 2016-10-05 | 2016 | 10 | 5 |
| 7 | 7 | 8 | 22422820 | 2016-11-14 | 2016 | 11 | 14 |
| 8 | 8 | 9 | 23642588 | 2017-04-07 | 2017 | 4 | 7 |
| 9 | 9 | 10 | 23705466 | 2016-10-21 | 2016 | 10 | 21 |

The table provides dates on which a user, identified by the "uid" visits the website after signing up on a date given in the signups table, shown earlier. Note that a user could visit the website on multiple dates, so that the 'uid' column in the visits.csv file would not be unique, as opposed to in the signups.csv file.

In order to compute the proportion of users who visited the website in the n$^{th}$ week after signing up on a particular date, the following methodology is used:

(1) The total number of users who signed up on the given date is calculated and this information was stored in a table.

(2) Next, the total number of users who signed up in the n$^{th}$ week after the signup date, identified by unique uid's is computed and this information was stored in a second table

(3) An INNER JOIN is carried out of the two tables to get the intersection, that is to find out the number of users who signed up and who also visited the website in the nth week.

Figure 4 shows a heat up, denoting the proportion of visitors visiting the website every week after sign up date. The first date of sign up is 06/01/16, running until 10/30/16. In all, there are 152 sign up dates shown on the y-axis of Figure 4 and 24 weeks shown on the x-axis of the figure. The figure illustrates that throughout all sign up dates, the proportion of visitors visiting in approximately 66%. This proportion decreases as time goes on with roughly only 15% who visit in the 24th week since sign up.
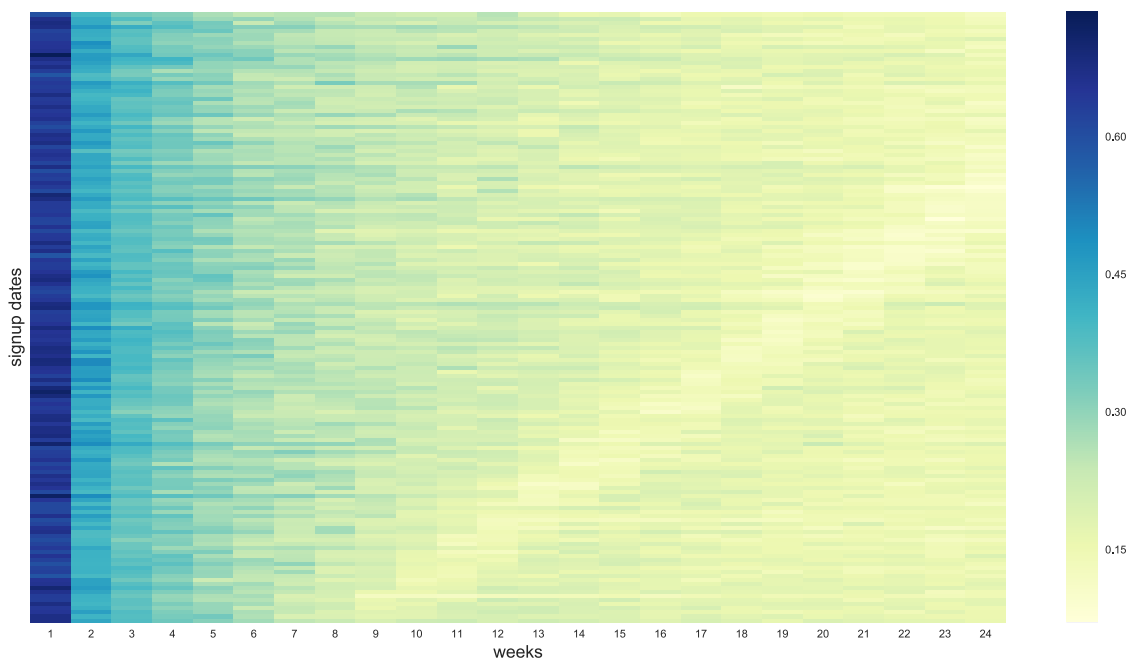


Figure 4 – Heat map of the proportion of visitors every week after signup date

The time history of the proportion of users visiting every week after the sign up date is plotted in Figure 5. To make the illustration less messy, only 4 specific dates have been chosen lying at roughly equal intervals between each other. The figure corroborates the conclusions of Figure 4 wherein the proportion of users exponentially decreases with each

passing week and reaches a steady state after week 20, implying that the visits remain approximately constant at 15% every week after week 20. The figure also shows that irrespective of the sign up date, the trend among all dates is approximately same.
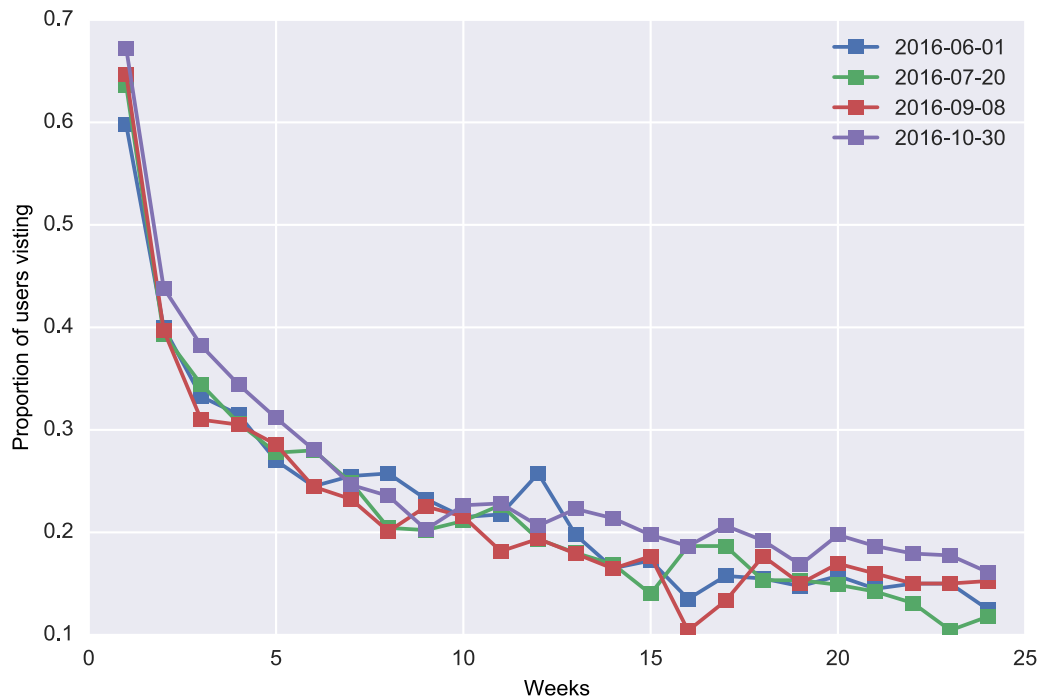


Figure 5 – Plot of proportion of users against number of week for 4 specific dates in the sign up calendar

Another important question which is analysed to visualize retention of users is how many users visit the website for the first within 1 week of signing up, within two weeks of signing up and so on. For example, if the sign up date was 06/01/16, the task is to evaluate how many visit the website for the first time within 06/02 – 06/08, how many visit for the first time within 06/02 – 06/15 and so on. Hence unlike in Figure 4 and Figure 5, the task is to calculate the cumulative of users who visited since the signing up, until 24 weeks have passed.
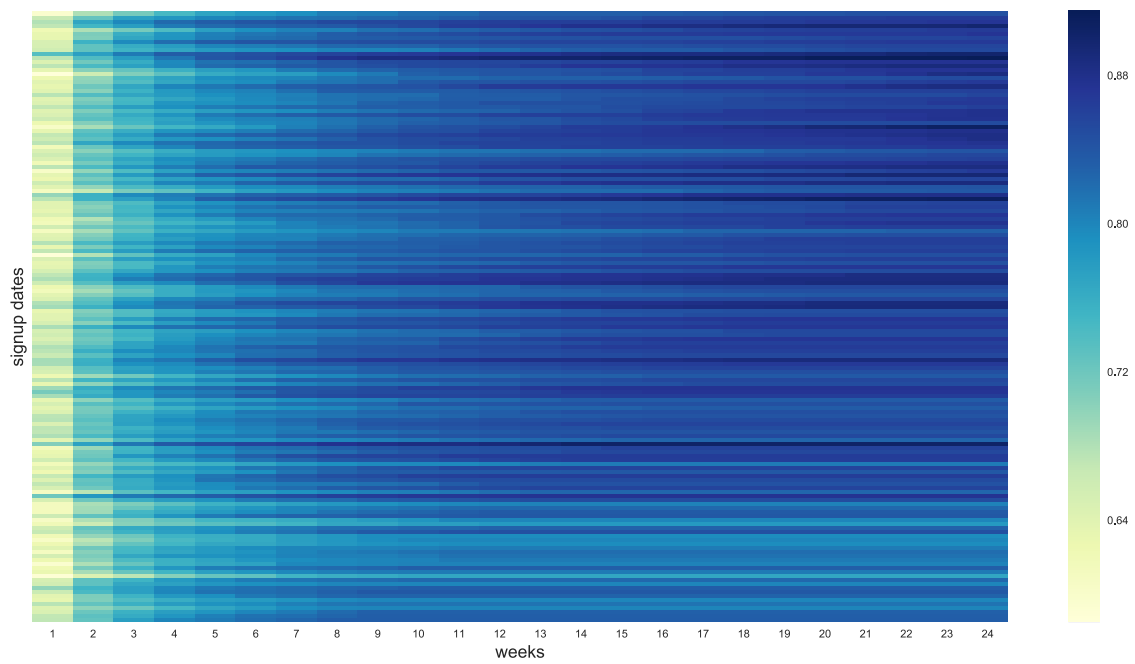
Figure 6 – Heat map of the proportion of visitors visiting for the first time after every n-weeks until 24 weeks.
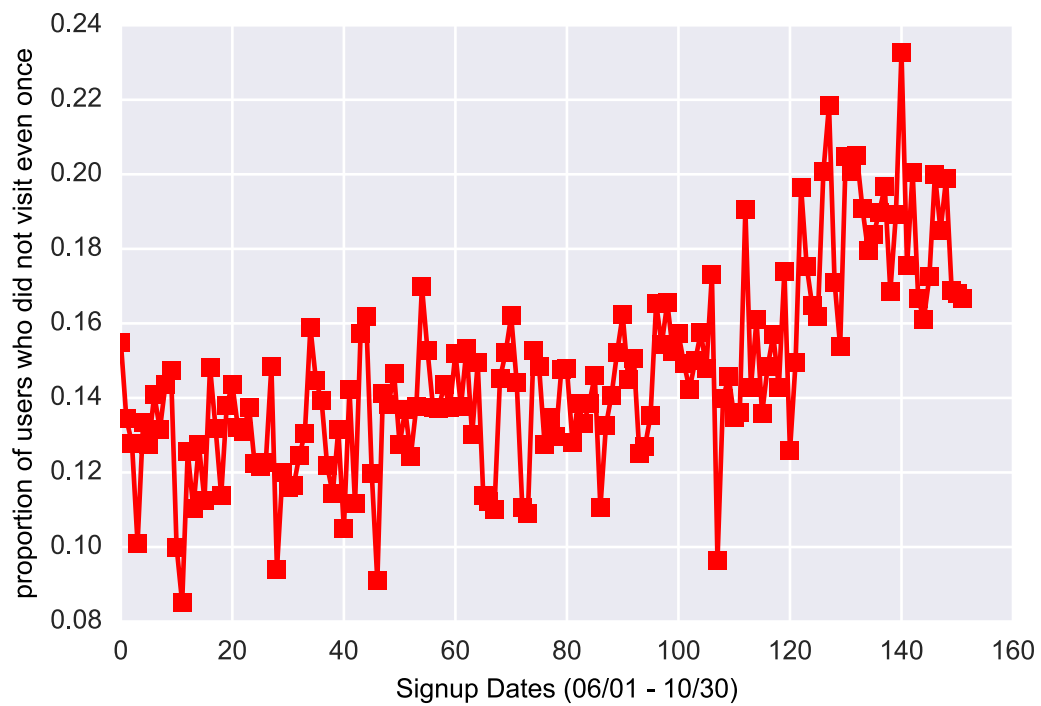


Figure 7 – Plot of the number of users who did not visit the website even once

Figure 6 shows a heat map of the proportion of users visiting within a cumulate time period starting from each sign up date, plotted on the y-axis and for 24 weeks in total, plotted on the x-axis. The heat map shows that total proportion of users who have visited the website at least one after 24 weeks is roughly 15% on average among all sign up dates. This can also be seen from the x-y plot in Figure 7 which shows a plot of the users who did not visit the website even once after signing up on any date from 06/01 – 10/30. On an average, approximately 14.6% of the users do not visit even once, and the proportion of non-visitors tends to increase with future sign up dates, as can be seen from Figure 7.

**Conclusions**

Customer retention and sign up behaviour for customers of an e-commerce website is analysed. Based on the study done, the following are the key conclusions:

(1) The frequency distribution of the number of users signing up on the website based on device type shows that device 1 and 5 are the most popular amongst users of the website. Majority of the users prefer auth_type 'C' as the mode of signing up using their preferred devices.

(2) The proportion of users visiting the website every week after sign up date decreases exponentially. However, the decay comes to a steady state after week 20.

(3) Irrespective of the sign up date, the proportion of users visiting at the end of week 24 if roughly the same across all sign up dates.

(4) Proportion of users who visit the website at least once after sign up, until the time span of 24 weeks is less than 90% on average. Only 60% of users visit the website a week after sign up.

(5) On an average, 14.6% of users do not visit the website even once since signing up on any given date.