seedbox_report

Diana Jovmir

2018-07-15

Intro

The following is the analysis of an experiment ran by Seedbox Technologies, aiming to explore the impact that 2 different modes of cancelling a subscription has on customer retention.

Experimental Setup

The experimental sample consists of subscribers to an internet service that visited the cancelation page of the website for the first time. Subjects were then split randomly into a "Control" and "Test Goups" following a Bernoulli sampling scheme.

The Control group could cancel their subscription using a web form while the Test group had to make a call to Customer service. Subjects were then followed for the remainder of their time as customers of the internet service and their subsequent transactions on the website were recorded in order to quantify impact of cancelling method on customer retention.

The sample distribution is described in Table 1:

Table 1: Sample Distribution

	n	Proportion	SE
Ctrl	44886	0.752	0.001
Test	14835	0.248	0.001

The Data

Transactions recorded were of 3 types: CHECKBACK, REFUND and REBILL. The amount of each transaction was also recorded along with a transaction number. I assumed that transaction numbers increased in time in order to be able to define a time dimension. Since it makes sense that REBILLs would happen monthly, I took my time dimension to be measured in months since visiting the cancelation page. I also found it helpful for analysis to impute a CANCEL transaction to the data that happens in the month following the last recorded transaction with an associated amount of 0. As such, a CANCEL transaction associated with a time stamp of 2 is assumed to have happend during the second month after visiting the cancelation page.

(Note: It's unclear to me if the cancel transaction should have counted in the month previous to a REFUND or REBILL as opposed to the following one. However, the number of these transactios is low enough that it shouldn't affect the calculation of proportions of returning customers.)

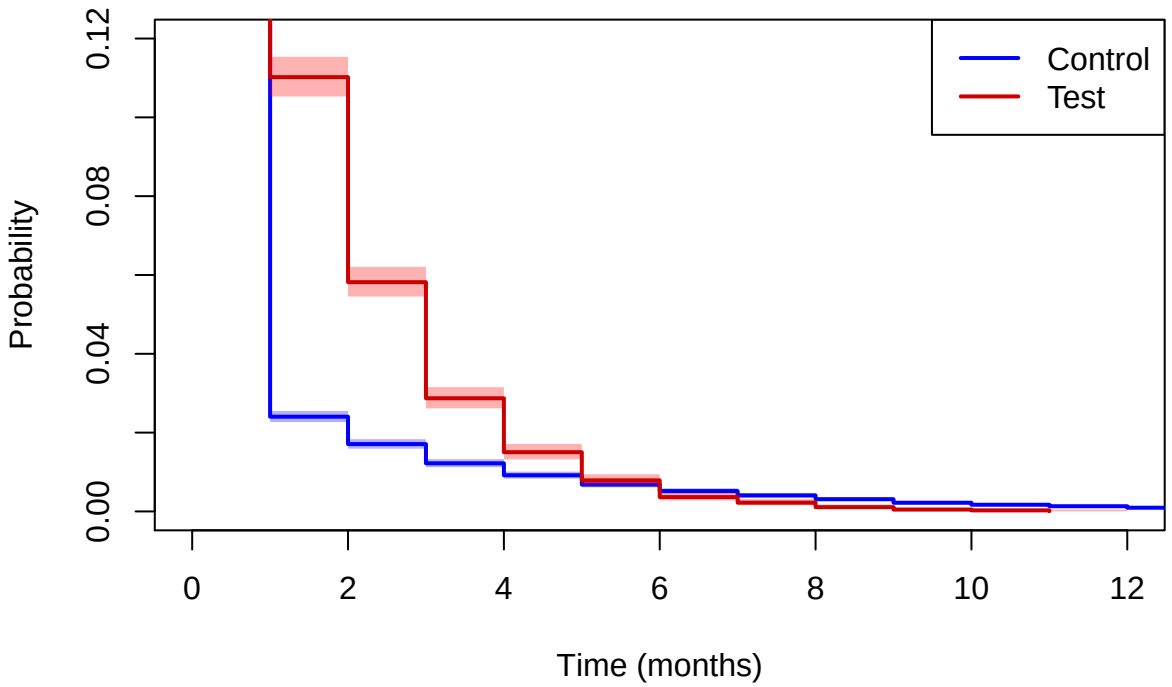
Table 2 shows a glance at the data set.

We found that the great majority of customers had cancelled their subscription 12 months after their initial visit to the cancel page with 43 customers remaining for the up to 22 months. Most customers cancelled (95%) their subcription during their first visit at the cancel page with the remaining ones cancelling gradually over the course of the experiment.

Analysis

To get an idea of the outcomes of our subject, we can take a look at the survival fit in Figure 1, with a "CANCEL" transaction being considered an event and all other transactions a non-event.

Figure 1: Probability of continued subscription



We can clearly see that the survival rates are higher for the Test group during the first 5 months, but resemble those for the Control group after 5 months, and, as expected, a non-parametric test assuming the curves come from the same distribution gives a p-value of approximately 0.

Is a user that must call-in to cancel more likely to generate at least 1 addition REBILL?

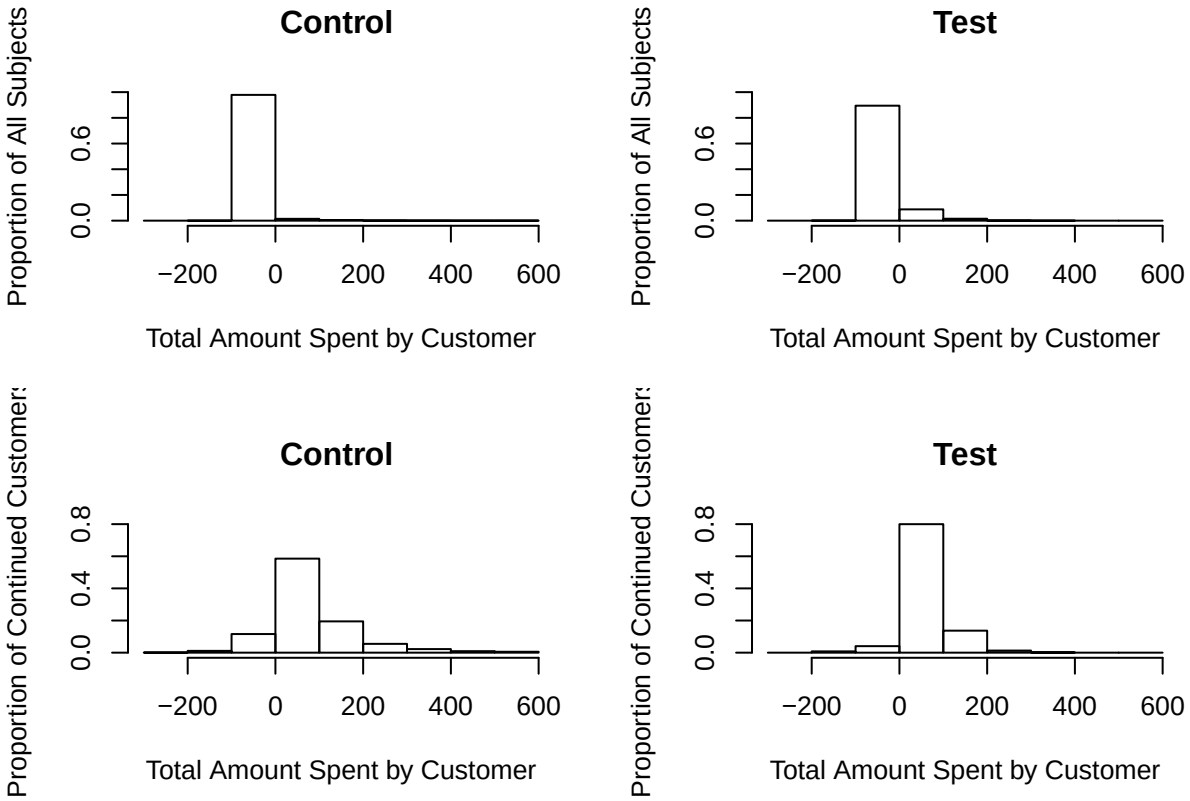
A test of equality of proportion indicates that a Test user has a significantly higher probability of subscribing to the service for at least 1 more month than a Control subjects. The results can be seen in Table 3.

Table 2: Test of Equality of Proportion of Continued Subscribers
After 1 month. H0: Test < Control

	Estimate	p-value
Control	0.021	0
Test	0.105	0

Is a user that must call-in to cancel more likely to generate more revenues?

The distribution of average spending per customer differs greatly when we consider all test subjects, as opposed to only those that didn't cancel in the first month. Let's take a look at Figure 2 to compare the 2 scenarios.



While the average amount billed per customer doesn't follow a normal distribution, we have a large enough sample to make do with a two sample t-test to find out how the distribution of means differ. We can see in Table 4 that Test subjects have a significantly higher spending overall but not when considering only those users that didn't cancel their subscriptions during the first month.

Table 3: t.test of Equality of Mean Spending of Continued Subscribers and All Test Subjects. H0: Test < Control

	Mean-all	Mean-subscribers
Control	2.00	83.34
Test	6.43	58.58
p-values	0.00	1.00

Is a user that must call-in more likely to produce a higher chargeback rate(CHARGEBACKs/REBILLs)?

In fact, users in the Test group appear to have a lower chargeback rate than those in the control group.

(NOTE: I don't really understand if chargebacks are supposed to include refunds. I' haven't included them at this point've decided not to include them.)

Table 4: Test of Equality of Proportion of Chargeback Rates; H0: Control > Test

	Estimate	p-value
Test	0.018	0.997
Control	0.028	0.997

Discussion

It seems clear from the above analysis that customers that must call in to cancel their subscription remain subscribed for longer than customers that can do so via a webform. They also generate more revenue on average, but that difference is mostly due to the initial delay in cancelling. Surprisingly, they also seem to generate a lower chargeback rate than the subjects in the control group.

However, we must be careful before deciding that it's more profitable overall to force people to call in to cancel their subscriptions. We haven't yet considered that former customers are also able to re-subscibe to a service. People that are un-motivated to stay might also be uncommited about quitting; however, having had a more difficult experience quitting the first time, subjects in the Test group might be reluctant to resubscribe. Data on return customers might be more difficult to gather, but it would be important to consider.