Time to User Churn

Julia Kirkpatrick • 2019-12-06

Research Population and Assumptions

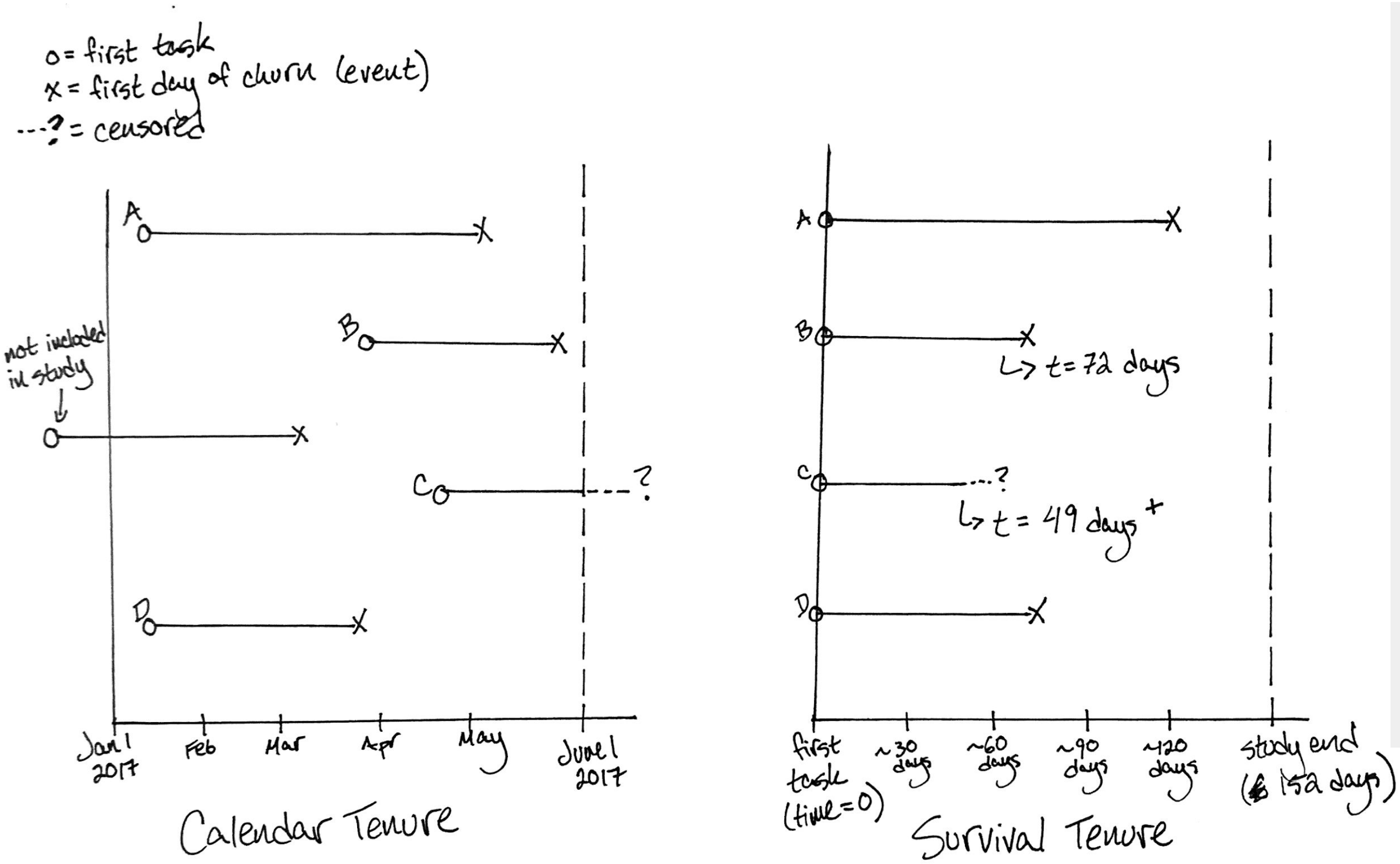
- All users are observed to have executed at least one task between Jan 1, 2017 and June 1, 2017.
- Users must have observations for at least 29 days - meaning they executed a task at least once and were observed for the full 28-day period after.
  - Our outcome measure is “first churn,” which is possible beginning at observation day 29.
- Data is structured and modeled to using survival analysis.
- Goal: describe attributes that contribute to increasing or decreasing the likelihood of churn (for the first time, not considering users who churn and subsequently become active again).

Method: Time to Event (Survival) Analysis

- Used to analyze data in which **time until the event**, in this case user first churn, is of interest. The response is generally referred to as survival time, retention time, or tenure.

- Allows for observation **censoring**. Censored users are those who remain active during the entire study period.
    - Status = churned: We know the exact time-until-churn for users who have churned in our study window: (first day of churn) - (day of first task) = tenure.
    - Status = censored: For users who have not yet churned, we only know that their tenure is at least equal to their tenure at the end of the study period, June 1 2017. The tenure for still-active users is (June 1 2017) - (day of first task) = tenure**+**.

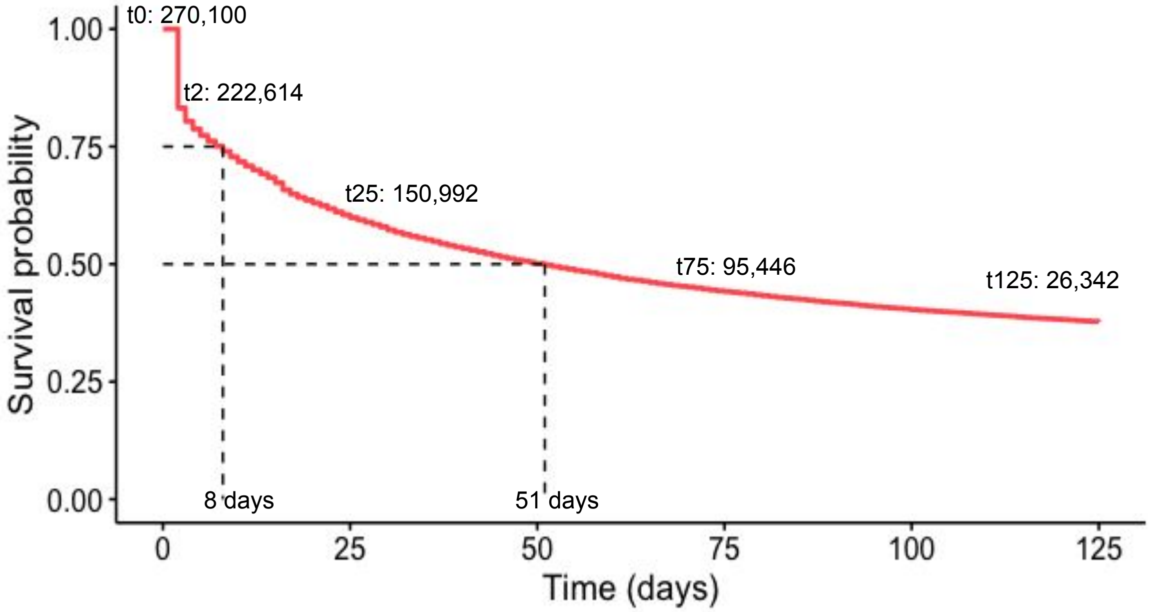
Method: Time to Event (Survival) Analysis



This visualization shows how users are normalized for the study. Each user's first-task-day is assigned as "time 0." This allows us to compare each user to each other, regardless of the calendar date on which they executed their first task.

Additionally, if a user is still active on the last day of our study, we are able to include their tenure and other information through censoring, denoted by "+".

Population Retention Curve

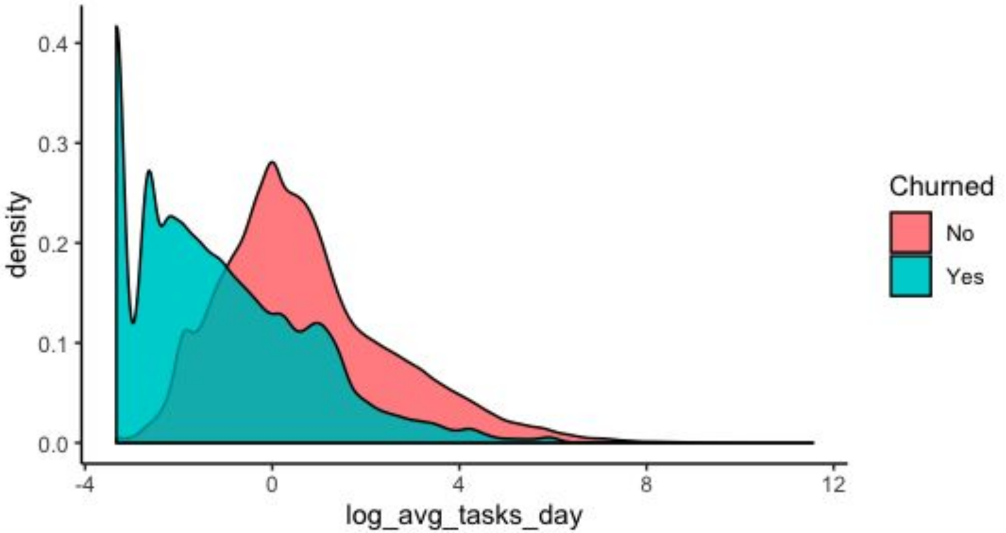


Time 0 corresponds with user's 28th day of observation post-first-task. All users are active at time 0, so survival probability is 1. User churn patterns shape the curve - as more users churn for the first time, the curve trends downward (probability of 'survival' decreases with time).

Information about users who never churn is also included in the curve. Active users do not affect the shape of the curve but increase the precision of the curve's estimates. The active user data ('censored observations') is primarily used to estimate feature effects in models.

We see a large drop-off from time 0 to time 2: this shows that **about 20% of new users execute tasks on their first two days and then never again**. A quarter of users ultimately churn after only 8 days of task-activity, and half of users after 51 days.

Potential Features Affecting Churn:
Average Tasks per Day



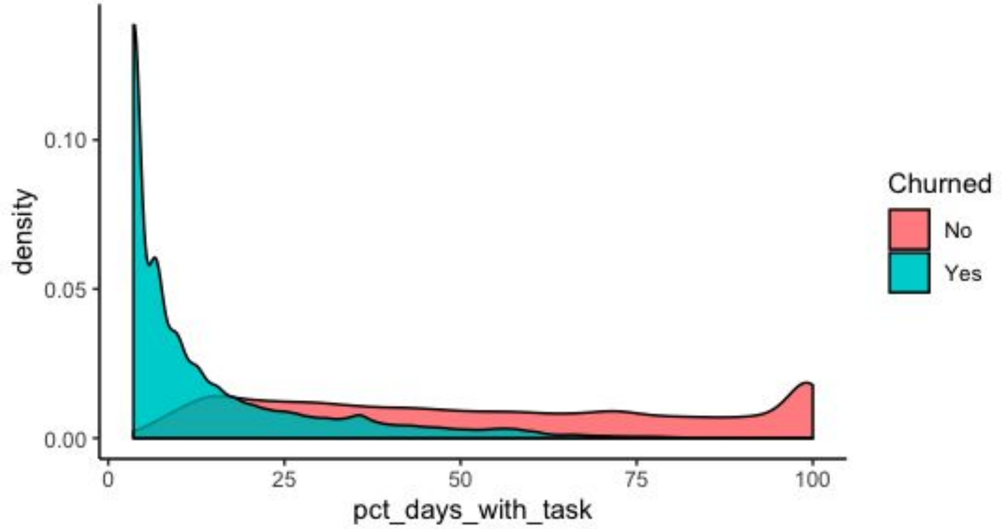
Group	n	Min	Median	Mean	Max	Std Dev
Population	270,100	0.04	0.63	14.8	104,462	331.8
Still active	115,017	0.04	1.52	28.5	104,462	505.4
Churned	155,083	0.04	0.23	4.65	8,816	45.06

We calculate average tasks per day as the average number of tasks the user executes during their first-active period.

Descriptive statistics by group are displayed below the graph. Because the data is so skewed, we use log(average tasks per day) as our feature which makes it easier to see that, unsurprisingly, the still-active population executes more tasks/day than the churned population did while they were active.

Potential Features Affecting Churn:

Percent of Days with Task



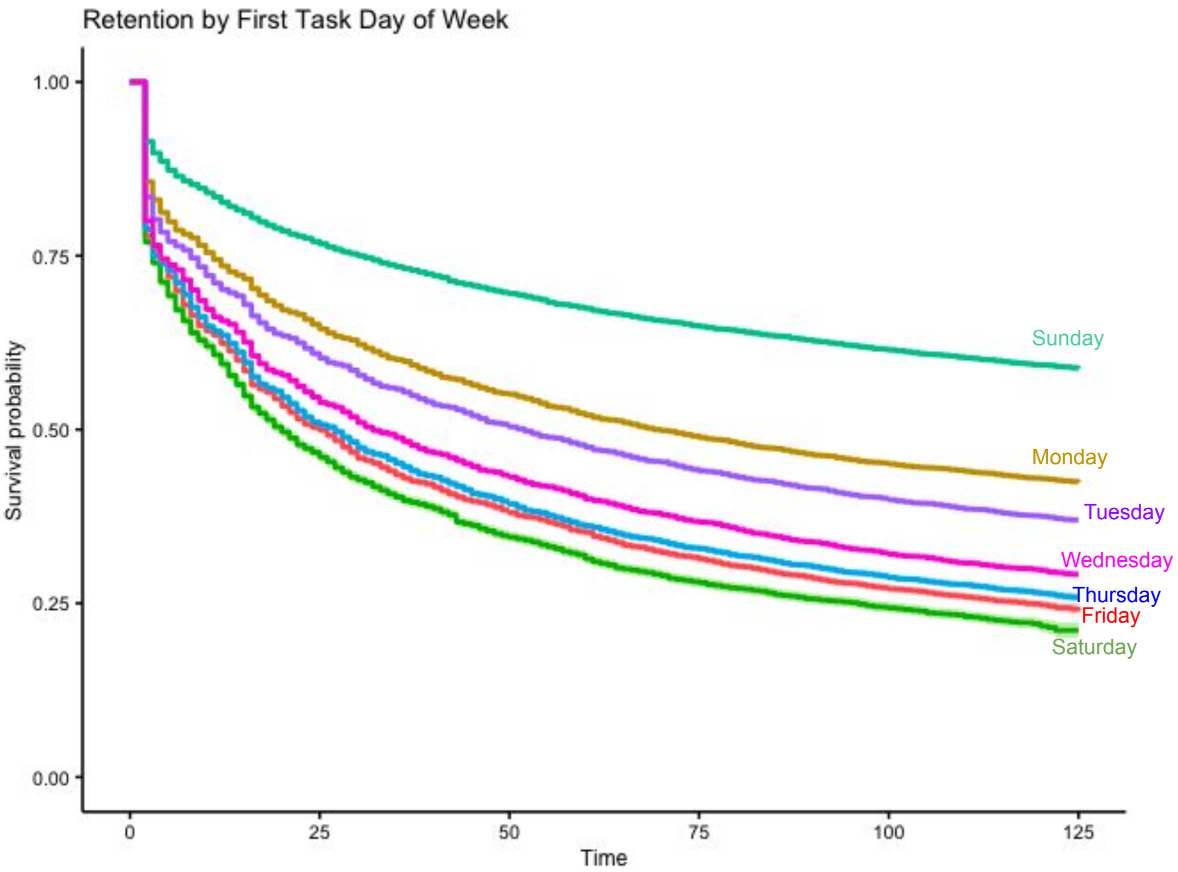
Group	n	Min	Median	Mean	Max	Std Dev
Population	270,100	3.6%	19.1%	30.8%	100%	29.4%
Still active	115,017	3.6%	49.7%	52.8%	100%	29.7%
Churned	155,083	3.6%	7.9%	14.6%	82.1%	15.0%

We calculate the user's percent of days with task as the number of days the user executes a task out of the days they are considered active.

Descriptive statistics by group are displayed below the graph. The still active group, pink, executes tasks on a higher percentage of their active days than the churned group, blue. The blue spike on the left reaffirms that many of the churned users executed tasks on only the first day or two out of the 28 days they are considered active.

Potential Features Affecting Churn:

First Task Day of Week



We are also able to group the users by the day of the week on which they executed their first task. The survival curve shows that users who execute their first task earlier in the week are much less likely to churn than users who execute their first task later in the week. Users executing their first task on Sunday have a much higher retention rate than those who first execute a task on Saturday.

Users who execute their first task on Sunday may be inherently different than those who choose to execute their first task later in the week, so a formal experiment is needed to test if first task day of week is a significant driver of retention or if it's just an interesting coincidence.

Features Affecting Churn

Using a random forest model we are able to assess which of these features have the highest impact on user first churn.

1. **Percent of days with task** was found to be the most important feature in our model in predicting first churn. Based on our exploratory analysis, users executing tasks on a higher percentage of their days while active are less likely to churn.

2. Average tasks per day is the second most important feature out of the three that were tested, but has a small relative importance compared to percent of days with task.

3. The weekday of first task is the least important predictor of first churn.

Insights, Recommendations, Next Steps

Given dates and quantities of user task activity allowed us to determine that **users who execute tasks more frequently** are less likely to churn.

The average number of tasks per day had a much smaller impact - another modeling approach could be used to determine the direct effect that average number of tasks per day has on churn likelihood, but was not successful using this model.

The weekday of first task, while it looked promising, did not have much of any importance in the model.

A next step for exploring this data in isolation would be to create an outcome, "**churn within 1 month**" to build another machine learning model that would predict more than just time to **first churn** as was explored here.

Integrating more user demographic information, such as account type, operating system, or method of acquisition, etc. would further refine the analysis of any model.