## **Introduction**

I am sure you have heard the phrase - retention is the key! Indeed, strong retention of users over time is a great indicator of product-market fit, something all start-ups intend to achieve.

To improve retention, it is imperative to understand leading indicators of churn and then improve the product accordingly.

So in this post, we look at a fictious SaaS company’s 3 months data (generated randomly) to understand how to do churn analysis.

## **Objective**

Our objective is to suggest product changes needed to reduce churn

## **Conceptual Overview**

In subscription businesses retention is the most important metric. Churn is like a hole in the boat, no matter how many customers a start-up acquires, it will not thrive without sticky features that retain users.

As a Product Manager, it is important to look deeper and analyse big numbers with scrutiny.

There are two kinds of churn -

1. **Customer Churn** – Defined simply as number of customers churned compared to total number of customers. Focus is on retaining high percentage of customers for a longer period
2. **Revenue Churn** – Defined as revenue churned compared to overall revenue. It is a measure of lost revenue. It is critical to understand why this metric is important compared to customer churn. Not all customers are equal so if a start-up is losing high value customers the impact on revenue churn will be much higher than if it is losing low value customers

In this post, we will look at the following metrics for TakeZero (our fictitious company) –

* **Monthly churn rate** – This is a basic month-on-month (MOM) customer churn.
* **Churn rates by customer segmentation** – Churn rates are reported with some customer segmentation. For instance, customers from different acquisition channels, region, subscription plan lead to different churn rates. Therefore, it is critical to understand what kind of customers have high churn rates
* **Churn rates by customer behaviour** – A user takes a series of steps after subscribing. It is important to find features which are sticky and drive retention in short term and long term. Drive usage of that feature to other customers as well
* **Cohort Analysis** – Cohorts are group of users sharing a common characteristic such as acquired in same month, acquisition channel etc. Over time as the product is getting improved, one needs to ensure that younger cohorts are showing better retention
* **Churn Prediction** – Once we identify predictor variables which are significant to predict churn, we can use any of Machine Learning techniques such as Logistic Regression, Decision Tree to predict customer churn

**Note:** This post will focus on analysing customer churn and not revenue churn (we will do that in another post)

## **Company Details & Data**

TakeZero is a fictitious SaaS company (like no-code platforms) which allows teams to collaborate and build basic no code web and mobile apps through its web application. We have data points for Q4 2019 i.e. October, November, December. Data and description of all the columns can be found here -

File - <https://drive.google.com/file/d/1okQlxFfRKLR8Sakk0QjNtqA8frGCqZH2/view?usp=sharing>

## **Data Overview**

Here is a snapshot of the first 2 rows of the data –

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Purchase month** | **purchase plan - # of seats** | **Team members added** | **Customer Region** | **avg number of web app sessions ran per month** | **avg number of mobile app sessions ran per month** | **Month churned** |
| 1 | October | 2 | 2 | US | 30 | 16 |  |
| 2 | October | 10 | 4 | US | 68 | 12 | November |

Column descriptions are the following -

|  |  |
| --- | --- |
| **Column Name** | **Column Description** |
| ID | Serial number of the customer |
| Purchase month | Month in which subscription plan was purchased, we have 3 months data: October, November, and December |
| Purchase plan - # of seats | Customers can buy *individual plan (1 seat)* or *team plans ( 2, 5 or 10 seats)*. Seats refer to maximum number of users that can be onboarded on the platform by the user |
| Team members added | Actual number of users added by the customer (capped to the number of seats in the bought subscription plan) |
| Web app sessions per month | Avg. number of monthly sessions building no code web apps by the customer |
| Mobile app sessions per month | Avg. number of monthly sessions building no code mobile by the customer |
| Customer Region | Continent to which the customer belongs. |
| Month churned | Month in which customer cancelled his plan. If empty, it means user did not churn. Churn data is available for November, December, & January |

## **Implementation**

**Monthly Churn Rate -** Total customers churned in that period / Avg. number of customers in that period

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **new customers** | **start of month** | **churn** | **end of month** | **avg. number of customers** | **Monthly churn rate** |
| **October** | 300 | 0 | 0 | 300 |  |  |
| **November** | 401 | 701 | 60 | 641 | 671 | 2% |
| **December** | 499 | 1140 | 130 | 1010 | 1075 | 3% |
| **January** | 0 | 1010 | 170 | 840 | 925 | 5% |

According to calculations, our monthly churn rate is around 2-3% which is on the much lower end of avg. churn rates in early stage SaaS companies.

## **Churn rates by user segment**

We have two data points on customer segmentation – region to which they belong, and purchase plan (# of seats they purchase). Let us look at churn rates by region and purchase plan to see if some kind customer segment is churning.

**Churn rate by region** – Observations,

* 69% of customers from Asia churn within 3 months. This is much higher than 16-18% churn of European and US customers in the same time period.

|  |  |  |  |
| --- | --- | --- | --- |
| **Region** | **New customers** | **Churned** | **Churn%** |
| **Asia** | **284** | **197** | 69% |
| **Europe** | **334** | **55** | 16% |
| **US** | **582** | **108** | 19% |
| **Grand Total** | **1200** | **360** | 30% |

**Churn rate by purchase plan** – 2 important observations here,

* Customers who buy individual plans (1 seat) churn at much higher rate (47%) than customers who buy team plans (avg around 22-23%)
* Customers who buy team plans and add less than 60% of the team members in 1st 3 months churn at a much higher rate. For example, 100% of the customers who bought 10 seat plan but added less than 5 members, churned within first 3 months

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Purchase Plan-# of seats bought** | **# of team members added** | **# of customers added** | **Churn** | **Churn%** |
| 1 | 1 | 358 | 169 | 47% |
| **Total** | **358** | **169** | **47%** |
| 2 | 1 | 57 | 50 | 88% |
| 2 | 222 | 11 | 5% |
| **Total** | **279** | **61** | **22%** |
| 5 | 1 | 18 | 17 | 94% |
| 2 | 34 | 34 | 100% |
| 3 | 7 | 0 | 0% |
| 4 | 96 | 5 | 5% |
| 5 | 120 | 7 | 6% |
| **Total** | **275** | **63** | **23%** |
| 10 | 1 | 2 | 2 | 100% |
| 2 | 16 | 16 | 100% |
| 3 | 22 | 22 | 100% |
| 4 | 20 | 20 | 100% |
| 7 | 59 | 1 | 2% |
| 8 | 53 | 2 | 4% |
| 9 | 63 | 1 | 2% |
| 10 | 53 | 3 | 6% |
| **Total** | **288** | **67** | **23%** |

## **Churn rates by customer behaviour**

**Churn rates by team members added**

**Churn rates by avg number of web app development sessions per month –** Observations,

* We know on an avg how many web app development sessions user does per month. We will normalize this metric to remove the effect of the number of seats team plan has i.e. avg number of web app development sessions per month per added user
* We don’t see much difference in the web app dev sessions done per month between churned and retained user. Therefore, there is no significant association between this metric and churn

|  |  |
| --- | --- |
| **Churn** | **avg web app dev sessions ran per month per added user** |
| 0 | 10.21071429 |
| 1 | 9.913888889 |
| **Grand Total** | **10.12166667** |

**Churn rates by avg number of web app development sessions per month –** Observations,

* We know on an avg how many mobile app development sessions user does per month. We will normalize this metric to remove the effect of the number of seats team plan has i.e. avg number of mobile app development sessions per month per added user
* We see a big difference between averages of churn and retained users. It seems retained users are doing much more mobile app dev sessions per month than churned out users. It shows how mobile app dev on the TakeZero platform is a sticky feature for short term retention (This is an important insight!!)

|  |  |
| --- | --- |
| **Churn** | **avg mobile app dev sessions ran per month per added user** |
| 0 | 8.997619048 |
| 1 | 3.083333333 |
| **Grand Total** | **7.223333333** |

## **Cohort Analysis**

* Cohort are groups of users sharing by a common characteristic. For instance, cohorts by acquisition date, daily, weekly, monthly cohorts etc
* Once can compare metrics such as engagement, retention, or conversion across cohorts
* Through cohort analysis, we want to understand if our product changes, campaigns or other investments are leading to positive retention among younger cohorts

For improving retention, there are two broadly types of cohorts

* Acquisition Cohorts - Group user by their acquisition day, week or month and measure retention. This helps to understand when we are losing customers and when does churn stablise
* Behavioural Cohorts – Group users by specific behaviours they have or haven’t take on the product within a given timeframe. (For example – inviting a team member in 1st month, Following 5 users in 1st week etc). Then we can track retention across such cohorts and find out which are sticky features or which features lead to +ve retention

**Acquisition Cohorts**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **% of customers retained in lifetime month** | | | | |
|  | **0** | **1** | **2** | **3** |
| **October** | 100% | 80% | 77% | 75% |
| **November** | 100% | 70% | 67% |  |
| **December** | 100% | 69% |  |  |
| **Avg** | 100% | 73% | 72% | 75% |



Observations,

* If we look at cohorts of users acquired in month October, Nov, December – we can understand that we lose majority of customers in 1st month and churn stabilises around 2nd-3rd month.

**Behavioural Cohorts**

Based on what we saw in “Churn rate by Customer Behaviour” section, we can form two hypotheses

Hypothesis 1: We believe team plan subscribers who fill more than 60% of the bought seats within first 3 months retain better

* Around 20% of the customers who fill < 60% of the bought seats in team plan churn within 1st month. Whereas, 0% churn in customers who fill > 60% seats bought in team plan
* Therefore, it is essential to improve team member invitation feature to ensure customers add more team members and fill as many seats bought in first month after onboarding

Hypothesis 2: We believe customers who do more mobile app dev sessions per month per user retain better

* Around 25% of the customers whose team does < 5 mobile app dev sessions per month per user churn within 1st month. Whereas, 0% churn in customers whose team does more than 5 mobile app dev sessions per month per user
* Therefore, mobile app dev seems to be a sticky feature for short term retention. Therefore, it is important to improve its discovery, usage, flow to ensure 100% of the customers start using this feature in 1st month after acquisition

## **Churn Prediction**

We can use historical data to build a churn prediction model which attaches churn risk score to new customers. Our exploratory analysis till now has revealed following predictors that could be significant in explaining churn

* Customer region
* # of seats bought
* Users added / # of seats bought
* Mobile app dev sessions per month per user

We can use any of the tools such as XLMiner (a paid Excel plugin) or Scikit-learn to perform classification machine learning algorithms to create a model. Simpler techniques like Logistic Regression or Decision Trees should product good results.

I have used Logistic Regression on XLMiner to do data partition (training (70%) & validation(30%)) and the resultant classification model provides following evaluation metrics on validation data

|  |  |
| --- | --- |
| **Metrics** | |
| **Metric** | **Value** |
| **Accuracy (#correct)** | 237 |
| **Accuracy (%correct)** | 98.75 |
| **Specificity** | 0.9878788 |
| **Sensitivity (Recall)** | 0.9866667 |
| **Precision** | 0.9736842 |
| **F1 score** | 0.9801325 |
| **Success Class** | 1 |
| **Success Probability** | 0.5 |

Since we are predicting churn we are optimising for low False -ves i.e. our model should not predict large number of customers who will churn as not churn. Therefore, Recall metric is primary model evaluation metric and a value of 98% on validation data is a good outcome.

Let us look at what predictors are significant at 95% confident interval (i.e. p-value < 0.05)

|  |  |
| --- | --- |
| Predictor | P-value |
| Intercept | **4.41E-6** |
| Number of user seats bought | **0.113** |
| User added/seats bought | **0.018** |
| Mobile sessions per month per added user | **5.83E-09** |
| Customer Region\_Asia | **0.044** |
| Customer Region\_Europe | **N/A** |
| Customer Region\_North America | **0.196** |

As evident from the table above, mobile sessions per month per added user, and user added/seats bought are two metrics that are highly associated with churn. Customer region – Asia, Purchase plan – number of seats bought are weak predictors of churn.