## **Why**

* To develop a framework on how to approach churn analysis problems as a Product Manager
* To improve skills in Excel
* To improve writing skills and whether I am able to convey my thought process
* Finally, currently focus is on improving metrics and analysis skills

What

* Take example of a fictious company TakeZero and dataset to do following analysis
  + Monthly churn rate
  + Acquisition Cohorts & retention curves
  + Behavioural Cohorts & Retention curves
  + Churn Prediction

How

* Using Excel & Pivot Tables

## **Introduction**

For a product, churn is like a hole in the boat, no matter how many users you acquire product will not thrive without strong retention of the acquired users. In fact, venture capitalists look at churn thoroughly as it is a strong indicator of whether you have a good product/market fit.

Churn analysis helps uncover patterns, product features, and behaviours that lead to user retention. Therefore, in this post we will learn how to do churn analysis by taking example of a fictious SaaS company, TakeZero, which allows developers to build no code web and mobile apps.

## **Objective**

Our objective is to build an understanding of what metrics we need to look at to understand churn and build hypothesis around how to reduce it.

## **Conceptual Overview** There are two kinds of churn -

1. **Customer Churn** – Defined as ratio of number of customers churned compared to total number of customers. Here, the focus is on retaining high percentage of customers for a longer period
2. **Revenue Churn** – Defined as ratio of revenue lost compared to overall revenue. Apart from customer churn we need to look at revenue churn because not all customers are equal. Hence, if a start-up is losing low churn but is mostly losing high value customers the impact on revenue will be much higher compared to if it was losing even high number of low value customers. Here, the focus is on retaining high value customers for a longer period or up-sell and cross-sell to decrease revenue churn

**Note:** For the sake of brevity this post will focus only on customer churn. We will focus on revenue churn in the next post.

We will look at the following metrics for TakeZero (our fictitious SaaS 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 have 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 for the product. It is important to find features which are sticky and drive retention in short term and long term.
* **Cohort Analysis** – Cohorts are group of users sharing a common characteristic such as acquisition month, acquisition channel etc. Over time as the product is getting improved, one needs to ensure that younger cohorts are showing better retention or lower churn
* **Churn Risk Score for every customer** – Through descriptive analysis we identify predictor variables which could be significant to predict churn and then we can use any of the ML techniques such as Logistic Regression, Decision Trees to predict customer churn and assign each new user a churn risk score. It helps us develop focus on high risk churn

## **Company Overview**

TakeZero is a fictitious SaaS company which provides an online platform to build no-code web and mobile apps. It works on a subscription model where customers can buy individual plans or team plans with pre-defined seats.

We have 2019 Q4 data. We have data on customers acquired in the month – October, November & December and which of these users churned by January month (i.e. churn within first 3 months).

## **Data Overview**

Here is a snapshot of the first 5 rows of the data –

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Purchase month** | **# of seats bought** | **Team members added** | **Customer Region** | **avg number of web app development sessions ran per month** | **avg number of mobile app development sessions ran per month** | **Month churned** |
| 1 | October | 2 | 2 | US | 30 | 16 |  |
| 2 | October | 10 | 4 | US | 68 | 12 | November |
| 3 | October | 10 | 4 | US | 40 | 4 | November |
| 4 | October | 2 | 1 | Europe | 16 | 20 |  |
| 5 | October | 10 | 2 | Asia | 12 | 8 | November |

Following are the column descriptions -

* **ID** – Serial number of the customer
* **Purchase Month** – Month in which subscription is bought
* **# of seats bought** – Customer can buy individual plan (1 seat) or team plans (2,5, or 10 seats). Seats refer to the maximum number of team members the customers can add on the platform
* **Team members added** – How many team members were added by the customer on the purchased plan
* **Customer Region** – Continent to which the customer belongs
* **Avg web app development sessions per month** - This number captures on an average how many web app development sessions did a customer (inclusive of all her team members) do in a month
* **Avg mobile app development sessions per month** - This number captures on an average how many mobile app development sessions did a customer (inclusive of all her team members) do in a month
* **Month Churned** – Month in which customer cancelled her plan. We have churn data for Nov, Dec, Jan. If empty, it means customer did NOT churn till end of the month January 2020

All the data is available in Excel format here -

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

## **Data Preparation**

I created following new columns by manipulating existing data –

* **Team members added/# of seats bought** – This ratio helps us determine of how many of team members are onboarded against the seats bought by the customer. Having more team members onboarded is a leading indicator engagement of the user and its team on the platform. On the vice versa, very few members onboarded means customer may not be finding value in the platform
* **Avg # of web app dev sessions per month per team member -** Since different customers buy different team plans to compare engagement with web app development feature across customers we normalize by calculating avg number of sessions done in a month per user
* **Avg # of mobile app dev sessions per month per team member -** Since different customers buy different team plans to compare engagement with mobile app development feature across customers, we normalize by calculating avg number of sessions done in a month per user
* **Churn Bool –** A Boolean variable with value 1 if user churned and 0 otherwise

## **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

|  |  |
| --- | --- |
| **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.

## **Key Learnings**

1. Customers who buy individual plans churn more
2. Customers from Asia region churn more
3. Majority of users who buy individual plans come from Asia
4. Churn stabilises by 2nd-3rd month
5. Customers who fill more than 60% of the seats bought within first 1st month of onboarding churn less.
6. Customers who use more do more than 5 mobile app dev sessions per month per user churn less
7. Therefore, features such as adding members to the team and mobile app dev sessions are sticky features leading to better short-term retention of the onboarded users

## **Recommendations**

It is important to understand that data analysis tells us “what” is happening but to understand “why” we should supplement it with qualitative user research.

1. Talk to customers who purchase individual plans and are leaving platform within first 3 month of on-boarding. Understand why they are leaving. We can form several hypotheses here –
   * It could be happy churn i.e. customers are satisfied with the product, but they still churn because their job is completed, and they don’t think they should pay for subscription anymore
   * It could be that our product is not suited to the needs of customer persona who buys individual purchase plans. I would want to go back to the user research in revalidate our assumptions and hypotheses there
   * Are they even the most profitable customers that we want to go after?
2. We have identified +ve correlation between features such as mobile app dev sessions, onboarding more team members against seats bought in the purchase plan and short-term retention.
   * We should try to look at more data from previous quarters and see if same pattern follows
   * Talk to customers who are using these features and understand what works for them
   * We should establish causal effect of these features to short-term retention through A/B experiments
   * If the above identified features indeed make the product sticky then we should focus on improving their discovery and usage through improvements in onboarding flow and nudge users towards these features early in their lifetime
3. Use the churn prediction model to assign churn risk score to every customer. Prepare a marketing plan through various channels such as mailers, push notifications etc to keep those users engaged on the platform

Retention is the most important metric in subscription businesses because a leaky product will never be able to build a critical mass of hooked users. As a Product Manager, it is important to look deeper then just the obvious metrics such as monthly churn rate and analyse big numbers with scrutiny. Also, it is critical to remember that data should inform us in building and validating better hypotheses, and what areas to focus on our qualitative research.