INDEX

[Introduction](#_Toc38700257)

[Objective](#_Toc38700258)

[Conceptual Overview](#_Toc38700259)

[Company Overview](#_Toc38700260)

[Data Overview](#_Toc38700261)

[Data Files](#_Toc38700262)

[Data Pre-processing](#_Toc38700263)

[Implementation](#_Toc38700264)

[Churn rates by user segments](#_Toc38700265)

[Churn rates by customer behaviour](#_Toc38700266)

[Cohort Analysis and Retention Curves](#_Toc38700267)

[Churn Prediction –Churn Risk Score for customers](#_Toc38700268)

[Recommendations to improve short term retention of TakeZero](#_Toc38700269)

## **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. Venture Capitalists have a thorough look at churn 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 an example of a fictitious 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 hypotheses around how to reduce it.

## **Conceptual Overview**

There are two kinds of churn -

* **Customer Churn** – Defined as the ratio of the number of customers churned compared to the total number of customers. Here, the focus is on retaining a high percentage of customers for a longer period
* **Revenue Churn** – Defined as the 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 has low churn but it is mostly losing high-value customers the impact on revenue will be much higher compared to if it was losing mostly low-value customers. Here, the focus is on retaining high-value customers for a longer period

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

In this post, we will look at the following metrics –

* **Monthly churn rate** – This is a basic month-on-month (MOM) customer churn.
* **Churn rates by user segments** – Churn rates are reported with some customer segmentation. For instance, customers from different acquisition channels, regions, subscription plans 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 to the product. It is important to find features which are sticky and drive retention in short-term and long-term.
* **Cohort Analysis & Retention Curves** – Cohorts are a 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
* **Churn Prediction Model** – Through descriptive analysis, we identify predictor variables which could be significant for churn prediction and use Logistic Regression to build a simple churn prediction model

## **Company Overview**

TakeZero is a fictitious SaaS company that 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

## **Data Files**

*The only way to learn data analysis is to do it!!*  To get your hands dirty download the data files and get started - [Churn Analysis Files](https://drive.google.com/open?id=1T59xFu-0tUlO-SaDqfUCO6bUJUqI24Gl). There are separate question and answer Excel files in case you want to practice before looking at the solution

## **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 # of web app development sessions ran per month** | **avg # 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 # of 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 # of 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 the customer cancelled her plan. We have churn data for Nov, Dec, Jan. If empty, it means the customer did NOT churn till end of the January 2020

## **Data Pre-processing**

I created following new columns by manipulating existing data –

* **Team members added/# of seats bought** – This ratio helps us determine how many team members are onboarded against the seats bought by the customer. Having more team members onboarded is a leading indicator of engagement of the user and its team on the platform. On the vice versa, very few members onboarded means the 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 of web app development feature across customers we normalize it by calculating a per team member ratio
* **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 it by calculating a per team member ratio
* **Churn Bool –** A Boolean variable with value 1 if user churned and 0 otherwise (You can find all these fields in the solution Excel file)

## **Implementation**

* **Monthly Churn Rate -** Total customers churned in the month / Avg. number of customers in month

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **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% |

**Observations** - Our monthly churn rate is around 2-3% (5% is anomalously higher because January user acquisition data is not available). For a SaaS company especially early to mid-stage this is at a much lower end of the spectrum.

## **Churn rates by user segments**

* **Churn rate by region** –

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

**Observations** - 69% of customers from Asia churn within 3 months. This is much higher than 16-19% churn of European and US customers in the same time period.

* **Churn rate by 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%** |

**Observations** –

* Customers who buy individual plans (1 seat) churn at a much higher rate (47%) than customers who buy team plans (avg around 22-23%)
* Customers who add only a small number of team members compared to seats bought tend to churn much more than customers who fill all the seats bought. For example – As seen in the table, customers who bought 10 seat team plan and added less than 5 members in the first 3 months, churn at 100%

## **Churn rates by customer behaviour**

* **avg # of web app development sessions per month per team member**

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

**Observations –** There is NO significant difference between the avg # of web app dev sessions done per month between a churned and retained customer. Therefore, according to our sample data, there is no significant association between this metric and churn

* **avg # of mobile app development sessions per month per team member**

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

**Observations** –

* Aha! there is a significant difference between avg # of mobile app dev sessions ran per month between a churned and retained user
* We can build and test a hypothesis that a greater number of mobile app dev sessions per month per user leads to higher short-term retention (a plausible *sticky feature!*)

## **Cohort Analysis and Retention Curves**

**Cohorts** are groups of users sharing by a common characteristic. For instance, cohorts by acquisition date, daily, weekly, monthly cohorts etc. One 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

To improve retention, we can use two 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 stabilise
* **Behavioural Cohorts** – Group users by specific behaviours they have or have not taken on the product within a given timeframe. (For example – inviting a team member in 1st month, following 5 users in 1st week etc). We can track retention across such cohorts and find out which features are sticky

**Acquisition Cohort**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **% of customers retained in lifetime months** | | | | |
| **Cohort** | **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 – it is evident that we lose most customers in 1st month and churn stabilises around 2nd-3rd month.

**Behavioural Cohorts**

* Cohort of customers who fill more than 60% of the bought seats within first 3 months vs those who do not



**Observations -** Around 20% of the customers who fill < 60% of the bought seats in team plan churn within 1st month compared to 0% churn in customers who fill > 60% seats in same timeframe

* Cohort of customers who do > 5 mobile app dev sessions per month per team member vs those who do not



**Observations** - Around 25% of the customers whose team does <= 5 mobile app dev sessions per month per team member churn within 1st month compared to 0% churn in customers whose team does > 5 mobile app dev sessions per month per team member

## **Churn Prediction**

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

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

I have used XLMiner (a paid Excel plugin) to run Logistic Regression model but same can be done in Python using Scikit-learn library. Logistic Regression model provides following model evaluation metrics on our 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 |

**Observation** - Since we want to optimise for False -ves, a 98.6% value of Recall 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 seats bought | 0.113 |
| Team members added/ # of seats bought | 0.018 |
| Mobile app dev 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 |

**Observation** - As evident from the table above, *mobile app dev sessions per month per team member, and team members added/# of seats bought* are two metrics that are highly associated with churn. *Customer region – Asia, and # of seats bought* are weak predictors of churn.

## **Recommendations to improve short term retention of TakeZero**

* Customers who buy individual plans churn more. Customer research should be done to understand why they are leaving. Is it happy churn (they are just done with their task and don’t want to pay for subscription) or product doesn’t suit their needs or pricing is the issue
* Establish causality b/w metrics - # of mobile app dev sessions, team members added and short-term retention through A/B test. Based on results, if these are sticky features, we should look at onboarding flow and product funnels to identify where users drop off or why they are not using these features. We should improve the discoverability and usability of the features to ensure 100% of customers use them within the first 3 months
* Engage users who have high churn risk score through various channels such as mailers, push notifications or sales calls to ensure their problems are resolved before they actually churn out

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 PM, it is important to look deeper than just the obvious metrics and analyse big numbers with scrutiny. Also, it is pertinent to remember that data should inform us in building and validating better hypotheses and should be backed by qualitative research before implementation.

If you have any comments or suggestions please comment below or reach out to me at - <https://twitter.com/kn_neeraj> or on [LinkedIn](https://www.linkedin.com/in/neerajkumar89/)