
Cohort Analysis and Layer Cake Chart

HAOYU SHEN

AUGUST 17, 2021

1 Introduction

In the 21st century, customers' data is playing a bigger role amongst all industries, SaaS companies will be able to provide the most appropriate service to customers based on the data they collected and analyzed. For this report, we are going to conduct a cohort study of the paid and refunded number of customers to monitor how acquired customers continue purchasing the product or service on a monthly or yearly basis, and what does the monthly revenue is highly dependent on.

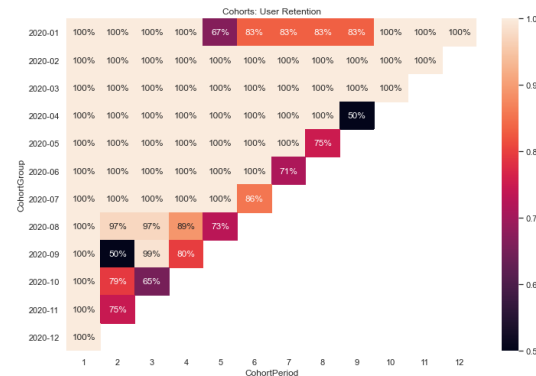
Cohort analysis is a type of behavioral analytic in which users based on their shared traits to better track and understand their actions. This kind of analytical tool allows user to ask more specific questions and make decisions that will reduce churn and increase revenue. Layer cake chart is an informative way to kick off retention analysis, but it doesn't paint the whole picture. It is produced by looking at cohorts in calendar time, can provide powerful insights into the health of business.

For further progress, we will conduct a cohort layer cake chart with no geographic separation. Layer cake chart is another type of chart used by SaaS(Software as a service). As they need to monitor monthly recurring revenue and annual recurring revenue to make decisions, it is essential to use a layer cake chart to monitor customers on a monthly or yearly basis.

2 Cohort Analysis

From the dataset, we find that the top 3 countries with the most customers are US, AU and CA. Therefore, we will conduct cohort analysis on each of them to study the customer retention within these countries better.

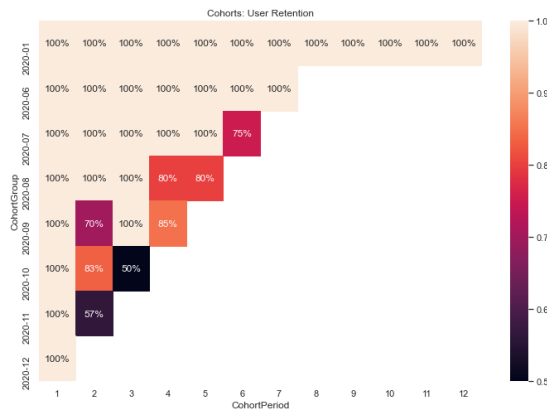
2.1 US Cohort Analysis



We can see the first period (column) of each cohort is 100 % as we assume each customer purchased in month 1. We can see from the above chart that customers who purchased from our company in February and March of 2020 are likely to return to our business later. The user retention rate can be read off from the chart directly, here is an example of how we make sense of the numbers from this cohort chart: in September 2020, we see that the 3rd cohort index has a value of 99 %, it means that 99 % of customer purchased in Sep 2020 and returned 3 months later. From the above chart, it is obvious that the lowest user retention rate occurred in April 2020, the 9th cohort index shows only 50 % customers returned to purchase 9 months after April 2020.

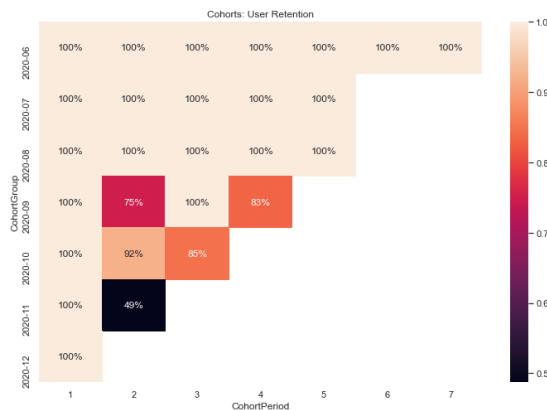
In addition, we see that many indices have a value of 100 %, this usually not happen in real life. A possible reason for this unrealistic result happening is that our data set is not big enough.

2.2 AU Cohort Analysis



Similarly, AU is the second country that our cohort analysis is looking into. It has nearly the same behavior as the US Cohort Analysis as the first column is 100 % for each cohort. Differently, AU has less customers from January to June because our data set is not large enough that some customers may be missing. From this graph, we see that the lowest index is the 3rd cohort from October 2020 with a value of 50 %, which means that only 50 % of customers who made their first purchase in October will purchase again in December.

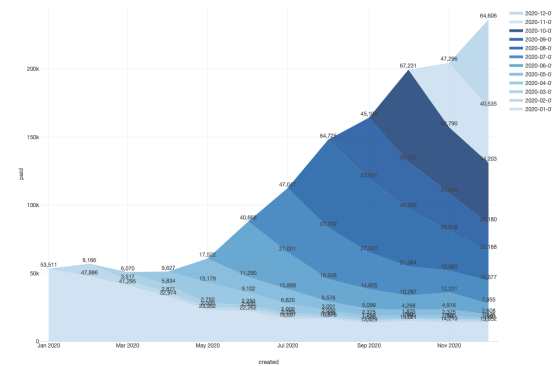
2.3 CA Cohort Analysis



Finally, we do cohort analysis on our last country which is CA. We see that both columns and rows of CA graph are much less than graph of

US and AU. This mainly because we arrange our graphs by descending order, which means the larger your data set is, the larger your cohort analysis graph may be. Although this graph is much smaller, we see that the second cohort which is the probability that customers who first purchased in November 2020 made their purchase again in December is 49 %. Similarly, this graph has the same shortcoming as US and AU that shortage in sample size results in inaccurate outcome of analysis.

3 Layer Cake Chart



The layer cake graph is generated using Exploratory, with data of payment date, paid amount and subscribed date. Exploratory is a tool to generate Layer Cake Chart using processed summary data we got from Python code. In the layer cake graph, the numbers appear on each dot explains the monthly revenue we generated from our customers, regardless of the origin. The x-axis shows the month and the y-axis shows the revenue amount, while different colors represent different first-time purchasing date for our customers. It tends to go down as the time goes by because customers churn. In Jan 2020, we see customers about to pay around 50k on their first purchase, and the purchase amount start to decline to around 10k at the end of 2020.

The graph tells us that the returned customers play a bigger role in our revenue comparing to new customers as we can see the area between

each cohort, which shows as different color, declines not at a fast rate. They kept paying month by month after their first purchase.

From the graph, it could be seen that monthly revenue is highly dependent on repeat business that many customers stick around, each ordering once per month. It seems like there was activities like promotion happened in Nov 2020 since there is a spike during that period despite the decreasing trend of our sales. This is a great customer retention chart as cohort band remains strong over time.

4 Discussion

Cohort analysis and layer cake chart are used to help us understand user's behaviour. Cohort analysis allows us to access the long-term trend in customer retention and calculate their life-time value which allows us to make a better decision. Layer cake chart is an informative way of representing retention analysis, it provides intuitive ideas directly from the graph that non-professionals can understand, only with some explanations.

In this report we used exploratory and python to complete the analysis. From the previous section, we can conclude that these three countries share several similarities and differences from our cohort analysis. For similarities, the probability that customers tend to purchase is positively relate to the time they made their first purchase. For all three countries, US, AU , and CA, customers who made their first purchase in Jan 2020 and June 2020 kept purchasing until the end of 2020 with a probability of nearly 100 %. For customers who made their first purchase by the end of 2020, November, for instance, all three countries have a probability lower than 80%.

For differences, we see that AU and CA have a point in common that the lowest probability of user retention was nearly the end of 2020, where as the lowest for US was in April which can be treated as the start of 2020. The reason that makes this difference can be related to the sam-

ple size of our data set that its not big enough that our conclusion is not guaranteed to be reliable.

From the layer cake graph, as we put all countries as a whole graph, we are able to see some other properties of our data set that cannot be captured by Cohort charts produced previously. It shows that monthly revenue, no matter which country it is, is highly dependent on repeat business as we mentioned above. This is a great layer cake chart for SaaS company as it shows customers tend to purchase repeatedly over the year.

5 Recommendation

In general, both the cohort analysis and the layer cake chart show pretty good outcomes as a business company that customers have a high retention rate. We still want to figure out how to acquire new customers and increase customer loyalty so we can both keep the old customers and expand our business. Firstly, we can offer greater deals for short-term purchase. We can see that customers who purchased at nearly the end of a year are less likely to purchase again compare to those customers who made their purchase at the beginning of a year, which means our long-term plan is working well and attracts costumers to come back to our business. Whereas some modification on the short-term plan need to be made for acquiring new businesses. Also, we can add credit points also with refund for old customers to increase customer loyalty that the longer they stay with the company, the more reward points they are able to retrieve. At last, we can expand our business by cooperating with other companies that customers can get a higher discount with card issued from our company.

6 Reflection

Accurate analysis results are based on accurate data and appropriate tools and methodologies.

From the analysis above, there is still area of improvement. For data, as mentioned before, the data set is not big enough and it may result in inaccurate analysis results. For instance, customers tend to purchase repeatedly with a probability of 100 %. This is not reasonable in real life as customers may change their minds or swap to other companies. Also, we can do user segmenting with a detailed data set. We can segment our customers by industry, monthly recurring revenue, or customer lifetime value, which gives us a better understanding of what our customers need.

For this report, we use heat maps and layer cake chart to show our data. The problem is these graphs are not intuitive enough as many important values are missed like monthly recurring revenue and lifetime value, we show only retention rate from our cohort analysis. We can add more graphs, such as the amount of active users and monthly revenue rate graphs, to our report. These graphs can help us to get an overview of what's happening from our customers, both in terms of user acquisition and activation. We can also track conversion rates and use segments to analyze different user groups. Overall, the more detailed information our analysis is composed of, the more accurate the decision that a company can make.