Cohort Analysis for The Main Customers

Yunying Zhu

October 22, 2021

1 Abstract

In this report, cohort analysis is conducted to assist companies in exploring the composition and behavior patterns of customer groups in their three main source countries, for the purpose of business decision-making and product upgrading. It consists of five parts, which are abstract, introduction, analysis, conclusion and future discussion.

2 Introduction

Business nowadays requires having more insight in their customers in order to make more profits. Many tools are in use, and one of the mostly widely used tool is Cohort analysis, which helps businesses to identify the loyal customers and figure out the rate of user retention. The dataset is sourced from the corporate records of daily business customer transactions, while the period during which the report was prepared spanned throughout 2020. In this report, python and Exploratory are applied to build heatmap and cohort layer cake graph as required to analyze the behavior of customers in the top 3 most-represented countries on the list.

2.1 Cohort Analysis

For an enterprise, customers make up the group that contributes directly to the revenue generated for the enterprise. Therefore, to study the patterns of customer behavior is beneficial for companies to find the solution to improving customer retention rate. This is a significant influencing factor in the sustainable growth of a company. In cohort analysis, the different customer data included in a data set is broken down into the related groups for subsequent analysis. The customer data classed into one group may show similar characteristics and users can be stratified in any dimension as expected. Rather than treating all users as a whole, it divides them into different groups. Group analysis can be conducted to explore the changes in corporate user engagement and retention rate over time. These data can be used to gain more intuitive understanding as to the churn situation of users on a daily basis, thus allowing companies to rethink whether their products fit with the market in real term. Group analysis contributes to distinguishing between growth indicators and engagement indicators, because it is easy for the growth of new users to cover up the insufficiency of existing users engagement.

2.2 Cohort Layer Cake Graph

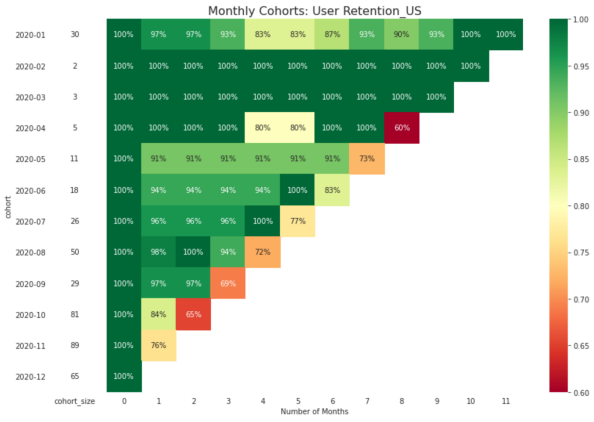
As a significant analysis tool of cohort analyst, the cohort layer cake graph classify customers by the time at which they make the first purchase. Through analysis and comparison of what percentage of sales is contributed to by new and existing customers at a specific point of time, it can help gain understanding as to the patterns of customer behavior. Meanwhile, an analysis of the company for its revenue structure can assist the examination on whether the sales structure is healthy and sustainable, which is crucial for business decision-making on whether publicity shall be increased to attract new customers, or it is best to concentrate on improving product services and continue strengthening the ties with existing customers.

3 Analysis

After doing some basic data cleaning and analysis using Python, it appears that the top 3 most-represented countries include the US, AU and CA. Therefore, a detailed analysis will be conducted of customer retention rate in these three countries later to gain insight into the patterns of customer behavior in these three countries.

3.1 US Cohort Analysis

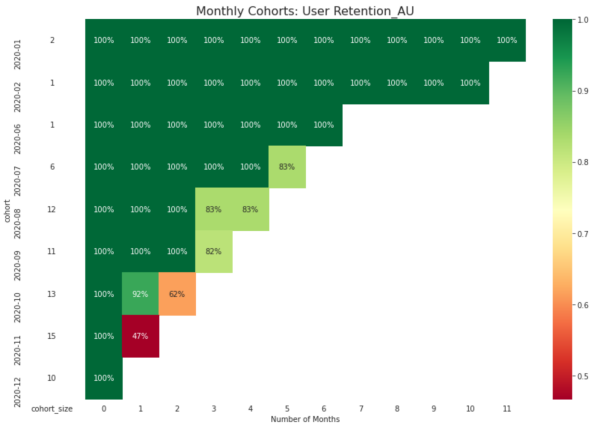
Firstly, the dataset includes 2352 pieces of consumer record related to US customers, ranking top among all countries concerned. Let's just look at the chart. The Y-axis represents the months in which the customers were acquired, and X-axis indicates the number of months after their first-time



purchase. Therefore, customer retention rate can be easily obtained from the chart above. Herein, a sample is provided to help understand the graph correctly. To be specific, the third cohort index in October is 65%, which means 65% of the customers who make the first purchase in October will return 3 months later and make purchase again in December. Additionally, it can be found out that in February and March 2020, all the indices reached 100%, which is rarely seen in reality. A possible reason for this is that the size of our dataset is too small. It could be seen from the heatmap that the cohort size is just 2 in February 2020 and 3 in March 2020. This size of cohort data is clearly insufficient to produce an reliable research results. At last, this table reveals that the lowest customer retention rate was reached in the last cohort index of April 2020. The customers who made their first purchase in April and remained active in December account for merely 50%. It is noteworthy that the user retention rate is supposed to decline from month to month, but not infinitely. That is to say, it should level off in a certain month. Ideally, the higher the value to reach a steady state, the

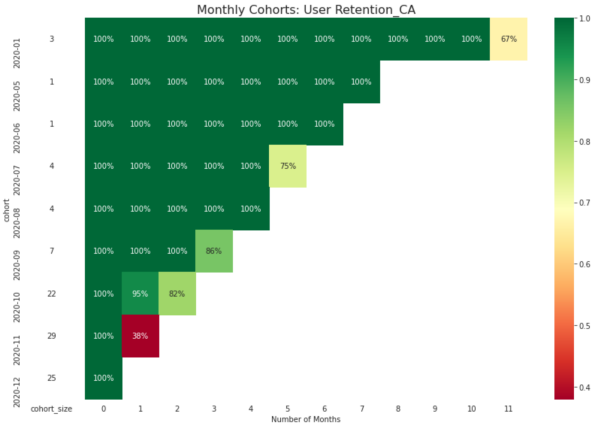
more able the company is to retain users. Due to the difference in business operations between individual companies, the period during which the stable value and the final stable value were reached is also different. However, these characteristics are not completely reflected in the outcome of this research, which results from the insufficient data size as well. Since a large amount of data may be lost during the process of data collection, it is difficult to produce reliable research results. In spite of this, we can still try to gather more useful information from the existing data to make sensible business decisions.

3.2 AU Cohort Analysis



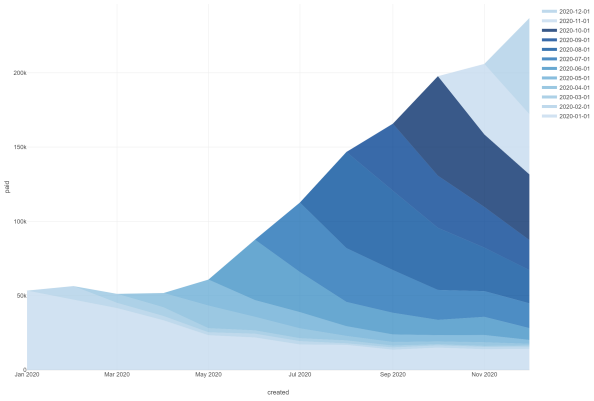
From the heatmap of AU, it can be found out that the amount of data is significantly smaller compared to the US. Therefore, AU heatmap shows more values of 100%. For example, it can be seen that the repurchase rate is extremely high in January, February and June. In fact, however, there are only one or two pieces of customer consumption records in these months. Such a small amount of data makes it difficult to draw reliable conclusions from the data collected during these months. Besides, it can be found out that the lowest customer retention record relates to the second cohort from November 2020. Only 47% of the customers made their first-time purchase in November and made purchase again in December. In the meantime, it can be seen that compared with other months, the number of consumption records from November is the highest, with 15 pieces of payment information recorded in total.

3.3 CA Cohort Analysis



The CA heatmap reveals the same characteristics as the AU and US data. The sample size is overly small, with a lot of 100% value shown in the table. Meanwhile, in the heatmap of CA, the minimum customer retention rate is also in November when the payment record listed is more than other months. Only 38% of customers remained active in December after making their first purchase made in November. Therefore, more data is needed to draw more reliable conclusions for this region.

3.4 Layer Cake Graph



Above is a layer cake graph that could help us gain an insight into the health of our company. In order to better understand the structure of our customer group, I build the layer cake graph with the date of payment as x axis and the sum of payment as y axis for grouping by the date of subscription. In this graph, our customers are divided into 12 "layers" according to the month in which their first purchase is made. The changes in amount pf customer payment over time can be observed. As shown in the figure, the new customers who make purchase for the first time account for the highest proportion in respect of total payment. After the first-time purchase, customer payment will decline sharply in the following four months, with the first month in particular. Afterwards, the pace of decline is slowed gradually and stabilizes after seven months. Moreover, there appears to be some promotion within October and November, with some of the customer making response.

4 Conclusion

Through a comparison between the heatmap and the layer cake graph, it can be found out that the sales of our company remain dependent on those new customers who make purchase for the first time. As time goes forward, the number of customers with intention to buy will decline from month to month, until it gradually stabilizes after 7 months. Both the heatmap of US and layer cake graph indicate that there were a small number of successful marketing campaigns held around November. Some existing customers responded to these activities, thus resulting in a short-term improvement of customer retention rate. Due to a lack of follow-up operations to maintain users Activeness, however, customer retention rate continued the declining trend subsequently. In addition, upon a comparison in the heatmap of AU and CA, it can be known that the lowest customer retention rates of the whole year relate to the second cohort in November, which may be attributable to the promotion held in November that attracted some non-target customer groups to make purchase. Notably, due to the small size of data, the results of cohort analysis are excessively optimistic, which means the results of this analysis shall be treated with rationality.

5 Future Discussion

First of all, since a successful promotion was conducted in November, a survey of all customers can be conducted to better understand their inner thoughts and iden-

tify the factors in customer retention or loss. In addition, measures can be taken to "remarket" lost customers within four months after their first-time purchase (especially within the first month of the purchase), disseminate promotional information through SMS, email, etc., and then continue with the observation of whether the customer retention rate would increase or not. On this basis, the best solution to winning back lost users can be worked out. Finally, in order to make a valuable research later, employees shall be required to record customer payment information clear.