Cohort Analysis on Retention Analysis

Given a set of user data for a new product, how to effectively analyze user behavior and the real user value? Conventional statistical approaches can mislead our decisions by demonstrating well-behaved average value and drawing some beautiful charts. Therefore, we are in urgent need of mastering an effective method to capture the real user value. Data never lies. However, sometimes the people who analyze the data do not do the appropriate analysis, which leads to the wrong interpretation of the user data! As an efficient data analysis method, cohort analysis was first used in medical research to see how the behavior of diverse groups of people varies over time.

What is cohort analysis in finance application?

A cohort is a subdivision of a user group, which refers to the group of users with common behavior characteristics in a specified time. Common behavioral characteristics refer to similar behaviors within a certain period, which can be categorized by different behaviors and contrasting times. For example, the users whose first purchase is in January 2021 may be grouped as a cohort. Or the frequency of use of product starts to drop in the last month of January 2021. Note that cohort analysis focuses on analyzing the differences between separate groups at the same stage of the customer life cycle.

How to Apply Cohort Analysis in Retention Analysis?

Cohort analysis can be used in various scenarios in commercial area, like retention analysis, churn analysis, renewal analysis, and advertising analysis. We take the retention analysis as an example to show how cohort analysis improve effectiveness for retention rate analysis.

月份	新增用户	每月的留存用户数										
		1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月
1月	80	80	78	75	72	70	69	67	66	66	65	64
2月	88		88	88	86	82	78	77	76	73	72	70
3月	105			105	103	103	98	94	92	90	86	82
4月	110				110	107	106	102	99	97	92	90
5月	115					115	114	112	105	98	97	96
6月	128						128	128	122	119	115	110
7月	137							137	136	129	122	118
8月	151								151	149	145	135
9月	161									161	158	154
10月	168										168	167
当月总下单用户数		80	166	268	371	477	593	717	847	982	1120	1086

Figure 1. The retention data for APP A in the time domain.

Acquisition cohorts:

Suppose your company has two new APPs (APP A, APP B) released in the same period and our task is to decide which APP should be put more energy on. The user data are shown in Table 1 and Table 2, which consists of the number of new users and the retention in each month. The retention rate is defined as the ratio of customers who are still frequently using this APP. So how to read this table? Firstly, let us look at the first row of data, in which there are 100 new registrations in the first month and 95 users remaining in the second month. In the third month, 3 more users fall away. By that analogy, the number of retentions is 45 in the 9th month. We can analyze the other row data in the same way. Then we can calculate the retention rate for each month and obtain Table 2. 2. We can know the trend of retention rate for each

group of newly registered users from each row. And beyond that, we can compare the retention rate of separate groups in the column domain.

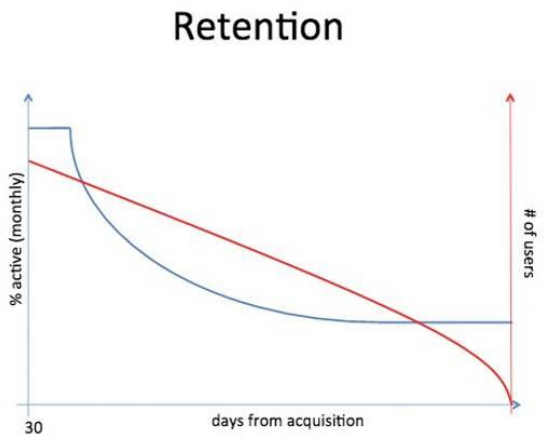

Figure 2. The visualization of retention rate for two products.

In the first few months, the retention rate of A is lower than that of B, and the simple conclusion is that APP A is not as good as APP B. Conversely, if we take a long-term view, we find that the retention rate of APP A continues to decline, while the retention rate of APP B slowly converges to a certain value. The retention trend of an excellent product should have the following characteristics:

1) the retention rate in the horizontal view should stay at a fixed value. For example, there are 100 inexperienced users in a certain month and the retention rate is stable at 50% after half a year, which means that this group of users will be valuable for the company. Otherwise, even though the retention rate declines very slowly, it may go to zero in the end. No matter how many inexperienced users there are, it is meaningless for the long-term development.
2) the retention data in the vertical view should be getting better. The company should constantly improve products and experiences based on users' feedback. The users who join later should enjoy a more excellent product and service, thus leading to a higher retention rate.

But if we think a little bit more carefully. It is found that the promotion of APP A is mostly online and A small part offline, while APP b is just the opposite. Consequently, if we want to analyze the data more precisely, we need to break it down a bit more.

Behavioral cohorts:

Another popular method to group the users is based on the behaviors. For example, after we obtain user data after they have registered the APP A, we can further divide the cohort based on the usage time or their preferred function when using this APP. Then, the columns of the table contain the cohort which is determined by the behavior, and the row data represents the observed activity in the APP. Compared to acquisition cohort, behavioral cohort depends more on the characteristics of the product and should be able to capture more valuable user information if it is utilized effectively. As Kelvin said, "If you cannot measure it, you cannot improve it."