

SPECIAL COVER

boodmo and OWOX BI: How to Evaluate Customer Acquisition Channels Using Cohort Analysis



Kailash Choudhary
Web analyst
at OWOX



Get a complete overview of your marketing performance

About boodmo

[boodmo.com](#) is the largest Indian online marketplace for auto parts. It's a platform where auto parts dealers offer genuine and aftermarket auto components, and customers can find the components they need in just a couple of clicks. boodmo has more than 400 suppliers around the country and focuses on the domestic market: «We're an Indian company, operating for Indian customers only.»

The product catalog on the [boodmo.com](#) features 18 categories of auto components from 42 popular brands (Audi, BMW, Ford, Honda etc.), a total of more than 35 million components. The website is adapted for mobile devices. Users can make purchases on their desktop computers, phones, tablets, or using a mobile app.

Goal

Founded in 2015, boodmo has already managed to occupy a free niche and become a major player in the Indian online market of auto spare parts. The company doesn't intend to rest on its laurels and strives to become «the #1 online destination for everybody interested in the auto service industry.»

To achieve this goal, boodmo needs to expand their market coverage by developing an effective customer relationship model.

Acquiring customers is great, but retaining customers and turning them into loyal clients is even better. That's why the boodmo team aimed to determine the LTV (lifetime value) of their customers and get a better understanding of how to improve it, by conducting a cohort analysis. To calculate the LTV, the company needs data about margins and customer acquisition costs for a certain time period. It's also essential to compare cohort performance across different sources and medium. This will allow for better understanding of which channels bring the most loyal customers, and which turn out to be unprofitable, bringing less revenue than what was invested in them.

Challenge

The company needed a single system to collect all the data required for the cohort analysis. Google Analytics is not suited for the challenge for a number of reasons:

- The information about the current status and content of orders is not updated in Google Analytics. If a user cancels their order or orders other products while the order is still processed, this data won't be reflected in Google Analytics reports.
- The data about margins and order completion rates isn't available in Google Analytics. This data is stored in the company's CRM system.
- Besides that, the cohort analysis report in Google Analytics doesn't show metrics the company wanted to see for each cohort. In other words, the metrics listed below can be viewed separately in other reports, but not in the cohort analysis report:
 - LTV (in Google Analytics, it's shown in a separate Lifetime Value report).
 - CAC (customer acquisition cost).
 - Number of registrations.
 - Number of users in a cohort who have placed one order, two orders, three or more orders.

Solution

To solve the challenge, analysts at boodmo:

- Collected all the necessary data in the [Google BigQuery](#) cloud data warehouse using [OWOX BI Pipeline](#).
- Used Google BigQuery to group users into cohorts and calculate the necessary metrics for each cohort: LTV, CAC, number of registrations, number of users who've made 1, 2, 3 and more purchases.
- Created automated cohort analysis reports by loading the data to Google Sheets.

This entire process is shown on the flowchart below:



Let's take a closer look at each step:

Step 1. Collect all the data in Google BigQuery

boodmo sends user behavior data from their website to Google BigQuery using OWOX BI Pipeline. The product was chosen for the following reasons:

- Data processing speed. With OWOX BI Pipeline, website data is sent to GBQ in near real time. The data becomes available for analysis in 5 minutes or less. It only takes a couple of seconds or minutes to process the data in Google BigQuery, depending on the data volume.
- OWOX BI Pipeline distributes costs across sessions when importing data to Google BigQuery. This allows for comparing advertising costs and revenue for each user and makes it easy to calculate the cost-effectiveness of user cohorts. This can't be done with Google Analytics alone — the data is aggregated and bound to advertising campaigns.
- OWOX BI Pipeline helps overcome a number of Google Analytics' limitations. The data loaded to Google BigQuery is unsampled regardless of its volume. There are also no limitations on the number of dimensions and metrics in reports, unlike Google Analytics reports with a maximum of 5 dimensions and 25 metrics.

The data about margin and order completion rates is loaded to Google BigQuery from the company's CRM system by using [POST requests](#). Information about other ways to load data can be found in the [BigQuery Documentation](#).

Step 2. Create cohorts and calculate metrics

After consolidating the data in Google BigQuery, boodmo created user cohorts and calculated the metrics for each cohort. All the necessary calculations were performed in Google BigQuery using SQL queries and additional UDF functions (for more information, refer to the [Google BigQuery Documentation](#)).

Before proceeding with Step 3, let's take a closer look at what cohort analysis is and what use it may have. In essence, cohort analysis is breaking users into cohorts and examining the behavior of each cohort at certain intervals. The basic difference between cohorts and segments is that cohorts are typically time-bound while segments are based on user behavior or characteristics (gender, age, location, device type etc.). In other words, users in a cohort share the same experience during a specific timeframe. An example of a cohort is, users who first visited the website in June. An example of a segment is, users who only visit the website on their tablets.

To conduct cohort analysis, it's important to determine the timeframe over which a cohort will be built: a day, a week, a month, etc. The cohort's distinctive feature (first visit, registration, first purchase, etc.) should be selected on the basis of what goals you strive to achieve and what metrics you're going to analyze. For example, if the goal is to calculate and improve retention rate, cohorts may be created based on the date of the first visit, registration, or app install. If the goal is to determine the LTV, cohorts should be created based on the time of the first purchase.

Besides that, a company needs to decide on the reporting period, during which each cohort will be explored. In this way, cohort analysis will help determine how the key performance indicators, such as LTV and CAC, change over time in an individually taken cohort.

Cohort analysis would help boodmo solve the following two tasks:

- Increase the customers' LTV. To do this, the company needed to break users into cohorts by the month of the first visit and see how LTV, CAC, the number of registrations, the number of users making purchases changed within each cohort during the reporting period. The company wanted to interact with each cohort separately, in the most appropriate way, based on the user behavior in the cohort.
- Optimize advertising investments by measuring the performance and payback period for each advertising channel. To do this, the company needed to segment cohorts by channel, and compare the LTVs and acquisition costs for each segment.

Step 3. Build cohort analysis reports

boodmo chose Google Sheets to visualize cohort analysis reports, because the size of the report table changes dynamically, new cohorts should be calculated automatically, and the existing cohorts should be automatically supplemented with new data over time.

[Google Apps Script](#) was used to send the data from Google BigQuery to Google Sheets, build reports, and enable automatic updates and formatting of the reports. Google Apps Script is a JavaScript cloud scripting language that provides a way to add custom functions to Google Sheets and other Google apps.

The following functions were added to the report:

- Filter by the type of metrics in the report. All the metrics for the cohorts can be calculated in both absolute and relative values.
- Filters by source and medium. Sources and medium can be specified using a regular expression, which provides an opportunity to analyze only the selected data.
- Custom formulas to apply automated conditional formatting based on the LTV/CAC ratio.

As a result, boodmo obtained a report that looks as follows:

Filter	Value
Columns	absolute values
Source	
Medium	

Month	Metric	N	LTV	CAC	LTV/CAC
2016-12 Collect user 90% 20-01-2017	ac LTV	0.00	0.00	0.00	0.00
	ac CAC	0.00	0.00	0.00	0.00
	Retention	21 494	0.07	0.00	0.00
	Registration	179	0.00	0.00	0.00
	1st Order	40	0.00	0.00	0.00
	2nd Order	0	0.00	0.00	0.00
2017-01 Collect user 90% 01-02-2017	ac LTV	0.00	0.00	0.00	0.00
	ac CAC	0.00	0.00	0.00	0.00
	Retention	21 447	0.00	0.00	0.00
	Registration	179	0.00	0.00	0.00
	1st Order	40	0.00	0.00	0.00
	2nd Order	0	0.00	0.00	0.00
2017-02 Collect user 90% 01-03-2017	ac LTV	0.00	0.00	0.00	0.00
	ac CAC	0.00	0.00	0.00	0.00
	Retention	40 740	0.00	0.00	0.00
	Registration	1 862	0.00	0.00	0.00
	1st Order	0	0.00	0.00	0.00
	2nd Order	0	0.00	0.00	0.00
2017-03 Collect user 90% 01-04-2017	ac LTV	0.00	0.00	0.00	0.00
	ac CAC	0.00	0.00	0.00	0.00
	Retention	34 770	0.00	0.00	0.00
	Registration	1 862	0.00	0.00	0.00
	1st Order	0	0.00	0.00	0.00
	2nd Order	0	0.00	0.00	0.00

The LTV and CAC metrics in the report were calculated as follows:

- $LTV = \text{cohort margin, divided by the number of users in the cohort.}$
- $CAC = \text{cohort acquisition costs, divided by the number of users in the cohort.}$

The green background color of the LTV cell in the report indicates that acquisition costs for the cohort 2016-12 were covered by the revenue they brought during the 3-month reporting period. Red cells indicate that LTV of users in the cohort is lower than CAC. If the indicators show that the acquisition costs are unlikely to pay off, it's worth paying more attention to users in that cohort. For example, send these users a promo email.

In addition, the company can assess the LTV performance across different customer acquisition channels, and replace the inefficient channel (CAC is higher than LTV) with a more efficient one.

Results

- boodmo received a fully automated solution for evaluating channel performance for each cohort of users grouped by month. The company can now measure both the performance of each individual channel and the performance across all channels.
- By analyzing the behavior of individually taken user cohorts, the company can take different approaches to remarketing and building relationships with their customers in each cohort.
- By evaluating channel performance, boodmo can optimize customer acquisition costs and thereby increase the LTV for each cohort.