

CLV, Customer Segmentation & RFM

Part 4: CLV Graded task

The task

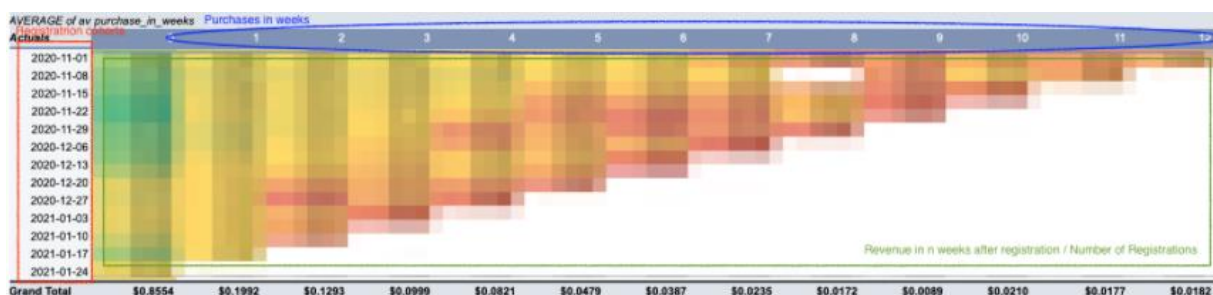
You got a follow up task from your manager as he read some article that calculating CLV using Shopify's formula is too simplistic. He has heard somewhere that using cohorts produces more reliable and actionable results. Once again please use [turing_data_analytics.raw_events](#) table to answer follow up questions from your manager.

TIP: imagine that current week is 2021-01-24 (the last weekly cohort you have in your dataset).

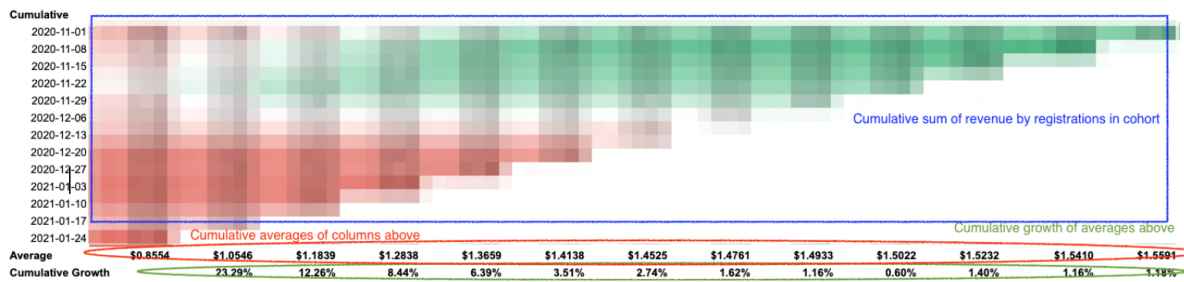
Additionally, he identified 2 problems with your previous analysis:

1. You included only customers who purchased something, while marketing is counting all user registrations that they manage to bring to your ecommerce site. Thus, you need to adjust your calculations to include all users who have been on your website, not only the one who purchased something.
2. Your average customer does not tend to stay with your ecommerce site for too long. He wants to see weekly picture using cohorts. He expects customers to stay no longer than 12 weeks with your ecommerce site.

As the first step you should write 1 or 2 queries to pull data of weekly revenue divided by registrations. Since in this particular site there is no concept of registration, we will simply use the first visit to our website as registration date (registration cohort). Do not forget to use `user_pseudo_id` to distinguish between users. Then divide revenue in consequent weeks by the number of weekly registration numbers. Once you apply conditional formatting to your data, the end result should look something like this:



Next you will produce the same chart, but the revenue / registrations for a particular week cohort will be expressed as a cumulative sum. For this you simply need to add previous week revenue to current week's revenue. Down below you will calculate averages for all week numbers (weeks since registration). Down below that you will calculate percentage growth, which will be based on those average numbers:



Basically, the chart above gives you growth of revenue by registered users in cohort for n weeks after registration. While numbers below summarize those values in monetary terms (red marked numbers) and percentage terms (green marked numbers). This provides you with a coherent view of how much revenue you can expect to grow based on your historical data.

Next, we will focus on the future and try to predict the missing data. In this case missing data is the revenue we should expect from later acquired user cohorts. For example, for users whose first session happened on 2021-01-24 week we have only their first week revenue which is 0.19USD per user who started their first session in this week. Though we are not sure what will happen in the next 12 weeks.

For this we will simply use previously calculated Cumulative growth % (red marked area in chart above) and predict all 12 future weeks values (ex. for this cohort we can calculate expected revenue for week 1 as $0.19\text{USD} \times (1 + 23.29\%) = 0.24\text{USD}$, for week 2 as $0.24\text{USD} \times (1 + 12.26\%) = 0.27\text{USD}$). Using Avg. cumulative growth for each week we can calculate that based on 0.19USD initial value we can expect 0.35USD as revenue on week 12. Provide a chart which calculates these numbers for all future weeks (up till week 12).

See example chart below which should show the following numbers for all weekly user cohorts:



The chart above is our final result. You should calculate the average of cumulative revenue for 12th week for all users who have been on your website. This not only provide better estimate of CLV for all your users who have been on your website (including the ones who did not purchase anything) but also allows you to see trends for weekly cohorts. Have a look at the conditional formatting you added to all 3 charts and be prepared to answer follow up questions.

Evaluation criteria for Graded project submission:

- SQL, correct columns identified to make analysis.
- SQL, correctly calculated main metrics: revenue, orders, customers.
- SQL, correctly calculated user counts by weekly cohorts.
- Google sheets, visualisation follows examples above, at least 3 different tables with conditional formatting applied are provided.
- Analysis, findings and calculations are correct and main trends identified.
- Analytical approach to the problem

During a task review, you may get asked questions that test your understanding of covered topics.

Sample questions:

- Does including all customers (not only the ones who purchased something) influence calculation of CLV?
- Is your previously calculated CLV (using Shopify formula) still accurate?
- How does cohort analysis help you understand what revenue you can expect in the future?
- If we knew that our business Take Rate (% of revenue we can actually treat as profit) is only 10%, would you use this percentage estimating Lifetime value of the customer? Could you compare it against customer acquisition costs provided by marketing?
- What caveats can you expect from such an approach?