Analysis of MockCompany's Transaction-Level Data

Sotiris Papasotiriou

Abstract

Abstract: This document presents a comprehensive analysis of MockCompany's transaction-level data to develop an investment proposal. The analysis includes data cleaning, cohort analysis, churn selection, projections, and parameter selection for potential funding.

Contents

1 Introduction

1.1 Definitions

These are the definitions that we'll be referring to throughout this document.

- **Cohort**: A cohort is a group of customers. Each cohort corresponds to a month. In the case of MockCompany, a customer's cohort is calculated as the month containing the customer's first nonzero payment.

- **Spend**: For a given cohort, the associated spend is the amount the company spent in Sales and Marketing that month. In the case of MockCompany, it is the associated "spend" per cohort in spend.csv.

- **Payment Period**: For a given cohort and a payment month, the Payment Period is the number of months between the payment month and the start of the cohort. For example, if we're looking at all payments made in Feb 2024 by the Feb 2024 cohort, those payments would have an associated Payment Period of 0. Similarly, the payments made in Feb 2024 by the December 2023 cohort would have an associated Payment Period of 2.

- **M0 / M[n]**: For a given cohort, M0 is the revenue the cohort has brought in in Payment Period 0, divided by the spend for that cohort. For a given cohort and positive integer n, M[n] (e.g. M1, M2,...) is the cumulative revenue brought in by the cohort in Payment Periods 0 through n, inclusive, divided by the spend for that cohort. For example, if the Jan 2024 cohort of customers paid $100 in Jan, $90 in Feb, and $80 in Mar, and had an associated spend of $1,000, then the M0 would be 10%, the M1 would be 19%, and the M2 would be 27%.

- **Delta M1 / Delta M[n]**: For a given cohort and nonnegative integer n, Delta M[n] is the revenue brought in by the cohort in Payment Period n, divided by the spend for that cohort. In other words, Delta M[n] = M[n] - M[n - 1]. For example, if the Jan 2024 cohort of customers paid $100 in Jan, $90 in Feb, and $80 in Mar, and had an associated spend of $1,000, then the Delta M1 would be 9% and the Delta M2 would be 8%.

- **Thresholds**: The levels of cohort performance that we are fine continuing to fund at. These can take all sorts of forms. For example–we might set the threshold parameters for MockCompany at 30% M0 and 1% Delta M[1-10], meaning that in order to continue funding MockCompany, we need each cohort we fund to have an M0 of at least 30% and a Delta M1, Delta M2, . . . , Delta M10 of at least 1% each.

1.2 Company Overview

MockCompany is a prospective company for investment. From our analysis, we have concluded that MockCompany operates a subscription-based service/product, offering customers the option to pay on a monthly or quarterly basis.

1.3 Objective

The objective of this analysis is to:

- Understand MockCompany's data and customer cohorts.
- Analyze cohort performance over time.
- Predict future cohort performance.
- Develop threshold parameters for an investment deal.
- Prepare an investment proposal based on the findings.

2 Data Overview

2.1 Marketing Spend Data

The marketing spend data captures MockCompany's monthly marketing expenditures from October 2022 to August 2024, categorized as follows:

- **Performance Marketing**: Spending focused on direct customer acquisition through measurable channels.
- **Brand Marketing**: Spending aimed at building brand awareness and fostering long-term customer engagement.

Each month's total marketing spend is the sum of Performance and Brand Marketing, with all values verified to ensure accuracy.

2.2 Customer Payment Data

The customer payment data (`customer_payments.csv`) contains transaction records starting from July 2021, with the following key columns:

- **customer_id**: A unique identifier for each customer.
- **payment_date**: The date on which each payment was made.
- **amount**: The amount paid in each transaction.
- **payment_frequency**: The payment schedule, either *monthly* or *quarterly*. Each customer is assigned one payment frequency.

The dataset includes cohorts from the first cohort on July 1, 2021, to the last cohort on September 1, 2024.

Key observations:

- No customers switched between subscription types (monthly to quarterly or vice versa).
- The dataset includes both monthly and quarterly subscribers.
- **Monthly customers** paid $9.99 from July 1, 2021, until January 1, 2024, when the fee increased to $11.99 (20% increase).

- **Quarterly customers** paid $24.99 from July 1, 2021, until January 1, 2024, when the fee increased to $29.99 (20% increase).
- Some customers signed up but never made any payments.
- Additionally, there are cases where customers signed up but took 1 or 3 months before making their first non-zero payment (1 for monthly customers, 3 for quarterly).
- It appears that each month in the dataset has at most 28 days. That is, the unique payment days range from 1 to 28. For the sake of the assessment, we assume this is due to the data creation process and not because data is missing.
- There appears to be one outlier customer, `cus_657763319983990`, with a total of 323 payments. All other customers have up to 39 payments.
- The dataset ends on September 15, 2024. Therefore, the September 2024 cohort is incomplete. Additionally, all past cohorts have incomplete payment data for September 2024.

3 Data Cleaning and Preparation

3.1 Processing Customer Payment Data

To prepare the customer payment data for analysis, the following steps were undertaken:

1. **Cohort Assignment**:
   - For each `customer_id`, identified the date of their first non-zero payment.
   - Assigned the **cohort** as the month and year of the first non-zero payment.
2. **Payment Period Calculation**:
   - Calculated the **payment_period** as the number of months between the cohort date and each subsequent payment date.
   - Note: Some customers may take several months after joining to make their first payment, resulting in negative payment periods. These entries are excluded from the main analysis but retained in the dataset for potential future analysis.
3. **Data Cleaning and Validation**:
   - **Customer Removal**: Removed customer with ID `cus_657763319983990` due to outlier behavior in payment patterns.
   - **Month Removal**: Excluded data for September 2024 as it represents an incomplete month, ensuring that only complete monthly data is used in the analysis.
   - **Zero Payment Removal**: Removed customers who only made payments of 0, as they do not belong in any cohort.
   - **Negative Period Exclusion**: Excluded data entries where the `payment_period` was negative, as these represent months before the first valid payment.
   - **Payment Gap Check**: Verified that no customer, whether on a monthly or quarterly plan, had skipped a payment period. This means that there is no churn and return customer.

3.2 Processing Marketing Spend Data

The marketing spend data was processed through the following steps:

1. **Cohort Removal**: Removed the September 2024 spend data as it represents an incomplete cohort, ensuring that spend data aligns properly with complete customer payment data.

2. **Data Verification**: Verified that the total spend equals the sum of performance marketing and brand marketing spend.

3.3 Finalized Data Structures

3.3.1 Customer Payments DataFrame

The customer paymentdata is structured as follows:

- **customer_id**: Unique identifier for each customer.
- **payment_frequency**: The payment frequency (e.g., monthly, quarterly).
- **payment_date**: Date of each payment.
- **amount**: The amount paid for each transaction.
- **first_nonzero_payment**: Date of the first non-zero payment.
- **cohort**: The cohort to which the customer belongs.
- **payment_period**: Payment period relative to the first payment.

3.3.2 Spend DataFrame

The spend data is structured as follows:

- **cohort**: The cohort for which the spend is recorded.
- **performance_marketing**: The amount spent on performance marketing.
- **brand_marketing**: The amount spent on brand marketing.
- **spend**: Total spend for the cohort.

4 Structured Analysis

4.1 Spend Analysis

4.1.1 Spend Breakdown vs Cohort Size

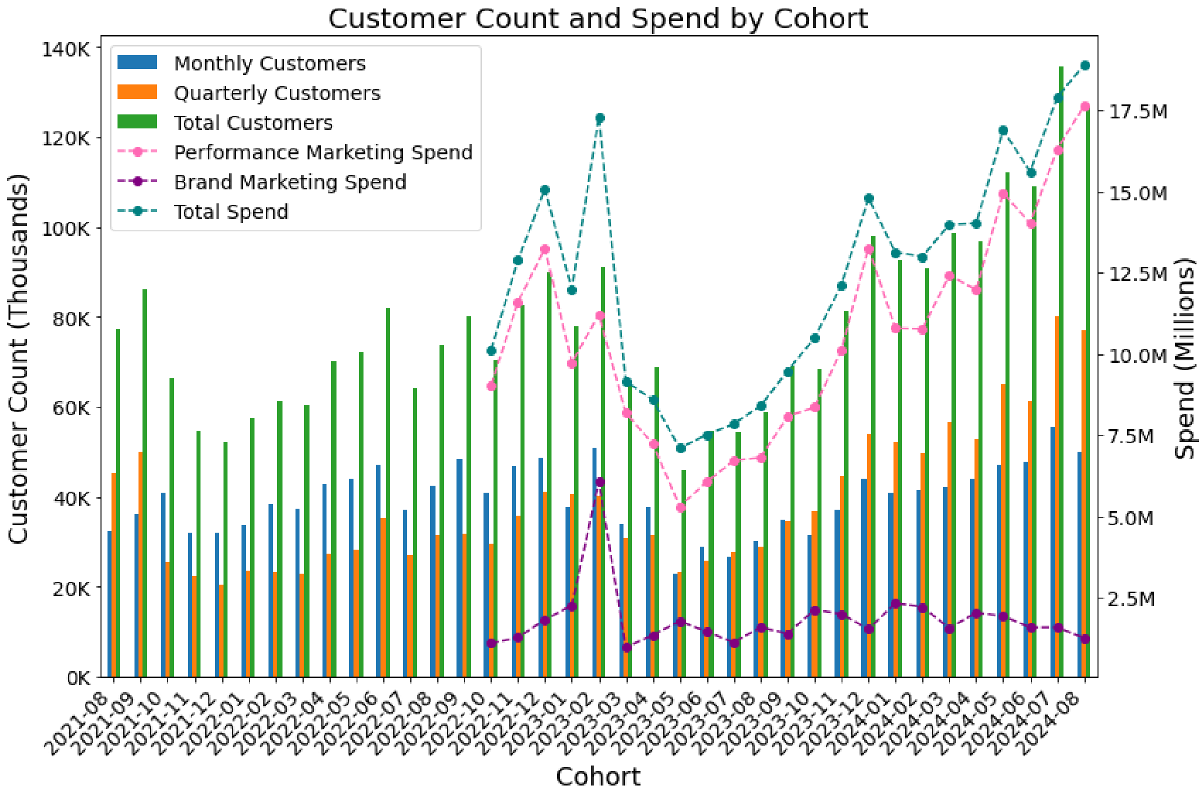

Figure 1: Spend and Cohort Breakdown

- **Excluded cohort**: We have excluded cohort 2021-07 as it has an unusual high amount of monthly customers. For the majority of the analysis we don't consider the cohort relevant. Normally, we would acquire more information about this cohort from MockCompany.

- **Performance marketing**: Performance marketing spend is always the majority of spend. It also appears to be very correlated with the cohort size.

- **Brand marketing**: Brand marketing is generally much lower and it doesn't fluctuate as much. The brand marketing spend of 2023-02 is unusually high - Perhaps this was a response to an event that hurt the companies image and this is a correction effort. Again, we would request more information about this from MockCompany.

- **Cohort breakdown**: Over time, we see a trend that is shifting from cohorts with more monthly customers to cohorts with more quarterly customers. This trend seems to pick up around the time that the subscription increase took place.

4.1.2 Estimated Missing Spend

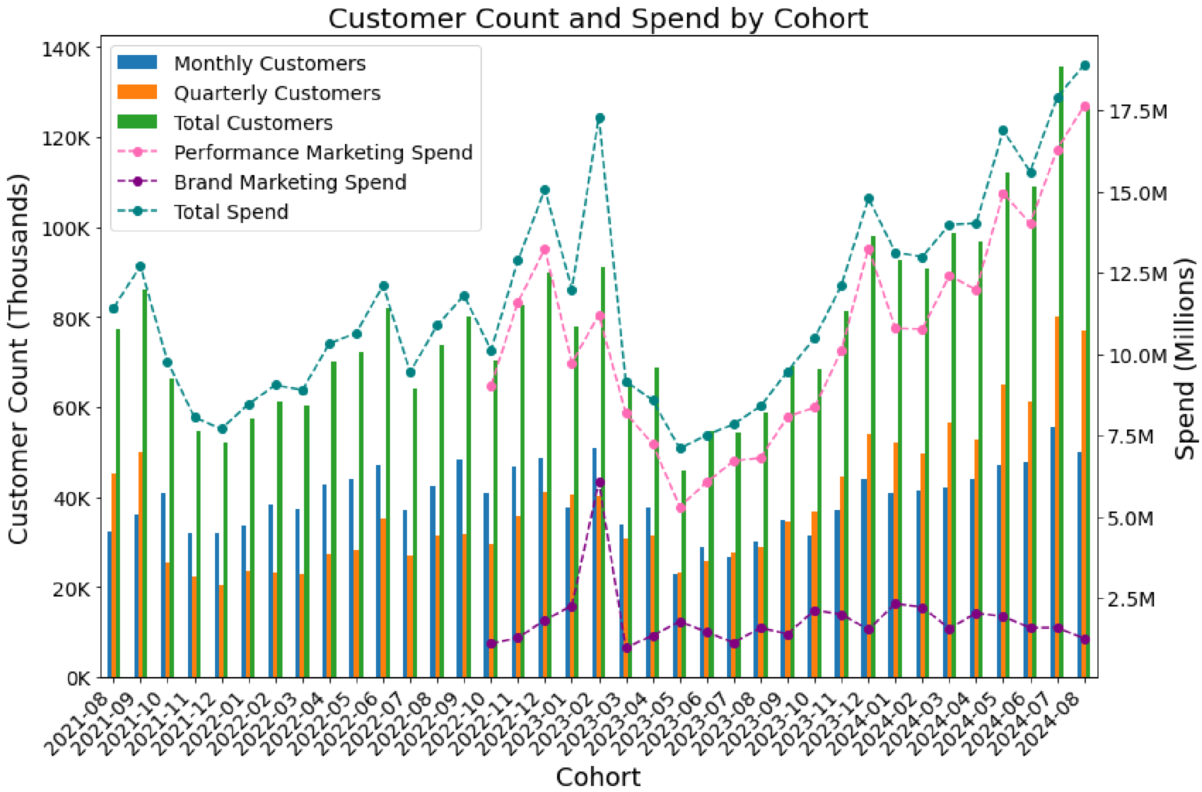

Figure 2: Estimated Missing Spend

- **Missing spend data**: We have calculated the estimated missing spend data by multiplying the average customer aquisition cost with the number of customers in the cohorts without spend data. The results above seem to be reasonable.

4.2 Payment Analysis

4.2.1 Payments by Payment Period

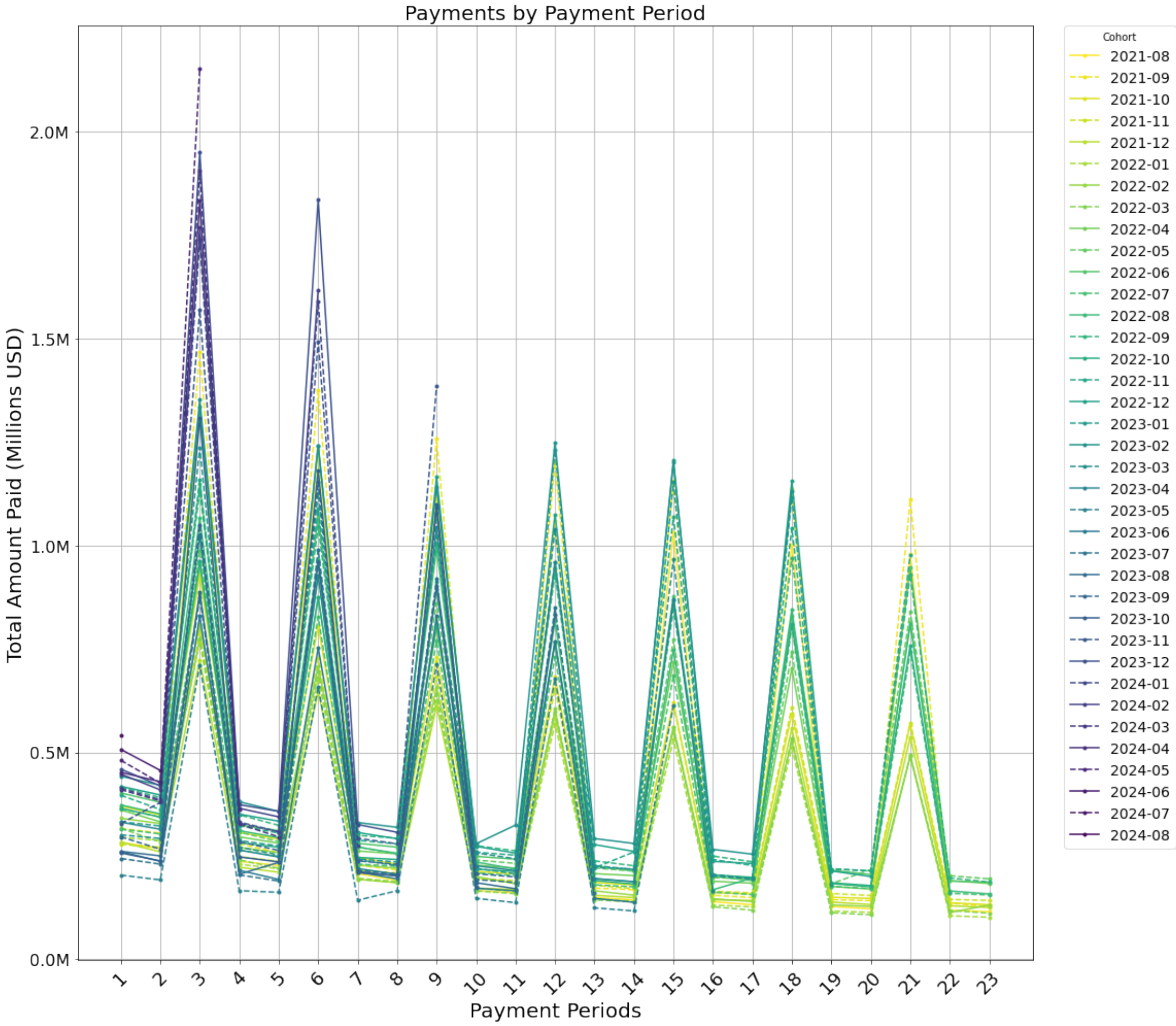

Figure 3: Payments by Payment Period

- **Payment Jumps**. As expected we see jumps in payments at 0 mod 3 payment periods. This is expected as these are the only payment periods where quarterly customers make payments. It is hard to draw conclusion from the combined payments by payments period plot as quarterly and monthly customers have different behaviors.

- **Declining Amount Payed over time**: We notice an decreasing trend in amount paid that we suspect is due to customers leaving in the first periods of cohorts life. We take a deeper look into this later.

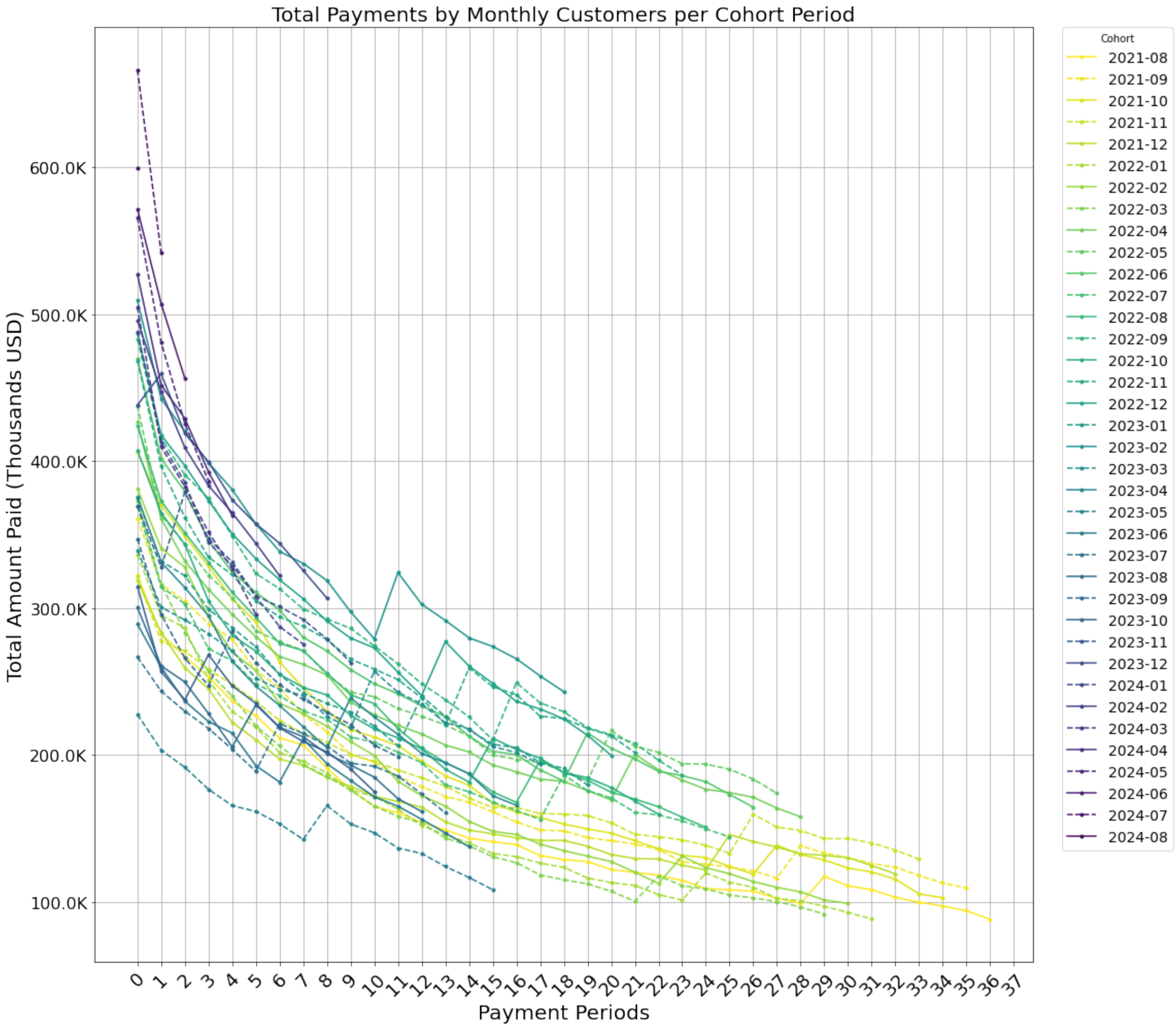

Figure 4: Monthly Payments by Payment Period

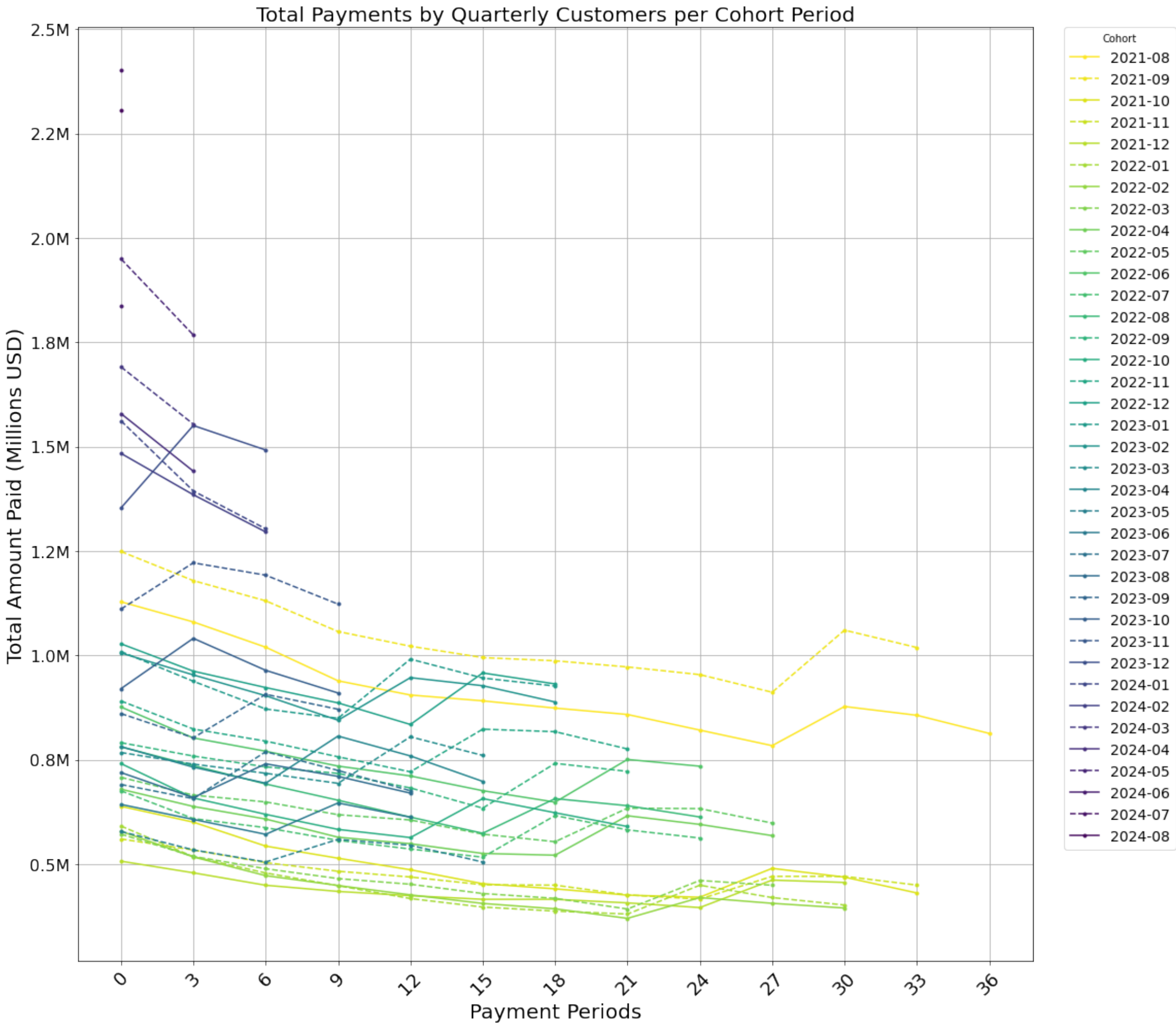

Figure 5: Quarterly Payments by Payment Period

Monthly and Quarterly Separately

- **Effect of Subscription Increase**. Once the separation is done, the effect of the subscription increase is clear. The jumps we see in both plots are a result of that increase.
- We can also see that the behavior of monthly and quarterly customers are different. There are clear differences in the rate at which payment by payment periods decrease throughout the life of the cohorts. We will see that later as well but quarterly customers appear to perform better.

4.2.2 ROAS

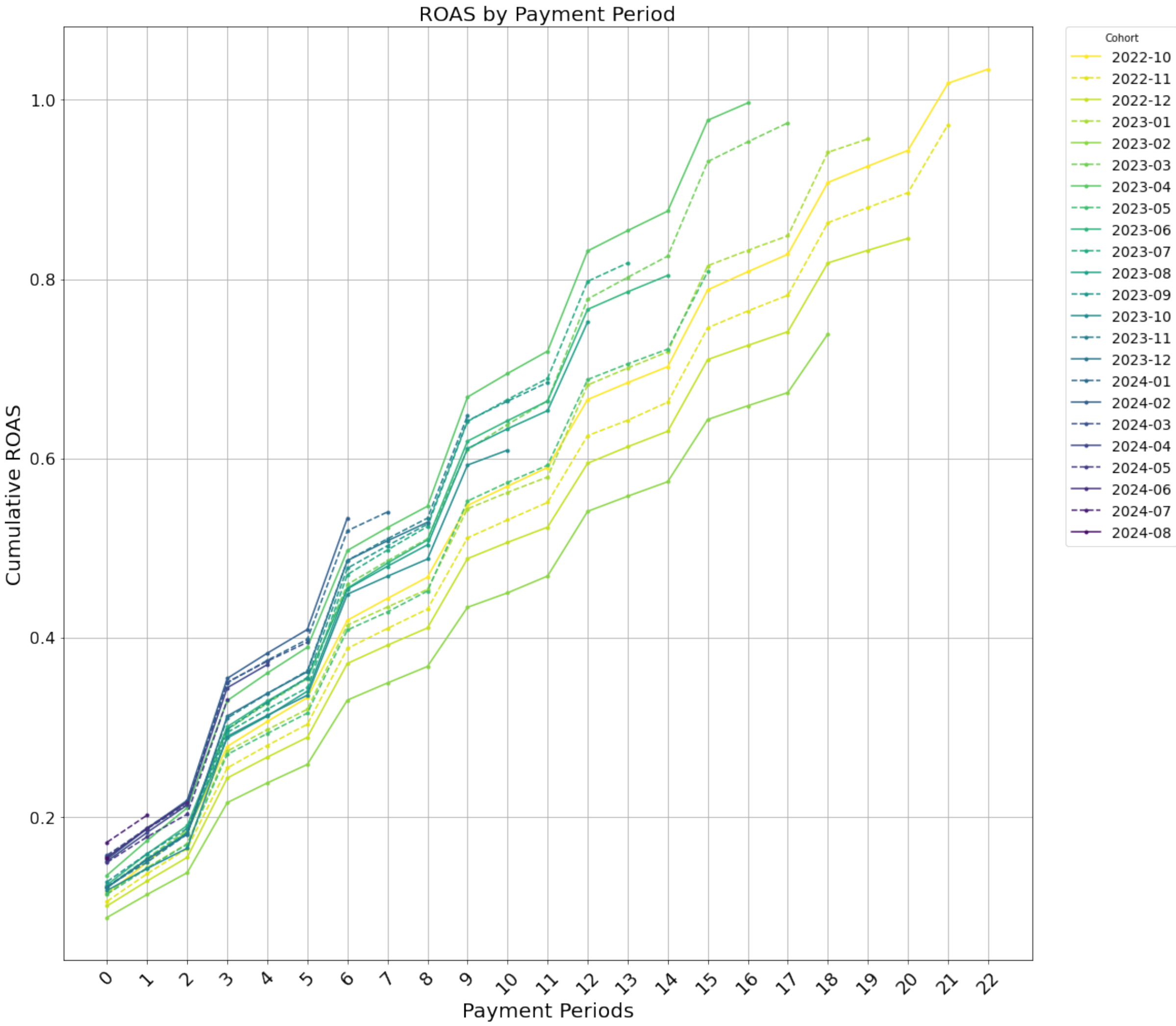

Figure 6: ROAS by Payment Period

- **Worse performing cohort**. It appears that the worse performing cohort is 2023-02. This is the cohort that had the jump in brand marketing spend (also mentioned in the Spend Analysis).
- We also see that the more recent cohorts seem to be performing better than the older ones. However, one should note that the newer cohorts had the increased subscription take effect from an earlier payment period and thus the takeaway is not conclusive.
- However, it is reasonable to assume that the increase in subscription was a positive change. Meaning that any potential decrease in cohort size or increase in customer churn is outweighed by the 20% increase in the value of each customer

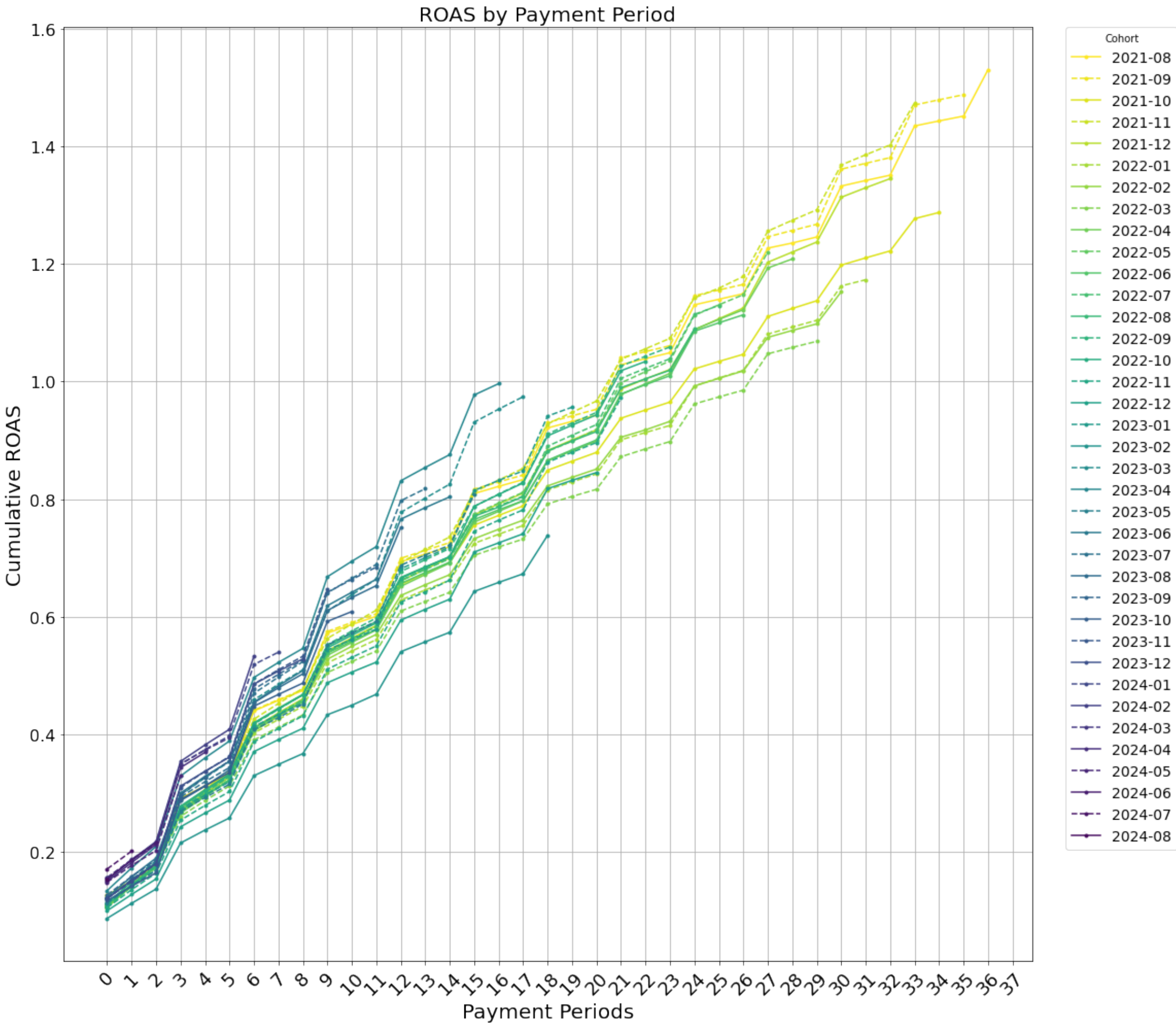

Figure 7: ROAS by Payment Period Including Estimated Spend

- We also include the ROAS plot after including the cohorts for which we estimated the spend amount to make sure that nothing seems out of the ordinary.

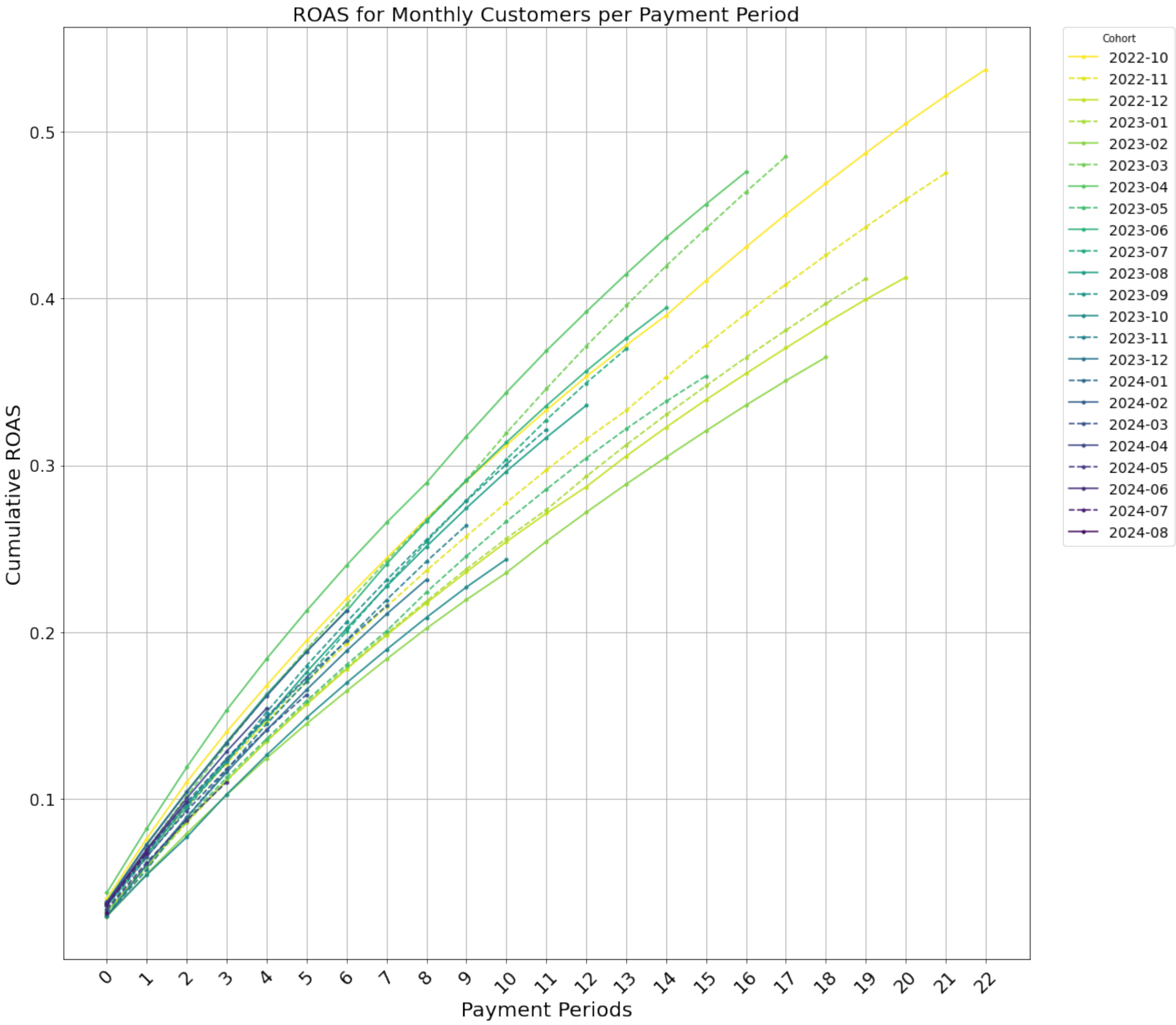

Figure 8: Monthly ROAS by Payment Period

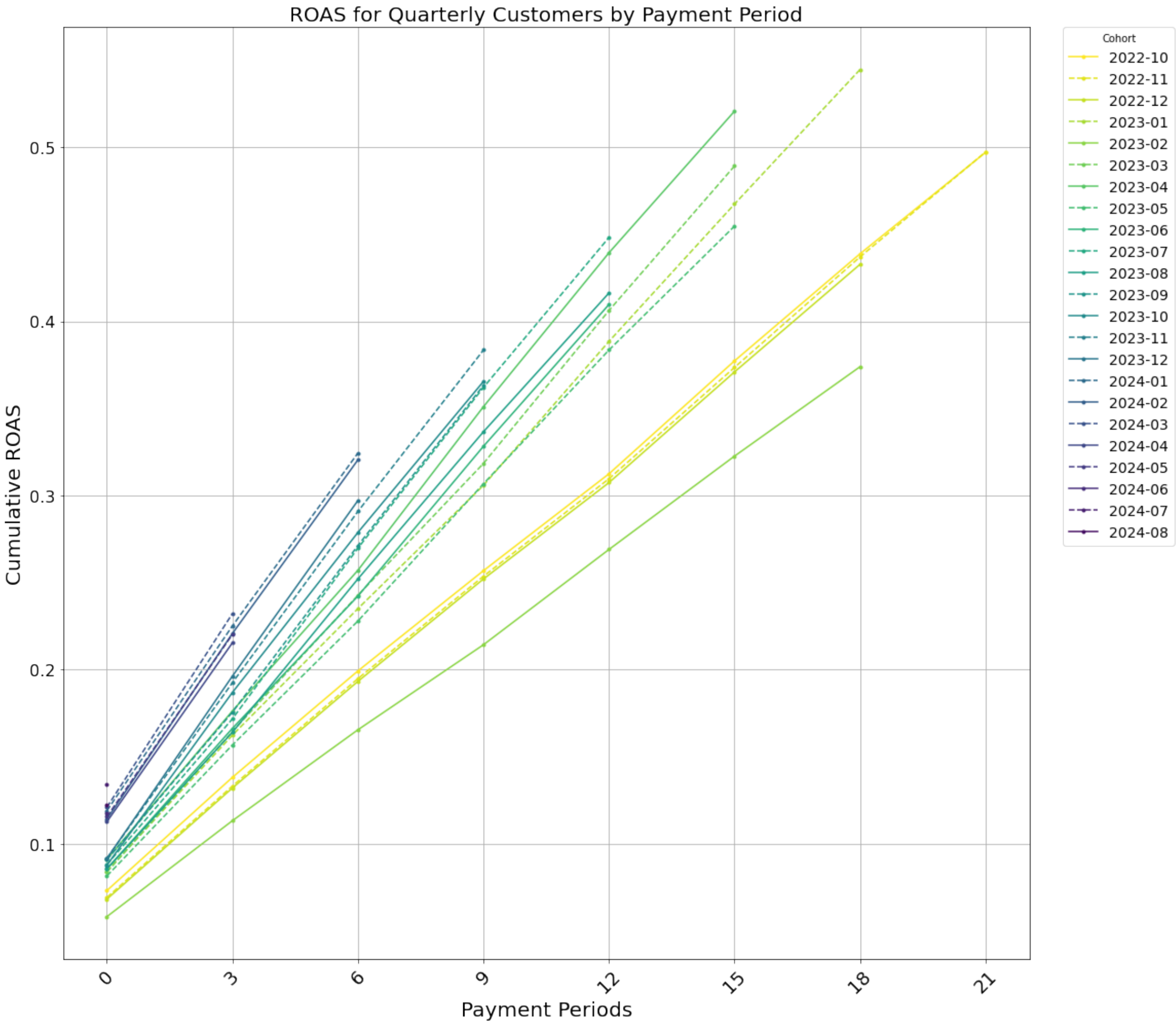

Figure 9: Quarterly ROAS by Payment Period

- Again, we see the differences in monthly and quarterly customer behaviors that were observed in the payments by payment period plots.
- It is worth mentioning that this difference is not unreasonable. One could argue that quarterly customers have a higher trust in MockCompanys product since they are willing to commit to the quarterly subscription. They are therefore going to churn at lower rates too - We will also see this later on.

4.2.3 Payment Churns

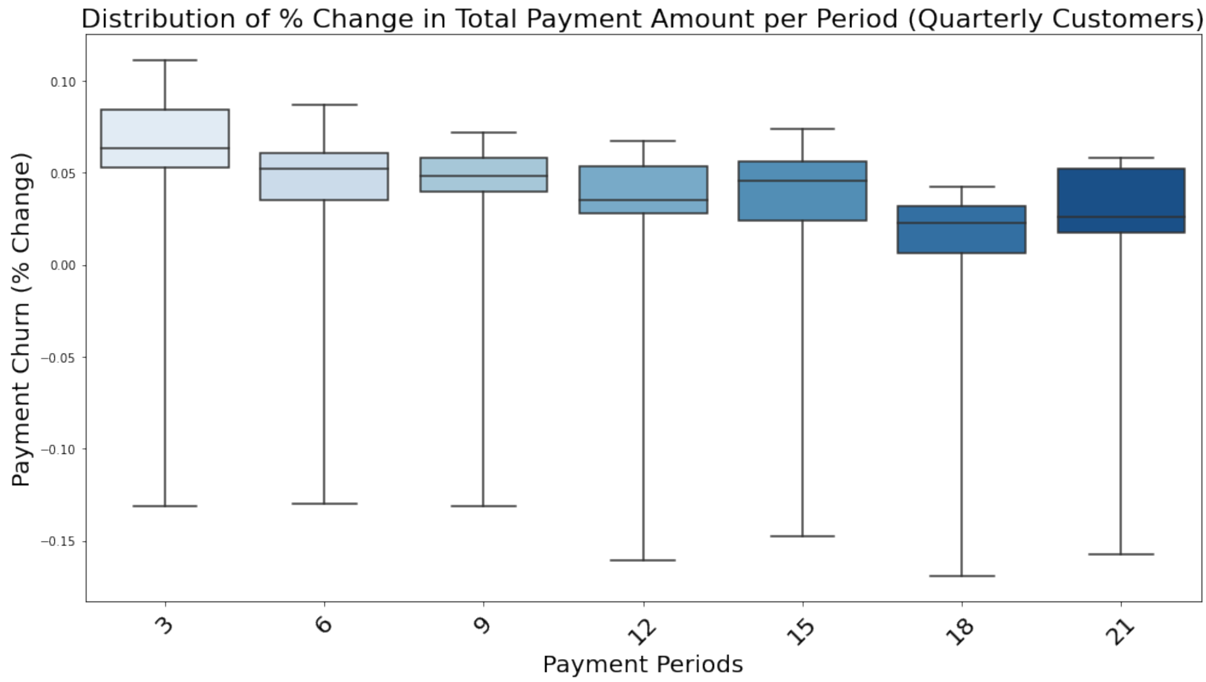

Figure 10: Quarterly Payment Churns

Here I have included the quarterly payment churns to highlight an issue. When looking at payment churns, we see that they can have negative values. This is the effect of the increase in the subscription price. If the subscription price was constant throughout time, we wouldn't observe negative churns since payments are just the number of costumers times the corresponding subscription amount. Note that the size of a cohort in number of customers can only decrease in time!

Churns will be a crucial focus in the upcoming phases of our analysis, as they are the main variable used to project cohort performance. Thus, it's essential to accurately understand the true churn patterns of our cohorts without them being obscured or seemingly improved by the increase of the subscription price.

For this reason, for the rest of the structure analysis sections, we will focus on cohort size instead of cohort payments.

4.3 Cohort Size

4.3.1 Cohort Size by Payment Period

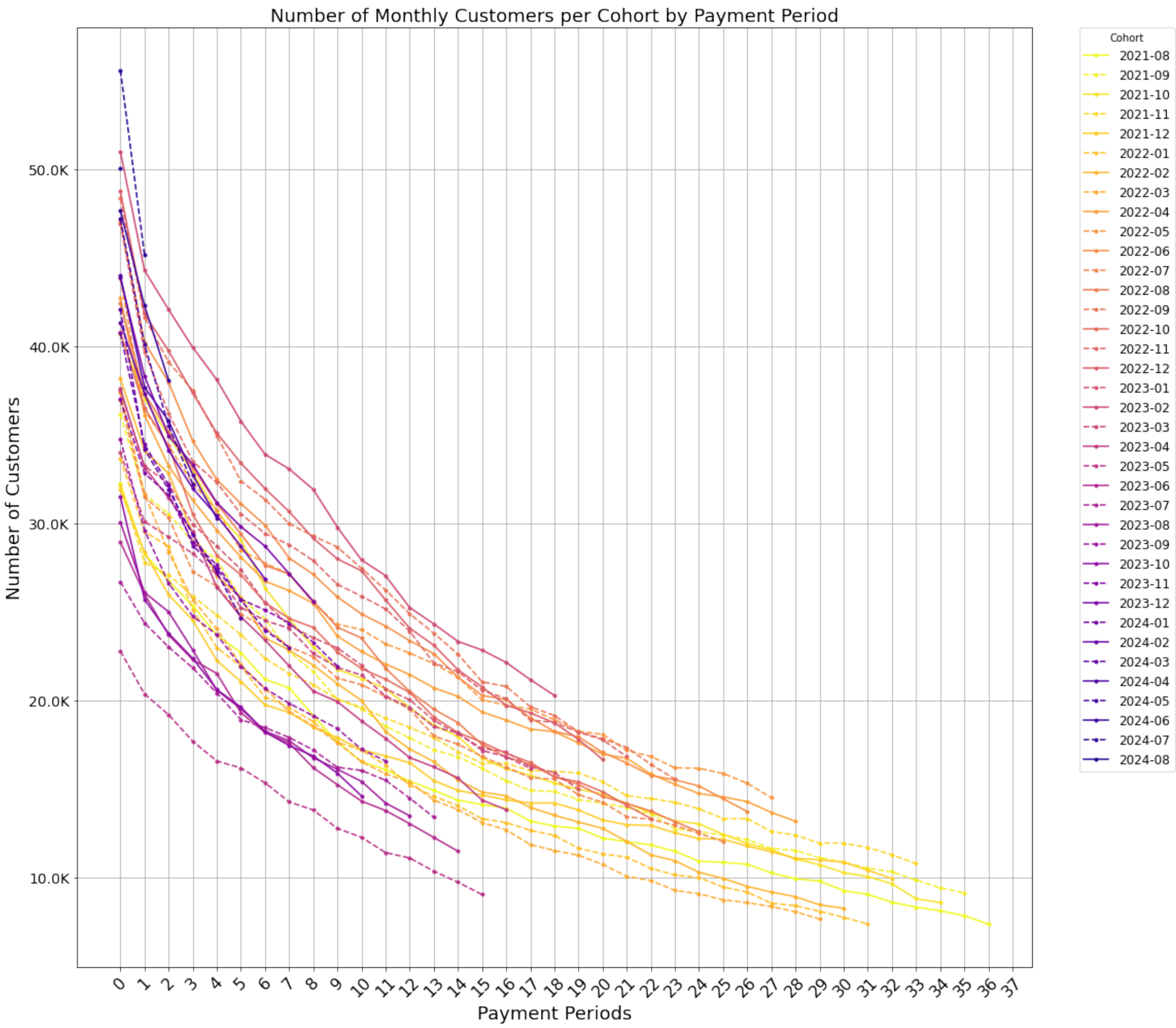

Figure 11: Monthly Cohort Size by Payment Period

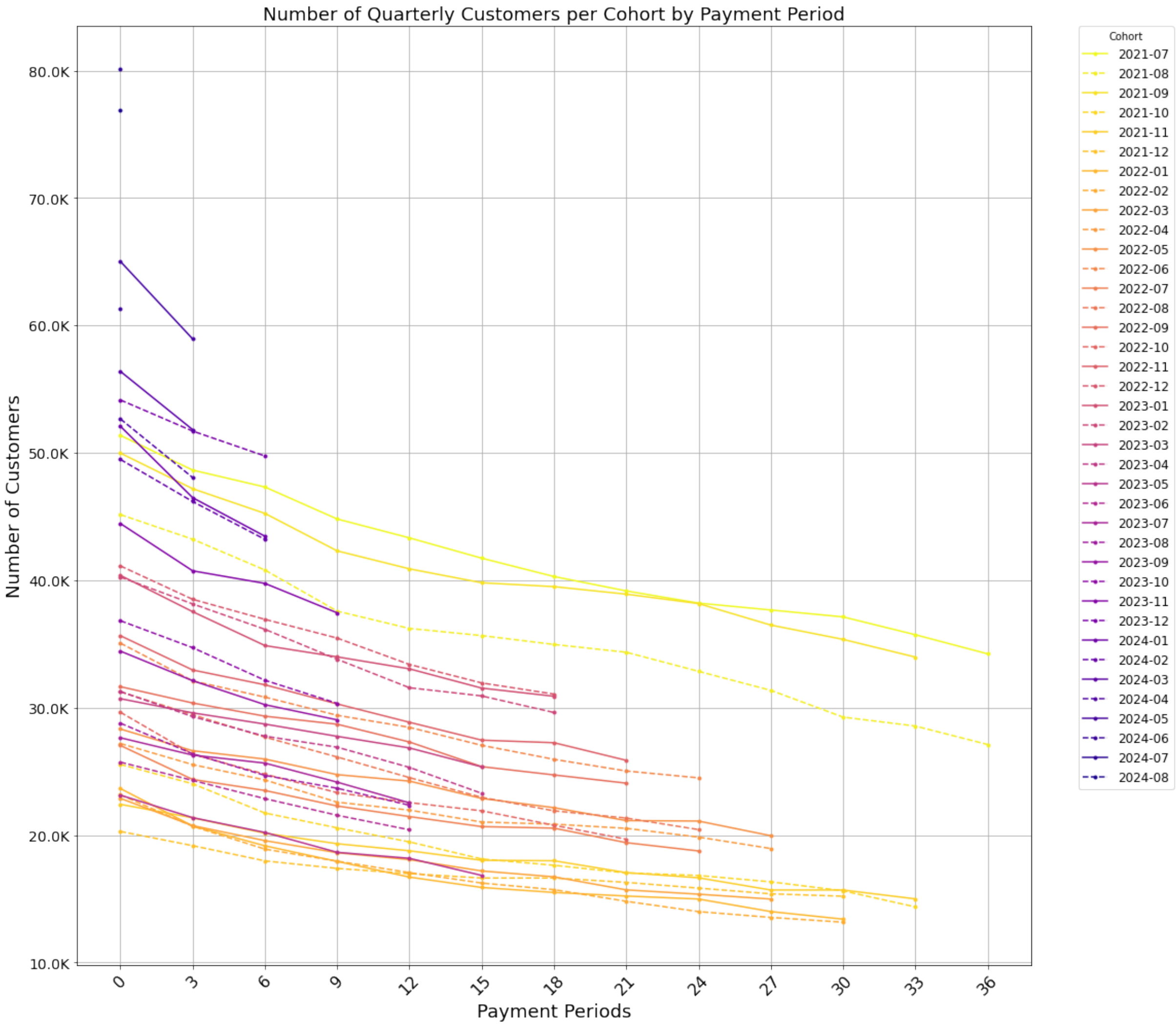

Figure 12: Quarterly Cohort Size by Payment Period

- As expected we see that the jumps we saw in the corresponding payment plot have disappeared as the effect of the subscription increase is not present.
- For both types of customers (monthly and quarterly), we see bigger drops in customer size during initial payment periods. Drops are definitely higher for monthly customers when compared to quarterly.

4.3.2 Cohort Size Churns

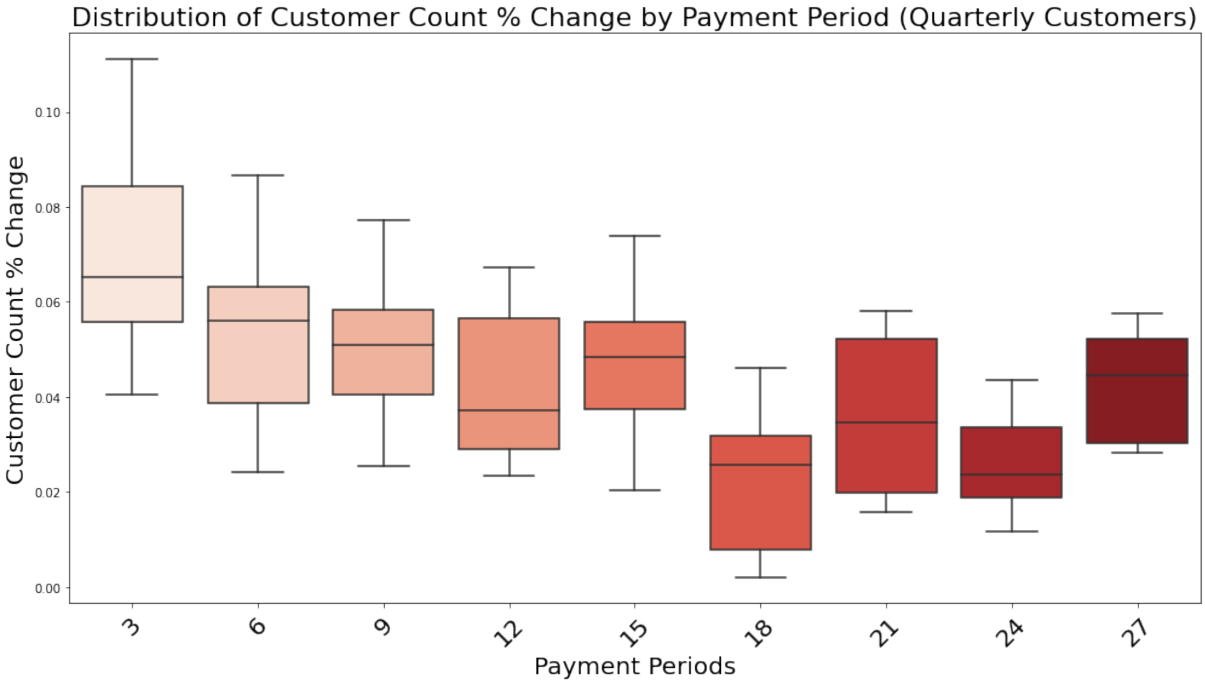

Figure 13: Quarterly Cohort Size Churns

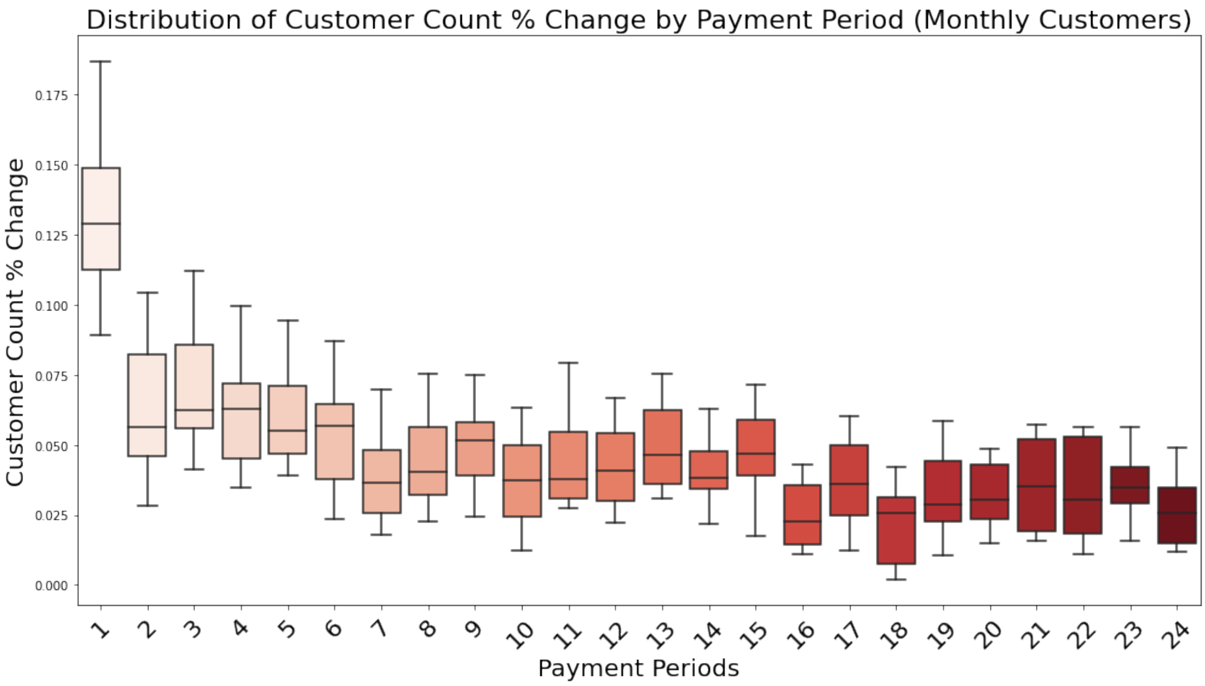

Figure 14: Monthly Cohort Size Churns

From the churn plots, we can observe that churn rates tend to follow a pattern over time. In the initial periods, churn rates are generally higher, as customers are still deciding whether the product meets their needs. As the months progress, the churn rate decreases as engaged customers tend to stay longer. The stabilization phase shows that churn plateaus after a certain point, indicating a more loyal customer base. This pattern holds true across both monthly and quarterly customers, although quarterly customers exhibit lower churn values and lower changes in churn values throughout.

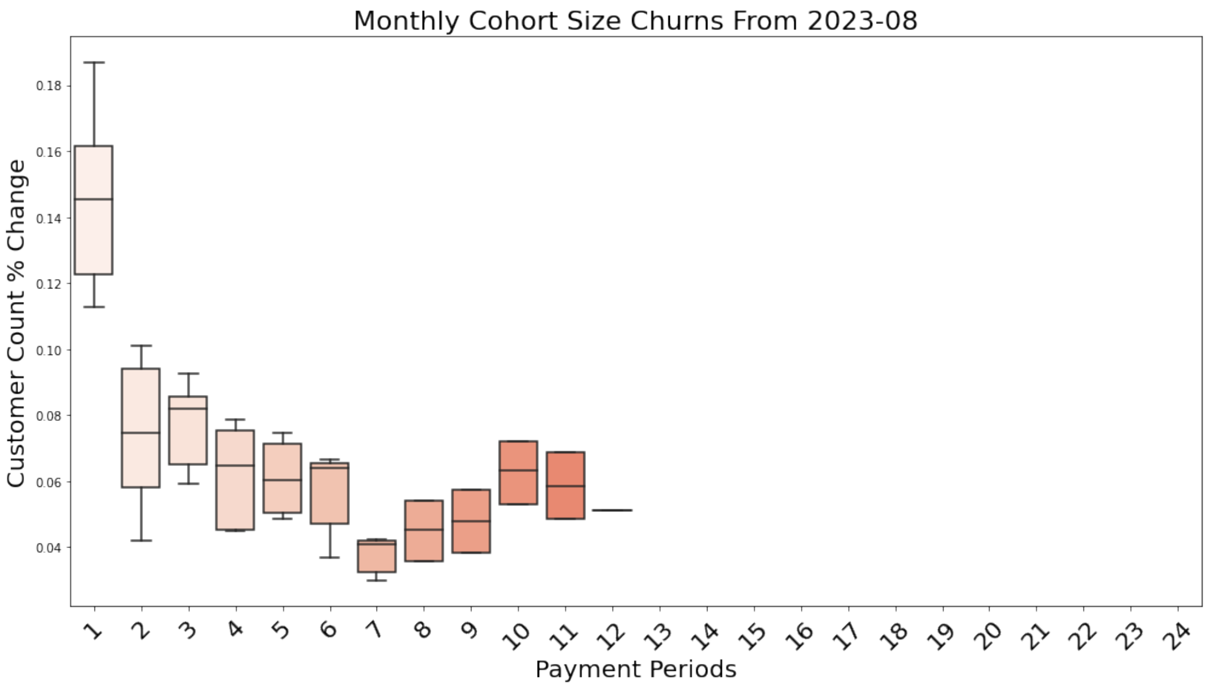

Figure 15: Monthly Cohort Size Churns - 2023-08 onwards

Above I include monthly cohort churns for cohorts starting in 2023-08 (The most recent year). During that year, the amount spend on marketing increased to historically high levels and the new subscription prices were introduced. I include this here to highlight how the churns of the first three payment periods seem to be affected by these factors more than the churns of the rest of the payment periods. This can be seen by comparing the median prices per payment period between the two Monthly Churn plots.

With the above insights in mind, we assume that customers go through different phases in their lifecycle:

1. **Onboarding (Months 0-3: C[1], C[2], C[3]):**
   This is the period where customers assess whether the product meets their needs and expectations. Churn during these months is likely influenced by factors such as subscription pricing and marketing spend. Customers are still deciding whether to commit to the product. Therefore, churn rates in these periods are more sensitive to external factors like marketing effectiveness.

2. **Engagement and Value Realization (Months 4-6: C[4], C[5], C[6]):**
   By months 4 to 6, customers who remain are more likely to have experienced the value of the product or service. These customers tend to integrate the product into their routine and are more likely to stay, leading to lower churn rates.

3. **Stabilization (Months 7+: C[7]+):**
   At this stage, customers have established routines around the product, and churn rates tend to stabilize. External factors become less influential, and historical performance of similar cohorts can provide better insights into predicting churn for these periods.

Before moving on to churn selection for projections and actually projecting the cohorts, I want to further analyze the effect of the subscription increase on our existing cohorts.

4.4 Effect of Subscription Increase

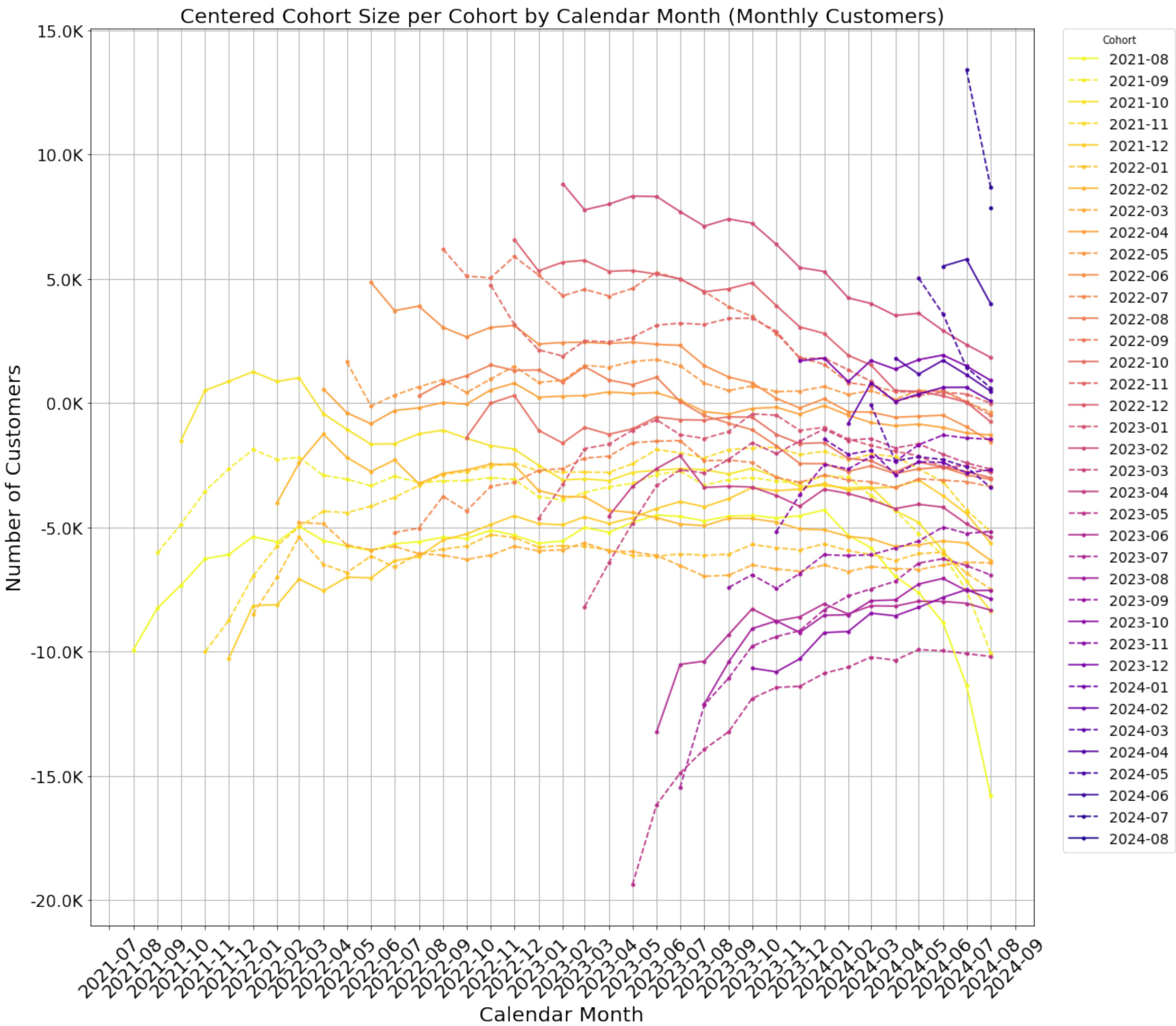

Figure 16: Centered Monthly Cohort Size - Calendar Format

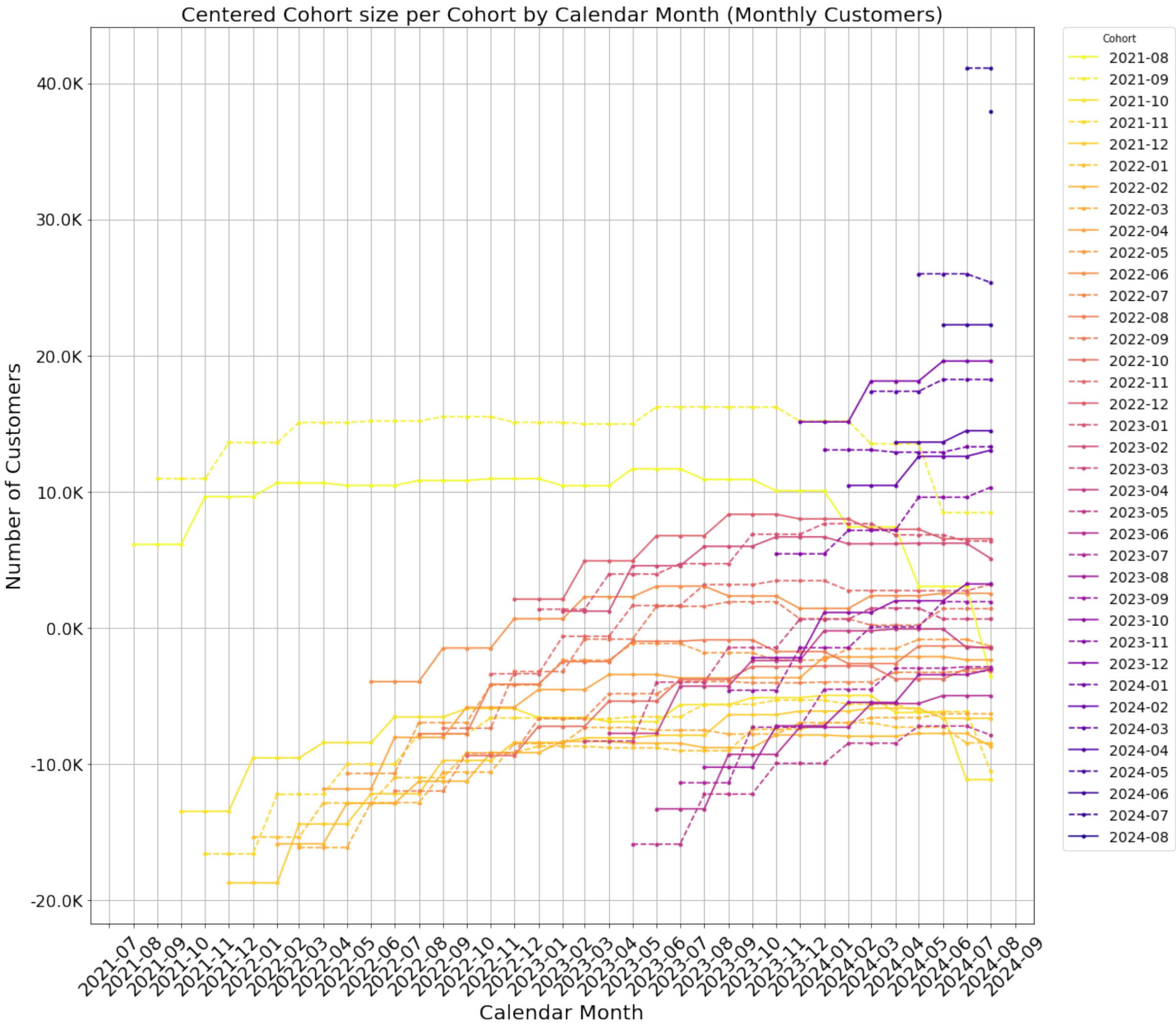

Figure 17: Centered Quarterly Cohort Size - Calendar Format

The above two plots were created as follows:

1. Center the cohort size by payment period data (subtract the mean of each column from the column).

2. Convert the rolling data frame to a calendar data frame. Instead of payment periods, the columns are now calendar months.

To analyze the effect of the subscription increase we need to view the cohort sizes by calendar months and observe unusual drops in cohort sizes after the month of the price increase (2024-01). However, different cohorts will be at different payment periods at month 2024-01. The effect that the payment period has in cohort size decrease will make it hard to draw conclusions about the effect of the subscription change. For example, cohort 2023-10 will have a much larger drop in size at month 2024-01 compared to cohort 2021-08 simply because it is at the first month of it's life when customers tend to leave at much higher rates. To remove part of this effect from our data we center it first. The centered rolling data has zero mean for each payment period. Therefore, the observed decrease in cohort size from payment period to the payment period is partially removed. Now we can turn the rolling data into calendar data and plot.

It is surprising to see that most cohorts seem to be unresponsive to the subscription increase. From the plots, we can tell that this change only affected a few of the oldest cohorts that started losing a lot more (than usual) number of customers after month 2024-01. This is promising for MockCompany as we can expect cohorts after the increase to behave similarly to cohorts before the increase.

5 Ad hoc Analysis - Churn Selection

5.1 Churn Selection using Historical Data

5.1.1 Problem Definition

In this section, we propose a framework to estimate churn rates. **C[M]** indicates the churn rate for a specific period. For example, **C[1]** represents the churn rate for the first period (the percentage of customers who were present in payment period 0 but left in payment period 1).

Our goal is to estimate **C[M]*** for each period by taking into account the different lifecycles of customers.

5.1.2 Methodology

To predict future churn rates, we analyze historical cohort data to estimate churn for each period (e.g., $C[1]^*$, $C[2]^*$, etc.). For the first 6 periods, we compute distinct churn rates. From period 7 onwards, we assume a constant churn rate.

For quarterly customers, a similar approach is used, where churn is computed for periods $C[3]^*$ and $C[6]^*$, with a constant rate applied starting from $C[9]^*$.

Churn projections for each period are based on the median, 75th, and 90th percentiles, calculated from historical data to allow us to make predictions for churn in medium, conservative, and very conservative cases. The data window for each period is adjusted as follows:

- **C[1]***: Recent data is used to capture immediate factors like marketing efforts or product updates, as churn in the initial period is highly sensitive to these influences.
- **C[2]*** **- C[3]***: A slightly broader time window is applied, balancing recent trends with additional historical data to ensure more reliable churn estimates as customer behavior starts to stabilize.
- **C[4]*** **- C[6]***: An expanded time frame is used to capture long-term engagement patterns. Churn in these periods reflects customers who have likely realized value from the product and are less influenced by short-term factors.
- **C[7]*****+**: For periods after 6, we average the churn rates from historical data beyond month 7. Once this average is calculated, we compute key percentiles (e.g., median, 75th, and 90th percentile) to project future churn.

In practice, $C[1]$ will only be unknown for a cohort that has just started and for which we don't have data. Similarly, $C[2]$ will be unknown for the last two cohorts, and so on.

It's important to note that this methodology can be applied to any company. However, it is bound to the recency of the data—we will need to re-run the analysis as new data becomes available to compute updated estimates for churn rates. Since the churn rates we compute to predict future performance heavily depend on recent data, this model cannot be static.

5.1.3 Churn Selection for the Projection Section

In this section, we describe the process we follow to select churn rates for the projection section. Our goal is to predict cohort performance starting from January 2023 until the payment period 30. Our last data point is August 2024.

Below is the matrix that shows the churn periods (C) for each cohort:

$$\begin{array}{ccccccccc}
\text{Cohort Jan 2023} & C[1] & C[2] & \cdots & C[18] & C[19] & C[20]^* & \cdots & C[30]^* \\
\text{Cohort Feb 2023} & C[1] & C[2] & \cdots & C[18] & C[19]^* & C[20]^* & \cdots & C[30]^* \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\text{Cohort July 2024} & C[1] & C[2]^* & \cdots & C[18]^* & C[19]^* & C[20]^* & \cdots & C[30]^* \\
\text{Cohort August 2024} & C[1]^* & C[2]^* & \cdots & C[18]^* & C[19]^* & C[20]^* & \cdots & C[30]^*
\end{array}$$

In this matrix, the values with C^* represent the periods for which we need to estimate churn rates using historical data. The earlier churn periods (e.g., $C[1]$, $C[2]$) for early cohorts are already known from actual data, and we do not need to predict them. However, for cohorts that begin later, even the early churn periods need to be estimated.

Given that we are conducting our analysis separately for monthly and quarterly customers, for quarterly customer predictions we will use $C[3]^*$, $C[6]^*$, and subsequent periods (up to $C[30]^*$).

We used the following data windows to estimate the churns:

For Monthly Customers:

- **C[1]***: August, 2023 – August, 2024
- **C[2]***: July, 2023 – August, 2024
- **C[3]***: June, 2023 – August, 2024
- **C[4]***: January, 2023 – August, 2024
- **C[5]***: January, 2023 – August, 2024
- **C[6]***: January, 2023 – August, 2024
- **C[7]+***: January, 2023 – August, 2024

For Quarterly Customers:

- **C[3]***: June, 2023 – August, 2024
- **C[6]***: January, 2023 – August, 2024
- **C[9]+***: June, 2022 – August, 2024

The values estimated for churn rates based on the median (50%), conservative (75%), and very conservative rates (90%) for both monthly and quarterly customers are as follows. Note that we rounded up values in some cases to be slightly more conservative.

For quick intuition, in the case of $C[1]^*$ under the median scenario, we used the time window from August 1, 2023 to August 1, 2024. We looked the actual values of churn rates for $C[1]$ within that window and selected the median value as our estimate.

Period	Median Churn	Conservative Churn	Very Conservative Churn
C[1]*	0.15	0.165	0.19
C[2]*	0.075	0.095	0.15
C[3]*	0.08	0.09	0.11
C[4]*	0.065	0.075	0.095
C[5]*	0.065	0.075	0.09
C[6]*	0.06	0.065	0.07
C[7+]*	0.05	0.06	0.07

Table 1: Monthly Churn Rates for Median, Conservative, and Very Conservative Cases (with * indicating estimated values)

Period	Median Churn	Conservative Churn	Very Conservative Churn
C[3]*	0.075	0.085	0.095
C[6]*	0.06	0.065	0.07
C[9+]*	0.05	0.06	0.065

Table 2: Quarterly Churn Rates for Median, Conservative, and Very Conservative Cases (with * indicating estimated values)

6 Projections

In this section, we provide predictions for the M30 values for each of the MockCompany cohorts beginning in 2023. We make these predictions under three different cases: a Median case, a Conservative case, and a Very Conservative case.

6.1 Methodology

For the three scenarios (Median, Conservative, and Very Conservative), we define churn rates that predict customer retention based on historical data (as explained in 5.1.3). These rates estimate how many customers will churn after each payment period. The aim is to use these churn rates to project cohort payment data/ cohort sizes for future months where no data is available.

The process starts by identifying the last month with available payment data for each cohort, which becomes the starting point for the projection. We then adjust the churn rates to align with the missing months, ensuring the correct churn rate is applied to the first month without data.

The extrapolation applies these churn rates to the last known cohort payment/size. For each subsequent month, we calculate the percentage of retained customers and multiply this retention rate cumulatively to model the cohort size over time. This generates a set of projected values for the missing months, providing a forecast of future cohort payments/sizes under each scenario.

6.2 Projection Results

6.2.1 ROAS Projections

The following figures present the projected ROAS for all cohort starting from January 2023 under three different scenarios: Very Conservative, Conservative, and Median cases.

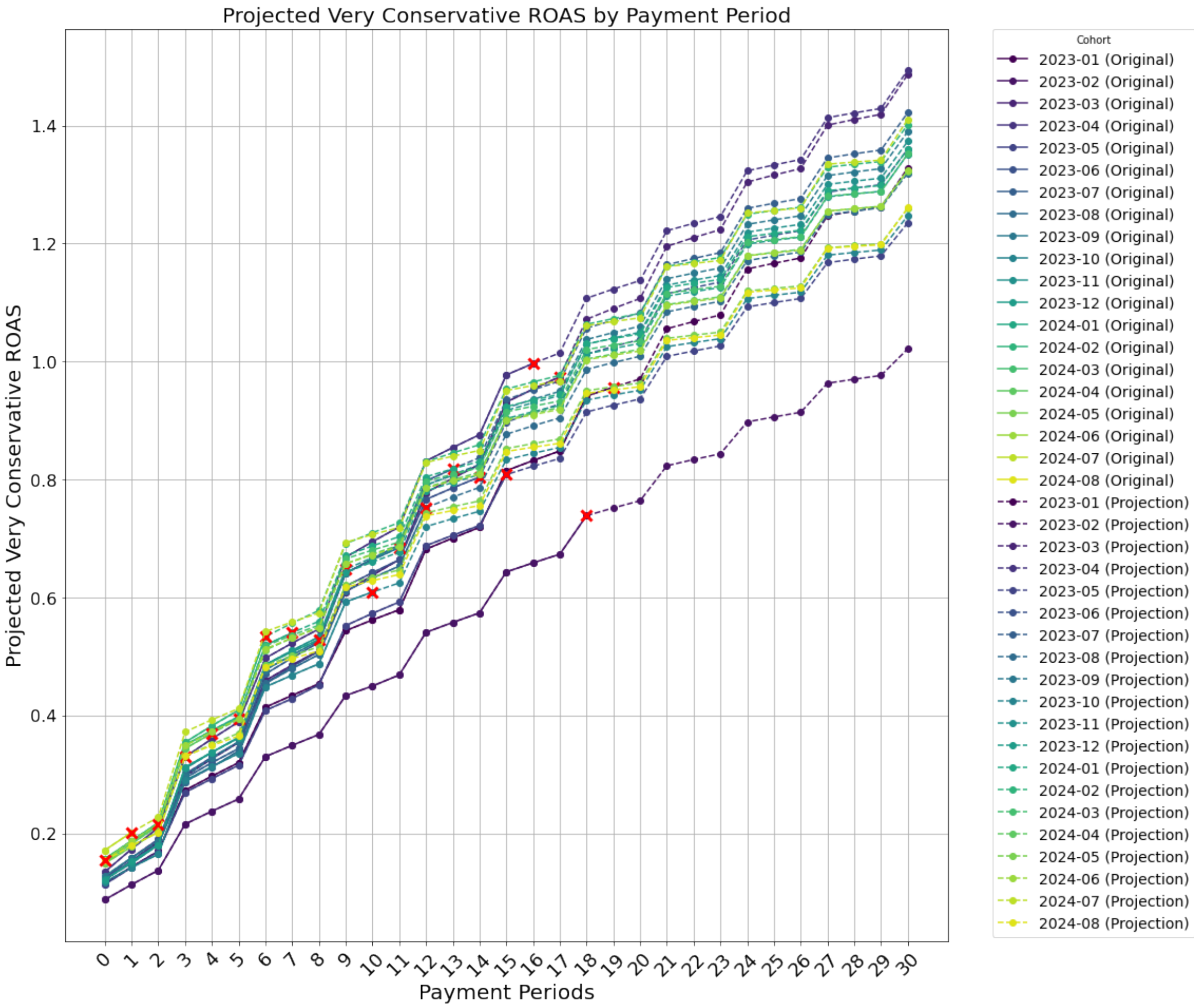

Figure 18: Projected Very Conservative ROAS by Payment Period

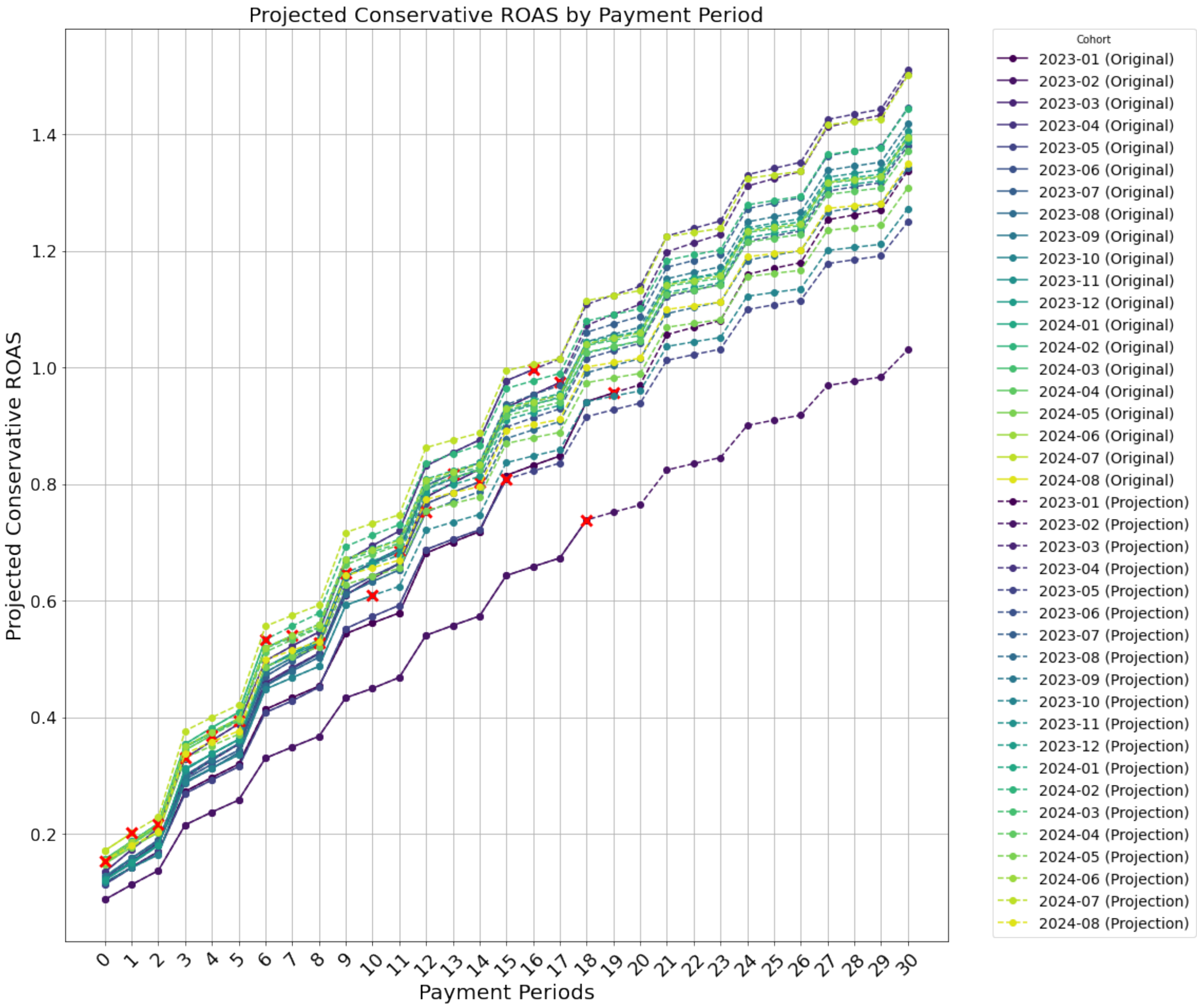

Figure 19: Projected Conservative ROAS by Payment Period

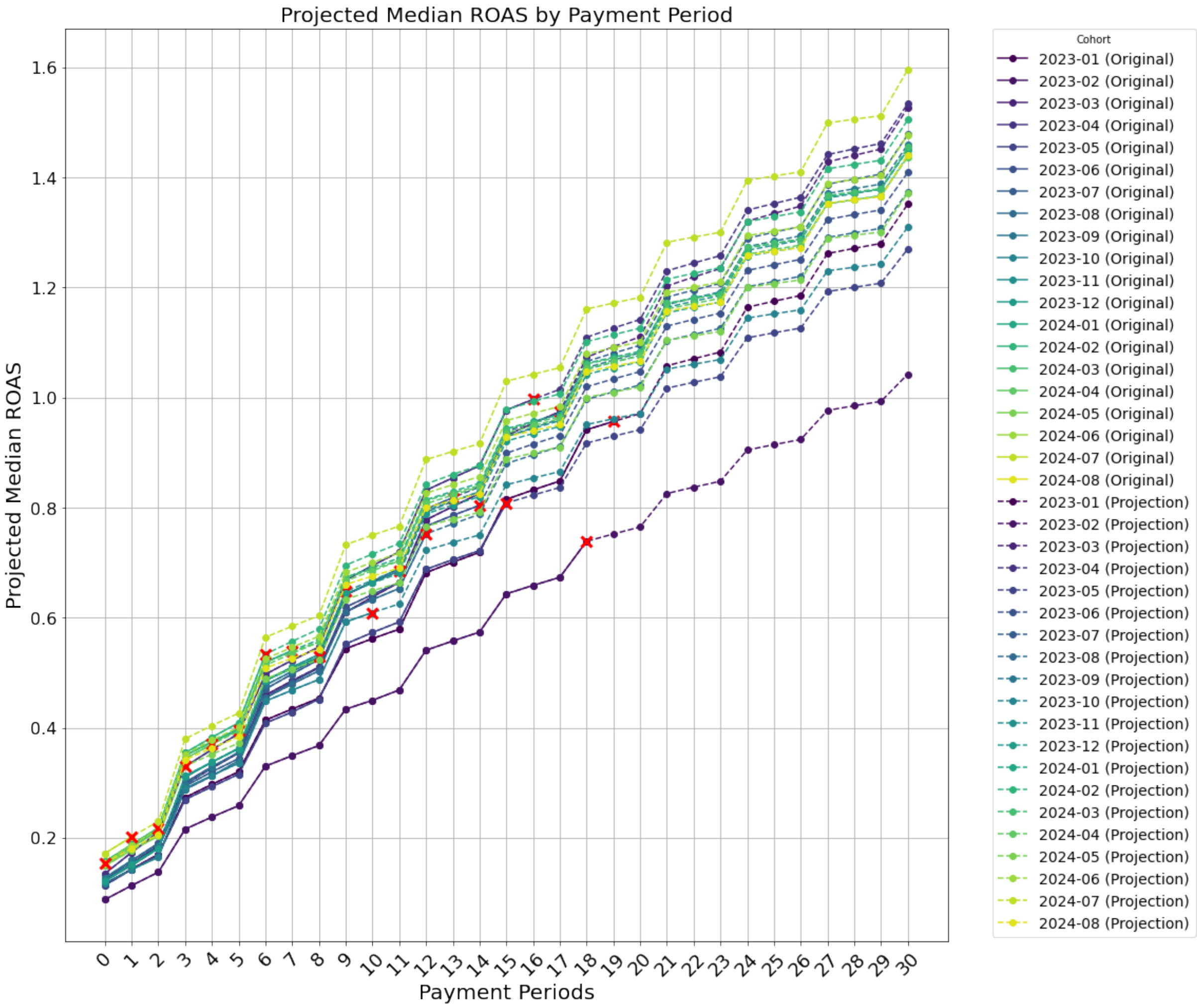

Figure 20: Projected Median ROAS by Payment Period

In the plots above solid lines are existing data, dashed lines present our projections and red x marks signify the start of the projections

As expected, variation between our different projected scenarios become more evident as the projected number of payment periods increases. To better understand the variation of our predictions we present a plot showing the difference between our very conservative and our median projection.

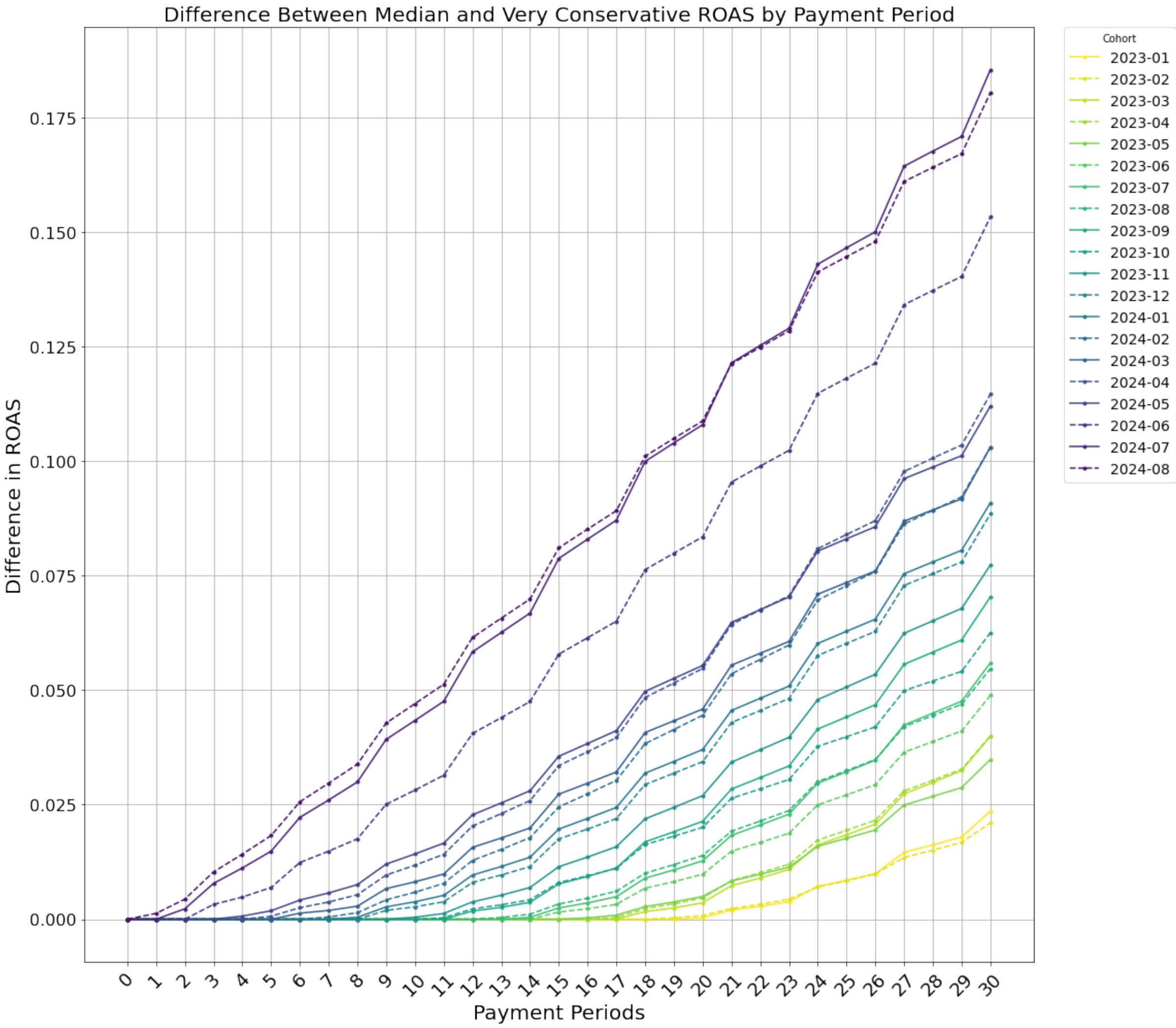

Figure 21: Median and Very Conservative projection difference

In the notebook, we have included additional projections for: cohort size by payment period, cohort size by payment period for monthly customers only, cohort size by payment period for quarterly customers only and ROAS projections as if the subscription prices were always at the increased values.

6.3 Projection Results on Existing Data Experiment

To evaluate the accuracy (or potential inaccuracy!) of our projections, we perform a backtest using existing data.

We consider cohorts starting from August 1, 2022, through August 2023, assuming that no data is available beyond this period. Additionally, we predict as far into the future as possible to ensure that we have complete cohort data without missing values (e.g., the August 2023 cohort only has data until August 2024). This results in a 12-month prediction period.

For the sake of this experiment, we predict monthly cohort sizes for 12 periods and compare the predictions with the actual values for the median and conservative case.

First, we collect the actual cohort size data (without missing values) and mask the bottom half of the prediction. Then, we estimate the churn rates for the median case using updated date ranges. The table below summarizes the date ranges used, as well as the median churn rate estimates for each period.

Churn Period	Data Window	Median Churn Rate
C[1]*	August 1, 2022 – August 1, 2023	0.1164
C[2]*	July 1, 2022 – August 1, 2023	0.0560
C[3]*	June 1, 2022 – August 1, 2023	0.0662
C[4]*	January 1, 2022 – August 1, 2023	0.0628
C[5]*	January 1, 2022 – August 1, 2023	0.0550
C[6]*	January 1, 2022 – August 1, 2023	0.0518
C[7+]*	January 1, 2022 – August 1, 2023	0.0394

Table 3: Churn Periods, Data Windows, and Median Churn Rates

Next, we compare the projected and actual cohort sizes using the results from our backtest. The plot below illustrates the differences between projected and actual cohort sizes across various payment periods.

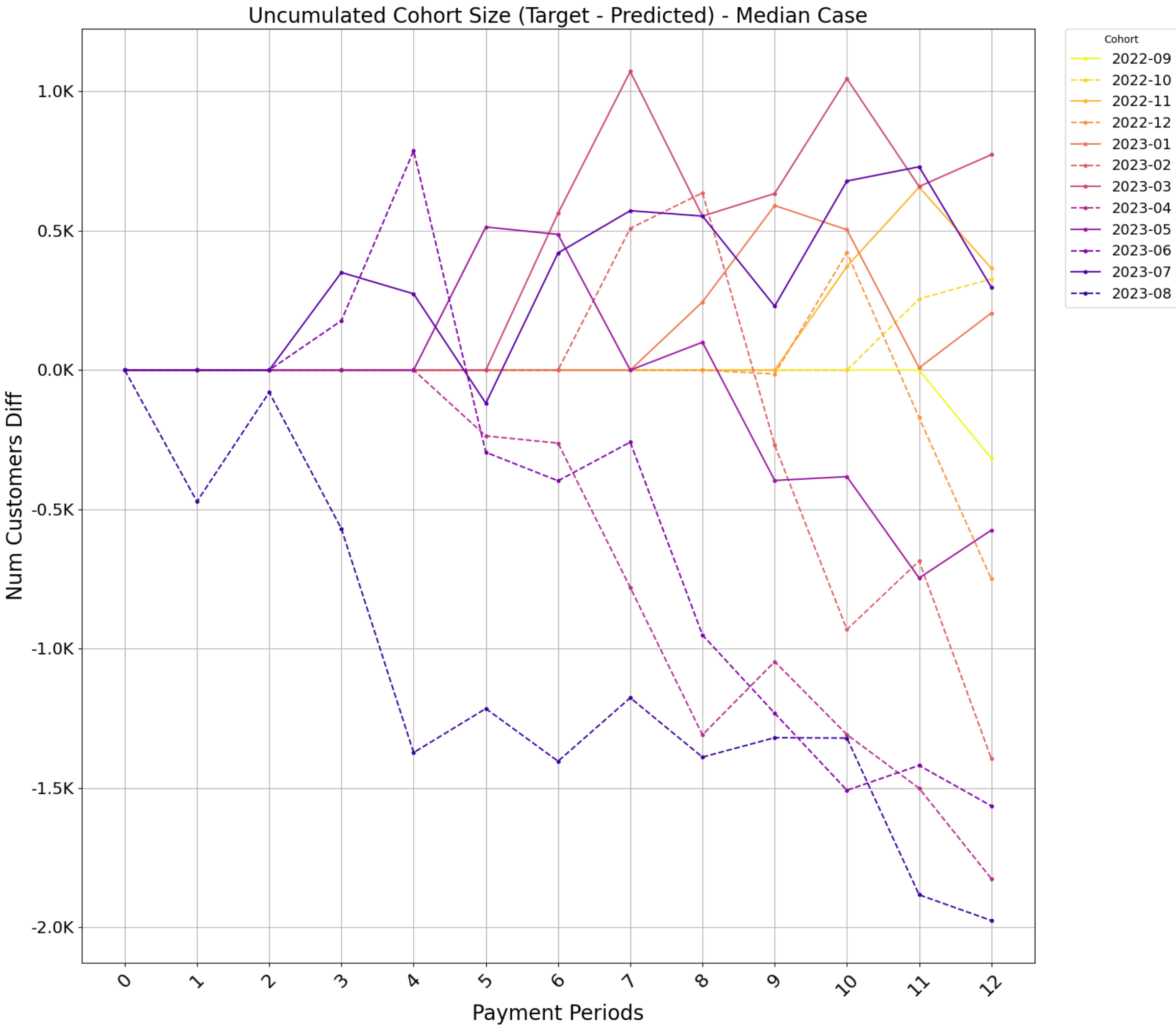

Figure 22: Uncumulated Cohort Size Prediction Difference Median Case

In the plot shown in Figure 22, we are visualizing the difference between the actual target values and the predicted values (target minus predicted). We observe that our predictions oscillate both below and above the target values, which is expected given that we are using median churn rates for this prediction. As anticipated, more recent cohorts (for which we had to estimate all churn rates) show larger deviations from the target values, such as the cohort from August 2023. Surprisingly, we achieved a decent prediction for the July 2023 cohort, despite having to estimate churn rates from period 2 through 12.

For most cohorts, we seem to be making reasonable predictions as the errors tend to fluctuate around the actual values over time. However, for certain cohorts, particularly the June 2023, April 2023, and August 2023 cohorts, our predictions overestimate the number of customers. This suggests that our estimated churn rates are lower than the actual churn rates for these cohorts.

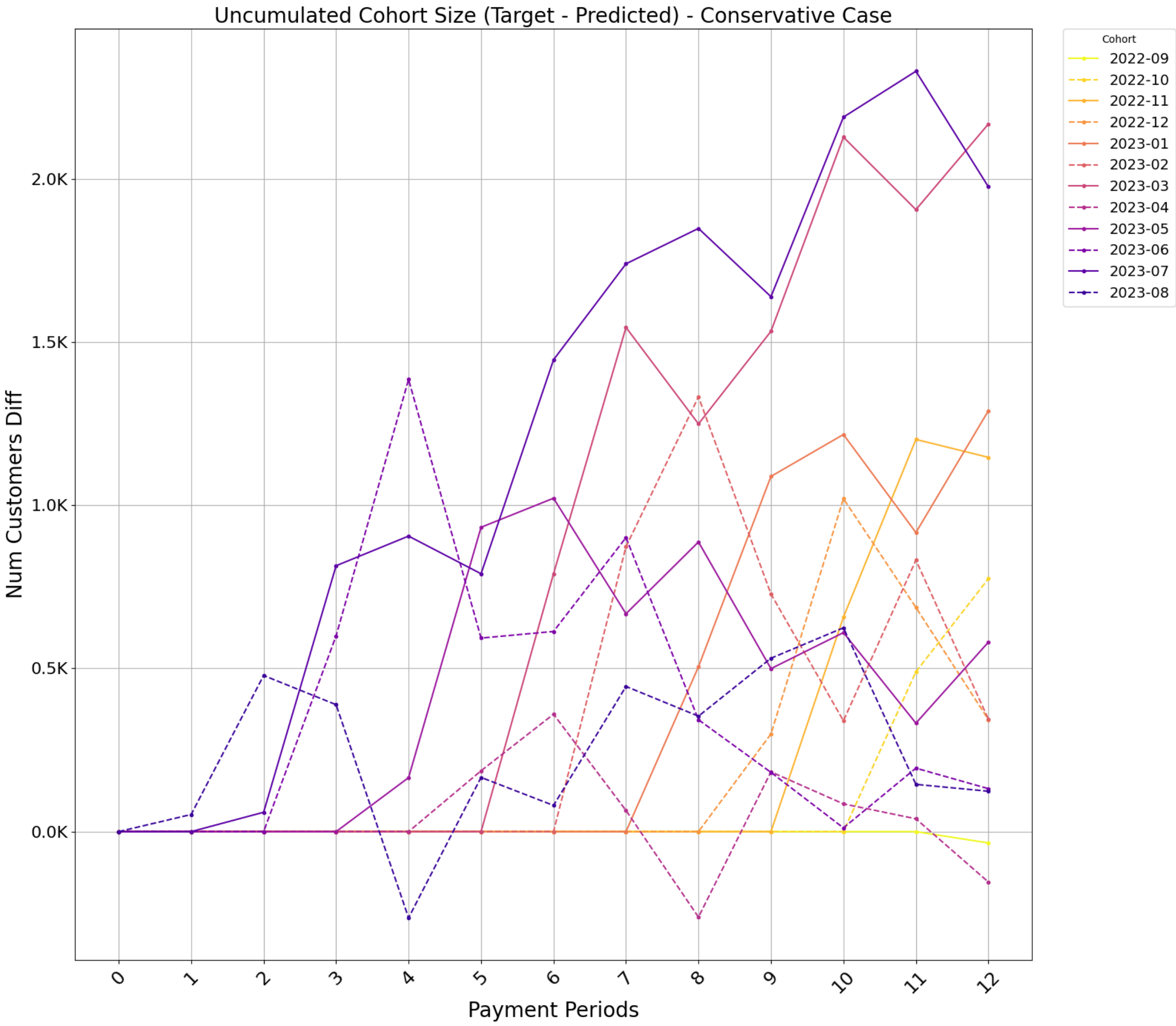

Figure 23: Uncumulated Cohort Size Prediction Difference Conservative Churns

We also make a prediction using the conservative churn estimates, and the results are shown in Figure 23. As expected, since we used more conservative churn values, our predictions consistently estimate fewer customers than there actually are (i.e., the actual number of customers is higher than our predicted values). This is consistent with the conservative approach, which assumes higher churn rates and therefore underestimates customer retention.

7 Parameter Choice

7.1 Thresholds

I decided to run a set of simulations to help me determine threshold values for each payment period. The simulations, project cohort ROAS fro differ M0 values up to M30. The goal is to find simulated cohorts with an M30 value at or just above 1.2.

I conducted two types of simulations with each type having two sub-categories:

- **Simulation A:** Cohort ROAS projections
  - Using conservative churns as selected in above section.

  - Using very conservative churns as selected in above section.

- **Simulation B:** Cohort ROAS projections with the option to collect 100% of revenue after a threshold has been violated.
  - Using conservative churns as selected in above section.
  - Using very conservative churns as selected in above section.

Before introducing the selected thresholds, I will describe the two simulation processes.
Simulation A Description:

- Simulate and project cohorts for multiple M0 values.
- Select the cohort with an M30 value just above or exactly at 1.2.
- Store the M[N] and Delta M[N] values of the selected cohort to aid with threshold selection

Note that if there is no increase in the collected amount, the selected cohort will provide a threshold guide for all M[N] values. If any M[i] value (i $\leq$ 30) was lower, the resulting M[30] would be less than 1.2 (Assuming churns remain constant).

Simulation B Description:

Unlike the previous type of simulation, the cohort with an M0 and churns that hits M[30] just at or above 1.2 is not enough to determine all M[N] thresholds. In fact, it is only able to give us insight on the M[0] threshold. To better understand this statement assume that we selected all thresholds based on this selected cohort. Then, If M0 was stressed (but not breached) we wouldn't collect 100% of the revenue from payment period 0 and we would fall short of 1.2 for M[30].

Therefore, once we select a threshold for M[i], we need to create a cohort that stresses M[i] and select a threshold for M[i+1] such that if 100% of revenue was collected from payment period i + 1 onwards we would still hit M[30] = 1.2.

In the simulation, I only repeat this process for i in [0, 3, 6, 9, 12]. These are the most important five payment periods as they are the earliest five that include payments from monthly and quarterly customers

Note that the complete process is included in the python notebook if further information is needed.

Introducing the selected thresholds:

After reviewing the results of both simulations, each for the conservative and very conservative case, I decided on the following guidelines for threshold selection.

Thresholds for payment periods 0-3 are represented as M0-M3 values. Payment period 4+ thresholds are set to Delta M[i] values.

Each threshold, is set to the Max(Conservative, very conservative) simulation result and the values are rounded and presented as percentages with .5% precision

Thresholds for quarterly payment periods are generally more important than the rest because revenue from both monthly and quarterly customers is present.

Thresholds for payment periods 12+ are not so important because for a 12-month investment period the corresponding violations will be breached after the investment period ends.

Let V[i] be the importance of the threshold violation in payment period i. Then

$$V0 > V1 > V2 > V3 > V6 > V9 > V12 > V4 > V5 > V7 > V8.$$

In general, violations in later payment periods are less significant because there are likely younger cohorts that would have trigger a threshold violation if the issue had been caused by an event with a broader impact.

Given the above considerations we propose the following thresholds:

- M[0] $\leq$ 14,5
- M[1] $\leq$ 17,5
- M[2] $\leq$ 20
- M[3] $\leq$ 32,5
- Delta M[6, 9, 12] $\leq$ 10%
- Delta M[4, 5, 7, 8, 10, 11] $\leq$2%

For the case where we collect 100% of the cohorts revenue after the violation we propose the following thresholds:

- M[0] $\leq$ 11,5
- M[1] $\leq$ 14
- M[2] $\leq$ 16,5
- M[3] $\leq$ 26,5
- Delta M[6, 9, 12] $\leq$ 10%
- Delta M[4, 5, 7, 8, 10, 11] $\leq$2%
- Delta[12, 15] $\leq$8%
- Delta[18, 21, 24, 27, 30] $\leq$7%

Notice that for this case, threshold for payment periods greater than 12 are also relevant since if breached, they will trigger the increased collection of revenue.

Finally, both cases, the wiggle room of our thresholds is the difference between the very conservative and conservative thresholds. With the conservative thresholds being our walkaway numbers. If needed, tests of non-quarterly payment periods could be dropped with the exception of payment periods 1 and 2. Quarterly payment periods give a stronger indication for each cohort while payment period 1 and 2 give us the ability to stop funding (or start collecting 100% of the revenue) early if we observe unreasonably high increases in initial churns rates. It is important to note that the thresholds for payment periods 1 and 2 are particularly important during the first three months of our partnership with MockCompany, until the first cohort we fund reaches its third month. After that, at any given time, there will always be a cohort that is tested against the more important payment period 3 threshold.

Below I include the simulation results used to create the thresholds up to payment period 12 with minimal processing.

Rolling month	0	1	2	3	4	5	6	7	8	9	10	11	12
M[N] Conservative (not 100%)	0.133	0.165	0.193	0.306									
Delta M[N] Conservative (not 100%)					0.02	0.02							
M[N] Conservative with 100%	0.106	0.131	0.155	0.248									
Delta M[N] Conservative 100%					0.02	0.02	0.09	0.02	0.02	0.08	0.01	0.01	0.08
M[N] Very Conservative (not 100%)	0.144	0.177	0.206	0.324									
Delta M[N] very Conservative (not 100%)					0.02	0.02	0.11	0.02	0.02	0.10	0.01	0.01	0.10
Range M[N] (not 100%)	0.133-0.144	0.165-0.177	0.193-0.206	0.306-0.324									
Range Delta M[N] (not 100%)					0.02-0.03	0.02-0.03	0.1-0.11	0.02	0.02	0.1	0.01-0.02	0.01-0.02	0.10
Range M[N] (with 100%)	0.106-0.115	0.131-0.142	0.155-0.164	0.248-0.263									
Range Delta M[N] (with 100%)					0.02	0.02	0.09	0.02	0.01-0.02	0.08-0.09	0.01	0.01	0.08

Table 4: ROAS and Delta values with and without 100% increase

7.2 Company Agreeability

Cohort	**0**	**1**	**2**	**3**	**4**	**5**	**6**
2024-01-01	0.156162	0.187631	0.217012	0.349419	0.374479	0.397938	0.519093
2024-02-01	0.152359	0.186808	0.218314	0.354395	0.382479	0.408966	0.533476
2024-03-01	0.157103	0.186432	0.213809	0.350152	0.373518	0.394677	NaN
2024-04-01	0.150239	0.182464	0.213076	0.343873	0.369772	NaN	NaN
2024-05-01	0.149052	0.177540	0.202769	0.330380	NaN	NaN	NaN
2024-06-01	0.154403	0.186909	0.216161	NaN	NaN	NaN	NaN
2024-07-01	0.171531	0.201812	NaN	NaN	NaN	NaN	NaN
2024-08-01	0.153833	NaN	NaN	NaN	NaN	NaN	NaN

Table 5: M[N] by Payment Periods for 2024+ cohorts

Cohort	**0**	**1**	**2**	**3**	**4**	**5**	**6**
2024-01-01	0.156162	0.031469	0.029381	0.132407	0.025060	0.023459	0.121155
2024-02-01	0.152359	0.034449	0.031506	0.136081	0.028084	0.026487	0.124510
2024-03-01	0.157103	0.029329	0.027376	0.136343	0.023365	0.021159	NaN
2024-04-01	0.150239	0.032225	0.030612	0.130797	0.025899	NaN	NaN
2024-05-01	0.149052	0.028488	0.025229	0.127611	NaN	NaN	NaN
2024-06-01	0.154403	0.032505	0.029253	NaN	NaN	NaN	NaN
2024-07-01	0.171531	0.030281	NaN	NaN	NaN	NaN	NaN
2024-08-01	0.153833	NaN	NaN	NaN	NaN	NaN	NaN

Table 6: Delta M[N] by Payment Periods for 2024+ cohorts

Thresholds are tested on every cohort starting from January 2024 onwards. The above tables show the M[N] and Delta M[N] values for these cohorts. All presented cohorts are above the proposed thresholds.

We selected cohorts starting from January 2024 because they are the only ones that had the increased subscription prices from the start of their lifetime.

If our cohort selection pool is considered too small by MockCompany, we will express the equivalent thresholds in terms of customer attrition rates that should not be exceeded for each relevant payment period. By doing that, the price increase will not reflect poorly on older cohorts that experienced it in later periods of their lifetime. In this case, we would also exempt cohort 2023-02 from all threshold tests as we consider it unrepresentative due to increase in brand spending for that month.

Lastly, we propose an additional condition: brand spending should not exceed 20% of the total budget. This is intended to prevent a repeat of poor performance caused by overspending on brand marketing. Additionally, since MockCompany is obligated to allocate 20% of the total spend, they are free to use the full amount on brand marketing if they choose.