Effects of subscription programme

Please note that the analysis below was ran on a simulated dataset and should be only considered in the context of the portfolio.

Question: What is the impact of joining the subscription programme on a business that cares about customer ordering frequency and value?

Executive summary

- The subscription plan significantly increases the frequency and value of orders in the whole sample when compared to the pre-subscription stage.
- The highest increases appear during the active subscription stage, but there are carry-over increases in the measures after cancellation, suggesting behavioural changes in the customers.
- The analysis revealed further differences between customers who retain their active subscription status and those who cancel around the end of the trial phase of the programme. Despite making up only 24 % of the whole sample the Retention group represents very valuable customers as they place significantly more valuable orders more frequently than the Churn group.
- The increases in the frequency and value of orders associated with the subscription programme translate into a clear growth of revenue. Within the Retention group we can see a linear increase in the frequency and value of orders with time, specifically when entering their fourth active subscription month.

Data analysis

One can re-frame the above question on a data analysis level in the following way: Does the monthly subscription stage affect ordering frequency and value?

All-analyses were conducted using *R* (Version 3.5.1). For the analysis script and exact test statistics please refer to the corresponding R MarkDown document.

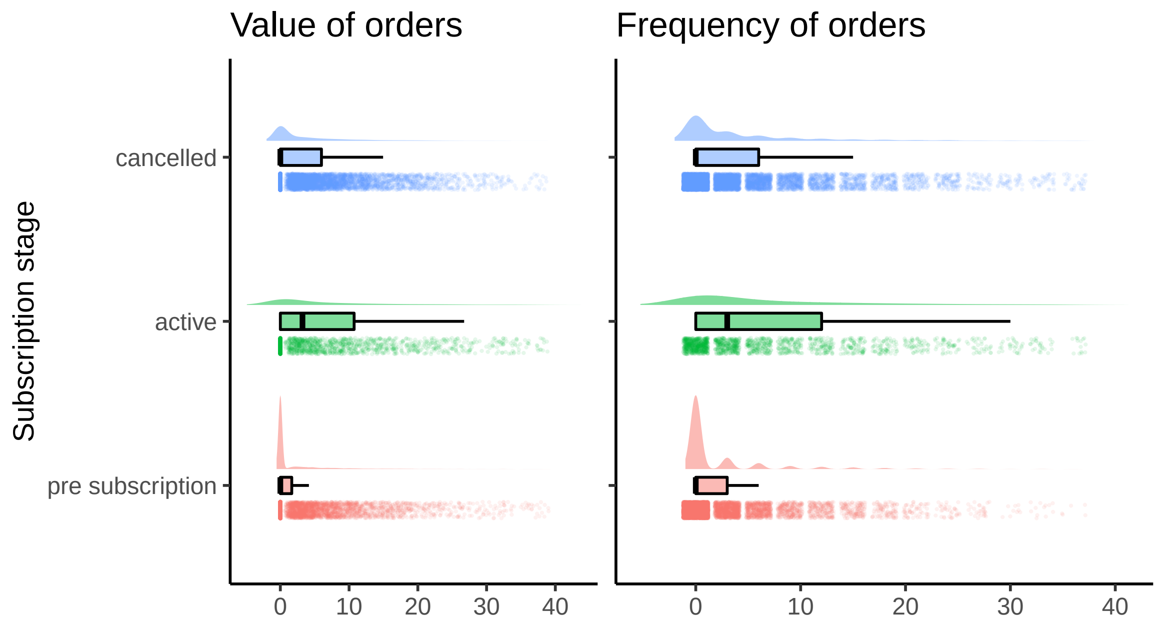
When exploring the data set, we see that it consists of different numbers of observation from N =1857 customers over 10 months of uninterrupted time-period. The descriptive statistics of the data set (see table below) reveal that on average, customers seem to increase both during an active subscription phase and after cancelling the programme. However, one needs to run test statistics to see if the different subscription states are statistically different from one another.

Table 1: Summary of average ordering frequency and value based on the monthly subscription status

Subscription status	Observations	Order frequency	SD order frequency	Order value	SD value order
pre subscription	8745	2.74	8.85	3.04	8.85
active	1982	13.00	21.13	13.38	21.13
cancelled	7945	5.43	9.23	4.99	9.23

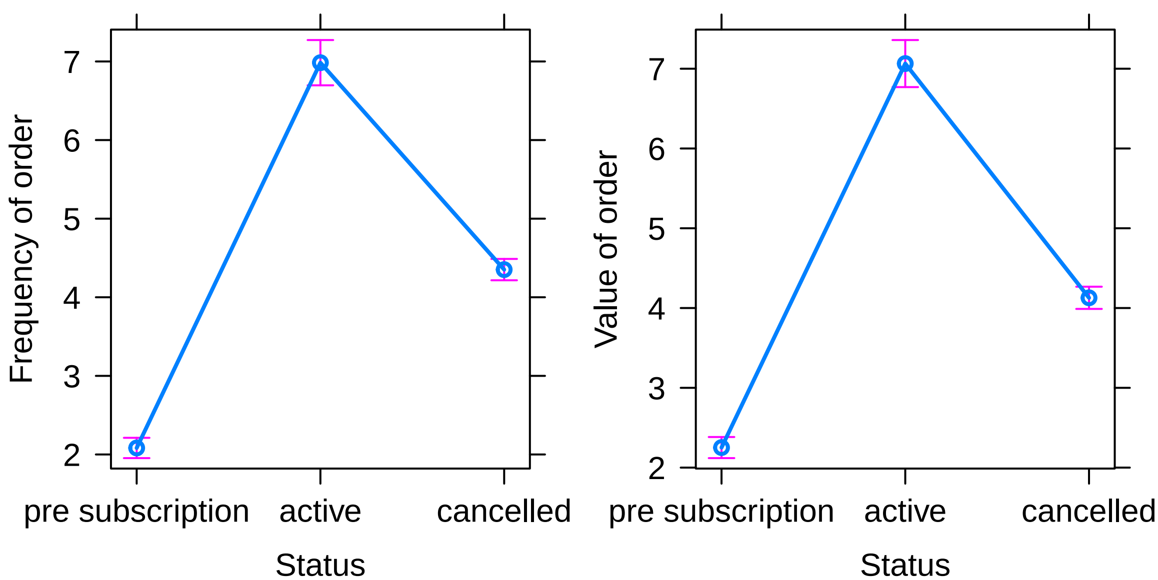
It is good practice to filter out outliers (here defined as values +/- 3 standard deviations away from the mean) before running statistical analyses, but here the data set remains the same after filtering.

Nonetheless, when plotting the distribution and individual data points with their central metrics (see figures below), it is clear that the distribution skewed (frequently the case with real-world data) and the frequency of 0 values are overrepresented in the sample. To address this, further filtering and transformations (or fitting models with non-normal distributions) could be done, but for this exercise I decided to keep the untransformed values as they are.



Given that monthly subscription state has 3 levels, linear regression was applied to see its effects on order frequency and value.

The analysis revealed a significant increase from pre-subscription stage both in the active and cancelled states for the frequency and value of the orders (see figure below). This result suggests that a subscription has facilitating effect on ordering behaviour, even after cancelling the programme. Both fitted models explained 6% of the variance for both outcome measures.



While the results seem to suggest the efficiency of the subscription programme the question arises whether this is true for all types of customers. It is possible that customers who were only interested in the two-week-long, free subscription period are different from the customers who retained their subscription.

At this stage customers were split into two groups such as the Retention group and Churn group based on the number of active monthly subscriptions. To avoid the inclusion of accidental customers who might have forgotten to cancel, only customers with more than 1 active month were included in the Retention group. The rest of the customers made up the Churn group.

The summary tables (see below) suggest that customers who had at least 1 active monthly subscription also kept their subscription for 1 other month at least. Specifically, this can be seen in Churn group where there was no-one in the Active stage throughout the 10 month period. We can see that both the frequency and the value of the average order per month increased during the active months for the Retention group. Interestingly, even after cancelling the subscription (i.e. after an active period or around the 2 week-long trial) the frequency and value seem to remain higher than in the pre-subscription period - suggesting behavioural change in the customer. Another, interesting observation is that the Retention and Churn group differ in terms of their baseline (i.e. pre-subscription). This data could be a useful starting point when preparing a strategy for targeting certain customers (but it is not under the scope of the present analysis).

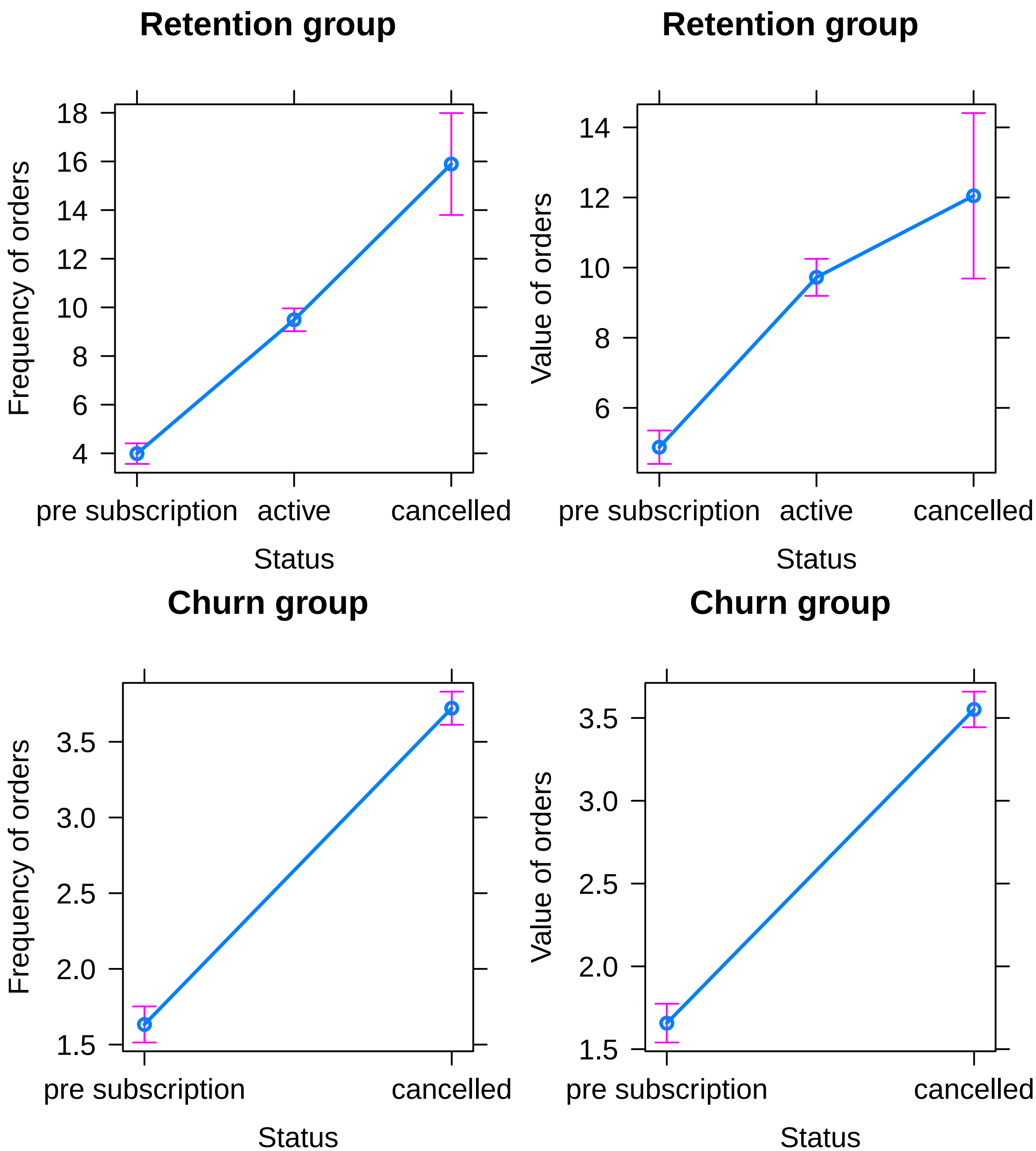
Table 2: Summary of average ordering behaviour for the Retention group

Subscription status	Observations	Order frequency	SD order frequency	Order value	SD value order
pre subscription	2307	3.99	10.55	4.88	10.55
active	1876	9.49	13.00	9.72	13.00
cancelled	94	15.89	9.63	12.05	9.63

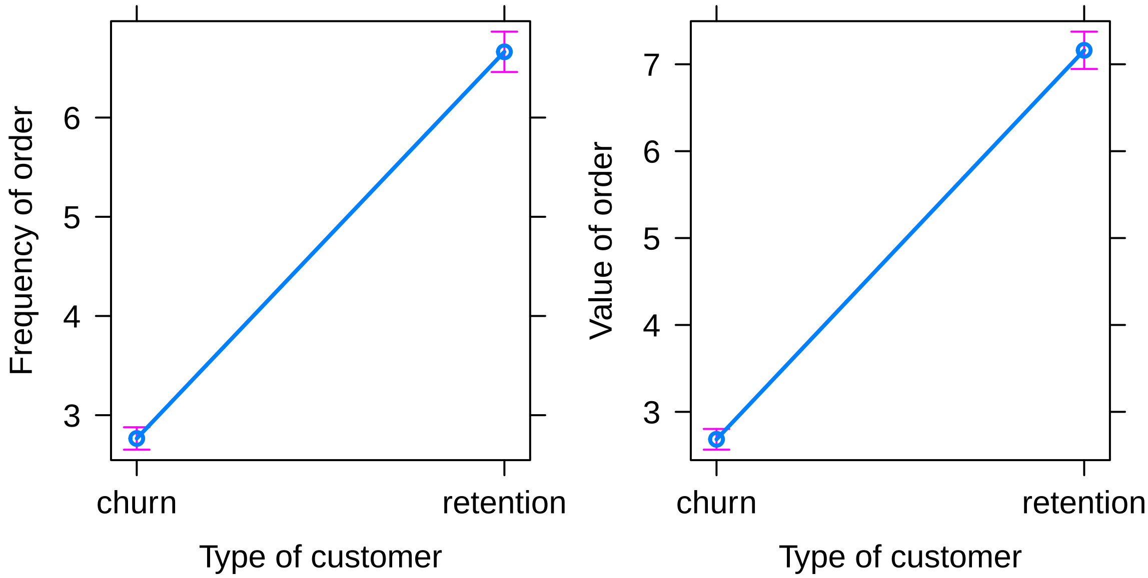
Table 3: Summary of average ordering behaviour for the Churn group

Subscription status	Observations	Order frequency	SD order frequency	Order value	SD value order
pre subscription	6348	1.63	3.88	1.66	3.88
cancelled	7519	3.72	5.38	3.55	5.38

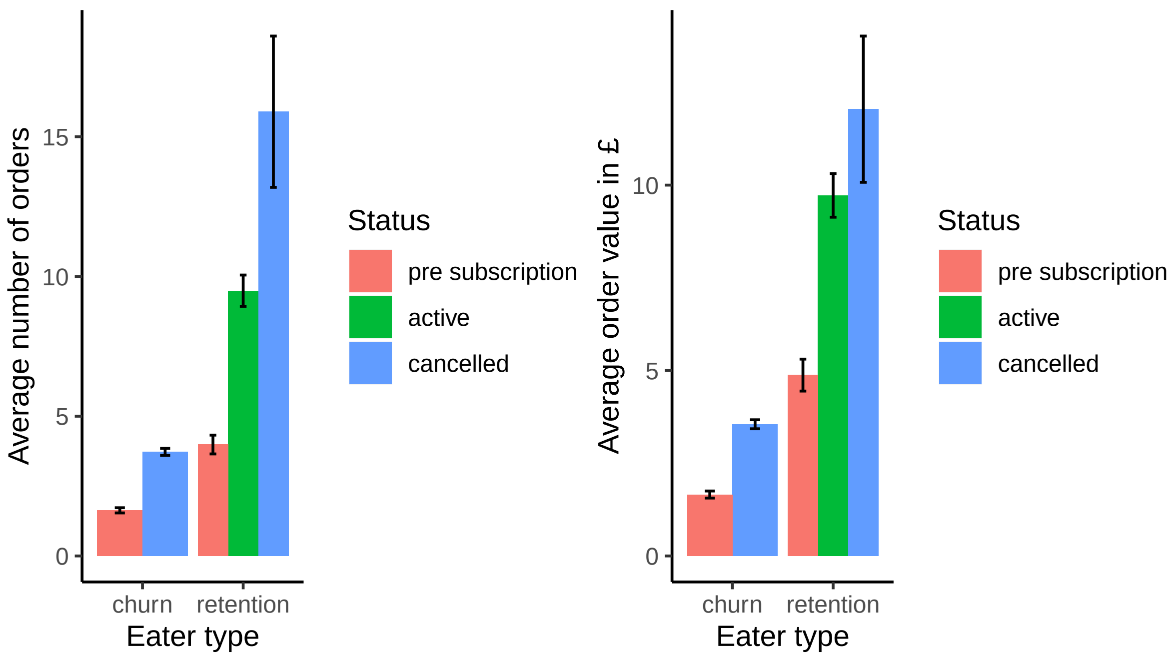
To test the effects of the monthly subscription state within the two groups separately, linear regressions were run on the data. The analysis revealed the significant increasing effect of monthly subscription state on both order value and frequency in both groups (to a higher extent in the Retention group).



To see if the two groups of customers were statistically different from one to another two linear regressions were run after combining the two groups into one, using a newly created categorical factor (i.e. group membership: Retention, Churn). Based on the results there is indeed a significant difference between these two kinds of customer at both outcome measures (see figure below).



When further investigating the two groups we can see that the majority of customers belong to the Churn group (76% of all customers). Despite being smaller, the Retention group represents very valuable customers as they place more frequent and valuable orders than the Churn group (see figure below).



We can look at the differences in terms of generated revenue between the two groups (see table below). For the calculations below we used £2.5 as the delivery fee payerd by the customer, £7.99 as the subscription fee and 20% comission rate. Based on the tables below it is clear that the subscription programme generates increased revenue during the active and cancelled states, and the Retention group customers are outstandingly valuable for the business (economically speaking).

revenue source	pre-subscription	active	cancelled

Table 4: Summary of the average monthly generated revenue in the Retention group

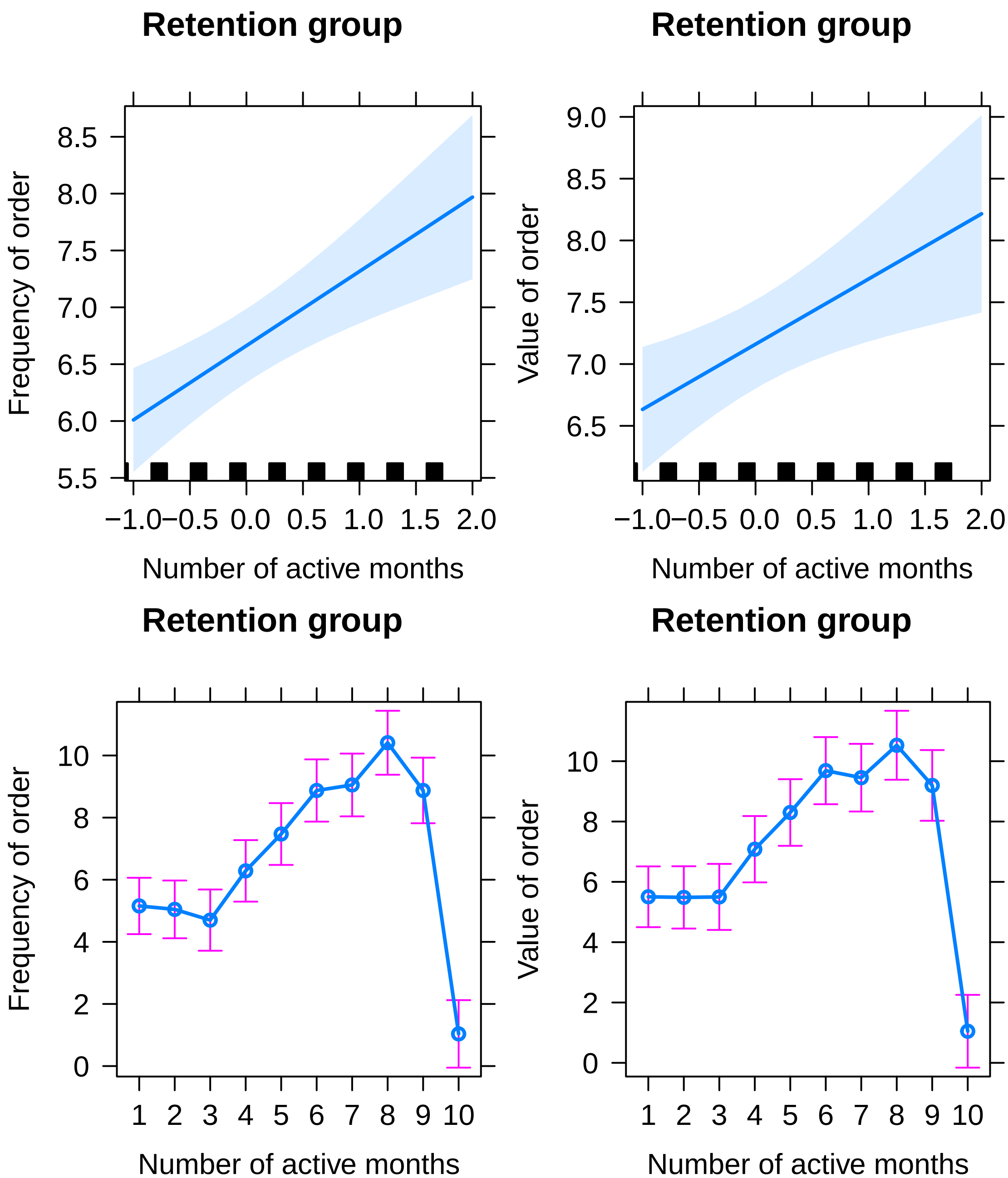
revenue source	pre-subscription	active	cancelled
Delivery fee	9.97	0.00	39.73
Subscription fee	0.00	7.99	0.00
Comission	3.89	18.46	38.30

Table 5: Summary of the average monthly generated revenue in the Churn group

revenue source	pre-subscription	cancelled
Delivery fee	4.08	9.30
Subscription fee	0.00	0.00
Comission	0.54	2.64

As the Retention group proved to be valuable, it is worth gaining a deeper understanding on the behaviour of this group. For instance, it can be interesting to test whether having an active subscription changes their ordering behaviour in the *long-term*. To see whether increasing number of months have a growing facilitation effect as time passes I created a variable coding for the accumulated number of active monthly subscriptions. After this, linear regressions were run where the months passed was coded both as a continuous and as a categorical predictor. While the first one is good for quantifying general effect of time the other analysis can help to identify break-points within the 10-month period.

The analysis revealed a significant positive linear relationship between the increasing number of active subscription months and increasing number of order frequency and value. When the active subscription months were coded as categories one can see the significant increases appear from the fourth active month. While in the tenth month there seems to be a significant drop it is likely to be arbitrary, probably because the data collection was finished at beginning of the tenth (final) month (see figures below).



Limitations and future steps

While the statistical tests provide confidence in the accuracy of these results it must be noted that the some of the assumptions these tests hold are violated by the distribution of the analysed sample. Accuracy could be further improved by acquiring exact data on monthly costs and revenue associated for every customer. Furthermore, collaboration with the product management and user research team would help with the interpretation and validation of the results regarding different types of subscribers. Finally, data on the customers' activity from the business's competitors would deepen the insights on the impact of the subscription programme.