Retail Analytics

Marketing efforts at Metaloop from 2023 to present date

# Bernardo Carvalho

# 07/04/2024

Table of Contents

[Bernardo Carvalho 1](#_Toc163602361)

[07/04/2024 1](#_Toc163602362)

[Introduction 3](#_Toc163602363)

[LTV Analysis 5](#_Toc163602364)

[Customer retention 7](#_Toc163602365)

[Campaign conversion analysis 9](#_Toc163602366)

[VAR model 11](#_Toc163602367)

[Analyzing time series 11](#_Toc163602368)

[GMV 12](#_Toc163602369)

[A graph of different types of data

Description automatically generated with medium confidence 12](#_Toc163602370)

[General Ads 12](#_Toc163602371)

[Brand Ads 13](#_Toc163602372)

[Display Ads 13](#_Toc163602373)

[Stationarity tests 14](#_Toc163602374)

[Conclusion 16](#_Toc163602375)

[Appendix 16](#_Toc163602376)

# Introduction

Metaloop is a metal recycling company based in Graz, Austria. We operate as a digital scrap-yard, which means we don't own or lease any warehouses where we store material. Instead, we do the matching of suppliers (also called "sellers") and buyers online, through a managed marketplace.

Our main sources of inbound marketing is a from a form that is available in our website. It's called internally the "Pro-form", and it's the form filled out by enterprise customers to sell a specified quantity of scrap metal. This form can be found on our main page, as shown in the image below:

A screenshot of a website

Description automatically generated

After clicking the button, the prospect is redirected to the form in the next image

A screenshot of a screen

Description automatically generated

After completing this form, an opportunity record is created on our CRM system and the sales department takes over. We do not directly sell anything from the website for wholesale customers. Our deals are of high involvement and the sales team always needs to qualify and follow up. Therefore, our most important conversion metric in the website is the completion of this form.

This is tracked in the Google Analytics suite via the tag manager. The event that records the submission of the pro-form is called "general-mql". Since no sale occurs on the website, Google analytics can't calculate ROI or LTV out of the box.

Our team was tasked with evaluating the overall performance of our Marketing efforts. In general, management is very interest in the LTV of the customers that come through inbound. Moreover, they want to know if there is any differentiation between those that come directly through a search ad.

WE emphasized that looking at the LTV of the customers that come only from search Ads might be a myopic approach, given that the marketing efforts build up over time, and issues like multiple channels simultaneously at play and attribution pose a challenge.

Considering this, we propose the application of a VAR model to our Sales data, in concurrence with three streams of digital marketing: General Ads, Brand Ads and Display Ads. We will elaborate on this on the **VAR model** session.

# LTV Analysis

Analyzing LTV at Metaloop can pose an additional challenge In comparison to other retail companies. Because we arbitrage a deal between a seller and a buyer of scrap metal, we can look at sales volume (or simply GMV) from a supplier or buyer perspective. Our market is usually constrained on the seller side, and our efforts are normally directed at the acquisition od new sellers. However, the actual revenue comes from the payment of buyers. The difference between the final sale value (GMV) and the price we paid for the supply (COGS) is called the Net Revenue. In order to estimate GMV and Net revenue on the seller side, several transformations to the data sources must be applied. These are detailed in Appendix 1: LTV analysis and Data preparation.

As cited in the introduction, when the a prospect completes the Pro form in our webpage, it generates an opportunity record in our CRM. After the deal is closed, one or many orders can come out of that opportunity. The image below shows that around 2/3 of our GMV comes from orders associated with an opportunity.

A pie chart with green and grey circles

Description automatically generated

In fact, the next plot shows that for a long time all our revenue was associated with opportunities (which were in most cases generated by the inbound channel). Recently, however, we have seen the growth of the outbound channel.

A graph with numbers and a line

Description automatically generated

However, when we look at the proportion of the GMV that comes from pro-form opportunities, we see that it is quite small.

A green graph with numbers and a green line

Description automatically generated with medium confidence

This shows that the orders generated not from the pro-from are usually larger than those from the form. This is expected from a business perspective, as the larger customers will not do the larger opportunities from the form. It is usual for large customers to do trial loads before they start committing to larger volumes.

We know that this is a limited approach (only looking at the opportunity level), since accounts could have started from the pro-form, but developed into a recurring customer.

## Customer retention

The customer retention is calculated based on the adjusted contribution margin (GMV – COGS – Transport costs). The margin needs to be adjusted to accommodate the Acquisition orders (i.e: orders that go into stock) from the Seller's perspective. Some orders with negative margin are expected (due to actual losses or bundled orders). They should even out across time.

We define the "cohort period", which is the period of the seller’s first order. We group by order creation date, since we are looking at the marketing efforts. The "cohort number" represents the period difference relative to the acquisition of the order. The following image plots the cohort margin by cohort number (in year-quarters)

A blue and white graph

Description automatically generated with medium confidenceSome noticeable trends are that the newer cohorts (i.e.: down across the y axis) have a higher value and better retention, although retention is visible throughout the entire base. The notable exception in cohort 2018Q4 refers to our most profitable account. To check if we have an overall retention across cohorts YoY, we plot a cohort retention line plot (excluding this year). We see that in general we retain most of our revenue over time, with 25% being our lowest retention rate from the initial cohort (we used a log scale for better visualization, given the demolition project started in 2018 that generates a huge outlier in retention)

A graph of different colored lines

Description automatically generated

Although we see a clear retention profile for our general base, the retention for pro-form exclusively sellers is much more limited.

A graph of blue and black bars

Description automatically generated with medium confidence

This indicates that for pro-form generated opportunities, the retention is negligible. However, as mentioned previously, this is likely since only the first trial loads are done from a pro-form opportunity. After this, the next opportunities are created after direct contact to our account managers.

**CLV with fixed horizon**

The CLV with fixed horizon aims at estimating the \_present\_ lifetime value of the new account. For marketing, putting it in terms of present value is relevant because all of the investment in campaigns is made upfront. The formula is:

Where is the expected margin at period t (in our case, this is constant), is the recurrance rate at period t, i is the discount rate (i.e: the discount rate per period of future value in comparison to present value) and N is the horizon.

We can also use a geometric approximation to arrive at an infinite-horizon formulation.

Applying the fixed horizon method to the orders that come from sellers that have interacted with the pro-form or had opportunities created by the inbound team (see Appendix II), we arrive at an estimated CLV of EUR 979

## Campaign conversion analysis

Having arrived at an LTV candidate, we can look at the conversion of the website and the conversion of opportunities as a whole and from the pro form. This will enable us to evaluate the click value, which should drive our future campaigns and strategy.

The following image shows the session conversion rate for Ads campaigns into the key events. As mentioned in the introduction, the key event is the “general\_mql” event. However, the other appointment related events are also considered conversion events in our inbound strategy

A screenshot of a web page

Description automatically generated

Based on these values, the conversion from session to inbound event is given by:

Therefore, if we consider the total conversion rate of the opportunities (27%), the expected return per click becomes EUR 1.32. In comparison to the EUR 0.7 CPC in the same period, we can say that we have slack on our bidding.

This conclusion still depends on a very loose assumption that the Ads expenditure generated all of this volume. That is, had we had no marketing campaigns, we would still observe the same volumes of inbound traffic, given that most of our traffic comes from organic and referred search? And how does the awareness role of marketing impact the organic traffic and in consequence the events not generated by Ads?

To try to answer this questions, we take a statistical approach and employ a VAR model, where we aim to assess the impact of the three types of campaign we do: General, Brand (both Ads) and Display campaigns.

# VAR model

A Vector Autoregressive (VAR) model is a statistical method used to analyze the dynamic relationships between multiple time series variables related to marketing activities. A VAR model treats each variable in the dataset as a function of its past values and the past values of all other variables in the system. It assumes that the variables influence each other mutually over time. It consists of a system of equations, with each equation representing one variable as a function of lagged values of itself and other variables. For example, if we have four variables:

1. Search Ads
2. Brand Ads
3. Display Ads
4. GMV

The VAR model would comprise three equations, each describing the behavior of one variable in terms of its own past values and the past values of the other two variables. The model captures the dynamic interactions and feedback loops between marketing variables. For instance, changes in advertising spending might impact sales revenue, which in turn affects future advertising decisions. The VAR model can quantify these interdependencies and provide insights into how changes in one marketing metric affect others over time.

## Analyzing time series

We start by plotting the time series of our data, along with the ACF and PACF plots. This first step serves the purpose of building intuition about the time-series behavior and what type of transformation, if any, we will need to achieve stationarity in the series.

### GMV

A graph of different types of data

Description automatically generated with medium confidence

### General Ads

A graph of a line graph

Description automatically generated with medium confidence

### Brand Ads

A graph of a line graph

Description automatically generated with medium confidence

### Display Ads

A graph of a line graph

Description automatically generated with medium confidence

## Stationarity tests

After analzing the time series, we perform unit root tests on our datasets. We are using the augmented Dickey-Fuller’s test, and we see from these results that only the Brand Ads is stationary.

Augmented Dickey-Fuller Test

data: LgmvTotal

Dickey-Fuller = 2.0515, Lag order = 4, p-value = 0.99

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

data: LgeneralAds

Dickey-Fuller = -0.48055, Lag order = 4, p-value = 0.98

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

data: LbrandAds

Dickey-Fuller = -3.8297, Lag order = 4, p-value = 0.02268

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

data: LdisplayAds

Dickey-Fuller = -1.8237, Lag order = 4, p-value = 0.6466

alternative hypothesis: stationary

To correct this, we perform a 1 period differentiation in the remaining data. Re-running the tests yields:

Augmented Dickey-Fuller Test

data: DgmvTotal

Dickey-Fuller = 0.25344, Lag order = 4, p-value = 0.99

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

data: DgeneralAds

Dickey-Fuller = -3.7555, Lag order = 4, p-value = 0.027

alternative hypothesis: stationary

Warning: p-value smaller than printed p-value

Augmented Dickey-Fuller Test

data: DdisplayAds

Dickey-Fuller = -5.319, Lag order = 4, p-value = 0.01

alternative hypothesis: stationary

We see that our GMV data is still non-stationary. Therefore, we must perform another differentiation. After this final step, the data also passes the stationarity test.

=================================================================

Dependent variable:

-----------------------------------

y

(1) (2) (3) (4)

-----------------------------------------------------------------

DdisplayAds\_cut.l1 0.062 0.015 -0.070 -0.391

(0.128) (0.032) (0.101) (0.286)

DgeneralAds\_cut.l1 1.037\*\* -0.182 0.338 0.884

(0.514) (0.127) (0.409) (1.153)

LbrandAds\_cut.l1 -0.043 -0.024 0.637\*\*\* -0.068

(0.107) (0.026) (0.085) (0.240)

D2gmvTotal\_ts.l1 0.016 -0.006 0.050 -0.518\*\*\*

(0.048) (0.012) (0.038) (0.107)

const 0.107 0.092 1.121\*\*\* 0.172

(0.334) (0.083) (0.265) (0.748)

-----------------------------------------------------------------

Observations 66 66 66 66

R2 0.069 0.050 0.491 0.352

Adjusted R2 0.008 -0.013 0.457 0.309

Residual Std. Error (df = 61) 0.858 0.212 0.682 1.922

F Statistic (df = 4; 61) 1.133 0.795 14.691\*\*\* 8.275\*\*\*

=================================================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The var model presents some challenging results. We will ignore the significance score for this analysis, given the reduced size of our dataset. First, we look at the carry-over effects, which are on the diagonal of the matrix. We see that displayAds and brandAds seem to have a carry over, long term effect on the GMV, whereas the generalAds don’t. This implies that the entire effect of the general-purpose ads lies on it’s immediate impact on the GMV. We should note here that we differenced GMV by 2 periods, and therefore immediate relates to the next period in this case.

Next, we turn to the direct impact on the target variable (column 4). It is unexpected that we have negative coefficients there, since it means that these campaigns have a negative short-term effect on the target. However, if we return to the time series, we see that most of our expenditure goes to the general Ads, so this result could be caused by randomness in the data.

The feedback effect of the target variable on the input variables is also

# Conclusion

1. We should definitely improve our tracking
2. We should re-run the VAR model after we start putting more money and effort into marketing

# Appendix

[file:///Appendix 1/ LTV analysis and Data preparation](file:///Appendix%201/%20LTV%20analysis%20and%20Data%20preparation)