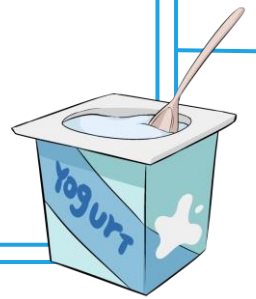
Professors: Lucio Lamberti & Piercesare Secchi
Tutor: Francesco Cherubini

A.Y 2021-22

LAB PROJECT

MARKETING MIX MODELING

Performances review and media budget allocation



POLITECNICO
MILANO 1863

INTRODUCTION

GOAL SETTING & PROJECT OVERVIEW

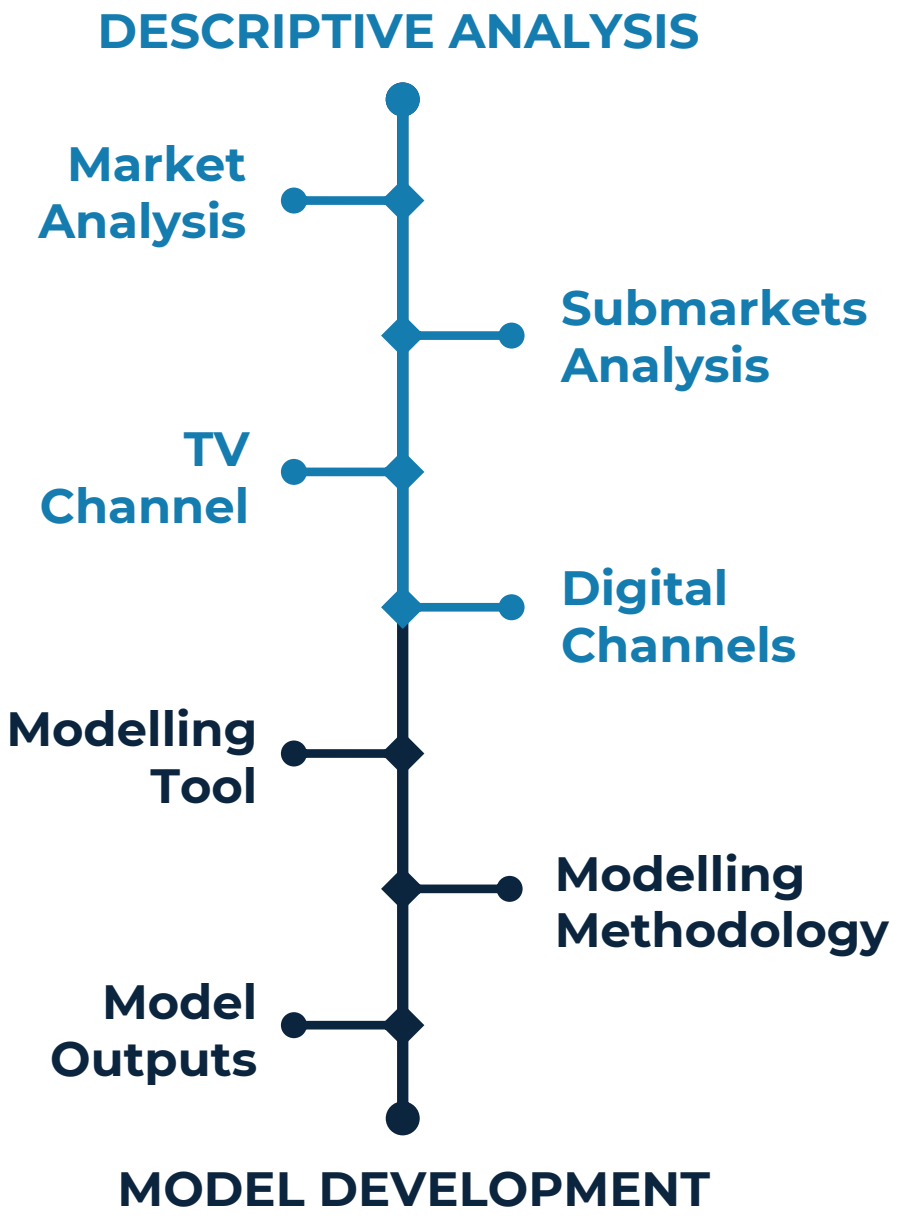
The **objective** of the work is twofold. On one hand, it is crucial to study the **characteristics of the market of reference** and to go deeper in understanding **the peculiarities of the Submarkets** where the Focus Brand (FB) is operating. On the other hand, it is relevant to define **how to allocate the media budget**, integrating investments data with media, television and sales performances. The main focus of the study regards the **Submarket 1**, but a part of the analysis and modelling is also dedicated to the other two Submarkets.

According to the information provided, **different degrees of detail** are given. In fact, sales are available for all the actors in all Submarkets, but it is not the case for media and television data. Media impressions/clicks and investments data are **only related to the Focus Brand** in the Submarket 1 while few more information on competitors, in the 3 Submarkets, are given for television performances. The presentation is composed by two main parts: the first one deals with the ***descriptive analytics*** of the data available, while the second part is on the ***model development*** in order to drive conclusions about budget allocation, media effectiveness and return and contribution of different levers on sales.

INITIAL ASSUMPTIONS

Before proceeding with the descriptive analysis, it is necessary to provide guidance on the **assumptions** introduced and the procedures used in the development of the study. The main ones are reported below:

- ✓ **2017* & 2021*** data **are not complete**, for this reason, both semi-annual and yearly computations and comparisons are taken into consideration. In particular, data of the second semester are provided for 2017 while for 2021 only those of the first semester are given;
- ✓ Both **weekly** and **monthly** data are considered. The choice between the 2 alternatives is driven by what results to be more effective to reach the specific goal. Moreover, the **month information** is **used to aggregate the weekly data**, already available, into monthly data. Unfortunately, some weeks are in between two different months, but since it occurs seldom, this simplification is introduced;
- ✓ The **main tools** used to develop the analysis are **Excel** and **R**. Dealing with the model building, a central and important role is given to a *specific library for marketing mix modelling*: **Robyn**.



OVERALL MARKET ANALYSIS

MARKET OVERVIEW

First of all, we are interested in understanding the **market size**, both in terms of volumes and monetary value, and the evolution that it experienced over the last 5 years. The tables on the right report both annual and half-yearly sales; the values in bold correspond to the best performances registered along the years. It is worth noticing that the best years for the market date back to the pre-pandemic period, in fact the overall revenues recorded a **decreasing trend** from 2019 onwards.
The yoghurt market consists of 3 main Submarkets, each one characterized by different volumes and revenues, level of competition and growth rate. Submarket 2 presents the highest revenues and volumes, followed by Submarket 3 and 1 which are more niche-oriented.

The graph at the bottom helps exploring the decreasing trend affecting the total market. From 2019 to 2020 volume sales decreased by **-8,3%** and continued to slow down also in the first semester of 2021. This loss is mainly due to a sales contraction of Submarket 3 (-16,8%) and Submarket 2 (-7%) whereas a slight increasing pattern distinguishes Submarket 1. Category 1, in fact, did not suffer the Covid period like the other categories, because it compensated its losses in the first months of the year with remarkable sales in the second semester, and it is also the only one which remained moderately stable also in the first semester of 2021.

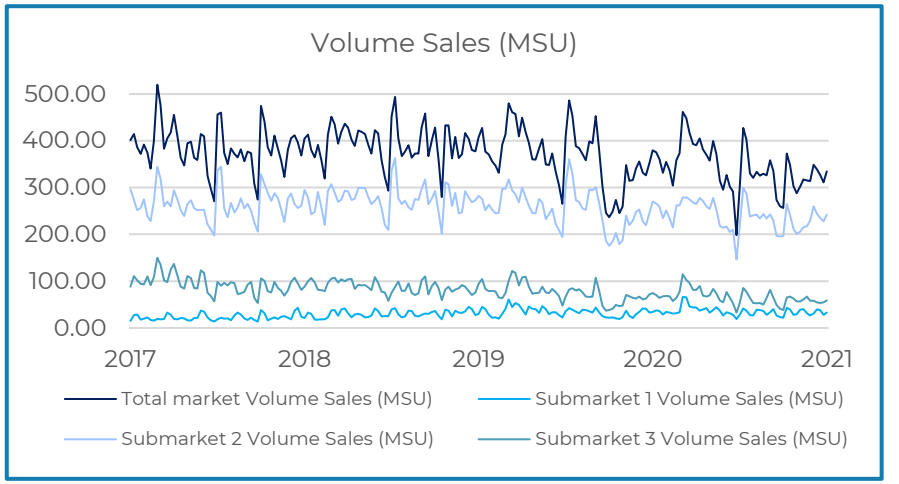
SUBMARKETS OVERVIEW

The three Submarkets show different competition models, and the competitive behaviors of the brands are moderately variegated. Submarket 1 and 3 are dominated by the **Focus Brand** and **Competitor 1**, which hold most of the sales and volumes shares; instead, Submarket 2 shows a fiercer competition, due to the **higher fragmentation** of shares and the lack of one or a few dominant brands. While the competitive situation in Submarket 2 and 3 remains approximately unchanged over time, Submarket 1 shows a progressive shift in competitive behavior due to a change in the balance of power between brands.
In the next phase, each Submarket is going to be described and analyzed and an important specification should be done for the residual term that will be cited in the following graphs. "**Residuals**" stands for the delta between the overall volume and value sales of each Submarket provided in aggregated data and the actual sum of all competitors' volume and value sales. For the sake of the exercise, we assume that this difference identifies smaller brands that represent a much more fragmented part of the market and will hereby be referred to as "residual". While in Submarket 1 and 3 residuals represent a negligible part of the market, in Submarket 2 they own, altogether, the greatest sales share.

Year	Volume Sales (MSU)	VALUE Sales (MLC)
2017	/	/
2018	20.177	**243.537**
2019	**20.250**	239.382
2020	18.580	221.555
2021	/	/

1° Semester	Volume Sales (MSU)	Value Sales (MLC)
2017	0	0
2018	9.942	**118.734**
2019	**10.271**	121.470
2020	8.978	108.000
2021	8.502	99.915

2° Semester	Volume Sales (MSU)	Volume Sales (MLC)
2017	10.092	120.950
2018	10.235	**124.803**
2019	**9.979**	117.912
2020	9.603	113.555
2021	0	0



SUBMARKET 1

COVID EFFECT

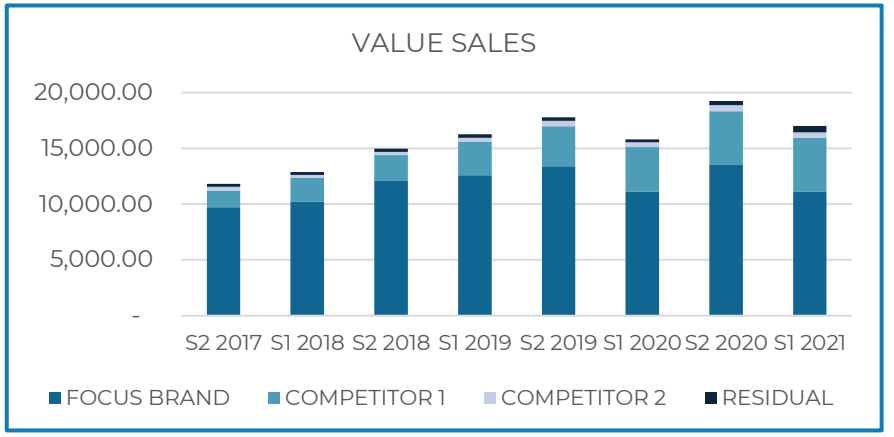
Overall, this market experienced a **double-digit growth** in the first 4 periods. Between March and May 2020, characterized by the **strictest Covid-19 regulations** (lockdown period), the market suffered a **13.2% sales shrinkage** compared to the previous year then compensated by the highest sales peak in the second semester of the same year. Finally, between January and June 2021, volumes decreased with respect to the previous period, but increased compared to its corresponding in 2020.

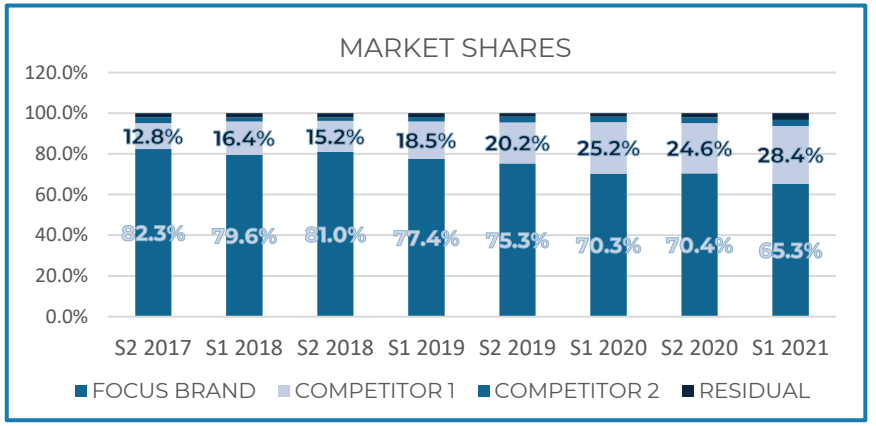
COMPETITIVE LANDSCAPE

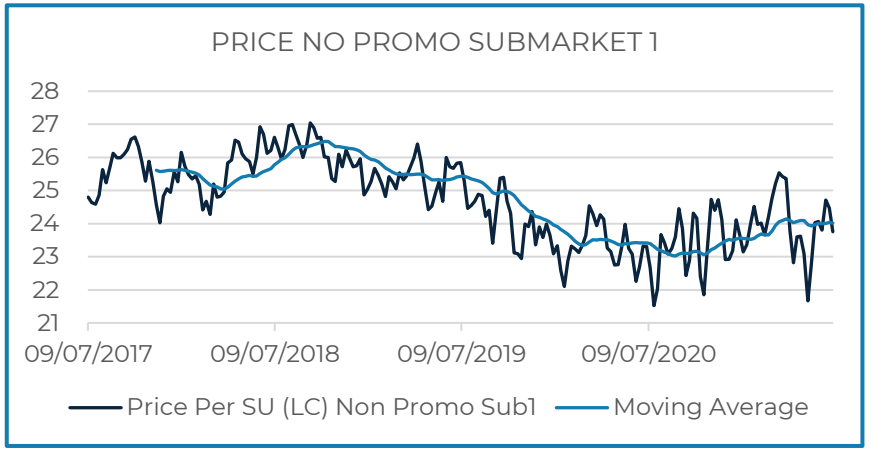
Submarket 1 is composed of **3 main competitors** and a minor percentage of smaller residual competitors. Among them, the **Focus Brand** and **Competitor 1** clearly dominate the market, owning together never less than **93.5%** of the total sales. However, they showed two diametrically opposite behaviors. Despite never reaching below 65% in the timeframe under analysis, the Focus Brand's initial 82.3% market share was **progressively lost to Competitor 1** in almost all periods; in fact, while the former recorded a 17% loss of market share in this timeframe, Competitor 1 has more than doubled its shares, shaking the Focus Brand monopolist position and reaching 28.4% at the end of the first semester of 2021, from an initial 12.8%. As for **Competitor 2**, since this player **never reached more than 3%** of the Submarket and didn't go through major changes in market share, nor other features, it was eventually excluded from further analyses, and deemed not relevant.

PRICE EVOLUTION

The full price (without promotions) of goods sold in the Submarket 1 clearly shows a **decreasing trend** over time, as it is noticeable in the graph beside. To make it more comprehensible, a simple moving average over 20 periods (about 5 months) was calculated. It is difficult to assess the role of Covid-19 in this downward price shift: in the period corresponding to the first lockdown price evidently declines, but this started way before the pandemic arrived. Furthermore, from August 2020 on, the trend slightly shifts upwards, and price starts to grow back. A more plausible cause for this behavior may be represented by the **competition between the Focus Brand and Competitor 1,** whose prices follow the pattern of the overall Submarket and start to decrease approximately at the same time. This may be a sign for a **price war**; moreover, the fact that the prices of C1 are quite lower than FB may explain their great gain in the market share, that speeds up starting from 2019 on, even during the pandemic period. This makes it very interesting to evaluate the magnitude of contribution to sales of both the Focus Brand's and the Competitor 1's prices.







SUBMARKET 1

VALUE SALES

Looking at the value sales year over year, the impact of **Covid-19** is particularly relevant in the period between March and May 2020, that coincides to the first lockdown. Also, lower levels of revenues during the rest of the year, with respect to the previous years, can be associated to the same root cause. Nothing certain can be stated, on the other hand, about the sales **seasonality**: although a quite recurrent peak during the month of September can be graphically noticed for all the years of the analysis, patterns of similarity in other periods are not observed. To investigate that, a time series decomposition on R was performed, which unfortunately showed autocorrelated residuals, making its results unreliable. Finally, as it can be noticed by the second graph where sales are represented weekly, there are a few very high peaks that are not explainable ex-ante.

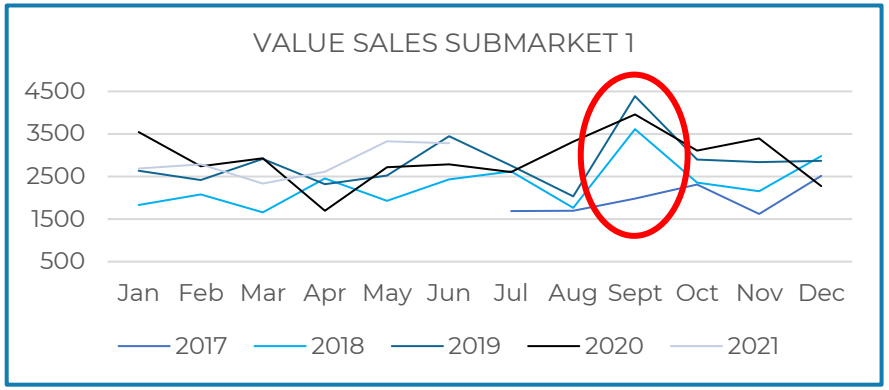
FOCUS BRAND – PERFORMANCE DRIVERS

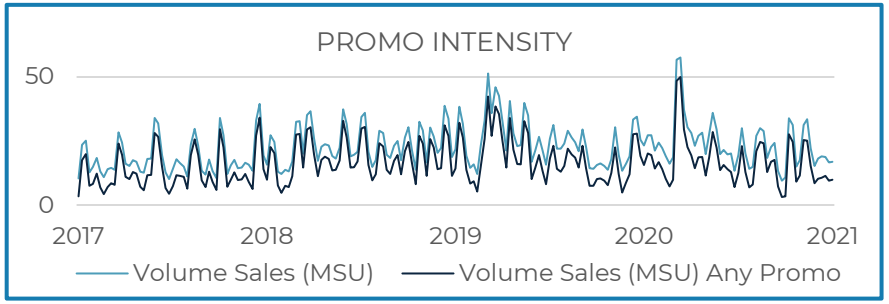
The Focus Brand's performances are heavily affected by **promotions**: in fact, between 2017 and 2018 **more than 50% of sales** were achieved under price cut promos. More particularly, moving into the last few years, in-store promotions become increasingly important representing, from 14% in the second half of 2017 up to 41% of sales in 2020.

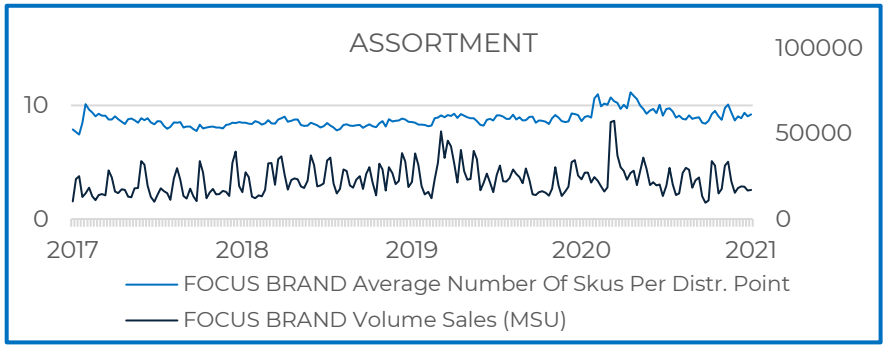
Assortment of products doesn't grow that much throughout the years (especially if we make a comparison with the other companies), so the **trend of sales does not seem to be directly relatable** by this variable. This is confirmed by the correlation test between the two features, whose result is **0,33**. However, it must be stated that the product assortment level of the Focus Brand is always high, therefore despite the correlation with sales is not so remarkable, it might still be a relevant component of the sales baseline that the FB has nurtured over time.

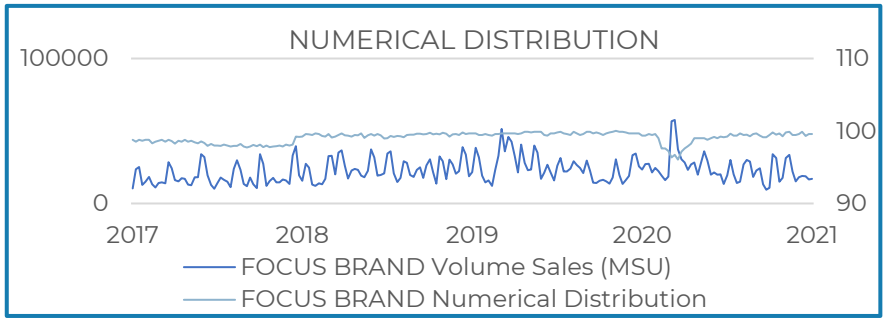
Numerical distribution is extremely **stable over the years** considered for the analysis. Therefore, distribution **does not seem a lever** that drives the brand's performance, despite it might again be part of the baseline. Again, this is confirmed by the correlation test, which is equal to **0,08**.

A similar result is obtained if the **weighted distribution** is analyzed in parallel with the value sales of the company: both graphically and statistically (correlation equal to about **0,18**), a relationship between the two variables is not suggested to be present. An interesting remark can be made on the **weighted distribution anypromo** variable, that represents the distribution of sales made in promotions, weighted on the amount of revenues made by the retail stores selling it. In this sense, this variable mixes the concepts of weighted distribution and promotion intensity, and can be useful to represent the brands' strategies in terms of promotion. Model-wise, it seems to make more sense to use this variable to model promotions because the company can know its future value, in contrast with promo intensity which fully depends on the consumers' behavior.





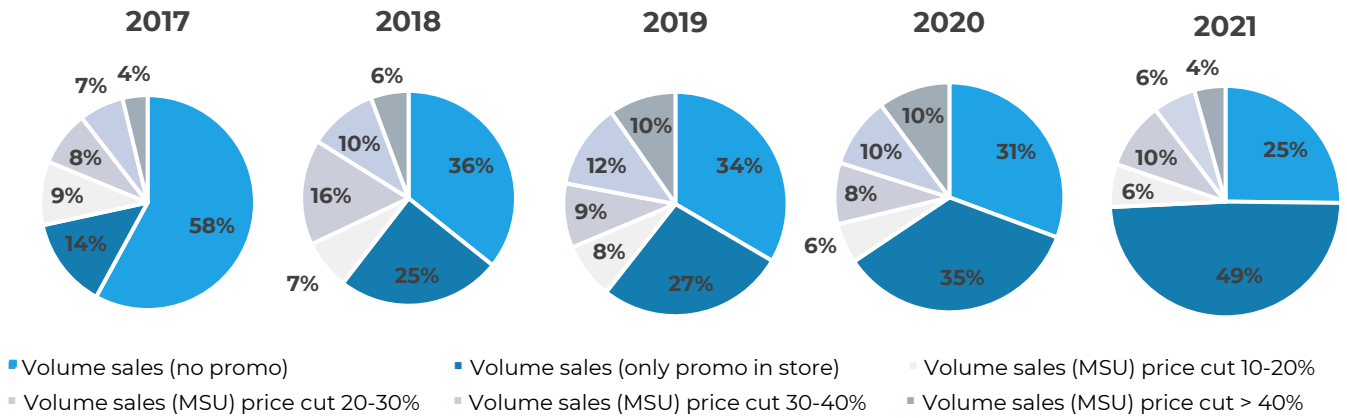




SUBMARKET 1

COMPETITOR 1 – PERFORMANCE DRIVERS

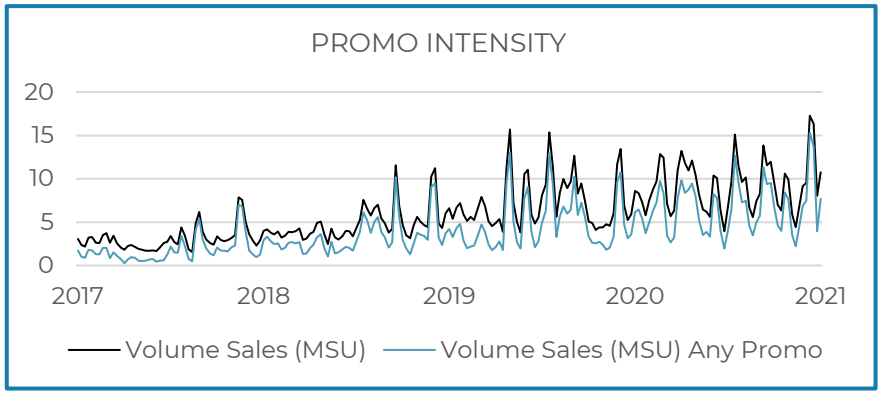
Also Competitor 1's **sales** are **strongly driven by promotions**, especially during the last two years, where a more aggressive price strategy seems to be applied and sales without any kind of promotions range only between 25% and 30%. The set of pie charts below highlight that the main competitor's sales on promotions have **evolved** over time, following **similar proportions to the Focus Brand's**. This might be caused either by a shift in their promotional strategy or by a variation in customer behavior: consumers might have moved towards increasingly buying C1's products in promotion.

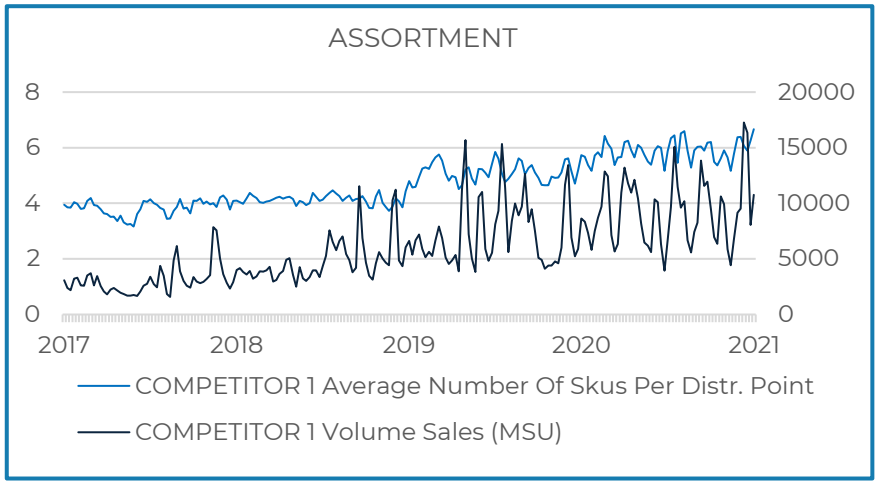


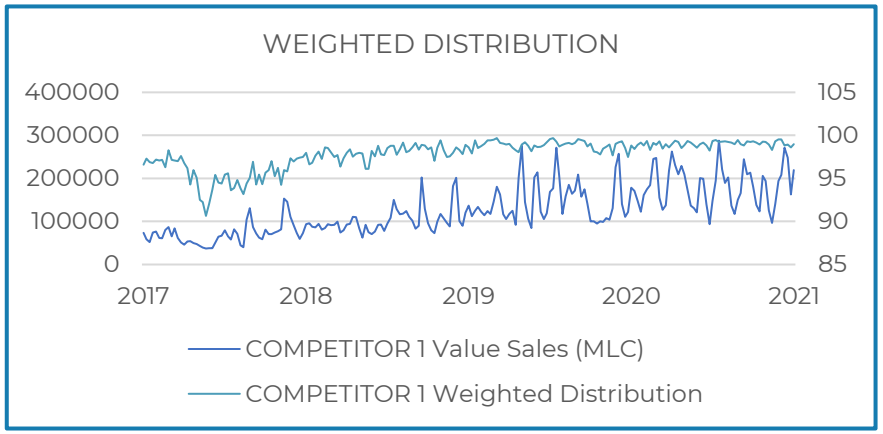
In terms of **assortment**, Competitor 1 has been gradually reducing the gap of average number of SKUs sold in the distribution points with the Focus Brand, **increasing its product diversification** in Submarket 1. In fact, Assortment of products ranges from a minimum of 3,16 to a maximum of 6,67. Its increase over time goes hand in hand with sales, as proved also through a correlation test (**0,7613**).

As for distribution, for both **numerical and weighted distribution** a certain relationship degree with volumes and value sales can be established, not so much as graphically as statistically. In fact, tests return a correlation equal to, respectively, **0,5956** and **0,6378**: this suggests that sales and distribution have similar patterns over time.

This chapter shows how Competitor 1 is eroding a consistent part of the Focus Brand's market share by aligning its strategy with the aforementioned. This makes it interesting to understand, during the the modelling phase, which of these competitive levers has an impact on the FB's market share erosion and which instead is more marginal. It is therefore important to separately take into account them, to see where the main threats for the Focus Brand come from.







SUBMARKETS 2 & 3

SUBMARKET 2

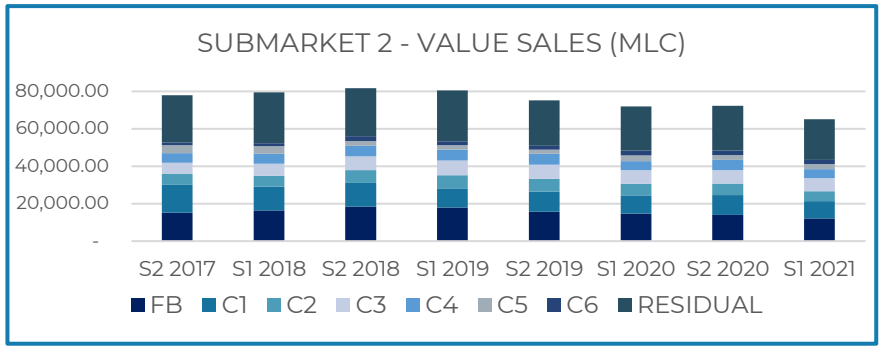
Compared to Submarket 1, Submarket 2 is very different. In fact, while the first one was characterized by a duopoly, here the market is **much more fragmented**, **thus more competitive**. This situation has not significantly changed over time, since the brands with most shares in the market remained the same (Focus Brand, Competitor 1, Competitor 2, and Competitor 3), and their shares fluctuated with no significant changes in the market positions of the brands. In this case, the **residual share** is very high, meaning that, besides the main competitors, also many more and smaller brands compete, constituting a consistent amount of the sales.

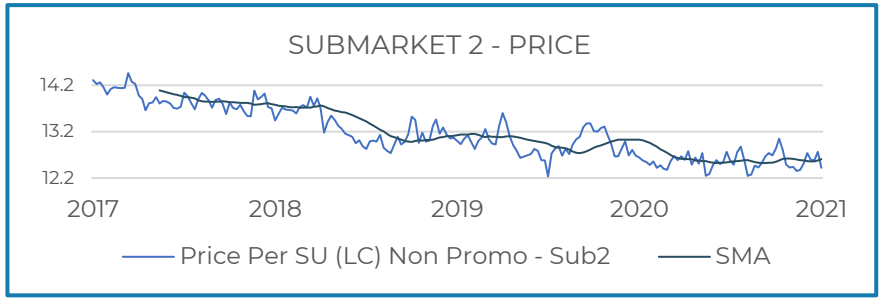
Price (no promo) in the overall Submarket 2 has a **strong decreasing trend** throughout the entire period considered for the analysis, that seems to **flatten** in the **last months of S1 of 2021.** Once again, to highlight this trend, a simple moving average has been plotted onto the price curve. This behavior is mainly caused by the Focus Brand, the price of which is suffering a severe slump also in this market, while other competitors (e.g., C1) started increasing their prices in that exact same period. In terms of Covid-19, its impact is evident in the graph on top, showing a downturn in S1 2020.

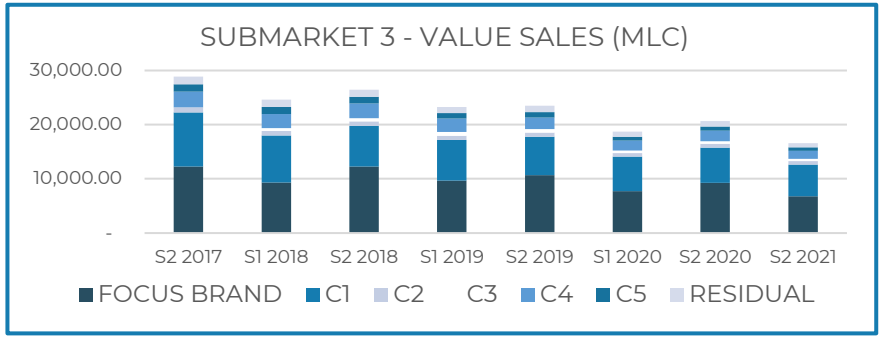
SUBMARKET 3

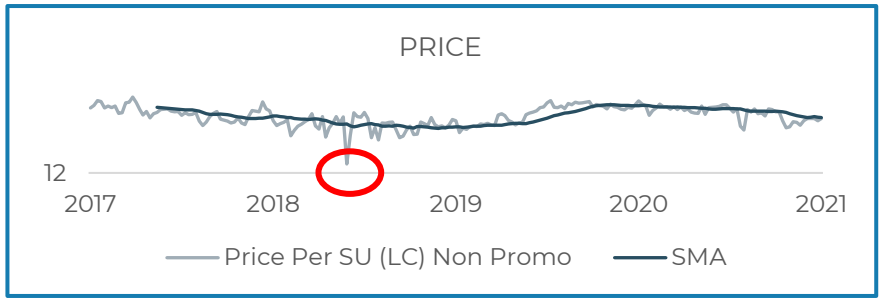
Submarket 3 sees 3 competitors over 6 (Focus Brand, Competitor 1, and Competitor 4) to share at least **85%** of sales in all periods. Among them, the FB is always the most important brand in terms of share, but it presents an **alternating behavior** all along the 8 periods. It can be noticed that in the first semesters of these 4 years the FB always earn shares at the expense of Competitor 1, which then in the second semesters earns back all (or part of) its shares. Instead, the third most important competitor in the market (Competitor 4) presents a steady market share, despite a decrease in the last periods that brought to the lowest level in this timeframe, that is **8.7%**.

Submarket 3 behaves strangely compared to the other two markets in terms of price. In fact, the price fluctuates and goes up and down throughout the years without a clear pace: a deeper analysis on data aggregated by month, for example, **doesn't show particular patterns** or seasonality. Moreover, an abnormal value is observed in the first part of 2018, as depicted in the graph on the right. Also in this case, Covid-19 has certainly impacted on the market, as it can be evidently seen from the drop of sales, generally quite constant over the years, experienced in the first semester of 2020









TV CHANNEL

TV PERFORMANCES

The TV GRP stands for the number of target people that have been seeing a spot at least once. The higher this value, the greater the probability that sales will be boosted thanks to TV investments. The graph on the right shows the sum of GRP's on TV channels of the Focus Brand in Submarket 1. Every year, the best performances are always achieved in the **second semester**.
Because investments in TV are much more relevant and continuous than any digital medium's (as shown in next slide), and given the type of good sold by the brand, that is a typical FMCG, we expect TV investments to have a much higher impact on sales compared to digital media. If considering MLC as thousand euros, TV investments in the 4-year timespan represent about **8.63%** of revenues.

PRIME TIME VS DAY TIME

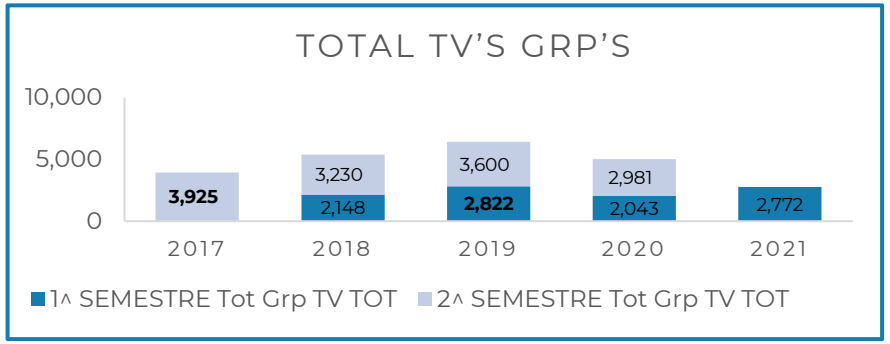
The graph helps to understand how effective the advertisement is, in function of the time of the spot placement. Evaluating it in a qualitative way, it is pretty evident that daytime values are always higher than prime time ones. However, we should consider how many hours each period of time is composed by. Knowing that the primetime spans from 19.30 to 23.30 and assuming that daytime lasts from 7.30 to 19.30, we are comparing 12 hours in the day and only 4 hours in the evening, so prime spots result to be **more effective in terms of performances** confronting with daily GRP's. At the same time, the primetime cost per GRP is 29% higher. Given this tradeoff, in order to assess the true effectiveness of the 2 strategies, further investigation is required.

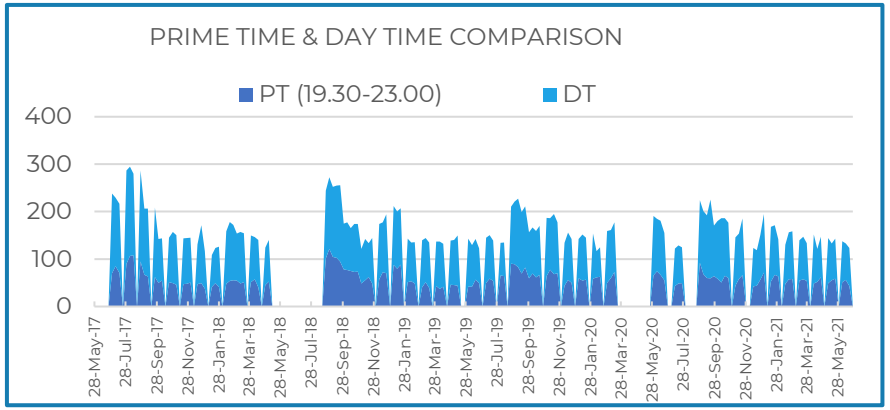
TV CHANNELS

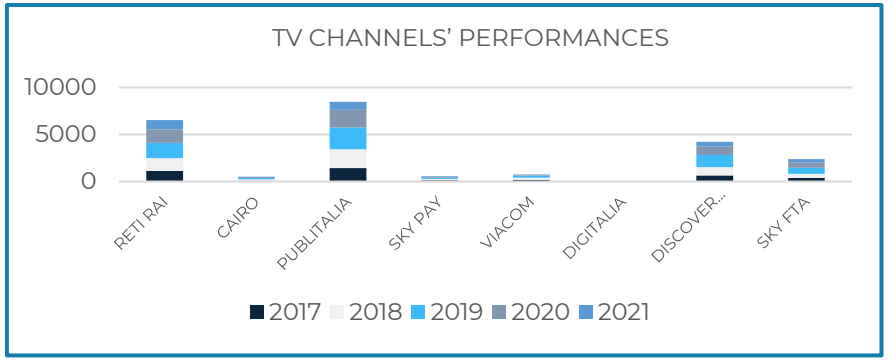
Analysing the graph on the right, respectively Publitalia, Reti Rai, Discovery Media and Sky Fta seem the most effective channels, thus suggesting an investment strategy mainly based on those television stations. Besides, considering the evolution of Total GRP's along the years, it is possible to highlight the **decreasing trend that affected TV**. The Television GRPs can be aggregated as TV GEN (rai, pubblicitalia, Cairo,) and TV DIG (all the others). The cost per impression of TV GEN is 47% higher than TV DIG.

COPY LENGTH

Speaking about the length of copies used by the *Focus Brand in* the Submarket 1, only spots of 10, 20, 30 seconds are used and those of 20 seconds are the spots which collected a much higher number of vision. From the table on the right, copy 30" is mostly used in 2017, so it will be **neglected**. The main experts suggested assuming an equal cost per GRP between copy 20 and copy 30.







Year	Total GRP's 10"	Total GRP's 20"	Total GRP's 30"
2017		1.681	2.244
2018	301	5.075	
2019	624	5.796	
2020	2.819	2.190	27
2021	1.700	1.074	
Total	5.444	**15.816**	2.271

*best performances are highlighted in bold

DIGITAL CHANNEL

For the Digital Channel data are provided on a weekly basis and refer to the investments made, and the performances reached in terms of Clicks (AMS & Google) and Impressions (10 different media). Since the investments are made on impressions, not much analysis was carried out on clicks, as suggested by domain experts. In order to understand the Strategy adopted by the FB on the digital marketing side we have analyzed the shares of investment dedicated to each tool, year over year in each semester. From a general overview different conclusions have been reached:

- A **reduction in the complexity** of tools used. With the beginning of the Covid pandemic the number of tools the Company has invested in the same week has decreased drastically.
- Higher **continuity and uniformity** among semester 1 and semester 2 after the pandemic. In the previous year, on the other side, there is a remarkable gap between the two.
- **Data granularity** on some media is low: It often occurred that was spread uniformly in weeks, creating 4-periods series with the same investment and impression levels.

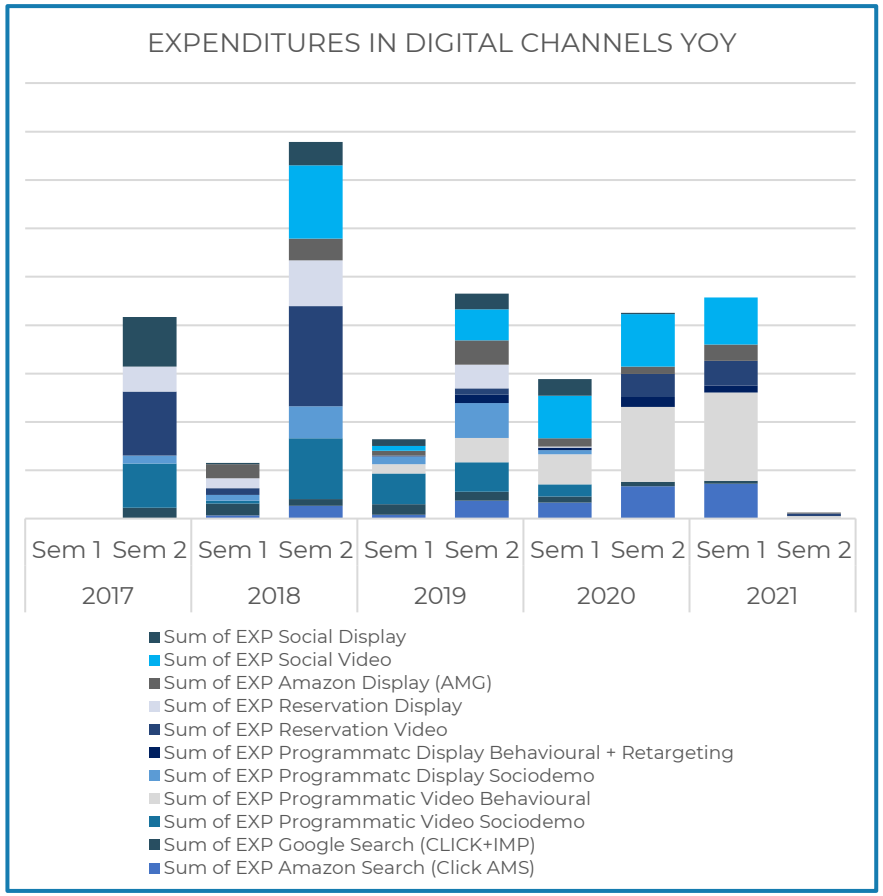
DIGITAL VS TV CHANNEL

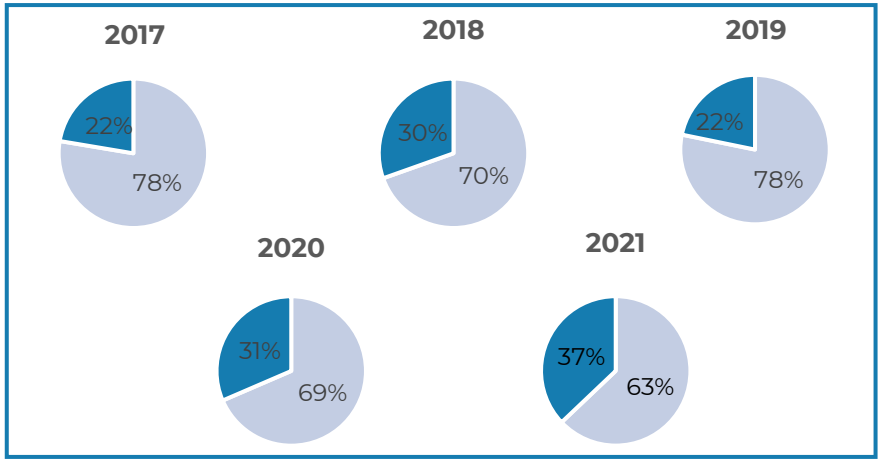
When comparing investments in TV with those in Digital Channels it is possible to clearly state that investments in **TV** are showing a **decreasing trend** (Digital Channel Appendix). On the other side investments in Digital show a more dynamic trend. Consequently we can notice that the share of investments in Digital is growing but still it represents a lower % compared to TV expenses. Similarly when comparing the Expenses in Tv and Digital with the **Revenues** in the same period the share for Digital is much lower than the TV one. It is interesting to notice how the different channel have been impacted in an opposite way by the pandemic, as can be seen when considering years 2019 – 2020: a decreasing trend in Tv compared to the increasing one for the digital channel.

DIGITAL – TOWARDS THE MODEL

Once a general analysis on the dataset was performed, we have narrowed our objective on trying to identify the most relevant variables to be kept in the model. In particular, our objective is to understand if and how it is possible to merge some of the Digital variables, mostly the ones related to the impressions, in order to build a more robust model while facilitating the analyses. With this regard we have used three approaches, which results are reported in the Digital Channel Appendix:

1. Building a **correlation matrix**: by understanding which variables are more correlated it is possible to merge them and thus avoid multicollinearity problems. In particular the analysis has been performed considering the investments in the different tools.
2. **PCA analysis**
3. Consider the type of content and type of media **semanticity**: for instance, merging all the Behavioural ones, the Sociodemo ones.





MODELLING TOOL

GOALS OF THE MODEL

For the modelling phase, the goal of the team is to create a robust data-driven statistical instrument that produces **actionable strategic** and **operational** insights for the company. On the strategic side, we want to provide a comprehensive understanding of the market, by including internal (media-related and non-media related) and external (competitor-related) variables. On the operational side, we want to give practical guidance on how to optimally allocate the marketing budget across different channels in the short and mid-term to maximize the ROI and boost the sales.

META'S ROBYN FOR MARKETING MIX MODELLING

Robyn is an open-source R library developed by *Meta* and included in the CRAN that allows a semi-automated MMM process reducing the human bias, that proved to be successful by several data-driven companies. It uses statistical and ML techniques (Ridge regression, multi-objective evolutionary algorithm for hyperparameter optimisation, …) to quantitatively define media channel efficiency and effectiveness, explore adstock rates and saturation curves.

ROBYN INPUTS

In order to adopt the library, the following inputs are required:

- **Dependent variable**: this is the primary metric that the MMM measures against.
- **Paid media variables**: The exposure (clicks, GRPs, impressions…) of any media variables with a defined marketing spend falls into this category. Transformation techniques will be applied to paid media variables to reflect carryover effects (adstock).
- **Organic variables**: Any marketing activity without a clear marketing spend fall into this category. Typically, this includes newsletters, push notifications, social media posts, and so on. Transformation techniques will be applied as well to paid organic variables to reflect carryover effects.
- **Contextual variables**: These include other variables that are not paid or organic media that can help explain the dependent variable. The difference lays in the absence of carryover effects for these vars. The most common examples of context variables include competitor activity, price & promotional activity, and so on.
- **Prophet variables**: *Prophet* is a *Meta* original procedure used for forecasting and decomposing time series data. The variables used for the decomposition are represented by the trend, the seasonality, the holidays and by an extra regressor called *events* that can be chosen by the user for modeling events that impact the business (Black Friday, company events, …).
- **Window_start**, **window_end**: This window determines the date range of the data period within the dataset that is used to specifically regress the effects of media, organic and context variables on the dependent variable
- **Adstock**: Robyn allows to an accurate representation of the real carryover effect of marketing campaigns. It reflects the theory that the effects of advertising can lag and decay following initial exposure. There are two techniques that can be adopted to model the adstock of media:
  1. Geometric: The biggest advantage of the Geometric transformation is its simplicity. It only requires one parameter called 'theta' that is the fixed linear decay rate applied to the effect of one media channel on the following periods. In this case lagged effects cannot be modeled.
  2. Weibull: Whereas a geometric adstock has just one hyperparameter (theta), Weibull is a two-parameter function that allows the modeling of a changing decay rate over time. The two hyperparameters needed to model the adstock are the shape and scale; the shape parameter controls the shape of the decay curve, that could be more L or S shaped, while the scale parameter controls the inflexion point of the decay curve.

Geometric Adstock: $Decay_{t,j} = x_{t,j} + \theta * decay_{t-1,j}$ **Weibull Adstock:** $Decay_{t,j} = x_{t,j} + 1 - e^{\frac{decay_{t-1,j}}{\alpha}}$

where $x_{t,j}$ is the investment made at time t for the media j, while θ and α are the hyperparameters

MODELLING TOOL

REGRESSION MODEL

In order to address multicollinearity among many regressors and prevent overfitting, a regularization technique to reduce variance at the cost of introducing some bias is applied. In particular, cross-validation Ridge regression with a lambda associated to one standard error from the minimum error (lambda.1se) is adopted. The final formula used for the modeling process is:

$$Sales_t = Intercept + \beta_j * \left(decay_{t,j}\right)^{\alpha} / \left(decay_{t,j}^{\alpha} + \gamma^{\alpha}\right) + \beta_{holiday} * holiday_t + \beta_{season} * season_t + \beta_{trend} + \cdots + \varepsilon$$

- Y_t = revenues at time t
- β_j is the coefficient for the regressor j
- decay t,j comes from the adstock trasformation
- Diminishing return trasformation : β_j * (decay t,j ^ α)/ (decay t,j ^ α + γ ^ α), where α and γ are the hyperparameters needed
- ε is the error

ADSTOCKS AND SATURATION CURVES HYPERPARAMETER OPTIMIZATION WITH NEVERGRAD

Robyn also allows to compute the diminishing return curves for the returns of media investments, which is done through the Hill function. This last is a two-parametric function where one parameter controls the shape of the curve between exponential and s-shaped and the other one controls the inflexion point (alpha and gamma in the formula above).

For the computation of the adstocks' and saturation curves' hyperparameters, it is necessary to priorly define for each media a range of plausible values and the parameters dimensionality increases proportionally with the total number of marketing channels to be measured. Thus, it is necessary to deal with a high dimensionality parameter space where, the greater the number of parameters, the greater the model complexity, its dimensionality and computational requirements.

In order to achieve computational efficiency while optimizing overall model accuracy, *Robyn* leverages on *Meta's Nevergrad* gradient-free optimization platform. *Nevergrad* is used to perform a multi-objective optimization problem in order to achieve two objectives:

- **Model fit**: It aims to minimize the model's prediction error (NRMSE, the normalized root-mean-square error)
- **Business fit**: It aims to minimize the decomposition distance (DECOMP.RSSD, decomposition root-sum-square distance). The distance accounts for a relationship between spend share and a channel's coefficient decomposition share. If the distance is too far, its result can be too unrealistic (media activity with the smallest spending gets the largest effect), in this way the best models should be in a certain way coherent to the decisions taken till now from the top management for the allocation of the media budget.

In this way *Nevergrad* eliminates the majority of "bad models" (**large prediction error and/or unrealistic media effect**) and identifies a few clusters of Pareto-optimal models. The final user can then analyze and study the most representative models (centers of the clusters) and take actions.

MODELLING TOOL

BUDGET ALLOCATION

After the analysis and the selection of the optimal model, *Robyn* offers the possibility to optimally allocate the budget for the media channels using the *robyn_allocator()* function. The re-allocation is made according to what the model has learnt, in particular according to the saturation curves previously built and to some parameters chosen by the final user. The faster the curves reach an inflection point and thus a flat slope, the quicker they will saturate with each extra euro spent and consequently the lower will be the budget allocated to that specific channels.
In order to customize the re-allocation budget strategy according to the specific business context, the final user can choose the overall maximum budget that is required to be allocated, the related time frame to spend it, and the minimum and maximum monetary constraints to invest in each media channel expressed as percentage of the historical mean spend. For example, if the mean spend in Google in the model's timeframe of reference was 1000 euros and the monetary constraints are 0,7 and 1,5 *Robyn*, according to the saturation level of this channel, will allocate an amount of money included in the range 700 and 1500.

MODEL CALIBRATION

In addition, *Robyn* offers the possibility to "**calibrate**" (i.e., validate) the model adopting the ***Incrementality testing*** methodology to establish the true incremental response of the dependent variable caused by each media activity during a specific time frame. Calibration requires additional data coming from experiments, that are based on models created with this library.
The experiments require to split the target audience into two homogenous groups: a test group and a control group. The test group is exposed to the media activity that has to be analysed, while the control group is not exposed to it. The difference registered in the response variable between the two groups represents the incremental effect of the media activity on the sales. Conceptually, this method is like a Bayesian method, where experiment results are used as a prior to calibrate the coefficients of media variables. For more accurate and realistic models they should be constantly calibrated.
In the code the *LiftStartDate* and the *LiftEndDate* must be provided, and they represent the time window in which the experiment took place and it that must be within the input data range defined by *window start* and *window end* defined previously. The spends sustained in the time windows of the experiments conducted on the different media channels must be provided as well and also the part of revenues attributed to them with the related confidence (if I repeat the same experiment 100 times how many times will I obtain the same results?).
In the end, after the model is built the *robyn_refresh()* function can be used to continuously retrain and refresh the model when new data arrives, making MMM more actionable.

METHODOLOGICAL STEPS DEFINITION

At this point, we aggregated all the knowledge created during the descriptive analysis with the comprehension of Robyn's characteristics and potential, and we designed a rigorous methodology to find the best way to create a marketing mix model. This methodology consists of 4 steps:

1) **Define different sets of values** for Robyn's input variables based on the results of the descriptive analysis.
2) **Set up** as many **trials** as the number of combinations of all the different values of different variables defined in the previous step, and execute them in Robyn.
3) **Define selection criteria** to select the best model, referring to the main goal of the modelling phase.
4) **Select the model and analyse its outputs** to answer the questions and allocate the company's future media budget.

The next three slides explain the decisions we made in these three steps that brought to the choice of the marketing mix model we used to answer our questions.

MODELLING METHODOLOGY

STEP 1: VARIABLE SETS DEFINITION

Relating to the description of Robyn's input variables we decided to use as dependent variables the FB's value sales (measured in LC), and as other inputs:

1) **Prophet Variables**: in all the trials we included the trend, seasonality, and holiday variables. The descriptive analysis highlights the growing volumes and sales trend (apart from the lockdown semester), so this variable is for sure fundamental to explain part of the sales' behaviour. Despite the timeseries decomposition showed autocorrelation in its residuals, we still decided to include the seasonality as well to model sales, because it still made sense to look if and how models were influenced by it. Finally, holidays were not evaluated in the first part of the project, but since Robyn made it possible to easily include them, for us and for the client it could be interesting to see how sales behave in holiday periods as well.
2) **Context Variables**: because we aimed at providing the company with **actionable** insights, we decided to include in the context variables levers that the company can decide to move which are only partially influenced by the customer behaviour, and variables that synthetically describe the competitive context, as a monitor for the competitive situation. We decided not to change them in all the trials. These variables can be divided in three groups: Focus Brand's variables, Competitor 1's variables, and other context variables.

   For modelling the price, promotion, and distribution strategy of the **Focus Brand**, we used:

   1) **Product assortment** (in number of SKUs): despite the assortment does not change much throughout the years, being the sales referred to a physical retailer sales, the product variety is a fundamental sales driver; in fact, the assortment is positively correlated with sales.
   2) **Weighted distribution any promo**: the descriptive analysis shows that this Submarket is characterized by a constantly high level of promotions, which remarkably drive the company sales. Among the various promotion variables the WD anypromo is the chosen one, because it is a lever that the company has direct control on, following our goal of using variables that thoroughly represent the brands' strategies and choices.
   3) **Price non promo**: for considering price, we decided to reject the average price (as it would include promotions, that are already modelled by WD anypromo), and we used the price without the promotions, assuming that this price is a lever that the company can move.

   Concerning the **Competitor 1's** variables, we considered:

   1) **Product assortment**: not only it is a fundamental driver of sales (for the same reasons as the Focus Brand's), but the descriptive analysis shows how the Competitor 1's strategy has changed over time in terms of assortment, reducing the gap with the FB. In fact, it is strongly negatively correlated with the Focus Brand's sales.
   2) **Weighted distribution any promo**: as for the Focus Brand, we decided to model promotions with the weighted distribution any promo variable. The competitor's promotion strategy has changed over time as well (aligning with the FB's one), boosting its competitiveness in Submarket 1.
   3) **Competitor's value sales**: despite it is not a lever that the company can freely move, this variable models both the main competitor's average price and volumes, therefore it must be included in our model.

   Other context variables:

   - **Lockdown**: the lockdown period has been modelled with a dummy variable that has value "1" in the periods more heavily influenced by he covid restrictions, that are March, April, and May 2020. In that timeframe, restrictions have been homogeneous and nation-wide, contrarily to November 2020-march 2021.
3) **Adstock**: the adstock functions we could select on were three: geometric, Weibull without lag, Weibull with lag. The last one was excluded because digital media are rarely capable of having a consistent lag, and domain experts excluded the possibility that the TV's lag could have a relevant impact on sales. Nevertheless, we carried out experiments with it and noticed that results were insufficient, both statistically and business-wise. Therefore, the adstock functions we included in our trials were two: **geometric** and **Weibull without lag**.

MODELLING METHODOLOGY

4) **Timeframe**: in our trials we decided to use two different timeframes: **4 years** and **2 years** (2nd semester 2019 – 2nd semester 2021). The two-year timeframe seemed a good compromise because it allowed not to completely exclude the pre-coronavirus period and at the same time give a higher weight (in terms of number of data points) to the post-covid one. Shorter periods (1 year and 6 months) were excluded because prior experiments showed bad results, both in terms of R^2 and residuals; having a comparable number of rows and columns is in general not a good practice, the Robyn documentation suggests not to go below the 10:1 row-column ratio.
5) **Media Variables**: of the 12 media variables that we were provided with, 3 have been excluded ex-ante, for various different reasons:
   - Amazon variables (Amazon Clicks and Amazon Display): they have been excluded after confronting with the domain experts: they stated that these investments were made mainly to drive e-commerce sales in some strategic periods for the company, and they suggested to exclude them because our sales data refer to physical sales, and these data might invalidate the analysis.
   - Google Clicks: the domain experts stated that the company's investment in this channel is on impressions, not clicks; as our goal is to provide actionable insights – by using variables that the company can directly influence to model reality – clicks were excluded and google impressions were used.

Coherently with our goals, we set up different aggregation criteria for the digital media variables (TV was always left alone). The actionability of our results can only be achieved if variables are **not too aggregated** (e.g., the digital variables altogether), because that way the decision-maker could only understand if investing on digital media or TV, with no detail on which digital media to choose. Nevertheless, the completely granular solution could not be our only option, because as already stated, the investments on digital media are very fragmented and without aggregating we would risk of underestimating their impact only because of this discontinuity. The aggregation criteria we used are **4**; in all of them, Google impressions were always left alone, first because investments were continuously made on it, and secondly because they are hardly associable to any other digital media.

Aggregation	Aggregated variables (Digital Media)	Explanation
Granular	None	No aggregation was performed
Strategic	• Programmatic Sociodemo Video + Display • Programmatic Behavioural Video + Display • Reservation Video + Social Video • Reservation Display + Social Display	The variables with a high correlation in terms of investment strategies and similar loadings (i.e., the ones which the brand invests in at the same time), introduced in slide 12 and whose data is in the Digital Channel Appendix.
Media Type	• Social Video + Display • Programmatic Behavioural + Sociodemo, Video + Display • Reservation Video + Display	The media of the same type are aggregated. This aggregation was driven by the assumption that similar media have similar impacts on consumers.
Content Type	• Social Video + Display • Reservation + Programmatic Video • Reservation + Programmatic Display	The media that present the same content type are aggregated. This last aggregation criteria was driven by the importance of adstock in Robyn's regression equation: we presumed that similar types of contents present similar adstocks in consumers' minds. It must be noted how social videos and display have been kept together: this was done under the domain experts' suggestions, as according to them social media content is different compared to the other types of content.

MODELLING METHODOLOGY

STEP 2: TRIALS SETTING AND EXECUTION

Having set 4 different media aggregations (granular, strategic, media type, content type), 2 different timeframes (4 years, 2 years), and 2 types of adstocks (geometric, Weibull without lag), the total number of trials set and executed was **16**.

STEP 3: MODEL SELECTION CRITERIA

Coherently with our goals, we set some selection criteria to achieve a model that is both **statistically robust** and consistent with the **business context**. In particular:

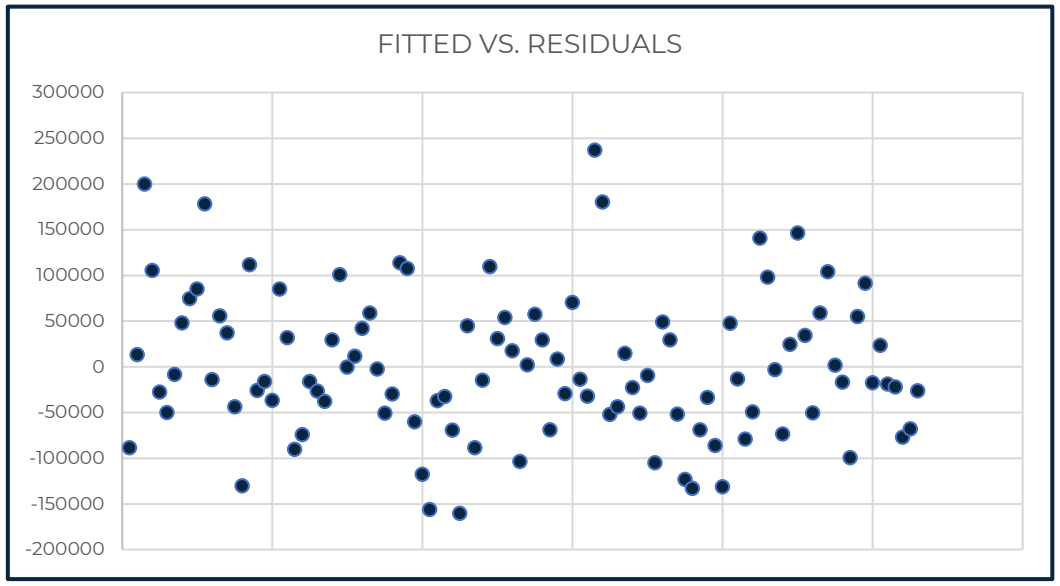
- **Statistically**, a reasonable R^2 and homoscedastic and normal residuals.
- **Business-wise**, a model that makes general business sense, and that is coherent with the main findings of the initial descriptive analysis and with the business domain experts' considerations. Two dimensions we can leverage to understand in broad terms if a model makes sense business-wise are:
    - Variable contributions to sales: because sales are made only in physical supermarkets, we expect the assortment and promotions to contribute to sales more than TV and, most of all, digital media. We believe that, especially digital media, can hardly have an important effect on sales. Moreover, the Competitor 1's variables should all have a negative effect on sales.
    - Adstocks: we expect the TV's carryover effect of the investment to be higher than the digital media's one, which should decay much faster. This because investments in TV are made more continuously, and because the effect that they have on our minds is much more enduring.

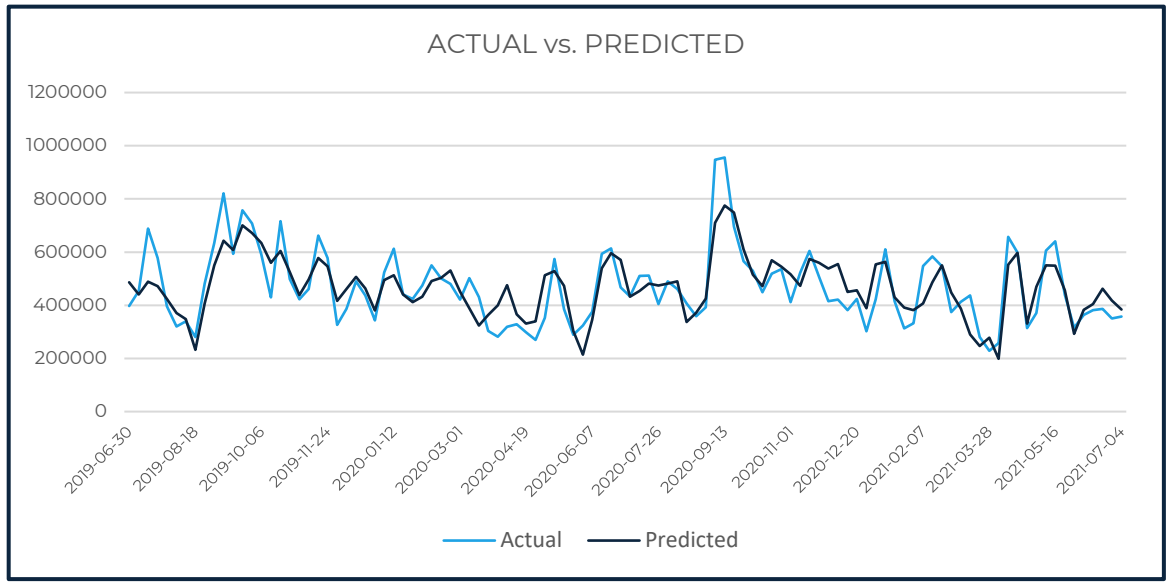
In absolute terms, these criteria and, more in general, the methodological steps we have designed and followed in this work are not enough for selecting and using a model in reality. In fact, after choosing a model trained on the given data, and eventually refreshing it with a few weeks of newly acquired information (thanks to the "refresh" function, introduced above), the model should be **calibrated**. This means that, after reaching (one or) a few promising models, their results should be tested through experiments; a common practice to do so with Robyn is **incrementality testing**, which – by dividing the population in more subsets – allows to understand the causal impact of different advertising interventions with respect to a baseline, that is one of the subset of the initial population purposely not exposed to any advertisement. The data gathered through this experiment should complement the initial model and validate – or reject – its results, and show if the parameters obtained are, in reality, significant or not. For the purpose of this exercise, and because we did not have neither the possibility nor the data to carry the calibration out, we did not include the model validation outputs as criteria for the model selection. One idea could have been to use a number of final weeks as a calibration set; however, given how Robyn's calibration is built, that would not make sense because to validate eventual hypotheses, a part of the population should have not been exposed to the media, making it impossible to evaluate the real impact on the exposed population.
As one of the Robyn creator states, being Robyn a practice and business-oriented library, the only way to obtain the "true values" for MMM with this tool is to formulate hypotheses with the model created in the first steps of our methodology, and then verify them with real experiments. This said, the results we obtained for some models in terms of R^2 and residuals seem promising, and coherent with many of the business hypotheses derived along the descriptive analysis; in fact, we believe that the opinions of domain experts that we could question along the project, and the considerations coming from our knowledge of the market through the descriptive analysis, can partially compensate the lack of the model calibration.

MODELLING METHODOLOGY

STEP 4: MODEL SELECTION AND CRITERIA VERIFICATION

The model selected for carrying out the exercise is trained on a **2-year** timeframe, sees the media variables aggregated by promotion **strategy** (i.e., investment correlation), and has a **geometric** adstock. Digging deeper on the statistical side, it presents a solid **63% R^2** and its residuals are shown below; apart from a few positive sales peaks that the model is not able to capture (corresponding to a few exceptional periods, as anticipated in slide 6), residuals are normal (p-value of the Shapiro-Wilk test is 21.3%) and look homoscedastic; apart from very few anomalies in the pattern of the residuals, their variance seems to be constant over time by analysing the Fitted vs Residuals plot.





It is worth noticing that the model we have chosen is not the best one in terms of R^2 : in fact, two models with the same timeframe and adstock but with different media aggregations returned a 64% R^2 and equally good residuals. This brings us to the business side considerations: these models were discarded because some of their results **did not make sense**. More specifically, two media's ROIs, which in both models reached absurd values, like 2200% (Google Impressions) and 1450% (reservation + social videos), therefore they were discarded.

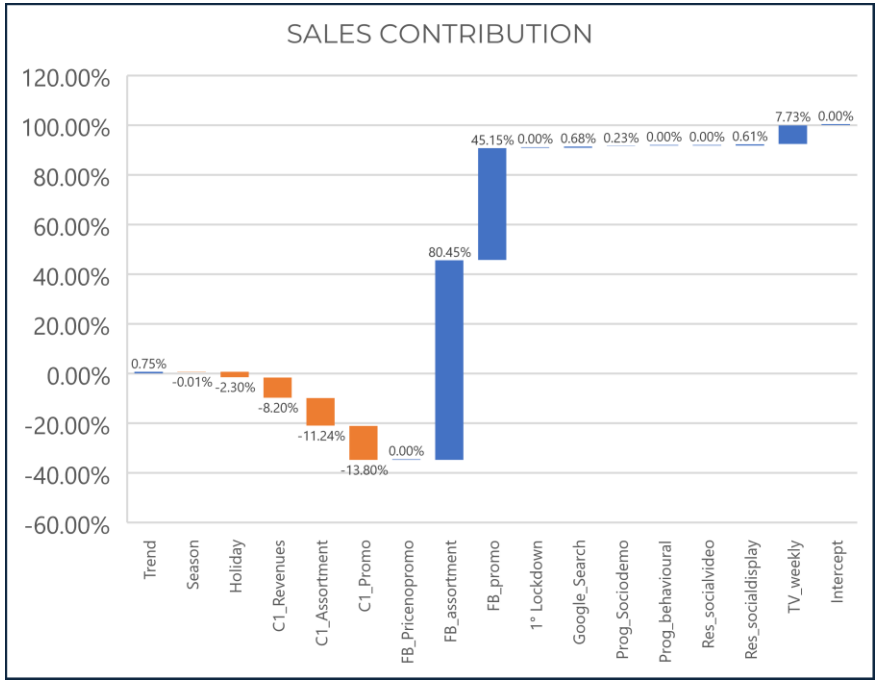
The chosen model instead has the contributions to sales heavily balanced in favor of promotions, distribution, and assortment, while media account for 9.25% of the total sales, with television having the most relevant impact among media variables (83.5% of the total media contribution). The competitor's contribution is negative for all its included variables, with promotions being the most important lever of competition. Again, this makes sense business-wise because C1 has relevantly increased its sales in the last few years, and in the same timeframe it evolved its promotion strategy towards the FB's one. In terms of adstocks, as expected TV has by far the greatest carryover, while google impressions the lowest ones, showing how the continuous investments on TV created a consistent adstock.

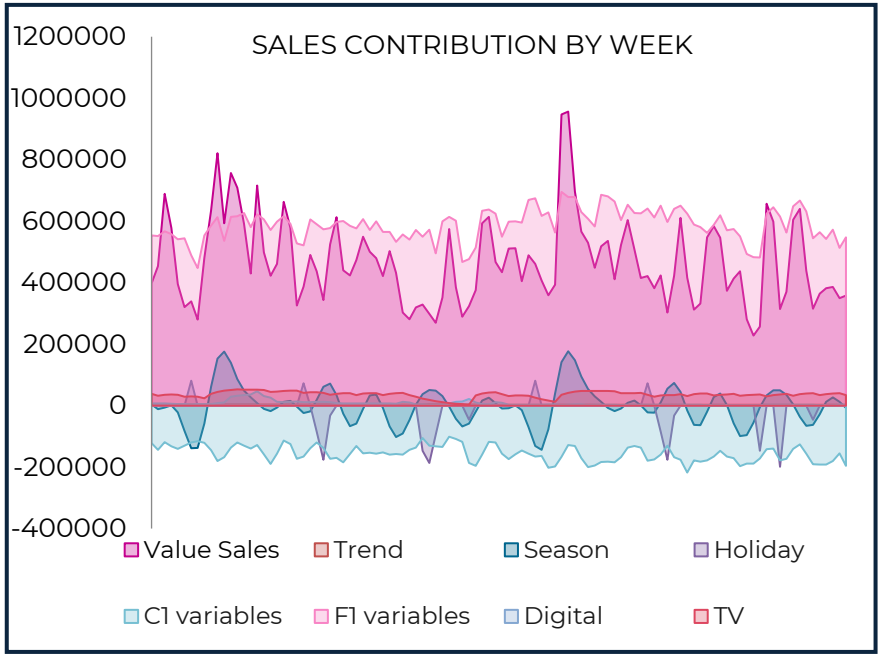
MODEL OUTPUTS

QUESTIONS 1 & 2

The graph on top shows the contributions of all variables (context, prophet, and media) to the Focus Brand's sales. To answer question 1, the media contribution accounts to **9.25%** on the total sales, of which the digital channels represents **16.43%**, and TV **85.57%**. After TV, the most relevant medium is Google search, whose contribution however is low, **0.68%**, closely followed by the reservation and social display advertisement, that contributes to **0.61%** of sales. Apart from the programmatic sociodemo media, which impact is **0.23%**, the other two media seem not to bring any influence on sales. As already anticipated, the variables that mostly contribute to sales are the assortment, promotions and distribution, which is unsurprising given the nature of the good and the fact that the value sales only account for sales in retail stores. Not by chance, the level of baseline, composed by the non-media related variables, contributes to **90.75%** of sales. In particular, it is made up of:

- **Competitor-related variables**: A negative impact amounting to **–33.2%** of the competitor-related variables. More specifically, the most contributing variable is the competitor's promotion, showing that, as hypothesized in the descriptive analysis, its new promotion strategy seems effective in eroding the Focus Brand's market shares. The second most impacting variable is the assortment, which is close to promotions for its impact, and that proves to be important for the competitor as well. Finally, its value sales: because the growth rate of this Submarket is positive but it is lower than the sales' growth of the Focus Brand's main competitor, we expected that its value sales negatively impact the Focus Brand's ones, furtherly confirming how the erosion of the Focus Brand's value sales is attributable to the fiercer competition.
- **Focus Brand's variables**: compared to other business models, above all online platforms that advertise online, the media variables' contribution to sales is significantly low: if we consider a consumer that shops in a grocery store, that is exposed to in-store price promotions or a wide assortment, it is easy to picture that these variables can be much more impactful in its decision-making than digital advertising for a FMCG as yoghurt. So, assortment represents by far the most important contribution to sales (**80.45%**), accounting for the greatest slice of the baseline. The message for the brand is clear: its high number of SKUs in the grocery stores' shelves makes a huge impact in selling. Promotions instead are still very relevant (**45.15%**), but not as assortment is.
- **Prophet variables**: despite seasonality has a low cumulated impact on sales (graph on top, the bottom figure shows that seasonality's impact is very relevant in some periods. The positive and negative peaks shown there sum up to an aggregated **-2.3%** contribution, which certainly does not well represent the real impact that seasonality has on sales. Finally, it must be noted that the trend variable has a slightly positive impact; this can be both a symptom of low growth of the company brand's sales, but it can also be due to the heavy weight of assortment, which presents a slightly increasing trend in this timeframe, and the model might have partially underestimated the real effect of the pure market trend, which after COVID showed very good results.





MODEL OUTPUTS

QUESTION 3

In order to calculate the ROI and effectiveness of TV and digital channels year by year, and consequently evaluate the changes occurred, it was decided to apply the Robyn model with the same variables and time-span (2 years) starting from 2017. The results were three models (2017-2019, 2018-2020, 2019-2021) that allow to understand how the context changed along the years. The ROI, calculated as written below, was given by the Robyn model itself, while the effectiveness was calculated starting from the incremental effect on value estimated by the model and by doing the following calculation:

$$\boldsymbol{ROI} = \frac{Incremental\ Value\ Sales\ [€]}{Media\ Costs[€]} \qquad \boldsymbol{Eff.} = \frac{Incremental\ Volume\ Sales\ [Units]}{Impressions\ (or\ GRP)} = \frac{\text{Incremenal effect on value [\%]} * \text{Total Value[€]} * 1000}{Mean\ Price\ [€] * Impressions\ (or\ GRP)}$$

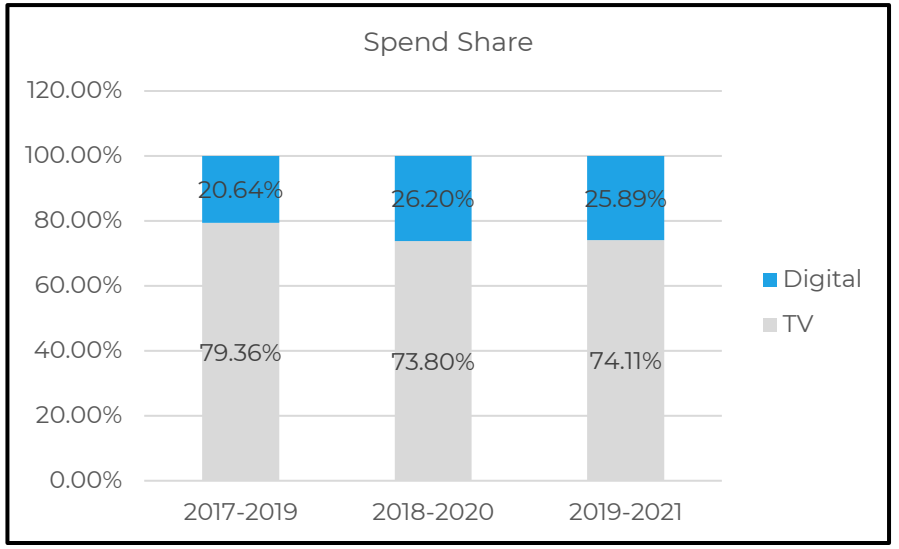
RESULTS

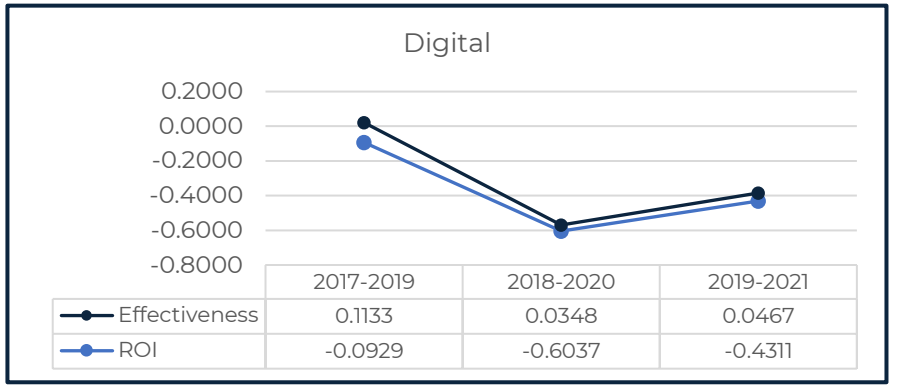
Before deep-diving in the result, we must make some considerations related to:

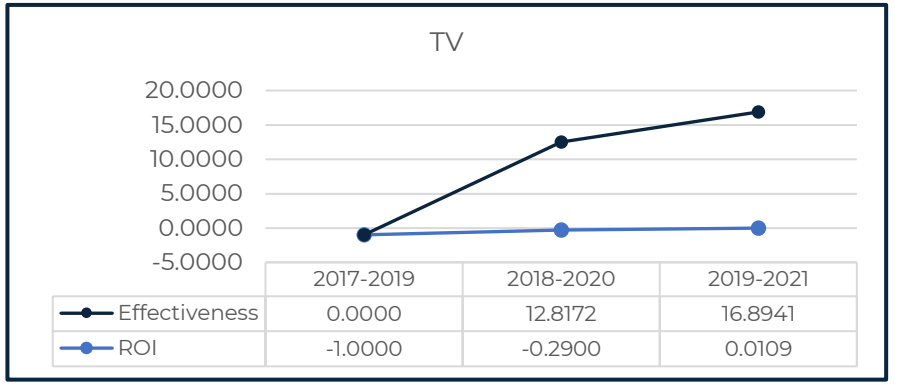
- **The context**: the company works in a market where sales are only in shops and consequently all advertising, both TV and digital, has a very low impact on final sales.
- **The covid period**: which impacted both sales, even if only to a small extent, and advertising, which was interrupted or reduced, perhaps due to the uncertainty of the context or some other causes.
- **The company's use of advertising**: considering the context mentioned above, the company in the period of analysis adopted different channels and different tactics, moreover not in a continuous manner as it might be in other contexts.

These elements certainly contributed to the **lack of continuity of results** in the different media, making interpretation complex.

The results obtained, aggregated into total TV and digital, show particular shapes, and a deep comprehension of the context is needed to explain them. Starting from the TV, the initial performances were bad and maybe they were the results of bad campaigns and excessive budgets' allocations which guaranteed flat GRPs that the model could not detect. In the following years, probably due to new tactics, the performances improved until gaining a positive ROI while lowering the investments. As regarding the digital channels, in 2017-2019 the results showed good performances, which probably are a consequence of the low budget allocated which guaranteed discrete results. Instead the following years, they proved to be less effective but it could be the result of an higher budget invested after Covid-19 and its wrong allocation, by pushing some channels instead of others and by changing frequently media.







MODEL OUTPUTS

QUESTION 5

SATURATION CURVES

Saturation curves are built through the Hill function's parameter definition during the hyperparameters optimization, based on the best model defined with our methodology (so, on a 2-year timeframe). As it can be noticed from the graph on top, some curves are obtained in their entireness, while for some others Robyn struggles to define the typical C or S-shape; for example, the reservation + social display and Google search curve. We can distinguish three distinct media groups:

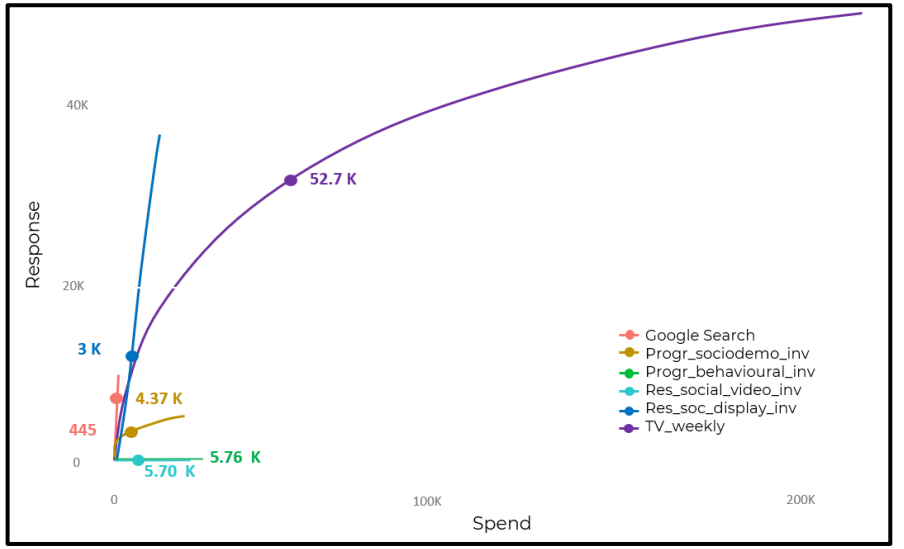
- **Unsaturated media**: Google Search and Reservation + Social display's curves seem at the beginning of a classical C shape, where the slope is still very high.
- **Saturated media**: TV and programmatic sociodemo seem to have reached a level in which higher investments would not yield satisfactory returns, since their slope is getting flatter.
- **Zero-response media**: according to the model, reservation + social video and programmatic behavioral do not contribute to sales, therefore their saturation curves are flat: investing in these media would be useless.

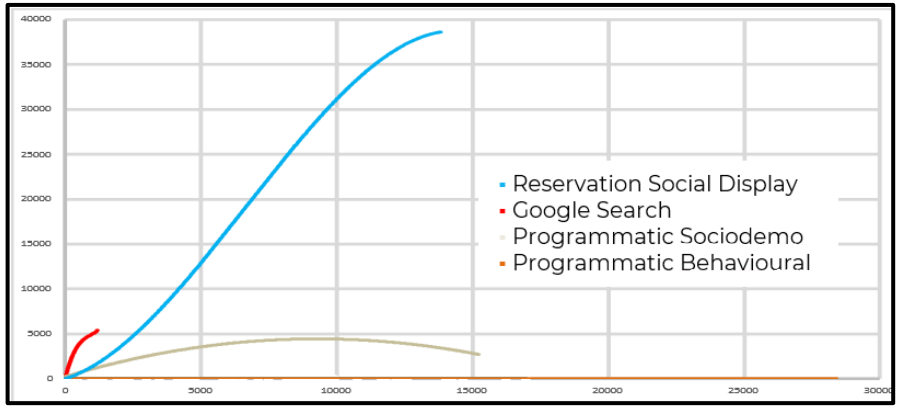
Given the huge differences between these three groups and the different scales they present and given that Robyn builds these curves by estimating two hyperparameters and using them in a formula, we furtherly explored the saturation curve situation by using a different method for tracing them (i.e., polynomial trendlines based on the weekly datapoints). While most curves did not change, the reservation + social display curve assumes an S-shape with a different behavior compared to Robyn's (graph on the side). Nevertheless, its saturation is still very low and thus it remains a medium worth investing on.

As a final consideration, it must be remarked that more continuous investments and data collection are required to make the saturation curves' results more reliable, because as for now, they are built upon not so many data points, making it hard to derive saturation curves with both methods.

DECAY RATES

Defining the carryover as 1 – decay rate, we can see from the table aside (which refers to the 2-year model selected through our methodology) that the greatest carryover is reached by TV. Not by chance, TV is the only medium in which the Focus Brand's investments over the 4 years have been continuative. Most of the digital media present a similar behavior - with a carryover between 20% and 30%, apart from the programmatic behavioural media which presents a much lower adstock. Finally, despite the impressive ROI, Google as well shows a low carryover effect.



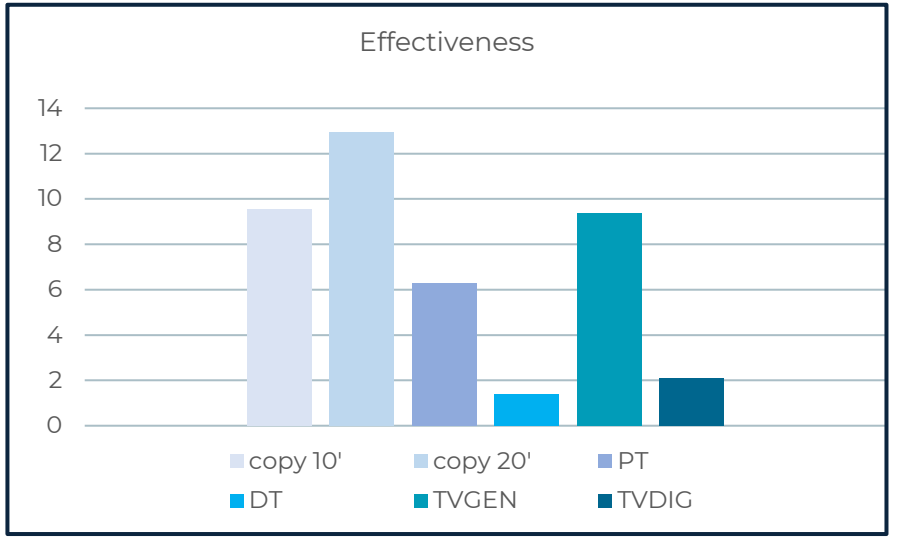


Media	Carryover
TV	62.71%
Reservation + Social Video	23.33%
Reservation + Social Display	20.54%
Programmatic Sociodemo	29.15%
Programmatic Behavioural	3.41%
Google	4.37%

MODEL OUTPUTS

QUESTION 4

To calculate the effectiveness of the various television tactics, three models similar to the best found in the previous points were created, but with the difference that the impressions and expenses for television were **segmented by tactic**. This led to compute the sales contribution of each of them.
Considering the increase in sales in monetary terms, the calculation of effectiveness has been adjusted by the Ratio (formula below) – which basically represents an average price on a two-year timeframe, so the total sales and total volumes are only those used in the period of interest. The higher the difference in price per impression, the higher the difference in effectiveness. The worst effectiveness is during day-time, while the best effectiveness is when the length of the advertisement is 20 seconds. As anticipated in the descriptive part, the 30" copy was not considered in this analysis since it was not very much used, especially in the model's timeframe of reference.



$$\boldsymbol{Effectiveness} = \frac{Incremental\ Volumes\ due\ to\ Media\ [Units]}{GRP}$$

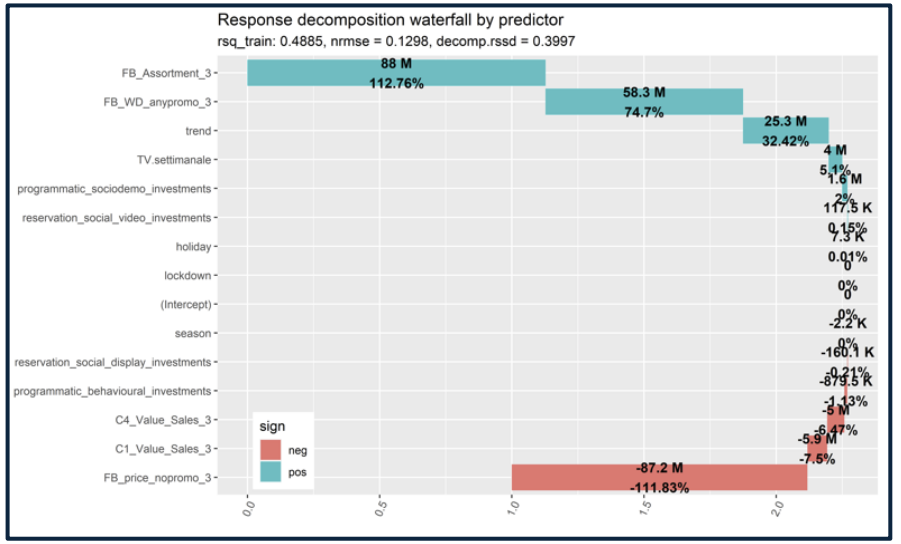
$$\boldsymbol{Ratio}\ (Avg.Price) = \frac{Total\ Sales\ [€]}{Total\ Volumes\ [Units]}$$

QUESTION 6

We applied the same methodology we designed for defining the best model to Submarkets 2 and 3. However, since in these Submarkets there are more relevant competitors, we only used their value sales to model their behavior (without the assortment, weighted distribution anypromo, and prices), to keep a lower number of columns. By applying the same statistical and business criteria to select the best model, we noticed **that Submarket 2 never met both at the same time**, therefore the conclusions we could draw from them were either not very relevant statistically (15-30% R^2) or absurd business-wise (too high contributions from both TV and social media that brought to huge ROIs, above 1000%). Instead, the model chosen for Submarket 3 is sufficiently good for the purpose of the exercise: its R^2 is almost 49%, negatively influenced by the scarce ability of the model to capture the positive sales peaks, as shown by its residuals (question 6 appendix).
The media effects on sales are very different:

- Reservation + social video, reservation + social display, and programmatic behavioral investments present **almost-0 contributions** to sales.
- Television's halo effect is not negligible, since it yields a 5.1% contribution, meaning that the Focus Brand's Submarket 3 sales **benefit from the investments in TV**.
- Programmatic sociodemo seems very effective in terms of sales generated, reaching a **high contribution** (2%), compared to the one showed for Submarket 1.

It must be highlighted how some results are not very convincing: regarding programmatic sociodemo, it seems that investments on this media made for Submarket 1 yield huge returns in Submarket 3, thus much higher compared to the ones in the Submarket they are directed to. The step forward to understand these values' significance is to proceed with the model calibration.



CONCLUSIONS

BUDGET ALLOCATION

The tool's budget allocation function has been used on the selected model to understand how to optimally allocate the media budget in the next 4 weeks to maximize the ROI. As weekly budget, we considered the average media budget of the last 2 years (model's reference timeframe), that is **51.4k€**; the model suggested to **reduce the media investments by 28.6%**, which is mainly due to the null contribution of two media to the brand's sales. If the remaining 36.7k€ are spent as suggested, Robyn forecasts a 2.6% media-contributed sales (i.e., 9.25% of the total sales) increase, which corresponds to roughly 1k€ per week. This results show that the company's current budget on media is too high: with almost 30% less, a slight increase in revenues can be achieved.

MODEL BENEFITS AND FUTURE STEPS

Because of a new "restriction on cookie policy," the accuracy of almost all the tracking systems that adopted a *Multi-touch attribution* technique dramatically decreased and this constituted a serious challenge for many marketers. For this reason, the importance and the impact on business performances of MMM increased in the very last few years, and it is likely to keep growing.

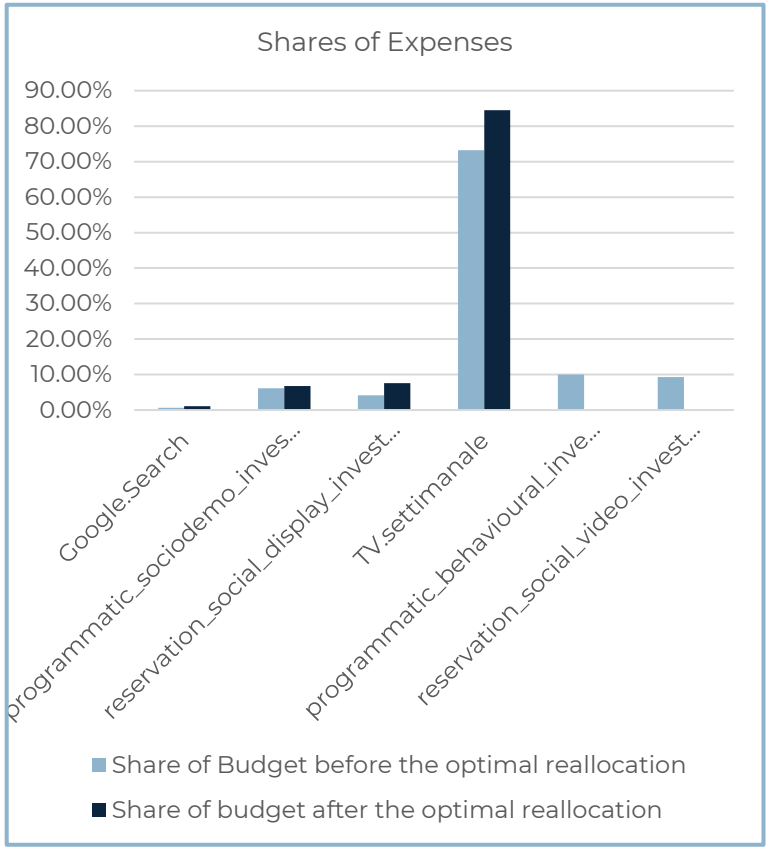
As a group, we decided to adopt *Meta's Robyn* tool because it proved to be successful in real business contexts and because it is an instrument that all companies can use. Indeed, it is an open-source library that allows at the same time a high level of automatization and to have a customized Marketing Mix Model that is constantly updated on new data. Moreover, especially small/medium-sized companies encountered big barriers in the adoption of Marketing Mix Models because generally these projects are promoted by big consulting companies that charge very high prices.

Finally, it is important to underline that, in order to build a robust and reliable model and accomplish the tasks required the work is not done and the further improvements for the Focus Brand are:

1) **Validate the model** through the frequently mentioned calibration process, by testing it with real data (incrementality testing), retraining it, and confronting its results with the company's domain experts to finetune its parameters.
2) **A more accurate data collection and warehousing process**. Concerning sales, the data provided were collected at a weekly granular level, while for investment data the granular level was monthly and to convert everything on a weekly basis the average was adopted, reducing in this way the variability of the data.
3) **Industrialize the model**: redesign the current MMM definition and media budget allocation process, including the software in the company's architecture to feed it with straight-from-source data, and upskill/reskill the company's marketers and data scientists to use it, update it (thanks to *refresh*), and fix it where needed.

More detail on both the budget allocation and the future steps for the company are provided in the **business presentation**, the other relevant deliverable of this project.

In conclusion, it should be highlighted that MMM is more suitable for other types of businesses, especially for the e-commerce ones, where the importance of media is higher



REFERENCES AND ACKNOWLEDGMENTS

A special thanks to Cristian Nozzi, founder and CTO of *Cassandra*, for the time he dedicated to clarify some of our methodological doubts.

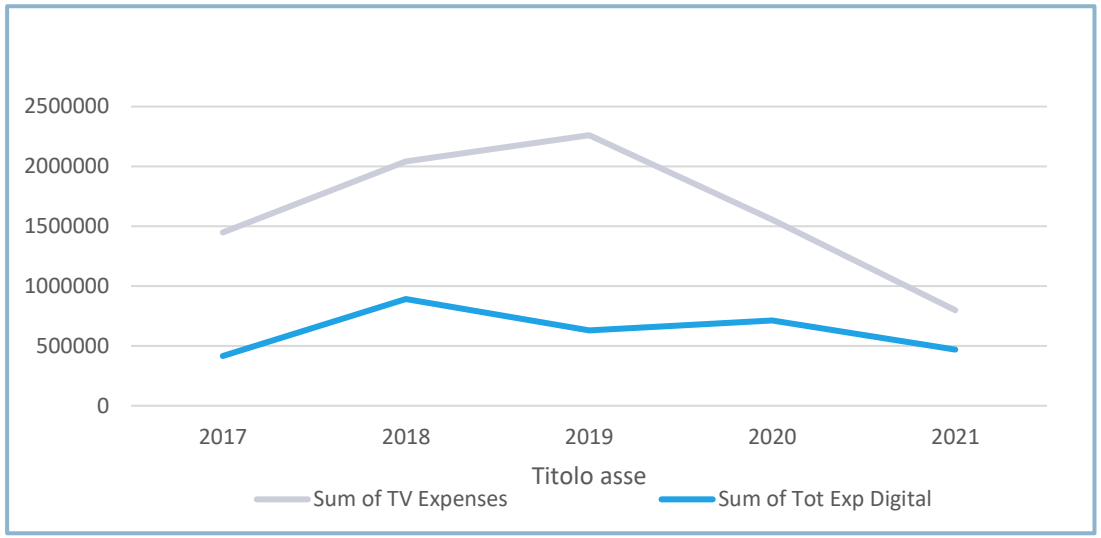
SITOGRAPHY

- "All you need to know about Marketing Mix Modelling" Course – by Cassandra
- "Facebook Robyn" Masterclass by Cassandra
- "Marketing Mix Modelling" Masterclass by MASS Analytics
- https://www.linkedin.com/pulse/how-use-mmm-increase-your-marketing-roi-gabriele-franco/?trackingId=ZaDDn6DIF5c3EdjaOPoVLA%3D%3D
- https://cran.r-project.org/web/packages/Robyn/Robyn.pdf
- https://facebookexperimental.github.io/Robyn/docs/analysts-guide-to-MMM
- https://facebookexperimental.github.io/Robyn/docs/features
- https://github.com/facebookexperimental/Robyn/issues/131
- https://github.com/gufengzhou

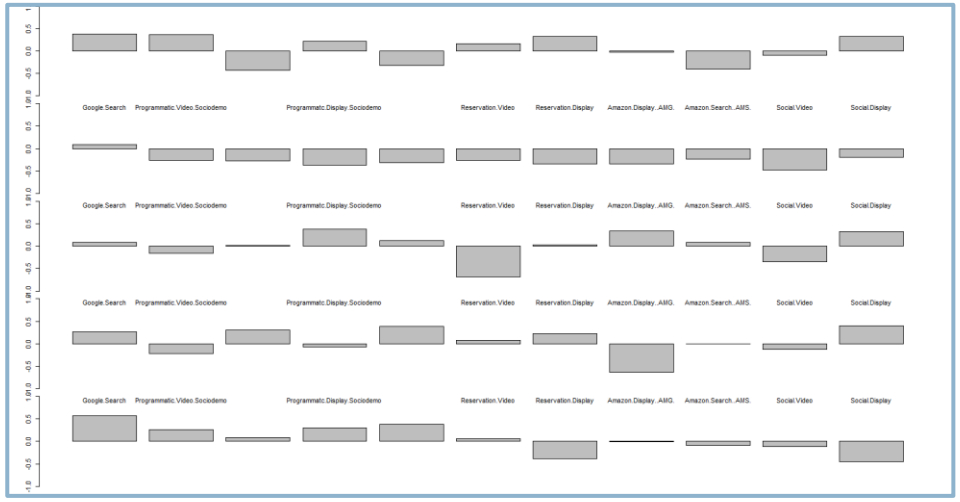
APPENDIX: DIGITAL CHANNEL

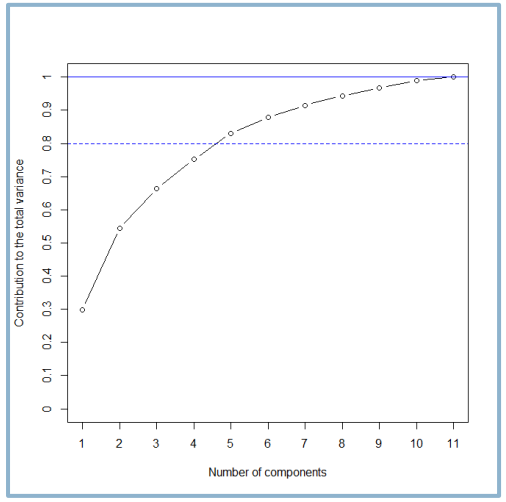
DIGITAL VS. TV INVESTMENTS COMPARISON (FIGURES IN €)

Year	TV Expenditures	Digital Expenditures	Value Sales (x1000)	% Tv Expenses on Revenues	% Digital Expenditures on Revenues
2017	1,447,659	417,017	9,724,302	15%	4%
Sem 1	-	-	-	-	-
Sem 2	1,447,659	417,017	9,724,302	15%	4%
2018	2,042,368	893,373	22,386,971	9%	4%
2019	2,259,797	629,453	25,984,914	9%	2%
2020	1,554,402	713,991	24,290,156	6%	3%
2021	798,013	470,380	11,538,691	7%	4%
Sem 1	798,013	457,334	11,181,098	7%	4%
Sem 2	-	13,046	357,592	0%	4%
Grand Total	**8,102,239.00**	**3,124,214.31**	**93,925,034.06**	**9%**	**3%**

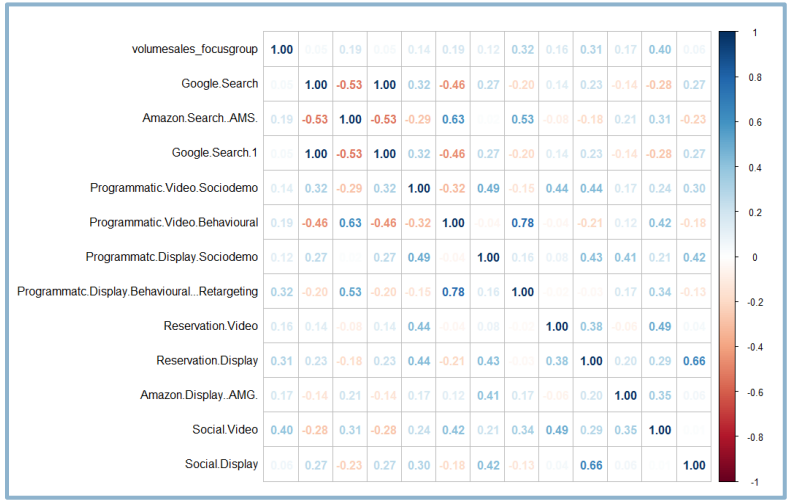


AGGREGATION – PRINCIPAL COMPONENT ANALYSIS





AGGREGATION – CORRELATION MATRIX



APPENDIX: QUESTION 6

MODEL'S RESIDUALS