Digital Marketing Attribution Model

Unilever Team: Jim Keyt, RL Rosqueta & Molly Baker
Program Mentor: Michelle Li
Faculty Advisor: Retsef Levi

Danying Xiao, MBAn '20

Jonathan Chan, MBAn '20

About Unilever

Unilever is a multinational consumer goods company, headquartered in London, England Its products include food, energy drink, ice cream, tea, cleaning agents, beauty products, and personal care products.

Project Objective

Develop the ability to better understand which, and by how much, marketing and in-store trade promotion efforts influence Breyers and Hellmann's sales so that marketing teams can better focus their initiatives on high conversion activities.

Product Functionality

Business Needs	Corresponding Functionality
Frequently, real-time	• Reports automatically generated quarterly
Data Gathering	• Leverage existing data in Domo (technical) • Minimize number of flat files to be uploaded to Domo (business process)
Data Validation	• Engage Mindshare and Digital Hub Measurement Lead to interpret each report (business process) • Generate reports at the campaign (if possible) and channel level to balance actionability and reliability (technical)
Understand Campaign Impacts	• Using lens 1 model outputs to develop dashboard 1 for tracking attribution by campaign
Understand Campaign Performances	• Using model 1 outputs to calculate campaign performance metrics (incl. ROI, CPMs, Engagements, Impressions, Spend, Effectiveness), visualize in dashboard 2
Understand Social Media User Activities	• Using model 2 results to develop dashboards for tracking attribution by online marketing metric
Comparison	• Develop QvQ (e.g. Q2 2020 vs. Q1 2020) reports • Develop QvY (e.g. Q2 2020 vs. All 2019) reports

Business Impact

Challenges	Solutions
• **Time lag** inhibits ability to action on insights	• Reports generated **quarterly**
• Significant effort to **collect and validate data**	• Most datasets **update automatically** in Domo • Simple **file drop** for other datasets
• Unclear on how to best allocate spend given the **disruption of BAU**	• More **frequent** pulse on performance
• Diagnostic reports are not indicative of the future (ie. **prediction**)	• Machine Learning pipeline established, allowing predictions with more **data collected**
• Reports are brand specific and do not **contextualize** across all brands	• Simple to **scale up** to other brands

Project Scope & Timeline

- **Feb 2020** — Defining Project Scoping & Data Collection
- **Mar 2020** — Data Preprocessing & Feature Engineering
- **May 2020** — Model Development & Implementation
- **Jun 2020** — ETL Construction
- **July 2020** — Product Handover

Approach Overview

Raw Data	*Feature Engineering*	*Exploratory Analysis*	*Modelling*
• Sales (IRI)	Engineered weekly sums and weekly averages as features from the raw data	Considered various Sales baseline models and decided to use Incremental Sales from IRI	Explored Ridge, Lasso and OLS models and cross-validated, chose the final model based on the balance between interpretability and model accuracy to develop Dashboard 1
• Trade Promotion	Created campaign impression time series table from paid media and offline marketing data based on campaign, platform and duration	Generated correlation matrices for each feature with Incremental Sales	Develop Dashboard 2 using campaign performance metrics
• Paid Digital Media (Mindshare)	Engineered performance metrics such as ROI, Effectiveness and CPM	Reviewed Campaign placement name and filtered out added-value campaigns (Campaigns that had 0 spends)	
• Print • TV	Lag data for every week up to 8 weeks in order to consider the lagging effects in marketing	Selected 20 highly correlated features from lens 2 datasets based on correlations with incremental sales and marketing domain knowledge	Develop Ridge, Lasso and OLS models to develop Dashboard 3, eliminating features with negative coefficients and low statistical significance
• NA Stacked Clean • NA Social New Followers • Google Analytics F&R Output			

All Dashboards | Dashboard 1 &2 | Dashboard 3

Modelling

The below two formulas represents the backbone of our regression models. We explored different regression methods, including Ridge, Non-Negative and Ordinary Least-Squared Regression with recursive feature elimination to find the best model based on model accuracy and interpretability.

Lens 1 Model: Incremental Sales Attribution by Campaigns

Incremental Sales ~ a_1 $Campaign_1$ + …+ a_N $Campaign_N$ + b_2 Trade

Variables:

- *Dependent Variable (Time Series, Continuous):*
  Incremental Sales (in dollar amount)
- *Independent Variable:*
  $Campaign_{i,t}$ (Time Series, Continuous):
  Campaign Impressions at Week t

Lens 2 Model: Incremental Sales Attribution by Social Media User Activities

Incremental Sales Volume ~ a_1 **Email** Session Time + …+ a_N **Paid Social Media** Page Views + b_1 Trade

Variables:

- *Dependent Variable (Time Series, Continuous):*
  Incremental Sales (in dollar amount)
- *Independent Variables (Time Series, Continuous):*
  Different Social Media Performance Measurements

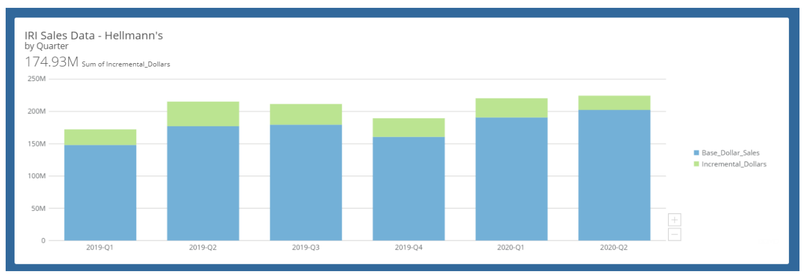
The final model that was chosen as default is the Non-Negative Regression model.

Results for Hellmann's

Below are the dashboards that have been created for Hellmann's as of July 23, 2020. Breyer's page has the exact same types of visualizations. We are showing Hellmann's results as a sample analysis on the insights.

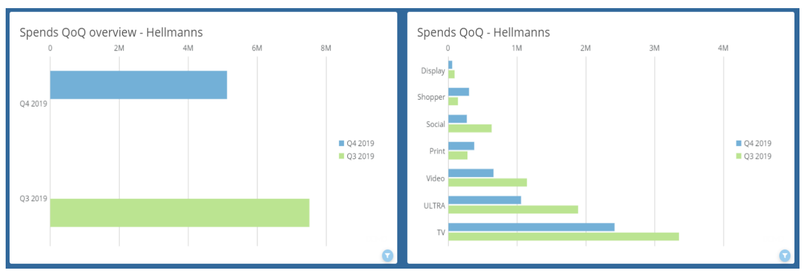
Sales Overview

Figure below show the sales quarterly breakdown in base dollar amount and incremental dollar amount.



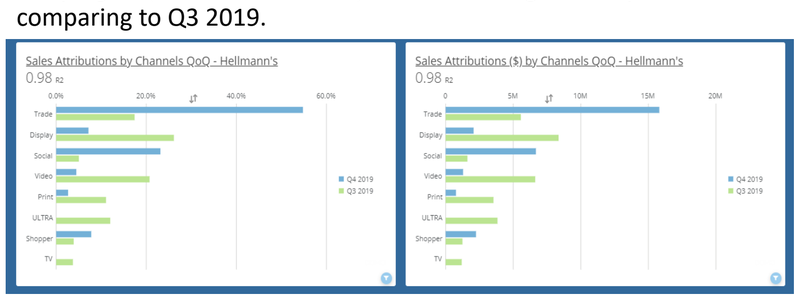
Spends Overview

Figure below shows the digital media marketing spending by quarters and channels. From this, it can be seen that the 33% decrease in spend is largely in its highest spend channels such as TV, Ultra and Video. Shopper increased in Q4 due to US Thanksgiving. Spend is an important metric as higher spends generally correlate with increased sales (ie. as spend increases, sales also increase)



Lens 1: Incremental Sales Attribution by Channels

Below are the outputs of a non-negative regression. It shows the sales attributed to each of the channels in percentage and dollar amount. As we can see that the incremental sales attributed from *media campaigns (other than Trade)* decreased significantly in Q4 2019 comparing to Q3 2019.



Lens 1: Top Performing Campaigns Leader Board (QoQ)

The below performance metrics are calculated by normalizing contributions by impressions/spends, in order to compare channel performance more fairly.

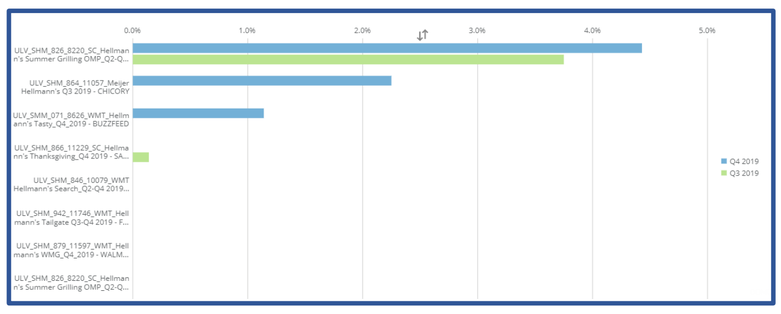
- Effectiveness = $\frac{\text{Total Channel Incremental Sales Contribution}}{\text{Total Channel Impressions}}$
- ROI = $\frac{\text{Total Channel Incremental Sales Contribution}}{\text{Total Channel Spends}}$

Measurements Rank - Hellmann's

	Q4 2019		Q3 2019	
Channel	ROI Rank	Effectiveness Rank	ROI Rank	Effectiveness Rank
Display	1	1	1	1
Social	2	2	4	5
Shopper	3	3	2	3
Video	4	4	3	2
ULTRA	5	5	5	4

Lens 1: Drill-down View on Campaign Contributions

The figure below shows a more granular view on the contributions (i.e. at the shopper campaign level). This drill-down view is deigned to help brand manager to have a further understanding on each campaign and give extra evidence to support the decision of re-running or removing certain campaigns.



Lens 1: Top Performing Campaigns Leader Board (QoQ)

The table below shows the top performing campaigns across the current quarter and previous quarter. The results shown are a mixture of the three models generated in order to achieve results stability. This is valuable to Marketing teams as it drills down on which campaigns are successful. It may help teams decide whether to re-run successful campaigns and to investigate why certain channels and sites are underperforming for the same campaign.

Top Performing Campaigns for Current Quarter

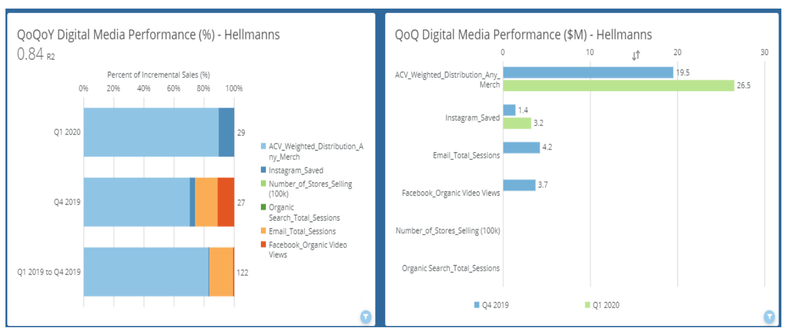
Channel	Campaign&Site
Social	UNE_HL_628_Hellmann's Franklin 2019 - PINTEREST
Social	UNE_HL_628_Hellmann's Franklin 2019 - BELIEVE ENTERMNT GRP
Video	UNE_HL_628_Hellmann's Franklin 2019 - HEARST NETWORK
Print	PEOPLE MAGAZINE_Print_Hellmann's
Video	UNE_HL_628_Hellmann's Franklin 2019 - DISCOVERY NETWORKS
Shopper	ULV_SHM_942_11746_WMT_Hellmann's Tailgate Q3-Q4 2019 - FACEBOOK
Social	UNE_HL_628_Hellmann's Franklin 2019 - BRAND NETWORKS
Shopper	ULV_SMM_071_8626_WMT_Hellmann's Tasty_Q4_2019 - BUZZFEED
Print	COUNTRY LIVING_Print_Hellmann's
Print	BETTER HOMES&GARDENS_Print_Hellmann's
Trade	ACV_Weighted_Distribution_Any_Merch
Display	UNE_HL_628_Hellmann's Franklin 2019 - MEREDITH CORP

Top Performing Campaigns for Previous Quarter

Channel	Campaign&Site
Video	UNE_HL_628_Hellmann's Franklin 2019 - HEARST NETWORK
Display	UNE_HL_628_Hellmann's Franklin 2019 - MEREDITH CORP
ULTRA	UNE_HL_628_Hellmann's Franklin 2019 - THE TRADE DESK INC
Social	UNE_HL_628_Hellmann's Franklin 2019 - PINTEREST
Shopper	ULV_SHM_826_8220_SC_Hellmann's Summer Grilling OMP_Q2-Q3 2019 - SAM'S C
Social	UNE_HL_628_Hellmann's Franklin 2019 - BRAND NETWORKS
Print	EATING WELL MAGAZINE_Print_Hellmann's
Print	GOOD HOUSEKEEPING_Print_Hellmann's
Print	BETTER HOMES&GARDENS_Print_Hellmann's
Trade	ACV_Weighted_Distribution_Any_Merch
Video	UNE_HL_628_Hellmann's Franklin 2019 - DISCOVERY NETWORKS
Shopper	ULV_SHM_864_11057_Meijer Hellmann's Q3 2019 - CHICORY

Lens 2: Incremental Sales Attribution by Digital Media Activities

The figure below shows the attribution to the highly correlated features over the current quarter, last quarter and last year. On the left, the contribution is shown as percentages, and on the right, it is shown in dollar amount. What we can observe is that in Q1 2020, there was little digital media attribution compared to Q4 2019 and all of 2019. It appears that Email Total Sessions and Facebook Organic Video Views, which have historically contributed, did not contribute at all to Q1 2020. Insights like these allow the Marketing team to dig deeper and understand why these metrics did not perform in that period.



Lens 2: Correlations between Incremental Sales and User Activities Overtime

This last figure shows the correlation of top, key features with incremental sales and the number of datapoints available. The features are color coded based on their dataset.

It aims to track digital activities' correlations with sales overtime. It does not imply causation, but if consistent trends are observed across these variables overtime, it can be potentially defined as a leading indicator of incremental sales for a brand manager to track more closely, instead of getting overwhelmed by thousands of different metrics that are currently being extracted from all the APIs.