Sales Forecast with ML - PoC

June 2023

The proof-of-concept would aim to cover the 2 use-cases of “LL forecast” and “trend forecast”

	Long-term forecast LRF	**Mid-term** Campaign/Demand forecast LL	**Short-term** Trending/Early sales reforecast
Projection horizon	7-12 months	6 to 3 months ahead	End of campaign month
Review Frequency	Monthly	Rolling revision. Focus on lag 3	Daily refresh; with focus on Day 3/7/14
Benchmarks		▸ LL forecast + Foresight = official estimates, validated	▸ R-factor method (using Cash Forecasts)
Proposed pilot scope	*Out of scope*	**POLAND** ▸ Vehicles : Brochures only (exclude other distribution channels for the PoC) ▸ All categories	**ROMANIA** ▸ Multi-vehicles distribution (brochure, online, special offers), each with their own timing ▸ All categories

Main focus for today

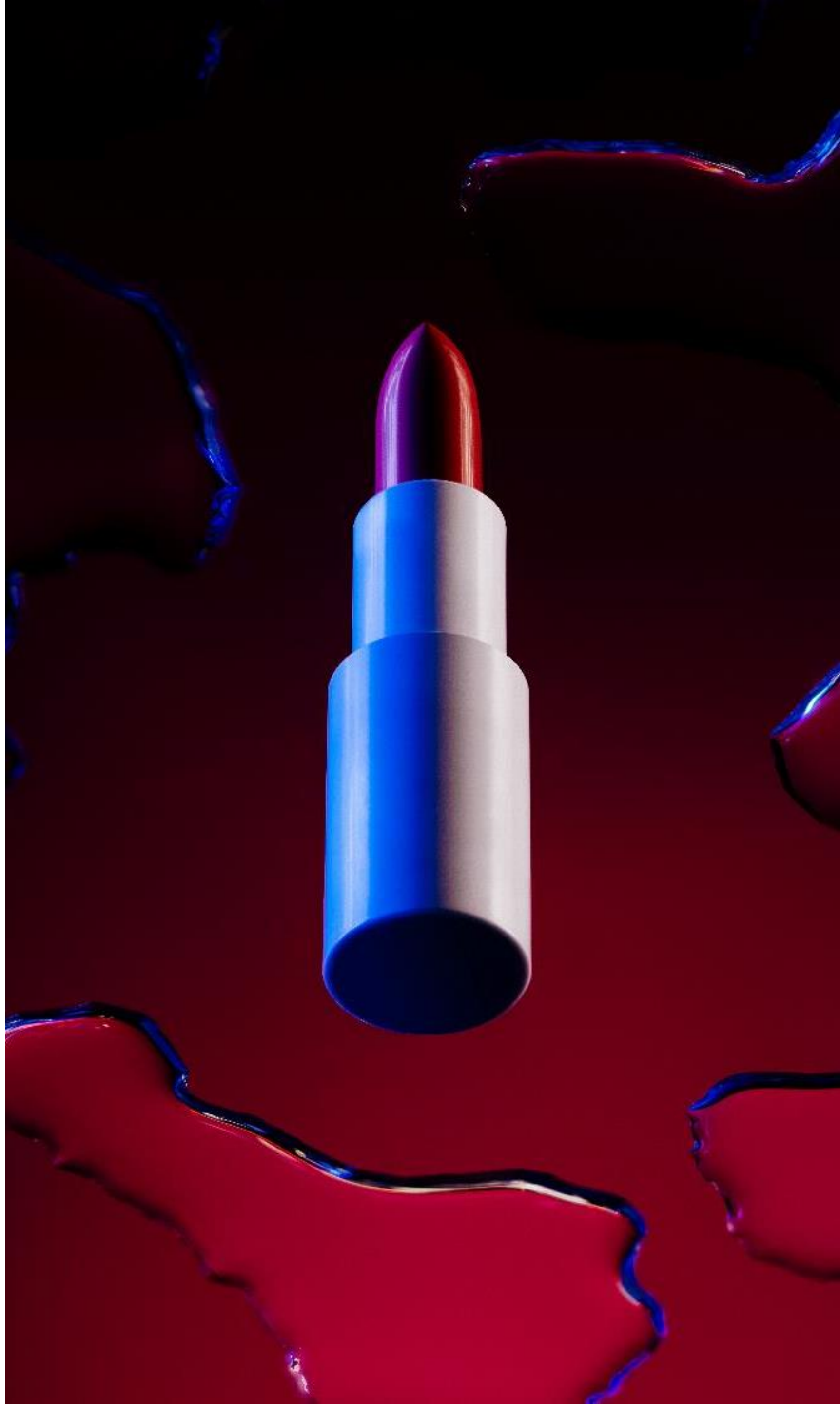
IRIS by Argon&Co

Today is our second iteration for the « Trending » PoC

	May	June	July
0. Framing & data collection	May 4th Kick-off; **Data Collection**: Shared as flat files extracts; Data discovery & catalog. **Framing**: Validate expected output and benchmark methods for both use-cases (1 & 2)	today	
1. PoC model – Early sales / Trend	**Model development**: Customization from our Python librairies; Training on AVON's data; Performance backtesting & iterations + sharing with AVON team	**Model development** (continued)	Go/No-Go for Prototype — 7th July
2. PoC – Base forecast		***Go for DEV*** *Actual start depending on IRIS' team availabilty*; **Model development**: Customization from our Python librairies; Training on AVON's data; Performance backtesting & iterations	**ReX & Next steps**

Weekly update · Workshop · Milestone

Contents



Our last model incorporates advanced features to better address various limitations in previous models

	Model split	Level	Algorithm	Features								Other		
				# of features	Actuals @Day 3	Previous campaigns	SKU features	Price	Offer features	Calendar features	Global Ratio	Retrain frequency	Hyper-parameters tuning	Feature selection
R-factor proxy	Market	Campaign	R-factor	2	Basic	No	No	No	No	No	No	Na	Na	Na
Model 10 - baseline	Market	Day	LightGBM	1	Basic	No	No	No	No	No	No	1 month	No	No
Model 12 – 1st iteration	Market	Day	LightGBM	16	Basic	Basic	Basic	No	Basic	No	No	1 month	No	No
Model 244 2nd iteration	Market	Campaign	LightGBM	~50	Basic	Advanced	Basic	Advanced	Advanced	No	Yes	1 month	No	No

Unless specified otherwise, this model is used in the rest of the presentation

Overview of results on backtesting period : +9 FA points compared to last model submission

Performance at aggregated level, at Day 3 of campaign
Backtesting Feb. 2021 to July 2022, Model 244, Excl. unplanned offers

Quick figures

5607 Nb of Nodes	484 # of days
60,17K Nb of Offer SKU line	44 837 712 Total sales qty
43,3 % % Nodes eligible	444,3 Avg daily sales per Node
	10,7 Nb of OSL per SKU

Performance

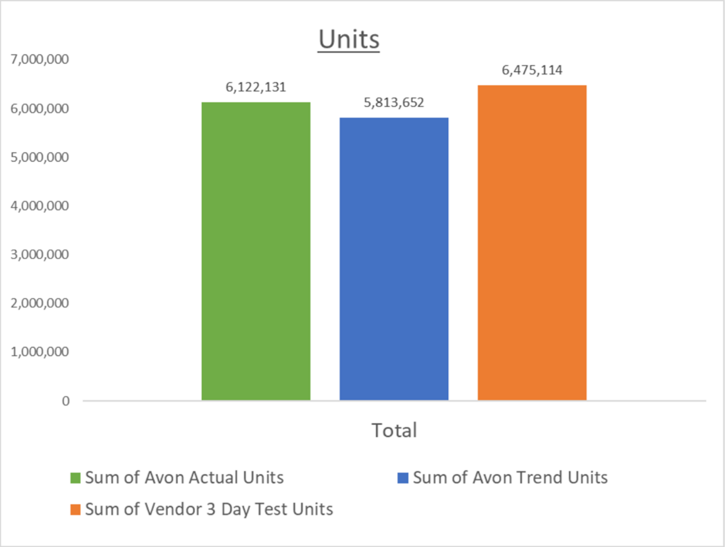
	by SKU x Campaign			By Offer SKU line x Day
	Accuracy	Bias	Outliers %	Accuracy
Gain *Model vs. R-factor*	-0,7 %	0,6 %	0,7 %	-2,2 %
Model	83,5 %	-0,6 %	1,5 %	51,8 %
~R-factor	84,3 %	0,0 %	0,8 %	54,0 %
Extrap'	52,7 %	38,8 %	0,7 %	28,0 %

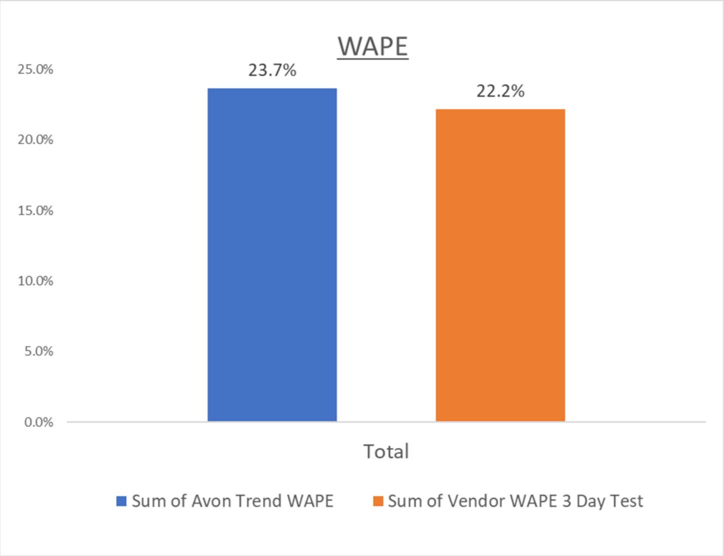
- We find our benchmark performance of ~84% accuracy for the R-factor proxy method
- The model performance is close to the R-factor proxy method. As a reminder, this method uses the *actual sales* of the campaign instead of the *cash forecast* (not available)

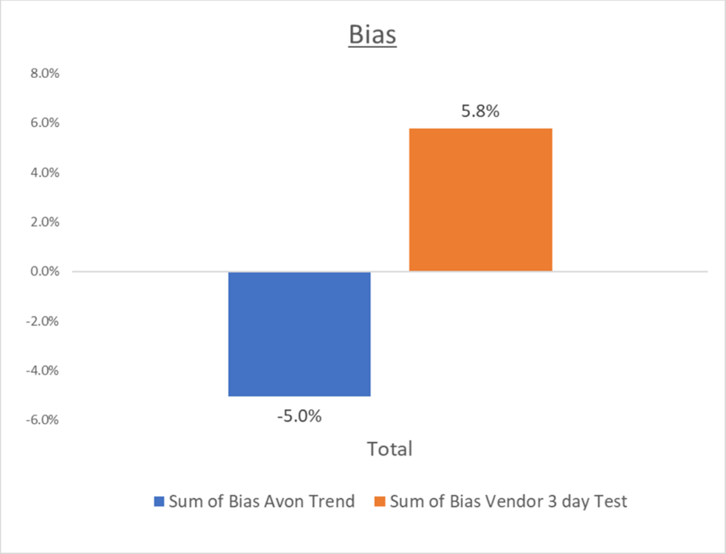
Test performance on C8, C9 and C10 of 2022

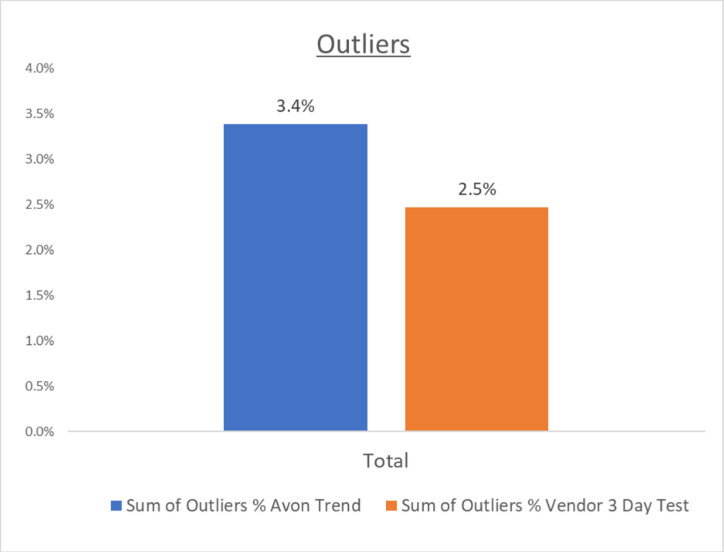
- **+ 1,5pts Accuracy gain** compared to trend forecast by the AVON team, leveraging the cash forecasts
- **Positive ~5% bias** instead of negative, which will reduce lost sales
- Significant improvement in **outliers %**
- Better performance on C8, slight improvement on C10, on par for C9 (except for bias which is improved from -11,1 to +3,2%)

Aggregated level (SKU, Campaign), at Day 3



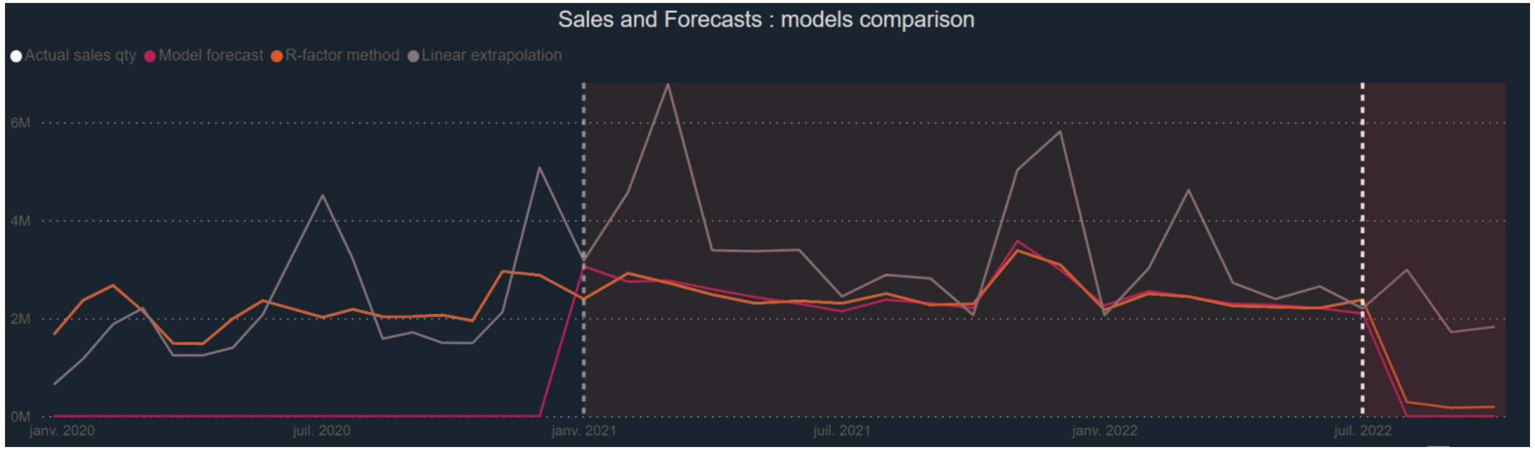


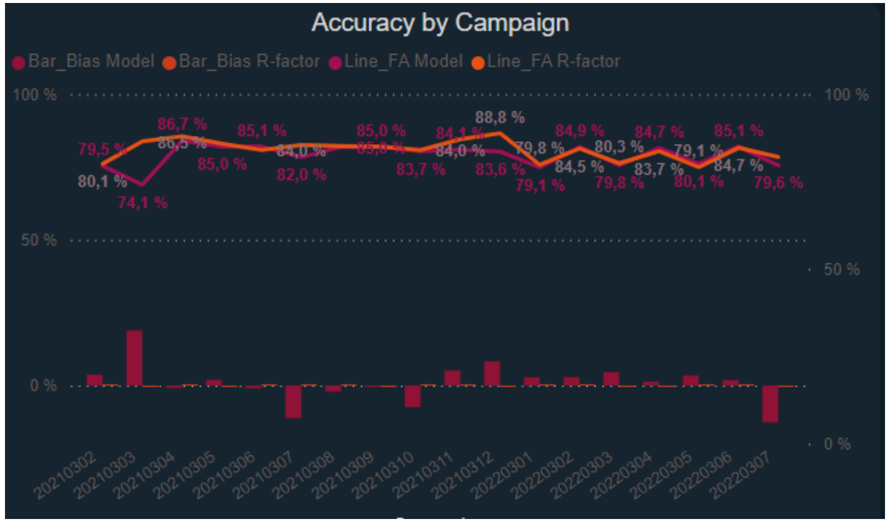




Overview of performance by campaign

Performance at aggregated level, at Day 3 of campaign
Backtesting Feb. 2021 to July 2022

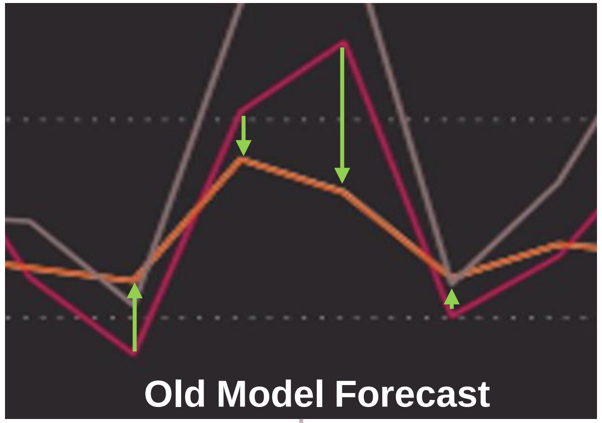


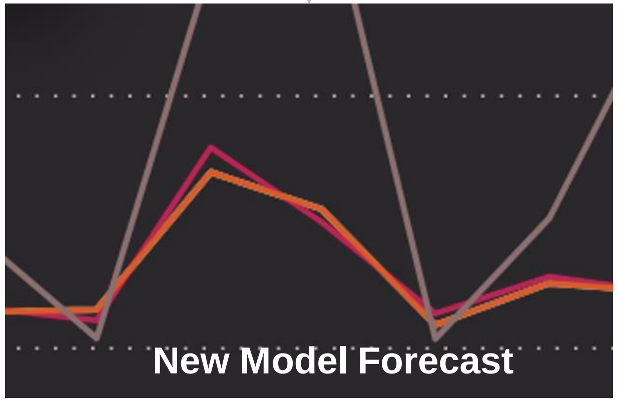


- The bias is relatively low on every campaign, which can be seen visually on the timeseries lines
- There is still an under-forecasting biais for July and October campaigns, although greatly reduced (from ~35% to ~12%)
- Performance is improving over time, as the model benefits from more and more historical data (see from July 2022)

Introducing new features : R-Factor inference

- **Intuition & approach** :
  - Correct overall bias caused by unfortunate low sales quantities in first 3 days
  - To do that, we need a “base level” to ground the forecast. We saw that the R-factor proxy is quite good at providing this base level
  - We need to forecast our own R-factor to feed the model. Ultimately, it could be replaced by the cash forecast, which could be more accurate



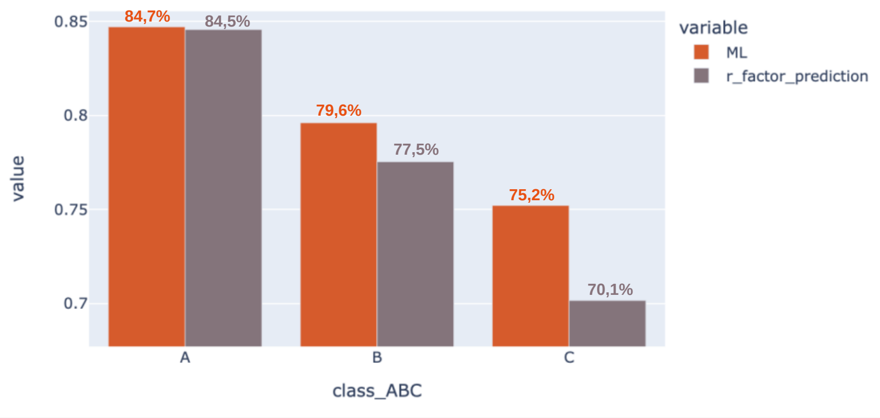


- **Calculation Method** :
  - Step 1 : Make a simple aggregated campaign sales forecast with 2 months moving average
  - Step 2 : Correct moving average forecast to take into account end of year seasonality : Replace November, December and January by Y-1 sales
  - Step 3 : Calculate R-Factor on first 3 days and project it on previously aggregated forecast

R-factor alone	ML with R-factor feature
82,9% FA	83,5% FA

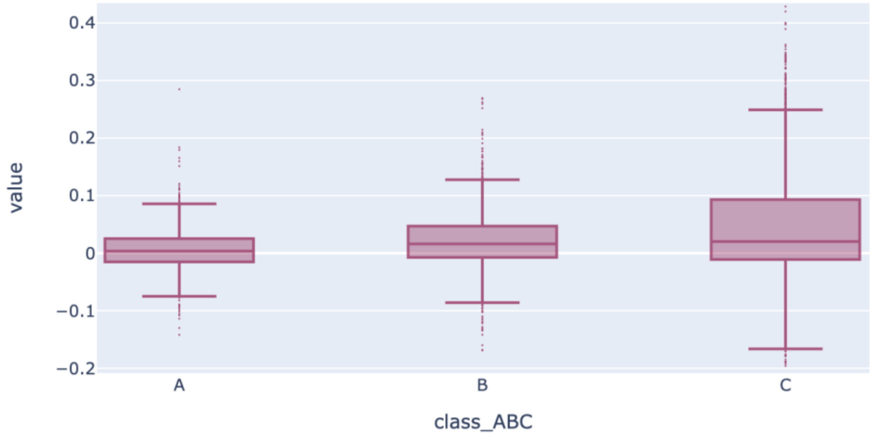
R-Factor inference Vs ML Performance

Forecast Accuracy by ABC class (ML Vs R-Factor)



Accuracy Distribution per SKU (ML Vs R-Factor)

Delta FA distribution (FA_{ML} – $FA_{R\text{-}Factor}$)



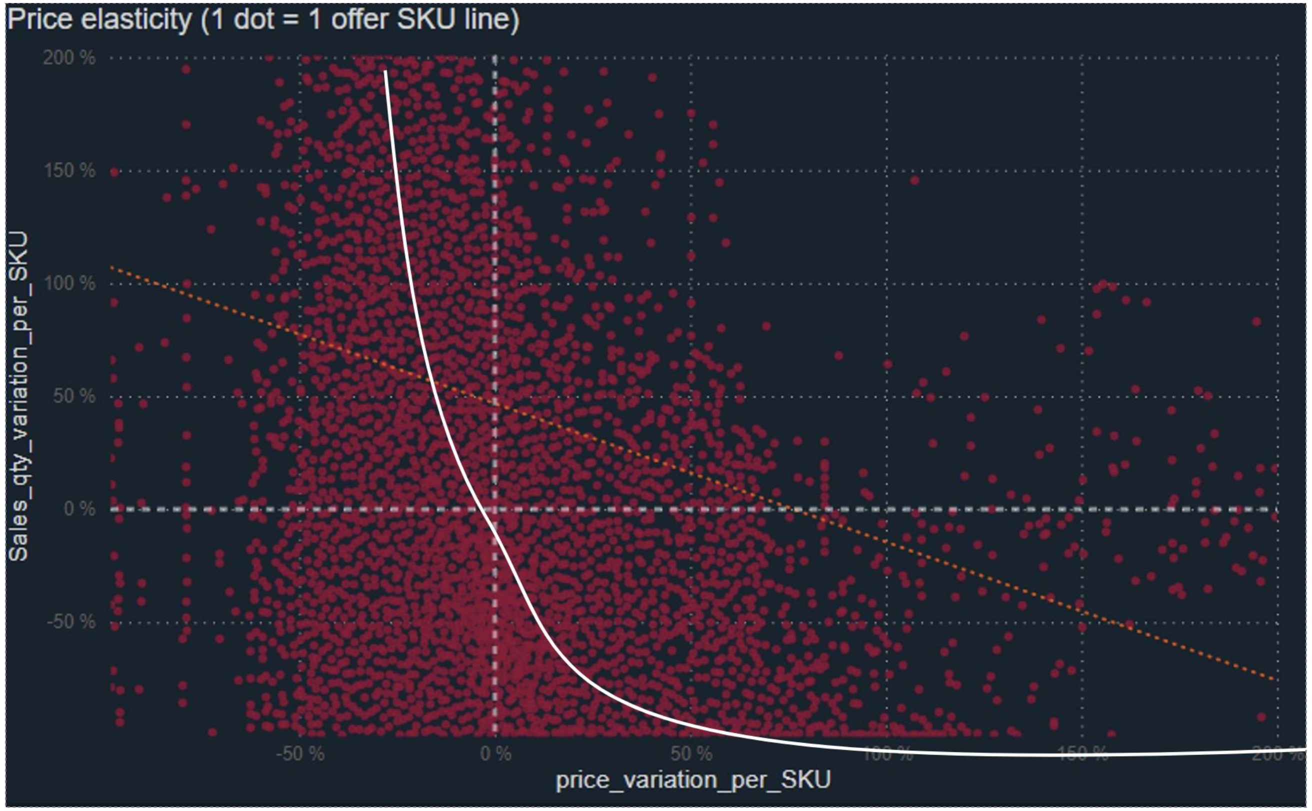
- **Conclusion** :
  - ML Model improves the accuracy **by ~2% on class B** and ~**5% on class C**
  - **~65% of class B and C** SKUs have an increased forecast accuracy, representing ~**10%** of sales quantities
  - **~55% of class A SKUs** are improved; while the degradation is < -7% in 90% of cases

IRIS by Argon&Co

Introducing new features : Sales/Price Elasticity

- **Intuition** : Give the model insights about the impact of a price variation on sales qty

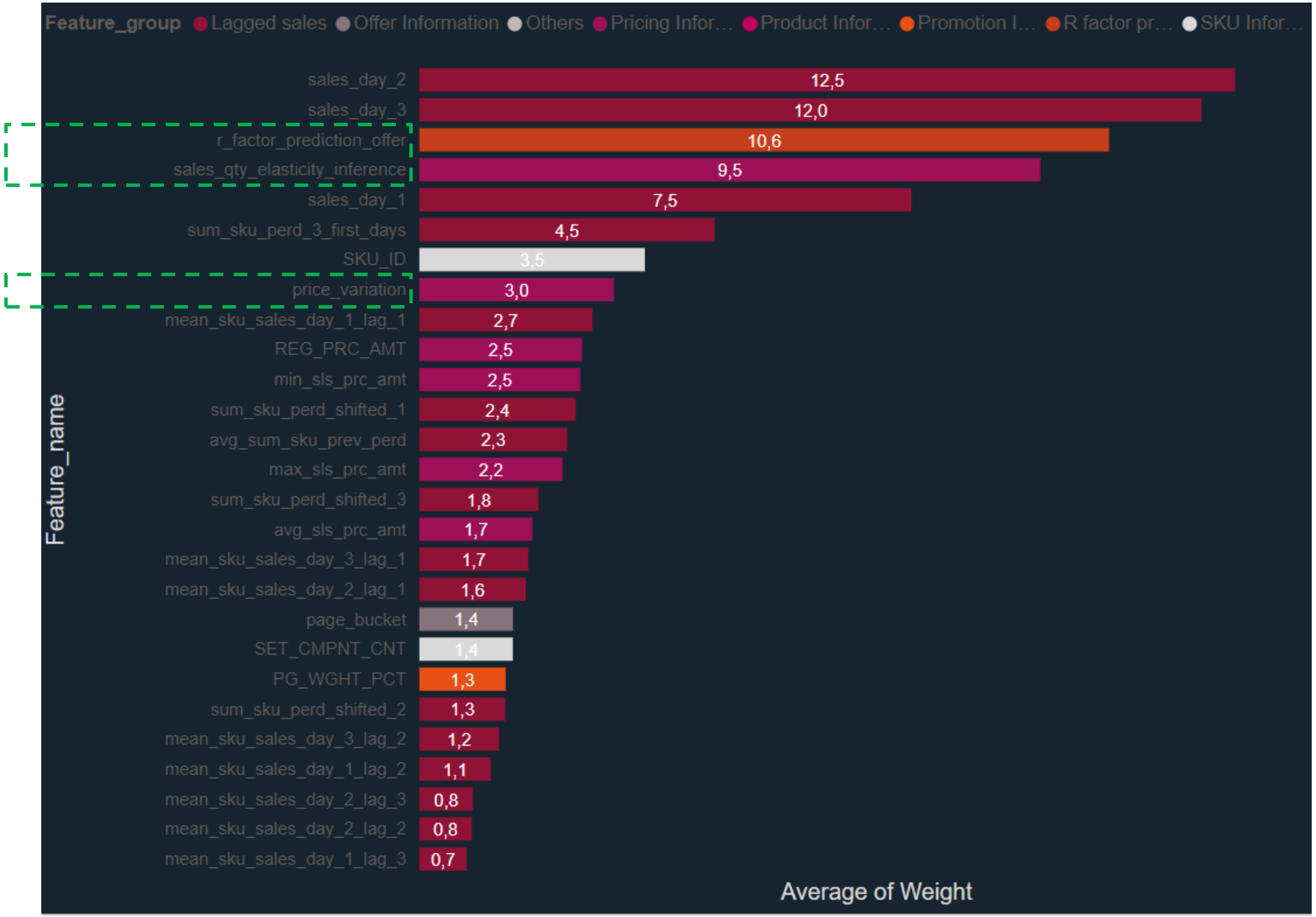
- **Calculation Method** :
  - Step 1 : Compute Sku price variation compared to a sliding mean price
  - Step 2 : Compute Sku sales variation compared to a sliding mean sales qty
  - Step 3 : For each CATGRY_ID, apply a logarithmic regression where X is price variation and Y is sales qty variation
  - Step 4 : Create a sales qty prediction based on the price variation



IRIS by Argon&Co

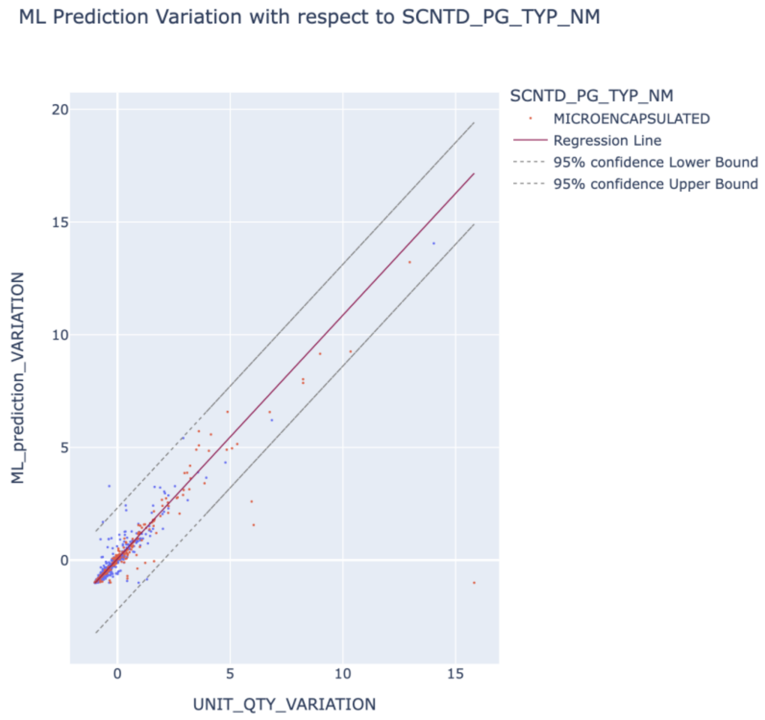
Features Importances

- **Early Signals** : The first 3 days of sales are the most important features in the prediction
- **R-Factor Prediction** : Including R-Factor prediction into features helps avoiding strong negative bias caused by low sales in the initial days
- **Sales Elasticity** : This features translates the impact of the price variation on sales based on historical impacts.
- **Price Variation** : Gives insights about the price being superior or inferior to the mean price in the same campaign

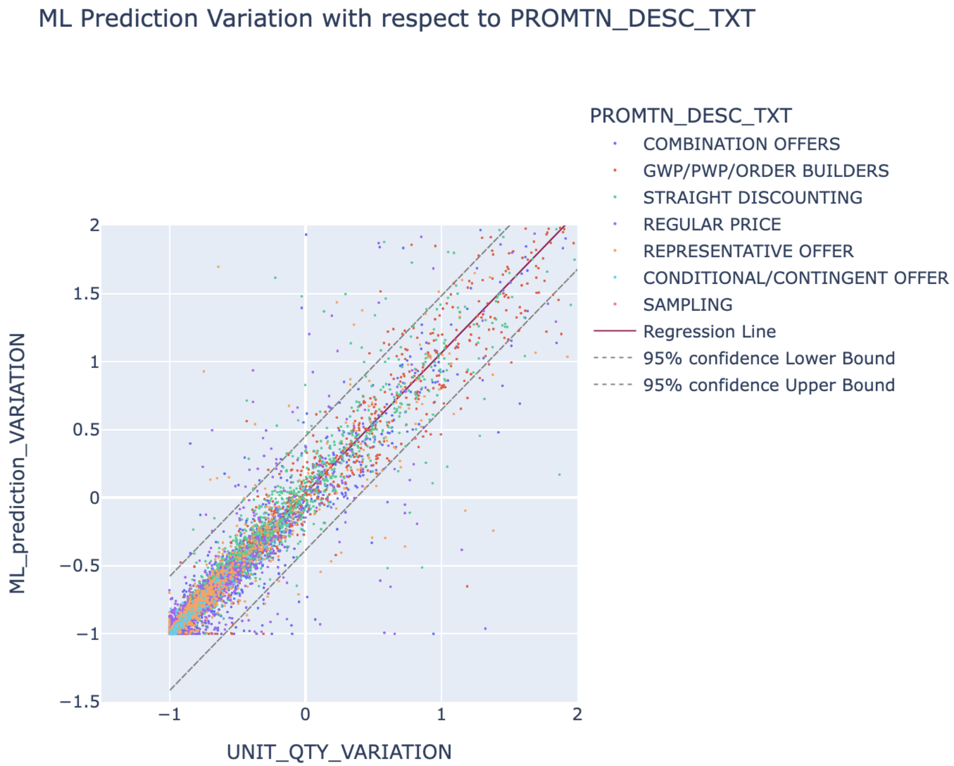


Examples of Features Impact on ML Forecast

Impact of Microencapsulation



Impact of Promotion type



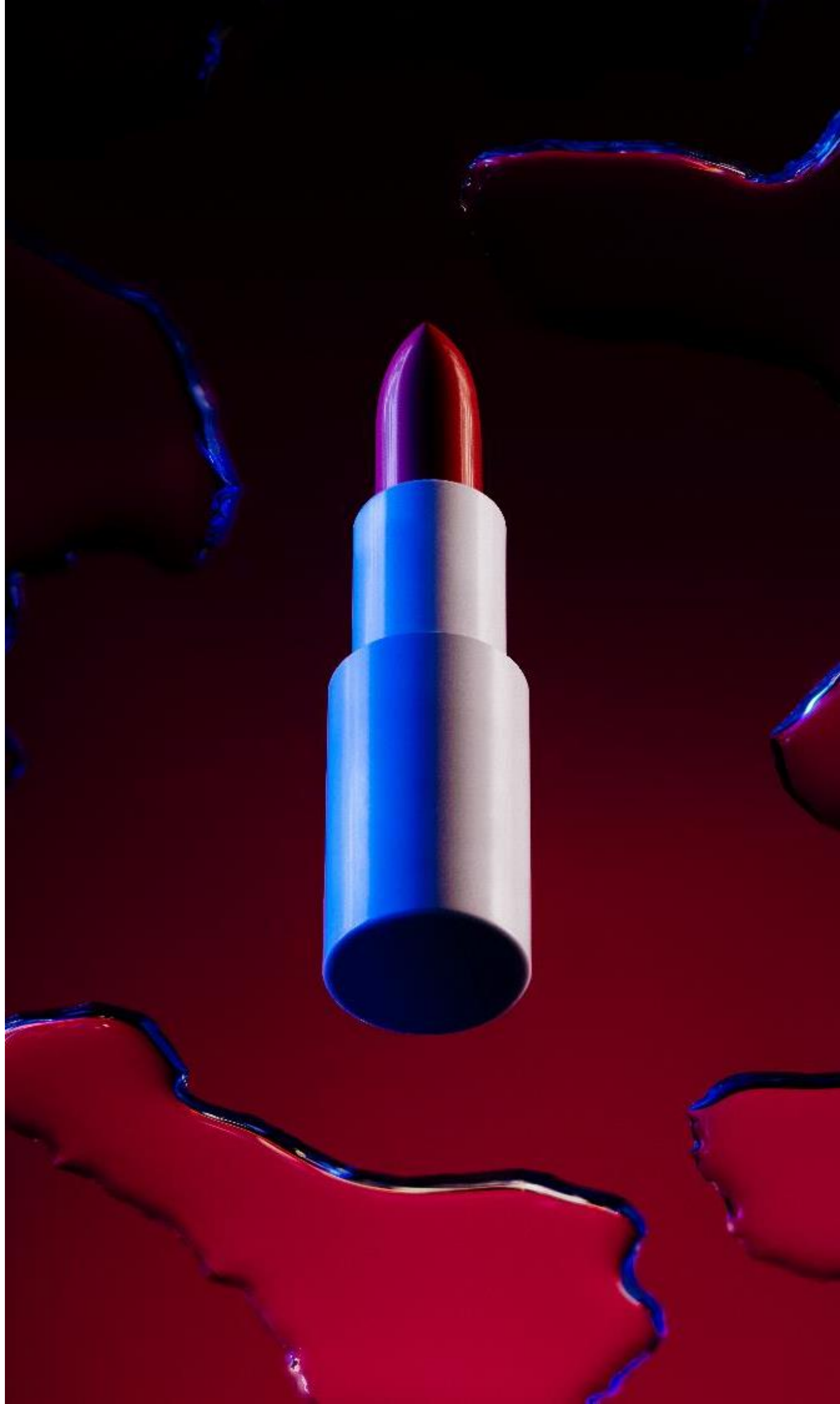
- We define the *variation* as the percentage of higher or lower sales quantities compared to mean sales quantity
- The variations of sales quantity with respect to Promotions is similar to the variation applied by the ML forecast
- ML model will forecast more or less depending on the value of promotion or micro-encapsulation

IRIS by Argon&Co

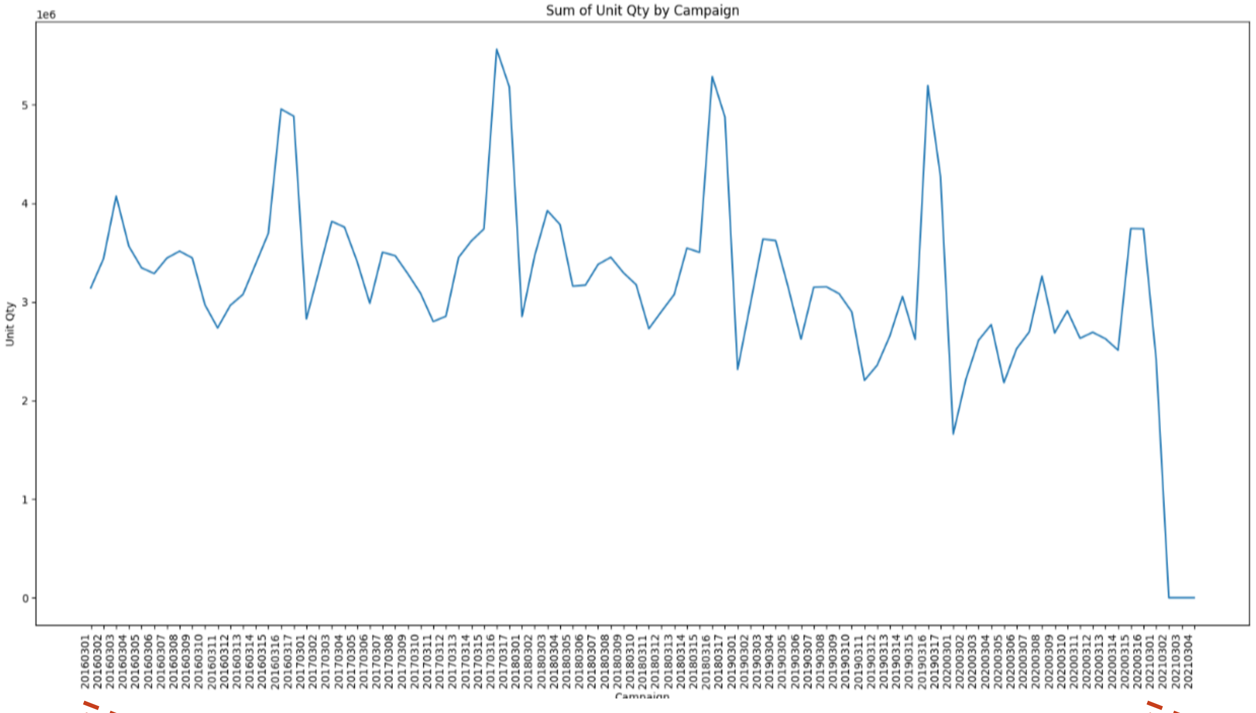
Other ideas

Pre-process the first 3 days of sales to “clean them” when they are biased8	No obvious influencer of the first 3 days patterns vs rest of campaign	inconclusive
Post-process the forecast to bound the ML forecast inside a moving average + standard deviation	Helps correcting the bias in extreme situations +1% FA and -1ps bias overall	successful
Forecast daily weights to split the campaign-level forecast to the day level	Would need to compare relevance against the weights currently used in Supply Chain SI	Not tested
Leverage R-factor (cash forecast) or its components (# of reps, # of orders) as model features	Would need to have the forecasted values for the testing period of C8,9,10 2022	Not tested

Contents



Provided with 5 years of historical data (2016 to Jan 2021), we're expected to forecast C5->C11 of 2021



2016 C01

2021 C01

… And evaluated at the ***aggregated*** node level

Forecast granularity (node)

- Market (Poland)
- SKU
- Campaign

Forecasts are generated at the ***elementary*** node level…

Forecast granularity (node)

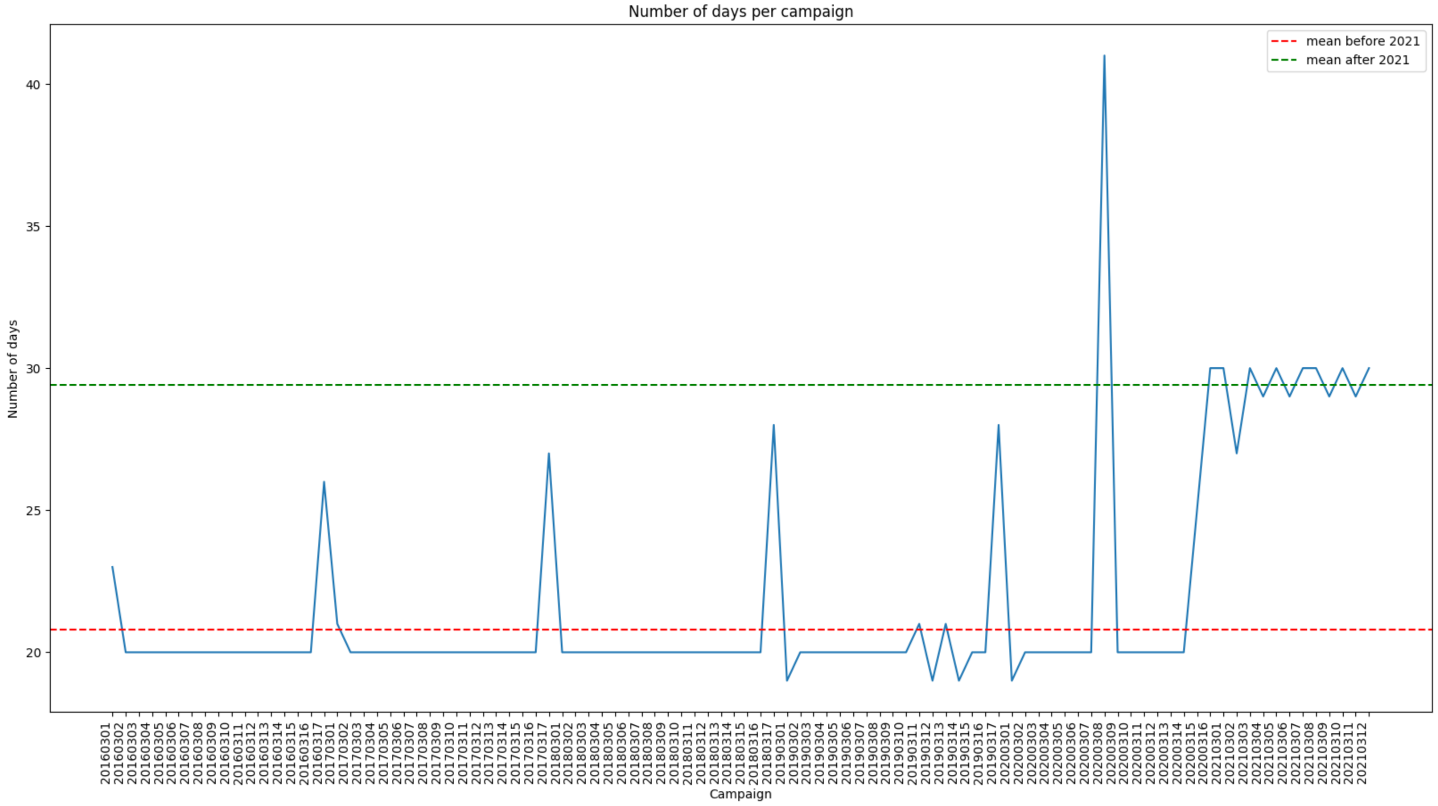
- Market (Poland)
- Offer SKU Line ID
- Campaign

- We can hold back a **backtesting period of 1 year (2020)** to test our performance before inferring the forecasts on C5-11 of 2021
- Our "**lags** of interest" will be 4 – 6 – 10 months for backtesting

Demand forecast - First ideas

How to manage variable campaign sizes

- The dataset offers much more historical data than for the *Trending* forecast
- The campaign duration is variable over time (not based on full months before 2021)
- We need to have the model forecast a full campaign quantity while considering its duration



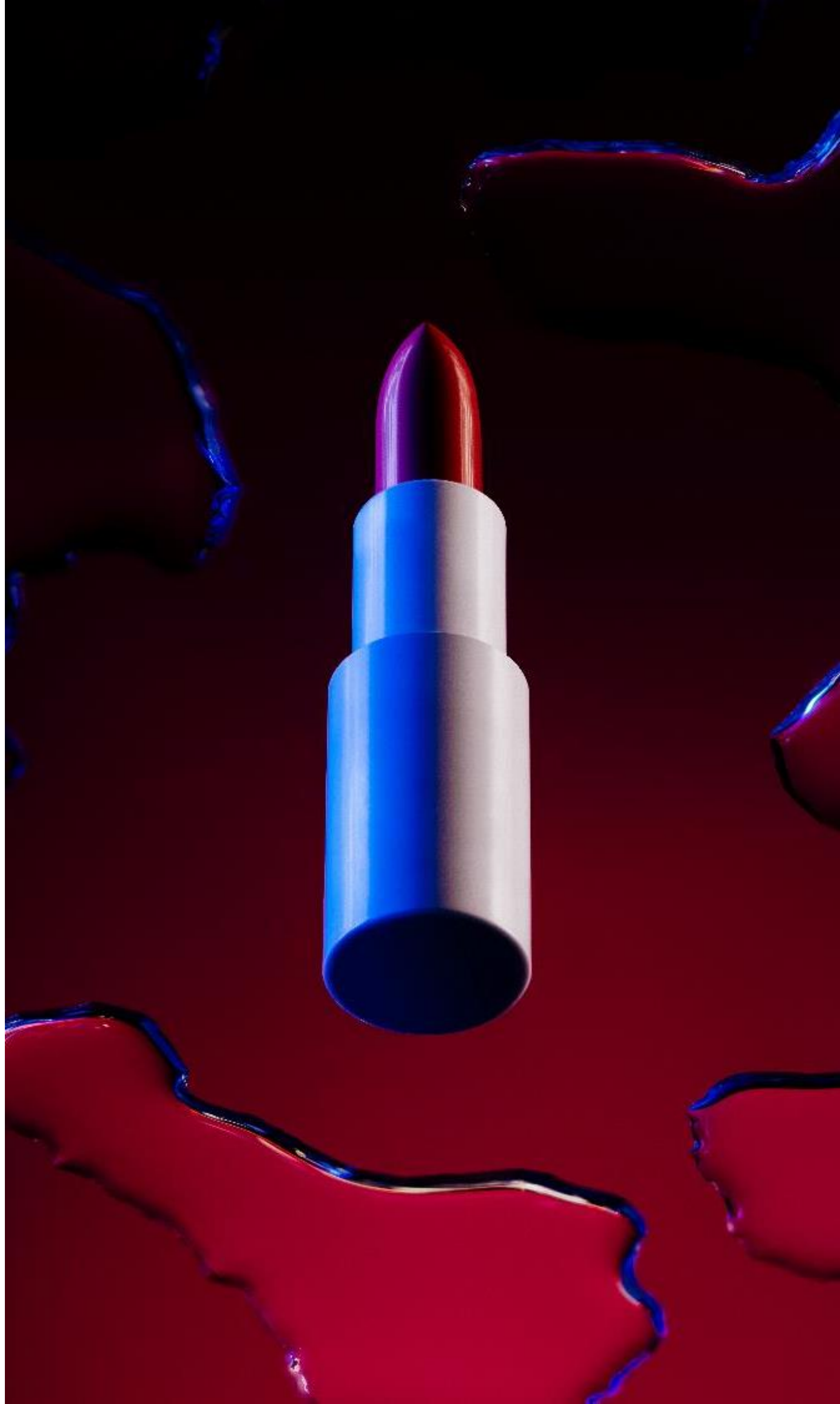
Method 1

- Normalize the sales for a full month by # of days prorate
- Attach a campaign to a “main month” to get the equivalent number of days

Method 2

- Change the model target to an average daily units per day of the campaign
- Multiply by the actual number of days in post-processing

Contents



Our next iteration for *Trend forecast* could optimize the set of features

	Model split	Level	Algorithm	Features								Other		
				# of features	Actuals @Day 3	Previous campaigns	SKU features	Price	Offer features	Calendar features	Global Ratio	Retrain frequency	Hyper-parameters tuning	Feature selection
R-factor proxy	Market	Campaign	R-factor	2	Basic	No	No	No	No	No	No	Na	Na	Na
Model 10 - baseline	Market	Day	LightGBM	1	Basic	No	No	No	No	No	No	1 month	No	No
Model 12 – 1st iteration	Market	Day	LightGBM	16	Basic	Basic	Basic	No	Basic	No	No	1 month	No	No
Model 244 2nd iteration	Market	Campaign	LightGBM	~50	Basic	Advanced	Basic	Advanced	Advanced	No	Yes	1 month	No	No
Next iteration	Market	Campaign	LightGBM	~50	Basic	Advanced	Basic	Advanced	Advanced	Yes	Yes	1 month	Yes	Yes
For the next iteration	Brochure / rest of							Normalize price for inflation	Add Number of representatives and orders	Optimize Split by day using calendar features	Feed R-Factor with Cash Fcst Prediction		Typically ~1% gain after Feature Selection and hyper parameters tuning	

IRIS by Argon&Co

Next steps

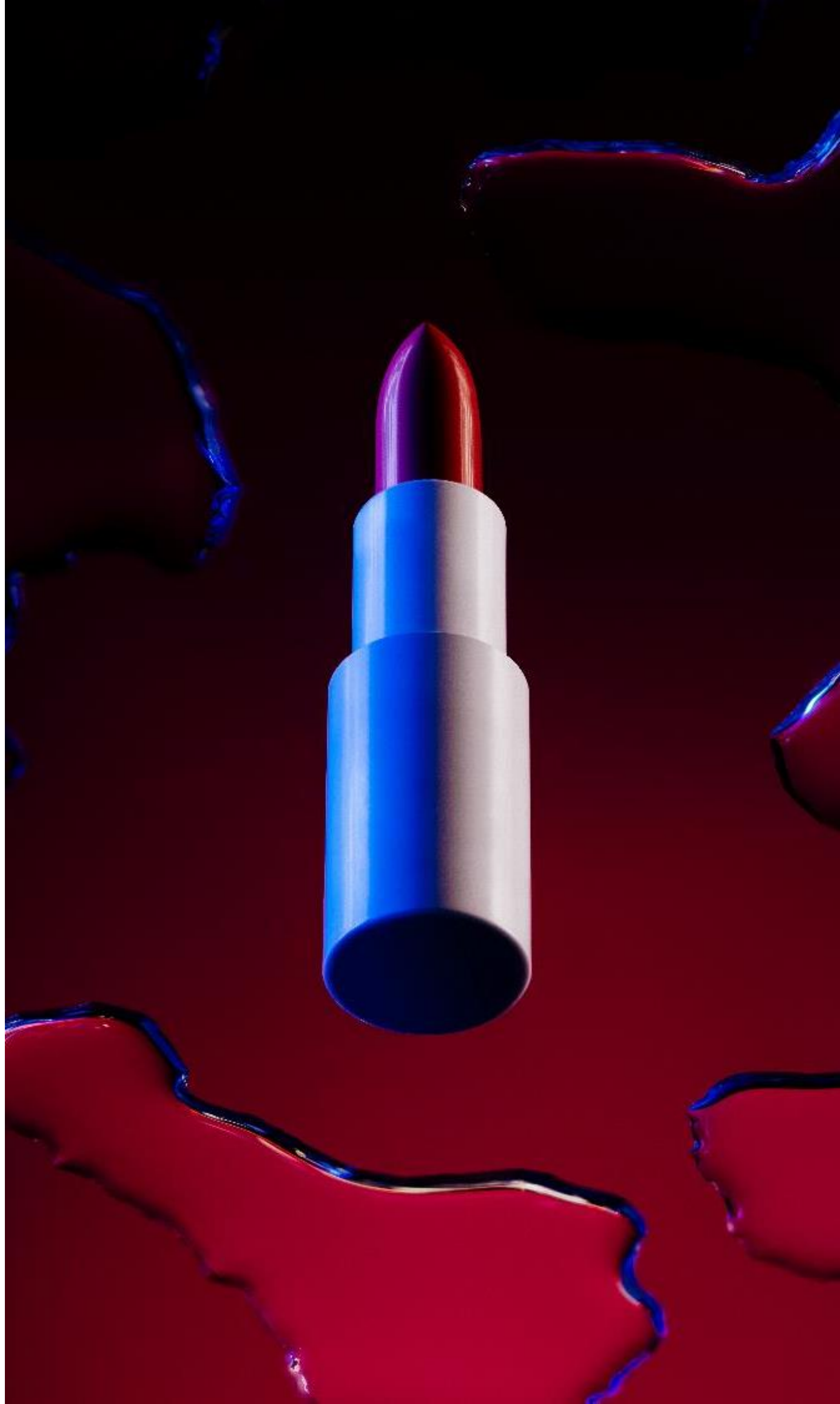
Trending PoC

- We have some room for further optimization of the model (see previous page)
- Incremental improvement in accuracy should be of a few %
- We suggest to spend time on the *Demand Forecast* first, before possibly coming back to the Trend forecast
- *Open discussion* : should we ultimately test and benchmark for 7 and 14 days forecasts?

Demand forecast PoC

- Build end-to-end pipeline to generate and back-test forecasts
- Re-use all relevant features from *Trending* PoC
- *First submission* : next week

Contents



Demand Forecasting POC Approach

Data Avon will share

- 5.5 Years of historical data 2016 - January 2021 with full campaign Demand Actual Units, plus forecasting lag* period February - April 2021 without Units, all by **period** / **item / price point = Offer SKU Line level**. Full list of prod attributes, offer details, price, etc. File name: *Poland Core Brochure History File Jan 2016 to April 2021 (units up to Jan 2021) - Final Hardcoded - including OE forecasts.xlsx*
- Test data at Offer SKU Line Level with full data except units also provided for test period May to November 2021. File; *Poland Core Brochure Test Data File May - November 2021 - no units - Final Hardcoded.xlsx*
- Data to include Core Brochure vehicle only, all products, offers in campaign.
- Market: Poland

Appro ach

Approach for POC / for test campaigns (forecast on Offer SKU Line level):

- Vendor to provide forecasts for May, June, July, August, September, October & November 2021 (Campaigns 5-11).
- Forecast units to be provided at Offer SKU Line Level in test data file (May – November 2021) file. Every row in the test data file should be given a unit forecast (in the column ‚BH”). File (as above); *Poland Core Brochure Test Data File May - November 2021 - no units - Final Hardcoded.xlsx*

*Lag = 3 months between forecasts being created and actuals becoming available.

Trending POC approach

Data Avon will share

- 3 Years of historical data (2020, 2021, 2022 up to C7), plus selected days from test periods (C8,9,10 2022).
- Units, sales **by day** by **item / price point = SKU offer level**. Data details include; full list of product attributes, offer details, price, etc and daily Demand Actuals data. Files names: *Romania Trending POC Data - C1-16 2020 - Final.xlsx, Romania Trending POC Data - C1-12 2021 - Final.xlsx, Romania Trending POC Data - C1-7 2022 - Final.xlsx*
- Data to include all vehicles, all products and offers in the campaign.
- Market: Romania
- There are some offers (called „Agile offers”) with data available from C7 2021 onwards where offer duration is different than 1 month / campaign (you can find offers dates in the file).

Approach for POC / for test campaign: (forecast on Offer SKU Line Level, key accuracy measurement at SKU / Period level)

Approach

- Vendors to provide trend forecast for Aug, Sept, Oct’22 by Offer SKU Line and day in phases:
  1) Avon provides 3 first days of data, vendors to trend August, September & October ’22 (1st iteration). Files; Romania Trending POC Data - C8 2022 Test Data - 3 Days of Actuals.xlsx, Romania Trending POC Data - C9 2022 Test Data - 3 Days of Actuals.xlsx, Romania Trending POC Data - C10 2022 Test Data - 3 Days of Actuals.xlsx. After receiving trend from vendors to check accuracy:
  2) Avon provides 1 week of Aug data and ask to trend August, September & October ’22 (2nd iteration). After receiving trend from vendors to check accuracy:
  3) Avon provides 2 weeks of Aug data and ask to trend August, September & October ’22 (3rd iteration).
- Output format – Vendors to enter missing Unit Quantities (column: „D”) for Offer SKU Lines in every row (orange cells) for remaining days of August, September & October ’22 in relevant test files.