Causal Impact Analysis

Branded Search test Wegenwacht

Martijn Ras & Niels Horsman

06-11-2025

Short introduction

Niels Horsman

Data Scientist @ ACE team | ANWB

Fun fact:

Finally joined the club — called Wegenwacht for the first time this year! 🚙 ⚡

Wegenwacht = Roadside assistance

Short introduction

Martijn Ras

Web Analyst @ Acquisition team | ANWB

Long time fan of the Wegenwacht:

- Yes: Like to travel
- But: Old car

Is the ANWB wasting money on google ads?

Average Dutch person: =

Research question:

Are paid branded search ads necessary, or does this traffic also convert through organic search results?

Wegenwacht = Roadside assistance

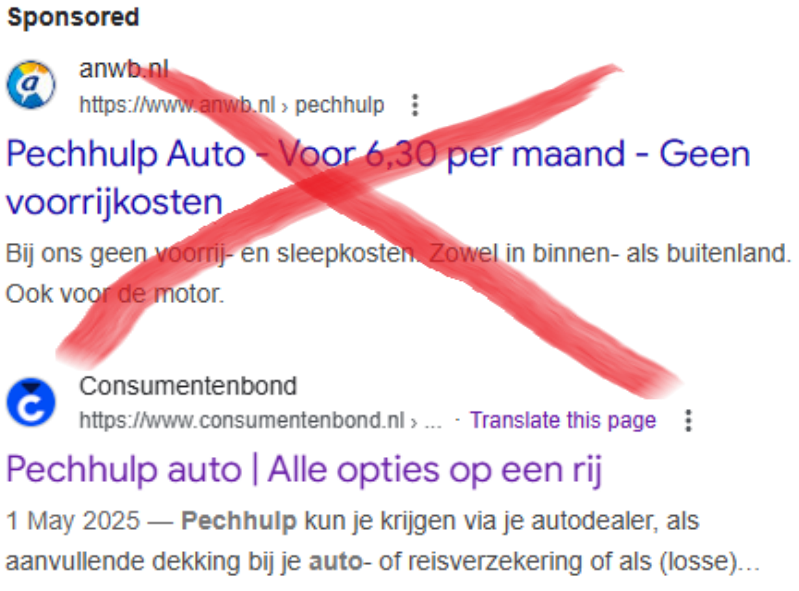
Standard procedure: A/B test

🚫 **Not Possible:** We cannot selectively disable ads for users randomly

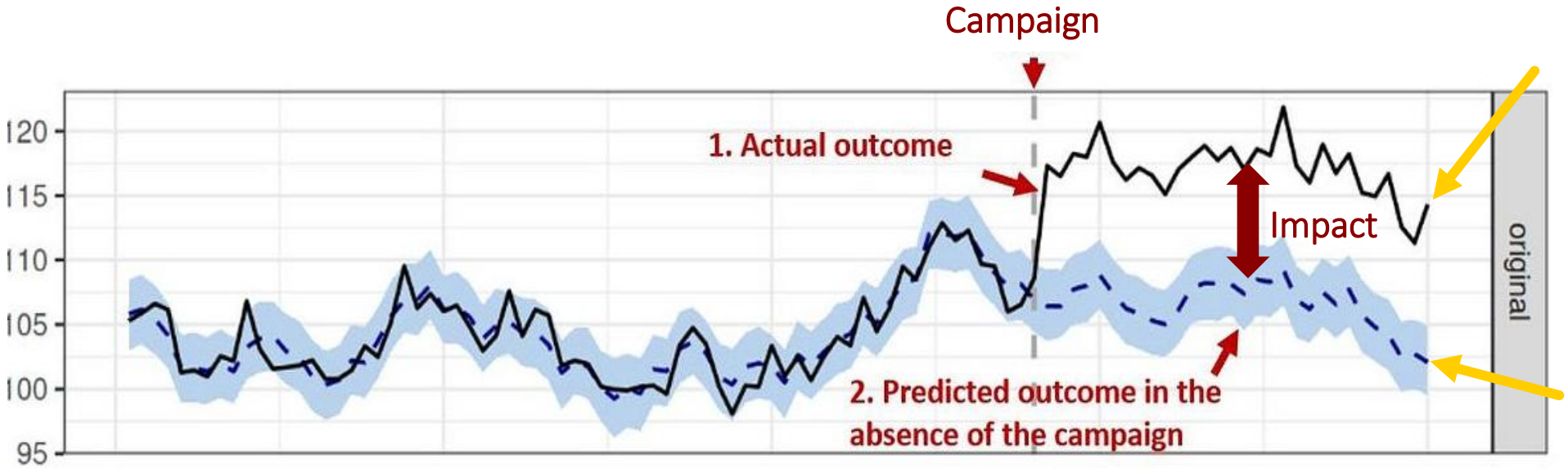
A: receives sponsored ads

compare

B: receives no sponsored ads



Solution: Causal Impact Analysis



B: receives no sponsored ads

A: receives sponsored ads

source

Method

Causal Impact Analysis requires 2 components:

1. **Forecast method:** Predicts what would happen without intervention
2. **Significance method:** Tests whether observed difference is statistically significant

Together: Forecast shows the baseline, significance test proves there is an actual effect

1. Forecast method: PROPHET

Prophet: Facebook's open-source forecasting tool. Makes prediction based on historic trends and external factors

Benefits:

- **Automatic seasonality detection**
  Models daily, weekly, and yearly patterns without manual configuration
- **Robust**
  Handles missing data and outliers effectively without the need for preprocessing
- **Easy integration of external factors**
  Holidays, events, and covariates can be directly incorporated into the model

2. Significance method

- **Calculate probability per observation**
  Compare each measured value to Prophet's forecast including uncertainty intervals

- **Significance threshold**
  p-value < 0.05 indicates a statistically significant deviation (impact)

Test design

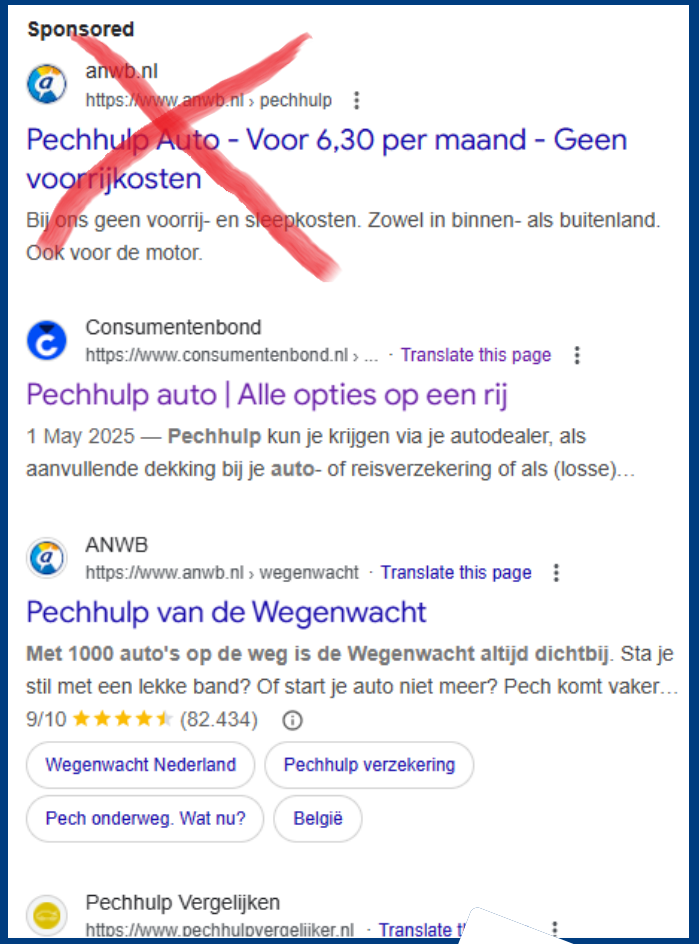
Duration:
3 March t/m 24 March 2025

Intervention:
All SEA-campaigns with search terms that contain ‘ANWB’ & ‘Wegenwacht’ or ‘ANWB’ & ‘pechhulp’ will be paused in Google Ads.

KPI: Wegenwacht sales

Excess budget branded search SEA:
Not allocated to other campaigns

Afterwards: Causal Impact Analysis



Predictors of sales Wegenwacht

Predictors

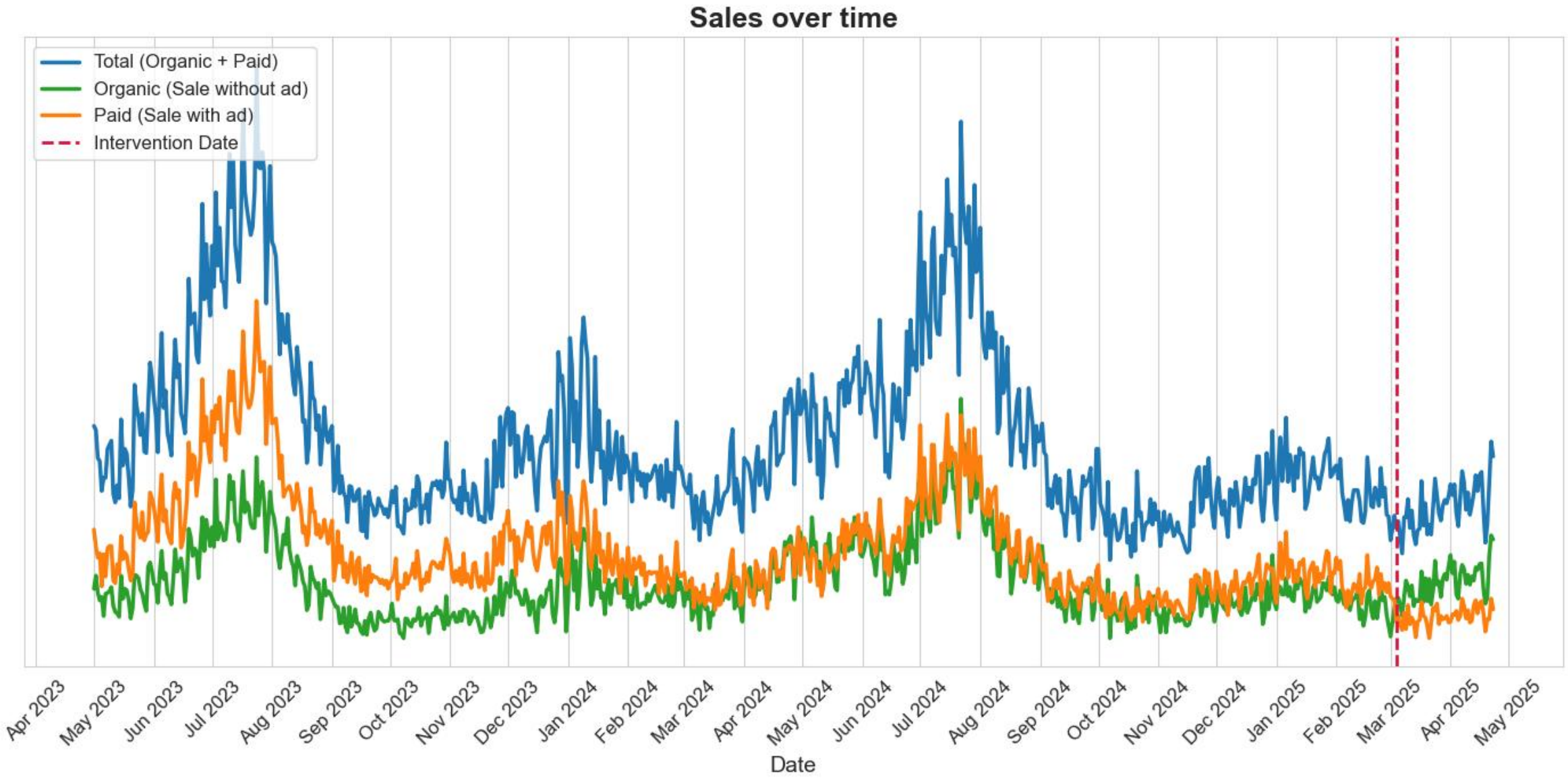
- KNMI: Temperature + Rainfall
- ANWB Onderweg app - ‘session_start’ event
- ANWB emergency center: Calls & Cases
- Google Trends: “ANWB” & “Wegenwacht”
- Marketing spend (SEA excluded)
- RDW: Car registrations
- ANWB.nl: #sessions
- ANWB.nl/wegenwacht: #sessions

KPI’s (*dependents*)

Sales Wegenwacht:

- *Sales WW - Total*
- *Sales WW - Paid Search*
- *Sales WW - Organic Search*

Raw data

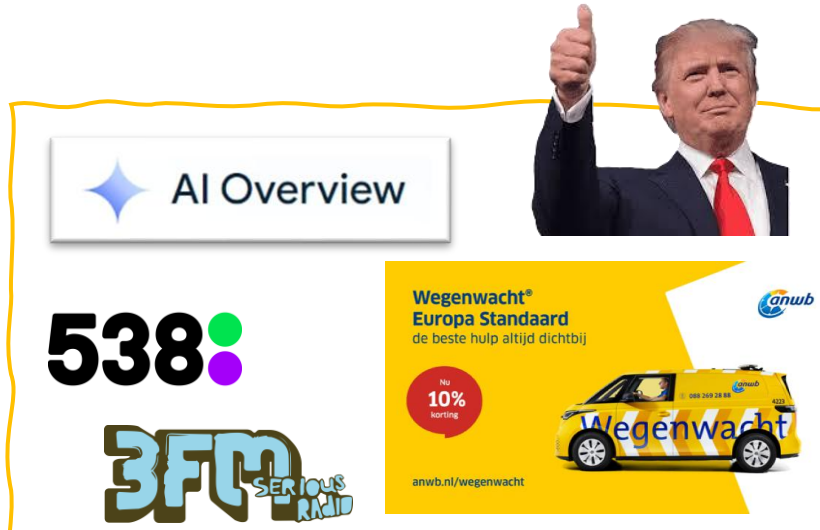

Sales over time
Total (Organic + Paid)
Organic (Sale without ad)
Paid (Sale with ad)
Intervention Date
Apr 2023
May 2023
Jun 2023
Jul 2023
Aug 2023
Sep 2023
Oct 2023
Nov 2023
Dec 2023
Jan 2024
Feb 2024
Mar 2024
Apr 2024
May 2024
Jun 2024
Jul 2024
Aug 2024
Sep 2024
Oct 2024
Nov 2024
Dec 2024
Jan 2025
Feb 2025
Mar 2025
Apr 2025
May 2025
Date

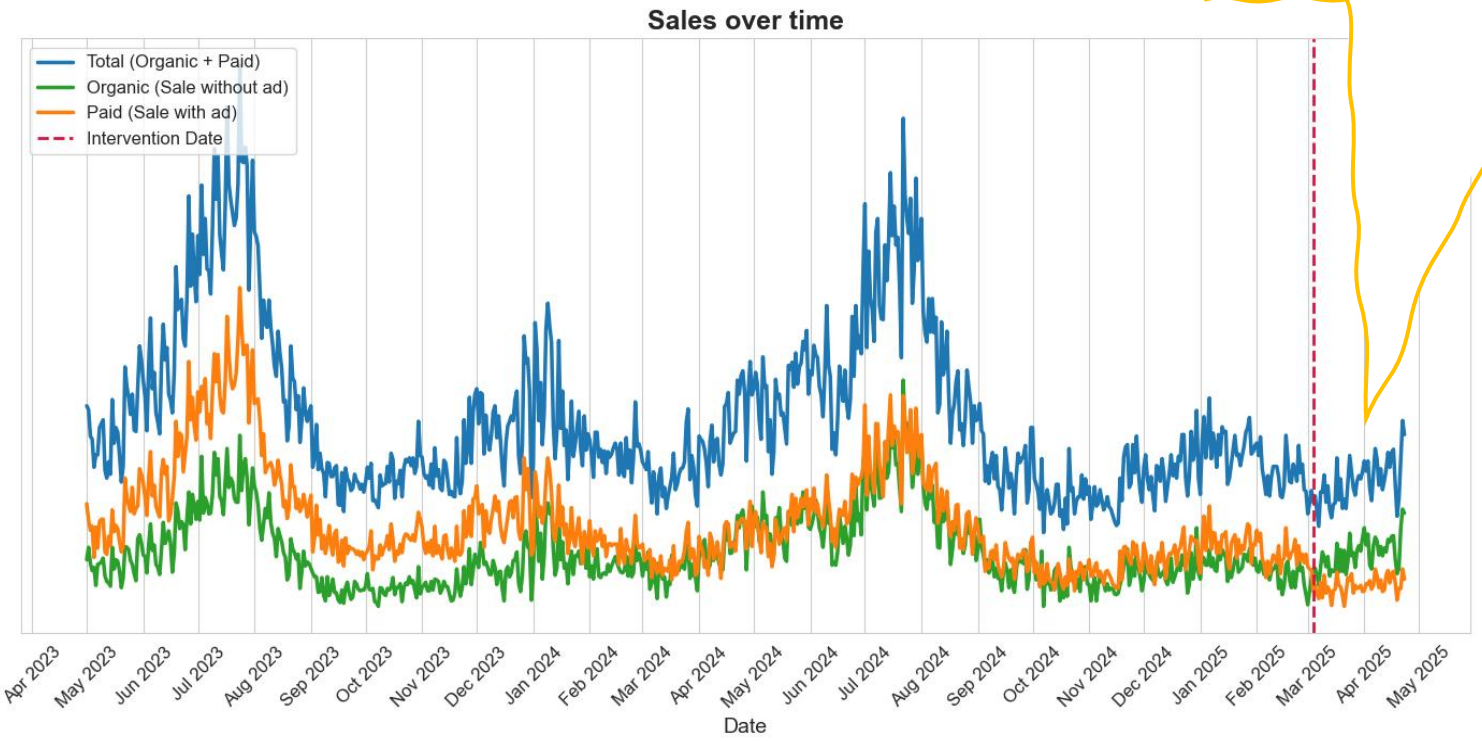
Raw data

Initial Assessment: No significant change observed on total sales
→ recommend pausing paid ads

Caveat: Results may be influenced by external factors

Next Step: Conduct rigorous analysis for objective validation



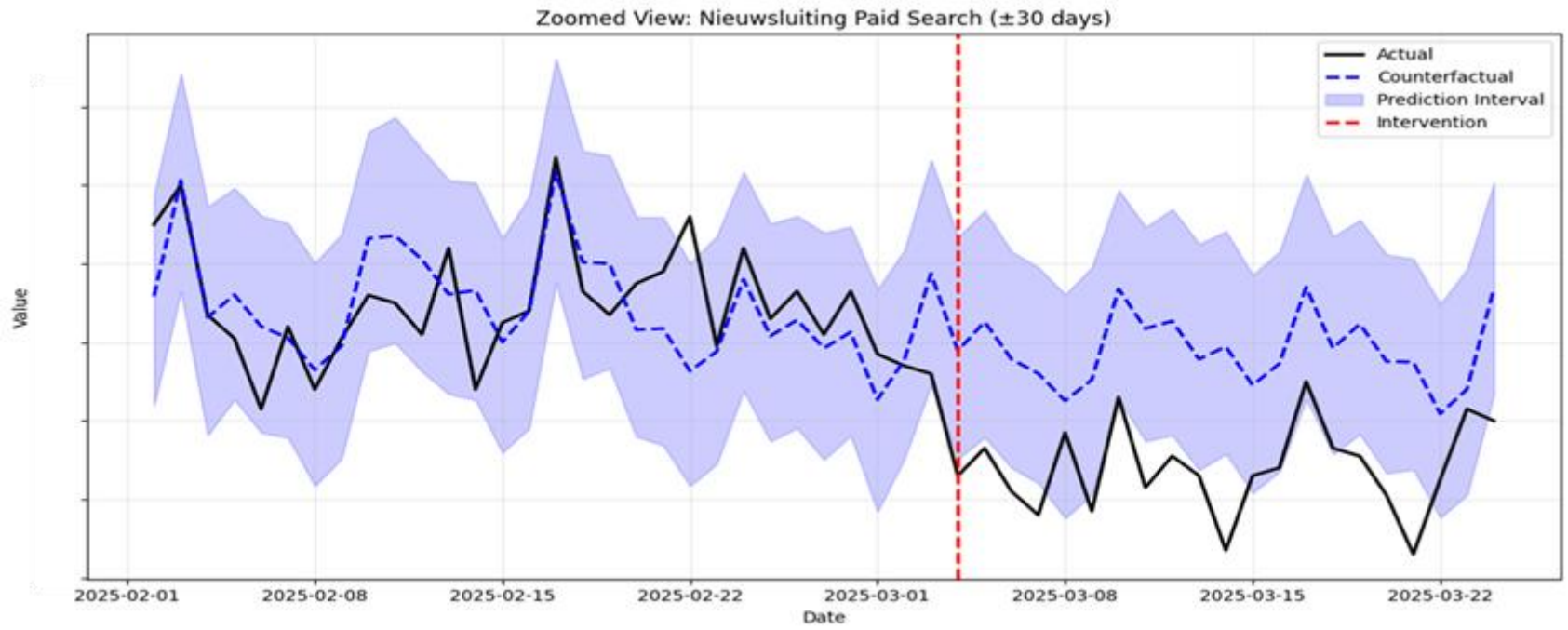


Results

Sales Wegenwacht Paid Search

Observed sales consistently outside 90% confidence interval of prediction

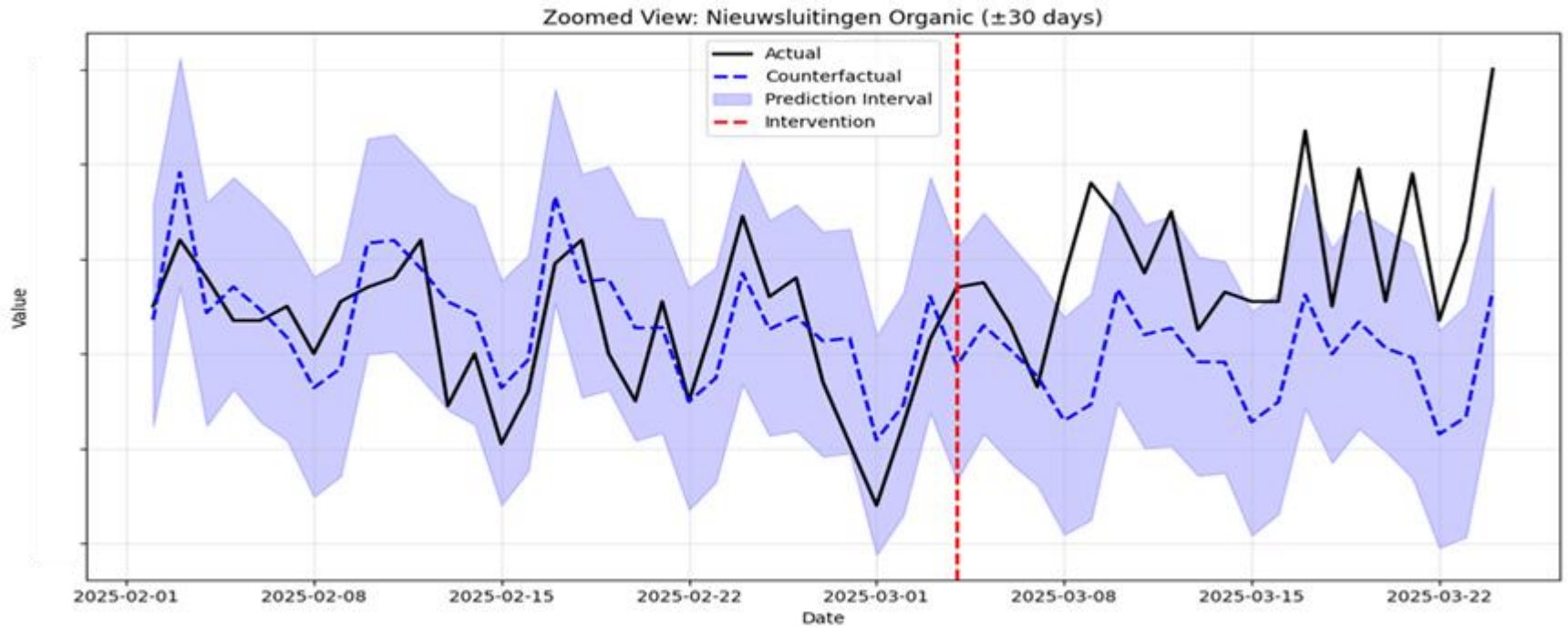
Significantly **lower** ✅



Sales Wegenwacht Organic Search

Observed sales often outside 90% confidence interval of prediction

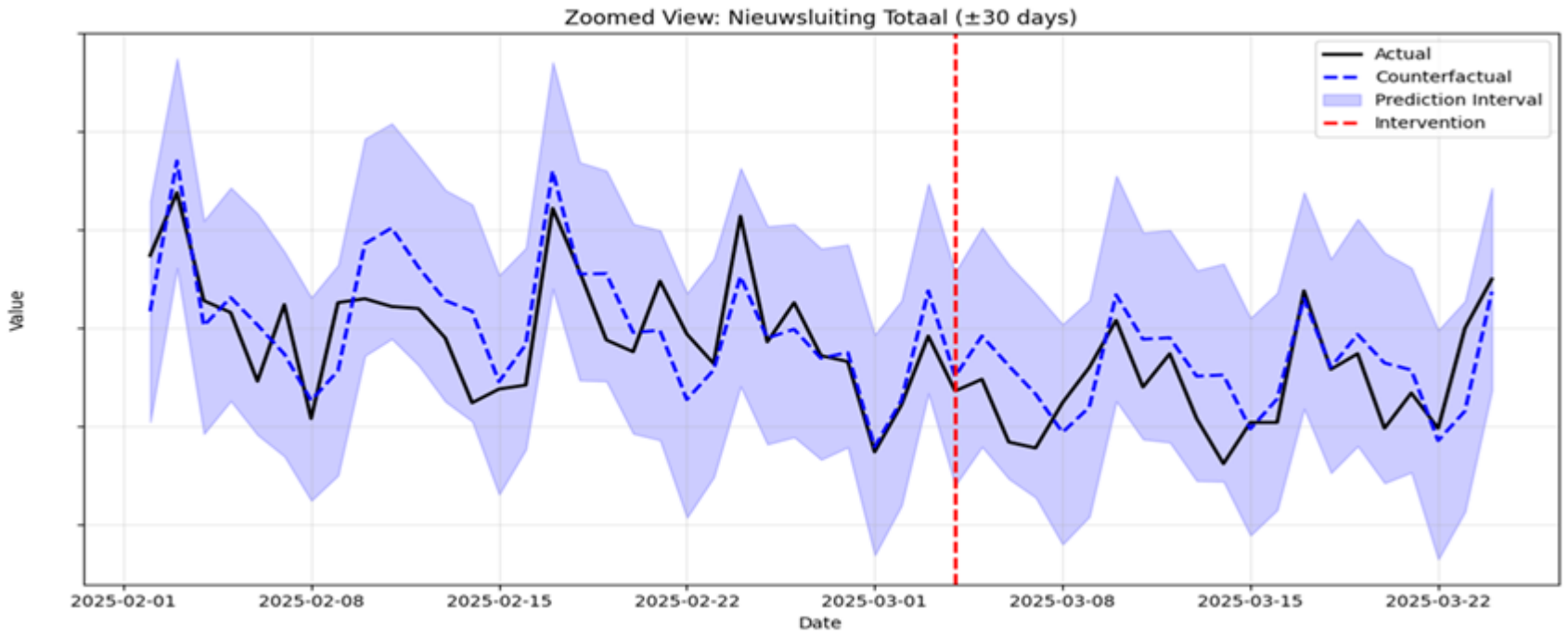
Significantly **higher** ✅



Sales Wegenwacht Total

Observed value follows the prediction.
Well within the confidence interval

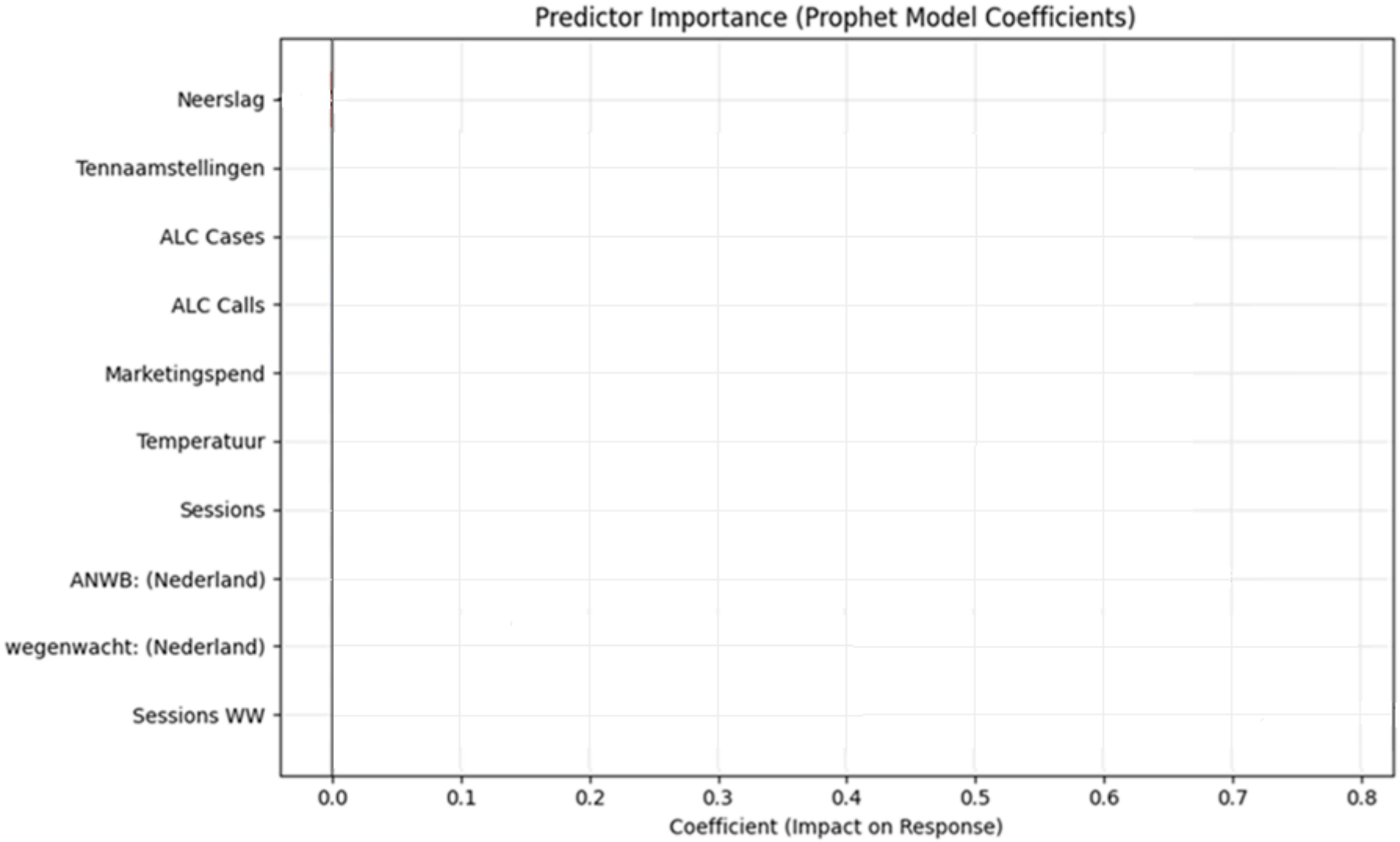
No significant change ✅



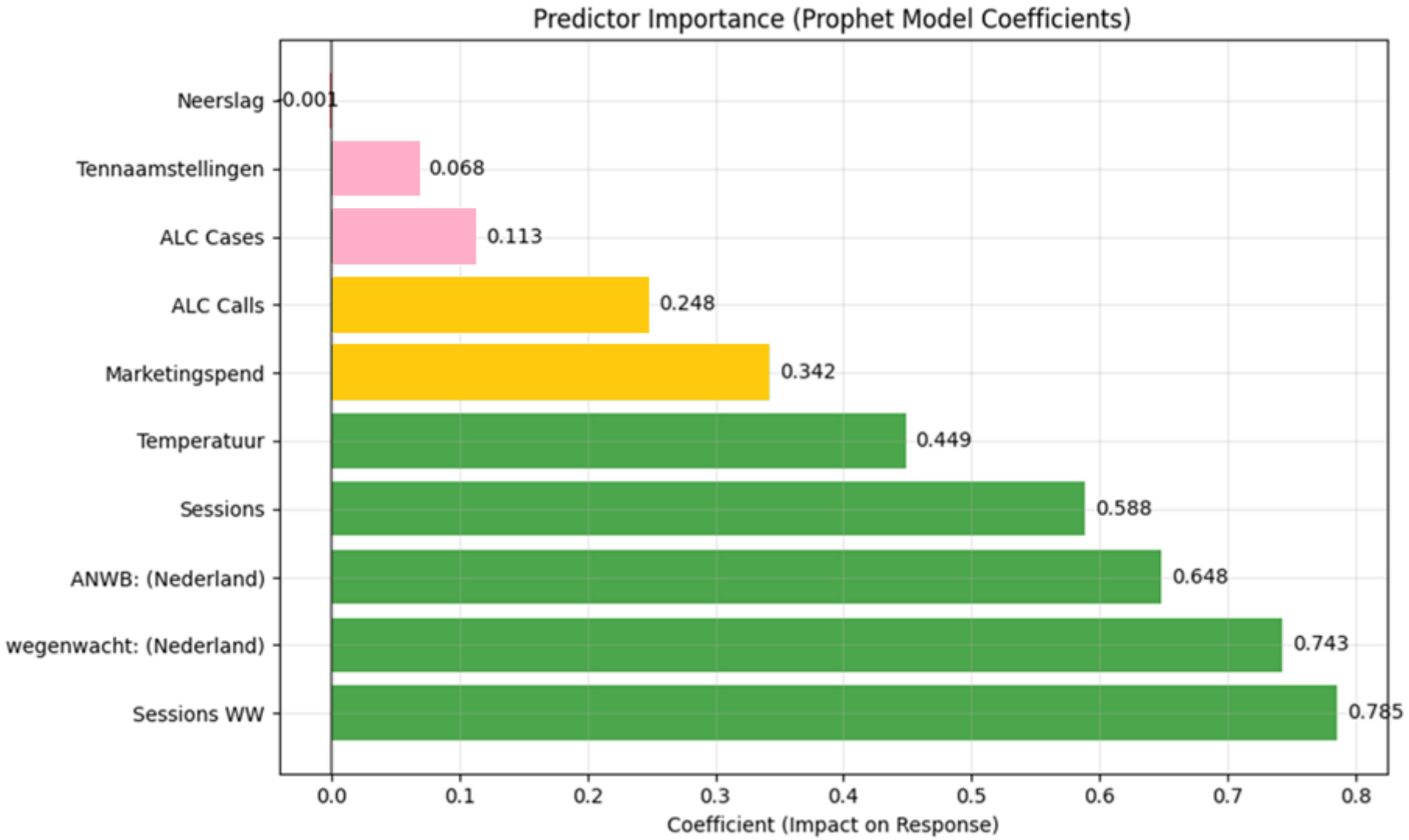
Conclusion:

- Turning off the Branded search terms had **no impact** on the total sales of Wegenwacht
- **40% decrease** of Branded Search costs

Importance of the external factors



Importance of the external factors



Recommendations

1. Turning off branded search terms is not for everyone

2. The same method is very usefull for other cases where you want to measure impact without A/B testing

   A few examples:
   - Launch of a new product
   - Nationwide marketing campaign (TV, printed media, Influencer collaboration)
   - Discounts, or price increase / decrease
   - Redesign
   - Etc.

Learnings & Next steps

- A/B testing is widely used (within ANWB),
  But when that is not possible, often little is learned that is truly backed by data
  - This method can be applied in those situations!
- Use cases were collected across the entire organization
  - Dozens of valuable use cases!
- Collaboration with Data Science
  - DIY step-by-step guide made available for ANWB
- Excitement throughout the organization (and beyond)
  - To learn from this and apply it more often.

Advice for you

- **Model selection**: Prophet is the most user-friendly and robust option (alternatives: CausalImpact, ARIMA & more)
- **Data requirements**: Minimum 2 years of historical data to capture seasonal patterns
- **Avoid multicollinearity**: Check for highly correlated variables, for example similar search terms
  Cluster by intent and select one representative term per cluster
- **Documentation**: Define why independent variables are predictive AND unaffected by the intervention.
  Take your time: This is often more complex than expected
- **Try it!** Causal Impact Analysis provides valuable insights when A/B testing isn't feasible

Thanks for your attention!

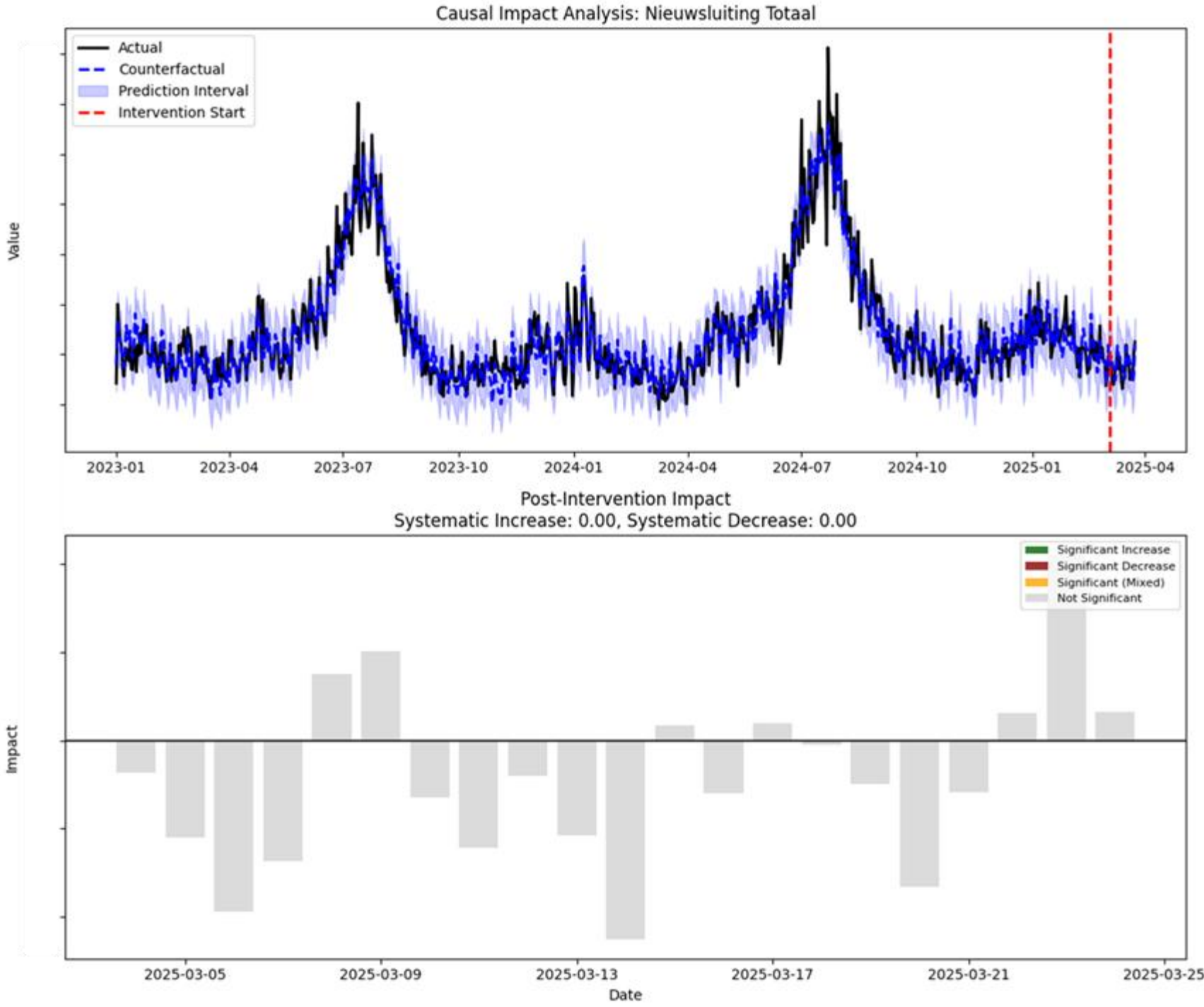
Questions?

Extra slides (for questions)

Results (more detailed)

Sales Total

- Real output looked like this.
- For presentation purpose we zoomed in on the last bit of time-line



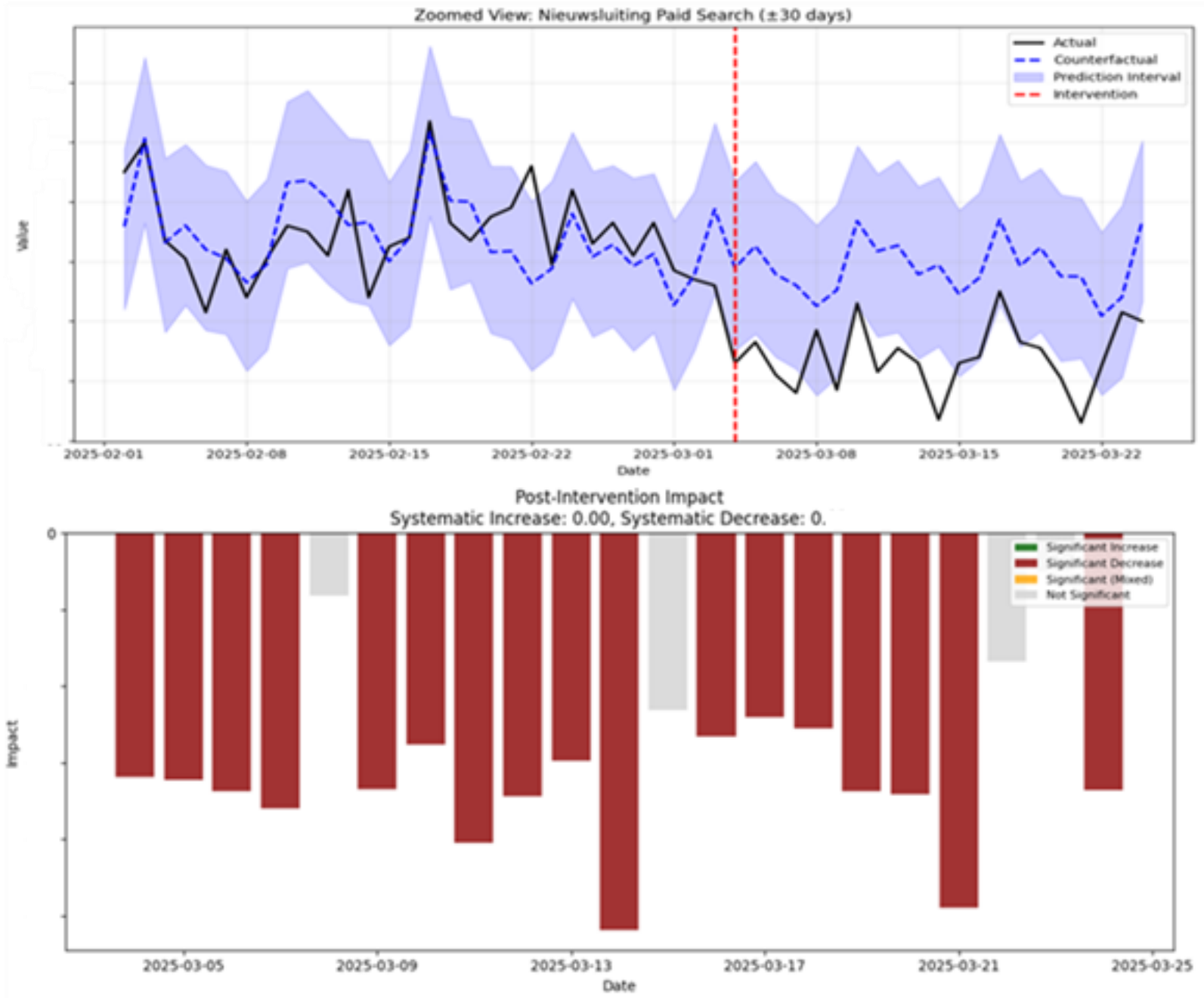
Nieuwsluitingen Paid Search

Observed sales consistently outside 90% confidence interval of prediction

Significantly lower ✅

```
Causal Impact Results:
Average daily impact:   -1,44%
Total cumulative impact: -30,27%
Overall significant: True
Combined p-value: 0.000000
```

Measured daily impact:
Often significant decrease



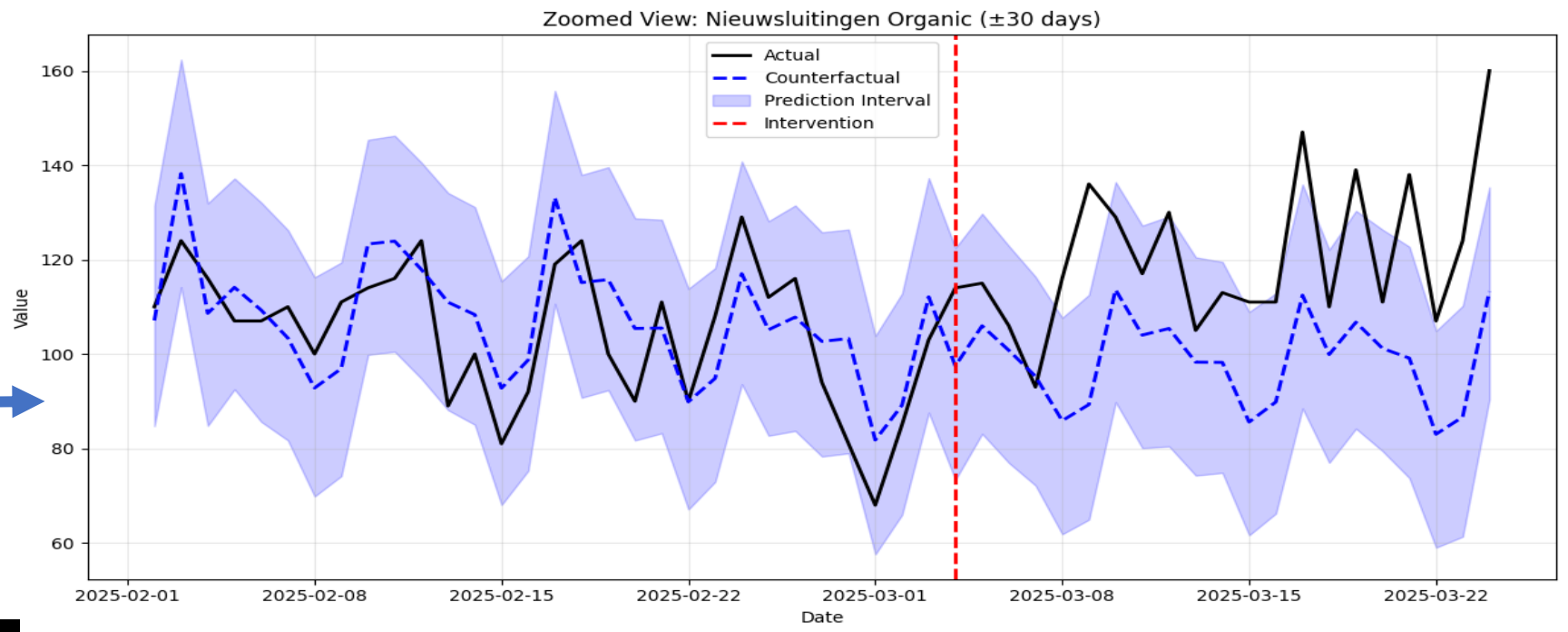
Nieuwsluitingen Organic Search

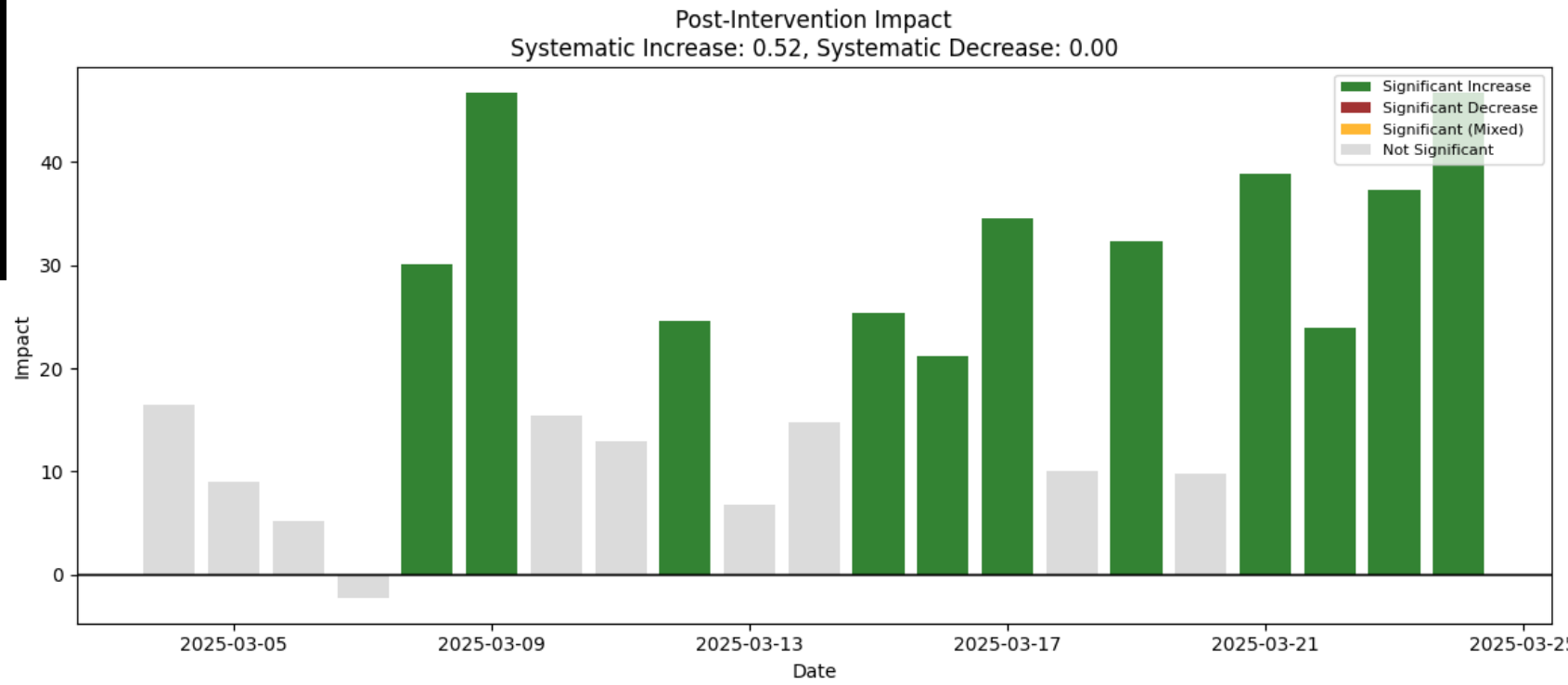
Observed sales consistently outside 90% confidence interval of prediction

Significantly higher ✅

```
Causal Impact Results:
Average daily impact:   +0,71%
Total cumulative impact: +14,89%
Overall significant: True
Combined p-value: 0.000000
```

Measured daily impact:
Often significant increase





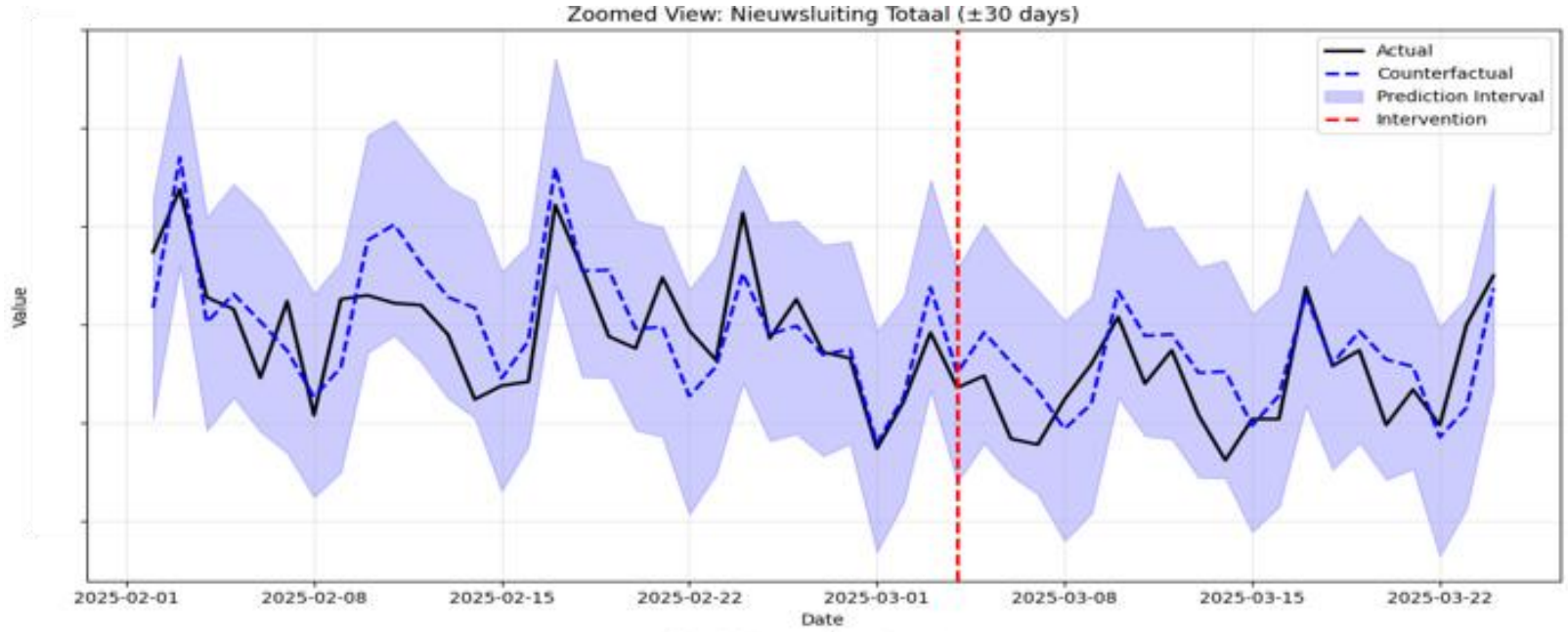
Nieuwsluitingen Totaal

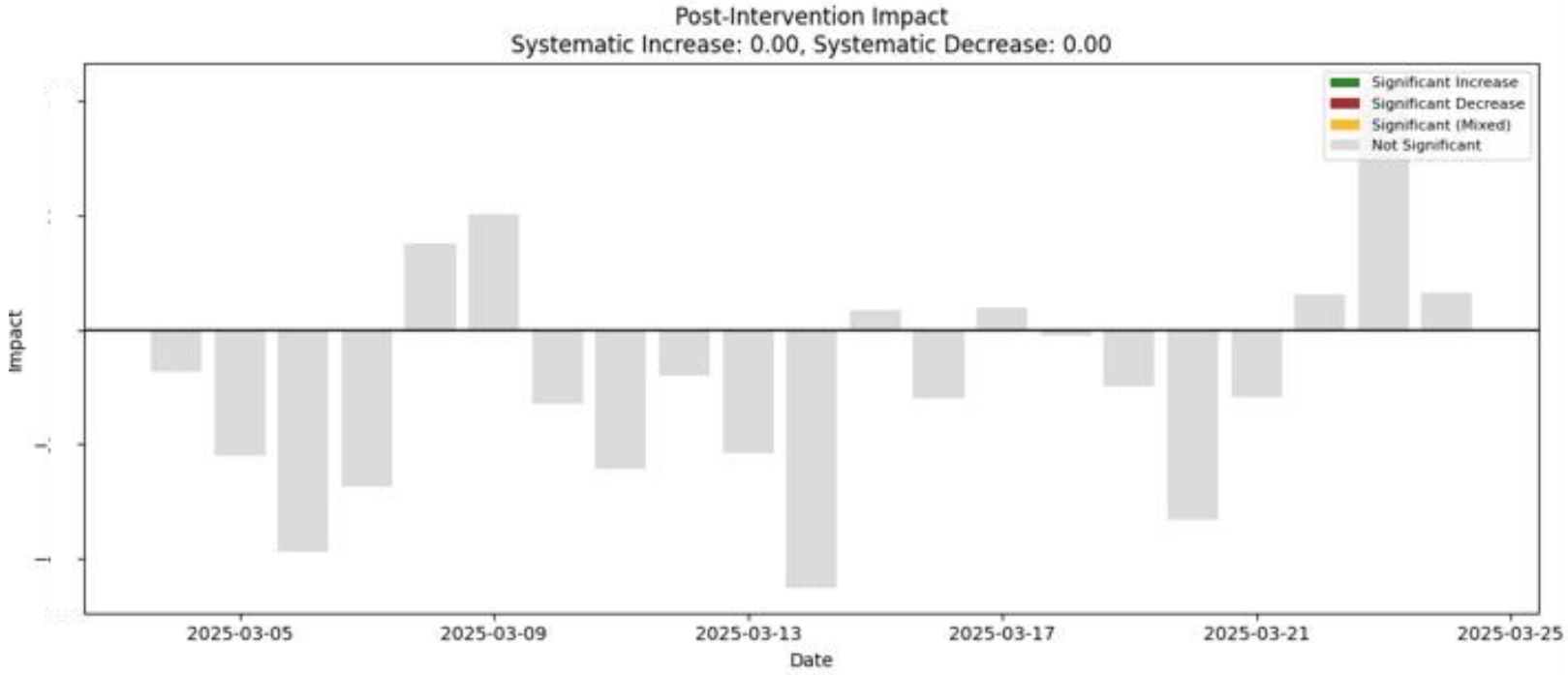
Observed value follows the prediction.
Well within the confidence interval

No significant change ✅

```
Causal Impact Results:
Average daily impact:  -0,12%
Total cumulative impact: -2,52%
Overall significant: False
Combined p-value: 0.904528
```

Measured daily impact:
Sometimes increase,
Sometimes decrease,
Not significant





Validation

Validation

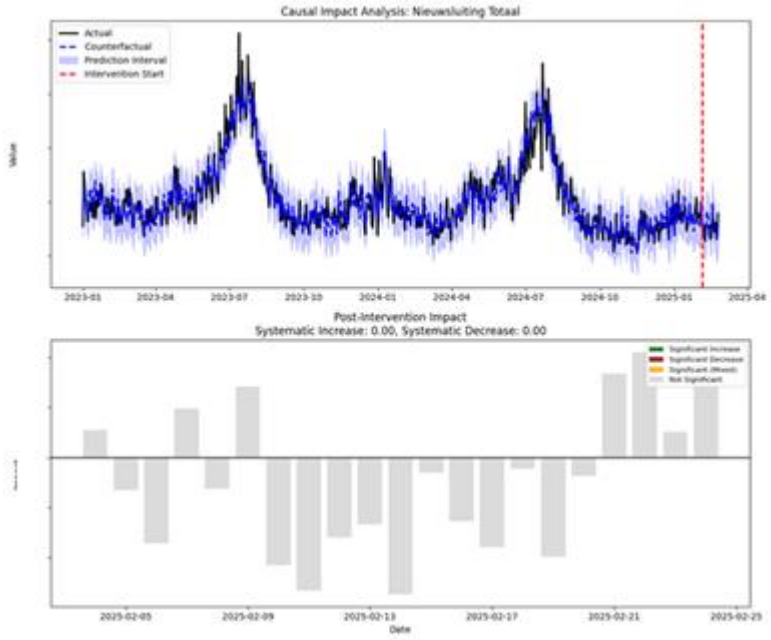
Method:

The same analysis with the intervention date shifted randomly
(for example: 1 month & 1 year earlier)

“Intervention”: 2025-02-04

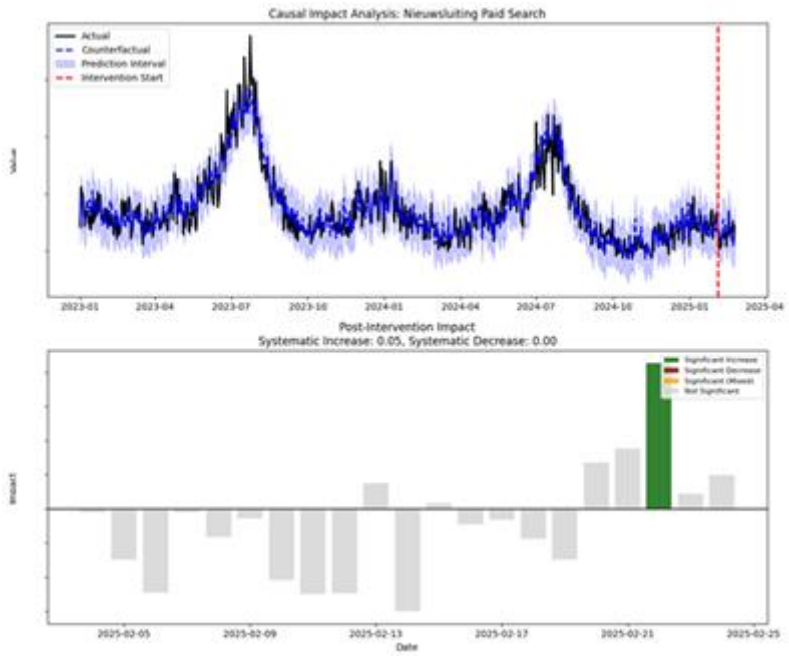
1. Total: NOT changed? ❌

```
Causal Impact Results:
Average daily impact: [redacted]
Total cumulative impact: [redacted]
Overall significant: False
Combined p-value: 0.867347
```



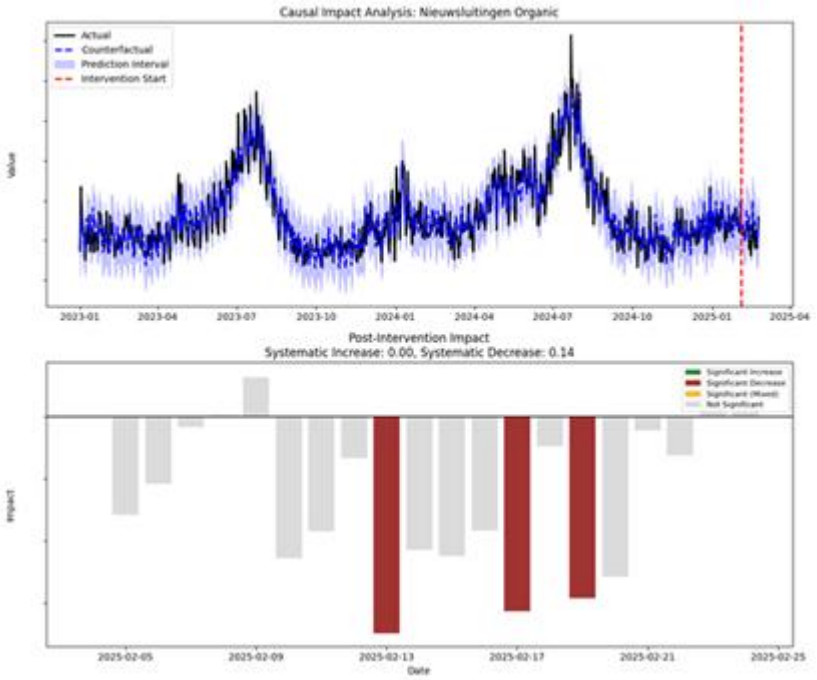
2. Paid Search lower? ❌

```
Causal Impact Results:
Average daily impact: [redacted]
Total cumulative impact: [redacted]
Overall significant: False
Combined p-value: 0.773974
```



3. Organic Search higher? ❌

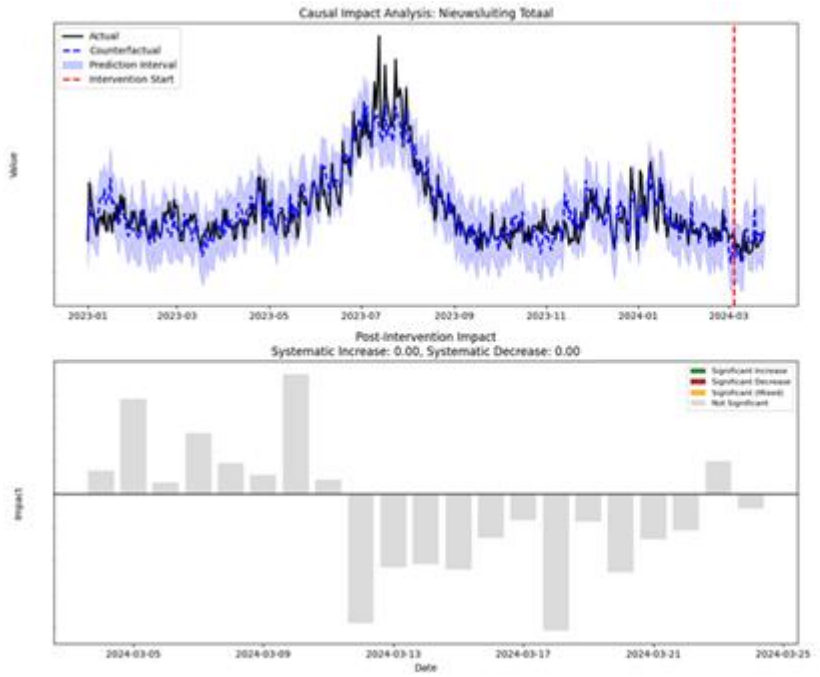
```
Causal Impact Results:
Average daily impact: [redacted]
Total cumulative impact: [redacted]
Overall significant: False
Combined p-value: 0.258480
```



“Intervention”: 2024-03-04

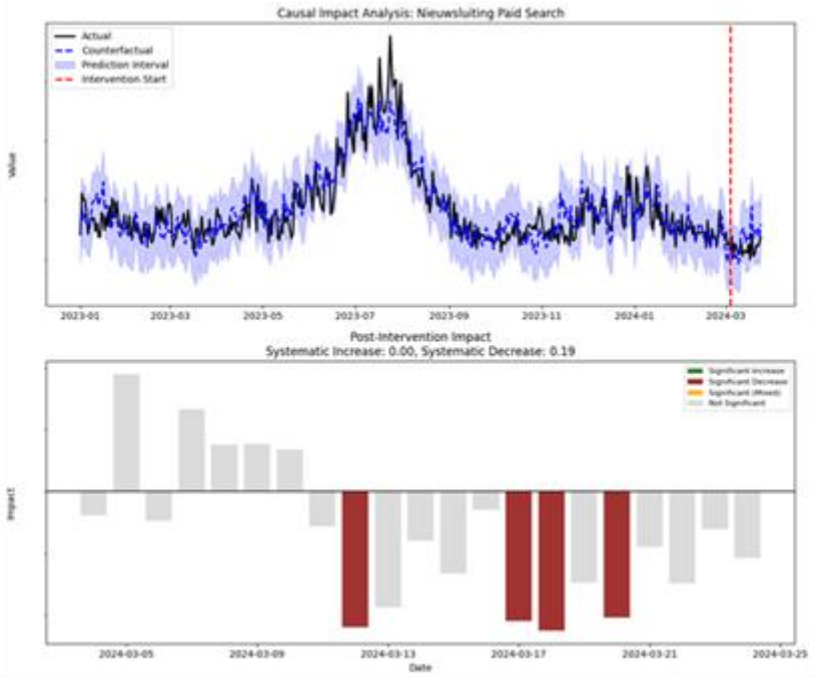
1. Total: NOT changed? ❌

```
Causal Impact Results:
Average daily impact:
Total cumulative impact:
Overall significant: False
Combined p-value: 0.866213
```



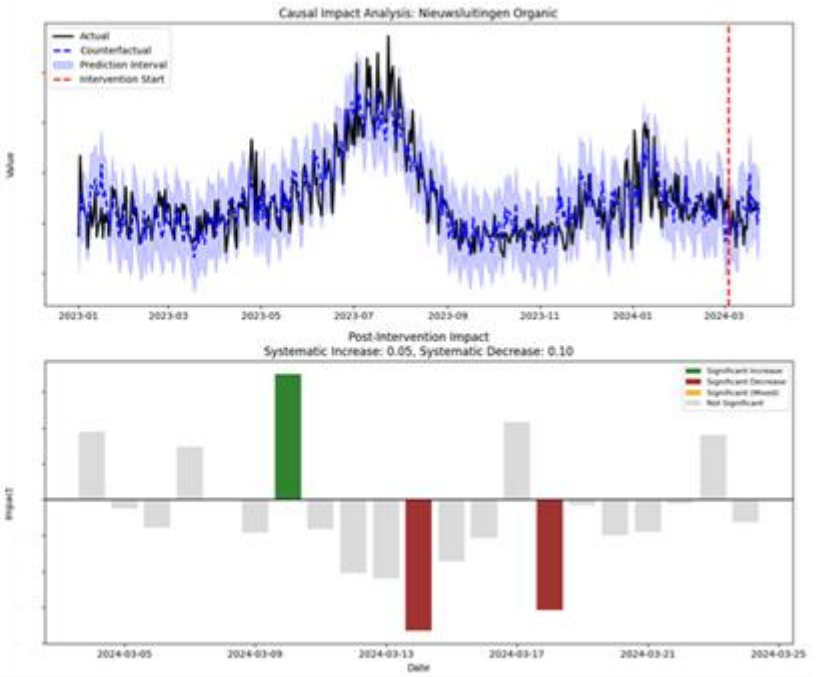
2. Paid Search lower? ❌

```
Causal Impact Results:
Average daily impact:
Total cumulative impact:
Overall significant: False
Combined p-value: 0.218250
```



3. Organic Search higher? ❌

```
Causal Impact Results:
Average daily impact:
Total cumulative impact:
Overall significant: False
Combined p-value: 0.490301
```



Learning moments

1. Model selection

Started out with Google's CausalImpact package.

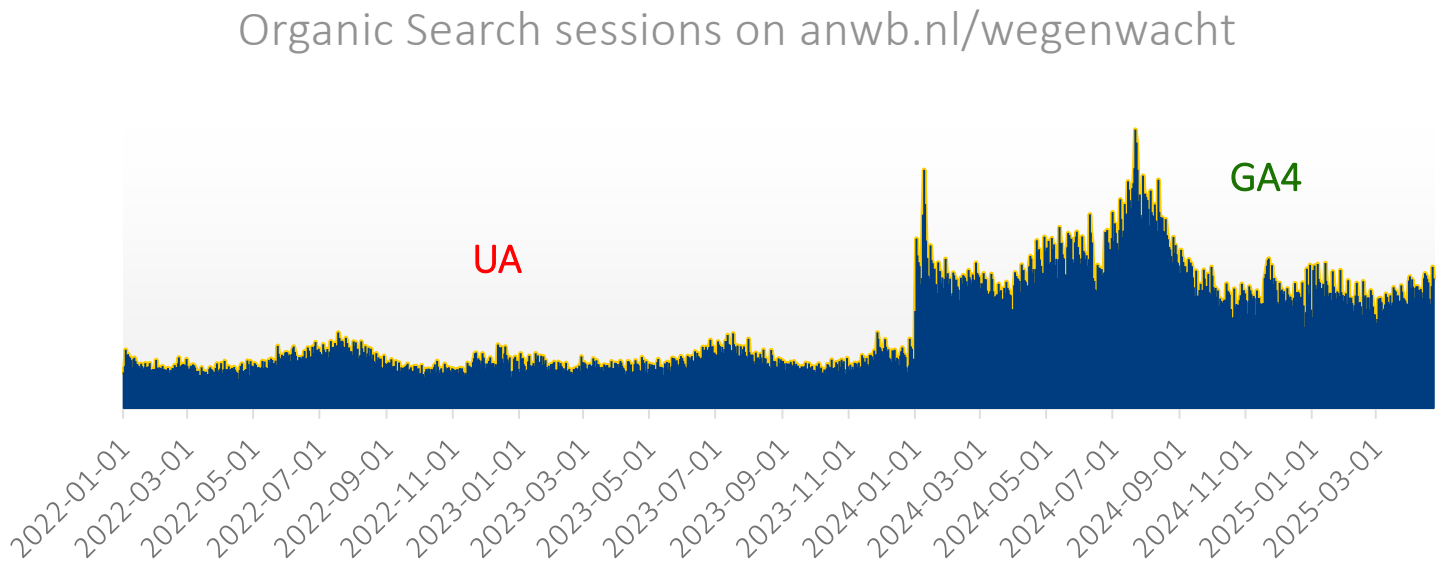
- Does not work well with seasonality. So it did not work so well
- Unfortunately a lot of time lost working out why it did not give us good results
- Lot of time lost on manual tuning of the datasets

Learning:
Just use Prophet

Model:	**CausalImpact**	**Prophet**	**ARIMA**
Type model:	Bayesian structureel	Additief	Lineair statistisch
Focus:	Causale impact	Forecasting	Forecasting
Seizoen patronen:	Nee	Ja	Beperkt (SARIMA)
Gebruiksgemak:	Matig	Hoog	Matig
Outlier robuust:	Ja	Ja	Nee
Interventieanalyse:	Ja	Nee	Nee

2. Problems with longer time-periods

Beforehand we thought: the longer the historic time periods in data, the better the prediction for the future. This gave a lot of problems with our change from Universal Analytics to GA4. Because of large differences in data:



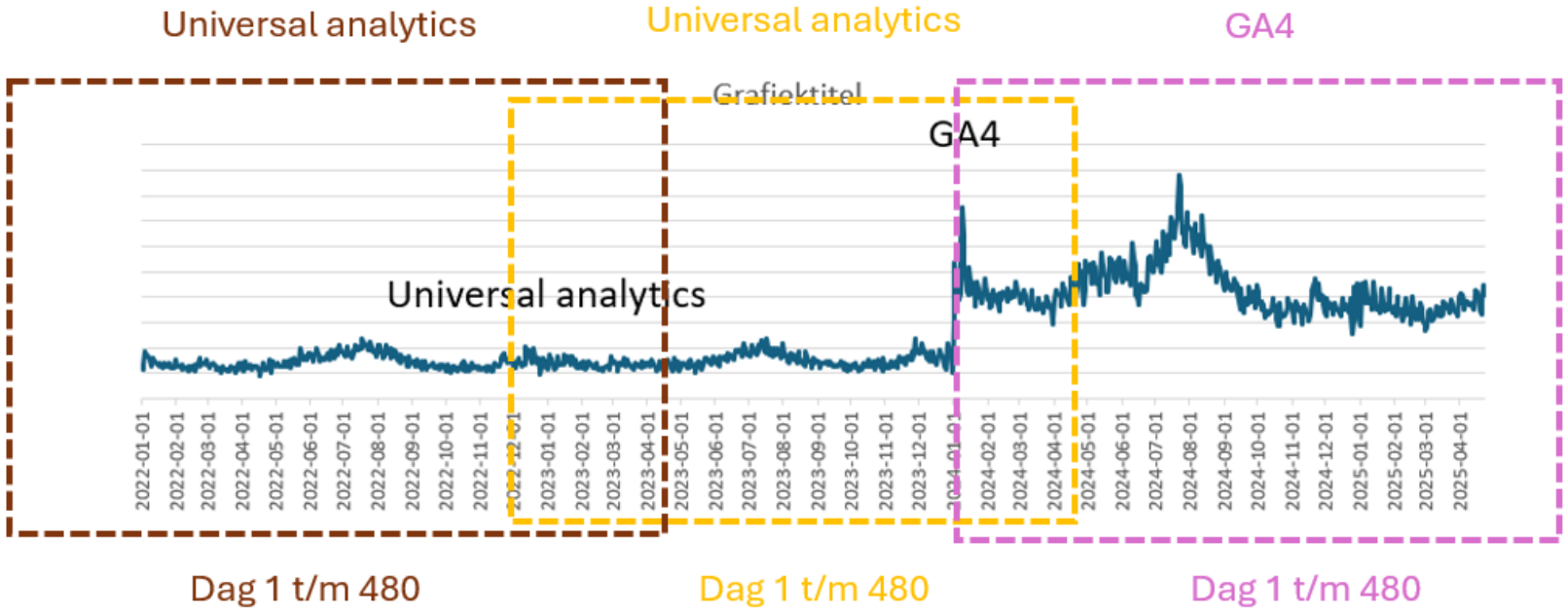
The difference between these 2 trendlines is 65.5% on average. Other datasets had differences in deviations.

Solution 1: Regression model on all datasets to normalise the data

Result: Not the result that we envisioned..

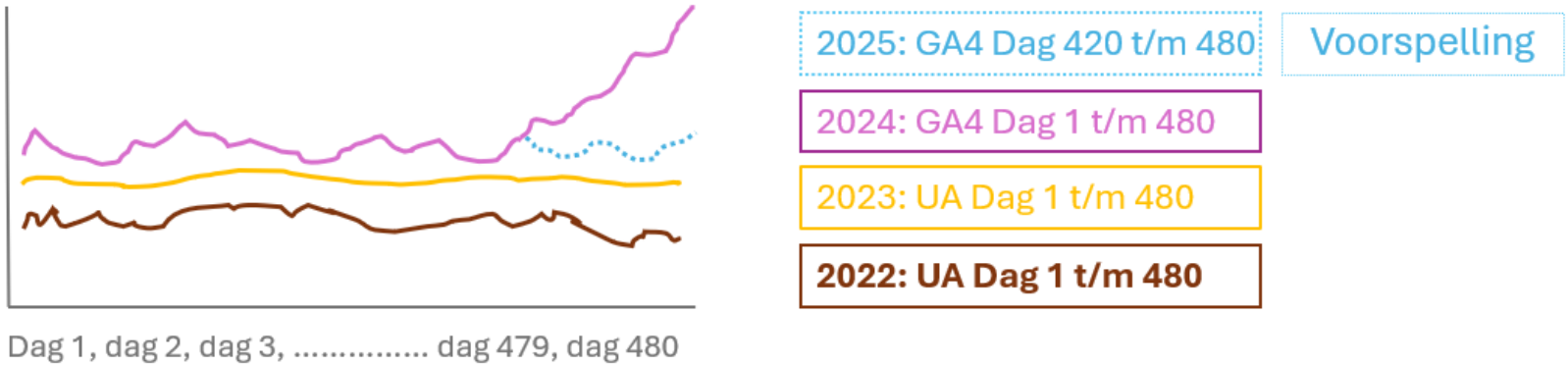
Solution 2:

Instead of 1 longer time-serie



Splice the timeline in 4 shorter time series as different predictors

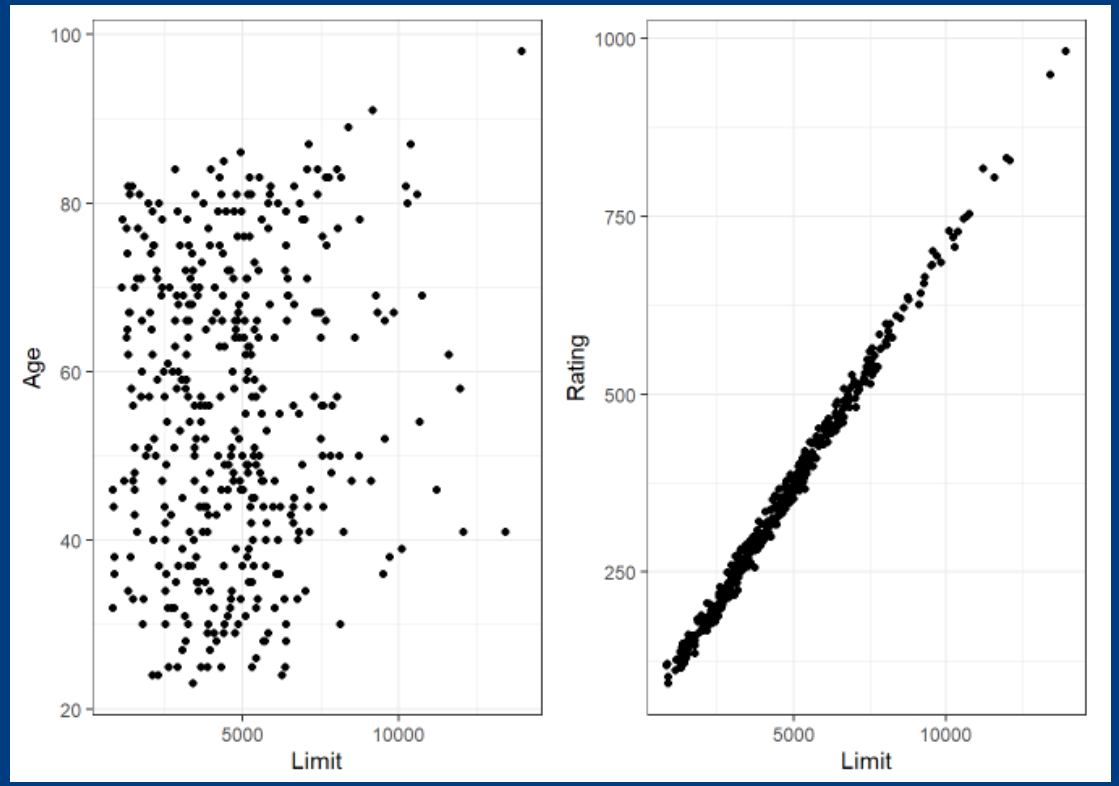
Result:
Not the result that we envisioned..



3. Multicollinearity

- Too many similar datasets with identical trend lines no longer add predictive value; instead, they increase the risk of multicollinearity
- In our case; Google Search terms were a good option for data selection, but terms like *“Wegenwacht”* and *“Pechhulp”* were too similar.
- Hence, use distinctive regressors, including unique keywords.

Learning:
Before applying the Prophet model, run a collinearity test on your datasets.



4. Daily vs weekly data

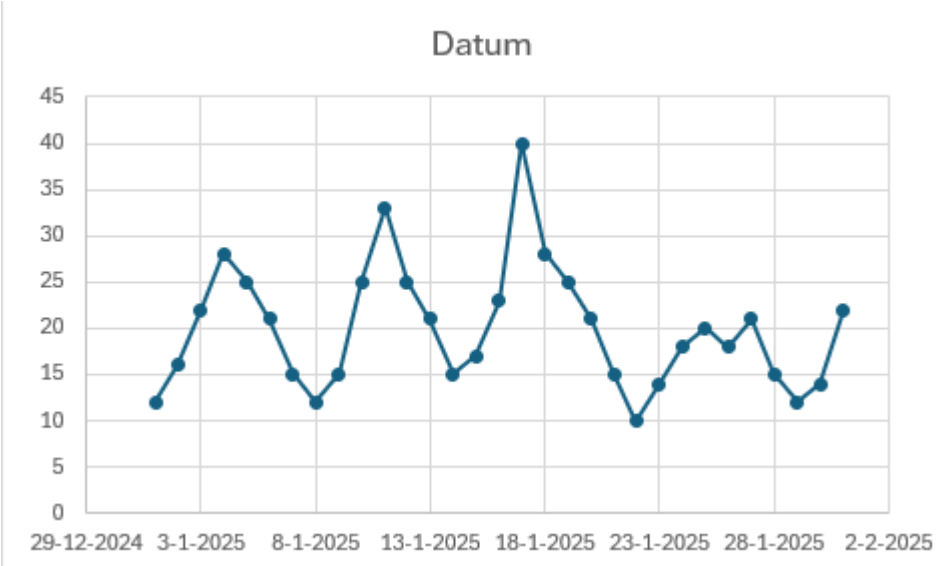
Some data is only available on a weekly basis.

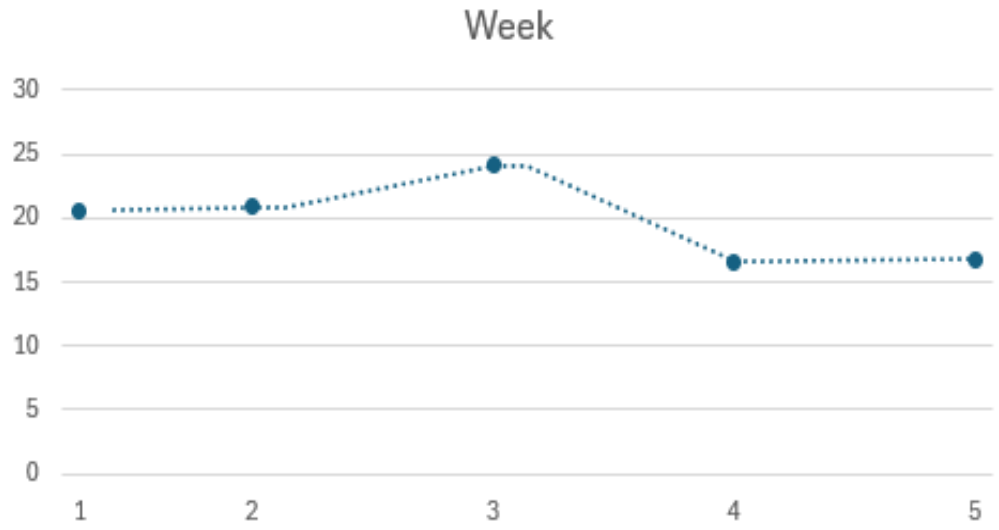
It is not possible to mix datasets at weekly and daily levels in the same model.

In that case, it's better to convert all data to **daily granularity** using Python (or Copilot) rather than converting everything to weekly.

For example, assign each day the **average value of that week**.

Learning: More data points = better predictive model.





5. Werk je data handmatig bij

- Identify extreme outliers by plotting your data in graphs and manually adjust those outliers
- Use Python to convert weekly data into daily data by evenly distributing the weekly value across individual days.
- For missing data (e.g. '*null*'):
  - Fill in a logical value for a missing day (e.g., an average, previous value, or zero)
  - Or use the last known value
  - If too many values are missing, it may be wise to exclude the dataset entirely

Learning:
Prophet is very robust, but it's better to have clean data as input and get better results.

6. Google Trends datasets (daily data) available in Python

Unfortunately, daily Google Trends data (Google search queries) only goes back one year

However, using Python, you can download datasets that retrieve historical search term data beyond one year. There is a daily limit, so if you need many datasets, patience is required.

Other useful datasets include:

- Public holidays
- Vacation periods
- Stock prices
- Weather data, etc.

Learning:
Check if the datasets you need are available in a public Python library (*Pandas*).