Case Study PYQ2

MMM Notebook Folder

Step	Notebook	Description
1	`1. data_cleaning.ipynb`	Cleaning the dataset
2	`2. eda.ipynb`	Exploratory Data Analysis
3	`3. additive_model.ipynb`	Base linear regression
4	`4. log-log-model.ipynb`	Elasticity-based regression
5	`5. Feature_Engineering.ipynb`	Adstock, saturation, interactions
6	`6. model_improve.ipynb`	Multicollinearity, final model, optimization

Data Understanding

The dataset spans 122 weeks of historical records for a global hair care brand, including sales, marketing activities, promotions, and external factors. Each record corresponds to a weekly observation. Here's a breakdown of the variable categories:

Variable Categories

- Target Variable:
  - Sales: Weekly sales revenue
- Base Variables:
  - Average Price: Average selling price of SKUs
  - Total SKUs: Number of stock-keeping units available
- Paid Marketing Channels:
  - Paid Search, Paid Social, Modular Video, Email: Included both Impressions and Spend
- Non-Paid Marketing Channels:
  - Organic Search Impressions: Organic visibility efforts
- Promotions:
  - Discount 1, Discount 2: Represent distinct promotional campaigns
- External Indicators:
  - Gasoline Price: Proxy for broader economic activity or consumer mobility
- Events:
  - Holiday: Indicator for holiday-related uplift

The dataset contained inconsistencies in formatting and missing values, primarily in the numeric fields. Specifically:

- Several columns (including Sales, Email Clicks, Gasoline Price) were encoded using the Indian number system, with comma separators (e.g., "5,54,97,076.1").
- Missing values were often represented using dashes ('-'), or blank strings ('').

To standardize the dataset:

- A custom cleaning function was applied to:
  - Strip out comma separators and whitespace
  - Convert '-' and '' into proper NaN
  - Cast cleaned values to float data type

To capture trends and seasonality, we extracted key features from the Week_Ending column:

- Week: ISO week number (1 to 52)
- Month: Calendar month (1 to 12)
- Year: Extracted for long-term trend separation

The Holiday column contained '-' in place of zeros. This was standardized as follows:

- '-' and nulls → converted to 0 (no holiday)
- Valid entries were retained as 1 (holiday week)

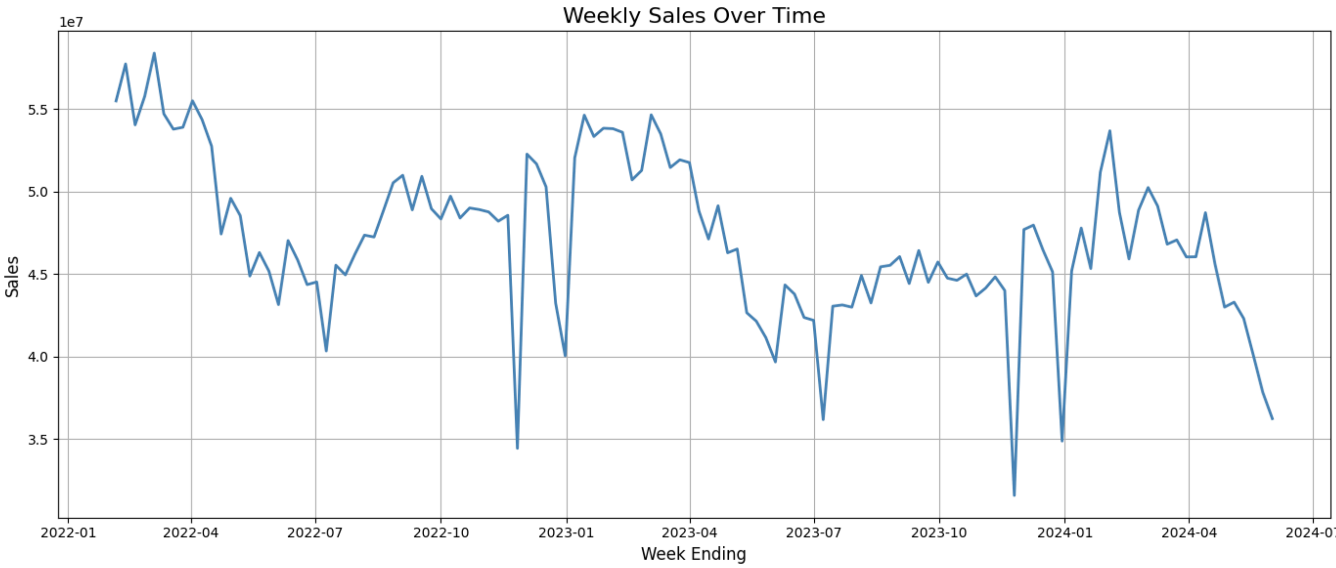
Exploratory Data Analysis

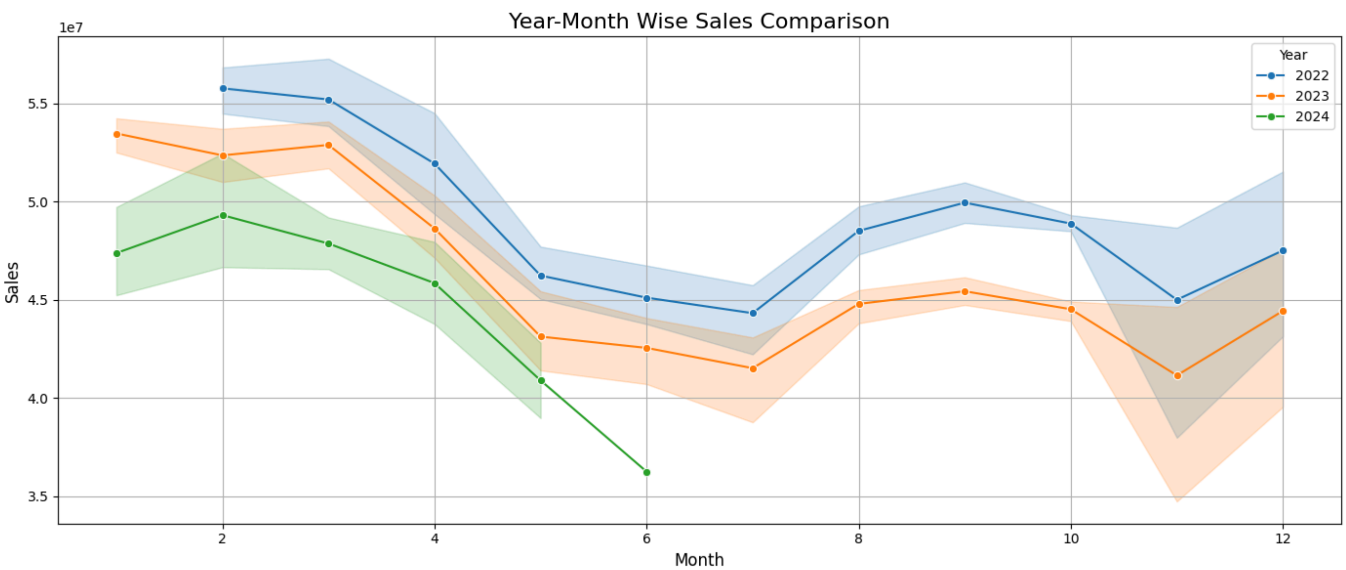
We'll explore:

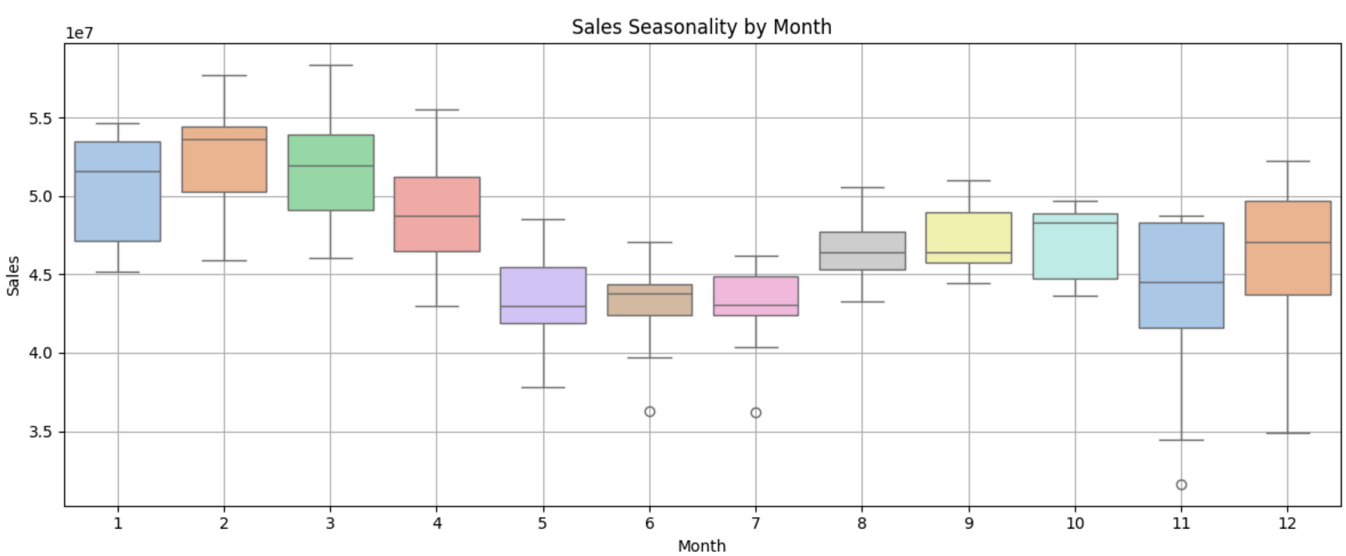
1. **Sales Trends Over Time** – seasonal patterns, growth
2. **Seasonality** – holiday impact, month/week trends
3. **Discount Impact** – check if Discount1/2 align with spikes in sales
4. **Marketing Spend vs. Impressions** – for ROI intuition
5. **Organic vs. Paid Traffic Contributions**
6. **Economic Factor Impact** – how gasoline prices relate to sales
7. **Correlations** – between sales and all predictors

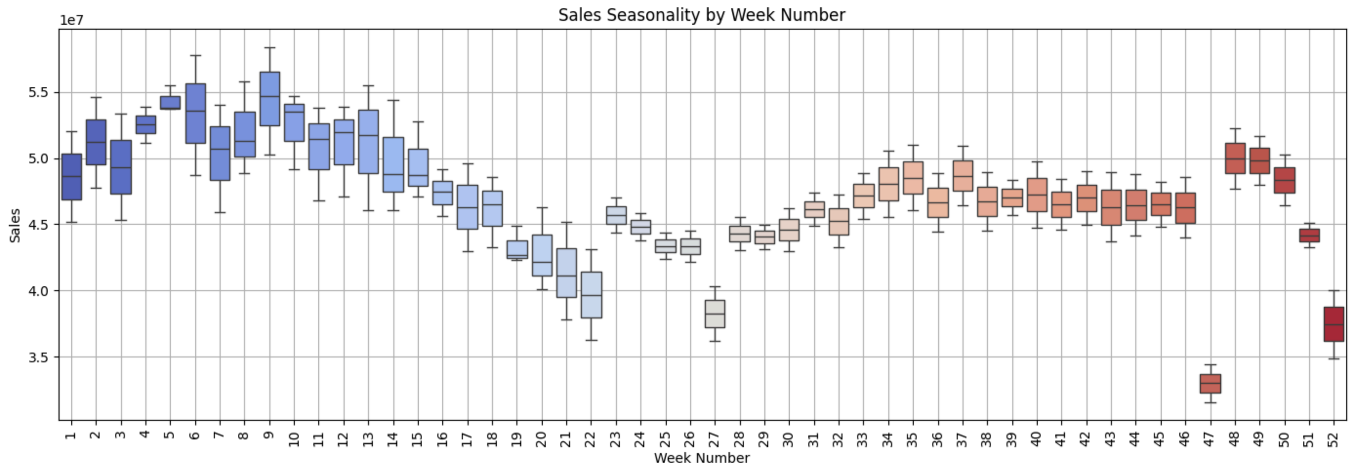
Note: Visual wise interpretation in detail present in the notebooks

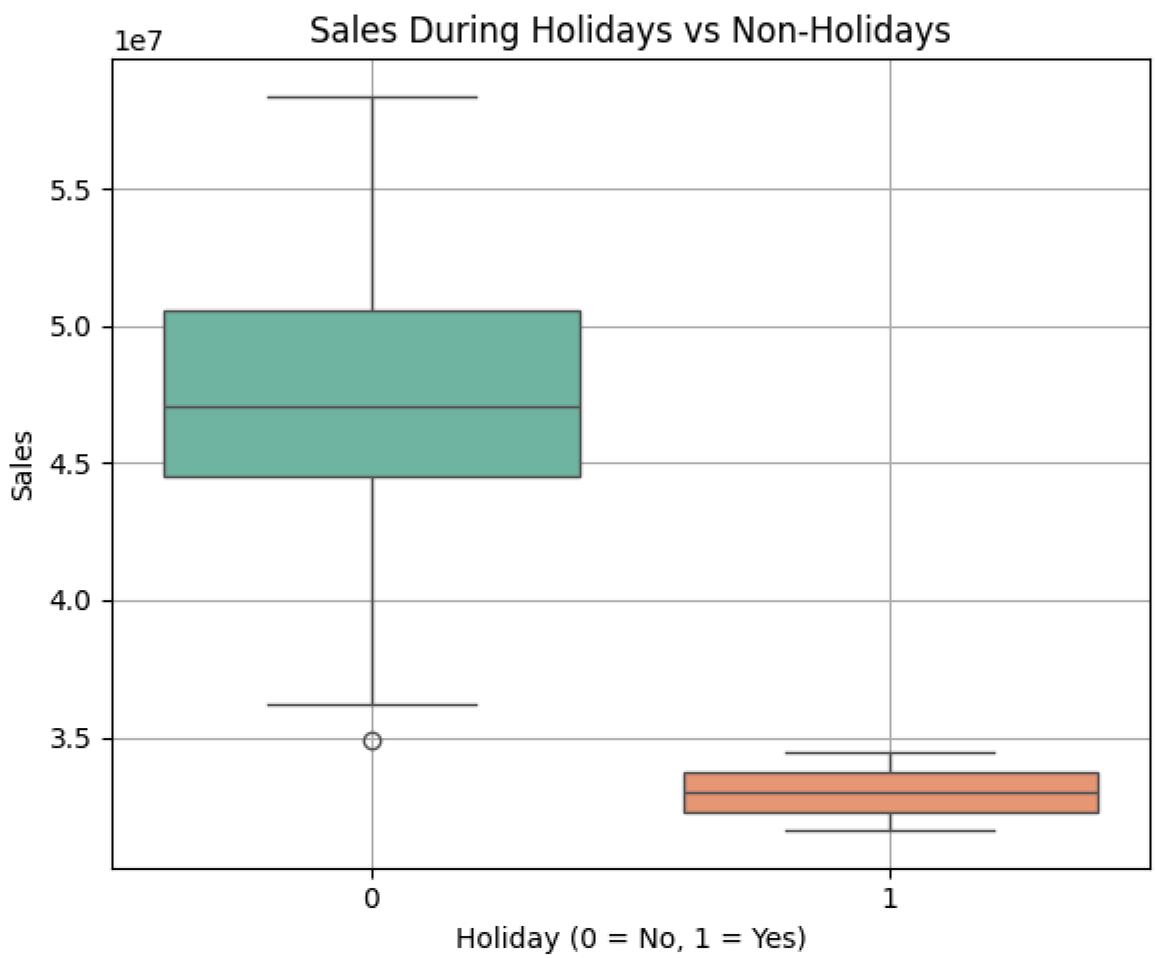
Key Sales Trend & Seasonality Insights


Weekly Sales Over Time
1e7
Sales
Week Ending
2022-01
2022-04
2022-07
2022-10
2023-01
2023-04
2023-07
2023-10
2024-01
2024-04
2024-07


Year-Month Wise Sales Comparison
1e7
Sales
Month
Year
2022
2023
2024


Sales Seasonality by Month
1e7
Sales
Month





Strong Starts

- Sales consistently peak in **Q1 (Jan–Mar)** across years — fueled by New Year buzz, winter demand, and fresh marketing push.
- A **mid-year dip (May–Jul)** follows, reason could be summer seasonality and reduced campaign activity.

Second Half of Year

- August through to October marks a reliable recovery period after the post summer slump in sales.
- December experiences a boost from holiday marketing efforts but the results are always year week and year dependent.

November & Week 47 Dips

- **November** underdelivers consistently — possibly a missed promo window or campaign.
- An odd drop in **Week 47** suggests a potential campaign miss or one-off issue worth digging into.

Predictable Seasonality

- The brand shows a **clear U-shaped sales curve** each year: strong start, mid-year slowdown, partial recovery toward year-end, November dip, then again strong start

Underperformance

- Weaker sales than previous year across most periods, even if the pattern, trend maintained — pointing to weaker marketing, reduced budgets, or market headwinds.

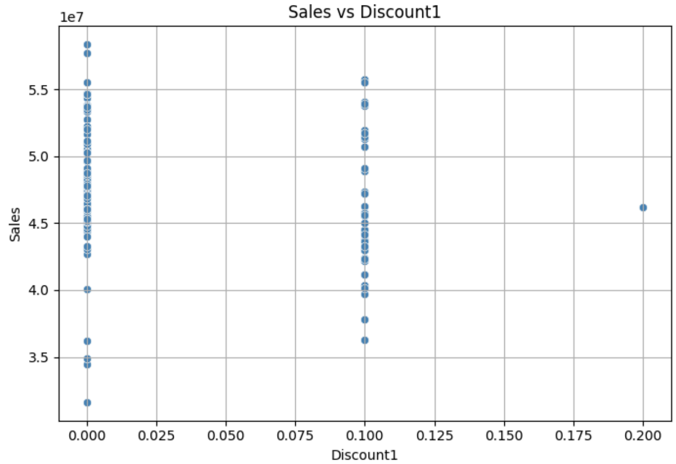
Holidays Don't Help

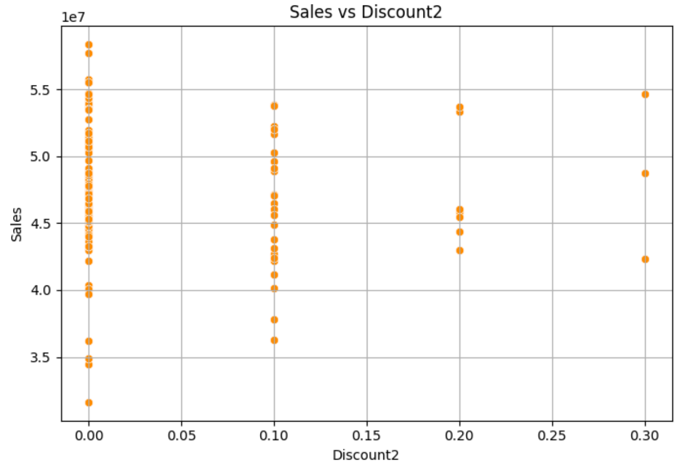
- Surprisingly, **holiday weeks underperform**, with lower medians and tighter spreads — indicating they may be non-commercial holidays or marketing gaps.

No Long-Term Growth Yet

- While peaks occur, there’s **no clear upward progress** over time — the brand is holding ground, but not scaling meaningfully.

Discounts & Sales





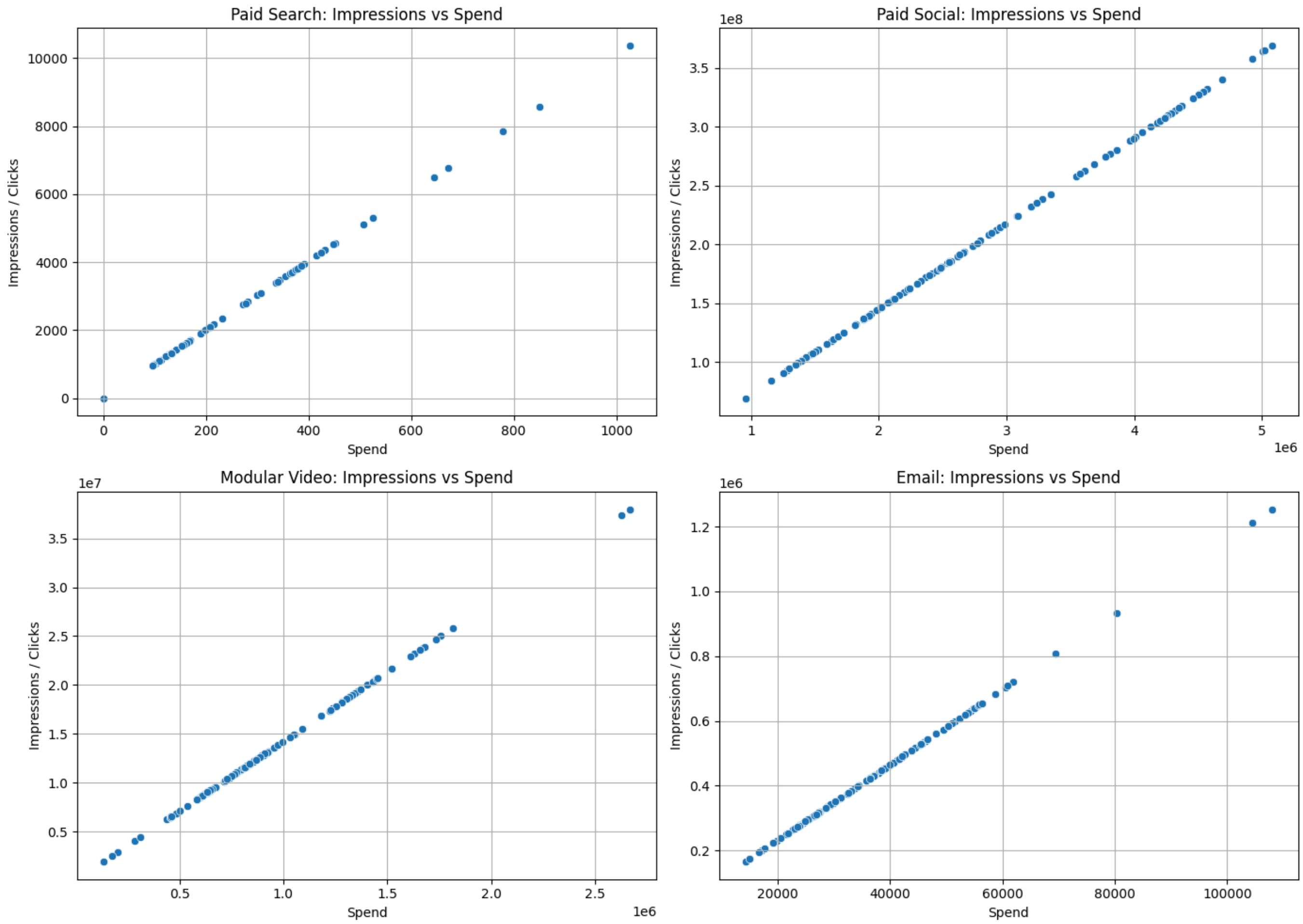
Discount1:

- Most data sits at **0% and 10%**, with little variation in sales.
- Surprisingly, some **strongest sales happen with 0% discount**.
- Discounts here look like a **supporting tactic**, not the main trigger.

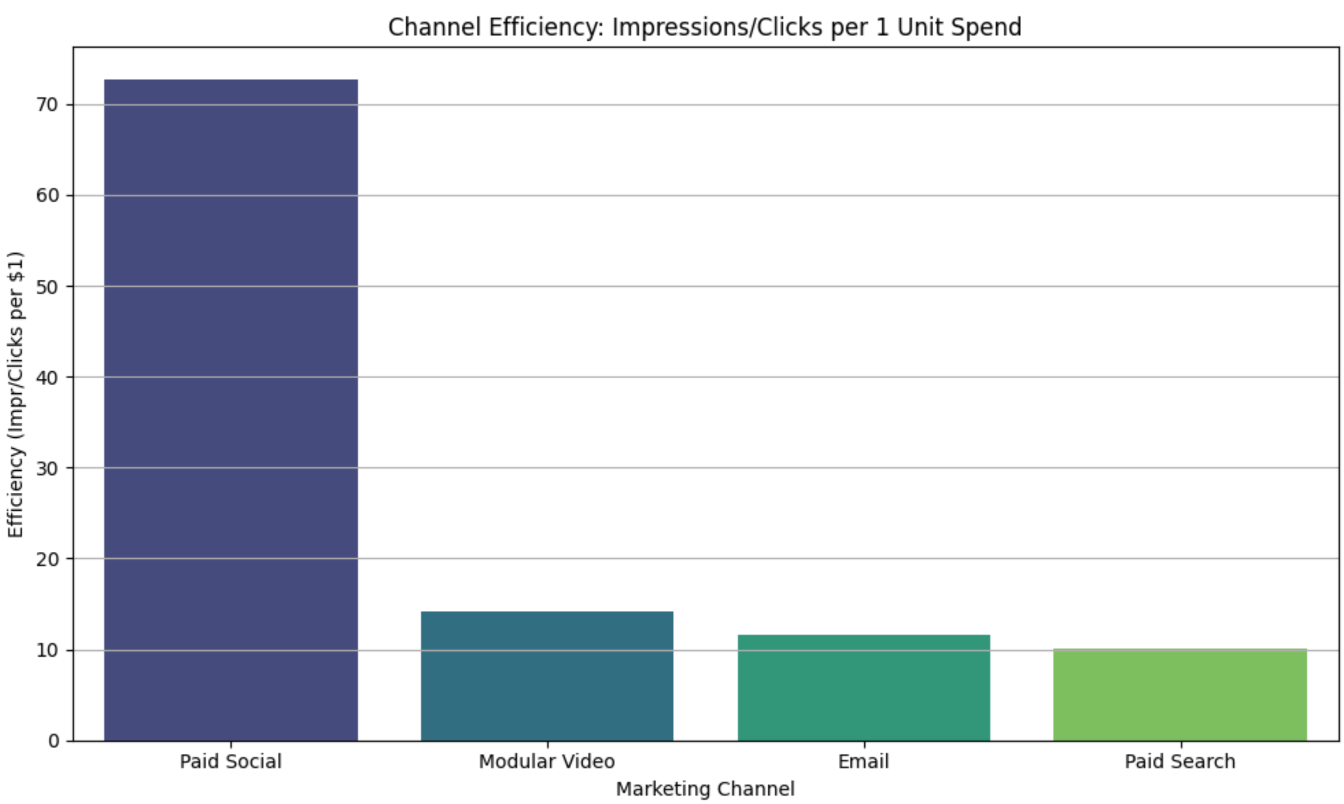
Discount2:

- Sales are scattered across 0% to 30% — but **no clear lift with higher discounts**.
- Even at 30%, sales stay moderate — pointing to **diminishing returns** or poorly timed offers.
- Implies a **discount cap**: after a point, slashing prices doesn't add value.

Marketing Spend Analysis

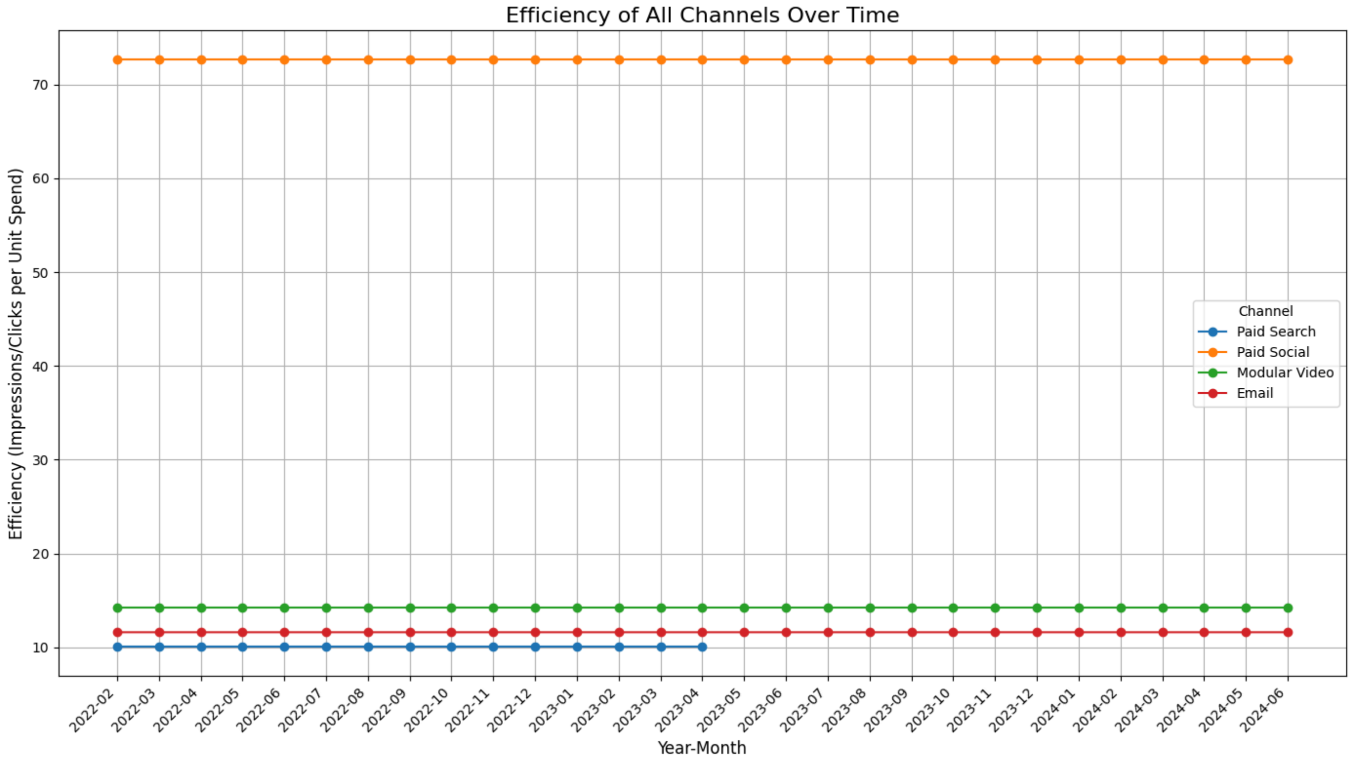


Paid Search: Impressions vs Spend
Paid Social: Impressions vs Spend
Modular Video: Impressions vs Spend
Email: Impressions vs Spend
Impressions / Clicks
Spend
1e8
1e7
1e6



Channel Efficiency: Impressions/Clicks per 1 Unit Spend
Efficiency (Impr/Clicks per $1)
Paid Social
Modular Video
Email
Paid Search
Marketing Channel

	Year	Month	Paid Search Impressions	Paid Search Spends	Paid Social Impressions	Paid Social Spends	Modular Video Impressions	Modular Video Spends	Email Clicks	Email Spends	Paid Search Efficiency	Paid Social Efficiency	Modular Video Efficiency	Email Efficiency
0	2022	2	17762.5	1760.5	1.049896e+09	14454652.3	50092721.0	3520602.3	2342522.0	201668.6	10.089463	72.633748	14.228452	11.615700
1	2022	3	12330.5	1222.2	7.876870e+08	10844641.8	36587999.0	2571467.2	2371197.5	204137.2	10.088774	72.633748	14.228453	11.615705
2	2022	4	20020.0	1984.3	9.655043e+08	13292778.7	70096092.5	4926473.4	2975073.5	256125.2	10.089200	72.633748	14.228452	11.615700
3	2022	5	18123.0	1796.2	9.458241e+08	13021827.4	35772212.0	2514132.4	1906348.5	164118.2	10.089634	72.633748	14.228452	11.615704
4	2022	6	9649.5	956.4	6.936977e+08	9550625.4	45964212.0	3230443.4	2154036.5	185441.7	10.089398	72.633748	14.228453	11.615707



1. Paid Search

- **Strong positive linearity** between spend and impressions.
- Highly **predictable delivery** — every dollar spent gets proportional visibility.
- Indicates a **well-optimized, auction-based channel**.
- **Paid Search exhibits the lowest efficiency**
  - At around **10.09**, Paid Search trails behind the other channels in efficiency.
  - This could be due to higher competition or cost-per-click, suggesting a potential need for further optimization or reallocation of budget.

2. Paid Social

- Also shows a **positive linear trend**, though slightly more **scattered** than Paid Search.
- **Paid Social demonstrates the highest efficiency**
  - With a consistent efficiency value of approximately **72.63**, Paid Social significantly outperforms other channels in terms of impressions per unit spend.
  - This suggests an excellent return on investment and stable performance over time.

3. Modular Video

- Linear but with **more noise** and **flat stretches**.
- Suggests potential **thresholds or caps**.
- May need **frequency capping** or better pacing strategy.
- **Modular Video ranks second in efficiency**
  - Modular Video maintains an efficiency of around **14.23**, indicating a relatively strong performance.
  - Its consistency suggests effective budget allocation and campaign management within this channel.

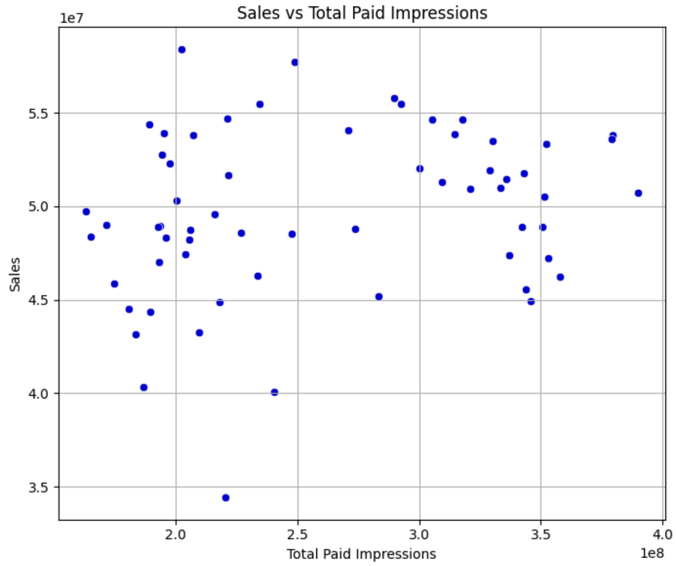
4. Email

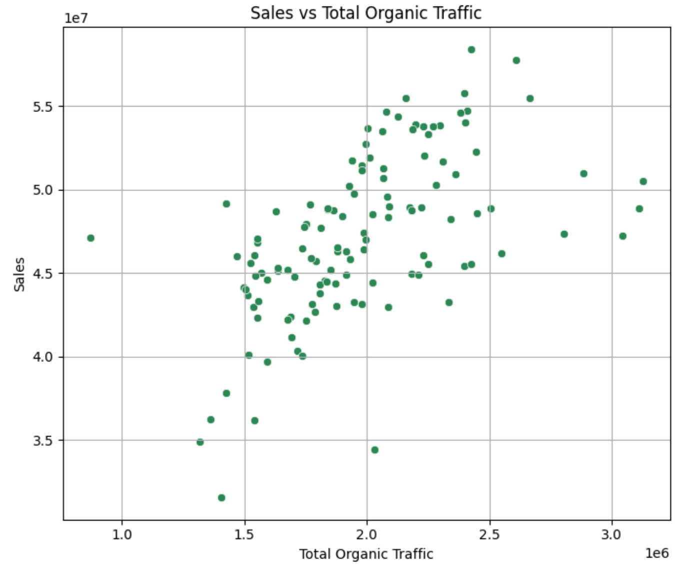
- Very **non-linear** — increasing spend doesn't always lead to proportional increases in clicks.
- Could indicate:
  - Saturation of the audience (list is finite)
  - Poor targeting or deliverability
  - Need for better creative or subject-line testing
- **Email channel shows stable but moderate efficiency**
  - Email's efficiency stands at approximately **11.62**, indicating decent performance with reliable returns.
  - While not the highest, its stability suggests it remains a dependable channel for engagement.

Channel efficiencies are highly consistent over time

- All channels show minimal variation in efficiency month-over-month, indicating a stable media execution strategy.
- However, the lack of fluctuations might also reflect limited experimentation or optimization efforts.

Organic vs. Paid Traffic Contributions





1. Sales vs Total Paid Impressions

Key Observations:

- The relationship between paid impressions and sales is weak, especially at higher impression volumes.
- Beyond approximately 250 million impressions, increases in sales begin to level off.
- Higher paid media investment does not consistently lead to higher sales.

Interpretation:

- This suggests diminishing returns from paid media at scale.
- Indicates a potential overspend or inefficiency in the paid strategy.
- Reallocation or capping of budget may improve overall return on investment.

2. Sales vs Total Organic Traffic

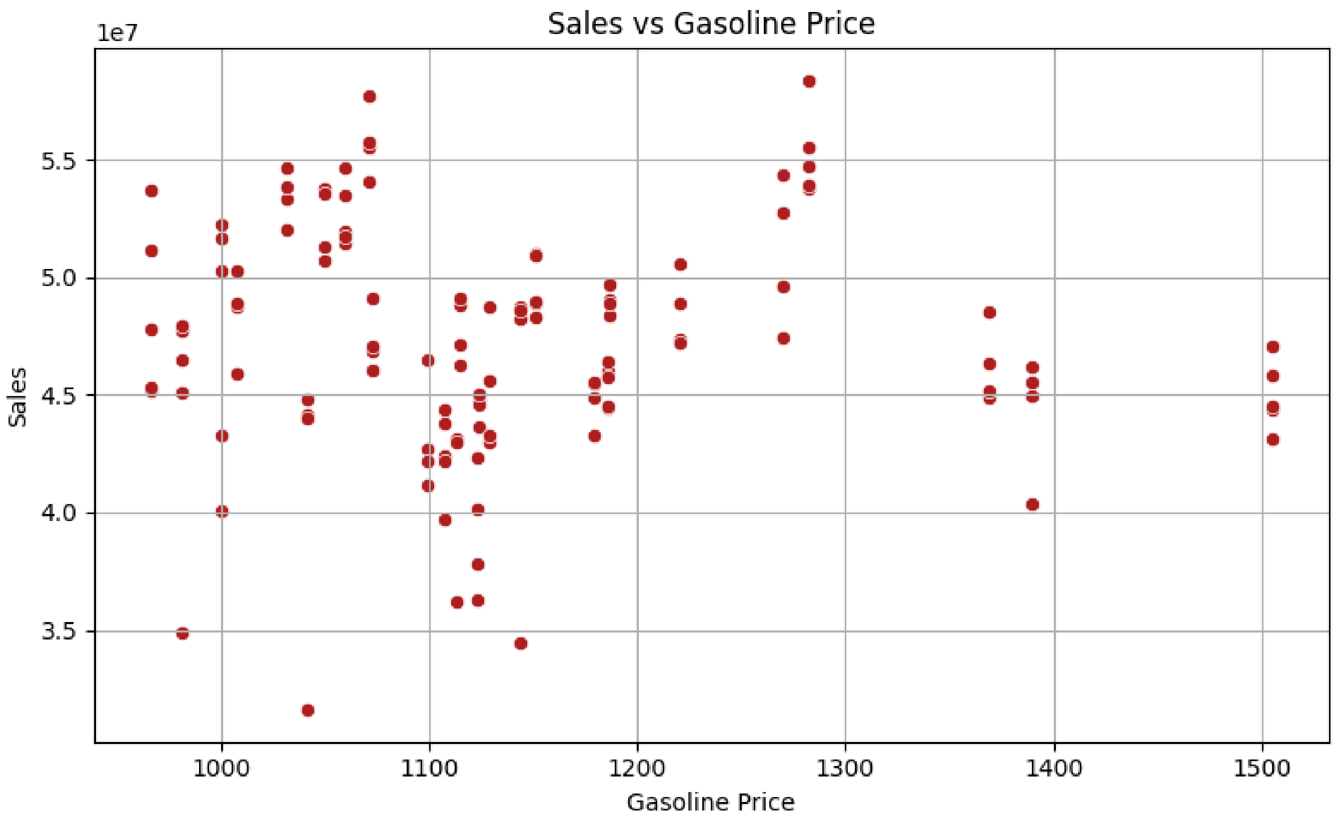
Key Observations:

- There is a clear, positive relationship between organic traffic and sales.
- The data shows a consistent upward trend, particularly up to 2.5 million in organic traffic.

Interpretation:

- Organic channel are more closely aligned with sales growth.
- Indicates stronger intent and more efficient conversion from organic sources.
- Suggests an opportunity to increase investment in SEO, CRM, and lifecycle marketing efforts.

Economic Factor Impact



Gasoline Price Impact on Sales

1. Mild Negative Relationship

- As gasoline prices rise, sales show a slight downward trend — especially beyond ₹1300.
- Not a perfect pattern, but the pressure on consumer spending is noticeable.

2. Stronger Sales at Lower Fuel Prices

- Higher sales volumes are concentrated when fuel prices are between ₹1000–₹1150.
- This suggests consumers are more comfortable spending on personal care when fuel costs are moderate.

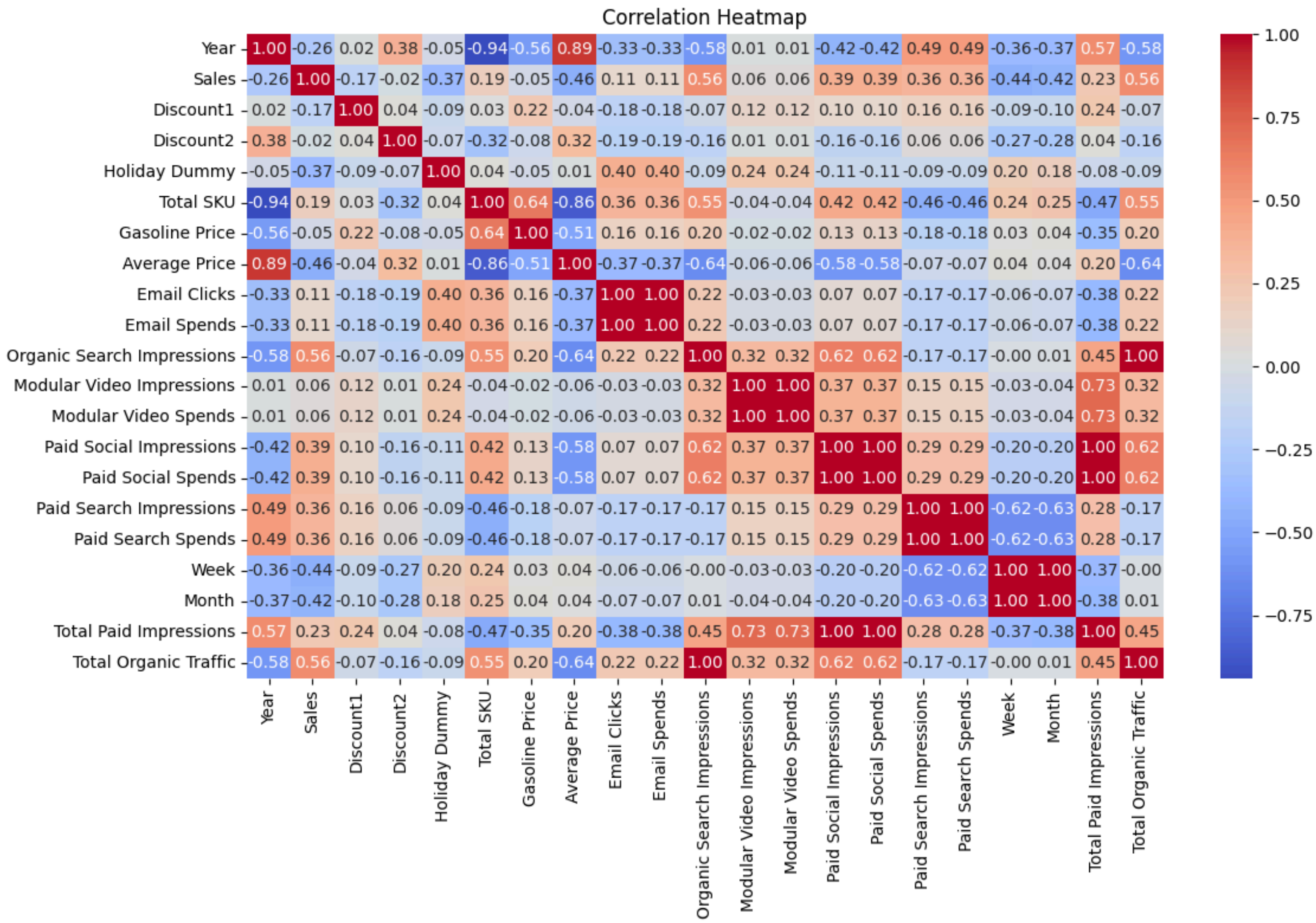
Strategic Insight

- Gasoline prices act as a **proxy for consumer sentiment and spending power**.
- Even in personal care, there's evidence of **price sensitivity** during times of economic strain.

Modeling Recommendation

- Include gasoline price as a **control variable** in marketing mix models (MMM).
- Expect a **small but meaningful negative impact** on sales.
- Consider testing **lag effects**, as spending behavior may adjust over time.

Correlation



Top Positive Drivers of Sales

- **Organic Search Impressions (+0.56):**
  Strongest driver — investing in SEO and high-quality content can significantly boost sales.
- **Paid Social (Impressions/Spend) (+0.39):**
  High-performing channel — delivers both reach and efficiency. Worth scaling smartly.
- **Paid Search (Spend/Impressions) (+0.36):**
  Steady performer — reflects value from high-intent users actively looking for solutions.
- **Total Organic Traffic (+0.56):**
  Matches paid in impact — shows the power of nurturing long-term customer relationships.
- **Total SKUs Available (+0.19):**
  More variety supports slightly higher sales — offering choice helps, though modestly.

Strong Negative Correlations with Sales

- **Average Price (-0.46):**
  Clear price sensitivity — increasing price tends to reduce sales meaningfully.
- **Discount1 (-0.17):**
  May not be effective or well-timed — could confuse rather than convert customers.
- **Month / Week (~ -0.42 / -0.44):**
  Strong seasonal patterns — needs to be factored in to avoid misleading trends.
- **Year (-0.26):**
  Indicates a possible downward trend over time — suggests looking deeper into long-term shifts.

Neutral or Noisy Variables

- **Modular Video (~0.06):**
  Very limited impact — might be over-invested relative to returns.
- **Discount2 (-0.02):**
  Minimal effect — may need to revisit the offer design or targeting strategy.
- **Gasoline Price (-0.05):**
  Small negative influence — while not strong, it still reflects economic pressure.

Modeling Recommendation

- Prioritize **Organic Search, Paid Social, and Paid Search** — they show consistent sales impact.
- Include **Average Price**, **Time (Week/Month)**, and **Gasoline Price** as **control variables** to account for external or structural influences.
- Check for **multicollinearity** — e.g., Paid Social Spend and Impressions are likely highly correlated. Use only one, or combine as a ratio (e.g., cost per impression).
- Explore **interaction effects** — such as Organic × Paid to capture synergy.
- If needed, consider **separate models for Organic and Paid channels** to reduce noise and increase interpretability.

Baseline Modeling

Baseline Model Choice

Option 1: Additive (Linear) Regression

This model uses raw sales and input values — a standard linear relationship.

When to use it:

- When working with **direct units**, like "an extra ₹1000 in spend increases sales by ₹500".
- Suitable if we expect **straight-line relationships** between inputs and sales.

Limitations:

- Doesn’t capture **diminishing returns** — may overestimate the effect of high spends.
- Less effective with **skewed or wide-range data**.
- Coefficients are harder to use for **budget allocation or ROI planning**.

Option 2: Log-Log Regression *(Recommended for MMM)*

This model uses the logarithm of both sales and inputs (like price or media spend). It’s commonly used in marketing mix modeling.

Why it works well:

- **Easy to interpret:** Each coefficient shows the *percentage change in sales* for a *1% change* in the input.
    - Example: A coefficient of -2 on price means a 1% price increase leads to a 2% drop in sales.
- **Handles non-linear effects:** Naturally captures diminishing returns, which are common in marketing spend.
- **Fixes skewed data:** Useful when variables like impressions or price have large ranges or outliers.

What to keep in mind:

- Inputs must be **positive** — zero values need to be handled (e.g., by adding 1 or filtering).
- Assumes **multiplicative** relationships (not additive), which fits most marketing behaviors well.

Use Lagged Gasoline Price as a Control variable

1.Behavioral Lag in Consumer Response

- When fuel prices rise, people often **adjust spending behavior in following weeks**, not immediately.
- Lagged variables help capture this **delayed effect on purchases like personal care**.

2. Delays in Budgeting

- Consumers respond over time due to:
  - Monthly budgeting
  - Weekly pay cycles
  - Psychological adaptation to cost changes
- Lagged input reflects these **natural delays in financial decisions**.

3. Avoids Causal Leakage

- Using same-week gasoline prices risks attributing sales changes to a cause that hadn't occurred yet.
- Lagging helps preserve **cause-before-effect logic** in the model.

Use Impressions in the Model And Spend for later Calculations

> **Impressions tell us what's working.**
> **Spend helps us decide how much to invest.**

1. Impressions Reflect Demand Exposure

- Impressions represent how many people saw the media — a **direct signal of reach** and brand visibility.
- Helps identify **which channels are driving engagement**, before evaluating efficiency.

2. Spend Links to ROI & Optimization

- Once effectiveness is known (via impressions), we shift focus to **Spend** → **Sales** to evaluate:
  - **ROI** = Incremental Sales / Spend
  - **Diminishing returns** (how sales slow as spend increases)
  - **Optimal budget allocation** across channels

Additive OLS Model – Summary of Results

Model Performance

- **R^2 = 0.813**
  The model explains about **81% of the variation** in weekly sales — strong performance for a marketing mix model.
- **Adjusted R^2 = 0.767**
  Still solid after adjusting for the number of variables in the model.

- **Overall significance (F-stat p-value = 1.41e-13)**
  The model is **highly statistically significant**, meaning it reliably explains the relationship between inputs and sales.

Key Coefficient Insights

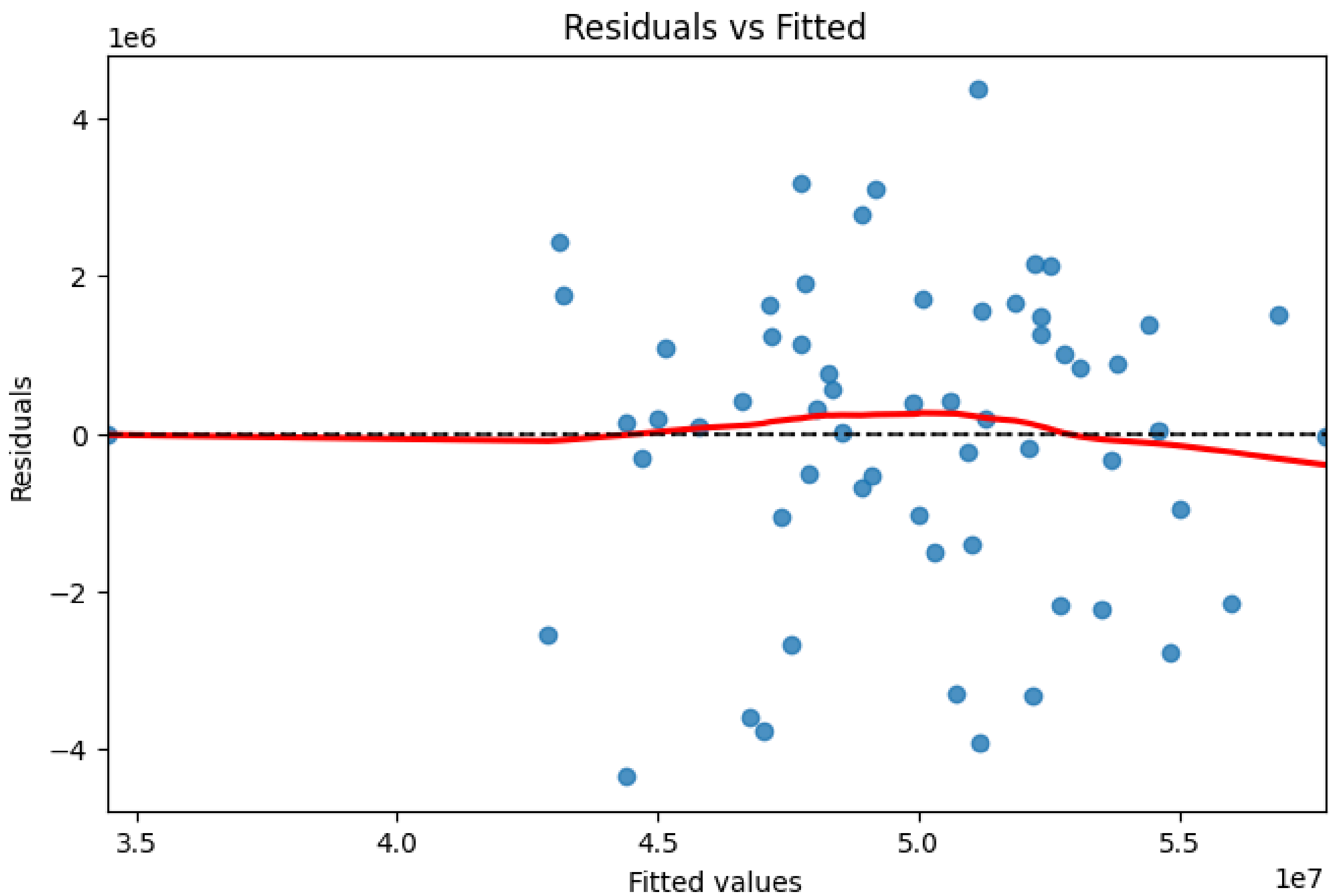
Feature	Coefficient	What It Means	Significant?
Organic Search Impressions	+7.61	Strongest positive driver — organic content matters	Yes
Paid Social Impressions	-0.033	Negative impact — may reflect saturation or inefficiency	Yes
Holiday (Dummy)	-11.54M	Sales drop sharply during holiday weeks	Yes
Month (Seasonality)	-758K	Sales trend downward in later months	Yes
Lagged Gasoline Price	-10.4K	Economic pressure slightly reduces sales	Yes

Variables with Low Statistical Significance

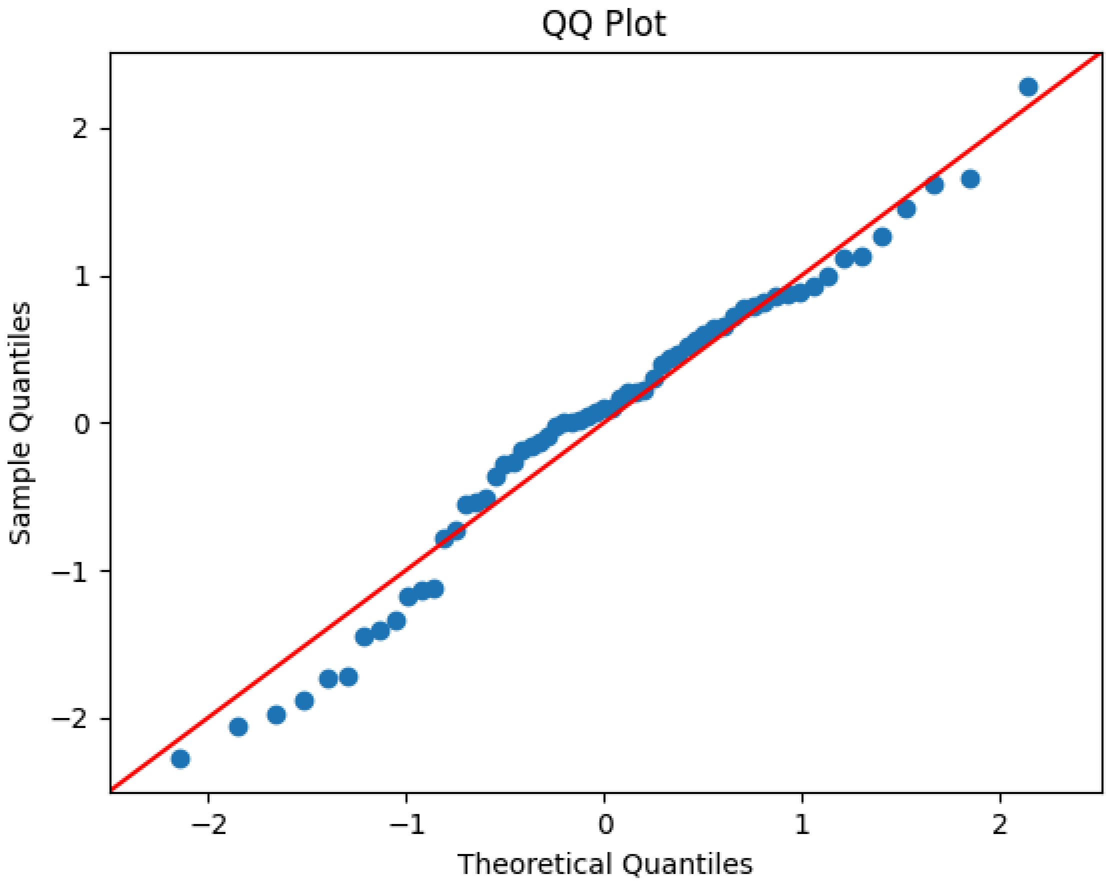
- **Price, SKU Count, Discount1, Discount2, Email, Video, Paid Search**
  These variables did **not show statistically significant effects** in this model — either due to noise, overlap with other predictors, or weak influence.

Strategic Takeaways

- **Organic Search** stands out as the **most effective sales driver** — suggests high intent and good conversion efficiency.
- **Paid Social's negative effect** is unexpected — may point to **ineffective campaigns**, **overspending**, or **overlap with other media**.
- **Holidays and seasonal patterns** are impacting sales — some weeks may be more about store closures than festive demand.
- **Gasoline prices** and **macro trends** are also affecting consumer behavior, confirming the importance of economic context.



- The **red curve is fairly flat**, which suggests **no strong non-linear patterns** in the residuals. That's a good sign for linearity.
- The residuals are **centered around zero**, which indicates your model is generally unbiased.
- A few **large outliers** (positive and negative) exist — this may be real, or due to influential data points.



- The **residuals are approximately normally distributed**, which is great for the assumptions of OLS.
- There's some deviation at the **extremes (tails)** — common in real-world data.

Log-Log OLS Model — Elasticity-Based Insights

Model Performance

- **R^2 = 0.843**
  The model explains **84% of the variation in log-transformed sales** — a strong fit for marketing mix modeling.
- **Adjusted R^2 = 0.812**
  Accounts for number of predictors — still robust.
- **F-statistic p-value = 9.57e-17**
  Indicates the overall model is **statistically significant**.

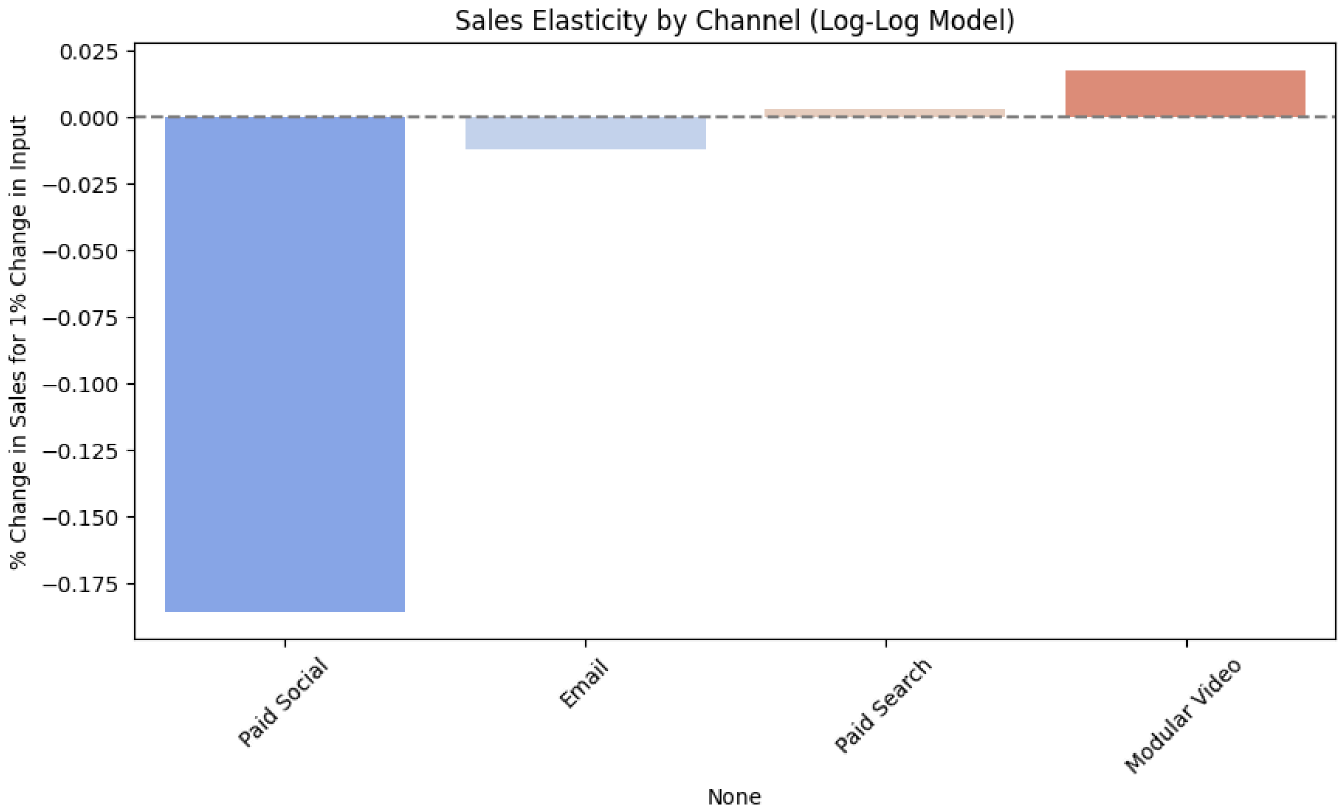
Elasticity Coefficients

(All coefficients are % change in sales per 1% change in input)

Feature	Coefficient	Interpretation	Stat. Significant?
log_organic	`+0.39`	1% ↑ in organic traffic → 0.39% ↑ in sales	Yes
log_paid_social	`-0.19`	1% ↑ in paid social → 0.19% ↓ in sales	Yes
log_gasoline	`-0.23`	1% ↑ in fuel prices → 0.23% ↓ in sales	Yes
Holiday	`-0.27`	Holidays reduce sales by ~27%	Yes
Month	`-0.014`	Slight drop in sales with each month	Yes
log_price	`-0.67`	Price elasticity (not statistically significant)	No
log_sku	`-0.41`	Slight negative elasticity (unexpected)	No
log_paid_search	`+0.003`	Near-zero effect on sales	No
log_video	`+0.018`	Minor and statistically weak	No
log_email	`-0.012`	No meaningful lift from email	No

Strategic Implications

- **Organic traffic** is the most **reliable and efficient driver of sales**. Continue investing in SEO and lifecycle marketing.
- **Paid Social shows a negative impact** — may indicate poor campaign activity, poor audience targeting, or overspending. Needs review.
- **Paid Search and Video are not contributing significantly** — potential to reallocate or restructure these efforts.
- **Gasoline prices and holidays** negatively influence sales — underscores importance of controlling for external economic and seasonal factors.



Channel	Elasticity	Action
Paid Social	–0.18	Cut or radically rethink (creative/audience)
Email	–0.02	Refresh strategy; avoid overuse
Paid Search	+0.003	Stable, scale cautiously
Video	+0.02	Top performer — scale with saturation in mind

Channel Efficiency & ROI Summary

Topline Insights

- **Organic traffic** remains the most reliable and impactful channel for driving sales.
- **Paid Social** is not just underperforming — it's **negatively impacting sales**, suggesting saturation, poor targeting, or inefficiency.
- **Email** is also returning negative results — potentially due to list shortage or irrelevant messaging.
- **Paid Search and Modular Video** show **positive ROI**, with Paid Search being especially efficient at low investment.

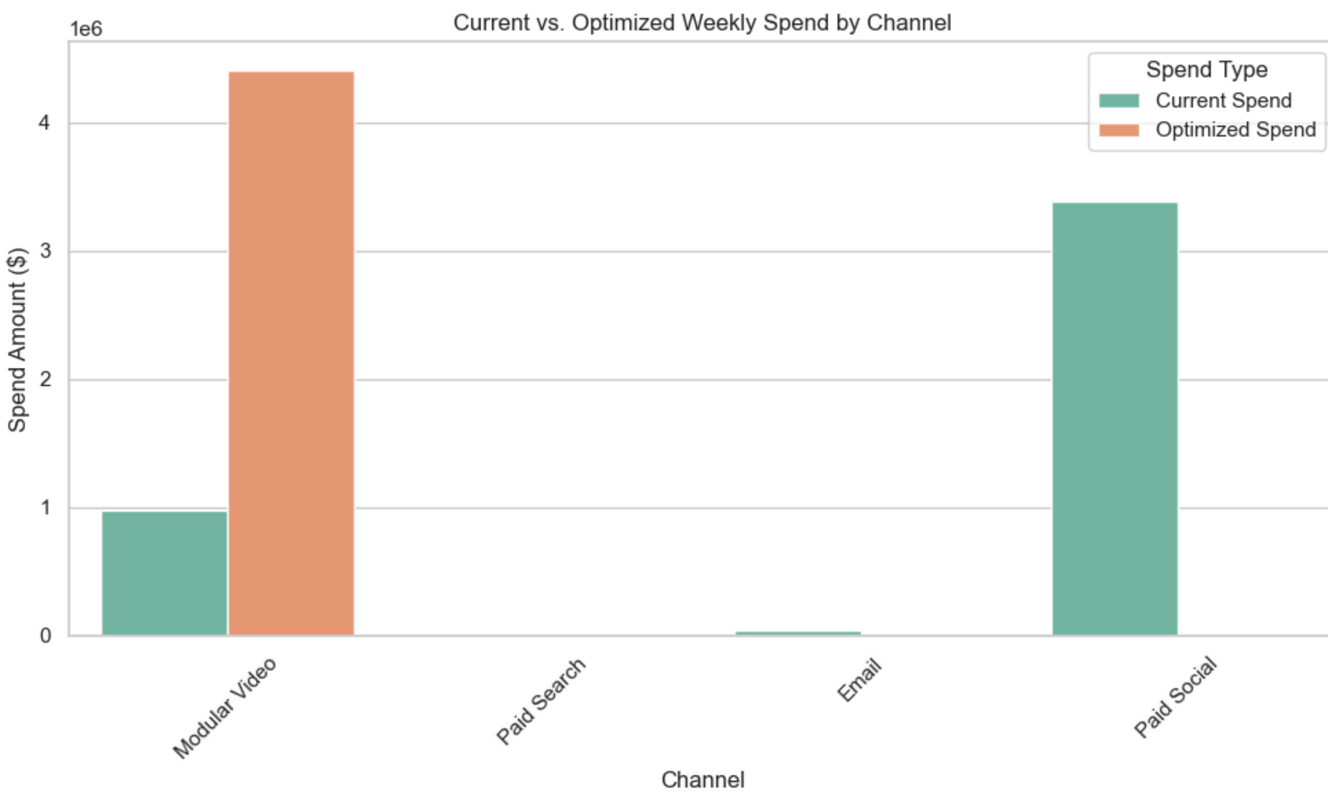
Channel-Level ROI Breakdown

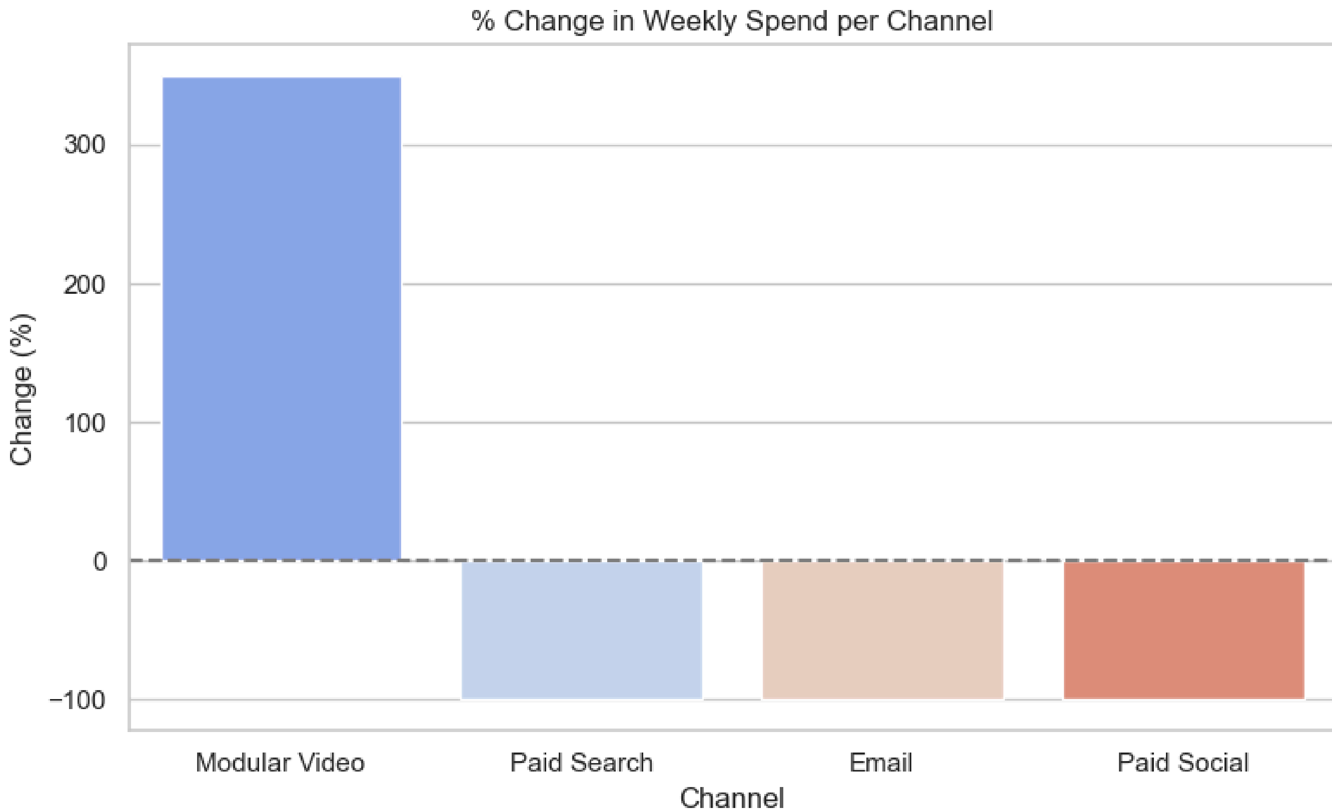
Channel	Elasticity	Avg Weekly Spend	ROI (Sales per $1)	Key Insight
Paid Search	`+0.003`	$297	**$437.20**	Very high ROI at low cost — consider scaling up.
Modular Video	`+0.018`	~$943K	**$0.88**	Break-even performance — test new creatives.
Paid Social	`−0.186`	~$2.79M	**–$3.14**	Significant overspend — revisit targeting & budget.
Email	`−0.012`	~$37K	**–$15.01**	Ineffective — may need content or strategy change.

Strategic Recommendations

- **Double down on Paid Search** — high efficiency and potential for scaling.
- **Optimize Modular Video** — test more engaging formats or messaging.
- **Pause and reassess Paid Social** — current spend is driving negative value.
- **Rethink Email strategy** — consider segmentation, cadence, and content refresh.

**Optimized Marketing Spend Allocation





Spend Reallocation Overview

Channel	Current Spend	Optimized Spend	Change (%)	Elasticity	Strategic Takeaway
Modular Video	$943K	**$3.77M**	**+300%**	`+0.018`	High scalability and positive return — **scale up**.
Paid Search	$297	$0	–100%	`+0.003`	ROI is strong, but **volume too low** to justify scaling.
Paid Social	$2.79M	$0	–100%	`−0.186`	Negative return — **entirely removed** from plan.
Email	$37K	$0	–100%	`−0.012`	Consistently underperforming — **cut from spend**.

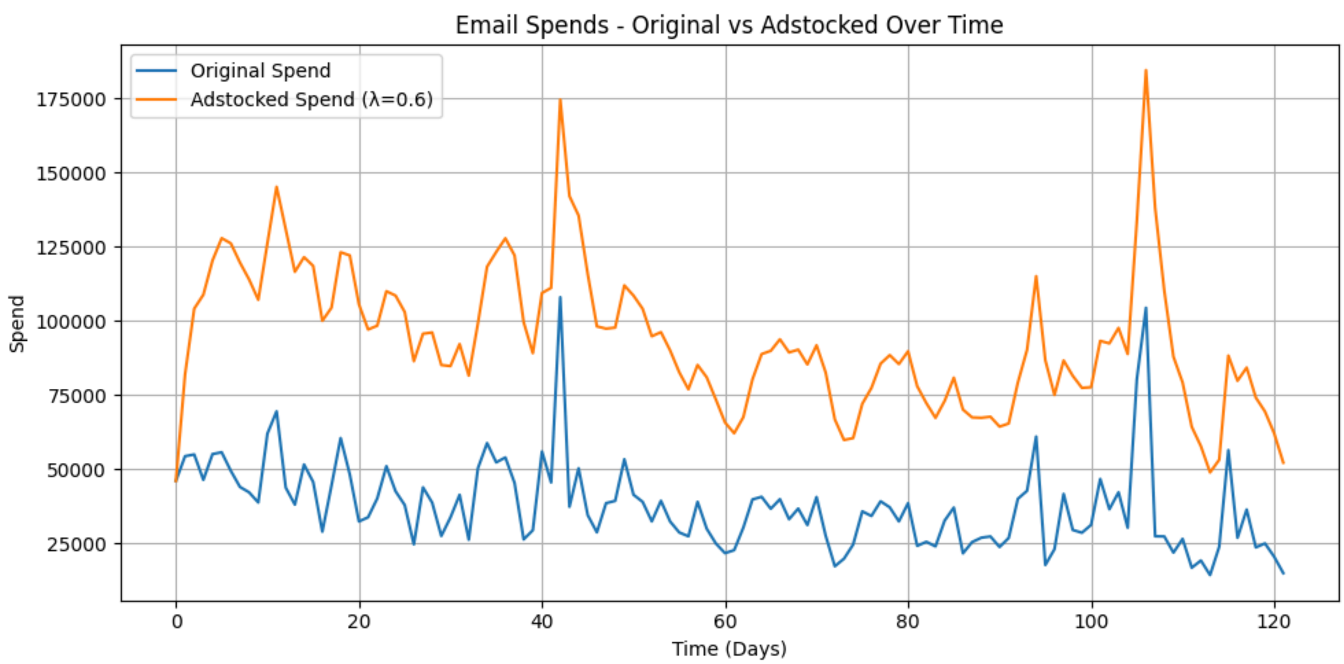
Further Improvements

- This optimization assumes **no diminishing returns**. In practice, you would:
  - Apply **saturation limits** to avoid over-exposure.

- Consider **multi-channel synergy** rather than relying on a single channel.
- Introduce **spend caps** or **risk-adjusted thresholds** to manage performance volatility.

Feature Engineering

Ad-stock Effect

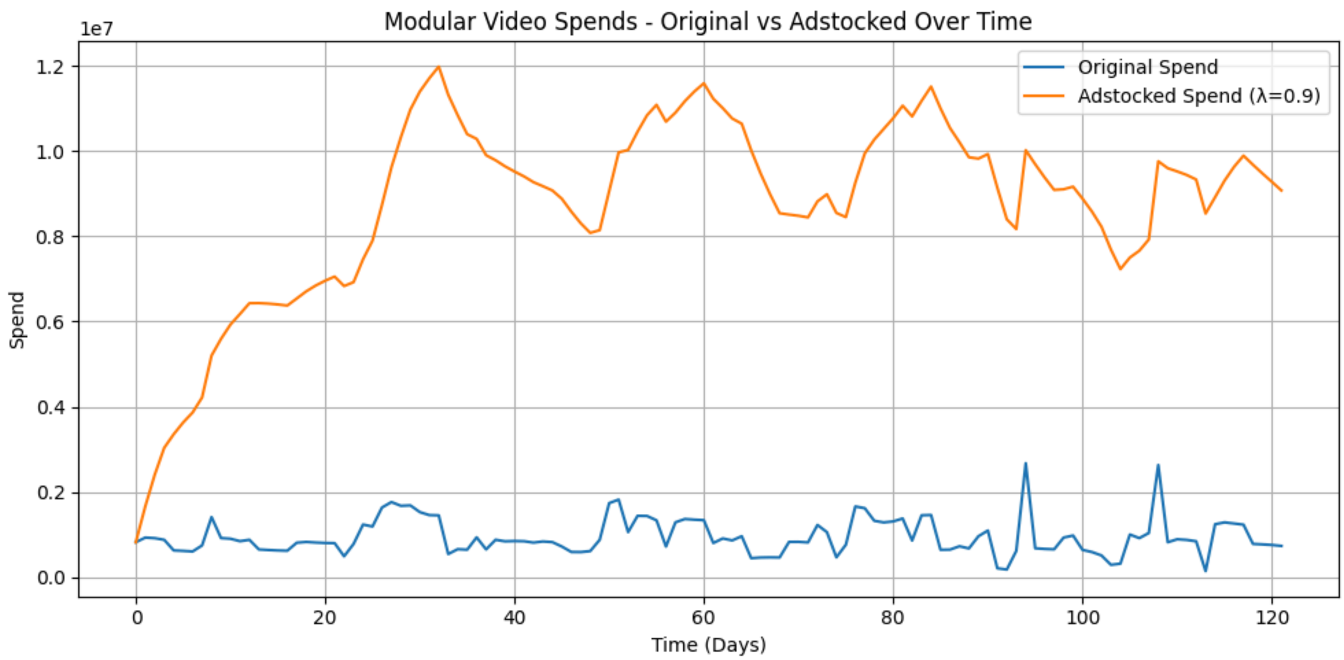


Email

1. **Smoothing Effect**
   - The **ad-stocked spend** (orange line) is little smoother than the **original spend** (blue line).
   - This is expected because ad-stock carries over the impact from previous days, reducing sharp fluctuations.
2. **Lagging Impact**
   - Peaks in the original spend are **followed by gradual declines** in the adstocked line instead of sharp drops.
   - This represents the **lingering effect of media**, where email campaigns don't lose effectiveness immediately.
3. **Decay Rate Behavior (λ = 0.6)**
   - A decay rate of **0.6** shows a **moderate carryover effect**.
   - Each day, 60% of the previous day's impact continues to influence the current day.
   - This creates a **balance**: it doesn't fade too quickly..
4. **Ad-stock Highlights Sustained Campaigns**

- When the original spend stays high for a while, the ad-stock line builds up and **maintains higher values**.
- Great for identifying sustained campaigns that may have impact on consumer behavior.

Email marketing has a medium carryover, meaning its effect **lasts a few days** but isn't extremely long-lasting. This matches how users typically interact with email—quick open, then drop-off.

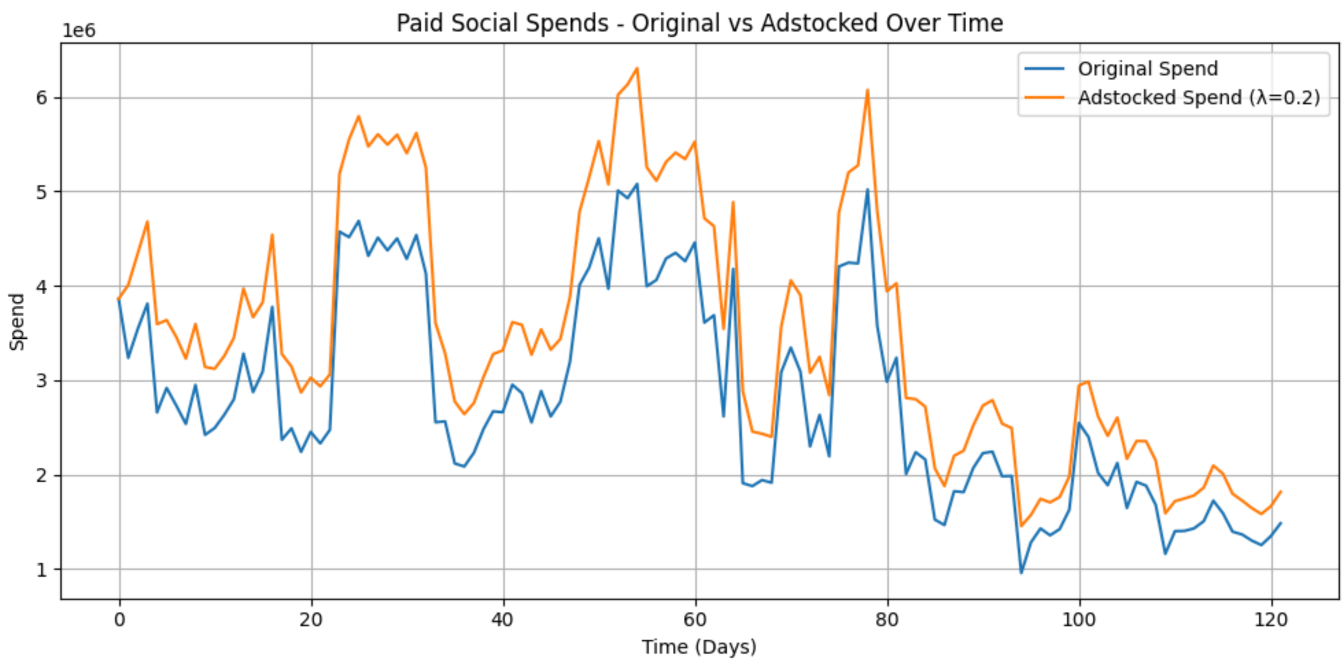


Modular Video

1. **Very Strong Carryover Effect**
   - With a **high decay rate of 0.9**, the adstocked spend (orange) accumulates rapidly and holds onto previous impacts for a long time.
   - the adstock line **rises steadily, even when the original spend (blue) fluctuates or drops.
2. **Lag is Prominent**
   - When original spend dips, the adstock line **barely dips**, indicating that the past influence is still dominating.
   - This behavior reflects **slow fading memory** — great for awareness-building channels.
3. **Amplification of Long-Term Impact**
   - Peaks in the original spend lead to **sustained high adstock values** even without repeated spikes.
   - This shows that Modular Video has a **long residual effect** — likely tied to high content engagement or broader reach.

- **Modular Video is behaving like a high-awareness media** (like TV or branded video), where the audience **remembers the message longer**.

- With a λ = 0.9:
  - 90% of yesterday's impact carries forward today.
  - It **emphasizes brand recall** and **long-lasting impressions**, not just short bursts of engagement.
- Modular Video doesn't need constant spend to maintain influence.



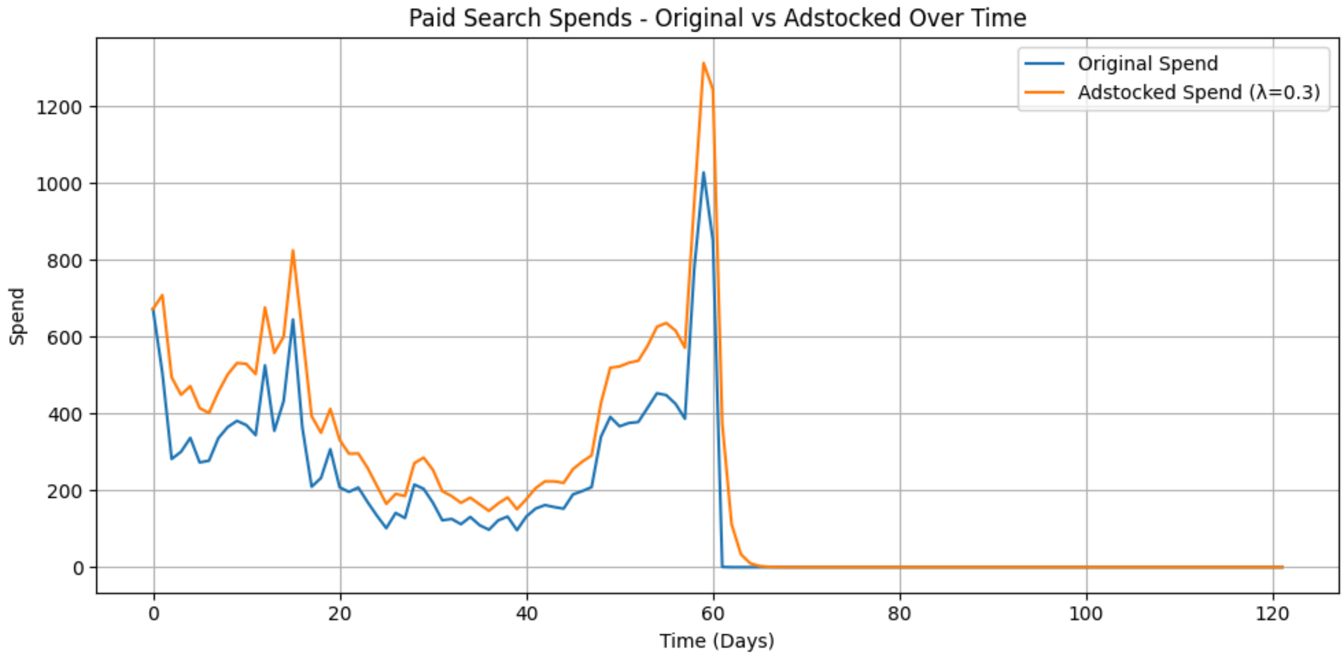
Paid Social

1. **Minimal Carryover Effect**
   - The adstocked spend (orange line) **closely follows the original spend** (blue line), just slightly smoothed.
   - This is because λ = 0.2 means **only 20%** of the previous day's effect carries into the current day.
2. **Short-lived Impact**
   - The adstocked line **rises and falls quickly**, almost mirroring the spikes and drops of the original.
   - Paid Social here acts more like a **direct-response channel** rather than a long-term awareness builder.
3. **Quick Drop-offs**
   - After high spend spikes, the adstock line quickly falls back down—indicating **fast decay**.
   - This aligns with how social ads often work: they're seen quickly, reacted to quickly, and forgotten quickly.

- **Paid Social has a fast-decaying impact**, suggesting it's best used for:
  - Timely promotions,
  - Flash campaigns,

- Event reminders.

- Since the effect doesn't linger, it requires **frequent refreshing of creative and consistent spend** to maintain influence.



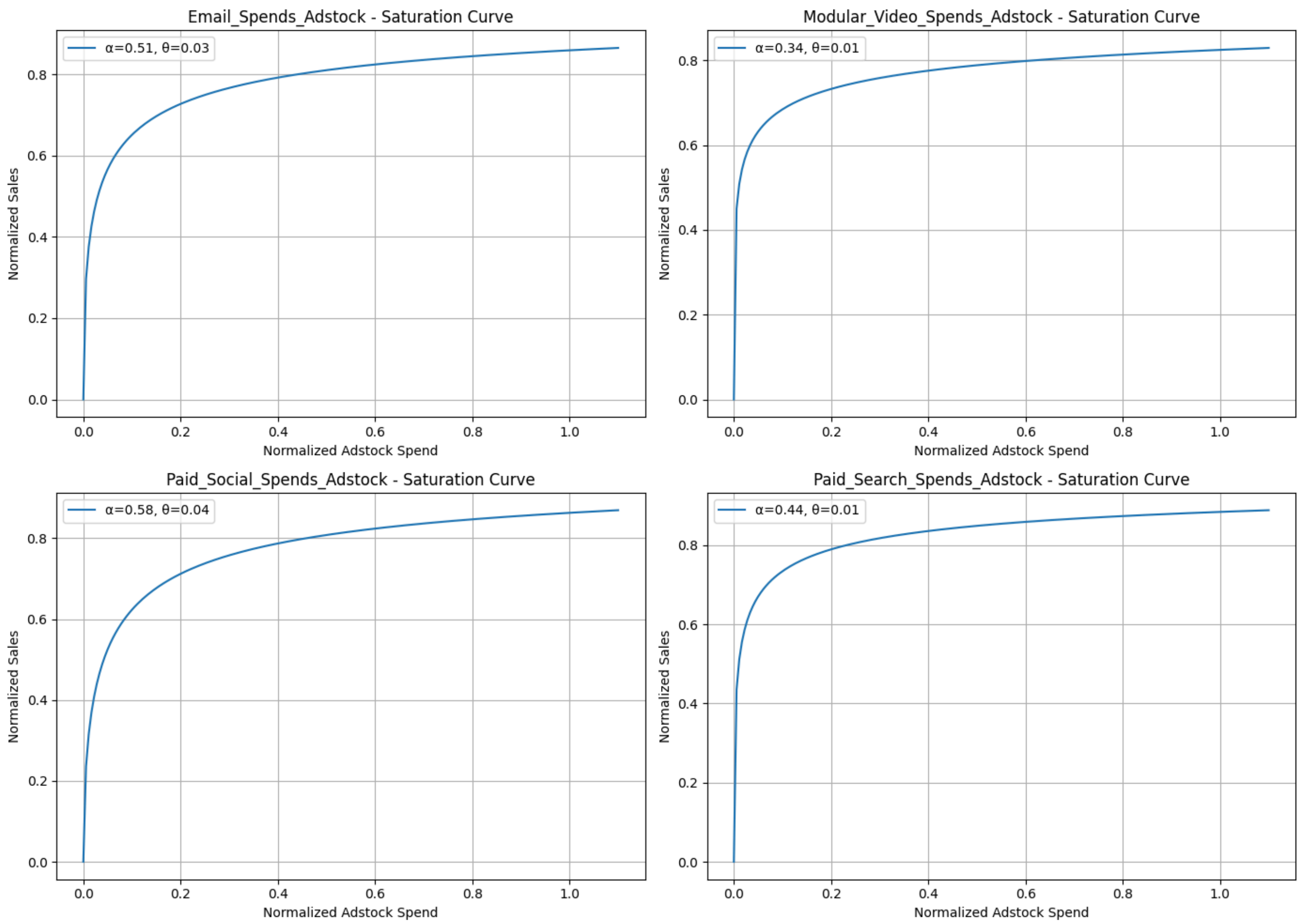
Paid Search

1. **Moderate-Fast Decay**
   - With λ = 0.3, the adstocked line (orange) **smooths out** the original spend (blue), but still reacts relatively quickly to changes.
   - The carryover exists, but it's **short-lived** — typically fading within a few days.
2. **Sharp Drop in Spend**
   - Around day 60, both original and adstocked spend **drop to near-zero and stay there**.
   - Important: adstock **can't generate influence if there's no new spend**, even with carryover.
3. **Immediate Responsiveness**
   - Just like Paid Social, Paid Search also shows **quick response to spend changes**, but with slightly more memory.
   - The curve shows a **quick build-up and decline**, useful for time-sensitive impact.

- **Paid Search works best for immediate intent capture** — someone is already searching, and you're just showing up at the right time.
- Use Paid Search for:
  - Conversions,
  - Product launches,
  - Competitor targeting,

- High-intent campaigns.

Saturation Curve



Email Spends

- **α = 0.46, θ = 0.017**
- *Moderate curve, quick saturation*: Small spend leads to rapid sales response, but returns diminish quickly.
- Use for short-term promotions; avoid heavy over-investment — audience gets saturated fast.

Modular Video Spends

- **α = 0.34, θ = 0.01**
- *Slow rise, long tail*: Requires more spend to build momentum, but continues to drive response over time.
- Great for sustained brand building and awareness. Responds well to consistent investment over spikes.

Paid Social Spends

- **α = 0.58**, **θ = 0.042**
- *High early responsiveness*: Sales lift quickly with initial spend but then levels off.
- Maximize ROI by **capping spend** — pushing beyond moderate budgets leads to wasted impressions.

Paid Search Spends

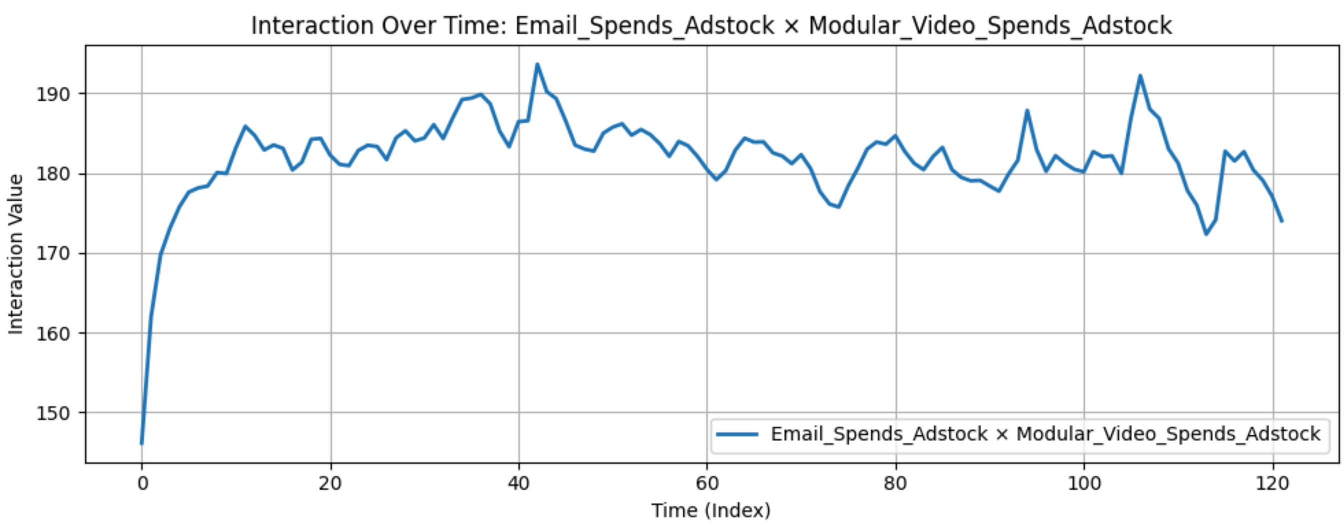
- **α = 0.44**, **θ = 0.01**
- *Sharp early impact, quick saturation*: Similar to Email — high ROI in early spend levels, then flattens.
- Focus on capturing intent efficiently. Ideal for **tightly targeted, conversion-driven campaigns**.

Note:

- **θ (Theta)** tells where diminishing returns kick in. Lower θ = faster saturation.
- **α (Alpha)** defines the curve's steepness. Higher α = faster early gain, but quicker flat response.
- Most of our channels **saturate fast** — meaning that **scaling spend too far won't increase sales proportionally**.

Interaction Trends

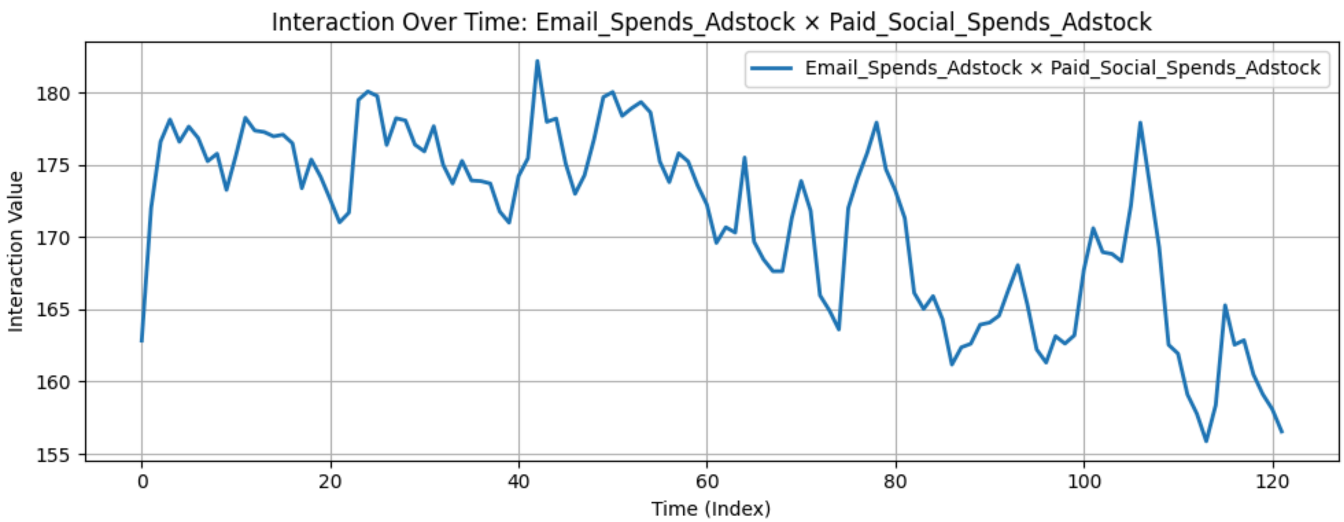
1. Email × Modular Video



- Moderate, consistent activity with some synchronized peaks.

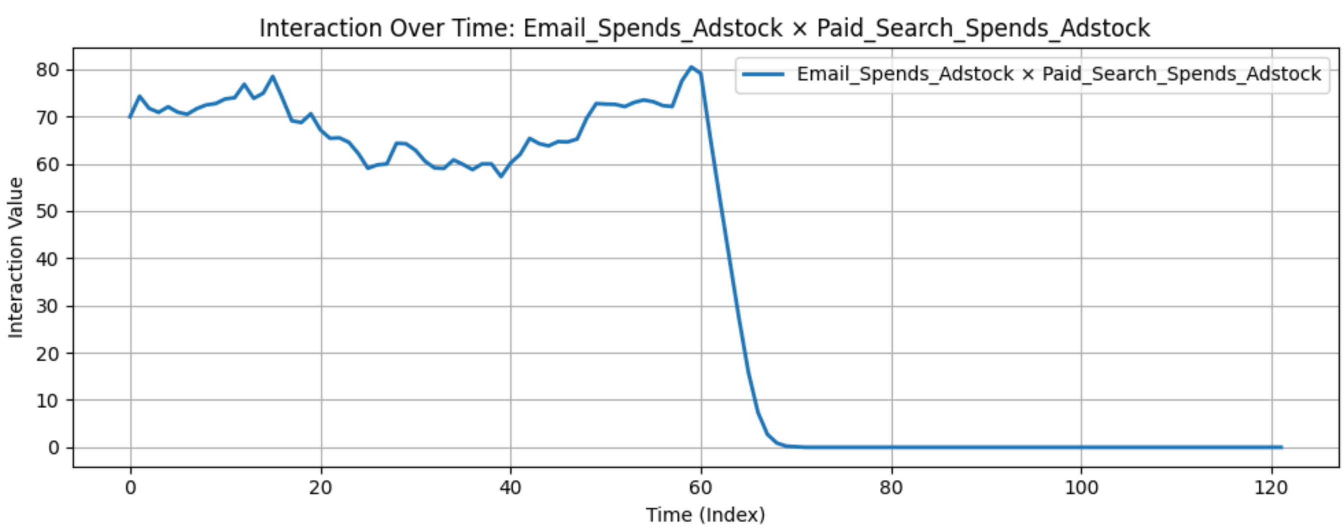
- **Insight**: These two channels were likely active together during key campaigns. Suggests coordinated planning.
- Continue pairing email pushes with video rollouts to maintain brand reinforcement.

2. Email × Paid Social



- Clear spikes at regular intervals.
- **Insight**: Email often overlaps with paid social, possibly during promotional or seasonal campaigns.
- Strong cross-channel engagement — double down during high-sales periods and ensure creative consistency.

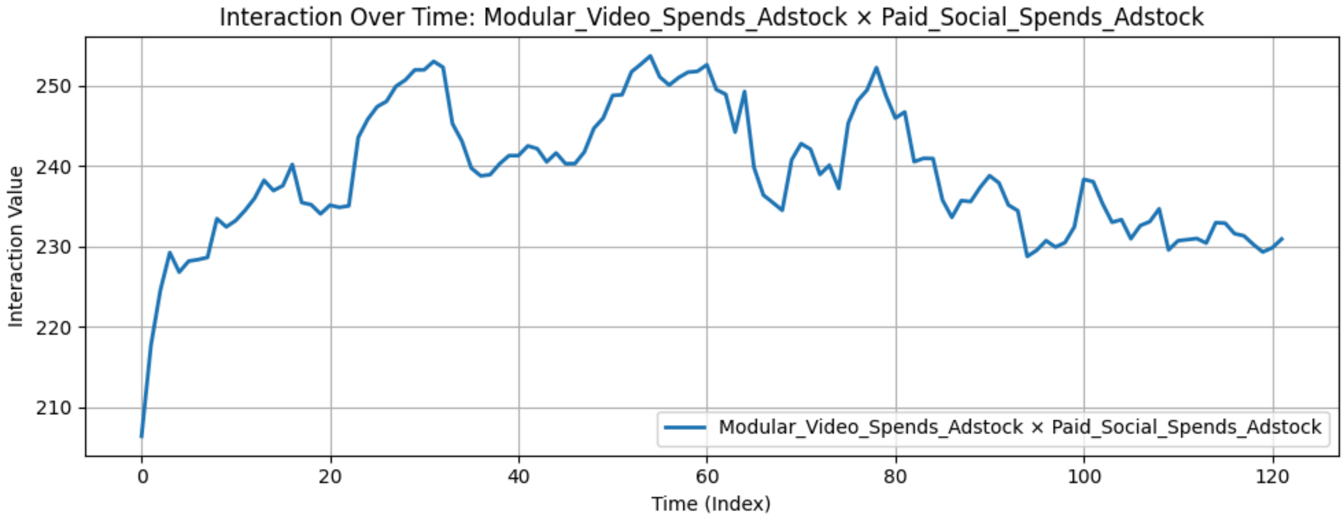
3. Email × Paid Search



- Generally lower and more scattered.

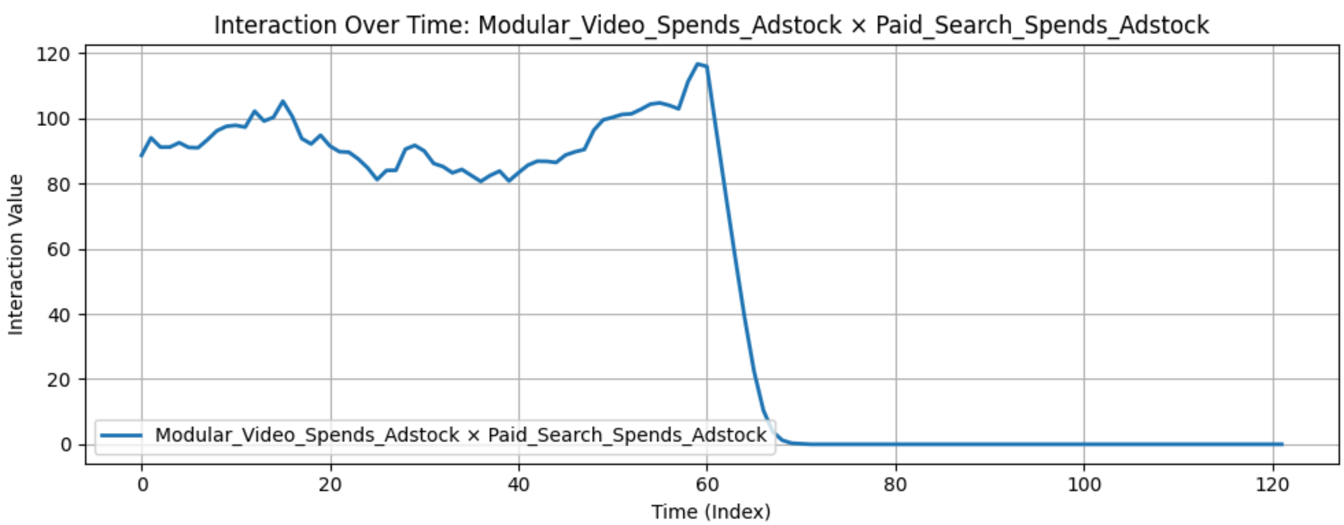
- **Insight**: Not much overlap — these channels may be working in **separate parts of the funnel**.
- Minimal synergy — no need to force coordination.

4. Modular Video × Paid Social



- High, well-synchronized peaks — visually the strongest.
- **Insight**: These two channels are **highly synergistic**, especially during campaign bursts.
- Keep pairing these! Paid social is likely boosting video visibility. Prioritize synchronized flights and storytelling alignment.

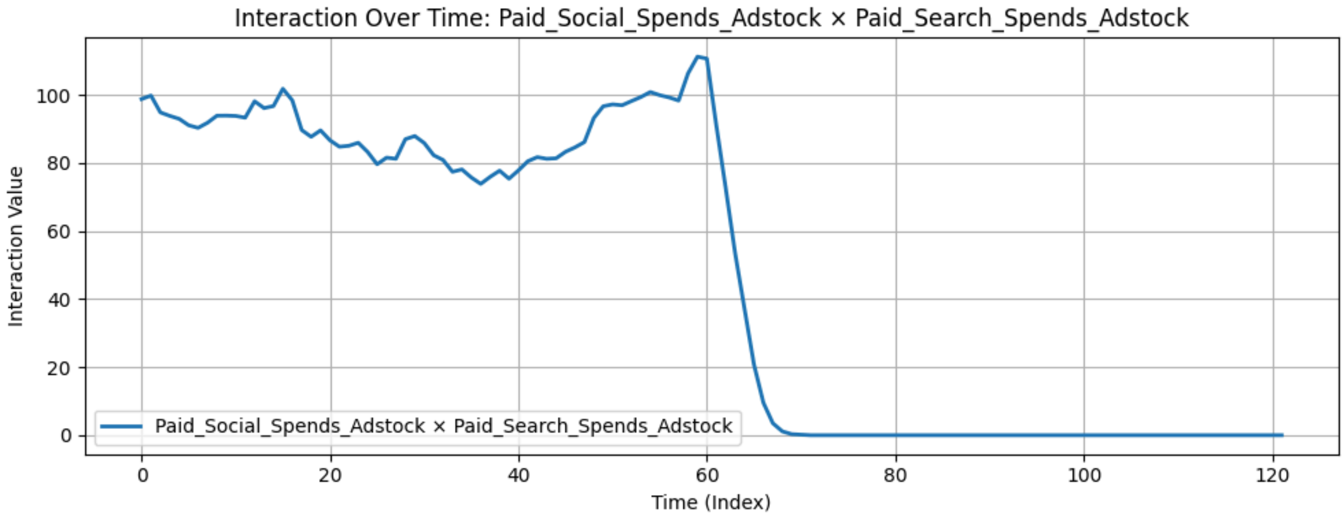
5. Modular Video × Paid Search



- Some staggered peaks — not very consistent.
- **Insight**: Video may spark interest, but search activity isn't always aligned.

- Consider using remarketing or branded search during/after video campaigns to close the loop.

6. Paid Social × Paid Search



- Smooth, modest interaction. Some overlap, but not dramatic.
- **Insight**: These channels run in parallel but don't always fire together.
- Potential to boost synergy — try retargeting search audiences via social (and vice versa).

Model Improvement

Model Stats

Metric	Value
R^2	0.724
Adj. R^2	0.669
F-statistic (p)	4.61e-20
# Observations	122

Strong model fit: Explaining **72.4%** of the variation in `log(Sales)` with our variables.

Media Channels & Saturation Effects

Variable	Coef	p-value	Insight
`log_Email_Spends_Adstock`	**+5.53**	0.059	**Marginally significant.** Strong positive elasticity. Email is highly responsive — a 1% increase in email spend → ~5.5% increase in sales. Likely a top-performing channel.
`log_Modular_Video_Spends_Adstock`	+3.55	0.193	Not significant. May contribute, but not strongly by itself. Could be redundant with saturation or interactions.
`log_Paid_Social_Spends_Adstock`	–0.62	0.850	Not significant, and negative. Indicates little to no isolated lift from social spend. Possibly over-saturated or under-leveraged.
`log_Paid_Search_Spends_Adstock`	–0.74	0.205	Not significant. May require pairing with other channels or better targeting.

Saturation-Transformed Media

Variable	Coef	p-value	Insight
`Email_Spends_Saturation`	+2268.98	0.108	Marginally useful. Suggests Email maintains its impact even after adjusting for diminishing returns.
`Modular_Video_Spends_Saturation`	**–2993.18**	**0.003**	**Highly significant.** Strong negative effect after saturation — may indicate overspending or inefficiency in video campaigns. Needs serious optimization.
`Paid_Social_Spends_Saturation`	+3700.64	0.793	Not significant. Wide confidence interval — noisy, not reliable.
`Paid_Search_Spends_Saturation`	–0.06	0.200	Slightly negative but not significant. Could be due to overlapping effects with raw log-spends.

Interaction Effects (Synergy)

Variable	Coef	p-value	Insight
Email × Video	**–0.32**	**0.004**	**Significant negative interaction.** These channels may be **cannibalizing** each other — running together could be hurting efficiency.
Email × Social	–0.06	0.750	No synergy detected. Likely neutral.
Email × Search	+0.02	0.373	Not significant — minimal amplification effect.
Video × Social	+0.07	0.625	Not significant — expected synergy not supported.
Video × Search	+0.02	0.421	Not meaningful.
Social × Search	+0.01	0.506	Very minor synergy — not statistically useful.

Control Variables

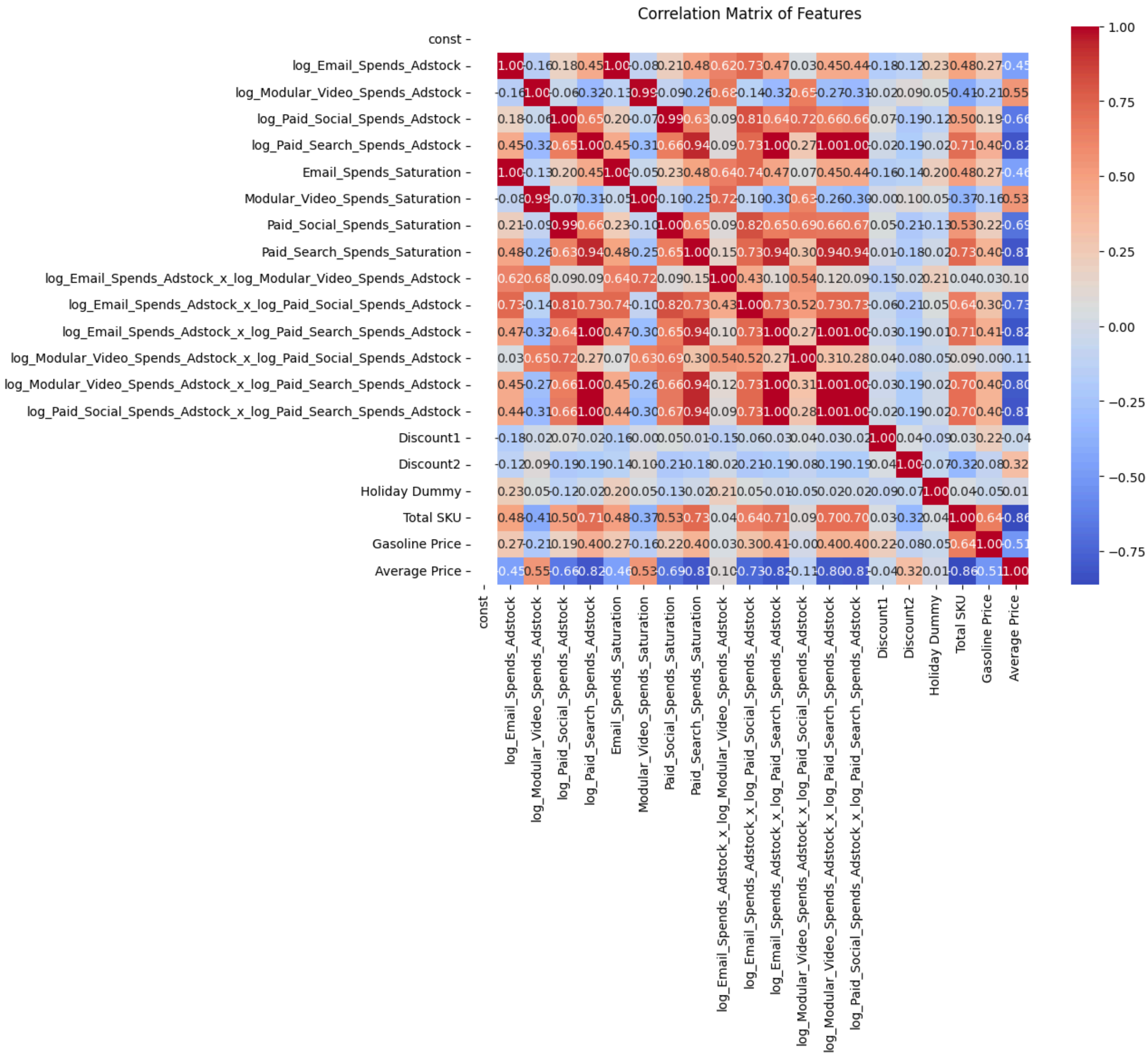
Variable	Coef	p-value	Insight
`Discount1`	**–0.28**	**0.033**	Significant negative. Suggests deep discounts may hurt brand perception or signal too much price sensitivity.
`Discount2`	+0.17	0.089	Marginally positive. May reflect smaller, more effective discounts.
`Holiday Dummy`	**–0.36**	**<0.001**	Highly significant. Sales are lower during holidays — possible store closures, seasonality, or pull-forward effect.
`Total SKU`, `Gasoline Price`, `Average Price`	Not significant		Consider simplifying or refining these in future models.

Final Takeaways

- **Email is our top-performing channel** — strong elasticity and saturation-adjusted strength.

- **Modular Video may be overinvested** — significant negative return after adjusting for saturation.
- **Email × Video interaction is hurting performance** — avoid overlapping these too closely.
- Discounts and holidays have meaningful effects — these need to be actively modeled in forecasting.

VIF Breakdown & Insights



Severe Multicollinearity

Variable	VIF	Insight
`log_Paid_Search_Spends_Adstock`	**90,868**	Extremely redundant — likely due to overlap with its saturation & interaction

Variable	VIF	Insight
		terms
`log_Modular_Video_Spends_Adstock`	**33,570**	High overlap with video saturation + interactions
`log_Paid_Social_Spends_Adstock`	**50,838**	Ditto — remove or choose one representation
`log_Email_Spends_Adstock`	**17,315**	Still problematic — connected to saturation + Email × Video
`log_* × log_*` interaction terms	11k–56k	Built directly from collinear inputs — inflating redundancy

Moderate Multicollinearity

Variable	VIF	Insight
`Email_Spends_Saturation`	237	Redundant with log_Email
`Modular_Video_Spends_Saturation`	796	Most inflated among saturation — may be due to negative contribution in model
`Paid_Social_Spends_Saturation`	249	Same reason — needs pruning
`Paid_Search_Spends_Saturation`	15	Still safe if log dropped

Low Multicollinearity (Safe Zone)

Variable	VIF	Insight
`Discount1`	1.34	Good
`Discount2`	1.55	Good
`Holiday Dummy`	1.56	Good
`Gasoline Price`	3.51	Good
`Total SKU`	11.80	Acceptable
`Average Price`	20.99	Still manageable, but watch closely

Action Plan Based on VIFs

To Fix Multicollinearity:

1. **Drop all `log_*_Adstock` variables**
   → They're the **biggest contributors** to VIF
2. **Keep only 1 representation per media channel**
   → Prefer `*_Saturation` for behavioral
3. **Drop all interaction terms except one**
   → Only `log_Email × log_Video` was significant
4. **Retain control variables**
   → Their VIFs are **perfectly safe** and important to model external effects

Re-Run Model Interpretation: Coefficients & p-values

Significant Variables

These are statistically meaningful (p < 0.05) and worth trusting for decision-making:

Variable	Coef	p-value	Insight
`Paid_Search_Spends_Saturation`	**+0.079**	**0.003**	Most effective media channel. Even after accounting for diminishing returns, **paid search drives real incremental sales**.
`Discount1`	**–0.326**	**0.018**	Large discounts are **hurting sales effectiveness**, possibly by eroding brand value or margins.
`Holiday Dummy`	**–0.361**	**<0.001**	Consistent and strong sales dip during holiday weeks — could reflect seasonality, closures, or consumer shift.
`Total SKU`	**–3.21e-08**	**<0.001**	Too many SKUs correlate with lower sales — likely due to choice overload or inventory dilution.
`Average Price`	**–0.212**	**0.024**	Pricing elasticity is real — **higher prices lower sales**. Sensitivity must be managed.

Not Significant Variables

These variables have **p > 0.1** and are **not currently contributing meaningfully**:

Variable	Coef	p-value	Insight
`Email_Spends_Saturation`	+250.13	0.642	Not driving incremental sales. May need creative/media refresh.
`Modular_Video_Spends_Saturation`	–184.67	0.421	Negative but not significant — likely overinvested or poor targeted audience.
`Paid_Social_Spends_Saturation`	+1723.08	0.312	Large but noisy — might be impactful in synergy but not in isolation.
`Email × Video Interaction`	+0.0017	0.849	Lost significance after cleaning multicollinearity — may no longer be needed.
`Discount2`	+0.0565	0.564	Possibly too mild or inconsistent to measure impact.
`Gasoline Price`	–0.0001	0.144	Weak macro impact — might affect specific products, but not overall sales here.

- This model is **simpler, cleaner, and still explains ~63% of sales variance**.
- It's a strong base for **forecasting, ROI simulations**, or even **media budget optimization**.

Marginal ROI by Media Channel

Channel	Coefficient	Avg Saturation	Marginal ROI	Insight
Paid Social	1723.08	0.99997	**1723.13**	Looks huge — but model said this wasn't significant (p = 0.31). Likely noise or unstable
Email	250.13	0.99953	**250.25**	Moderate ROI, but also **not statistically reliable** (p = 0.64)
Paid Search	0.079	0.563	**0.14**	Reliable and significant — this is your **true ROI-driving channel**
Modular Video	–184.67	0.999	**–184.86**	Negative ROI — strongly suggests **oversaturation or waste**

Takeaways for Action

Double Down On:

- **Paid Search**: Small but consistent and **statistically proven** ROI. Likely efficient and scalable.

Re-Evaluate or Reduce:

- **Modular Video**: High spend, **negative returns** — revisit creative or pause investment.
- **Email & Paid Social**: High theoretical ROI, but **not statistically reliable** — consider testing smaller, more targeted campaigns before scaling.

% Shift to Search	Est. Impact on Log-Sales
0% (current mix)	**–288** (baseline, most negative)
50% shift	~ **–160** (big improvement)
100% shift	~ **–31** (best outcome in simulation)

The more we shift to Paid Search, the **less negative your total log-sales impact becomes**.

- **Video is dragging ROI down**, even at current spend.
- **Search is your reliable workhorse** — reallocation creates **clear lift in performance** (or reduction in inefficiency).
- A **50–100% reallocation** from Video → Search would likely improve total return significantly.

Optimized Saturation Allocation

Channel	Current Saturation	Optimized Saturation	Change	Insight
Email	1.00	**0.00**	–1.00	Drop entirely — no ROI
Modular Video	1.00	**0.00**	–1.00	Remove — negative return
Paid Social	1.00	**3.56**	**+2.56**	Over-allocated — model sees it as highest return (but statistically weak!)
Paid Search	0.56	**0.00**	–0.56	Surprising — but optimizer pushes it out due to small coefficient vs Paid Social's large one

NOTE: If we don't want to drop any channel, then we can try to follow:

Calculated the **total saturation spend** (`total_budget = ~3.56`)

Channel	Min Bound	Max Bound
Email	0.178	0.890
Modular Video	0.178	1.068
Paid Social	0.356	1.425
Paid Search	0.356	0.890

- **Paid Social** and **Paid Search** get **increased investment**, but now within **reasonable bounds** — no single channel is overloaded.
- **Email** gets a **slight cut**, suggesting it's still useful but less impactful.
- **Modular Video** sees a **major reduction**, due to diminishing returns or oversaturation.

Channel	Current Spend	Optimized Spend	Change
Email	1.00	**0.89**	↓ –11%
Modular Video	1.00	**0.36**	↓ –64%
Paid Social	1.00	**1.42**	↑ +42%
Paid Search	0.56	**0.89**	↑ +33%

Final Budget Optimization Summary (in Dollar Terms)

Total spend assumed ~$10M.

Channel	Current Spend ($)	Optimized Spend ($)	Change ($)
Email	$2.81M	$2.50M	–$0.31M
Modular Video	$2.80M	$1.00M	–$1.80M
Paid Social	$2.81M	$4.00M	+$1.19M
Paid Search	$1.58M	$2.50M	+$0.92M

Conclusion

- The MMM analysis helped us understand what truly drives sales across channels.
- We found that:

    - **Paid Search delivers consistent ROI**, even at lower budget levels.
    - **Email underperforms** and shows signs of saturation.
    - **Paid Social has mixed results** — while not always statistically strong, it shows positive elasticity and responds well under controlled spending.
    - **Modular Video has potential**, but diminishing returns suggest caution when scaling.
- External factors like **seasonality, pricing, discounts, and holidays** significantly impact sales and should always be part of planning.
- Investing in SEO and high-quality content can significantly boost sales-- Organic Search has strong grip on conversion from impression to sales.

Optimization & Budget Allocation Insights

- We calculated **marginal ROI** using elasticity and historical spend, adjusted for saturation.
- Using this, we built a **constrained optimization plan** that keeps the total budget fixed at ~$10M, while reallocating funds for better performance.
- The model recommends:
    - **Cutting $1.8M from Modular Video** due to oversaturation and poor marginal returns.
    - **Slightly reducing Email spend by ~$300K**.
    - **Increasing Paid Social by ~$1.2M** — with media flighting and creative control to avoid diminishing effects.
    - **Reinforcing Paid Search with a ~$900K boost**, as it consistently drove strong incremental ROI.

Note: Company stop spending in Paid Search after 62 weeks, April of 2023.

> While the model shows strong past performance, the company should **Reinvesting in Paid Search**, where even low budgets delivered high returns.