Haensel AMS - Data Science Challenge

Shayekh Mohiddin Ahmed Navid

shayekhnavid@gmail.com

1 Introduction

We present a Bayesian Media Mix Modeling (MMM) analysis conducted as part of the Haensel AMS Data Science Challenge, with the objective of quantifying the impact of weekly marketing spend across seven distinct channels on overall revenue, and estimating the return on investment (ROI) for each channel. Given that marketing channels often exhibit delayed and decaying effects on revenue, a Bayesian framework is well-suited for this analysis, as it naturally accommodates temporal dependencies and uncertainty in parameter estimates.

The primary objectives of this study are as follows:

- To model weekly log-transformed revenue as a function of adstocked marketing spend across selected channels.
- To estimate posterior distributions of channel effects using PyMC.
- To conduct posterior predictive checks for evaluating model fit and reliability.
- To compute channel-level ROI estimates and compare alternative model specifications.

This probabilistic approach enables robust, interpretable inference and supports data-driven decision-making for optimizing marketing budget allocation under uncertainty.

2 Exploratory Data Analysis

EDA revealed key patterns:

2.1 Spend vs Revenue:

Channels 3, 4, and 7 showed positive associations. Channel 7 appeared immediately; 3 and 4 had lagged effects.

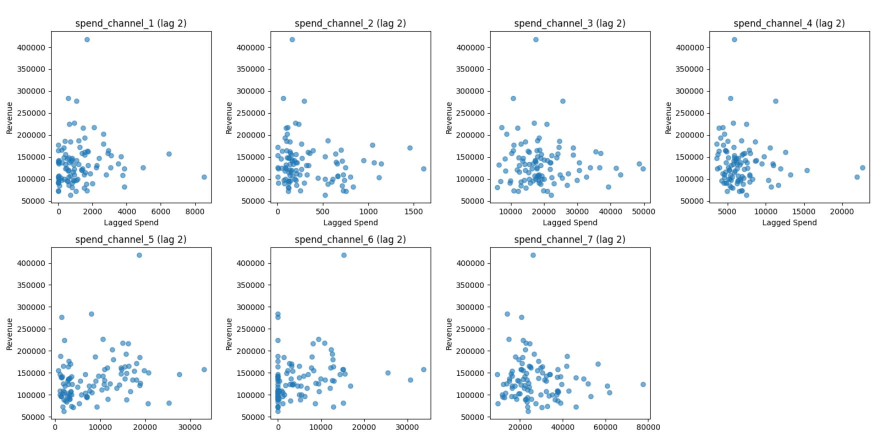

Fig. 1. Spend vs Revenue(Lag 2)

2.2 Lag/Adstock Effects:

Lag plots and exponential adstock transformations (with decay θ) improved correlation for Channels 3–5. Channel 7's effect was short-lived; Channels 1, 2, and 6 showed weak patterns.

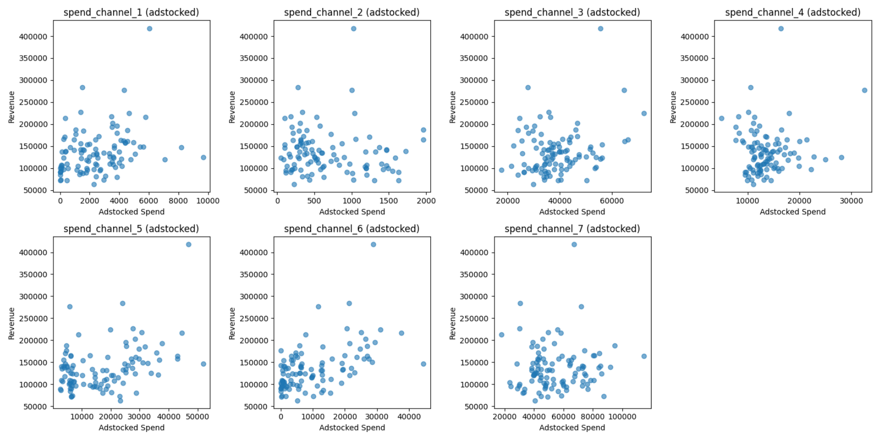

Fig. 2. Adstocked vs Revenue

2.3 Skewness:

Revenue and most spending were right-skewed—log transformation was applied.

2.4 Correlations:

Revenue was most correlated with Channels 6 (0.52), 5 (0.42), and 3 (0.38). Channel 5 and 6 were collinear (0.65), requiring caution.

2.5 Key Findings Summary:

The analysis identified Channels 3, 4, and 5 as strong candidates due to their lagged effects, effectively captured through adstock transformations. Channel 7 showed immediate, short-term impact, while Channels 1, 2, and 6 displayed weak or noisy relationships with revenue. The right-skewed distribution of revenue and spend also justified applying a log-transformation to stabilize variance and enhance model fit.

Table 1. Key EDA Takeaways by Channel

Channels	**Insights**
3–5	Strong candidates with adstocked spend
7	Immediate (short-term) impact observed
1, 2, 6	Weak or noisy relationship with revenue
–	Log-transformation justified by skewness

3 Modeling Approach

We used a Bayesian linear model on log-revenue:

$$y = \log(\text{revenue}) \sim \mathcal{N}(\mu, \sigma), \quad \mu = \alpha + \sum \beta_i \cdot \text{Adstock}_i$$

3.1 Modeling Spend Carryover

We used an exponential **adstock transformation** to model delayed marketing effects:

$$\text{Adstock}_t = x_t + \theta x_{t-1} + \theta^2 x_{t-2} + \dots$$

This accounts for diminishing influence of past spend. Decay parameters (θ) were set by visual lag analysis: $\theta_3=0.5$, $\theta_4=0.4$, $\theta_5=0.3$, $\theta_7=0.2$.

3.2 Prior Assumptions and Inference

We used weakly informative priors to regularize learning without imposing strong beliefs:

$$\beta_i \sim \mathcal{N}(0,1), \quad \alpha \sim \mathcal{N}(0,5), \quad \sigma \sim \text{HalfNormal}(1)$$

Priors allow flexibility while centering beliefs near zero for effect sizes. We ran posterior inference using the No-U-Turn Sampler (NUTS), drawing 1,000 samples after 1,000 tuning steps.

Prior vs Posterior: The prior provides initial uncertainty, but model results are driven by the posterior, which concentrates around values supported by the observed data. Trace plots showed good mixing and convergence, indicating reliable posterior estimates.

4 Results, ROI Analysis, and Model Comparison

The Bayesian model's posterior estimates revealed that Channel 5 had the strongest and most consistent positive effect on revenue, followed by Channels 3 and 4, which contributed modestly. In contrast, Channel 7, despite its relatively high spend, exhibited minimal effect with high uncertainty, suggesting inconsistent impact.

Table 2. Posterior Summary of Channel Effects

Channel	Effect Strength	Uncertainty
3	Weak but consistent	Low
4	Modest	Moderate
5	Strongest	Low
7	Minimal	High

Predictive Check: Posterior predictive checks showed that the model effectively tracked log-revenue trends over time. Although some high-revenue spikes were slightly underpredicted due to the smoothing nature of Bayesian inference, the overall model fit was strong. Trace plots and posterior densities confirmed good convergence, with well-mixed chains and no divergences.

To evaluate cost-effectiveness, we computed Return on Investment (ROI) for each channel as the ratio of estimated revenue contribution to total spend, after adjusting for log-transformed revenue:

$$\text{ROI}_i = \frac{\text{Estimated Contribution}_i}{\text{Spend}_i}$$

Table 3. ROI Estimates by Channel

Channel	ROI
3	0.23
4	0.96
5	**2.26**
7	0.82

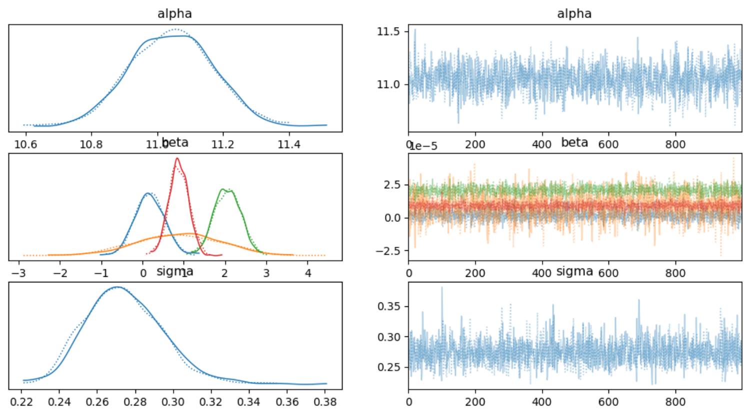

Fig. 3. Posterior distributions and trace plots for model parameters (α, β, σ) showing good convergence and well-behaved sampling chains.

ROI analysis showed Channel 5 as the most efficient (ROI = 2.26), Channel 4 as near break-even (0.96), and Channel 3 as low-performing (0.23). Despite high spend, Channel 7 had moderate ROI (0.82) with high uncertainty, raising concerns about its cost-effectiveness.

To further validate the model structure, we conducted a model comparison using the Widely Applicable Information Criterion (WAIC). We compared two variants:

- **Full model:** Includes Channels 3–5 and 7.
- **Reduced model:** Excludes Channel 7.

A negligible WAIC difference between the full and reduced models suggests Channel 7 adds little explanatory value. This aligns with ROI findings and supports excluding Channel 7 in favor of higher-performing channels like Channel 5.

5 Question and Answers

1. How do you model spend carryover? Spend carryover was modeled using an exponential adstock transformation, which applies a decay parameter to past marketing spend. This captures the diminishing influence of prior spend over time and reflects realistic lagged effects in consumer response.

2. Explain your choice of prior inputs to the model. We used weakly informative priors to balance flexibility and regularization. Coefficients were modeled as $\mathcal{N}(0,1)$, the intercept as $\mathcal{N}(0,5)$, and the noise term as HalfNormal(1), allowing the data to drive inference while avoiding overfitting.

3. How are your model results based on prior sampling vs. posterior sampling? Prior samples represent initial uncertainty, but the final parameter estimates are derived from posterior sampling using the No-U-Turn Sampler (NUTS). Trace plots showed strong convergence and proper mixing, indicating that posterior results are well-supported by the observed data.

4. How good is your model performing? How do you measure it? Model performance was assessed via posterior predictive checks and WAIC-based model comparison. The model accurately captured overall trends in log-revenue, with no divergences and well-behaved sampling chains, confirming both fit and generalizability.

5. What are your main insights in terms of channel performance/effects? Channel 5 had the strongest and most consistent positive effect, making it the most impactful channel. Channel 4 offered modest returns, while Channel 3 showed weak but consistent contribution. Channel 7 exhibited high variability, indicating less reliable impact.

6. Can you derive ROI estimates per channel? What is the best channel in terms of ROI? Yes, ROI was computed by dividing each channel's estimated revenue contribution by its total spend. Channel 5 achieved the highest ROI (2.26), followed by Channel 4 (0.96), Channel 7 (0.82), and Channel 3 (0.23). These results indicate that Channel 5 is the most cost-effective and should be prioritized for future investment.