

Marketing Mix Modeling with Mediation

1. Introduction

This project focuses on modeling weekly revenue by analyzing the effects of paid media channels, direct traffic, and organic search.

Social/Display Spend → Google Ads Spend → Revenue

We apply regularized regression techniques to isolate important contributors while controlling for seasonality and other factors.

2. Data Preparation

Load and Clean the Data

Dataset: cleaned_data.csv, containing two years of weekly data.

Convert the week column to datetime format for proper time series handling.

Handle Missing and Duplicate Entries

Insert rows with zero values for missing weeks to maintain continuity.

Remove any duplicate rows to ensure clean and reliable time series data.

Feature Engineering

Adstock Transformation: Captures the decaying effect of media spend over time.

Lag Features: Includes 1 to 4 week lags for spends and revenue to account for delayed responses.

Seasonality: Uses Fourier series (e.g., annual_sin_1, annual_cos_1) to model yearly patterns.

Time Features: Adds variables like t, week_of_year, month, and year.

Log Transformation: Applies log1p to revenue, spends, and price to reduce skewness and improve scalability.

Scaling

All numeric features are standardized using StandardScaler to ensure uniformity.

Output

A processed dataset ready for modeling with structured, scaled, and engineered features.

3. Modeling Approach

Stage 1 – RidgeCV Model to Predict Google Ads Spend (Mediator)

Goal: Estimate Google spend based on social media spends, adstock features, controls, seasonality, and other factors.

Why RidgeCV: Handles multicollinearity and overfitting through regularization.

Validation: Uses TimeSeriesSplit with 5 folds to account for temporal dependencies.

Output: Generates predicted Google spend (google_pred) to be used in Stage 2.

Stage 2 – ElasticNetCV Model to Predict Revenue

Goal: Model revenue by incorporating predicted Google spend along with other features like social media spend.

Why ElasticNetCV: Combines Lasso and Ridge regression advantages to perform feature selection and regularization.

4. Causal Framework

Stage 1 isolates the mediator (Google Ads spend) and avoids introducing bias by modeling it separately.

Stage 2 incorporates the predicted Google spend as an input, which maintains proper causal interpretation.

This structured approach aligns with marketing causality principles and ensures valid inference.

5. Key Diagnostics

Residual Analysis

Residuals on the test set exhibit no systematic patterns, indicating that the models effectively capture the underlying trends.

Google Spend Prediction

Stage 1 delivers accurate predictions, validating the effectiveness of the mediator model.

Feature Importance

ElasticNet coefficients highlight the most influential revenue drivers:

Social media spends (especially Facebook and TikTok)

Predicted Google Ads spend

Promotions

Average price

Random Forest confirms these rankings, supporting the robustness of the model.

6. Insights & Recommendations

Top Influencers

Google Ads spend plays a key mediating role in converting social media investments into revenue.

Promotions and pricing strategies have a direct and substantial impact on revenue.

Diminishing Returns

Adstock features reveal that higher media spending yields diminishing marginal returns.

Strategic Guidance

Optimize advertising budgets across social and search channels.

Leverage adstock insights to fine-tune campaign timing and frequency.

Regularly monitor pricing and promotional activities to maximize revenue outcomes.

7. Conclusion

This project successfully applies a two-stage causal modeling framework to analyze revenue drivers. T