Marketing Mix Modeling (MMM) with Mediation Assumption – Revenue Prediction

1. Introduction

This project aims to model weekly revenue as a function of paid media metrics, direct response channels (email/SMS), price, promotions, and social media followers. The analysis uses a causal perspective, where Google spend is treated as a mediator between social/display spend and revenue:

Social/Display → Google Spend → Revenue

We leverage regularized regression to identify key drivers while controlling for seasonality, trends, and media lag effects.

2. Data Preparation

Load Preprocessed Data

- Dataset: `cleaned_data.csv` (weekly, 2 years).
- Convert `week` column to datetime.

Handle Missing Weeks & Duplicates

- Insert zero-valued rows for missing weeks.
- Remove duplicates to ensure clean time series.

Feature Engineering

- **Adstock Transformation:** Models diminishing returns of media spend.
- **Lag Features:** Include 1–4 week lags for spends and revenue.
- **Seasonality:** Fourier series (`annual_sin_1`, `annual_cos_1`, etc.) for annual patterns.
- **Time Features:** `t`, `week_of_year`, `month`, `year`.
- **Log Transformation:** `log1p` applied to revenue, spends, and price for better scaling.

Scaling

- All numeric features standardized using `StandardScaler`.

Output

- Cleaned dataset ready for modeling.

3. Modeling Approach

Stage 1: RidgeCV – Predict Google Spend (Mediator)

- **Objective:** Model Google spend using social media spends, adstock, controls, seasonality, and time features.
- **Reason:** `RidgeCV` handles collinearity and prevents overfitting.
- **Validation:** `TimeSeriesSplit` with 5 folds.
- **Output:** Predicted Google spend (`google_pred`) used in Stage 2.

Stage 2: ElasticNetCV – Predict Revenue

- **Objective:** Model revenue using predicted Google spend, social media spend, adstock, lag features, price, promotions, and other controls.
- **Reason:** `ElasticNet` performs feature selection and handles multicollinearity.

4. Causal Framing

- Stage 1 models mediator (Google spend) separately.
- Stage 2 uses predicted Google spend as a feature to model revenue.
- This approach respects back-door paths and avoids leakage.
- Ensures the model aligns with marketing causality assumptions.

5. Key Diagnostics

Residual Analysis

- Residuals on test set show no systematic patterns → model captures trend/seasonality.

Google Predicted vs Actual

- Stage 1 accurately predicts Google spend, validating mediator model.

Feature Importance

- ElasticNet coefficients highlight top revenue drivers:
  - Social spends (Facebook, TikTok)
  - Predicted Google spend
  - Promotions
  - Average price
- RandomForest confirms ElasticNet feature ranking.

6. Insights & Recommendations

Top Drivers

- Google spend mediates social spend impact.
- Promotions and price significantly influence revenue.

Diminishing Returns

- Adstock features show decreasing marginal returns for high media spends.

Marketing Recommendations

- Allocate budget across social and search efficiently.
- Use adstock insights to optimize campaign timing.
- Monitor pricing and promotions for maximum revenue impact.

7. Conclusion

- Successfully modeled revenue using a two-stage causal framework.
- Models are interpretable, handle seasonality, lag effects, and collinearity.
- Cleaned dataset (`cleaned_data.csv`) and models (`stage1_google_model.pkl`, `stage2_revenue_model.pkl`) ensure reproducibility.