Report on Marketing Mix Modeling Using Linear Regression

1. Introduction

This report presents the analysis and results of a linear regression model applied to the marketing mix data of a company, referred to as Company X. The objective of the analysis is to understand the effectiveness of various advertising channels on weekly sales. Company X utilizes seven different paid advertising channels and aims to evaluate the impact of spending on these channels on future sales. This report answers key questions regarding the modeling approach, performance, and insights derived from the analysis.

2. Modeling Spend Carry Over

2.1 Adstock Transformation

To model the carry-over effect of advertising spend, we employed the adstock transformation. Adstock is a concept used to account for the delayed and diminishing impact of advertising over time. The transformation smoothens the spending data, incorporating the effect of past expenditures into the current period.

The adstock formula used is:

adstock[t] = spend [t] + decay_rate * adstock [t-1]

Where:

- spend[t] is the spending at time \(t\).

- decay_rate is a parameter between 0 and 1 that controls the rate at which the impact of past spending diminishes over time.

For this analysis, we set the decay rate to 0.5, meaning that the impact of the advertising spend decreases by 50% each week.

2.2 Implementation

The adstock transformation was applied to the spend data for each of the seven channels, resulting in adstock-adjusted spending data used as inputs to the regression model.

3. Choice of Inputs to the Model

3.1 Feature Selection

The input data for the polynomial regression model were inventory-adjusted advertising expenditure for each of the seven channels. These characteristics were chosen because they represent the cumulative and diminishing impact of advertising expenditure on sales, in line with the objective of understanding the long-term effects of marketing efforts.

3.2 Target Variable

The target variable was the weekly revenue generated by the company, as this is the primary metric of interest for evaluating the effectiveness of advertising spend.

3.3 Data Splitting

The dataset was split into training and testing sets (80% training, 20% testing) to allow for model evaluation and to prevent overfitting.

4. Why choose polynomial regression instead of the Bayesian model?

Polynomial regression is an extension of classical linear regression, where the relationships between the independent variables and the dependent variable are modeled as polynomials. It is a deterministic regression method where the parameters are estimated by classical error minimization techniques, such as ordinary least squares.

4.1 Differences between Polynomial Regression and Bayesian Models

I. Polynomial regression

- Nature : Deterministic.
- Parameters : Parameters are estimated directly from the data by minimizing a cost function (often the mean square error).
- Approach : Fit a polynomial equation ($y=\beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n$) to the data.
- Estimation : Uses techniques such as least squares to find the best coefficients that minimize the error between predictions and observations.

II. Bayesian models

- Nature: Probabilistic.
- Parameters: Parameters are treated as random variables with a priori distributions.

- Approach: Uses Bayes' theorem to update a priori beliefs about parameters after observing data. The a posteriori distribution of the parameters is calculated.
- Estimation: Uses methods such as Markov Chain Monte Carlo (MCMC) sampling to estimate the a posteriori distribution of the parameters.

5. Model Performance

5.1 Evaluation Metrics

The performance of the linear regression model was evaluated using two metrics:

1. Root Mean Squared Error (RMSE): This metric measures the average magnitude of the errors between predicted and actual values.

2. R^2 Score: This metric indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

5.2 Results

- RMSE: `27080.19`

- R^2 : `0.69`

These results indicate that the model explains 69% of the variance in the revenue data and has a reasonably low average error.

6. Main Insights on Channel Performance

6.1 Coefficient Analysis

The coefficients of the linear regression model represent the marginal effect of a unit increase in adstock-adjusted spending for each channel on the weekly revenue.

6.2 Interpretation

- Channel 1: Positive impact, coefficient = `15.38`

- Channel 2: Negative impact, coefficient = `-53.40`

- Channel 3: Negative impact, coefficient = `-5.10`

- Channel 4: Negative impact, coefficient = `-19.65`

- Channel 5: Negative impact, coefficient = `-1.03`

- Channel 6: Neutral impact, coefficient = `6.43`

- Channel 7: Positive impact, coefficient = `4.53`

6.3 Insights

- Channels 1, 6 and 7 have positive impacts on revenue, with Channel 1 having the highest impact.

- Channel 2,3,4 and 5 have negative impact, indicating that spending on this channel may not be efficient.

7. ROI Estimates per Channel

7.1 ROI Calculation

Return on Investment (ROI) for each channel was calculated as the ratio of the total impact on revenue to the total spend on that channel.

7.2 Best Channel in Terms of ROI

The best channel in terms of ROI is Channel 1, with an ROI of 1.45. This indicates that for every unit of currency spent on Channel 1, the company earns 1.45 units in revenue.

8. Conclusion

Channel 1 was identified as the most effective channel, offering the highest ROI. These insights can help Company X optimize its advertising spend to maximize revenue.

This analysis demonstrates the utility of linear regression in marketing mix modeling and provides a foundation for further refinement, including the exploration of non-linear effects and interaction terms in future models.