### 1. Introduction

The purpose of this assignment is to analyze the impact of various advertising expenditures (TV, online, and print ads) and pricing on the sales volume of a blender brand. By comparing different models, we aim to identify key factors that drive sales and provide data-driven recommendations for decision-making. We selected both parametric and machine learning models to explore differences in interpretability and predictive accuracy.

We opted for **Linear Regression, Ridge Regression,** **Lasso Regression** as the parametric model, allowing for nonlinear relationships between variables, and selected **Random Forest**, **Gradient Boosting**, and **Support Vector Machine Regression** as machine learning models, known for capturing complex, nonlinear patterns. After training, validating, and evaluating each model, we selected the best parametric and machine learning model for further analysis.

### **2. Model Proposal**

For this analysis, we examined the effects of advertising spend and price using the following models:

* **Parametric Model**: Linear Regression, Ridge Regression, Lasso Regression, which can handle nonlinear relationships between advertising spend and price.
* **Machine Learning Models**: We used Random Forest, Gradient Boosting, and Support Vector Regression to capture potentially complex nonlinear relationships and enhance predictive accuracy.

**Interaction Terms** between advertising spend and price were considered in the models. We split the data into an 80-20 ratio for training and testing to ensure model generalization, where the training set was used for model tuning, and the test set was reserved for model validation.

#### i. Training and Testing

The dataset was split into 80% training data and 20% testing data. The training data was used for fitting models, while the testing data evaluated model performance.

#### ii. Nonlinear Factors

In Lasso Regression, second-order transformations of advertising and price were applied to capture nonlinear relationships. The machine learning models used their intrinsic nonlinear abilities to manage these complex interactions.

#### iii. Model Evaluation and Selection

Each model's performance was evaluated using Mean Squared Error (MSE) and the R² score.

Based on these results, **Rideg Regression** and **Gradient Boosting** were selected as the best-performing parametric and machine learning models, respectively.

### 3. Model Comparison and Selection

When comparing Polynomial Regression and Gradient Boosting, we considered the following factors:

* **Predictive Accuracy**: Gradient Boosting showed slightly better predictive accuracy with lower MSE and higher R².
* **Interpretability**: Polynomial Regression is more interpretable, as each coefficient directly reflects the influence of each variable on sales. Although Gradient Boosting had higher accuracy, it lacks interpretability.
* **Estimation Process**: Polynomial Regression has a straightforward estimation process and requires fewer computations, while Gradient Boosting, although more complex, is beneficial for predicting on large datasets.

Considering both accuracy and interpretability, **Ridge Regression** was chosen as the final model for in-depth analysis.

### 4. ROI and Elasticity Calculation

Using the coefficients of the Rideg Regression model, we calculated the ROI and elasticity for advertising and price.

### Analysis

**TV\_ads**:

* 1. **ROI**: 0.000307, meaning each dollar spent on TV ads yields approximately $0.000307 in additional sales. This low ROI shows limited incremental revenue from TV ads.
  2. **Elasticity**: 1.675859, indicating a 1% increase in TV ad spending would increase sales by approximately 1.68%. Since elasticity is greater than 1, TV ads have a relatively high impact on sales growth, showing positive responsiveness.

**online\_ads**:

* 1. **ROI**: 0.000200, suggesting each dollar spent on online ads adds about $0.0002 to sales, representing a low return.
  2. **Elasticity**: 0.465965, meaning a 1% increase in online ad spending leads to only a 0.47% increase in sales. With elasticity less than 1, online ads have a minor effect on sales.

**Print\_ads**:

* 1. **ROI**: 0.004409, which is higher than TV and online ads but still yields a small return per dollar.
  2. **Elasticity**: 0.220916, showing that a 1% increase in print ad spending is associated with only a 0.22% increase in sales. This low elasticity indicates print ads have minimal influence on sales.

**Price**:

* 1. **ROI**: -481.129797, a negative ROI indicating that higher prices reduce sales, where price increases significantly impact revenue negatively.
  2. **Elasticity**: -1.657789, meaning a 1% increase in price would lead to about a 1.66% decrease in sales. This high (negative) elasticity suggests that sales are very sensitive to price changes, with price increases leading to considerable sales declines.

### Summary

* **Advertising Impact**: TV ads remain the most responsive among advertising channels, with the highest elasticity, suggesting they could contribute more significantly to sales growth. Although ROI is low for all ads, TV ads may offer the best potential for scaling sales.
* **Price Sensitivity**: Sales demonstrate high sensitivity to price changes, with even small price increases likely to lead to significant sales reductions. This indicates a highly price-sensitive market where pricing adjustments are critical to maintaining demand.

### 5. Actionable Insights

Based on the analysis above, we recommend the following actions:

**(1)Optimize TV Advertising**:

* **Action**: Increase investment in TV ads, as it shows the highest elasticity (1.68), indicating that TV ads significantly impact sales growth.
* **Rationale**: Even with a low ROI per dollar, the positive elasticity suggests TV ads effectively drive demand. Consider reallocating budget from lower-impact channels to TV ads to maximize sales lift.

**(2)Re-evaluate Online and Print Ad Spending**:

* **Action**: Minimize spending on online and print ads, or test alternative approaches within these channels to improve their effectiveness.
* **Rationale**: Both online (0.47) and print ads (0.22) have elasticity below 1, meaning they provide limited sales responsiveness. Reducing investment here may free up resources for more impactful areas.

**(3)Price Optimization Strategy**:

* **Action**: Implement a pricing strategy focused on stability or gradual price adjustments, and avoid sudden price increases to prevent sales decline.
* **Rationale**: With high negative elasticity (-1.66), even a small price increase is likely to reduce sales. Testing price reductions or offering discounts could boost demand, especially in a price-sensitive market.

**(4)Consider Bundling or Promotional Offers**:

* **Action**: Create bundle offers or periodic promotions, especially when adjusting prices, to counterbalance potential sales declines from price sensitivity.
* **Rationale**: Since price increases lead to significant sales drops, a well-timed promotional strategy can help mitigate the negative impact, maintain customer interest, and drive incremental sales.

**(5)Conduct Further Testing and Analysis**:

* **Action**: Perform A/B testing for different advertising budget allocations and price points to refine the strategy.
* **Rationale**: By experimenting with ad budgets and prices, you can better identify the optimal investment levels and pricing that maximize ROI and minimize customer churn in response to price changes.

### 6. Appendices

Table 1: Parametric regression and machine learning regression results

| **Model** | **MSE** | **R2 Score** |
| --- | --- | --- |
| Linear Regression | 6.864558e+11 | 0.916871 |
| Ridge Regression | 6.864451e+11 | 0.916872 |
| Lasso Regression | 6.864558e+11 | 0.916871 |
| Random Forest | 9.115160e+11 | 0.889617 |
| Gradient Boosting | 4.924807e+11 | 0.940361 |

Table 2: ROI and Elasticity

| **Variable** | **ROI** | **Elasticity** |
| --- | --- | --- |
| TV\_ads | 0.000307 | 1.675847 |
| online\_ads | 0.000200 | 0.465965 |
| Print\_ads | 0.004409 | 0.220914 |
| Price | -481.151678 | -1.657864 |

Fig 1: Scatter plot of Sales vs Advertising channels

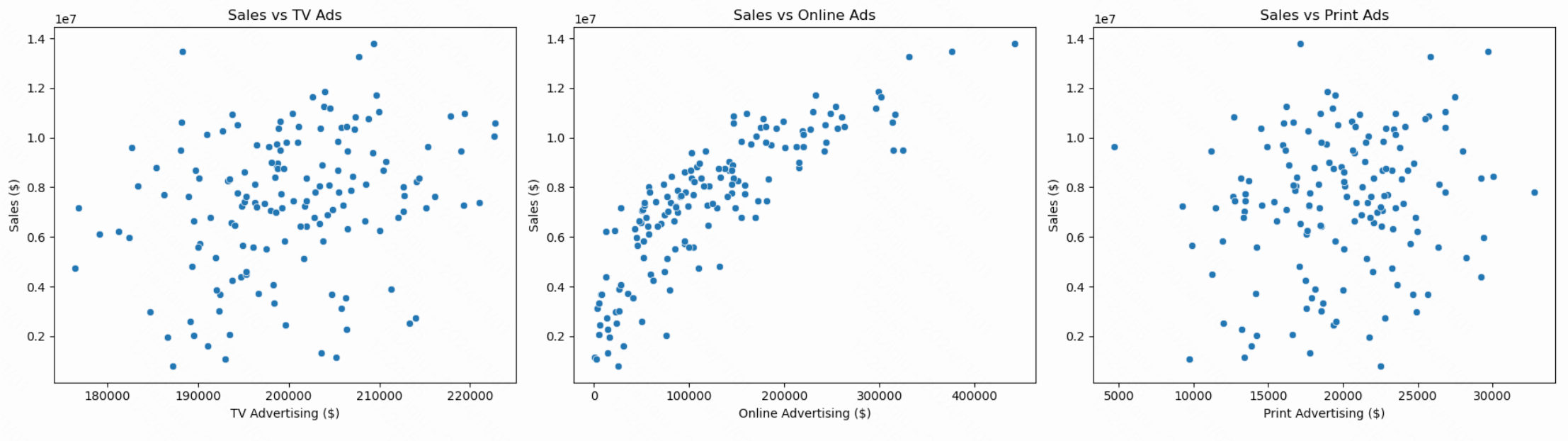


Fig 2: Scatter plot of Price vs Sales

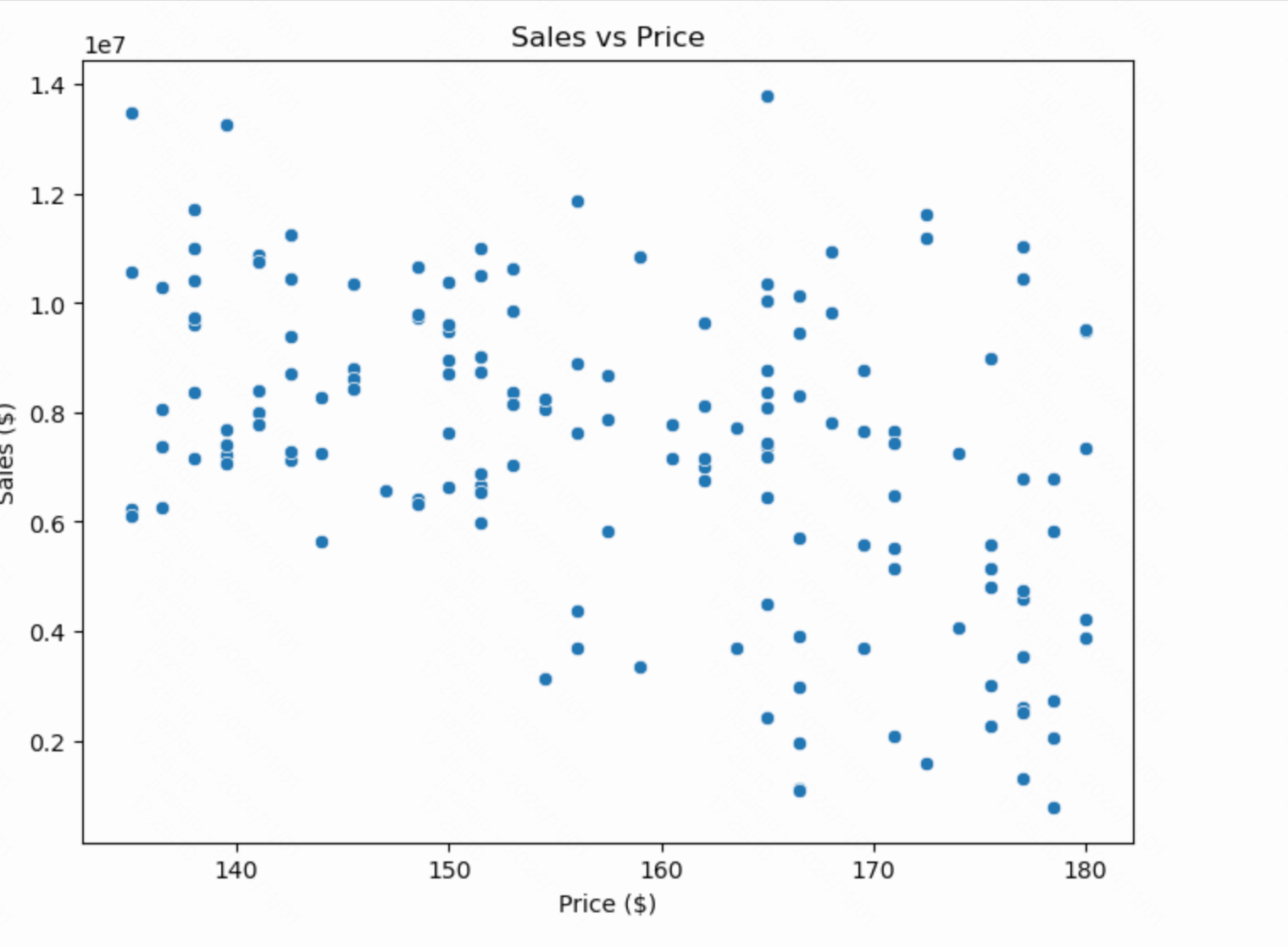


Fig 3: Bar chart of MSE for each model

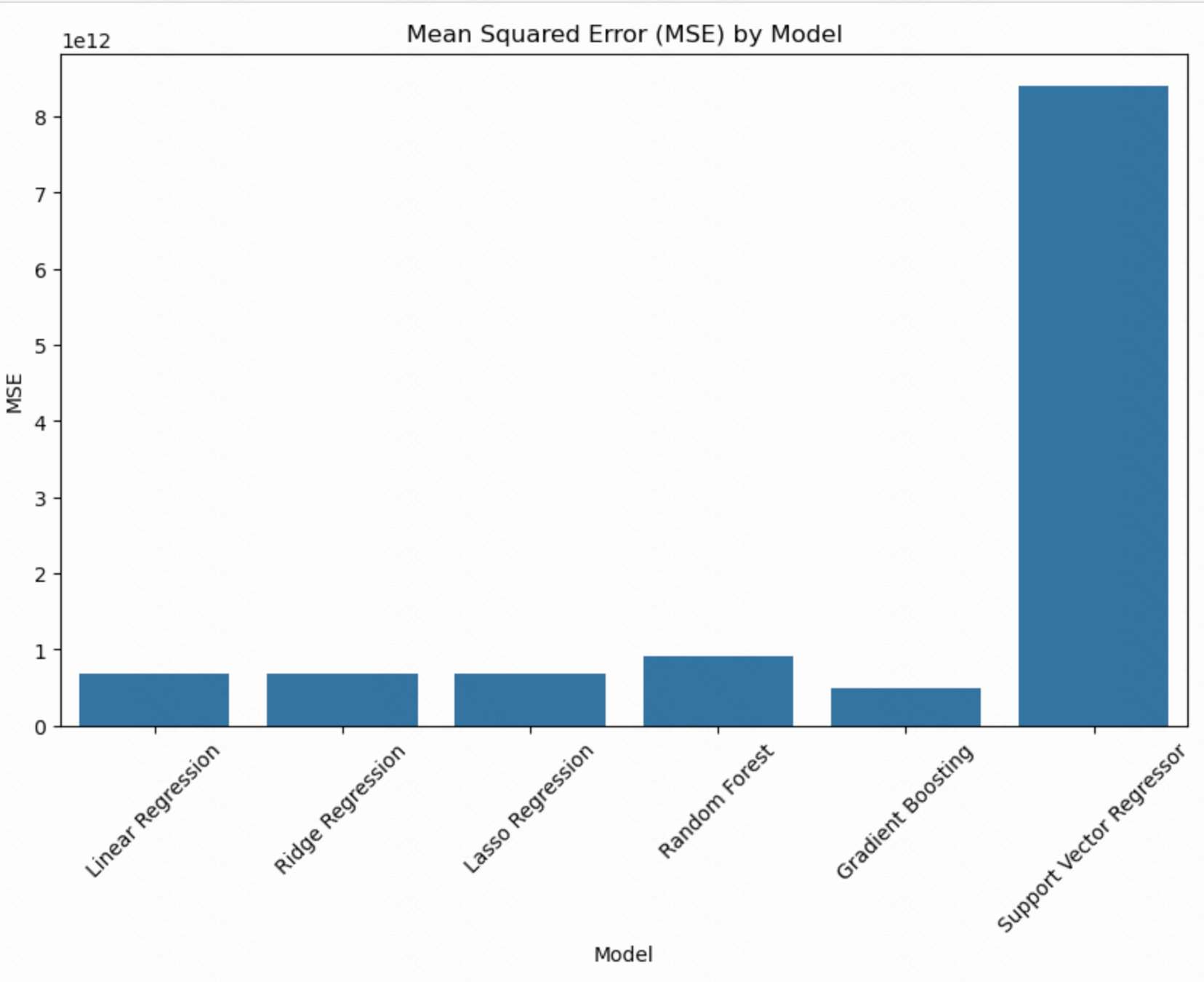


Fig 4: Bar chart of R\*2 Score for each model

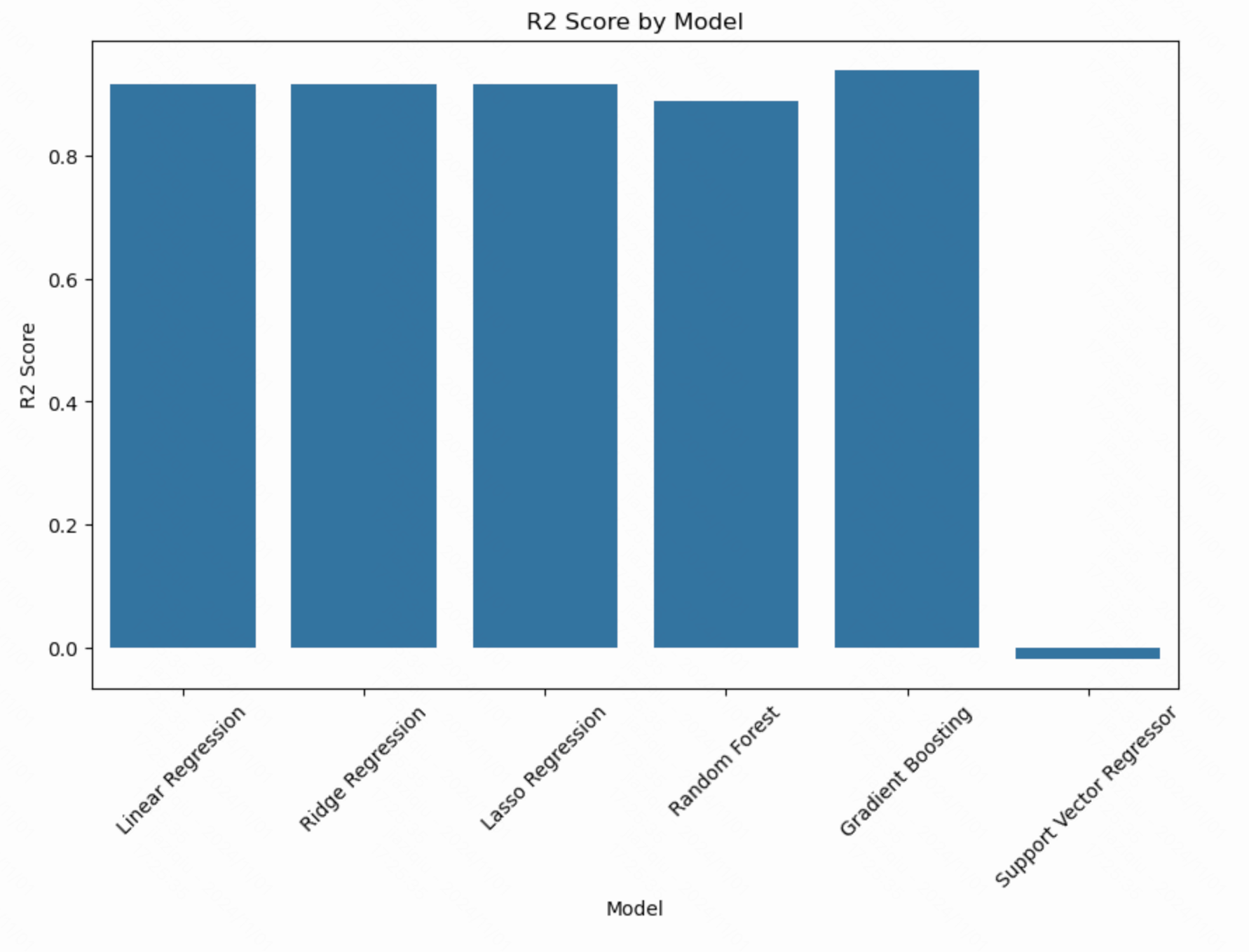


Fig 5: Residual plot for the best-performing model (Polynomial Regression in this case)

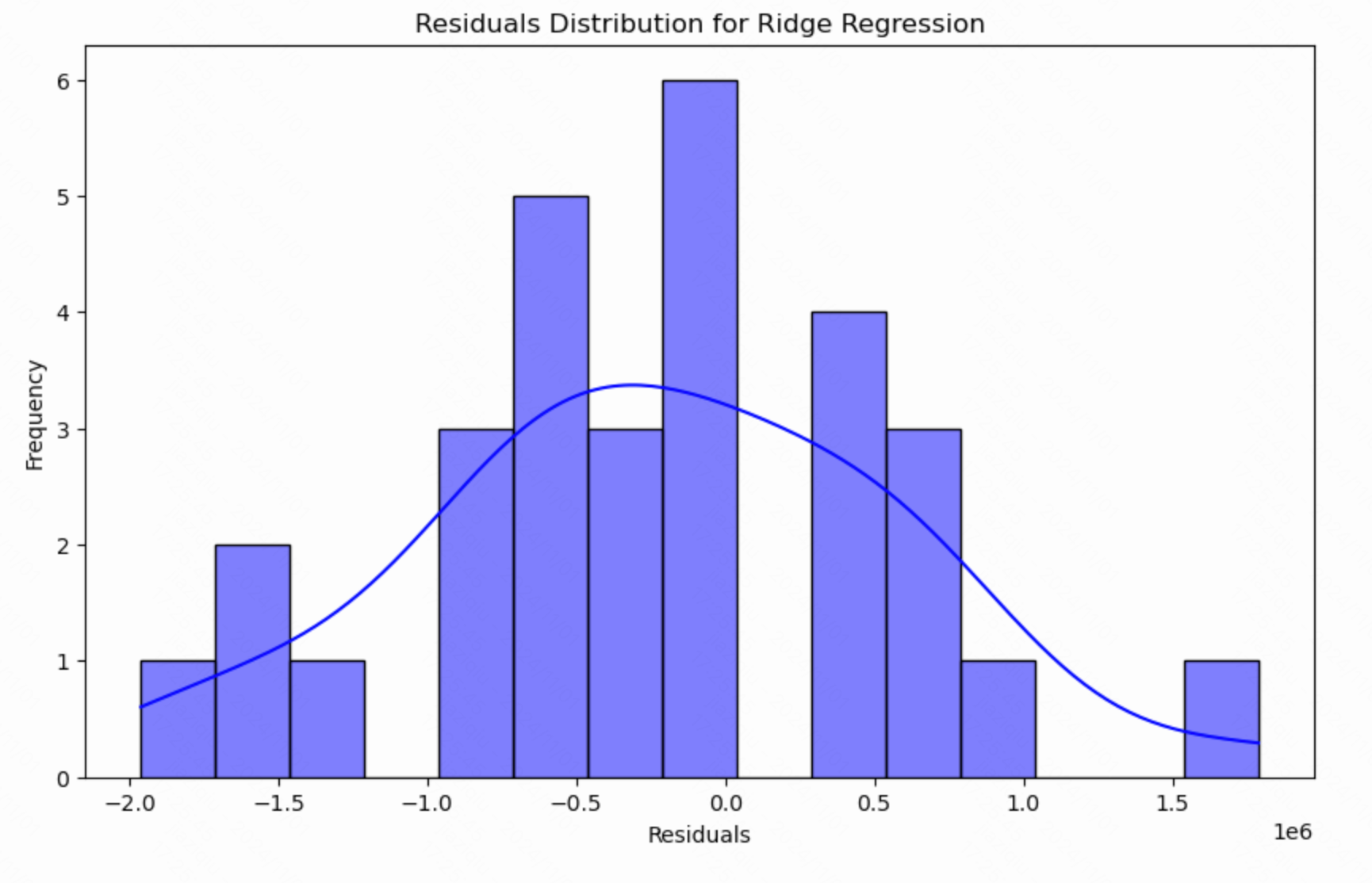


Fig 6: Model Performance Comparison: MSE and R² Score

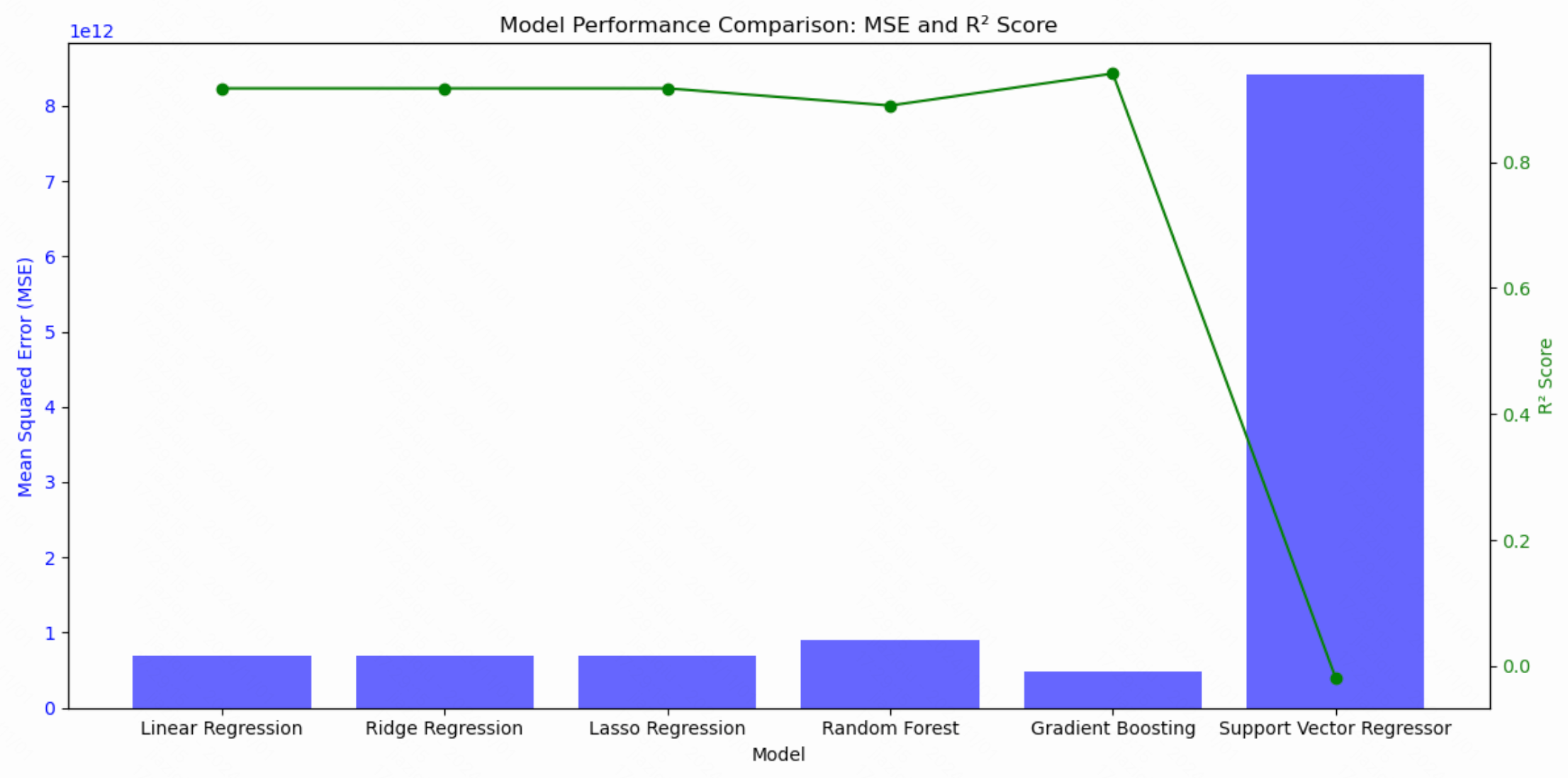


Fig 7: Elasticity and ROI Comparison Across Ad Channels

