NIET Journal of Management (NIETJOM)
Volume 15, Issue 1 Winter / Summer 2024 ISSN: 0975-7643 (Print)

Impact of Advertising Media on Sales Using Linear Regression Analysis

Neeraj Mittal[1], Aman Kumar[2], Amarjeet Singh, Amit Kumar, Amit raj

[1] Noida Institute of Engineering & Technology (MCA Institute), Greater Noida
[2] Noida Institute of Engineering & Technology (MCA Institute), Greater Noida

Abstract: *Advertising plays a crucial role in influencing customer purchasing behaviour and enhancing overall business revenue. This study examines the relationship between advertising expenditure across three major media channels—TV, Radio, and Newspaper—and product sales using a supervised machine learning approach. A linear regression model is developed to predict sales performance based on the allocated advertising budgets. The dataset used in this study comprises 200 observations and four variables. Various techniques such as exploratory data analysis, model training, performance evaluation, and data visualization are applied to gain meaningful insights. The findings reveal that TV and Radio advertisements have a significantly stronger impact on sales compared to Newspaper advertisements. Overall, the proposed model demonstrates satisfactory prediction accuracy and offers valuable insights to support effective marketing decision-making.*

Keywords - Advertising Analytics, Linear Regression, Sales Prediction, Marketing Data Analysis, Machine Learning

1.INTRODUCTION

In today's highly competitive and rapidly evolving business environment, organizations invest substantial financial resources in advertising to enhance brand visibility, attract potential customers, and ultimately increase sales performance. Advertising has become a strategic tool that not only supports product promotion but also strengthens customer relationships and market positioning. However, despite significant investments, all advertising channels do not generate equal returns. Some media platforms may offer higher customer engagement and conversion rates, while others may result in limited impact. Therefore, identifying the most effective advertising channels has become a critical challenge for businesses seeking to maximize return on investment (ROI) and optimize marketing budgets.

This research focuses on analysing the impact of advertising expenditure across three major traditional media platforms — Television (TV), Radio, and Newspaper — on product sales performance. By applying a multiple linear regression model, the study aims to quantify the relationship between advertising spending and sales outcomes. The model helps in identifying which advertising medium contributes the most to sales growth and how effectively sales can be predicted using historical advertising data. The findings of this research provide

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

valuable insights for marketing managers and decision-makers, enabling them to allocate advertising budgets more efficiently and develop data-driven marketing strategies that improve overall business performance

I. Dataset Description

The dataset used in this study contains **200 records** with the following attributes:

Variable Name	Description
TV	Advertising budget spent on television (in thousands)
Radio	Advertising budget spent on radio (in thousands)
Newspaper	Advertising budget spent on newspapers (in thousands)
Sales	Product sales (in thousands of units)

2.1 Data Characteristics

- Total observations: 200
- Total features: 4
- Data type: Float values
- Missing values: None detected
- Memory usage: Approximately 6.4 KB

The dataset is clean and suitable for regression analysis without requiring additional preprocessing.

II. Methodology

3.1 Tools and Libraries Used

The following Python libraries were used in the analysis:

- Pandas – Data handling and manipulation
- NumPy – Numerical computation
- Matplotlib – Data visualization
- Seaborn – Statistical plotting
- Scikit-learn – Machine learning model implementation

3.2 Exploratory Data Analysis (EDA)

To understand the data distribution and relationships among variables, the following techniques were applied:

Histogram Analysis:

Histograms were plotted for all numerical variables to examine their distributions.

Observations:

- TV advertising budgets show wider variation compared to Radio and Newspaper.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

- Sales values appear normally distributed.
- Newspaper advertising shows relatively lower variation and weaker correlation with sales.

3.3 Model Development

A **Multiple Linear Regression Model** was implemented using the following steps:

1. Independent variables selected:
    - TV
    - Radio
    - Newspaper
2. Dependent variable:
    - Sales
3. Data split:
    - Training set: 75%
    - Testing set: 25%
4. Random state was fixed for reproducibility.

3.4 Model Training

The training dataset was used to train the linear regression model, which allowed it to discover the fundamental connection between advertising spending across various media channels and the ensuing sales output. By doing this, the model found trends and measured how adjustments to advertising spending affect sales results.

3.5 Model Evaluation Metrics

The model's performance was evaluated using two essential metrics: mean squared error (MSE) and coefficient of determination (R^2 score). The Mean Squared Error (MSE) calculates the average of the squared discrepancies between actual and forecast values, revealing how much the model's predictions stray from genuine sales data. A lower MSE value indicates that the model has fewer prediction mistakes and hence more accuracy. The R^2 value indicates how much of the variance in the dependent variable (sales) is explained by the independent variables (advertising expenditures). It offers information on how well the model matches the data.

4.Results and Analysis

4.1 Actual vs Predicted Sales (TV Advertising)

A scatter plot was created between:

- X-axis: TV advertising budget
- Y-axis: Sales values

Two datasets were plotted:

- Actual Sales
- Predicted Sales

Observation:

The predicted sales values closely follow the actual sales trend, indicating that TV advertising is a strong predictor of sales performance.

4.2 Actual vs Predicted Sales (Radio Advertising)

Another scatter plot was generated using Radio advertising budget and corresponding sales.

Observation:

Radio advertising also shows a strong positive relationship with sales. The predicted values align well with actual sales points, proving the importance of radio marketing campaigns.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

4.3 Influence of Newspaper Advertising

Compared to TV and Radio, Newspaper advertising demonstrates weaker correlation with sales. This suggests that print media has a comparatively lower impact on product sales in the given dataset.

5.Discussion

The results clearly indicate that:

- TV advertising has the highest influence on sales.
- Radio advertising also contributes significantly.
- Newspaper advertising has minimal impact.

This implies that businesses should prioritize TV and Radio platforms while allocating marketing budgets. The regression model successfully captured the underlying relationships and provided reliable predictions.

6.Conclusion

This research demonstrates the application of linear regression for sales prediction using advertising expenditure data. The developed model performs well and offers valuable insights into marketing strategy optimization.

- **Key Findings:**
  - TV and Radio advertisements are the most effective sales drivers.
  - Linear regression is suitable for modelling advertising-sales relationships.
  - Data-driven decisions can significantly improve marketing ROI.

- **Future Scope**

Future improvements can include:

- Using advanced models such as Random Forest and XGBoost
- Including seasonal and regional data
- Performing feature importance analysis
- Applying time-series forecasting methods

REFERENCES:

[1] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.
https://www.statlearning.com/
[2] Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression*. Wiley.
https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473
[3] Tellis, G. J. (2009). *Effective Advertising: Understanding When, How, and Why Advertising Works*. Sage.
https://us.sagepub.com/en-us/nam/effective-advertising/book226847
[4] Nielsen (2019). *Global Trust in Advertising Report*. Nielsen.
https://www.nielsen.com/us/en/insights/report/2019/global-trust-in-advertising-report/
[5] Dataset repository (GitHub). *Advertising.csv* (2023).
https://github.com/selva86/datasets/blob/master/Advertising.csv

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Volume 15, Issue 1 Winter / Summer 2024 ISSN: 0975-7643 (Print)

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.