



The Impact of Different Types of Advertising on Product Sales

Yizheng Wang*

Department of Finance, Beijing University of Technology, Beijing, China

*Corresponding author: 22373225@emails.bjut.edu.cn

Abstract. This study explores the influence of various advertising types on product sales, with a particular emphasis on the effectiveness of television, radio, and newspaper advertising strategies. The analysis is based on a public sales dataset sourced from Kaggle. In the preliminary stages, data preprocessing and exploratory data analysis (EDA) were conducted to assess the relationship between the expenditures on different types of advertising and product sales. Subsequently, a multiple linear regression model was developed to estimate the regression coefficients associated with each advertising type. This model facilitates a quantitative evaluation of the extent to which each advertising medium influences sales outcomes. The findings suggest that television advertising exerts the most substantial effect on sales, while newspaper advertising has the least influence. Radio advertising's impact is intermediate, indicating a moderate correlation with sales performance. These results underscore the varying degrees of effectiveness of different advertising mediums, providing valuable insights for marketers seeking to optimize their advertising strategies.

Keywords: Multiple linear regression modeling, advertising, e-commerce, marketing, python.

1 Introduction

Advertising is an indispensable part of the business world. The history of advertising can be traced back to the end of ancient primitive societies, where it appeared along with the production and exchange of goods. In today's business environment, advertising still plays an important role, as companies place various types of advertisements to enhance brand awareness and increase sales. Among the many advertising strategies, television, radio and newspapers have long been generally recognized as the most dominant and effective advertising media. However, different types of advertising strategies often have different impacts on product sales, and therefore, recognizing and quantifying this impact relationship is crucial for marketers to develop appropriate marketing policies.

This study examines the effectiveness of television, radio and newspaper advertising in driving product sales. By obtaining and analyzing a public sales dataset on the kaggle platform, this study aims to quantify the relationship between spending on the three

advertising channels and the sales growth achieved. Multiple linear regression models were used in the study to provide the basis for an in-depth analysis of the relationship between advertising spend and sales outcomes. With this model, we aim to identify which type of advertising contributes the most to sales growth and to quantitatively compare the differences in effectiveness between different types of advertising.

The findings of this study will provide valuable suggestions for companies to optimize their advertising budgets and provide theoretical guidance for relevant marketers to adjust the allocation of advertising resources for maximum return on investment. The results of the study will not only enrich marketing theory, but also have profound implications for real-world company operations.

2 Literature Review

Advertising, as an important means of increasing product sales, has been receiving attention from scholars all over the world, and quantitative analysis of its effects has been a hot topic in marketing research. Yang Zhou (2021) used a simple linear regression analysis to explore the strong correlation that exists between YouTube, as an effective medium of advertising, and sales [1]. John R. Hauser (1982) quantified the dynamic effects of managerial variables, such as advertising, on purchasing behavior and sales by integrating multi-domain theories and data and using regression analysis techniques [2]. Julie R. Irwin (2001) identified the moderating relationship (interaction) as the core of marketing [3].

Multiple regression modeling originated in the 19th century, and its earliest incarnation was the method of least squares, first systematically introduced by Gauss in 1801. and was extended to a multivariate form by Youl in the early 20th century. Fisher's theoretical refinements in the 1930s and the advancement of computer technology in the mid-to-late 20th century made it a cornerstone tool in marketing research. Several scholars have refined and used multivariate regression modeling in the 20th and 21st centuries and applied it in various academic fields. For example, Brian Dalessandro (2001) proposed a moderated multiple regression model to demonstrate the interaction effect between two independent variables by including a product term [3], Roidah Lina (2022) used multiple linear regression for the linear relationship between product quality and user loyalty [4]. Guoquan Ye (2021) investigated the Netflix marketing how it affects customers' attitudes towards products through multiple factors [5]. Yang Cheng (2022) explored the relationship between relationship marketing and consumer behavioral intentions [6].

Some scholars have also tried to combine multiple regression modeling with marketing. For example, Fine F. Leung (2022) found through research that a variety of factors such as the sender of the message, the receiver of the message, and the message itself affect the effectiveness of marketing behavior [7]. Andrew N. Mason (2020) used multiple regression models to analyze the gradual increase in the importance of social media marketing in the wake of the New Crown Virus, in order to tell the story of how the various elements of the pandemic's production affect the consumer behavior [8].

Loes Janssen (2021) explores how multiple elements of Netflix as an effective advertising tool, such as its own image and product fit, affect marketing effectiveness [9].

However, the data used in these studies mentioned above were obtained through questionnaires, and there is uncertainty about their authenticity and reliability, and these studies did not examine the quantitative analysis of different types of advertisements on product sales. Therefore, this experiment aims to fill this gap by quantitatively analyzing the impact of different ad types on product sales through a publicly available real data set obtained from kaggle and building a multiple linear regression model based on the data set.

3 Data

3.1 Description of the Data Set

The Kaggle platform provides the dataset used in this survey, which contains historical data on spending on different ad types and corresponding product sales. The key features of the dataset include the spending on different ad types and the corresponding increase in sales. These data are used to analyze the impact of different ad types on sales, train multiple linear regression models, and validate the results.

3.2 Data Preprocessing

Detects possible missing values in the dataset and presents basic information about the statistics, including the mean, standard deviation, minimum and maximum values of the data.

3.3 Exploratory Data Analysis (EDA)

Plot the relationship between different types of advertising expenditures and sales.

Figure 1 shows the results of the EDA analysis, which has resulted in the relationship between different types of advertising expenditures and product sales revenues. Television advertising expenditures and sales revenue may show a significant positive correlation, i.e., as expenditures increase, sales revenue also increases significantly. Radio advertising expenditures may also be positively correlated with sales, but the slope is more moderate than that of television advertising. The relationship between newspaper advertising expenditures and sales revenues is weaker and may be close to horizontal, indicating a relatively small impact on sales.

Calculate the correlation between different types of ad spending and sales and plot it as a correlation heat map.

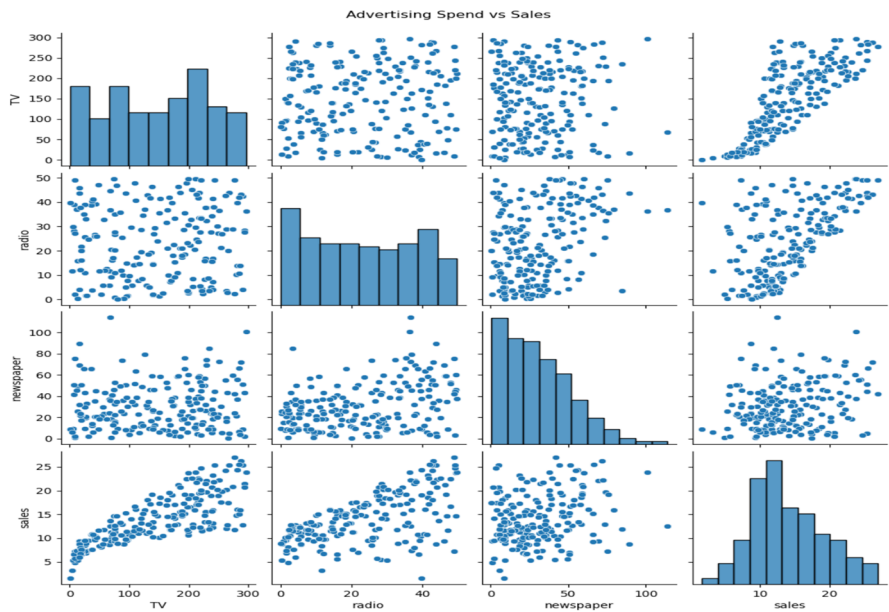


Fig. 1. EDA analysis results

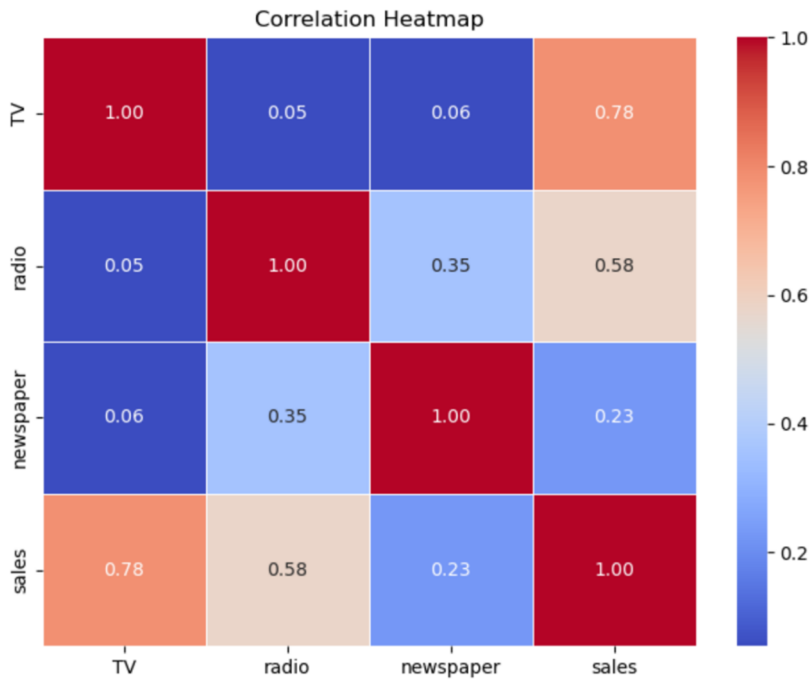


Fig. 2. Correlation heat map

Figure 2 is a correlation heat map that shows the correlation between different variables by color depth and value. Red color represents strong positive correlation and blue color represents weak or negative correlation. As can be seen from the figure, the correlation between different types of advertising expenditures is weak, indicating that the model's multicollinearity problem is relatively mild.

4 Methodology

4.1 Advantages of Multiple Linear Regression Models

The main model used in this study is the multiple linear regression model. One of the advantages of the multiple regression model is that it is highly interpretable, and the regression coefficients of the model give a clear idea of the direction and magnitude of the effect of each independent variable on the dependent variable, and this interpretability is particularly important in many fields, especially in economics.

Another advantage of the multiple linear regression model is its multicollinearity detection, which can help identify and solve multicollinearity problems by calculating indicators such as variance inflation factor (VIF).

In addition, the multiple linear regression model can be easily extended to a variety of forms, such as polynomial regression, ridge regression, Lasso regression, etc., which is highly scalable [10-11].

4.2 Structure of the Multiple Linear Regression Model

The basic structure of a multiple linear regression model is a linear equation that predicts the dependent variable through a set of independent variables. The basic structure of the model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (1)$$

Where Y belongs to the dependent variable, X_1, X_2, X_n belong to the independent variables, β_0 belongs to the intercept term, $\beta_1, \beta_2, \beta_n$ are regression coefficients and ε is the error term.

4.3 Model Training

The dataset is first divided into a training set and a test set, and the proportion of the training set and the test set is 80% and 20%, respectively. Then a basic multiple linear regression model is constructed and the model is trained using different types of advertising data and sales data in the training set.

4.4 Prediction Process

Next, this study uses the expenditure data of different ad types in the test set to predict the changes of the corresponding sales in the test set, and compares it with the actual

values in the test set to calculate the mean square error of the model, as well as to output the regression coefficients corresponding to different types of ads.

5 Results and Discussion

5.1 Experimental Results

By training the multiple linear regression model, giving the predicted values and comparing them with the actual values, this study obtained the mean score error of the model and the regression coefficients corresponding to different types of advertisements.

```
Mean Squared Error (MSE): 3.174097353976106  
  
regression coefficient:  
TV: 3.8306059858246497  
radio: 2.8019117101643043  
newspaper: 0.05998274078890284
```

Fig. 3. Mean score error and regression coefficients

Figure 3 shows that the mean square error of the model is about 3.17 (%) and the three regression coefficients of the model are 3.83, 2.80 and 0.05, corresponding to the three types of advertisements: television, radio and newspaper.

5.2 Results Analysis

The results show that the multivariate linear regression model constructed in this experiment is more effective as demonstrated by its smaller mean square error, which shows that the predicted values obtained through the model differ less from the actual values, and therefore the regression coefficients corresponding to the different types of advertisements of this experimental campaign are also more accurate as compared to the real values. The results of the regression coefficients show that TV advertisements have the greatest impact on sales, with one unit of TV advertisement expenditure often resulting in more than three units of sales growth. Newspaper advertising, on the other hand, has the smallest impact on sales, with one unit of newspaper advertising expenditure tending to bring about only 0.05 units of sales growth. Radio advertising falls between the two, and its sales growth effect is closer to that of television advertising and substantially better than that of newspaper advertising.

6 Conclusion

The purpose of this study is to construct and use a multiple linear regression model to explore the effect of different types of advertisements on the sales of a product. The

advertisements consist of three categories, which are television advertisements, radio advertisements, and newspaper advertisements. In this study, the real public dataset is obtained from kaggle, after data preprocessing and EDA analysis, the dataset is divided into two parts, the training set and the test set, which account for 80% and 20% respectively, by using the data in the training set to train the basic multiple linear regression model that is constructed, and the model is applied to the test set, and compared with the actual values in the test set and calculates the model's mean squared error and get the regression coefficients corresponding to different types of advertisements in the model. The experimental results show that, for the same expenditure, TV advertisements lead to the largest increase in sales, newspaper advertisements lead to the smallest increase in sales, and radio advertisements fall in between.

References

1. Zhou, Y., Ahmad, Z., Alsuhabi, H., Yusuf, M., Alkhairy, I., & Sharawy, A. M. (2021). Impact of YouTube Advertising on Sales with Regression Analysis and Statistical Modeling: Usefulness of Online Media in Business. *Computational Intelligence and Neuroscience*, 2021(1), 9863155.
2. Hauser, J. R., & Wisniewski, K. J. (1982). Dynamic analysis of consumer response to marketing strategies. *Management Science*, 28(5), 455-486.
3. Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models.
4. Lina, R. (2022). Improving product quality and satisfaction as fundamental strategies in strengthening customer loyalty. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 2(1), 19-26.
5. Ye, G., Hudders, L., De Jans, S., & De Veirman, M. (2021). The value of influencer marketing for business: A bibliometric analysis and managerial implications. *Journal of advertising*, 50(2), 160-178.
6. Cheng, Y., & Jiang, H. (2022). Customer-brand relationship in the era of artificial intelligence: understanding the role of chatbot marketing efforts. *Journal of Product & Brand Management*, 31(2), 252-264.
7. Leung, F. F., Gu, F. F., Li, Y., Zhang, J. Z., & Palmatier, R. W. (2022). Influencer marketing effectiveness. *Journal of marketing*, 86(6), 93-115.
8. Mason, A. N., Narcum, J., & Mason, K. (2021). Social media marketing gains importance after Covid-19. *Cogent Business & Management*, 8(1), 1870797.
9. Janssen, L., Schouten, A. P., & Croes, E. A. (2022). Influencer advertising on Instagram: product-influencer fit and number of followers affect advertising outcomes and influencer evaluations via credibility and identification. *International journal of advertising*, 41(1), 101-127.
10. Meng, W., Ming-liang, S., Jian-jun, Z., Si-xin, W., & Min, C. (2024). Least Square Estimation for Multiple Functional Linear Model with Autoregressive Errors. <https://doi.org/https://doi.org/10.1007/s10255-024-1143-2>
11. D'Haultfœuille, X., Gaillac, C., & Maurel, A. (2022). Partially Linear Models under Data Combination. <https://doi.org/https://doi.org/10.48550/arxiv.2204.05175>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

