The Impact of Advertising on Sales

Contents

[1 Executive Summary 2](#_Toc176612697)

[2. Business Context and Objective 2](#_Toc176612698)

[Objective: 2](#_Toc176612699)

[Scope: 2](#_Toc176612700)

[3. Dataset Overview 2](#_Toc176612701)

[Summary Statistics: 3](#_Toc176612702)

[Key Observations: 3](#_Toc176612703)

[4. Exploratory Data Analysis (EDA) 3](#_Toc176612704)

[A. Distribution of Variables 3](#_Toc176612705)

[B. Correlation Analysis 4](#_Toc176612706)

[C. Pair Plot Analysis 4](#_Toc176612707)

[D. Boxplot for Outlier Detection 5](#_Toc176612708)

[5. Linear Regression Model 6](#_Toc176612709)

[Model Setup: 6](#_Toc176612710)

[Model Results: 6](#_Toc176612711)

[Interpretation of Coefficient: 6](#_Toc176612712)

[6. Model Evaluation 6](#_Toc176612713)

[Residual Analysis: 7](#_Toc176612714)

[7. Findings and Business Implications 7](#_Toc176612715)

[TV Advertising Drives Sales: 7](#_Toc176612716)

[Other Mediums: 8](#_Toc176612717)

[Model Accuracy: 8](#_Toc176612718)

[8. Conclusion 8](#_Toc176612719)

## 1 Executive Summary

This business analysis report provides an in-depth analysis of the impact of **TV advertising expenditure** on **sales revenue**. The analysis uses a simple linear regression model, with **TV advertising** as the sole predictor variable and **sales revenue** as the dependent variable. The dataset includes data for 200 advertising campaigns across three mediums: TV, Radio, and Newspaper. However, this analysis focuses solely on the effect of TV advertising on sales performance.

Key insights suggest that TV advertising has a **significant positive impact** on sales. For every additional **$1000** invested in TV advertising, the model predicts an increase of approximately **$46,000** in sales revenue. The model is evaluated using various metrics, and the residual analysis confirms that the model does not suffer from bias or major issues.

## 2. Business Context and Objective

Advertising budgets are crucial for companies looking to increase sales and market penetration. Companies often allocate funds across different media such as TV, Radio, and Newspaper. Understanding the return on investment (ROI) for each advertising channel can help businesses make data-driven decisions and optimize their marketing budgets.

### Objective:

* To analyze the impact of **TV advertising** expenditure on **sales revenue**.
* To quantify the relationship between TV ad spend and sales using a linear regression model.
* To provide business recommendations based on the findings.

### Scope:

While the dataset includes TV, Radio, and Newspaper advertising data, this report focuses exclusively on **TV advertising** as the predictor of sales.

## 3. Dataset Overview

The dataset consists of **200 records**, with each row representing a unique advertising campaign. The following columns are available:

* **TV**: The amount spent on TV advertising in thousands of dollars.
* **Radio**: The amount spent on Radio advertising in thousands of dollars.
* **Newspaper**: The amount spent on Newspaper advertising in thousands of dollars.
* **Sales**: The sales revenue generated by the campaign in millions of dollars.

### Summary Statistics:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | TV (in 1000$) | Radio (in 1000$) | Newspaper (in 1000$) | Sales (in M$) |
| Count | 200 | 200 | 200 | 200 |
| Mean | 147.04 | 23.26 | 30.55 | 14.02 |
| Min | 0.70 | 0.00 | 0.30 | 1.60 |
| Max | 296.40 | 49.60 | 114.00 | 27.00 |
| Std Dev | 85.85 | 14.85 | 21.78 | 5.21 |

### Key Observations:

1. **TV advertising** has the largest budget allocations, with an average of **$147,040** and a maximum of **$296,400**.
2. **Radio advertising** has a much lower mean budget of **$23,260**.
3. **Newspaper advertising** spending ranges from **$300** to **$114,000** with an average of **$30,550**.
4. **Sales** revenue generated by these campaigns varies between **$1.6M** and **$27M**, with an average sales revenue of **$14.02M**.

## 4. Exploratory Data Analysis (EDA)

EDA is an essential part of data analysis that helps us understand the distribution of variables and the relationships between them.

### A. Distribution of Variables

1. **TV Advertising**:
   * TV advertising budgets show a **broad distribution**, with some campaigns spending as little as **$700** and others as much as **$296,400**.
   * The distribution is roughly symmetrical, which suggests a variety of TV ad spend strategies, with no extreme skew.
2. **Sales Revenue**:
   * Sales values are distributed between **$1.6M** and **$27M**, with most values concentrated between **$10M** and **$20M**.
   * The sales distribution is slightly skewed to the right, meaning that there are more campaigns with sales in the lower to mid-range, but some high-sales campaigns push the mean up.

### B. Correlation Analysis

To understand the relationships between variables, a **correlation heatmap** was generated. The heatmap shows the degree of linear association between each pair of variables, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation).

* **TV and Sales**: A **strong positive correlation** of **0.78**, suggesting that higher TV ad spending is closely linked to higher sales.
* **Radio and Sales**: A weaker but still positive correlation of **0.58**.
* **Newspaper and Sales**: The correlation between Newspaper ad spend and sales is very low (**0.23**), suggesting that Newspaper ads are not as impactful on sales as TV or Radio.

**A screenshot of a graph

Description automatically generated**

### C. Pair Plot Analysis

The **pair plot** visualizes relationships between variables. The plot reveals a clear **linear relationship** between **TV advertising** and **sales**. Campaigns with higher TV ad spend tend to generate higher sales.

In contrast, the pair plots between Radio/Newspaper and Sales show **more scattered relationships**, indicating that these mediums are less predictive of sales performance.

A collage of blue and white graphs

Description automatically generated

### D. Boxplot for Outlier Detection

Boxplots were used to check for outliers:

* **TV Advertising**: There are no significant outliers, indicating that TV budgets are generally within a reasonable range.
* **Radio and Newspaper Advertising**: Some outliers are present, particularly for Newspaper, which has a few high-budget campaigns that may not conform to the overall trend.

A graph of a bar chart

Description automatically generated with medium confidence

## 5. Linear Regression Model

A simple **linear regression model** was built to quantify the relationship between **TV advertising** and **sales**. The model aims to answer the question: **How much do sales increase with each additional dollar spent on TV advertising?**

### Model Setup:

* **Dependent Variable (Y)**: Sales (in millions of dollars).
* **Independent Variable (X)**: TV advertising budget (in thousands of dollars).

**Model Equation**:

Sales=β0+β1×TV

Where:

* β0​ is the **intercept** (expected sales when TV advertising spend is 0).
* β1​ is the **slope coefficient** (the increase in sales for each additional $1000 spent on TV).

### Model Results:

* **TV Coefficient (β1​)**: **0.046**
  + This means that for every **$1000** increase in TV advertising, sales increase by **$46,000**.
  + This coefficient is **statistically significant**, showing a meaningful positive relationship between TV spending and sales.
* **Intercept (β0​)**: This represents the expected sales when **no money is spent on TV advertising**. It provides a baseline level of sales that might come from organic channels, brand equity, or other forms of marketing.

### Interpretation of Coefficient:

* The model indicates that TV advertising is a strong predictor of sales. For every **$1000** spent on TV ads, sales are expected to increase by **$46,000**.

## 6. Model Evaluation

The model was evaluated using multiple performance metrics to ensure the predictions are accurate and reliable.

|  |  |
| --- | --- |
| Metric | Value |
| MSE (Mean Squared Error) | 8.97 |
| MAE (Mean Absolute Error) | 2.28 |
| RMSE (Root Mean Squared Error) | 2.99 |
| MedAE (Median Absolute Error) | 1.66 |

* **MSE (Mean Squared Error)**: The average squared difference between actual and predicted sales. The value of **8.97** indicates the variance between actual sales and predicted sales.
* **MAE (Mean Absolute Error)**: This is the average absolute difference between the actual and predicted sales. The **MAE of 2.28** implies that, on average, the model’s sales predictions are off by about **$2.28M**.
* **RMSE (Root Mean Squared Error)**: The square root of the average squared errors. The RMSE of **2.99M** shows that on average, the prediction error is about **$2.99M**.
* **MedAE (Median Absolute Error)**: The median absolute error shows that **half of the predictions** are within **$1.66M** of the actual sales values.

### Residual Analysis:

* A residual plot shows that the residuals (the difference between actual and predicted sales) are **centered around zero** and randomly distributed. This indicates that the model does not have any systematic bias and is well-calibrated.

A blue line graph with numbers

Description automatically generated

## 7. Findings and Business Implications

### TV Advertising Drives Sales:

* TV advertising has a significant positive effect on sales. For every additional **$1000** spent on TV ads, sales increase by **$46,000**.
* This means that TV should be considered a **primary advertising channel** for businesses seeking to boost their sales.

### Other Mediums:

* While this analysis focused solely on TV, the weak correlations between **Radio** and **Newspaper** advertising and sales suggest that these channels may not be as effective. Future studies could investigate combinations of mediums to understand their combined effect.

### Model Accuracy:

* The model has reasonable predictive accuracy, with most sales predictions falling within **$2.28M** of the actual values.
* The **RMSE** of **$2.99M** suggests that while the model works well in general, there may be other factors affecting sales that are not captured by TV ad spending alone.

## 8. Conclusion

The linear regression analysis reveals a **significant positive relationship** between **TV advertising** and **sales**. TV ads are a powerful driver of sales, with a clear return on investment of **$46,000** for every **$1000** spent. Companies should focus on optimizing their TV ad budgets to maximize sales revenue.

Future analysis could explore the combined effect of multiple advertising mediums, interactions between variables, and incorporate other factors such as seasonality, digital marketing, or product type to improve the model’s predictive power.