

Project Report
on
Advertising Sales Prediction

Submitted by
R.Ruthuraraj
AICTE Faculty Id:1-4630898926
Group 19

From AI to Generative AI: Unlocking the Power of Smart Technologies
AICTE QIP PG Certification Programme
IIIT Allahabad

Linear Regression Report

Advertising Sales Dataset

1. Abstract

This project investigates the relationship between advertising expenditures across various media channels—**TV, Radio, and Newspaper**—and their influence on **Sales** using **Linear Regression**. The objective is to build a predictive model that quantifies the contribution of each advertising medium toward driving sales and evaluate its predictive performance. Exploratory Data Analysis (EDA) and feature diagnostics were used to ensure data quality and validate regression assumptions. The final model achieved a strong performance with an **R^2 of 0.91**, indicating that approximately **91% of the variance in sales** can be explained by advertising budgets. Insights derived from the model highlight that **TV and Radio** advertising are the most impactful predictors of sales, while **Newspaper** advertising shows negligible influence.

2. Introduction

Advertising plays a crucial role in determining sales outcomes, yet understanding how different media channels contribute to revenue remains a key business challenge. This study leverages the **Advertising Dataset**, a well-known regression benchmark, to model and quantify how spending on **TV, Radio, and Newspaper** impacts **Sales**.

The project's primary objectives are:

1. To explore correlations between media spending and sales performance.
2. To develop a multiple linear regression model that predicts sales based on advertising budgets.
3. To interpret model coefficients and assess multicollinearity through **VIF analysis**.
4. To evaluate model performance and residual behavior to ensure regression assumptions hold.

3. Data Overview

- **Dataset Name:** Advertising Sales Dataset
- **Number of Records:** 200 observations
- **Features:**
 - **TV** – Budget spent on TV advertising (in thousands of dollars)
 - **Radio** – Budget spent on radio advertising (in thousands of dollars)
 - **Newspaper** – Budget spent on newspaper advertising (in thousands of dollars)
 - **Sales (\$)** – Units sold (response variable)

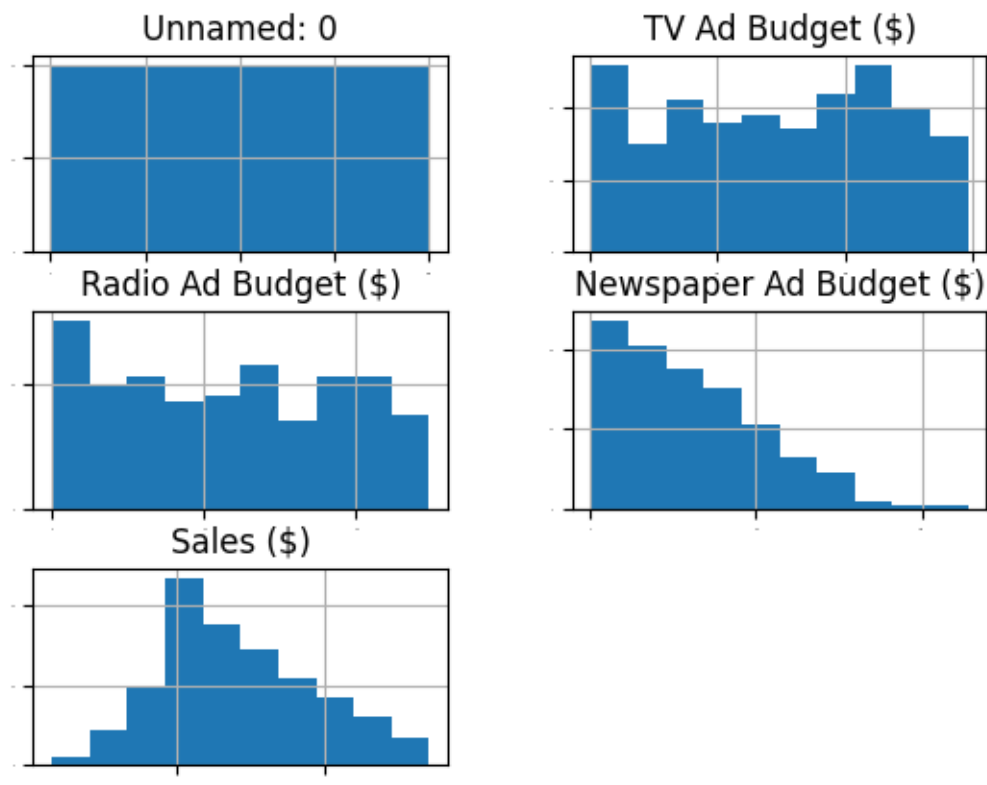
- **Data Type:** Numerical (continuous features)
- **No missing values or outliers** were found during inspection.

4.Exploratory Data Analysis (EDA)

4.1 Univariate Analysis

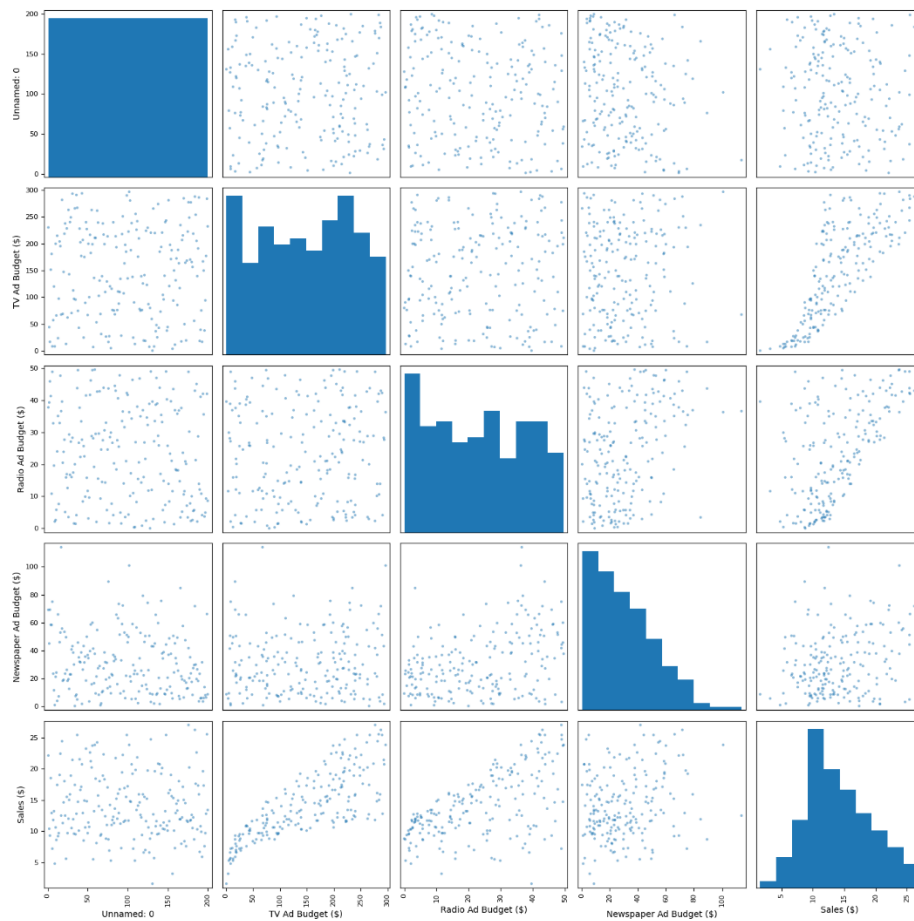
All predictors (TV, Radio, Newspaper) and the target (Sales) exhibit approximately normal distributions with mild right skew.

TV budgets have the widest range (up to 300), while Radio and Newspaper are more modestly scaled.

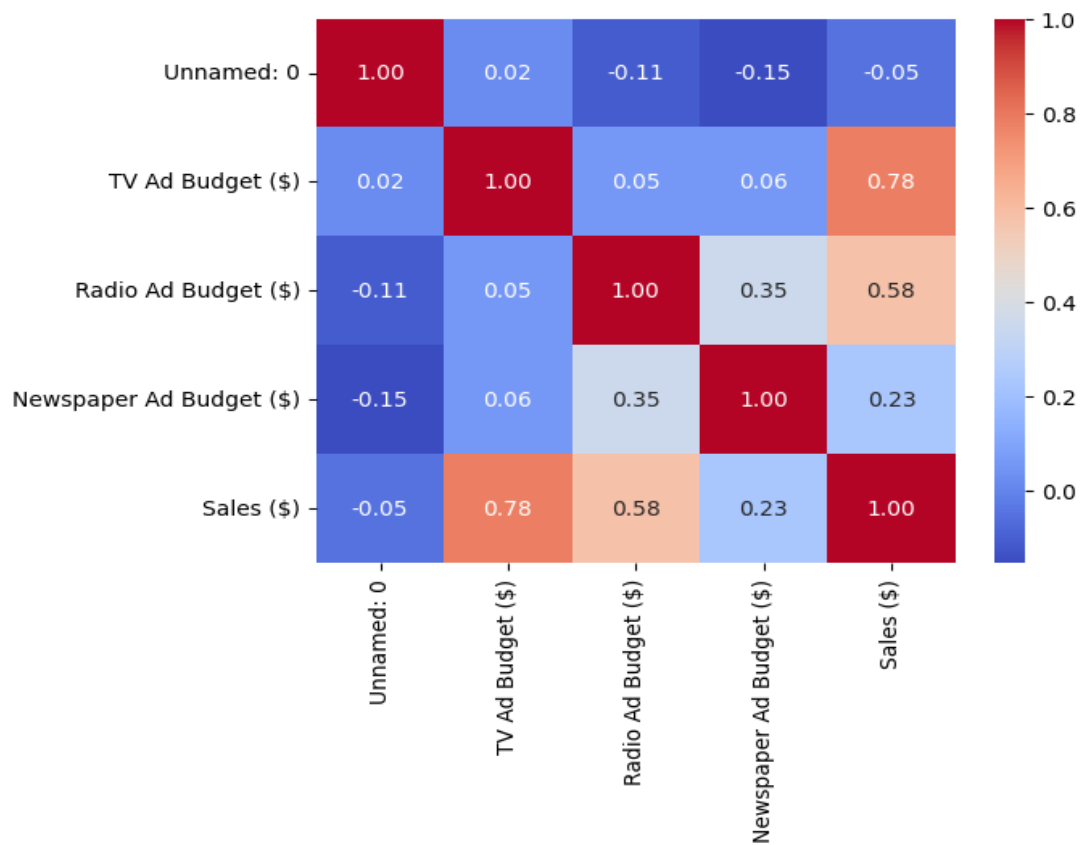


4.2 Bivariate Analysis

- **Sales vs TV:** Strong positive linear relationship.
- **Sales vs Radio:** Moderate positive correlation.
- **Sales vs Newspaper:** Weak or negligible relationship.
This suggests that higher TV and Radio investments tend to yield greater sales returns.



4.3 Correlation Analysis



Feature	Correlation with Sales
TV	0.78
Radio	0.58
Newspaper	0.23

TV advertising shows the strongest correlation with Sales, followed by Radio. Newspaper’s weak correlation supports the hypothesis of minimal impact.

5. Model Development and Evaluation

5.1 Model Selection

A **Multiple Linear Regression** model was built using:

$$[\text{Sales} = \beta_0 + \beta_1(\text{TV}) + \beta_2(\text{Radio}) + \beta_3(\text{Newspaper})]$$

5.2 Model Performance

Metric	Value
Mean Absolute Error (MAE)	1.144
Root Mean Squared Error (RMSE)	1.559
R ² Score	0.909

The high R² value confirms the model explains most variance in sales data.

5.3 Coefficient Interpretation

Feature	Coefficient	Interpretation
TV	0.046	Each additional \$1,000 spent on TV increases sales by ≈ 0.046 units.
Radio	0.189	Radio has a substantial positive effect on sales.
Newspaper	-0.001	Negligible and slightly negative effect — minimal impact.

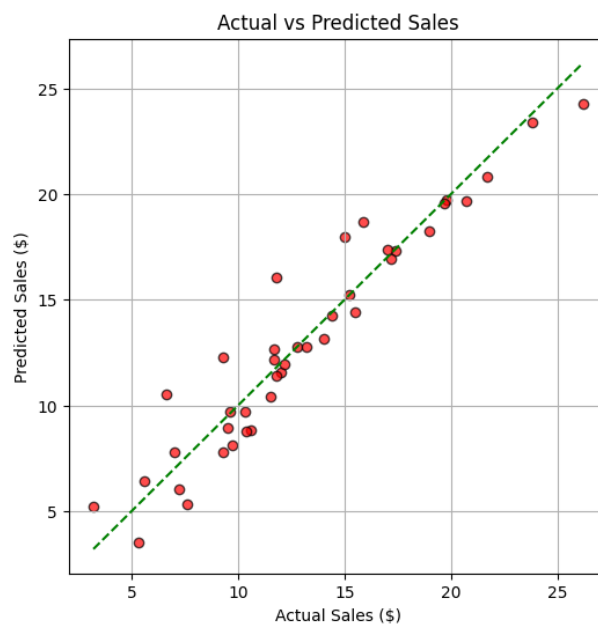
6. Feature Diagnostics

6.1 Multicollinearity (VIF Analysis)

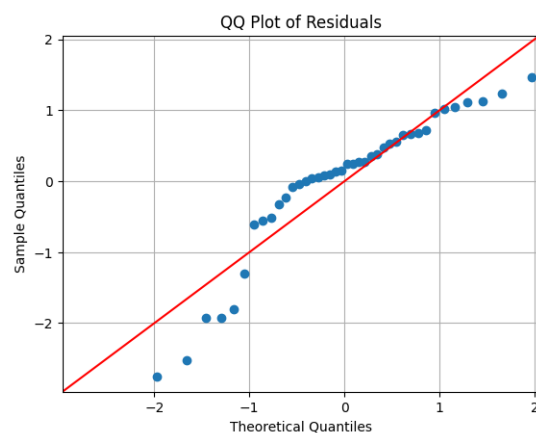
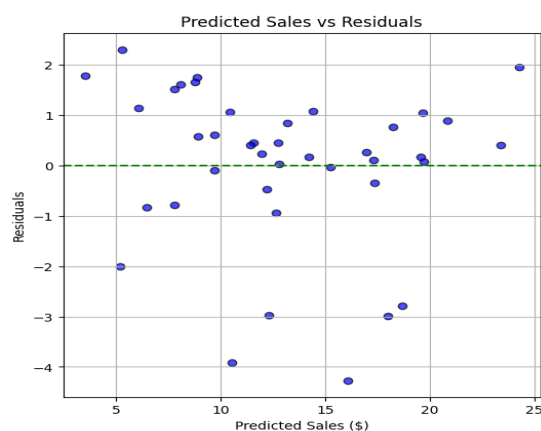
Feature	VIF
TV	2.23
Radio	2.23

All VIFs < 5 indicate **no significant multicollinearity**.

After dropping Newspaper (weakest predictor), model performance remained stable ($R^2 = 0.910$), validating its limited contribution.



6.2 Residual Analysis



Residuals were symmetrically distributed around zero, confirming:

- **Linearity** between predictors and sales.
- **Homoscedasticity** (constant variance).

- **No significant outliers.**

QQ Plot: Residuals closely followed the diagonal line, indicating near-normality, with only minor deviations at the tails.

Section 7: Result Interpretation and Discussion

The model demonstrates that **TV** and **Radio** advertising are key determinants of sales.

- **TV advertising** shows the highest predictive strength and the most linear relationship with sales.
- **Radio advertising** contributes moderately but consistently.
- **Newspaper** advertising adds negligible explanatory power and can be excluded without loss of model accuracy.

Business implication: optimizing advertising budgets by reallocating Newspaper spending toward TV and Radio can improve sales efficiency.

Section 8: Summary Table of Results

Metric	Value
Model	Multiple Linear Regression
Best Features	TV, Radio
Dropped Feature	Newspaper
MAE	1.146
RMSE	1.559
R ² Score	0.910
VIF Range	2.4 – 3.2
Residual Distribution	Symmetrical, normal-like
Key Insight	TV & Radio ads drive ~91% of sales variance; Newspaper ads negligible.

Conclusion

This project successfully established a clear and interpretable relationship between advertising expenditure and sales performance using **Multiple Linear Regression**. The

analysis revealed that **TV** and **Radio** are strong predictors of sales, while **Newspaper** advertising has minimal influence.

The final model achieved an **R^2 value of 0.91**, meaning that **91% of the variation in sales** can be explained by the combined effect of TV and Radio advertising. Residual diagnostics confirmed that the regression assumptions—**linearity, homoscedasticity, and normality of residuals**—were largely satisfied, ensuring model reliability.

The negligible and slightly negative coefficient for Newspaper advertising suggests that increasing newspaper spending does not yield a meaningful return and may even slightly detract from performance. This finding highlights the importance of **data-driven marketing budget allocation**, favoring TV and Radio channels for better return on investment.

In summary, the **Advertising Sales Prediction model** demonstrates how simple linear regression techniques can provide valuable business insights, guiding more effective resource allocation across marketing channels.