Homework #2

Group Members:
Jash Shah: js3297
Nisarg Shah: ns1452

Introduction:

This report investigates the advertising effectiveness of a specific product across 125 markets. We analyze a dataset with sales figures and corresponding advertising budgets allocated to three media channels: TV, Radio, and Newspaper. Our primary objective is to construct a precise model that predicts sales based on the advertising expenditure in each channel. This model will help us determine if there's a statistically significant association between advertising spending and sales.

This analysis employs hypothesis testing, confidence intervals, multiple regression, analysis of variance (ANOVA), and model selection criteria. We will leverage these techniques to assess the advertising expenditure's impact on sales and identify the most influential media channels for promoting the product.

Objectives:

Our analysis has three key objectives:

1. Model Building: Construct a multiple linear regression model to predict sales based on advertising budgets allocated to TV, Radio, and Newspaper media channels.
2. Association Analysis: Determine if a statistically significant association exists between advertising expenditure and sales.
3. Strategic Recommendations: Develop data-driven recommendations to optimize the advertising strategy based on the observed sales patterns in the model.

These objectives will allow us to assess the effectiveness of advertising spending on product sales and suggest improvements for future advertising campaigns.

Data Analysis:

To understand the impact of advertising on sales, we will perform exploratory data analysis (EDA) on the dataset. This includes:

1. Distribution Analysis: Examining the distribution of these variables using techniques like histograms to identify potential outliers, skewness, or non-normality.

2. Visualization: Creating scatter plots to visualize the relationships between advertising budgets for each media channel and sales. This can reveal potential correlations or trends.

3. Identifying Insights: Drawing critical insights from the EDA to understand the data's characteristics and potential relationships between advertising expenditure and sales.

These analyses will help prepare the data for model building and provide a foundation for interpreting the model's results.
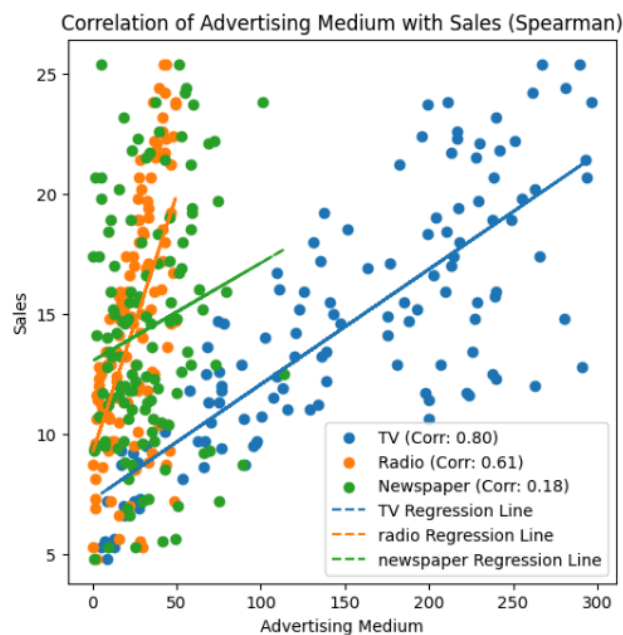
Modeling Analysis and Results:

This section discusses the methodology for constructing predictive models to estimate sales based on TV, Radio, and newspaper advertising budgets. We present the results of the constructed models, including regression coefficients, significance levels, and model fit statistics. Additionally, we evaluate the predictive capabilities of the models using appropriate evaluation metrics and provide insights gained from the modeling process.

● Methodology for Model Construction:

The methodology for model construction involved several steps:

1. Feature Selection: Initially, all three advertising mediums (TV, Radio, and Newspaper) were considered as predictors for sales. Subsequently, backward elimination was employed to select the most significant features based on p-values.

2. Model Training: Simple linear regression models were fitted for each advertising medium to predict sales. Additionally, multiple linear regression models were constructed using combinations of advertising budgets to assess their collective impact on sales.

3. Model Evaluation Criteria: The performance of the models was evaluated using key metrics such as the coefficient of determination (R-squared) and mean squared error (MSE).

- Results of the Constructed Models:

The results of the constructed models are plotted below:

Sales vs. Advertising Budgets (Individual Linear Regressions)

Sales vs. Advertising Budgets (Multiple Regression)
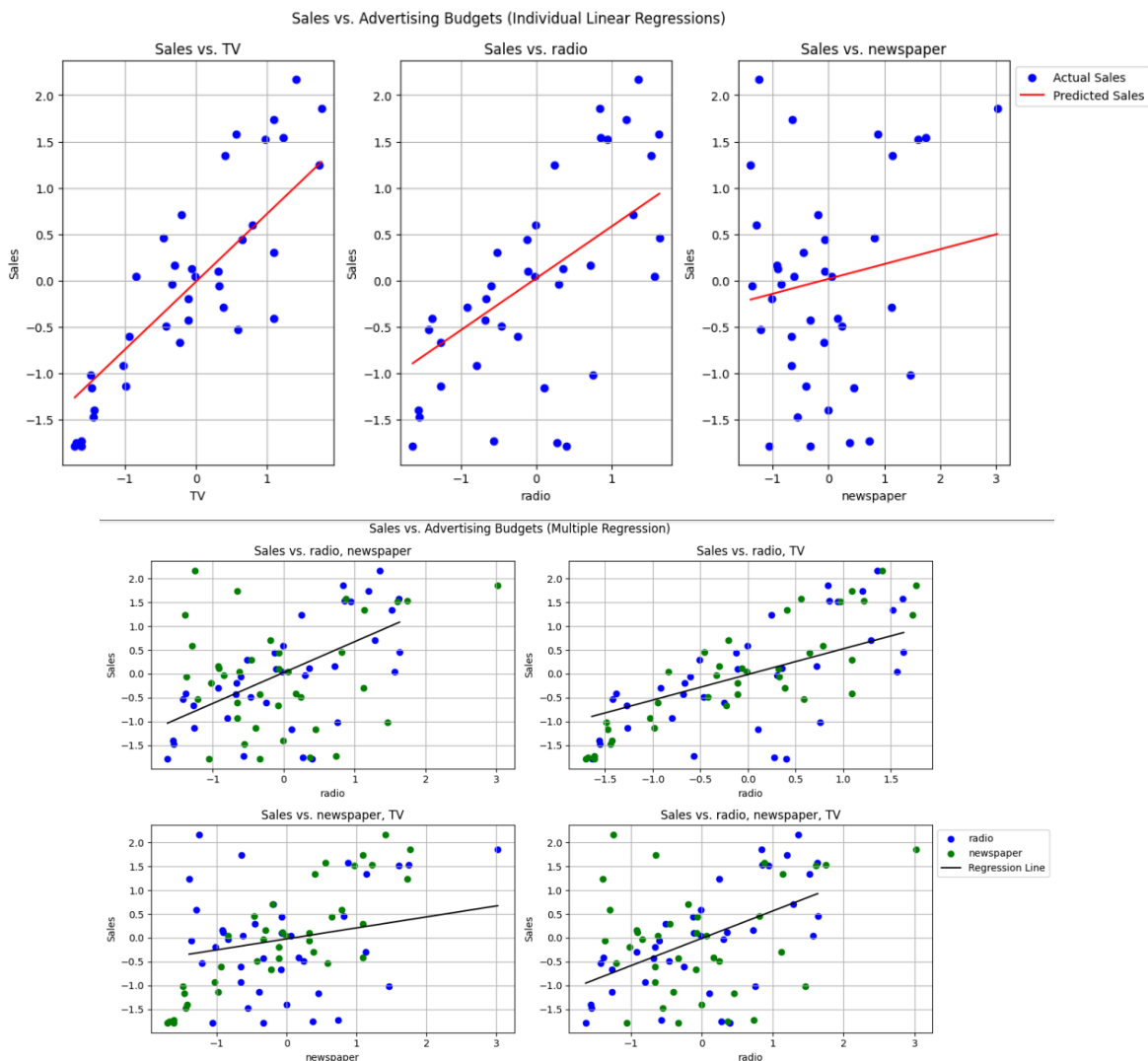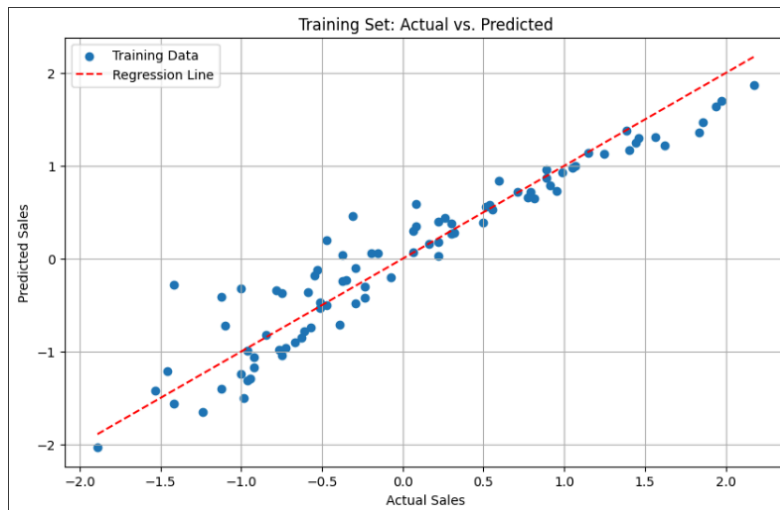
```
Simple Linear Regression (Sales vs TV)
Feature: TV
R-squared: 0.7100277855344144
Mean Squared Error: 0.3512020667783605

Simple Linear Regression (Sales vs radio)
Feature: radio
R-squared: 0.41482803584709405
Mean Squared Error: 0.7087355028481322

Simple Linear Regression (Sales vs newspaper)
Feature: newspaper
R-squared: 0.035136974554365596
Mean Squared Error: 1.1686012376014923
```

Conclusions and Findings:

- Main Findings:

1. Significant Associations: TV and radio advertising budgets exhibited stronger correlations with sales than newspaper advertising.

2. Optimal Budget Allocation: Allocating more budgets to TV and radio may lead to higher sales outcomes.

3. Predictive Capabilities: Models demonstrated varying degrees of predictive power, emphasizing the importance of appropriate predictors and evaluation criteria.