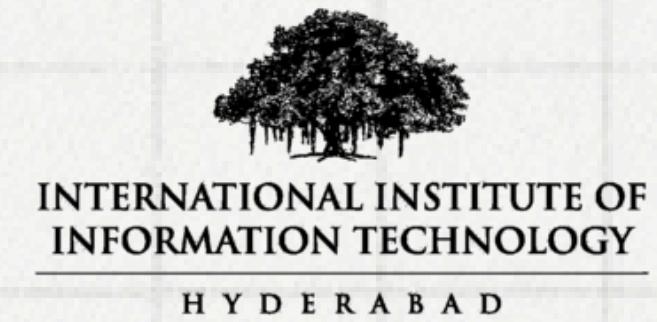


kaggle™

# Advertising Dataset

Mini Project 5

Vinit Mehta – 202211001



# About Dataset

**Published on:** [Kaggle](#)

**Year of Publish:** 2018–2019

**Link:** <https://www.kaggle.com/datasets/ashydv/advertising-dataset/data>

**Authors:** [Ashish](#)

This dataset shows the revenue generated from **sales** based on the amount spent on different types of **advertisements**.

The dataset contains **3 features TV, Radio, Newspaper** and target variable **Sales**.

Features TV, Radio, Newspaper are in unit of 1000\$ and target variable Sales in M\$.

Total number of **datapoints = 200** and there are **no missing data** in any of the columns.

BRSM: Advertising Dataset: Descriptive Statistics and Visualization

Vinit Mehta

## BRSM Report 1

Descriptive Statistics and Visualization of Advertising Dataset

21 February 2025

Name Vinit Mehta  
 Roll Number 2022111001  
 Mini Project 5 (Advertising Dataset)

### 1 Introduction

**Name of Dataset:** Advertising Dataset  
**Link to Dataset:** [kaggle](#)  
**About Dataset:**

- This dataset shows the revenue generated from sales based on the amount spent on different types of advertisements.
- The dataset contains 3 features *TV, Radio, Newspaper* and target variable *Sales*.
- Features *TV, Radio, Newspaper* are in unit of 1000\$ and target variable *Sales* in M\$.
- Total number of datapoints = 200 and there are no missing data in any of the columns.

#### • Assumptions:

1. The data is for different companies around the same time period.
2. All the companies belong to same sector/industry.

**Background:** Advertising has played a pivotal role in commerce since ancient times, evolving into a dominant force in capitalist economies by the mid-19th century, largely driven by newspapers and magazines [1]. The 20th century witnessed a rapid expansion of advertising across emerging media, including direct mail, radio, television, the internet, and mobile platforms. In the United States, advertising expenditure consistently averaged 2.2% of the Gross Domestic Product (GDP) between 1919 and 2007 [1].

The progression of advertising mediums follows a historical trajectory. The first newspaper advertisement in the United States appeared in the *Boston Newsletter* in 1704 [2], the first radio advertisement aired on August 22, 1922 [3], and the first paid television advertisement was broadcast on July 1, 1941, over New York's *WNBT* station [4]. The 1990s ushered in the era of digital marketing, leveraging search engines and database-driven customer targeting [5].

This dataset comprises 200 observations and explores the relationship between advertising expenditures across various media and sales performance. Historically, as new advertising mediums gained prominence, businesses adapted their spending patterns to optimize outreach and revenue generation. The dataset, created around 2018–2019, offers valuable insights into advertising expenditure trends and their impact on sales during that period.

Advertising is instrumental in driving product awareness and influencing consumer behavior. Businesses allocate budgets across multiple advertising channels such as television (TV), radio, and newspapers to maximize their market reach and sales potential. Understanding the impact of these expenditures on sales is critical for refining marketing strategies and enhancing return on investment (ROI). This study analyzes the dataset using various visualization techniques, descriptive statistics, and inferential statistical methods to determine the most effective advertising channel and provide strategic insights into advertising optimization.

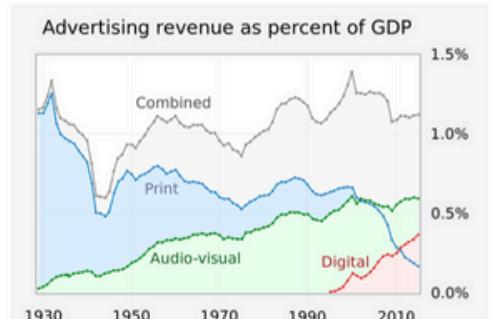


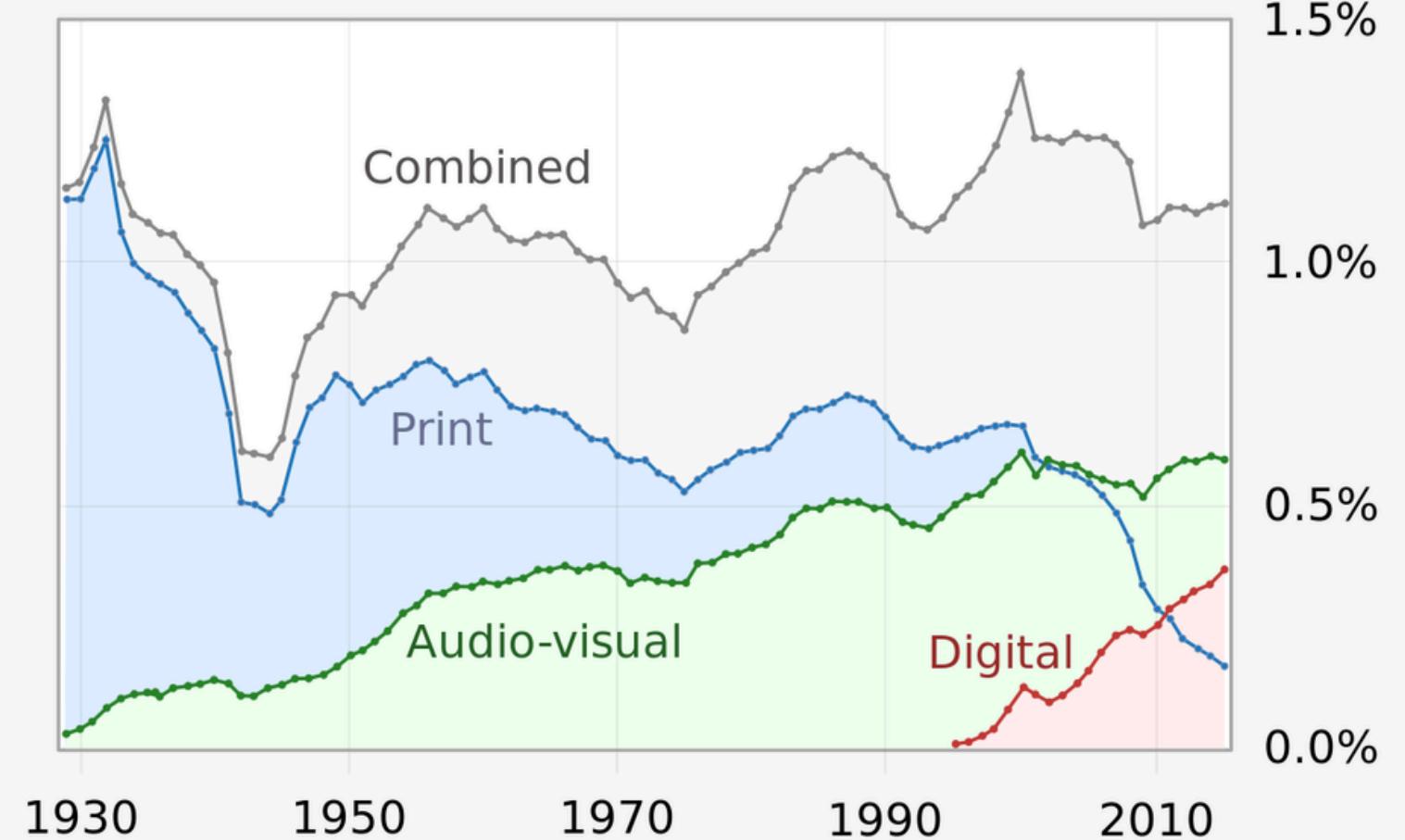
Figure 1: Advertising revenue as a percentage of US GDP (1929–2010) [1].

21 February 2025

# Background

**Advertising** has evolved from ancient commerce to a dominant force in capitalist economies, expanding across newspapers, radio, television, and digital platforms. In the U.S., ad spending averaged 2.2% of GDP between 1919 and 2007. This study analyses a dataset of **200 observations (2018–2019)** to explore the **relationship** between advertising **expenditures** across TV, radio, and newspapers and their **impact on sales**. By leveraging **visualisation** techniques and **statistical analysis**, the study identifies the most effective advertising channels, providing strategic **insights** for optimising marketing investments and maximising ROI.

Advertising revenue as percent of GDP



Advertising revenue as a percentage of US GDP (1929–2010)

# Project R1 vision and mission

To analyse the relationship between advertising expenditures across different media channels—TV, radio, and newspapers—and their impact on sales performance. This study aims to provide insights into optimising marketing investments by identifying the most effective advertising strategies through data-driven analysis.

01.

Used various **visualisation** technique like histograms, box plots, violin plots, pie charts, and QQ plots to explore the dataset, along with descriptive statistical analysis.

02.

Conducted **normality tests** to determine whether the features follow a normal distribution and, based on the results, calculated correlation metrics.

03.

**Answered key analytical questions** by interpreting initial visualisations and statistical checks to derive meaningful insights from the data.

# Methods

- Pie Chart
- Histogram
- Box Plot
- Violin Plot
- Scatter Plot
- A2S Ratio
- Q-Q Plot
- Normality Test
- Correlation
- Bar Graph
- Line of best fit
- Descriptive Stats



# Descriptive Statistics

Metric	TV Cost (1000\$)	Radio Cost (1000\$)	Newspaper Cost (1000\$)	Sales (M\$)
Mean	147.042	23.264	30.554	14.022
Median	149.75	22.9	25.75	12.9
Std Dev	85.639	14.810	21.724	5.204
Min	0.7	0.0	0.3	1.6
Max	296.4	49.6	114.0	27.0

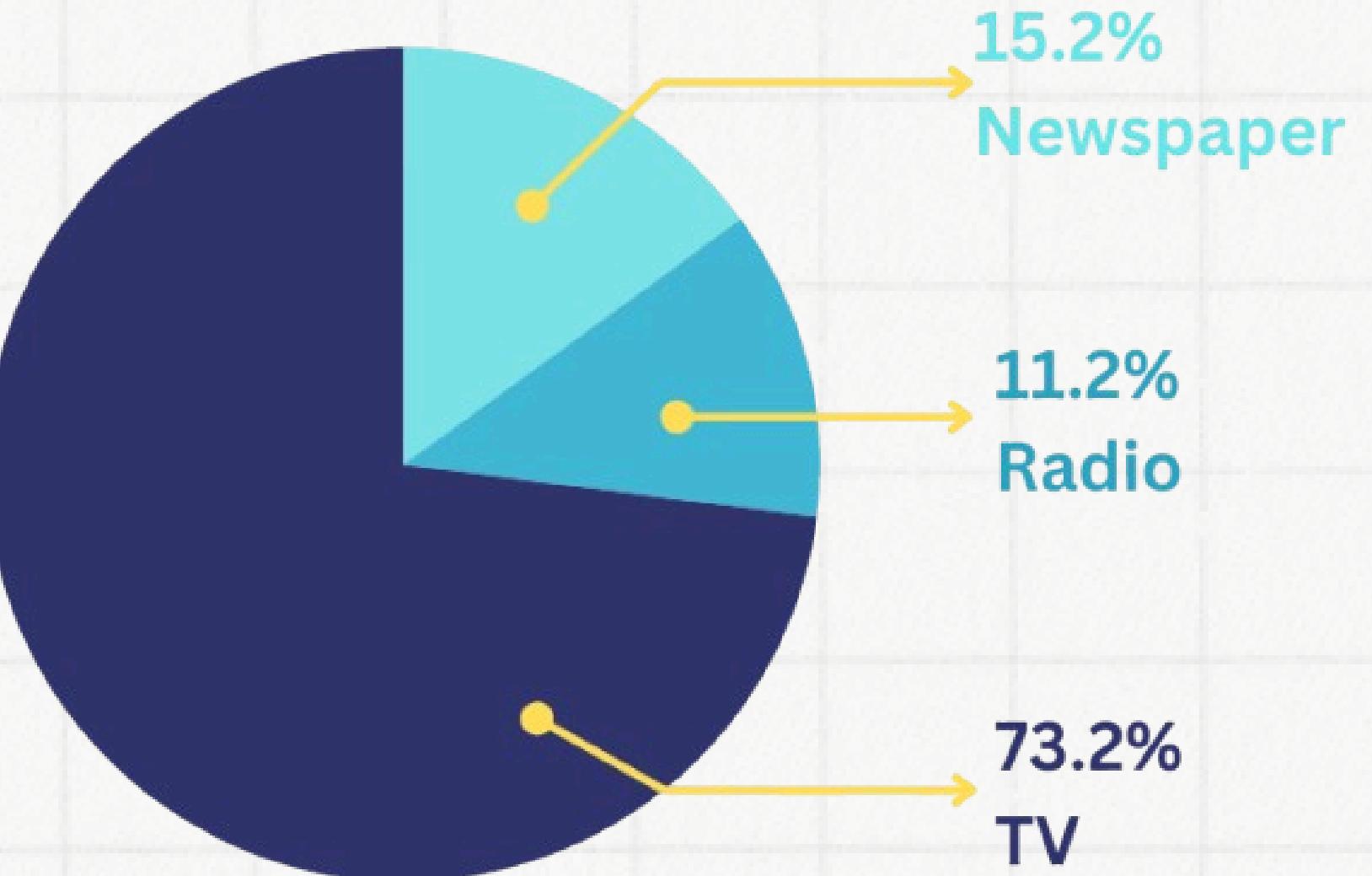


# Visualisations



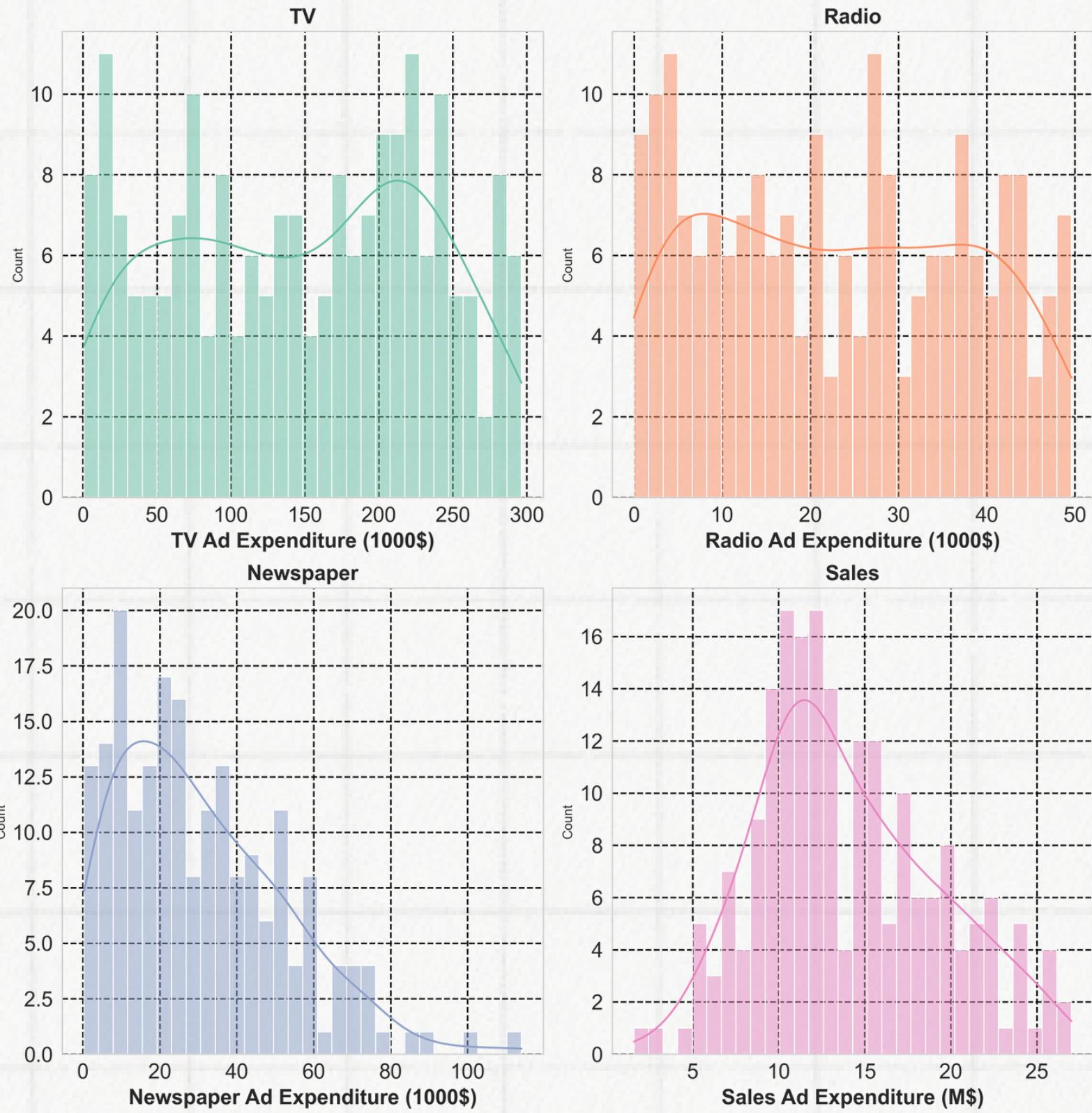
# Pie Chart

A pie chart visually **represents** **proportions** within a dataset by displaying categories as slices of a circle. On average, TV ads receive the highest investment at 73.2%, followed by newspapers at 15.2%, and radio at 11.2%, indicating a **strong preference for TV advertising** across all data points.



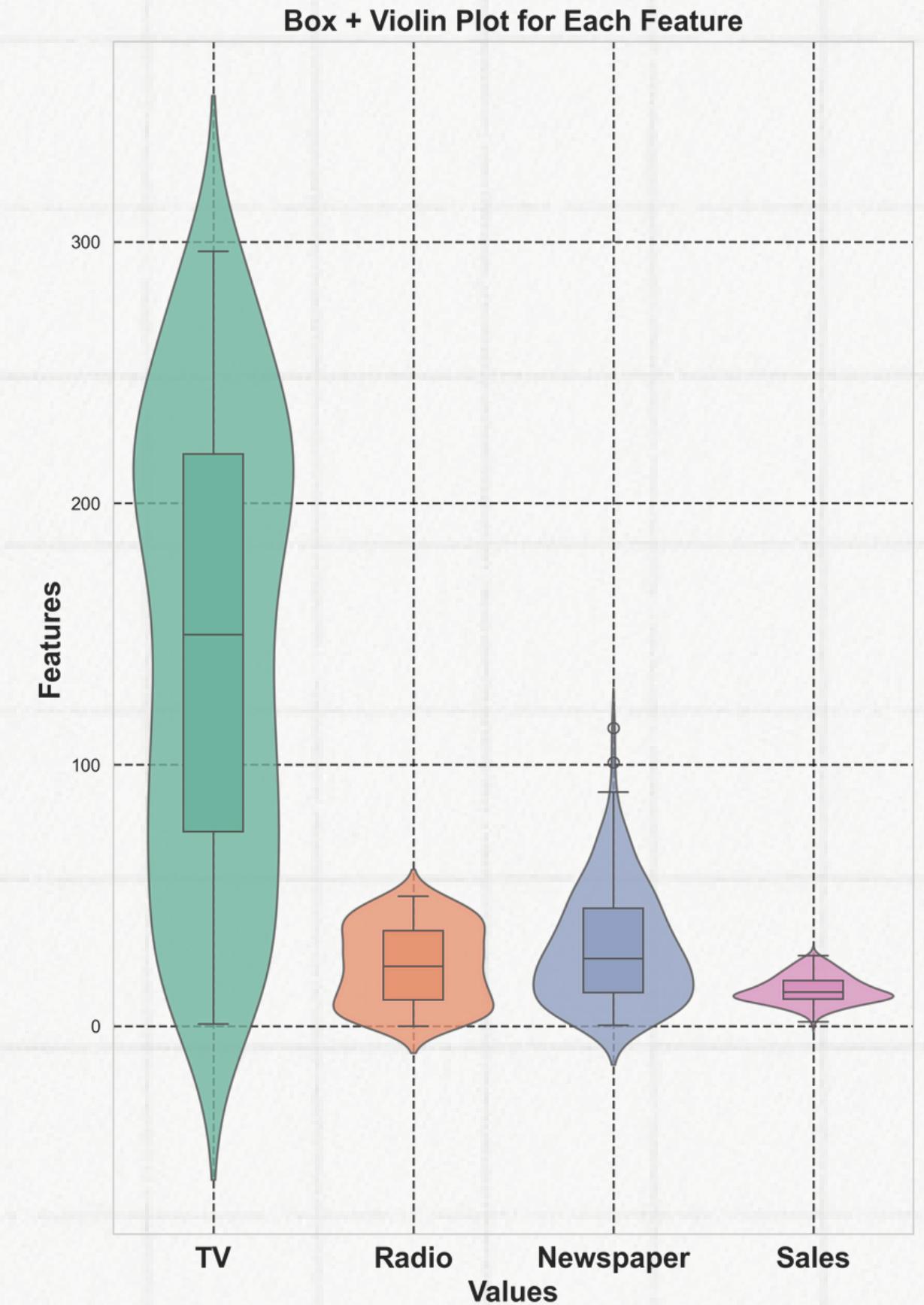
# Histogram

Histograms illustrate the distribution of numerical variables by segmenting continuous data into **bins** and counting the frequency of observations. Here histograms show **TV ads** have the **highest and most consistent** investment, while **newspaper** budgets are right-skewed, concentrated in the **lower range**. TV and radio spending are evenly distributed, whereas **newspaper and sales** data exhibit **right-skewed** patterns.



# Box+Violin Plot

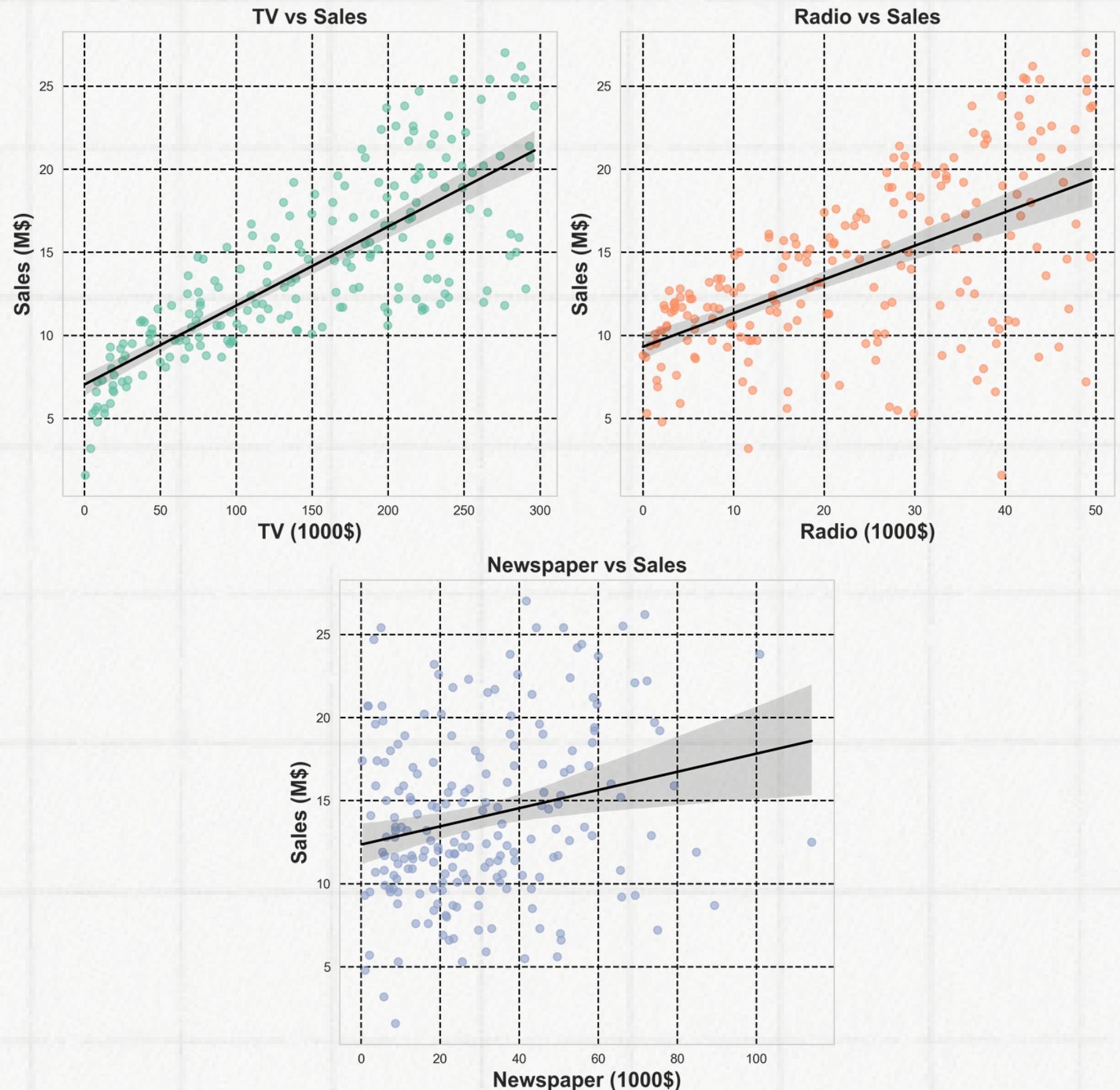
Box and violin plots visualize data distribution, with **box plots showing quartiles and violin plots adding density estimates**. TV ad spending is widely distributed with a **large IQR**, while **radio ads** are concentrated in the **lower range**. **Newspaper ads** are **right-skewed**, with most spending low and a few high-budget outliers. **Sales** show **moderate revenue** with **minimal variation**.



# Scatter Plot

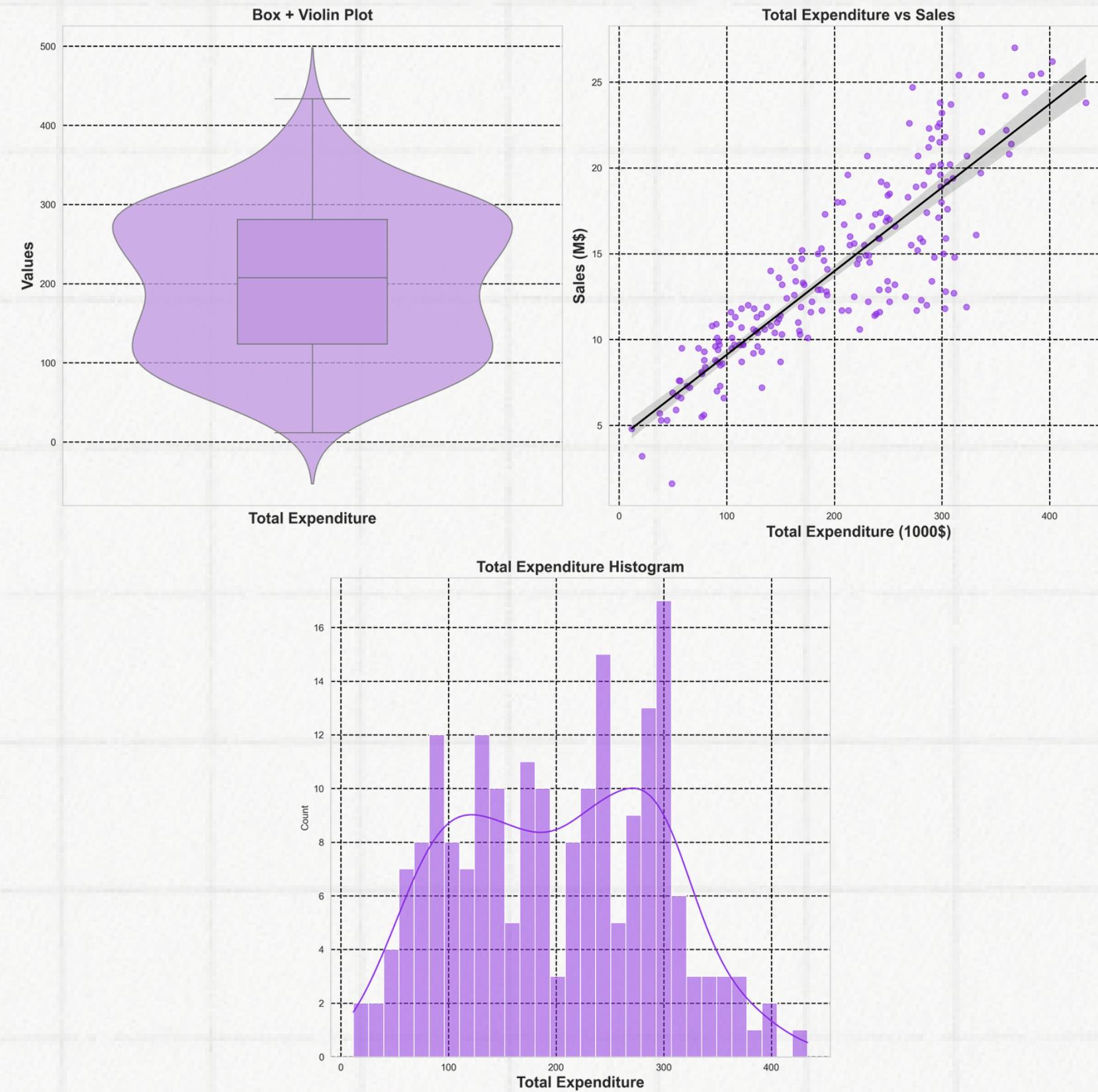
A scatter plot visualises **relationships between variables**, while the line of best fit models these relationships by **minimising prediction error**.

TV ads have the **strongest** and most direct impact on sales, making them the **most effective** channel. Radio ads contribute to sales but show variability, leading to **inconsistent revenue growth**. Newspaper ads have the **weakest impact**, suggesting they are the least reliable for driving sales.



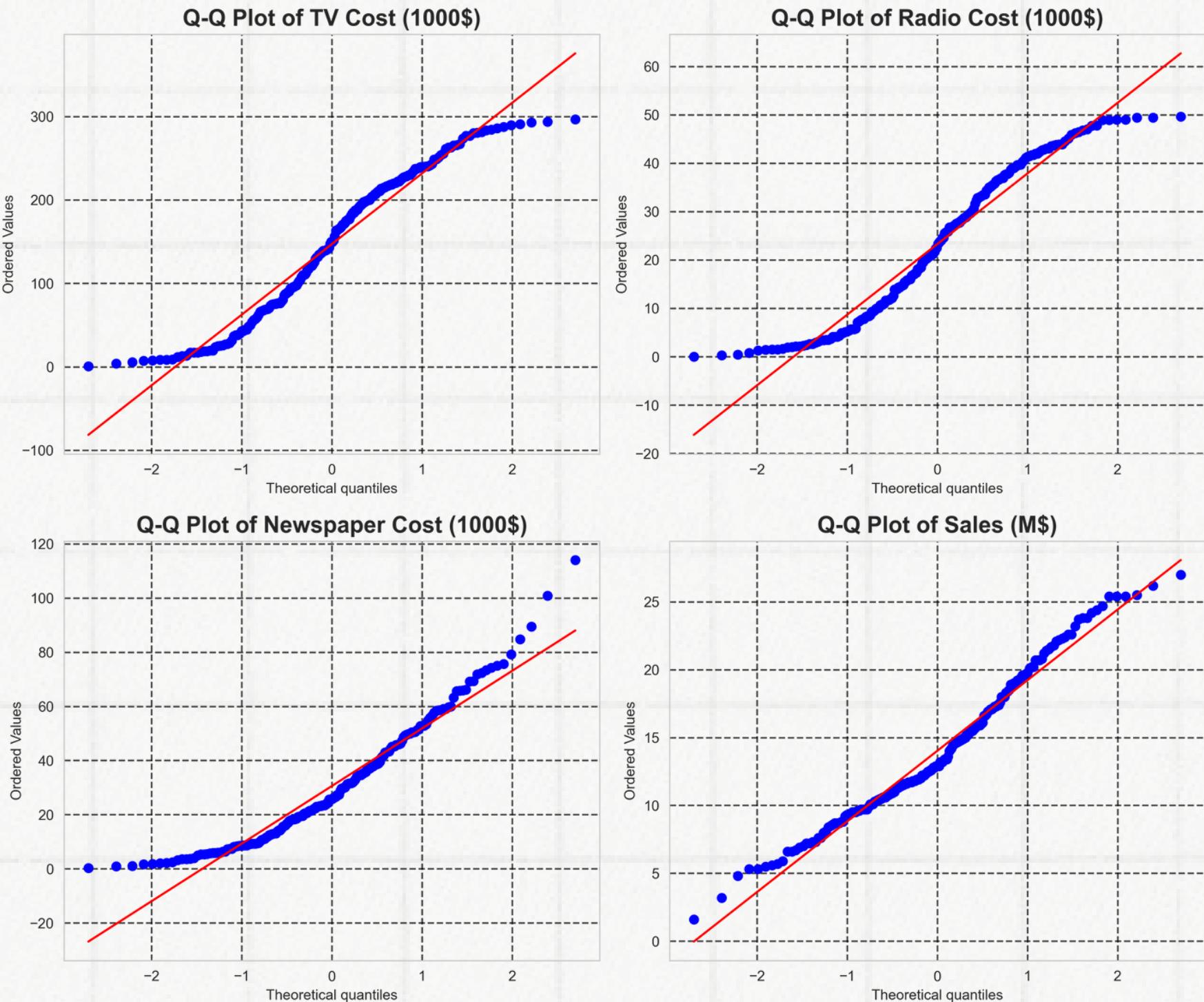
# Total Expenditure Visualisation

The distribution of total advertisement expenditure is **spread across a wide range**. Total expenditure shows a **strong positive correlation with sales**, confirming that higher ad spending boosts sales. The histogram reveals a **slightly right-skewed** distribution of total expenditure.



# Q-Q Plot

A Q-Q plot compares dataset quantiles to a normal distribution to **assess normality**. TV, Radio, and Newspaper costs show **right-skewed** distributions with **heavy left tails** and high-value outliers. Sales data aligns closely with the **normal** distribution, with minor deviations at the extremes. The **curvature in higher quantiles highlights potential outliers**, particularly in newspaper expenditures.



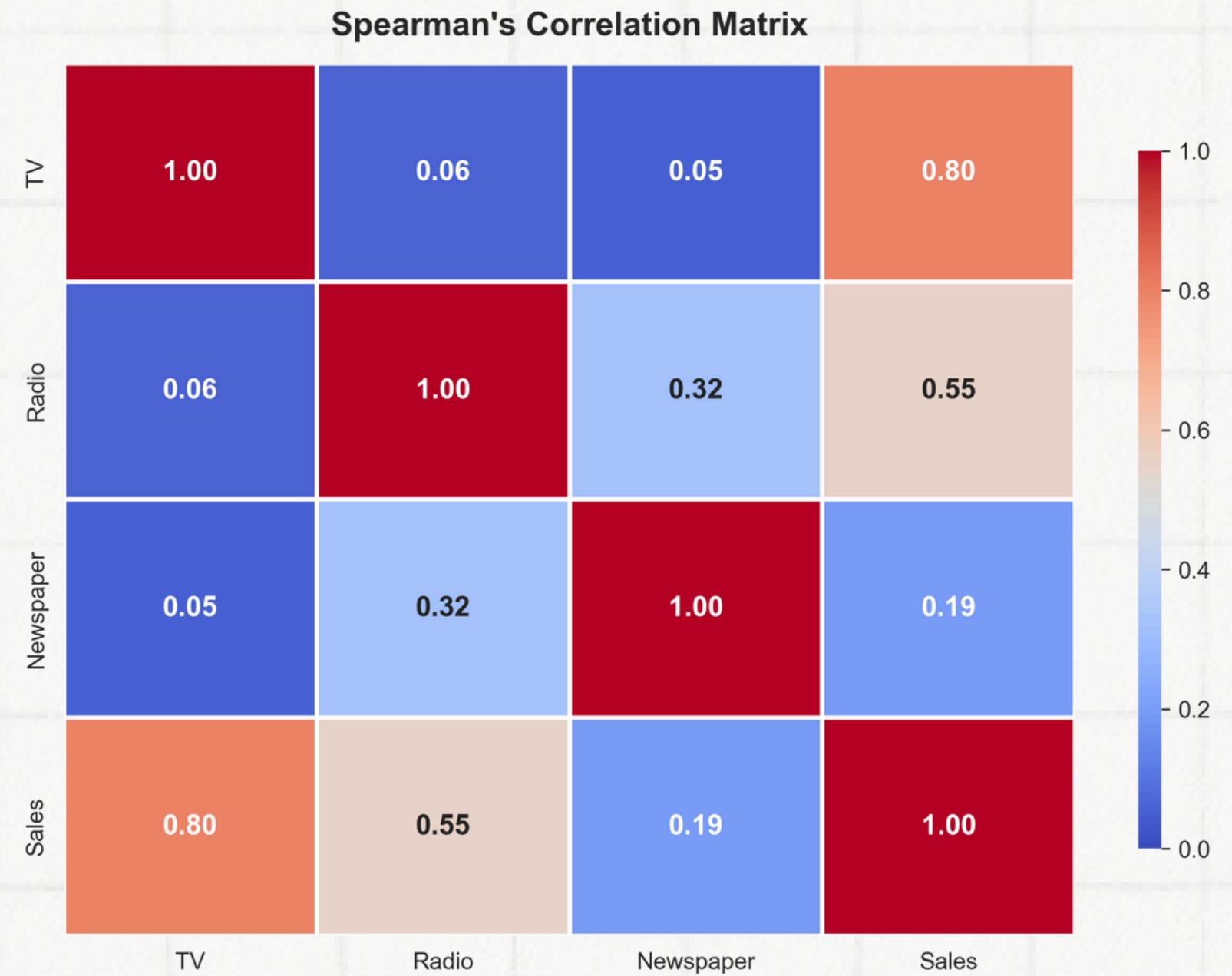
# Normality Test

TV and Radio ad expenditures follow a **normal distribution**, while Newspaper expenditures and overall sales **do not**. Normality tests like the **Kolmogorov-Smirnov test** (for  $n \geq 50$ ) and Shapiro-Wilk test (for  $n < 50$ ) confirm these assumptions. If data is not normally distributed, transformations or non-parametric tests may be used for further analysis.

```
1 For TV:  
2 KS Statistic: 0.087  
3 P-value: 0.0911  
4 Fail to reject the null hypothesis: Sample follows a normal distribution.  
5  
6 For Radio:  
7 KS Statistic: 0.084  
8 P-value: 0.1126  
9 Fail to reject the null hypothesis: Sample follows a normal distribution.  
10  
11 For Newspaper:  
12 KS Statistic: 0.0985  
13 P-value: 0.0384  
14 Reject the null hypothesis: Sample does NOT follow a normal distribution.  
15  
16 For Sales:  
17 KS Statistic: 0.0952  
18 P-value: 0.0499  
19 Reject the null hypothesis: Sample does NOT follow a normal distribution.
```

# Correlation Matrix

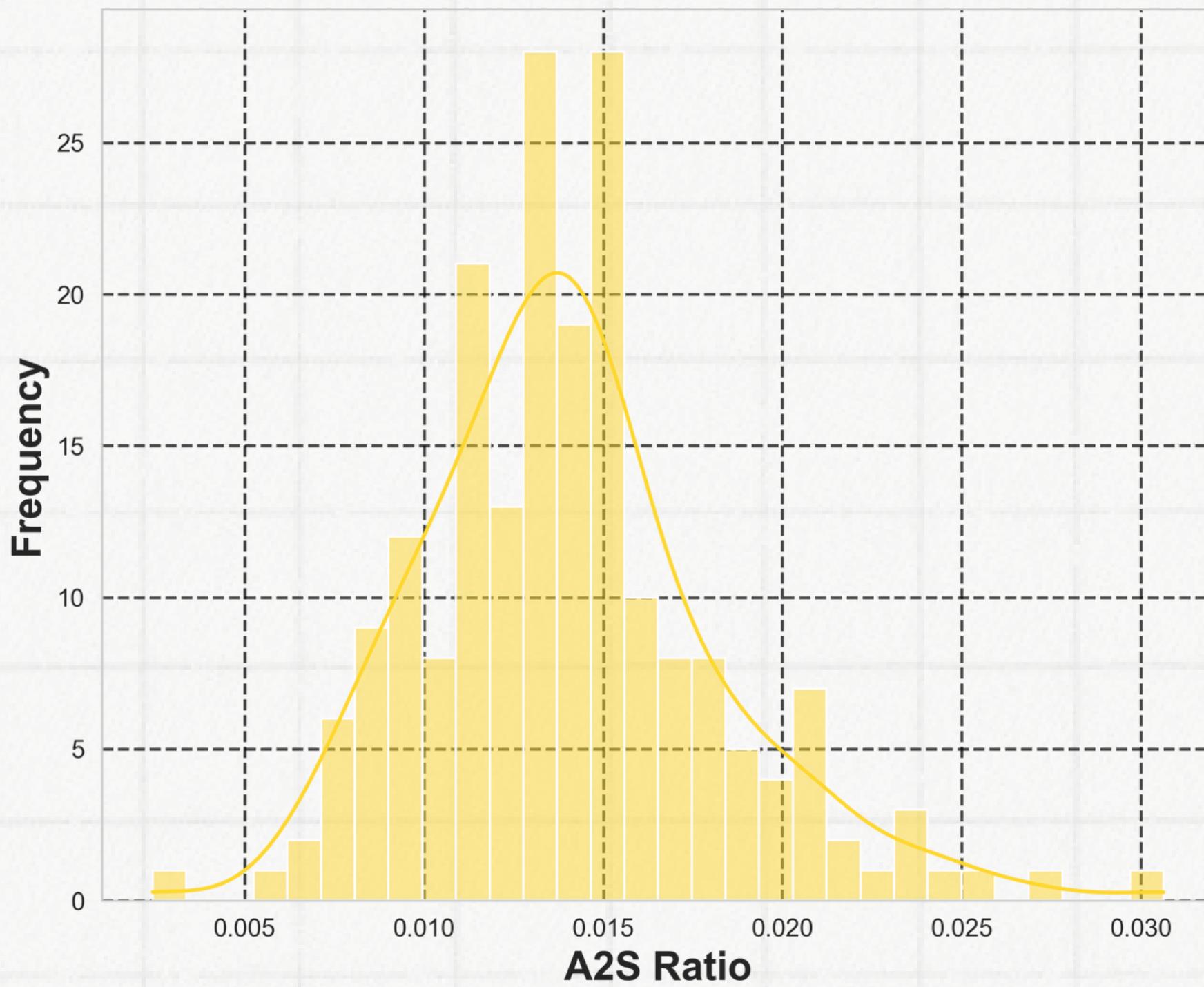
Correlation **measures the strength and direction** of relationships between variables. Spearman's correlation is used for **non-normal data**, while Pearson's assumes **normality**. TV ad spending has the **strongest correlation with sales** (0.80), followed by radio (0.55), while newspaper shows the weakest (0.19). TV ads have **minimal correlation with other channels**, suggesting independent budget allocations, while newspaper and radio show a mild correlation (0.32).



# Ad to Sales Ratio

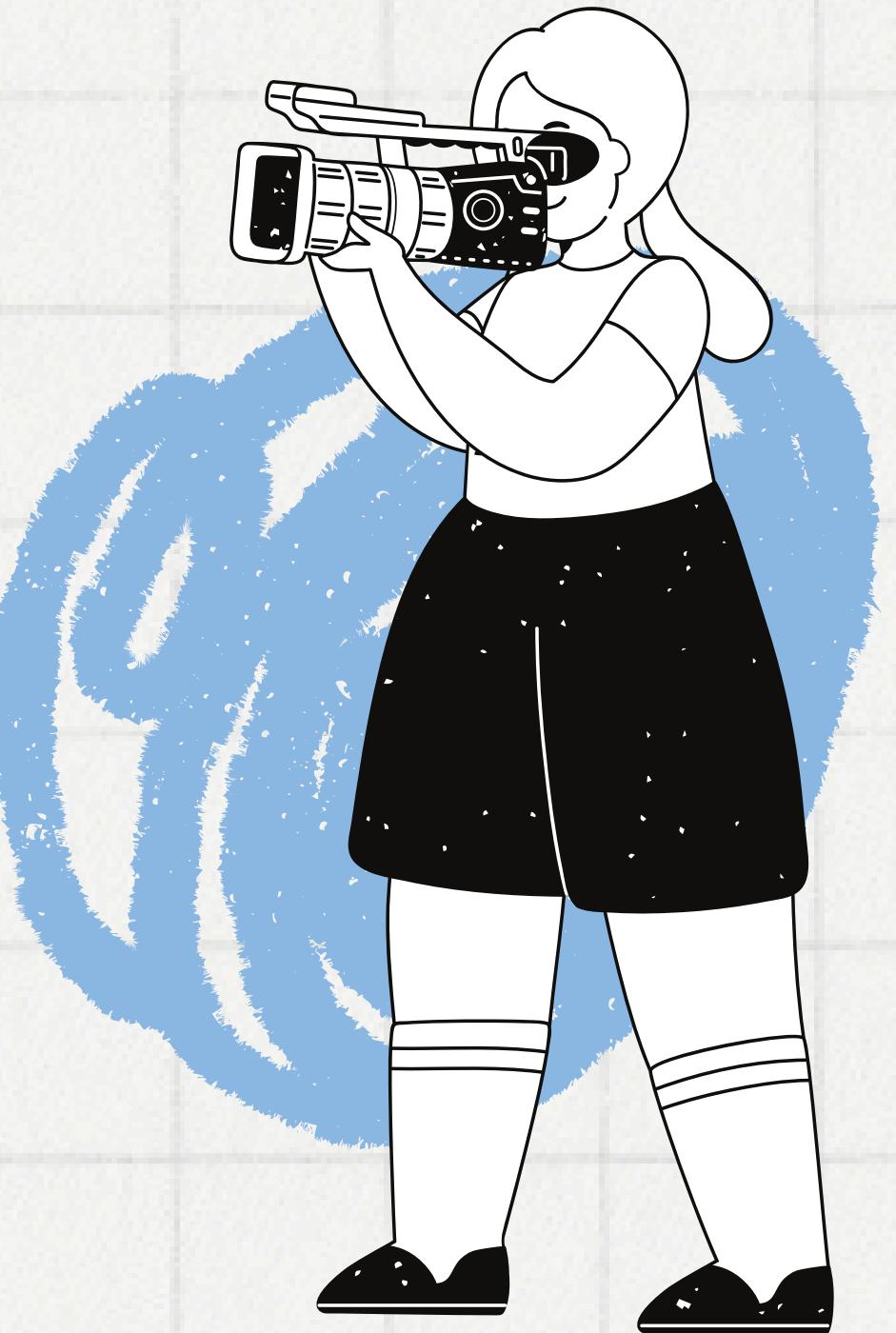
The Advertisement-to-Sales (A/S) Ratio measures the efficiency of ad spending in generating revenue, with lower values indicating higher efficiency. Most companies have an A/S ratio around 0.014, with few extreme values. Advertisements to sales ratio is a way to assess ROI.

Advertisement-to-Sales Ratio Distribution





# Questions & Answers

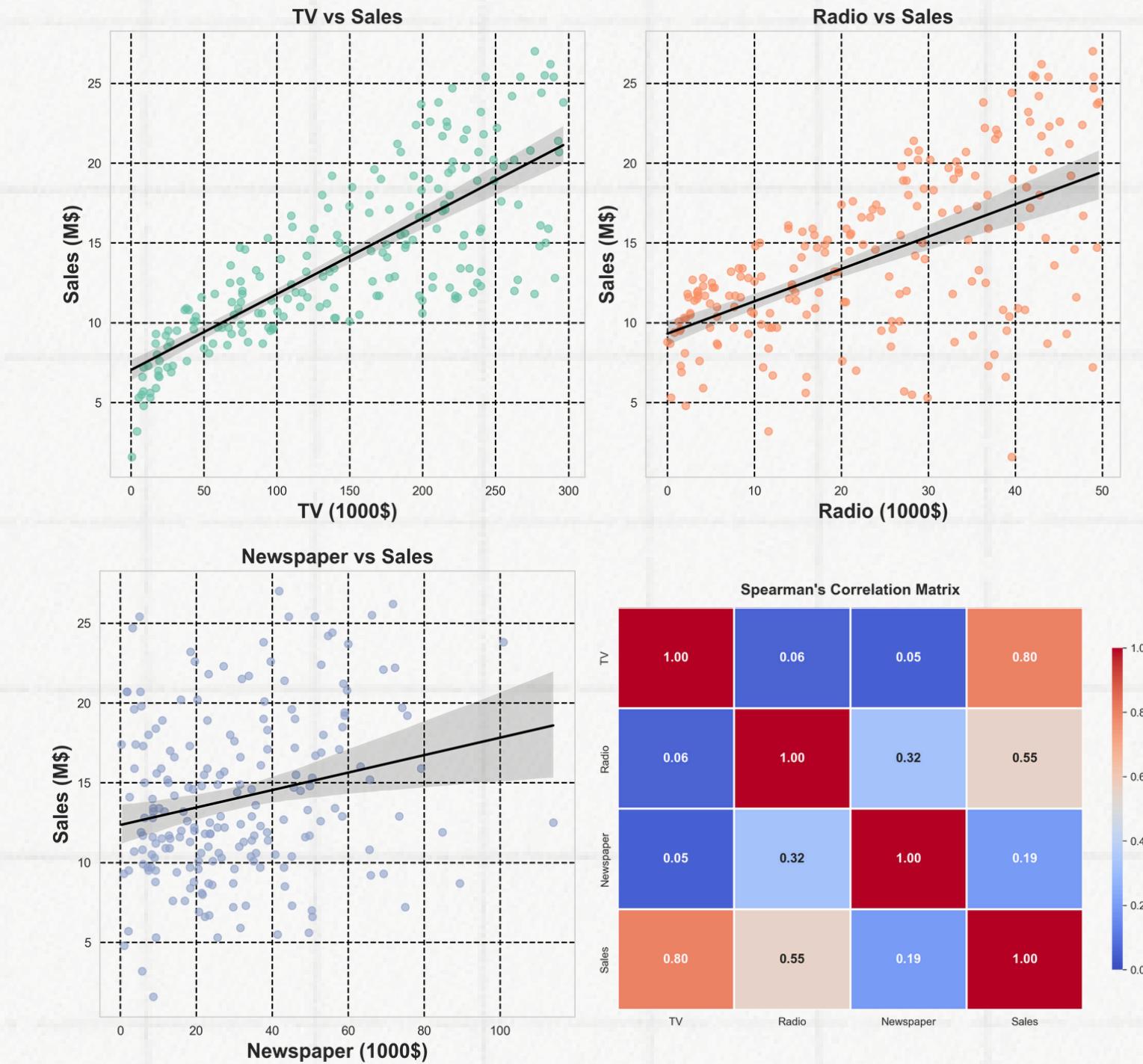


# Q1

Which platform yields the highest return on investment for advertising campaigns?

# A1

From the correlation heatmap and scatter plots, TV advertising shows the strongest positive correlation with sales. The regression line in scatter plots 5 also suggests that TV ad spend has the most direct impact on sales compared to Radio and Newspaper. Hence we can say that TV advertising provides the highest return on investment (ROI) among the three platforms.

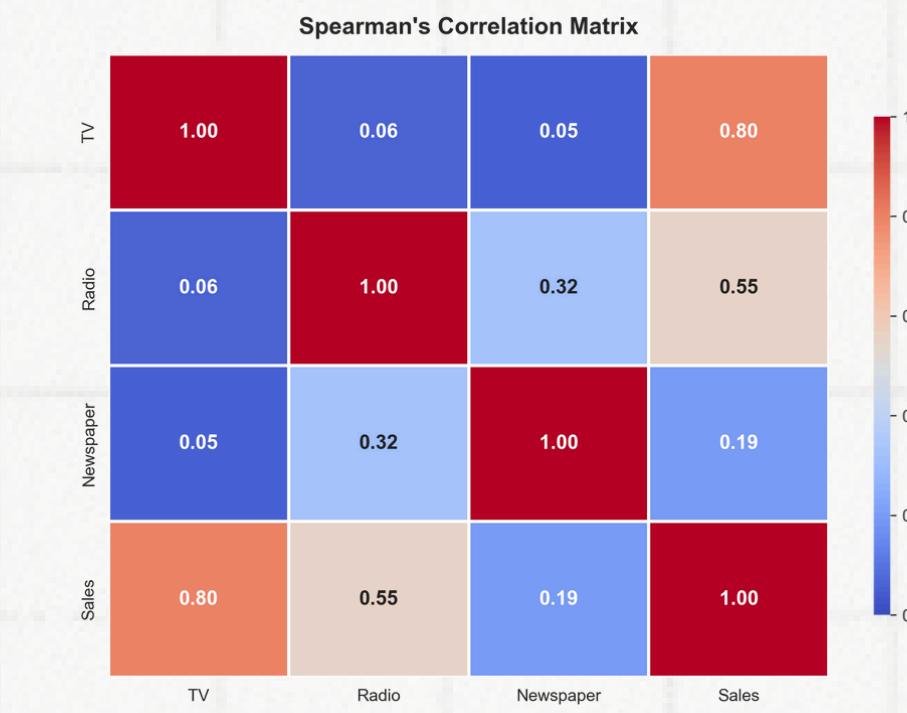
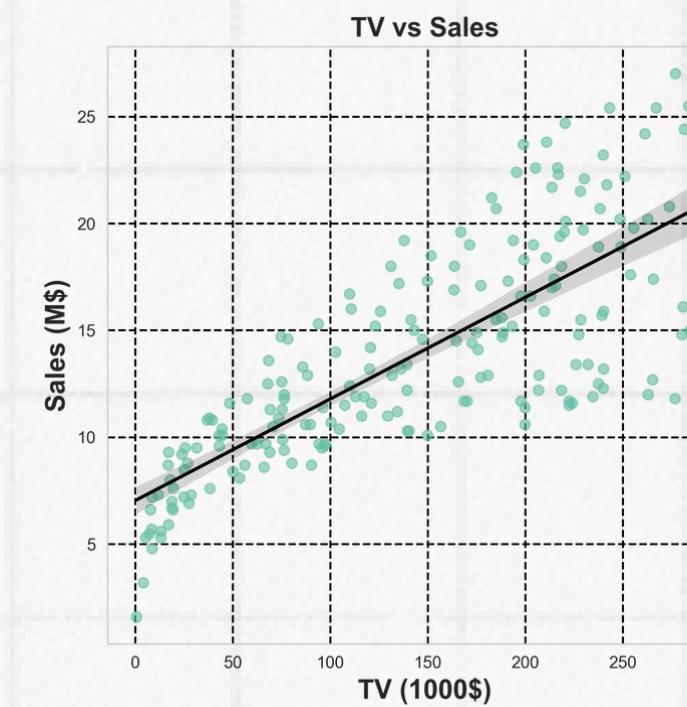


# Q2

**Is there a general trend between TV advertising spend and sales?**

# A2

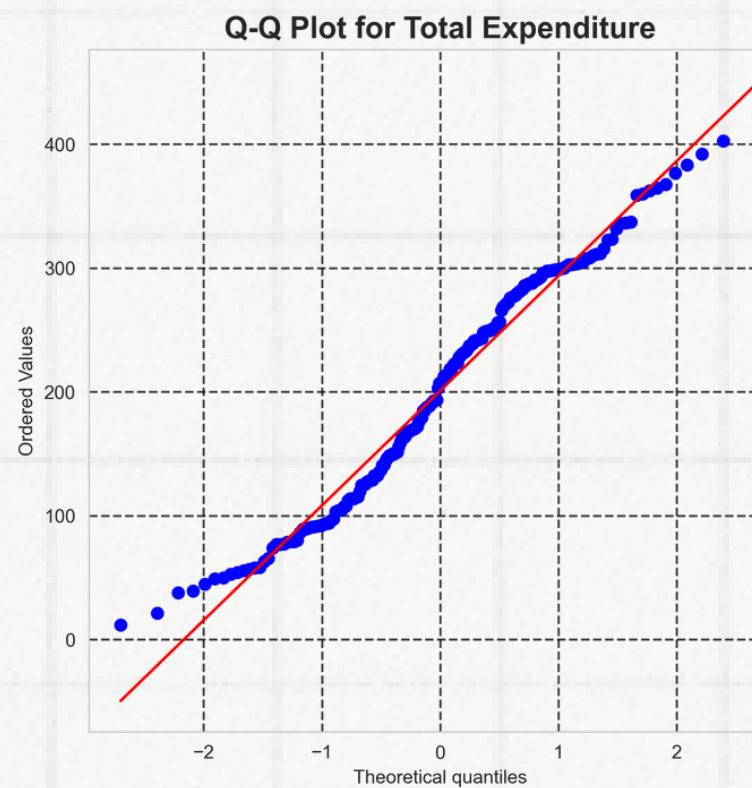
The scatter plot (TV vs. Sales) with regression line shows a clear upward trend, meaning higher TV ad spend generally leads to higher sales. The strong correlation coefficient (0.80) between expenditure on TV Ads and Sales also supports this relationship. Hence we can say that there is a strong positive trend between TV advertising spend and sales.



# Q3 What is the relationship between ad expenditure and sales?

## A3

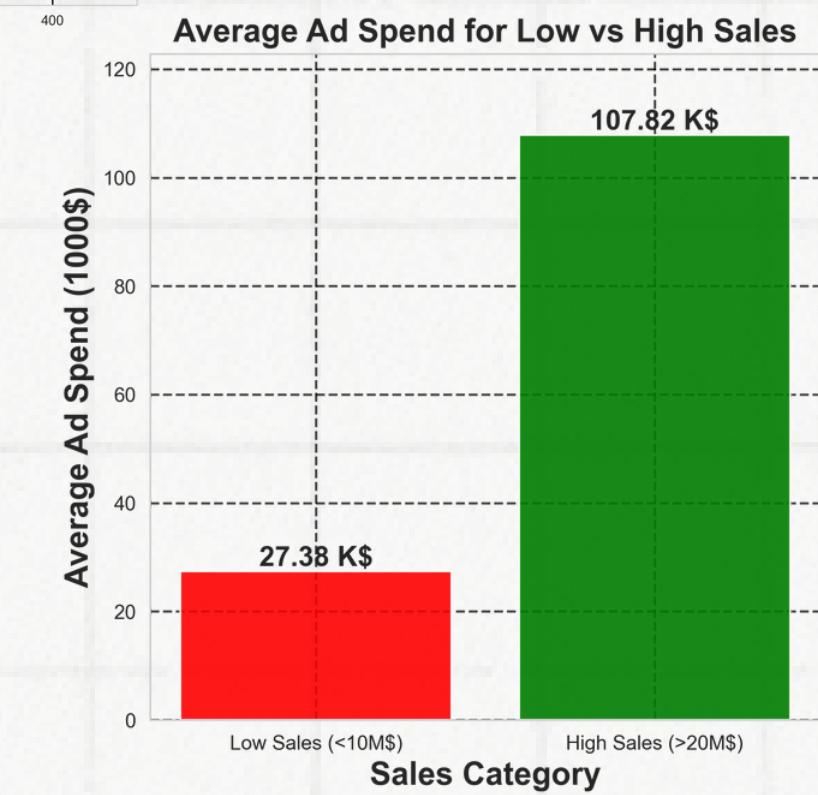
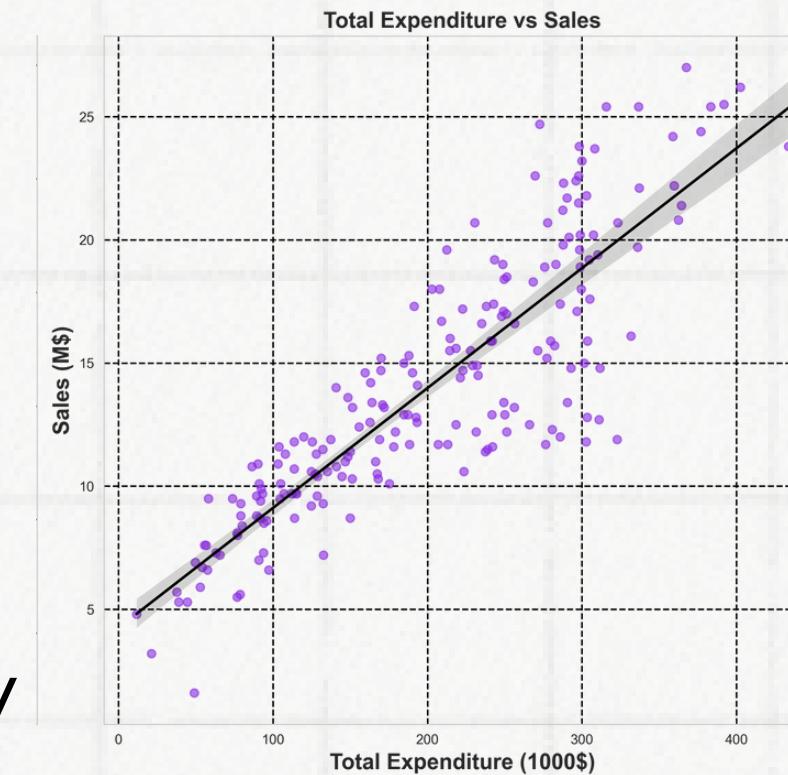
As can be seen from the scatter plot of total expenditure on advertisements vs revenue in sales in below figure that they have a strong positive linear relationship. And to confirm this we first did the normality check On calculating their correlation value turned out to be 0.87 hence confirming the linear positive correlation.



- 1 KS Statistic: 0.0702
- 2 P-value: 0.2657
- 3 Fail to reject the null hypothesis: Total Expenditure follows a normal distribution.
- 4
- 5 Spearman's Rank Correlation between Total Ad Expenditure and Sales: 0.8771

## Q4 How do the average advertising spends compare when sales are high versus when sales are low?

**A4** With reference to the scatter plot between total expenditure on Ads and revenue from Sales we can clearly see that when the sales are low the average advertising spends are also low and vice versa. To verify this I calculated the average total expenditure on advertisements and the results are given in beside figure. Please note that sales below \$10M are considered low and sales above \$20M are considered high. Hence they kind of show a direct proportionality.



# Conclusion

The analysis of advertising expenditures across TV, Radio, and Newspaper reveals that **TV ads receive the highest investment and have the strongest impact on sales (0.80 correlation)**, followed by radio, while newspaper ads contribute the least. A strong positive linear relationship between total ad spending and sales confirms that **higher investments generally drive revenue**, though newspaper spending is right-skewed with lower-budget campaigns. Companies with higher sales invest more in TV ads, highlighting its effectiveness. **Future research will explore optimal budget allocation, spending thresholds, and interaction effects between channels** using advanced statistical techniques like **Multiple Linear Regression** and **ANOVA** to refine advertising strategies and **maximise ROI**.



**Thank you  
very much!**

**vinit.mehta@research.iiit.ac.in**