

## Advertising and Sales

The Advertisement and Sales dataset includes 200 entries that show how different advertising methods, including TV, radio, and newspaper, impact sales. Each entry contains details about the amount of spending on these media channels, in thousands of dollars, and their corresponding sales figures (in thousands of units), giving a chance to dive into how effective these advertising strategies really are in boosting sales. To begin, we look at the summary of our dataset. Here we can see the minimum, maximum, and mean values being spent, in thousands of dollars, on each of the advertisement strategies. TV has the most spending, by far exceeding that of the other two methods. The sales column is in thousands of units, and ranges from 1.6k to 27k units.

```
summary(adsales)
```

```
##           ID           TV           Radio           Newspaper
##  Min.      : 1.00    Min.      : 0.70    Min.      : 0.00    Min.      : 0.30
## 1st Qu.: 50.75    1st Qu.: 74.38    1st Qu.:10.07    1st Qu.: 12.75
## Median :100.50    Median :149.75    Median :22.90    Median : 25.75
## Mean   :100.50    Mean   :147.03    Mean   :23.29    Mean   : 30.55
## 3rd Qu.:150.25    3rd Qu.:218.82    3rd Qu.:36.52    3rd Qu.: 45.10
## Max.   :200.00    Max.   :296.40    Max.   :49.60    Max.   :114.00
##           Sales
##  Min.      : 1.60
## 1st Qu.:10.40
## Median :12.90
## Mean   :14.04
## 3rd Qu.:17.40
## Max.   :27.00
```

I decided to create a faceted scatter plot with the three streams of advertisements to analyse the relationship between rising spending on the advertising method and its impact on the amount of sales. As can be seen in the figure below, sales increased the most with TV's where we can see a sharp positive correlation, with radio being less influential, and newspaper being the least influential on sales. As expected, all three advertising methods create more sales when more money is spent to push them, however TV is by clearly the most effective.

```
library(tidyr)
```

```
# Reshape the data to long format
```

```
ads_long <- adsales %>%
```

```
  pivot_longer(cols = c(TV, Radio, Newspaper), names_to = "Ad_Medium", values_to = "Spend")
```

```
ggplot(ads_long, aes(x = Spend, y = Sales)) +
```

```
  geom_point(alpha = 0.6, size = 3) +
```

```
  geom_smooth(method = "lm", se = FALSE) +
```

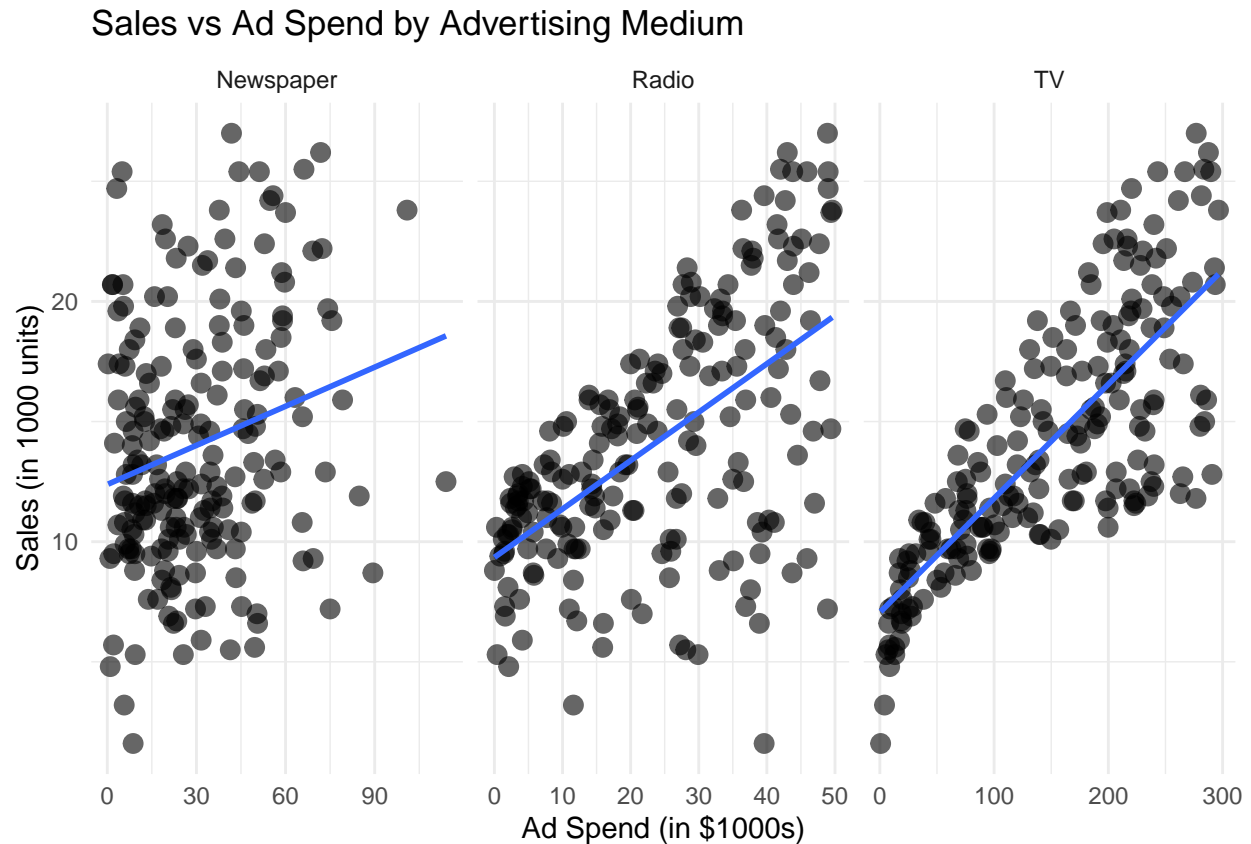
```
  labs(title = "Sales vs Ad Spend by Advertising Medium",
```

```
        x = "Ad Spend (in $1000s)",
```

```
        y = "Sales (in 1000 units)") +
```

```
theme_minimal() +
facet_wrap(~ Ad_Medium, scales = "free_x")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



To analyze the impact of the advertising spending, I built a linear regression model using TV, Radio, and Newspaper ad spend as predictors of Sales. The results show that TV and Radio ads have a statistically significant positive impact on sales, with Radio being the most effective (\$1000 spent on Radio increases sales by 0.185 units, compared to 0.0458 units for TV). In contrast, Newspaper advertising had no significant effect. The model explains 89.5% of the variance in sales ( $R^2 = 0.8951$ ), highlighting that TV and Radio are the most valuable channels for driving sales, while Newspaper spending is less effective.

```
# Split the data into 80% training and 20% testing
set.seed(123) # For reproducibility
train_index <- createDataPartition(adsales$Sales, p = 0.8, list = FALSE)

train_data <- adsales[train_index, ]
test_data <- adsales[-train_index, ]

# Train a linear regression model
sales_model <- lm(Sales ~ TV + Radio + Newspaper, data = train_data)

# View the model summary
summary(sales_model)
```

```
##
## Call:
## lm(formula = Sales ~ TV + Radio + Newspaper, data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.7570 -0.8217  0.2427  1.2018  3.4101
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.003317   0.351637   8.541 1.05e-14 ***
## TV           0.045798   0.001566  29.248 < 2e-16 ***
## Radio        0.184975   0.009611  19.246 < 2e-16 ***
## Newspaper   -0.000390   0.006526  -0.060  0.952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.713 on 158 degrees of freedom
## Multiple R-squared:  0.8951, Adjusted R-squared:  0.8931
## F-statistic: 449.3 on 3 and 158 DF, p-value: < 2.2e-16

# Predict sales on the test data
test_data$predicted_sales <- predict(sales_model, test_data)

# Evaluate model performance
model_performance <- postResample(pred = test_data$predicted_sales, obs = test_data$Sales)
model_performance

##      RMSE Rsquared      MAE
## 1.6200538 0.8986574 1.2687683
```

In the analysis of advertising return on investment (ROI) across our three mediums—TV, Radio, and Newspaper—I found that Radio consistently delivered the highest ROI, followed by Newspaper, with TV yielding the lowest returns. The mean ROI for Radio was just above 100%, indicating that, on average, each dollar spent on radio ads more than doubled in sales returns. Newspaper ads followed closely behind with an average ROI of around 90%. In contrast, TV ads had a much lower mean ROI, hovering around 10%, making it the least effective advertising channel in terms of return on investment. These results suggest that, for maximizing ROI, Radio and Newspaper are significantly more impactful than TV advertising within this dataset.

```
# Calculate ROI for each channel using actual sales instead of predicted revenue
test_data <- test_data %>%
  mutate(ROI_TV = Sales / TV * 100, # Directly use Sales for calculating ROI
         ROI_Radio = Sales / Radio * 100,
         ROI_Newspaper = Sales / Newspaper * 100)

# Summary of ROI
summary(test_data$ROI_TV)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.403   7.950   9.656  14.952  13.576  78.049
```

```
summary(test_data$ROI_Radio)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    20.26  50.77   71.04  116.79  108.79  486.67
```

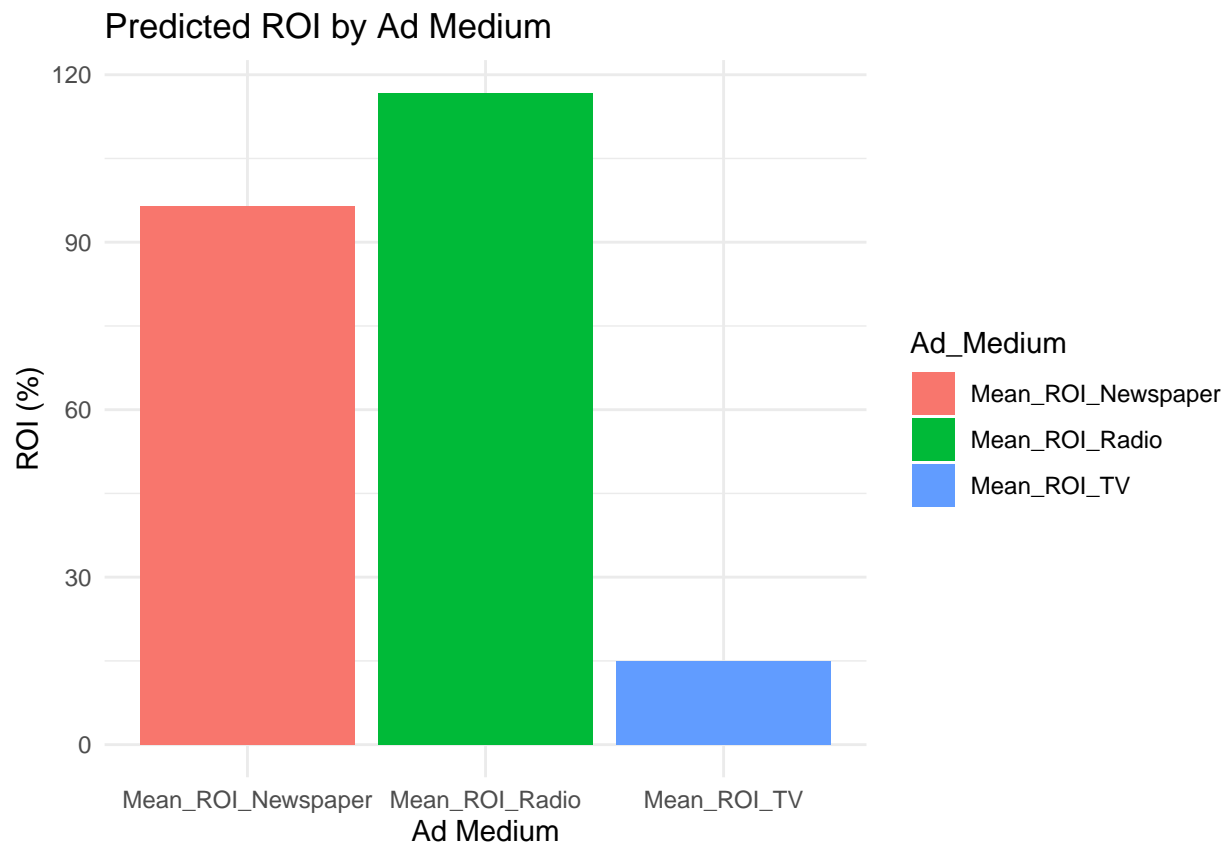
```
summary(test_data$ROI_Newspaper)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.42  33.71   54.34   96.52  104.67  771.88
```

```
# Bar plot for ROI
```

```
roi_long <- test_data %>%
  summarize(Mean_ROI_TV = mean(ROI_TV, na.rm = TRUE),
            Mean_ROI_Radio = mean(ROI_Radio, na.rm = TRUE),
            Mean_ROI_Newspaper = mean(ROI_Newspaper, na.rm = TRUE)) %>%
  pivot_longer(cols = everything(), names_to = "Ad_Medium", values_to = "Mean_ROI")

ggplot(roi_long, aes(x = Ad_Medium, y = Mean_ROI, fill = Ad_Medium)) +
  geom_bar(stat = "identity") +
  labs(title = "Predicted ROI by Ad Medium", x = "Ad Medium", y = "ROI (%)") +
  theme_minimal()
```



Based on our analysis, I recommend that the business prioritize investment in radio advertising, as it provides the highest return on investment (ROI) at around 120% while also significantly contributing to sales. TV

advertising, although yielding a lower ROI (~10%), has the strongest direct impact on sales based on the linear regression results, so maintaining a solid but controlled presence in TV ads is essential for driving overall sales. On the other hand, newspaper advertising shows minimal influence on sales and has a relatively lower ROI, so reducing or reallocating the budget from newspapers to other channels would be more cost-effective.

The dataset could benefit from additional information on product categories or target audience preferences. Understanding how different demographics respond to each ad channel, along with cost structure details for each product, could further refine the strategy and help optimize the marketing mix for maximum effectiveness.