Marketing Mix Regression Analysis

Devi Dahlia

1. Introduction & Business Context

Companies invest in multiple marketing channels such as Radio, TV, and In-Store promotions to drive product sales. However, without quantitative analysis, it is difficult to determine which channels truly generate value. This project applies simple and multiple linear regression to evaluate the effectiveness of different marketing channels on sales performance using real marketing mix data.

2. Business Questions

This project aims to answer the following questions:

1. Which individual marketing channels significantly impact sales?
2. How do these channels perform when evaluated together in a marketing mix?
3. How can the regression model be used to forecast sales under different budget scenarios?

3. Dataset Description

The dataset contains 104 observations of marketing and sales activity. After cleaning, 100 observations were retained. The dependent variable is **NewVolSales**, and the predictors include **Radio**, **TV**, **InStore**, and **Discount**.

4. Data Preparation

```
# Load dataset
marketing <- read.csv("~/Desktop/marketing-regression-project/Data/mktmix datasets.csv")

# Keep only relevant variables
marketing_small <- marketing[, c("NewVolSales", "Radio", "InStore", "TV", "Discount")]

# Remove rows with missing values
marketing_clean <- marketing_small[complete.cases(marketing_small), ]

# Check cleaned data
str(marketing_clean)
```

```
## 'data.frame':    100 obs. of  5 variables:
##  $ NewVolSales: int  19564 19387 23889 20055 20064 19481 19509 19033 20498 19509 ...
##  $ Radio      : num  245 314 324 298 279 259 235 290 315 318 ...
##  $ InStore    : num  15.5 16.4 62.7 16.6 41.5 ...
##  $ TV         : num  101.8 76.7 131.6 119.6 103.4 ...
##  $ Discount   : num  0 0 0.05 0 0.045 0 0 0 0.035 0.045 ...
```

```
summary(marketing_clean)
```

```
##   NewVolSales        Radio          InStore            TV
##  Min.   :17431   Min.   :  0.0   Min.   :10.78   Min.   : 37.66
```

```
##  1st Qu.:19112   1st Qu.:235.0   1st Qu.:21.89   1st Qu.:113.90
##  Median :19966   Median :278.5   Median :30.17   Median :138.04
##  Mean   :20219   Mean   :256.7   Mean   :32.48   Mean   :140.26
##  3rd Qu.:20960   3rd Qu.:313.2   3rd Qu.:40.17   3rd Qu.:175.68
##  Max.   :24944   Max.   :399.0   Max.   :68.12   Max.   :240.29
##     Discount
##  Min.   :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean   :0.02078
##  3rd Qu.:0.04859
##  Max.   :0.09076
```

5. Methodology

Simple linear regression was used to evaluate the individual impact of each marketing channel on sales. Multiple linear regression was then applied to assess the combined effect of all marketing channels simultaneously. Model performance was evaluated using p-values, R-squared, and adjusted R-squared.

6. Result

6.1 Simple Linear Regression: Radio

```
model_radio <- lm(NewVolSales ~ Radio, data = marketing_clean)
summary(model_radio)
```

```
##
## Call:
## lm(formula = NewVolSales ~ Radio, data = marketing_clean)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2757.0 -1115.0  -297.5   732.6  4679.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.000e+04  4.965e+02   40.28   <2e-16 ***
## Radio       8.611e-01  1.833e+00    0.47     0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1586 on 98 degrees of freedom
## Multiple R-squared:  0.002248,   Adjusted R-squared:  -0.007934
## F-statistic: 0.2208 on 1 and 98 DF,  p-value: 0.6395
```

6.2 Simple Linear Regression: TV

```
model_tv <- lm(NewVolSales ~ TV, data = marketing_clean)
summary(model_tv)
```

```
##
## Call:
## lm(formula = NewVolSales ~ TV, data = marketing_clean)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -3013.6 -1068.4  -172.3   576.4  4460.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19383.460    530.969  36.506   <2e-16 ***
## TV              5.955      3.617   1.646    0.103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1567 on 98 degrees of freedom
## Multiple R-squared:  0.02692,    Adjusted R-squared:  0.01699
## F-statistic: 2.711 on 1 and 98 DF,  p-value: 0.1029
```

6.3 Simple Linear Regression: InStore

```
model_instore <- lm(NewVolSales ~ InStore, data = marketing_clean)
summary(model_instore)
```

```
##
## Call:
## lm(formula = NewVolSales ~ InStore, data = marketing_clean)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2508.7  -860.4   -86.6   866.9  3955.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18435.70     361.85  50.948  < 2e-16 ***
## InStore        54.89      10.27   5.342 5.97e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1398 on 98 degrees of freedom
## Multiple R-squared:  0.2255, Adjusted R-squared:  0.2176
## F-statistic: 28.54 on 1 and 98 DF,  p-value: 5.97e-07
```

6.4 Multiple Linear Regression Model

```
model_multi <- lm(NewVolSales ~ Radio + InStore + TV + Discount,
data = marketing_clean)
summary(model_multi)
```

```
##
## Call:
## lm(formula = NewVolSales ~ Radio + InStore + TV + Discount, data = marketing_clean)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2360.9  -950.0  -172.0   775.5  3669.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 1.785e+04  8.014e+02  22.273   <2e-16 ***
## Radio       6.263e-01  1.607e+00   0.390   0.6976
## InStore     3.559e+01  1.453e+01   2.450   0.0161 *
## TV          5.436e+00  3.203e+00   1.697   0.0929 .
## Discount    1.391e+04  7.336e+03   1.896   0.0610 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1361 on 95 degrees of freedom
## Multiple R-squared:  0.2879, Adjusted R-squared:  0.258
## F-statistic: 9.604 on 4 and 95 DF,  p-value: 1.449e-06
```

7. Scenario Simulation & Forecasting

```
# Scenario 1: TV-heavy, Low InStore

plan_1 <- data.frame(Radio = 300, InStore = 20, TV = 220, Discount = 0.03)

# Scenario 2: High InStore

plan_2 <- data.frame(Radio = 200, InStore = 50, TV = 100, Discount = 0.03)

# Scenario 3: High Discount

plan_3 <- data.frame(Radio = 200, InStore = 35, TV = 120, Discount = 0.08)

# Predictions

predict(model_multi, plan_1)
```

```
##        1
## 20363.15
```

```
predict(model_multi, plan_2)
```

```
##        1
## 20715.95
```

```
predict(model_multi, plan_3)
```

```
##        1
## 20986.14
```

8. Business Recommendations

Based on the regression analysis and forecasting results, the company should prioritize In-Store promotion as it is the most consistent and statistically reliable driver of sales. Strategic discounting can provide additional short-term sales lift but should be applied cautiously due to marginal statistical confidence. TV advertising may be used as a supporting channel, while Radio advertising should be reconsidered due to its consistently weak performance.

9. Limitations

This analysis is subject to several limitations. First, the dataset contains only approximately 100 observations, which may limit statistical power. Second, external factors such as seasonality, competitor activity, pricing

strategy, and economic conditions were not included in the model. Finally, the linear regression model assumes linear relationships and does not capture non-linear patterns or interaction effects.

10. Conclusion

This project demonstrates how regression analysis can be applied to real-world marketing data to support evidence-based budget allocation decisions. The results consistently identify In-Store promotion as the strongest driver of sales, while Radio advertising shows no significant impact. Scenario simulations further illustrate how regression models can be used as practical decision-support tools in business planning.