MKTG_Assignment_5

MinJae Jo, Greyson Wheeler, Moris Bercian, Vivian McLaughlin, Brendan Pope

2025-11-21

Part 1

Step 1

```r
library(readxl)
library(GGally)
```

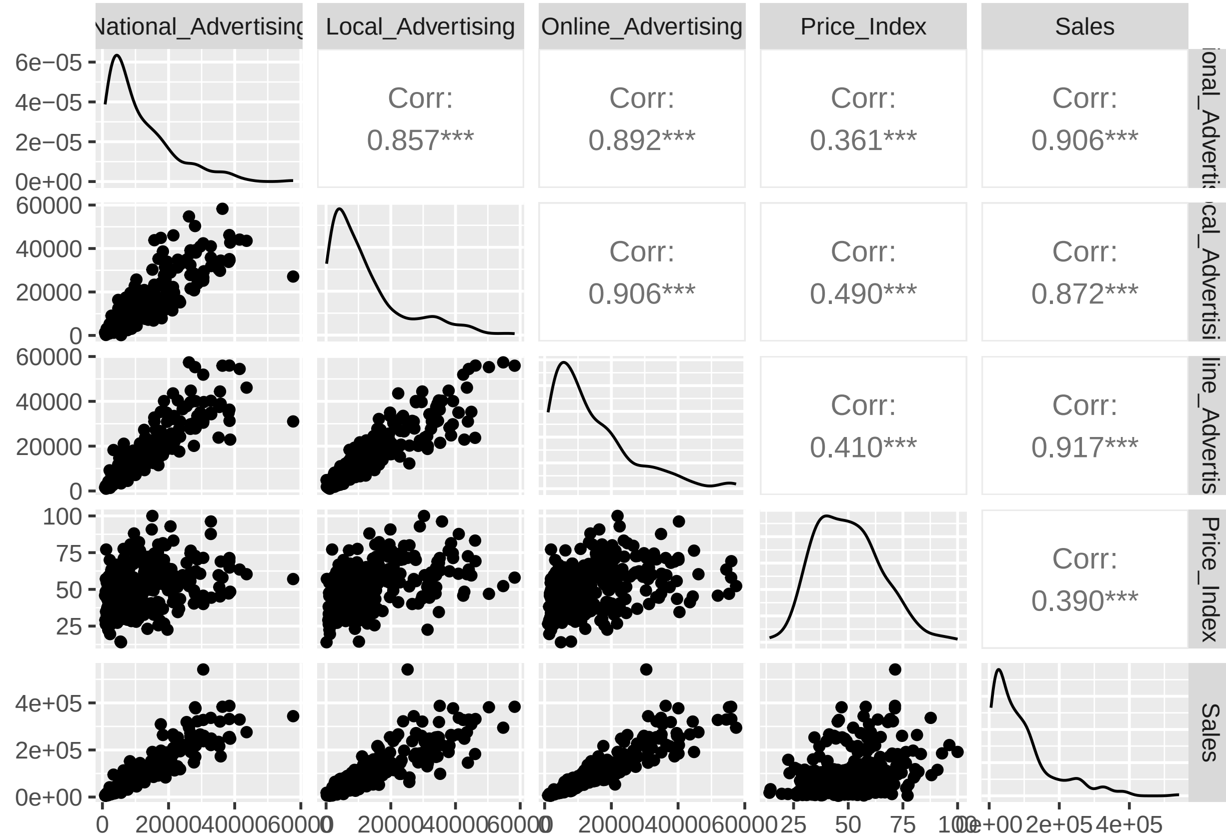
```
## Loading required package: ggplot2
```

```r
library(ggplot2)
fryo_data <- read_excel("Kreyers Froyo.xlsx", sheet = "Data")
```

Step2

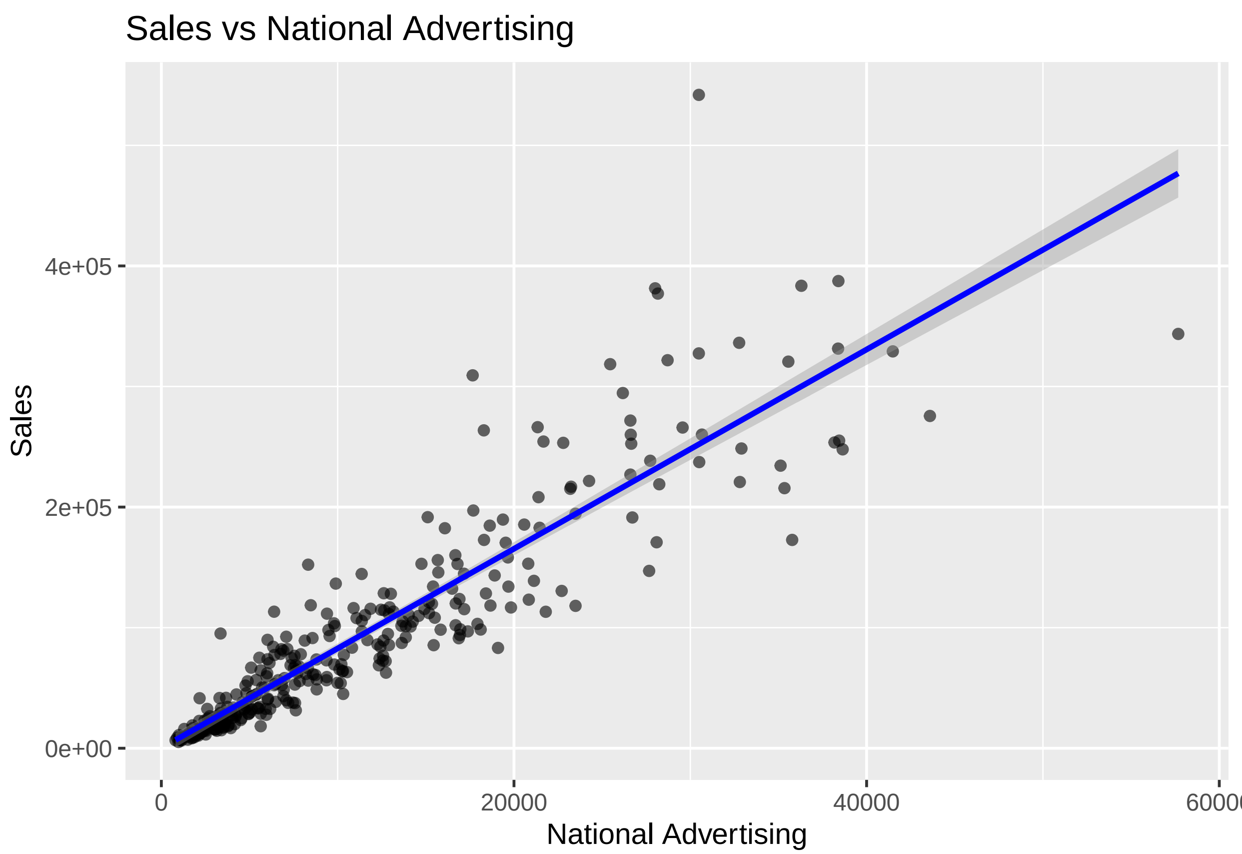
```
##                Variable        Mean          SD        Min   Max Count
## 1 National_Advertising 11191.04000  9665.24205  806.00000 57674   325
## 2    Local_Advertising 12359.49231 11658.05995  144.00000 58281   325
## 3   Online_Advertising 14109.52000 11796.07176 1020.00000 57395   325
## 4          Price_Index    49.97688    15.22343   14.00263   100   325
```

Step 3



Step 4

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



Step 5

```
fryo_data$log_Sales <- log(fryo_data$Sales)
fryo_data$log_National <- log(fryo_data$National_Advertising)
fryo_data$log_Local <- log(fryo_data$Local_Advertising)
fryo_data$log_Online <- log(fryo_data$Online_Advertising)
fryo_data$log_Price <- log(fryo_data$Price_Index)
log_model <- lm(log_Sales ~ log_National + log_Local + log_Online + log_Price,
                data = fryo_data)
```

Step 6

```
##
## Call:
## lm(formula = log_Sales ~ log_National + log_Local + log_Online +
##     log_Price, data = fryo_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54429 -0.15480 -0.02413  0.14062  0.90092
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.32076    0.18151   7.276 2.66e-12 ***
## log_National  0.40348    0.04131   9.766  < 2e-16 ***
```

```
## log_Local     0.13342    0.03223   4.139 4.47e-05 ***
## log_Online    0.56797    0.04851  11.709  < 2e-16 ***
## log_Price    -0.09392    0.04718  -1.991   0.0474 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2237 on 320 degrees of freedom
## Multiple R-squared:  0.9522, Adjusted R-squared:  0.9516
## F-statistic:  1595 on 4 and 320 DF,  p-value: < 2.2e-16
```

Part 2

1. In part 1 step 3, the graph shows that online advertising sales show the strongest relationship. Current graph figures show that the correlation between online advertising and sales is about 0.917, indicating that it is the highest of all advertising variables. This suggests that stores with high online advertising spending tend to consistently generate higher sales than those that rely mostly on national or local advertising. In addition, the upward trend of online advertising is evident in the scatterplot compared to other variables, supporting that online advertising is more directly related to store performance.

2. In the log regression model Step 6, Online Advertising has the largest coefficient, approximately 0.568. Because this is a log model, the coefficient represents elasticity. That is, a 1% increase in online advertising, while keeping the other variables constant, increases sales by about 0.57%. This elasticity is higher than the coefficients for national and local advertising of 0.40 and 0.13. Thus, online advertising not only correlates strongly with sales, but also provides the greatest proportional return when spending increases.

3. Yes, price sensitivity exists in the model. The coefficient of log Price_Index is approximately -0.094, which is negative and statistically significant at the 5% level ($P = 0.047$). This indicates that a 1% increase in the price index reduces sales by approximately 0.094%. Although the effect is weak compared to advertising variables, a negative sign shows that customers reduce their purchases in response to high relative prices. Therefore, price sensitivity exists, but is relatively insignificant compared to the effectiveness of online or national advertising.

4. Overall, logarithmic models account very well for sales. The model has an R-square of approximately 0.952 (and an adjusted R-square of 0.9516), showing that approximately 95% of store to store sales fluctuations are accounted for by the difference between advertising expenditure and relative prices. This is highly descriptive for real-world market data, suggesting that the four predictors of the model capture most of the main drivers of store performance. Furthermore, the F-statistics are very large (1595) with a p-value close to zero 1595. In other words, the entire model is statistically significant and unlikely to be described as random noise. Taken together, we find that response models are robust and reliable, and that advertising especially online advertising is the main and consistent determinant of sales outcomes, and that they play a small but meaningful role in price.