Term paper

Assignment in Marketing Analytics

Hand-in date:

19.02.2024

Campus:

BI Oslo

Examination code and name:

GRA4139 Marketing Analytics

Programme:

MSc in Business Analytics

Table of Contents

Question 1: Descriptive Analyses

Question 1a: Exploring the dataset

The crucial phase in the data analysis process is pre-processing the data. It is highly important to familiarize oneself with the dataset, examine it, and clean it before proceeding with the analysis. Such data wrangling ensures the obtention of meaningful results, free from skewed or faulty data. Hence, this part of the report focuses on dataset exploration and preparation for the subsequent model-building step. For the purpose of pre-processing the data and performing descriptive analyses, R programming is deployed.

1. Missing values & duplicated values: To begin with, it is discovered that the dataset does not contain any missing values or duplicated values; hence, further examination can be proceeded.

2. Outliers

*** Sales**

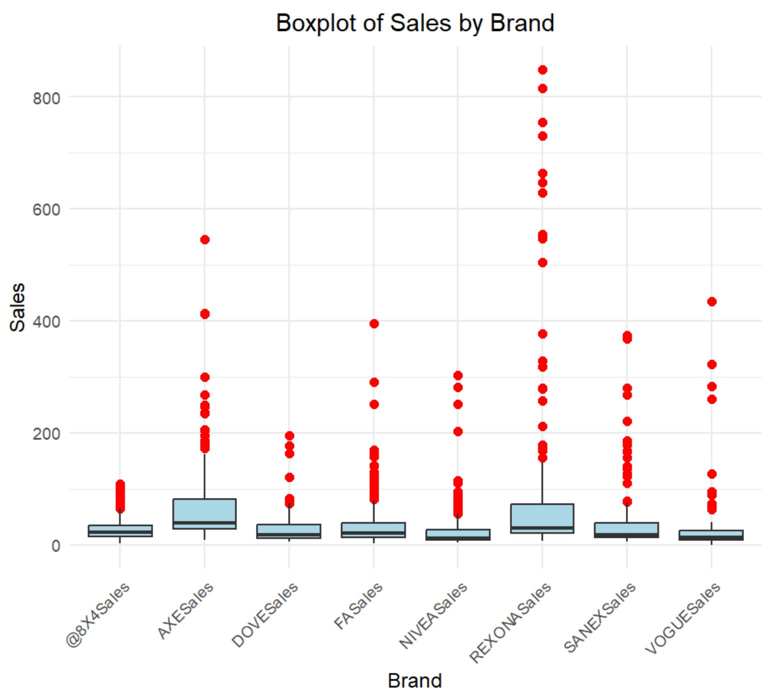

Figure 1.1. Boxplot of Sales by Brand

In Figure 1.1, it can be seen that outliers of sales are visible across all brands, with values significantly higher than the median, indicating some sales peaks during the period 2003 - 2006. REXONA and AXE show the greatest number of outliers with a wide range, especially for REXONA which recorded sales exceeding 800 in two separate weeks. These outliers could be present in the data for several reasons, such as promotional campaigns - which is the most likely scenario given the chain stores' involvement in a price war - or potential seasonal patterns. FA and NIVEA display a moderate frequency of outliers, while DOVE, VOGUE, and @8X4 have fewer outliers and more stable sales.

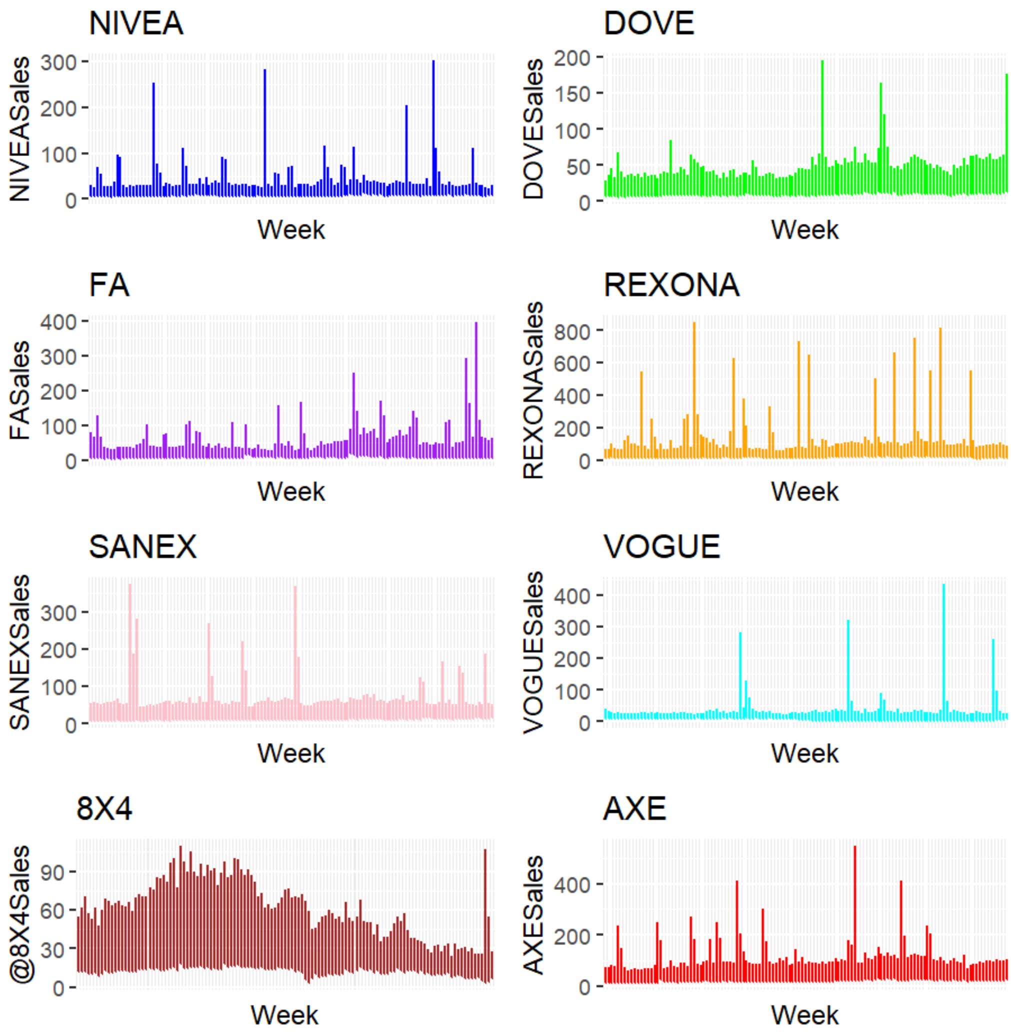

Figure 1.2. Weekly Sales Trends for Deodorant Brands (2003-2006)

The weekly sales trend analysis from 2003 to 2006 further supports these observations. Brands such as REXONA, FA, and NIVEA show frequent peaks in their sales patterns, suggesting they may have benefited from promotional campaigns or seasonal demand surges. In contrast, @8X4 demonstrates a relatively stable pattern with gradual growth and fewer sharp spikes, implying a more consistent customer base without reliance on promotions.

*** Price and Regular Price**

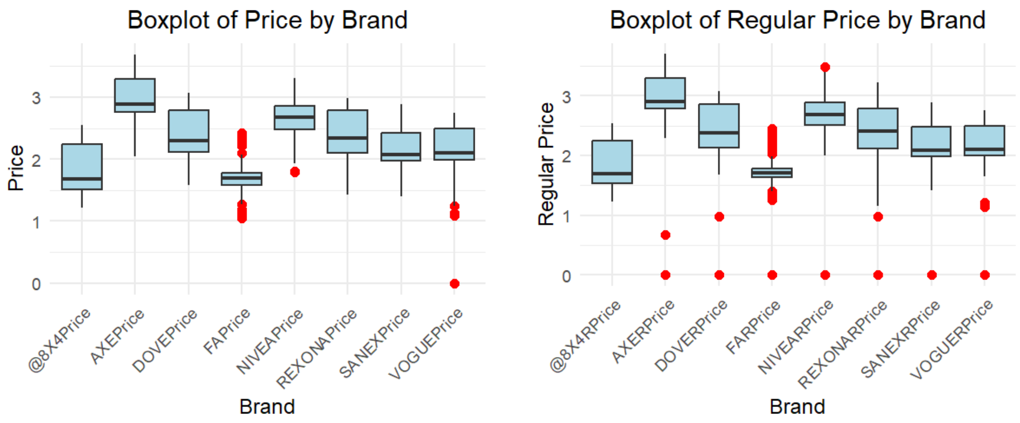

Figure 1.3. Boxplots of Price and Regular Price by Brand

The boxplot of price displays the price distribution for each deodorant brand, highlighting key patterns and outliers. It can be seen that FA and VOGUE show multiple outliers, indicating possible frequent promotions or price cuts. NIVEA has minor outliers, suggesting a stable pricing strategy with occasional discounts. SANEX maintains a stable price range with minimal variability.

In the boxplot of regular prices, which represent prices in stores without promotions and with some smoothing over time, we see fewer outliers overall, as expected due to the smoothing process. However, NIVEA exhibits a cluster of outliers below the lower whisker, indicating that even without promotions, there are some instances of lower-than-usual prices. Additionally, there are some outliers equal to 0 across several brands, which may indicate data anomalies or errors.

3. Logical inconsistencies

It has been observed that, although no negative sales or actual price values are present, 23 sales and 23 price values are recorded as 0. These values correspond to the VOGUE brand at the EDAH supermarket. This finding suggests that EDAH may not have sold the VOGUE brand during certain quarters.

Regarding regular prices, it has been identified that 61 regular price values are 0. Among these, 23 values belong to VOGUE, 1 to DOVER, 1 to FA, 7 to NIVEA, 11 to REXONA, 12 to SANEX, and 6 to AXE. This may be likely a data entry mistake, so it would be a safer choice to transform or exclude these observations from the dataset during further analyses.

Question 1b: Brand positioning in terms of price

1. Price

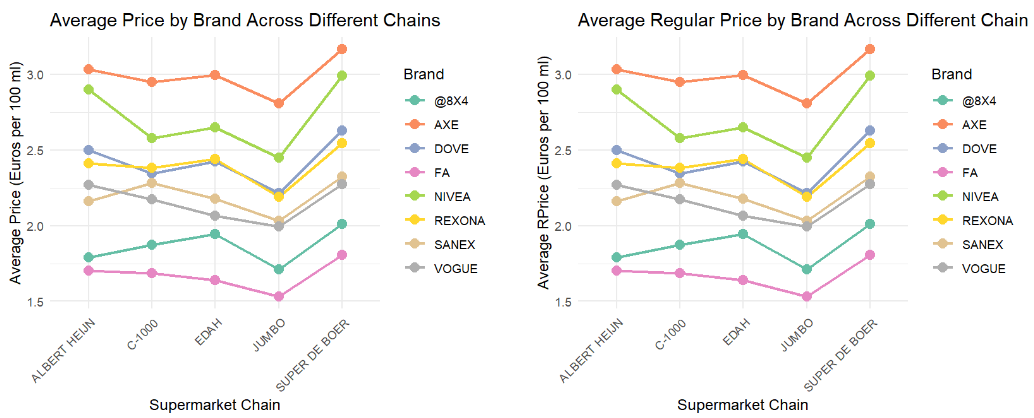

Figure 1.4. Average Price & Average Regular Price of Brands Across Chains

The boxplots (Figure 1.3) and line plots (Figure 1.4) illustrate the price positioning of deodorant brands and their variations across different supermarket chains. Despite differences in type of prices: actual vs regular price (calculated based on the prices for stores in which the brand is not promoted that week), the positioning of brands remains consistent across both metrics.

AXE consistently holds the highest average price, positioning it as the most premium brand, followed by NIVEA which shows some difference in the average price across different chains. In the mid-range segment, DOVE has more variation, with the highest price at SUPER DE BOER and the lowest at JUMBO. REXONA maintains a relatively stable pricing pattern across all chains while SANEX and VOGUE exhibit similar average prices.

FA and @8X4 clearly occupy the lowest-priced segment, with FA having the lowest overall average price and the least variation between chains. @8X4, while slightly more expensive than FA, shows notable differences across chains, with its lowest price at JUMBO and the highest at SUPER DE BOER.

2. Promotion

Regarding the promotion intensity, REXONA stands out with the highest promotion intensity across all chains, using a mix of display, feature, and combined promotions. Similarly, NIVEA, AXE, and DOVE show frequent promotions, with a strong focus on display, highlighting their emphasis on in-store visibility. In contrast, brands like @8X4 and VOGUE have minimal promotional activities, suggesting limited marketing efforts or alternative strategies.

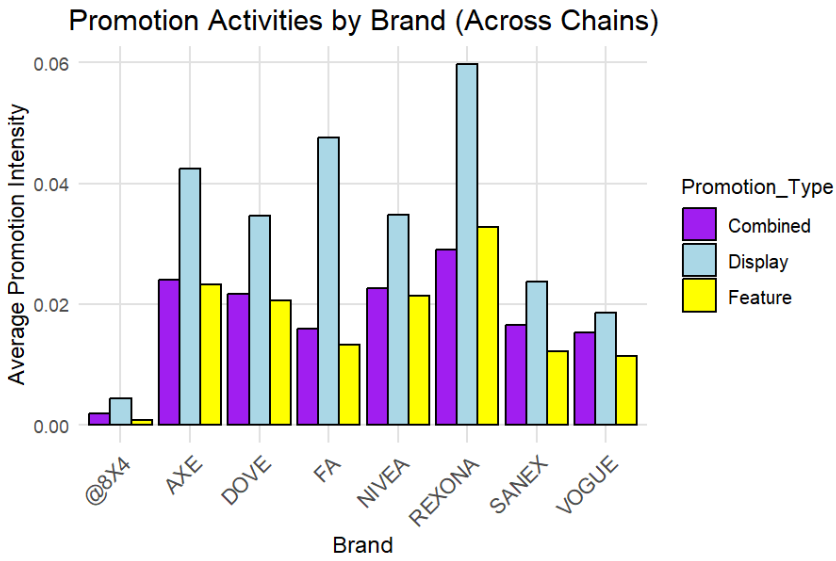

Figure 1.5. Promotion Activities by Brand (Across Chains)

Promotional intensity also varies by chain (Figure 1.5). ALBERT HEIJN and EDAH have the highest promotion activity, with brands like REXONA and NIVEA using frequent combined promotions (both display and feature) to increase visibility. SUPER DE BOER also shows significant promotion activity but with a stronger emphasis on feature promotions, indicating a preference for flyer and advertising campaigns. On the other hand, JUMBO has the lowest promotional intensity, with most brands relying on limited display promotions, suggesting a more price-driven retail strategy with fewer promotional campaigns.

The frequency and type of promotions vary by brand and chain. REXONA and NIVEA maintain consistent promotional efforts across all chains, while AXE and DOVE increase their activity in ALBERT HEIJN and EDAH, where in-store promotions are more common. FA and SANEX, though less promoted overall, show a slight preference for feature promotions in chains like SUPER DE BOER and EDAH.

Figure 1.6. Promotion Activities by Brand (Within Each Chain)

3. Conclusion

In conclusion, the deodorant spray market shows clear differences in price positioning and promotional intensity across brands. AXE and NIVEA are premium brands, with AXE having the highest price, followed by NIVEA, while FA and @8X4 are positioned as the lowest-priced brands. In the mid-range segment, DOVE, REXONA, SANEX, and VOGUE have similar average prices, with REXONA maintaining more stable pricing. In terms of promotions, REXONA leads with the highest intensity, using a mix of display, feature, and combined promotions, followed by NIVEA, AXE, and DOVE, which emphasize display promotions for in-store visibility. SANEX, VOGUE, FA, and @8X4 show significantly lower promotional activity, with FA and SANEX having a slight preference for feature promotions. Overall, premium brands engage in more frequent and diverse promotions, while mid-range and value brands have fewer, more selective campaigns.

Question 1c: Discovering seasonal patterns

1. Plotting Average Sales by Quarter

To assess whether there are any seasonal patterns in the data, we first plotted the average sales for each brand by quarter. From the line chart, we observed noticeable trends for some brands, particularly AXE and REXONA, which showed peak sales in Q2 and Q3, and a decline in Q4. Other brands, such as DOVE and FA, displayed more stable sales patterns.

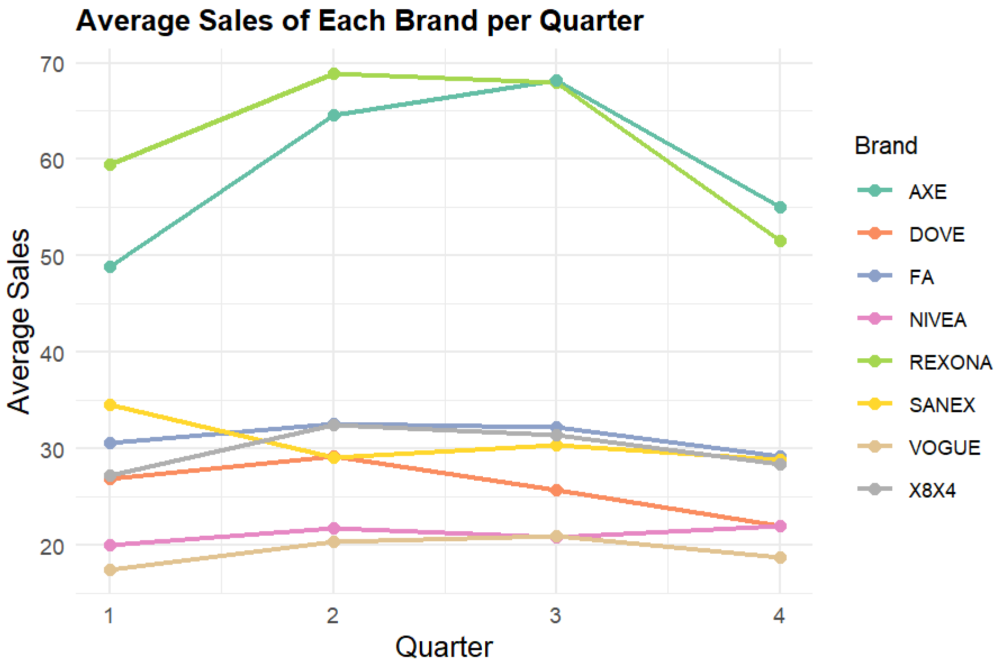

Figure 1.7. Average Sales of Each Brand per Quarter

2. Conducting Analysis of Variance (ANOVA)

Next, we performed one-way ANOVA tests on the sales of each brand to determine if there were statistically significant differences between quarters.

Brand	p-value	Significance
DOVE	0.0164	Significant (*)
FA	0.781	Not Significant
NIVEA	0.879	Not Significant
REXONA	0.332	Not Significant
SANEX	0.4	Not Significant
VOGUE	0.691	Not Significant
@8X4	0.0905	Marginally Significant (.)
AXE	0.00178	Highly Significant (**)

Table 1.1. ANOVA Result for Each Brand

Conclusion: AXE and DOVE show clear seasonal patterns, with AXE being the most prominent. @8X4 indicates a marginal trend. FA, NIVEA, REXONA, SANEX, and VOGUE show no significant quarterly differences.

Question 2: Promotional Activities Effectiveness

Model Selection Process

In order to identify the most effective marketing instrument and examine whether the effectiveness of promotions differ across chains, we first develop a multiplicative sales model for NIVEA.

1. Initial approach: Pooled and Unpooled models with new variable "NIVEADiscount"

- Define new variable: NIVEADiscount

$$NIVEADiscount_t = 1 - \frac{NIVEAPrice_t}{NIVEARPrice_t}$$

The reason why we define this new variable is to avoid collinearity issues. Since NIVEAPrice and NIVEARPrice have the correlation of 0.88996, including both variables separately in a regression model can cause multicollinearity. Additionally, when we include NIVEADiscount, the R-squared of the model is higher, which indicates that the discount variable explains more variance in NIVEA sales and improves the model's predictive power compared to using actual and regular prices separately.

- Pooled model:

$$NIVEASales_{it} = \beta_0 NIVEAPrice_{it}^{\beta_1} \beta_2^{NIVEADiscount_{it}} \beta_3^{NIVEADISP_{it}} \beta_4^{NIVEAFEAT_{it}} \beta_5^{NIVEADF_{it}} \varepsilon_t$$

(Note: i indicates chain, t indicates week)

- Unpooled models:

$$NIVEASales_{it} = \beta_{0i} NIVEAPrice_{it}^{\beta_{1i}} \beta_{2i}^{NIVEADiscount_{it}} \beta_{3i}^{NIVEADISP_{it}} \beta_{4i}^{NIVEAFEAT_{it}} \beta_{5i}^{NIVEADF_{it}} \varepsilon_{it}$$

(Note: i indicates chain, t indicates week)

- Findings and Issues: We ran both pooled and unpooled models for five chains and found that the R-squared of the pooled model is relatively low (0.3459), while the R-squared values for individual chains vary (some are high, some are low).

However, we encountered an inconsistency when performing the Chow Test. The **F-statistic is 160.2767**, indicating a significant difference between the pooled and unpooled models. Yet, the **p-value is 1**, suggesting that the coefficients and intercepts should remain the same across all chains. These two findings seem to contradict each other, raising concerns about the validity of pooling the data.
=> **Conclusion:** We do not choose this approach.

2. Second approach: Partially pooled model with indicator (dummy) variables for each chain

- Define new variables: We take Albert-Heijn as the reference group and define 4 dummy variables as follows:

- **CT**: 1 if the chain is **C-1000**, 0 otherwise.
- **ED**: 1 if the chain is **EDAH**, 0 otherwise.
- **SUP**: 1 if the chain is **Super De Boer**, 0 otherwise.
- **JUM**: 1 if the chain is **Jumbo**, 0 otherwise.

The reason for defining these new variables is as follows: First, while exploring the dataset, we observed that some chains have significantly higher sales than others. This suggests that these chains are either larger or more popular, attracting more customers. This finding indicates that "chain" can be a factor influencing sales levels. Hence, we create dummy variables for each chain to capture chain-specific effects, as different chains may have unique sales patterns, customer demographics, and pricing strategies. This improves model accuracy by accounting for unobserved heterogeneity and allows us to compare the impact of each chain relative to a reference category.

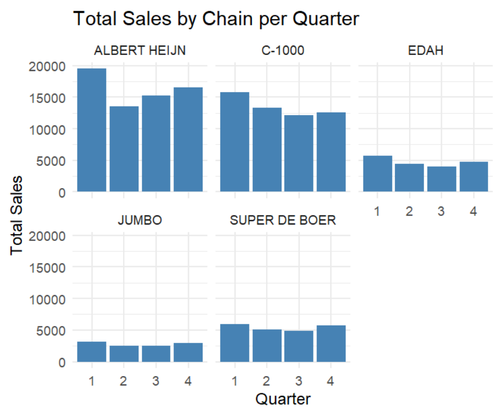

Figure 2.1. Total Sales by Chain (All Brands)

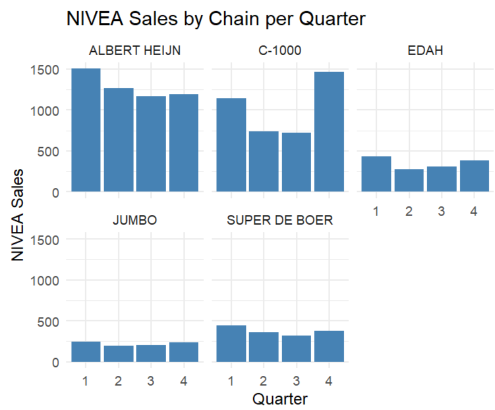

Figure 2.2. Total Sales by Chain (NIVEA only)

- Model 1:

$$NIVEASales_{it} = \beta_0 NIVEAPrice_{it}^{\beta_1} \beta_2^{NIVEADiscount_{it}} \beta_3^{NIVEADISP_{it}} \beta_4^{NIVEAFEAT_{it}} \beta_5^{NIVEADF_{it}} \beta_6^{CT_{it}} \beta_7^{ED_{it}} \beta_8^{SUP_{it}} \beta_9^{JUM_{it}} \varepsilon_t$$

- Log-transform of model 1:

$$ln(NIVEASales_{it}) = ln(\beta_0) + \beta_1 ln(NIVEAPrice_{it}) + NIVEADiscount_{it} ln(\beta_2) + NIVEADISP_{it} ln(\beta_3) + NIVEAFEAT_{it} ln(\beta_4) + NIVEADF_{it} ln(\beta_5) + CT_{it} ln(\beta_6) + ED_{it} ln(\beta_7) + SUP_{it} ln(\beta_8) + JUM_{it} ln(\beta_9) + ln(\varepsilon_t)$$

- Results:

```
Call:
lm(formula = log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount +
    NIVEADISP + NIVEAFEAT + NIVEADF + CT + ED + SUP + JUM, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97071 -0.14114 -0.01934  0.12433  1.60385

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.50980    0.18024  25.022  < 2e-16 ***
log(NIVEAPrice) -0.96997    0.16243  -5.972    4e-09 ***
NIVEADiscount    3.48785    0.31204  11.178  < 2e-16 ***
NIVEADISP        0.27590    0.12277   2.247 0.024978 *
NIVEAFEAT       -0.32274    0.14326  -2.253 0.024629 *
NIVEADF          0.55878    0.14381   3.886 0.000113 ***
CT              -0.38490    0.03783 -10.174  < 2e-16 ***
ED              -1.23344    0.03810 -32.372  < 2e-16 ***
SUP             -1.08015    0.03266 -33.073  < 2e-16 ***
JUM             -1.72444    0.04589 -37.574  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2501 on 605 degrees of freedom
Multiple R-squared:  0.8827,    Adjusted R-squared:  0.8809
F-statistic: 505.7 on 9 and 605 DF,  p-value: < 2.2e-16
```

Figure 2.3. Regression Results for NIVEA Sales Model 1

According to Figure 2.3, the Adjusted R^2 is 0.88, indicating the model explains 88% of the variance in log-sales, with a highly significant overall fit ($p < 0.001$). The regression results show that price negatively impacts sales, while discounts positively influence sales. Promotional effects vary, feature, display and combined feature & display significantly affect sales level. Additionally, sales performance differs significantly across chains, with some chains showing lower sales compared to the reference chain. The model explains a high proportion of variance in sales, confirming the strong impact of price, promotions, and chain-specific factors.

Question 2a: Identify the most effective marketing instrument

Based on the regression results, the most effective marketing instrument for NIVEA, is the Combined Feature & Display Promotion (NIVEADF), since 1% increase in combined feature and display promotions leads to a 0.56% increase in sales, suggesting a strong synergistic effect between in-store visibility and advertisements. Display promotion alone (NIVEADISP) also positively impacts sales, with a 0.28% increase per 1% increase in display promotions, indicating that in-store product placement helps attract consumers. However, Feature Advertisement alone (NIVEAFEAT) has a negative impact (-0.32%), meaning that increasing advertisements without additional in-store support may not be an effective strategy. This suggests that feature advertisements alone may not drive sales unless reinforced by physical store presence.

Variable	Estimate	Interpretation
log(NIVEAPrice)	-0.96997	A 1% increase in price leads to a 0.97% decrease in NIVEA sales.
NIVEADiscount	3.48785	A 1% increase in discount results in a 3.49% increase in sales.
NIVEADISP	0.27590	A 1% increase in display promotion results in a 0.28% increase in NIVEA sales.
NIVEAFEAT	-0.32274	A 1% increase in feature advertising results in a 0.32% decrease in NIVEA sales
NIVEADF	0.55878	A 1% increase in combined feature & display promotion leads to a 0.56% increase in NIVEA sales.

Table 2.1. Result Interpretations for NIVEA Sales Model

One note we need to mention here is that although Discount has the highest influence on sales, it is a variable we define based on the observed difference between actual and regular prices. However, we cannot determine whether this difference is due to NIVEA's pricing strategy or external factors; therefore, there is no clear evidence that Discount is a marketing strategy/instrument of NIVEA.

Hence, to answer the question which is the most effective marketing instrument of NIVEA, here we will only compare the effects of the three key marketing instruments: Feature, Display, and Combined Promotions.

Question 2b: Whether the effectiveness of promotions varies across grocery chains

While our previous model indicates that in-store promotions (especially Feature & Display combined) are key sales drivers, they assume the same effect of promotion activities applies across all chains (Intercepts are different in each chain but the coefficients are the same).

To assess whether the effectiveness of promotions varies across grocery chains, we employed a panel data approach using mixed-effects models. Given the hierarchical structure of the dataset, where weekly sales observations are nested within five supermarket chains, we used linear mixed-effects regression (LMM) to account for both fixed effects (price and promotions) and random effects, such as chain-specific variations.

We tested four different models and then applied ANOVA to compare their performance:

- **Model 2**: Different intercepts but same coefficients. This model follows the same structure as the previous model, Model 1, but allows for varying intercepts across different chains.
- **Model 3**: Different intercepts with different slopes for Display promotions. This model considers that the effect of Display promotions may differ by chain.
- **Model 4**: Different intercepts with different slopes for Feature promotions. Similar to Model 3, but here we allow the effect of Feature promotions to vary by chain.
- **Model 5**: Different intercepts with different slopes for Feature and Display promotions combined. This model accounts for the combined effect of Feature and Display promotions, allowing it to vary across chains.

```
> # compare the models
> anova(model_2, model_3, model_4, model_5)
refitting model(s) with ML (instead of REML)
Data: data
Models:
model_2: log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount + NIVEADISP + NIVEAFEAT + NIVEADF + (1 | Chain)
model_3: log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount + NIVEADISP + NIVEAFEAT + NIVEADF + (1 + NIVEADISP | Chain)
model_4: log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount + NIVEADISP + NIVEAFEAT + NIVEADF + (1 + NIVEAFEAT | Chain)
model_5: log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount + NIVEADISP + NIVEAFEAT + NIVEADF + (1 + NIVEADF | Chain)
        npar    AIC     BIC  logLik deviance   Chisq Df Pr(>Chisq)
model_2    8 84.698 120.071 -34.349   68.698
model_3   10 73.203 117.419 -26.601   53.203 15.4957  2  0.0004317 ***
model_4   10 47.909  92.125 -13.954   27.909 25.2941  0
model_5   10 44.122  88.338 -12.061   24.122  3.7867  0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2.5. Model Comparison Result

Based on the model comparison results, model_5 is the best choice as it has the lowest AIC (44.122) and BIC (88.338), indicating the best trade-off between model fit and complexity. This model includes a random intercept and slope for NIVEADF within Chain, which may better capture variations across chains. While model_3 was the only model that showed a statistically significant improvement over model_2 (p = 0.0004317), neither model_4 nor model_5 demonstrated a significant improvement beyond model_3. However, since rmodel_5 has the lowest AIC and BIC, it is still the most optimal choice from a model selection standpoint. If a simpler model is preferred for interpretability, model_3 could also be considered, as it was the only one that significantly improved upon the baseline model. Nonetheless, model_5 remains the strongest model overall due to its superior fit and complexity balance.

Here is the equation for the best-fit model that was estimated:

$$ln(NIVEASales_{it}) = ln(\beta_{0i}) + \beta_1 ln(NIVEAPrice_{it}) + NIVEADiscount_{it} ln(\beta_2) +$$
$$+ NIVEADISP_{it} ln(\beta_3) + NIVEAFEAT_{it}\ ln(\beta_4) + NIVEADF_{it} ln(\beta_{5i}) + ln(\varepsilon_{it})$$

```
> model_5 <- lmer(log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount+ NIVEADISP + NIVEAFEAT + NIVEADF
+                 + (1 + NIVEADF|Chain), data=data)
> summary(model_5)
Linear mixed model fit by REML ['lmerMod']
Formula: log(NIVEASales) ~ log(NIVEAPrice) + NIVEADiscount + NIVEADISP +
    NIVEAFEAT + NIVEADF + (1 + NIVEADF | Chain)
   Data: data

REML criterion at convergence: 33

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3423 -0.6084 -0.0585  0.5257  6.5921

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Chain    (Intercept) 0.46252  0.6801
          NIVEADF     0.76825  0.8765   0.84
 Residual             0.05756  0.2399
Number of obs: 615, groups:  Chain, 5

Fixed effects:
                Estimate Std. Error t value
(Intercept)       3.6184     0.3428  10.555
log(NIVEAPrice)  -0.9632     0.1558  -6.182
NIVEADiscount     2.5995     0.3256   7.983
NIVEADISP         0.5348     0.1229   4.352
NIVEAFEAT        -0.2903     0.1431  -2.028
NIVEADF           0.7782     0.4637   1.678

Correlation of Fixed Effects:
            (Intr) l(NIVE NIVEADs NIVEADI NIVEAF
lg(NIVEAPr) -0.460
NIVEADiscnt -0.261  0.561
NIVEADISP   -0.064  0.123 -0.271
NIVEAFEAT   -0.013  0.024  0.106  -0.123
NIVEADF      0.636 -0.019 -0.155   0.079  -0.297
```

Figure 2.4. Regression Results for NIVEA Sales Model 5

This specification suggests that the impact of using display and feature promotions combined on sales differs significantly across chains. The improvement in model fit indicates that allowing for heterogeneous responses to combined promotions better explains the observed variance in sales.

Question 2c: NIVEA's Main Competitors

Based on the dataset and findings from Question 1b, NIVEA's main competitors can be identified through their price positioning and promotional strategies across different grocery chains.

Analysis of price tiers indicates that NIVEA is positioned as a premium brand, competing closely with AXE, which maintains the highest average price across all chains. REXONA also falls within a similar price range, particularly in stores like Albert Heijn and EDAH, where both brands display comparable price levels. DOVE, while exhibiting some pricing variability, remains a competitor due to its overlap in price positioning across certain chains.

In terms of promotional intensity, NIVEA relies heavily on display promotions. AXE and REXONA exhibit similar promotional strategies, with REXONA using a mix of feature and display promotions to enhance visibility. DOVE also employs display promotions, aligning closely with NIVEA's approach. On the other hand, brands like 8X4 and SANEX demonstrate lower promotional intensity, with 8X4 engaging in minimal marketing activities and SANEX favoring selective feature promotions.

In conclusion, considering price positioning, promotional strategies, and competitive market dynamics, the primary competitors to NIVEA are AXE, REXONA, and DOVE. These brands share similarities with NIVEA in both pricing and promotional strategies, positioning them as the most significant competitors within the deodorant market.

Question 3: Managerial Recommendations

1. Maintain competitive pricing

Based on the regression results from the multiplicative sales model (Model 1), the analysis reveals that price sensitivity is high, as indicated by the significant negative coefficient for log(NIVEAPrice). This suggests that increasing prices would lead to a sharp decline in sales. Discount (the ratio indicating the difference between regular price and actual price) has the most substantial positive effect, meaning that offering price reductions significantly boosts sales. Therefore, NIVEA should maintain competitive pricing and consider targeted price promotions in key retail chains.

2. Increase Combined Feature and Display promotion and reduce Feature-only promotion (for all chains)

Promotional activities also play a crucial role in driving sales. Combined feature and display promotions (NIVEADF) show a strong positive impact, indicating that using multiple promotional elements together is particularly effective. Display promotions (NIVEADISP) contribute positively to sales, though to a lesser extent than dual-feature promotions. In contrast, feature-only promotions (NIVEAFEAT) appear to negatively impact sales, suggesting that the manager may have to be aware when to use this approach. Given these insights, NIVEA should prioritize combined feature and display promotion while reducing reliance on feature-only promotions. Suppose there is a budget for marketing activities, the brand manager should allocate an appropriate amount of money for Display-only promotion and combined feature and display promotion based on their elasticities, and reduce the amount of money spent on Feature-only activities.

3. Implement adjustments for Combined Feature and Display promotion in different chains

It is recommended to increase the use of combined feature and display promotions; however, the brand manager should carefully assess the extent to which these promotions should be increased across different retail chains. According to Model 5, the effectiveness of combined feature and display promotions varies by chain, suggesting that a one-size-fits-all approach may not be optimal. Therefore, the brand manager should prioritize increasing these promotions in chains where they have the **strongest positive impact on sales,** while adjusting the strategy in chains where the effect is weaker to maximize overall promotional effectiveness.

4. Increase more promotional activities in Albert-Heijn and C-1000

Based on the regression results (Model 1), we should increase promotional activities in Albert Heijn and C-1000 to maximize NIVEA's sales performance. The model indicates that display promotions (NIVEADISP) and dual-feature promotions (NIVEADF) have a significant positive impact on sales, with coefficients of 0.2759 ($p = 0.025$) and 0.5587 ($p < 0.001$), respectively. This suggests that increasing these types of promotions in key retail chains can drive noticeable growth.

When evaluating chain-specific effects, the negative coefficients for EDAH (ED) (-1.2344), Super de Boer (SUP) (-1.0801), and Jumbo (JUM) (-1.7244) indicate that these chains have lower baseline sales for NIVEA compared to Albert Heijn, which serves as the reference group. This does not imply a negative effect but rather reflects the fact that Albert Heijn has the highest baseline sales among the chains analyzed. In contrast, C-1000 also shows relatively stronger baseline sales. Since feature promotions (NIVEAFEAT) have a negative impact (-0.3224, $p = 0.0427$), relying solely on them is not recommended. Instead, NIVEA should focus on display and combined feature and display promotions to enhance visibility and engagement, particularly in stores where they have been found to be most effective.

Therefore, NIVEA should allocate more promotional resources to Albert Heijn and C-1000, prioritizing combined feature and display promotions, as they have been proven to generate the highest impact on sales. This targeted strategy will ensure that marketing investments are used efficiently to drive stronger performance across retail channels.

Question 4: Impact of Price War on The Model

The price war may have changed the relationships between the variables in our model, especially the sensitivity of NIVEA sales to price, promotions, and presence in different supermarkets. To determine whether these relationships remained constant before and after the price war, a Chow test can be applied to detect a possible structural break, in simple terms, we can evaluate a pooled model including all the data from fall 2003 to 2005 and compared it with an unpooled model of the market before and after the price war started, if the Chow test determines that the difference is significant, then we can conclude that the effect of our variables has changed and we need to adjust the model to the new state of the market.

The first step would be to split the data into two periods: one before October 2003 and one after. Then, the model would be fitted separately in each period and compared to a single model that includes the entire sample. If the coefficients of the model vary a lot between the two periods, this would indicate that the price war modified the relationships between the variables. Two separate regression models are then estimated to create a non pooled model:

- **Model 1:**

$$ln(NIVEASales_t) = \beta_0 + \beta_1 ln(NIVEAPrice_t) + \beta_2 NIVEADiscount_t + \beta_3 NIVEADISP_t + \beta_4 NIVEAFEAT_t + \beta_5 NIVEADF_t + \beta_6 CT_t + \beta_7 ED_t + \beta_8 SUP_t + \beta_9 JUM_t + \varepsilon_t$$

- **Model 2:**

$$ln(NIVEASales_t) = \alpha_0 + \alpha_1 ln(NIVEAPrice_t) + \alpha_2 NIVEADiscount_t + \alpha_3 NIVEADISP_t +$$

$$\alpha_4 NIVEAFEAT_t + \alpha_5 NIVEADF_t + \alpha_6 CT_t + \alpha_7 ED_t + \alpha_8 SUP_t + \alpha_9 JUM_t + v_t$$

We use a different math symbol in the two models to represent that the data is different for each of the models, this would be reflected in our programming code by splitting the data in two different groups. After making this process and calculating the respective errors we then create a pooled model:

- **Pooled Model:**

$$ln(NIVEASales_t) = \gamma_0 + \gamma_1 ln(NIVEAPrice_t) + \gamma_2 NIVEADiscount_t + \gamma_3 NIVEADISP_t +$$

$$\gamma_4 NIVEAFEAT_t + \gamma_5 NIVEADF_t + \gamma_6 CT_t + \gamma_7 ED_t + \gamma_8 SUP_t + \gamma_9 JUM_t + \gamma_{10} PostWar_t + v_t$$

It is important to mention that we have created a dummy variable named "Post-war" to represent the effect of the war in the model. Based on our previous work we can infer that it is going to be significant and will improve the predicting capacity of our model. After computing the respective errors and applying it to the chow test we can reach 2 different conclusions. First, the models are not different enough so there is no effect of the price wars in the parameters or the second option would be to adapt the model to the new conditions of the market because of the effect that the price war has created.

Another important consideration when evaluating the impact of the price war is the potential issue of multicollinearity. Given that price reductions, promotional activities, and supermarket presence are likely correlated as we saw in our previous model, multicollinearity could inflate standard errors and end up with an incorrect model. The VIF analysis can help diagnose this issue, and if high collinearity is detected, we could remove redundant variables, combining correlated predictors or applying regularization techniques such as the logarithms options as we explored in our partially pooled model.

An alternative approach to capturing the structural changes induced by the price war is to do the analysis with a panel data model, observing the variation across supermarkets and over time. This would allow us to control unobserved effects at the supermarket level, isolating the true effect of price and promotions on NIVEA sales while accounting for chain characteristics. A random-effects model could also be considered if we assume that unobserved differences across supermarkets are uncorrelated with the independent variables. This would provide a clearer causal interpretation of how the price war influenced consumer purchasing behavior beyond simple correlation.

In conclusion, the price war in Dutch grocery retailing could introduce structural changes in the relationships between NIVEA sales, price, promotions, and supermarket presence. By applying a Chow test, we can test whether these relationships remained stable or changed due to the new market state. If a structural break is detected, adjusting the model is necessary to better capture the market conditions. Additionally, addressing potential multicollinearity ensures that our estimates remain reliable and interpretable and that we are working with a meaningful model. Beyond traditional regression analysis, adopting a panel data framework could provide insights into how the price war affected consumer behavior and competitive strategies. Finally, refining the model based on these findings allows for a more accurate understanding of market responses and supports better decision-making in pricing and promotional strategies.