task

Jinran Li

2024-12-10

Data Processing

```
data <- read.csv("temu_product_sales_dataset.csv")
```

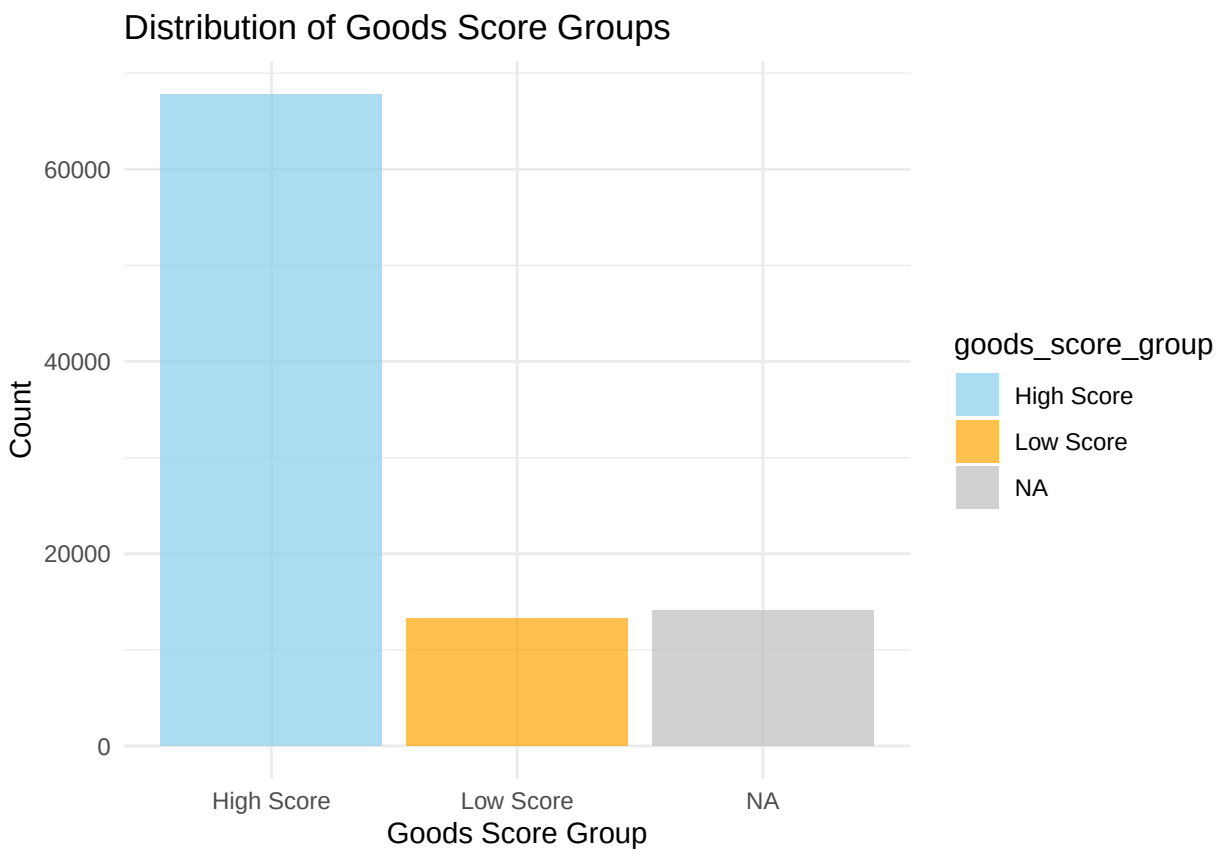
1.1 check the Na

```
#check the NA
sapply(data,function(x) sum(is.na(x)))
```

```
##   leve_1_category_id leve_1_category_name   leve_2_category_id
##                    0                    0                    0
## leve_2_category_name                title             goods_id
##                    0                    0                    0
##           sales_info          category_id              visible
##                    0                    0                    0
##         sales_volume            price_str                price
##                 1709                    0                    0
##          goods_score     comments_num_raw          comment_num
##                14119                    0                14119
```

Because the NA values are so large, there are 14,119 rating and review NA values, which is a huge amount of data, I've scratched the NA values into a new column, which is available below to exwplore why customers don't want to rate scores

```
#Creating Grouping Variables
data <- data %>%
  mutate(goods_score_group = case_when(
    is.na(goods_score) ~ "NA",          # NA
    goods_score <= 4.5 ~ "Low Score",       # Lowrating (<4.5)
    goods_score >4.5 ~ "High Score"     # Highrating (>4.5)
  ))


ggplot(data, aes(x = goods_score_group, fill = goods_score_group)) +
  geom_bar(alpha = 0.7) +
  labs(
    title = "Distribution of Goods Score Groups",
    x = "Goods Score Group",
    y = "Count"
  ) +
  scale_fill_manual(values = c("skyblue", "orange", "grey")) +
  theme_minimal()
```




```r
# Extract Low Score data
low_score_data <- data %>%
  filter(goods_score_group == "Low Score")

# Extract High Score data
high_score_data <- data %>%
  filter(goods_score_group == "High Score")

# Extract NA data
na_score_data <- data %>%
  filter(goods_score_group == "NA")

data <- data %>%
  mutate(
    low_score = ifelse(goods_score_group == "Low Score", 1, 0),
    high_score = ifelse(goods_score_group == "High Score", 1, 0),
    na_score = ifelse(goods_score_group == "NA", 1, 0)
  )
```

1.1.2Delete all the na value after i filled NA value into a new column I also removed the sales_valume for the next observation.

```r
data <- data %>% na.omit()

# check if there are still missing values
sapply(data, function(x) sum(is.na(x)))
```

```
##   leve_1_category_id leve_1_category_name   leve_2_category_id
##                    0                    0                    0
## leve_2_category_name                title             goods_id
##                    0                    0                    0
##           sales_info          category_id              visible
##                    0                    0                    0
##         sales_volume            price_str                price
##                    0                    0                    0
##          goods_score     comments_num_raw          comment_num
##                    0                    0                    0
##    goods_score_group            low_score           high_score
##                    0                    0                    0
##             na_score
##                    0
```

1.2 Basci information

to see how many category and how many product we have in level 1

```
table(data$leve_1_category_name)
```

```
##
##               Automotive          Bags & Luggages          Beauty & Health
##                     1941                     4376                     4569
## Electronics & Appliances       Health & Household           Home & Kitchen
##                     3062                      344                     7451
##  Industrial & Commercial    Jewelry & Accessories    Kid, Baby & Maternity
##                     1415                    11233                     8516
##           Men's Clothing      Musical Instruments         Office & School
##                     4564                      259                     1946
##     Patio, Lawn & Garden             Pet Supplies                    Shoes
##                      346                     3032                     3870
##        Sports & Outdoors Tools & Home Improvement             Toys & Games
##                     3856                      816                     3578
##    Underwear & Sleepwear         Women's Clothing     Women's Curve + Plus
##                     3582                    10642                      618
```

1.2.1same as previes

```
length(unique(data$leve_2_category_name))
```

```
## [1] 189
```

```
table(data$leve_2_category_name)
```

```
##
##          Abrasive & Finishing Products                           Accessories
##                                     66                                   102
##                   Adhesives & Sealants                   Air Quality & Irons
##                                     53                                    19
##                           Arts & Crafts                                  Bath
##                                    278                                   458
```

```
##                          Beauty Tools                                 Bedding
##                                   903                                    79
##                       Birds & Poultries                       Boy's Bottoms
##                                    52                                   147
##                             Boy's Coat                            Boy's Sets
##                                   156                                   682
##                             Boy's Shoes                           Boy's Tops
##                                   348                                   719
##                        Car Accessories               Car Storage & Organizers
##                                  1196                                    70
##                               Car Toys                    Cleaning Supplies
##                                    49                                   113
##                     Commercial Lighting                 Computer Accessories
##                                    37                                   434
##                  Curve + Plus Beachwear                 Curve + Plus Bottoms
##                                     9                                    87
##       Curve + Plus Dresses & Jumpsuits          Curve + Plus Tops & Outerwear
##                                   189                                   212
##            Curve + Plus Two Pieces Set      Curve + Plus Underwear & Sleepwear
##                                     9                                   112
##                           Cutting Tools                             Cycling
##                                   235                                   408
##                   Dogs & Cats Supplies                   Dolls & Plush Toys
##                                  1295                                   265
##                  Event & Party Supplies          Fish, Reptiles & Amphibians
##                                   710                                    40
##                                Fishing               Foot, Hand & Nail Care
##                                   716                                  1500
##               Gardening & Urban Farming                        Girl's  Shoes
##                                    92                                   393
##                          Girl's Bottoms                          Girl's Coat
##                                   238                                   151
##                          Girl's Dresses                          Girl's Sets
##                                   581                                   990
##                             Girl's Tops           Greeting Cards & Postcards
##                                   741                                   184
##                               Hair Care                             Hardware
##                                   240                                    57
##      Headphones, Earbuds & Accessories                Health Care Products
##                                   478                                    20
##                          Home Appliances                 Home Decor Products
##                                   141                                  1558
##  Home Use Medical Supplies & Equipment                   Household Supplies
##                                    38                                    55
##                     Industrial Materials                          Instruments
##                                   194                                   100
##                    Jewelry Accessories                 Keychains&Key Shells
##                                   147                                   182
##                      Kid's Accessories                    Kids' Electronics
##                                  2735                                    85
##                       Kids' Home Store                      Kitchen & Dining
##                                    20                                  2344
##    Kitchen, Bath Fixtures & Appliances                 Learning & Education
##                                    50                                   138
```

```
##               Lighting & Ceiling Fans          Lights & Lighting Accessories
##                                   280                                     78
##                 Luggage & Travel Gear                                Makeup
##                                   531                                   1114
##                    Maternity Clothing        Maternity, Baby Gear & Nursery
##                                    17                                    618
##                    Men's Accessories                      Men's Backbags
##                                   561                                   175
##                          Men's Belts                         Men's Boots
##                                    82                                    205
##                         Men's Bottoms          Men's Clothing & Accessories
##                                   412                                   325
##       Men's Exotic & Novelty  Clothing                        Men's Glasses
##                                   132                                   261
##                         Men's Handbags              Men's Hats, Caps & Wigs
##                                    42                                    582
##                    Men's Jackets&Coats                       Men's Jeans
##                                   519                                   203
##                         Men's Jewelry             Men's Loafers & Slip-Ons
##                                  2798                                   171
##                       Men's Loungewear                    Men's Mules&Clogs
##                                   152                                    45
##                 Men's Sandal & Slippers               Men's Scarves & Gloves
##                                   436                                    73
##                           Men's Shirts                          Men's Shoes
##                                  1071                                   974
##                    Men's Shoulder Bags               Men's Socks & Hosiery
##                                   169                                   886
##                    Men's Sports Shoes                          Men's Suits
##                                   106                                    54
##                       Men's Sweaters                    Men's Sweatshirts
##                                   151                                   747
##                       Men's T-Shirts                           Men's Tops
##                                  1190                                   217
##                       Men's Underwear             Men's Waist & Chest bags
##                                   379                                   243
##             Men's Wallets & Card Cases           Men's Watches & Accessories
##                                   204                                   375
##              Men's Work & Safety Shoes                           Motorcycles
##                                    39                                   174
##                    Novelty & Costumes  Occupational Health & Safety Products
##                                   149                                    79
##                Office & School Supplies                  Office Electronics
##                                   572                                    17
##                            Organization                       Other Animals
##                                   300                                    48
##              Other Industrial Supplies                   Other Toys & Games
##                                    75                                   757
##         Outdoor Lighting & Power Tools                       Outdoor Lights
##                                   103                                   113
##                    Outdoor Recreation     Outdoor Supplies & Patio Furniture
##                                   582                                   151
##      Paint, Wall Treatments & Supplies               Papers, Labels & Indexes
##                                   343                                   358
```

```
##                       Party Supplies                            Personal Care
##                                  325                                      392
##             Personal Care Electronics                     Personal Care Products
##                                   72                                       52
##             Pet Apparel & Accessories                    Pet Cleaning & Grooming
##                                 1221                                      376
##                    Phone Accessories                    Phone Cables & Chargers
##                                  325                                      238
##         Phone Cases & Screen Protector                         Power & Hand Tools
##                                 1008                                      443
##                Power Banks & Batteries                      Power Tools & Safety
##                                   61                                      137
##                          Pretend Play                    Puzzles & Building Toys
##                                   58                                      415
##                     Quadcopters & UAV                          Replacement Parts
##                                   34                                       18
##                        Seasonal Decor                         Shoe Accessories
##                                  975                                      132
##             Sports & Fitness Supplies                               Sports Bags
##                                  569                                      165
##    Stationery & Gift Wrapping Supplies                    Stickers & Crafts Tape
##                                  160                                     1059
##                 Storage & Organization                 Studio Recording & Stage Live
##                                  886                                       57
##                Test, Measure & Inspect                        Tools & Equipment
##                                  233                                      223
##                              Wall Art                         Wearable Technology
##                                  238                                      271
##         Wellness & Relaxation Products                         Wigs & Accessories
##                                   19                                      420
##                         Winter Sports                        Women's Accessories
##                                   41                                     1126
##                    Women's Activewear                         Women's Athleisure
##                                  585                                      211
##                      Women's Backbags                         Women's Beachwear
##                                  169                                      396
##                       Women's Blazer                          Women's Bodysuits
##                                   45                                       30
##                        Women's Boots                Women's Clutches & Evening Bags
##                                  324                                       47
##                 Women's Coat & Jacket                  Women's Cosplay Costume
##                                  401                                       12
##                 Women's Crossbody Bags                           Women's Denims
##                                  562                                      665
##                       Women's Dresses                  Women's Fashion Sneakers
##                                 2202                                      479
##                        Women's Flats                          Women's Glasses
##                                  153                                      118
##                      Women's Handbags                         Women's Hats & Caps
##                                  233                                      528
##                       Women's Jewelry                          Women's Jumpsuits
##                                 4389                                      123
##                      Women's Lingerie                             Women's Pants
##                                  803                                      295
```

```
##                         Women's Pumps                         Women's Sandals
##                                    70                                     115
##                   Women's Sexy Lingerie              Women's Shapewear & Others
##                                   720                                      94
##                         Women's Shorts                   Women's Shoulder Bags
##                                   140                                     422
##                         Women's Skirts                      Women's Sleepwear
##                                    86                                     416
##                        Women's Slippers                    Women's Sports Shoes
##                                   727                                     136
##             Women's Stockings & Hosiery                          Women's Suits
##                                  1008                                     147
##                        Women's Sweaters                    Women's Sweatshirts
##                                  1406                                     738
##                           Women's Tops                      Women's Tote Bags
##                                  2737                                     967
##              Women's Waist & Chest Bags            Women's Wallets & Card Cases
##                                   111                                     501
##                         Women's Watches Writing Supplies & Correction Supplies
##                                   193                                     515
##                              Yoga&Studio
##                                   110
```

1.3 revenue

Create a new column sales revenue

```
data <- data %>%
  mutate(revenue = price * sales_volume)
# Viewing the data structure after adding a new column
head(data)
```

```
##   leve_1_category_id leve_1_category_name leve_2_category_id
## 1                885 Patio, Lawn & Garden                888
## 2                885 Patio, Lawn & Garden                888
## 3                885 Patio, Lawn & Garden                888
## 4                885 Patio, Lawn & Garden                888
## 5                885 Patio, Lawn & Garden                888
## 6                885 Patio, Lawn & Garden                888
##                 leve_2_category_name
## 1 Outdoor Supplies & Patio Furniture
## 2 Outdoor Supplies & Patio Furniture
## 3 Outdoor Supplies & Patio Furniture
## 4 Outdoor Supplies & Patio Furniture
## 5 Outdoor Supplies & Patio Furniture
## 6 Outdoor Supplies & Patio Furniture
##                                                                                 title
## 1            1pc 30 LED Bubble Ball Lights Solar String Lights Outdoor Waterproof Lights
## 2                    50 LED Solar Sakura Lights Outdoor Waterproof Christmas Decorations
## 3       10m 20m 30m Solar Copper Wire String Light Holiday Supplies Christmas Decoration
## 4 LED String Lights, Battery Operated Copper Wire Fairy Lights For Garden And Home Party
## 5       1pc 39ft 100LED Solar String Lights, Waterproof Christmas Lights, Twinkle Lights
## 6                                          1pc Outdoor Decorative Stone Shaped Key Case
##       goods_id sales_info category_id visible sales_volume price_str price
```

```
## 1 6.017592e+15 5,560 sold         891    True         5560       655  6.55
## 2 6.017592e+15 3,569 sold         891    True         3569       589  5.89
## 3 6.017592e+15 2,364 sold         891    True         2364       489  4.89
## 4 6.010995e+14   538 sold         891    True          538       159  1.59
## 5 6.010995e+14   474 sold         891    True          474       839  8.39
## 6 6.017592e+15   469 sold         891    True          469       162  1.62
##   goods_score comments_num_raw comment_num goods_score_group low_score
## 1         4.7            (200)         200        High Score         0
## 2         4.7            (200)         200        High Score         0
## 3         4.7            (200)         200        High Score         0
## 4         5.0              (1)           1        High Score         0
## 5         4.7            (200)         200        High Score         0
## 6         4.6            (129)         129        High Score         0
##   high_score na_score  revenue
## 1          1        0 36418.00
## 2          1        0 21021.41
## 3          1        0 11559.96
## 4          1        0   855.42
## 5          1        0  3976.86
## 6          1        0   759.78
```

1.3.1 check the na

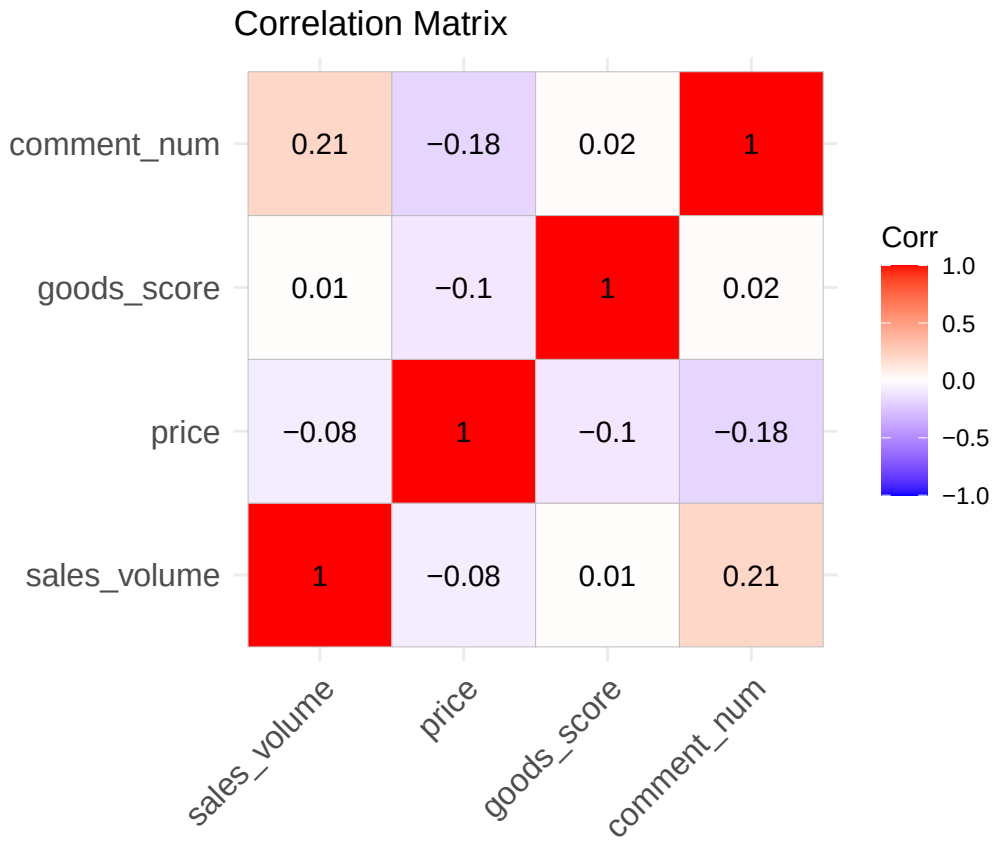
```
data <- data %>%
  mutate(price = as.numeric(price),
         sales_volume = as.numeric(sales_volume))

data <- data %>%
  filter(!is.na(price) & !is.na(sales_volume))
```

EDA

2.1 Correlation Matrix

The correlation matrix shows a weak linear relationship between the variables, with only sales_volume showing a slight positive correlation with comment_num (0.21) and a weak negative correlation with price (-0.08)

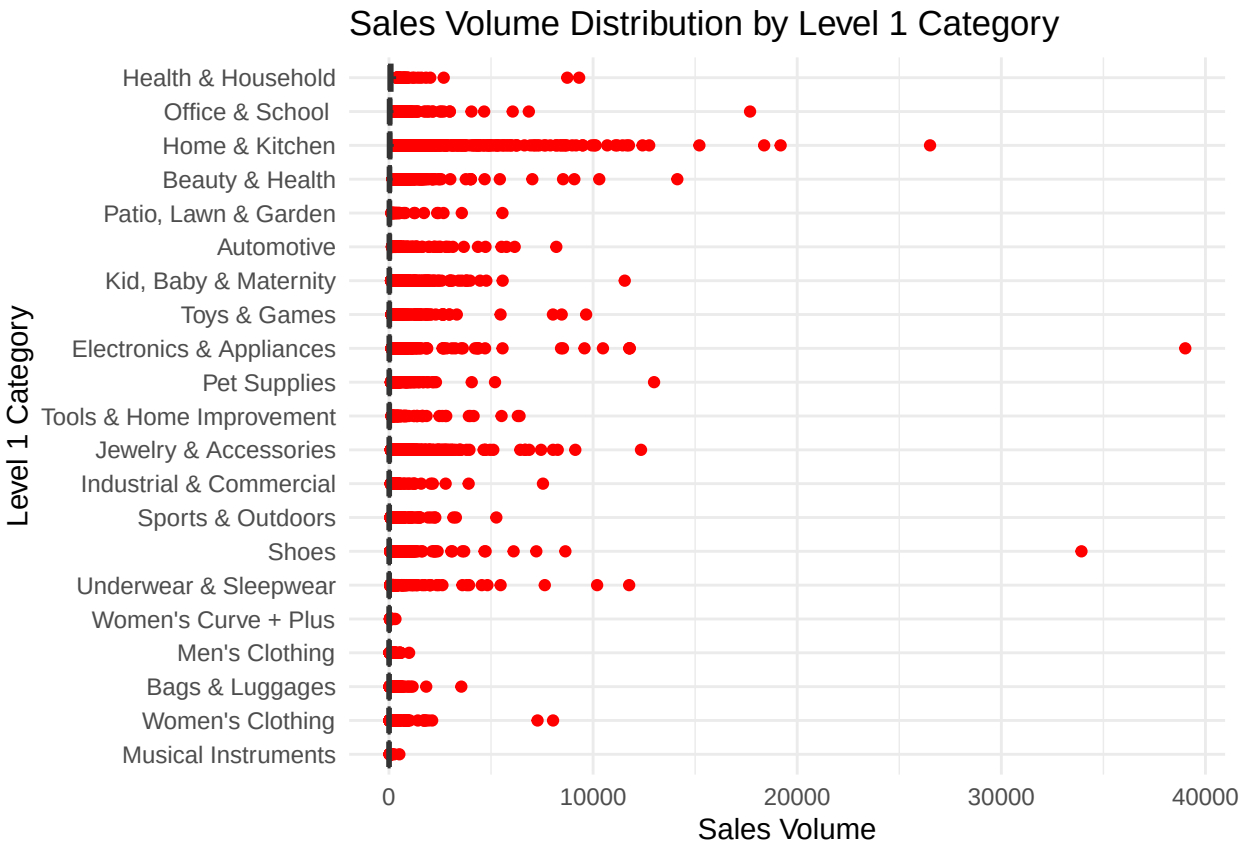
```
numeric_vars <- data[, c("sales_volume", "price", "goods_score", "comment_num")]
cor_matrix <- cor(numeric_vars, use = "complete.obs")

# Heat map the correlation matrix
ggcorrplot::ggcorrplot(cor_matrix, lab = TRUE, title = "Correlation Matrix")
```



2.2 distribution

shows the distribution of sales by level 1 category. Each red dot represents the sales volume of one product, the horizontal axis represents the sales volume, and the vertical axis lists the categories.

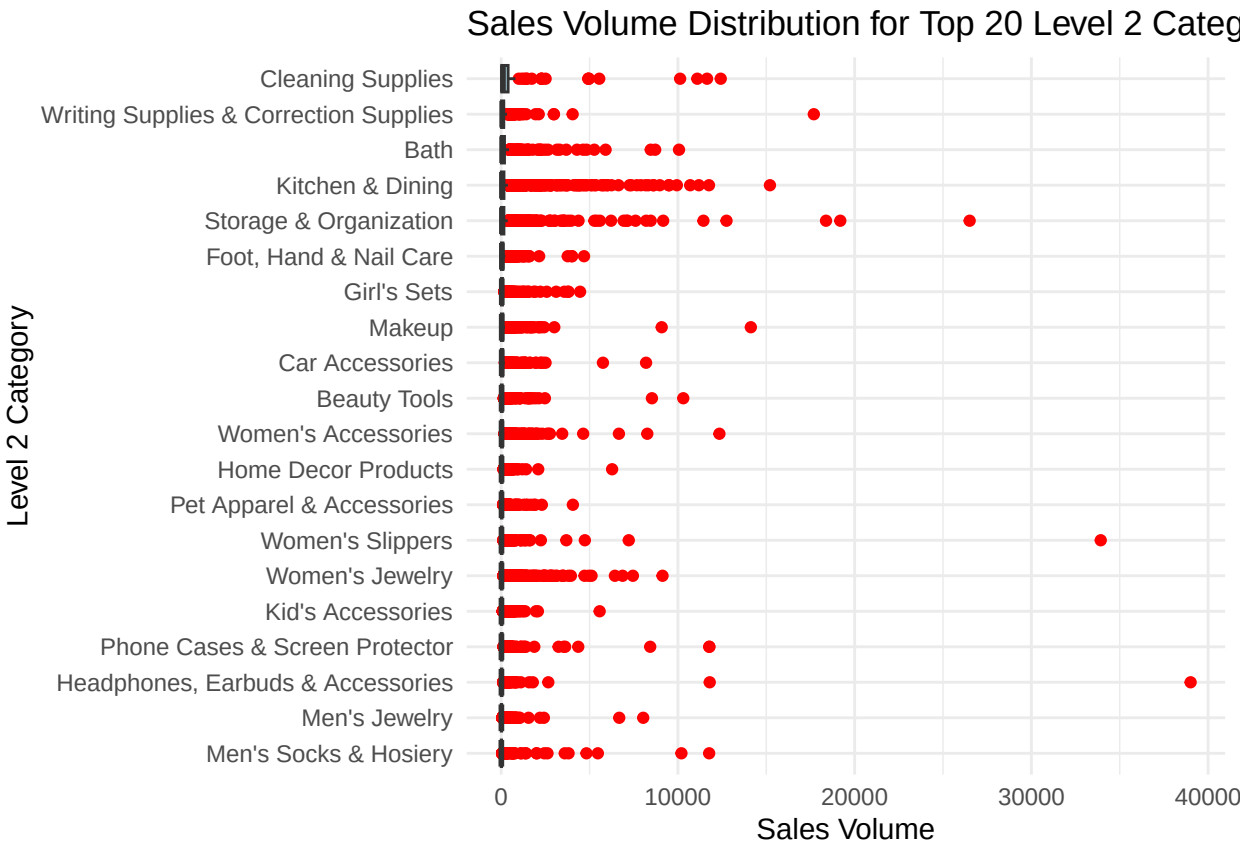
```
ggplot(data, aes(x = reorder(leve_1_category_name, sales_volume, FUN = median), y = sales_volume)) +
  geom_boxplot(outlier.color = "red", fill = "lightblue") +
  coord_flip() +
  labs(title = "Sales Volume Distribution by Level 1 Category",
       x = "Level 1 Category", y = "Sales Volume") +
  theme_minimal()
```



2.2.1 lv2 Sales volume distribution

shows the distribution of sales volume for the top 20 secondary categories, providing a more detailed view of product performance within the subcategories. Each red dot represents the sales volume of one product, with categories such as "Women's Slippers" and "Headphones, Earbuds & Accessories" having the highest sales volume, including outliers of over 30,000 units.

```
# Top 20 high-demand categories by total sales volume
top_level2 <- data %>%
  group_by(leve_2_category_name) %>%
  summarise(total_sales = sum(sales_volume, na.rm = TRUE)) %>%
  arrange(desc(total_sales)) %>%
  slice_head(n = 20)  # Get the first 20 categories

# Filter out the top 20 categories
filtered_data <- data %>%
  filter(leve_2_category_name %in% top_level2$leve_2_category_name)

# Box line plots showing distribution of sales in high demand categories
ggplot(filtered_data, aes(x = reorder(leve_2_category_name, sales_volume, FUN = median), y = sales_volu
  geom_boxplot(outlier.color = "red", fill = "lightblue") +
  coord_flip() +
  labs(title = "Sales Volume Distribution for Top 20 Level 2 Categories",
       x = "Level 2 Category", y = "Sales Volume") +
  theme_minimal()
```

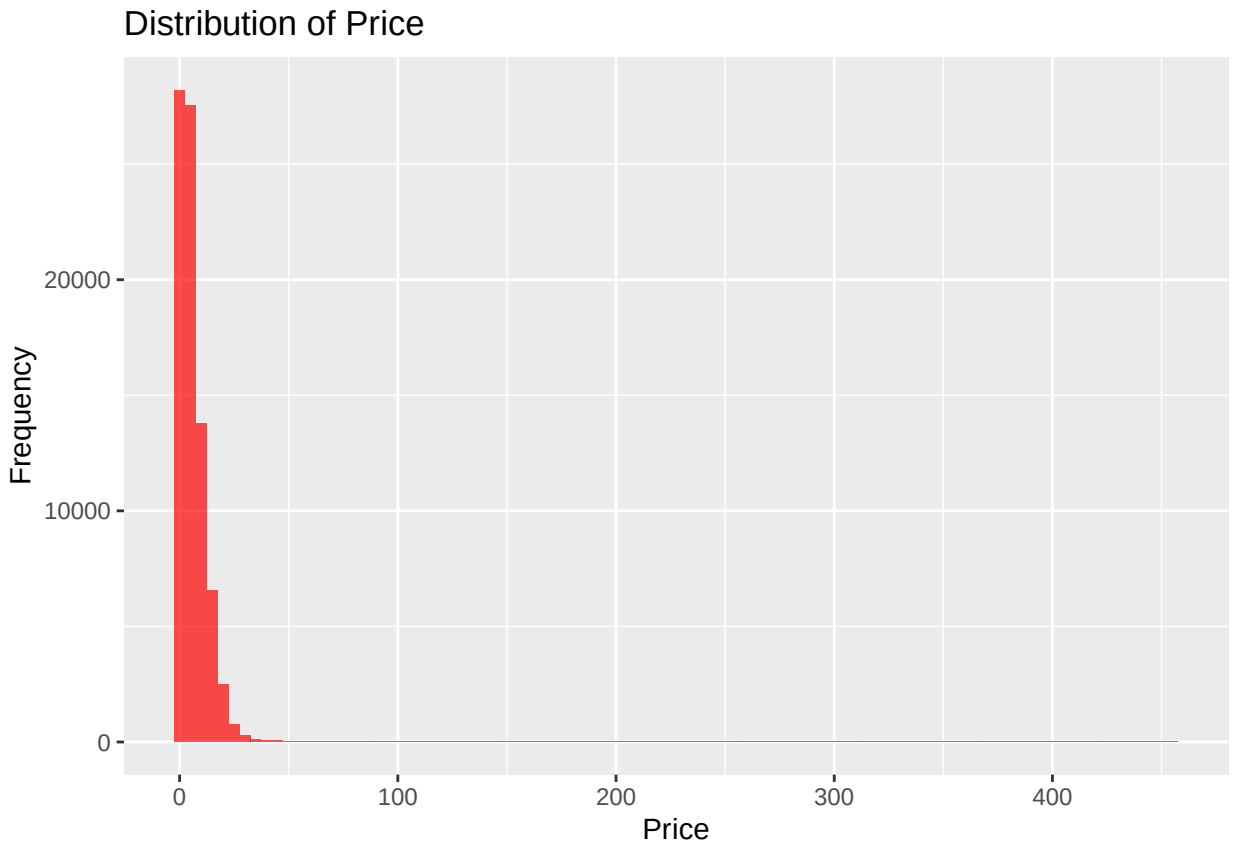


2.2.2 Price, Score , Comment_num

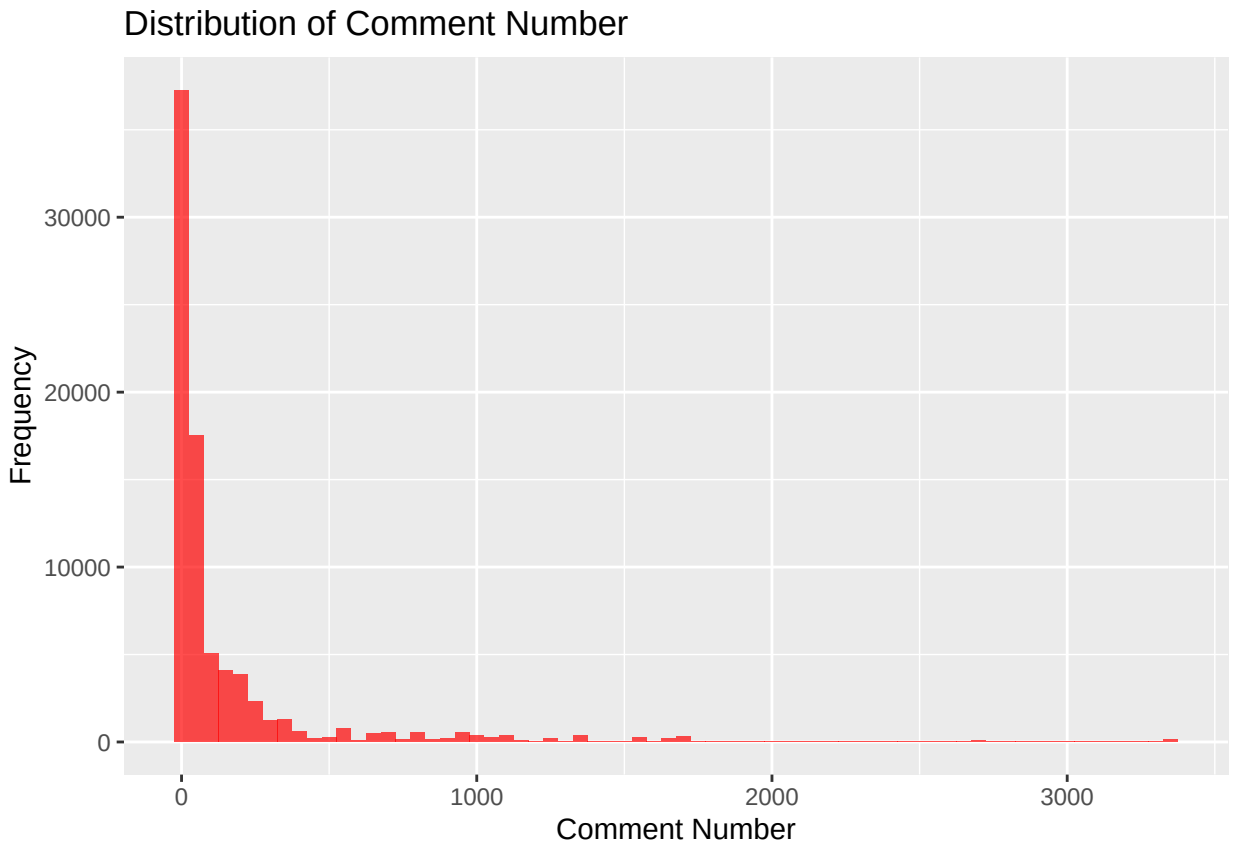
The distribution of prices and reviews is highly right-skewed, with most item prices and reviews clustered in the lower range and a small number of items with higher prices and reviews.

The distribution of item ratings is relatively concentrated, with most items rated in the higher range, showing more consistent user ratings.

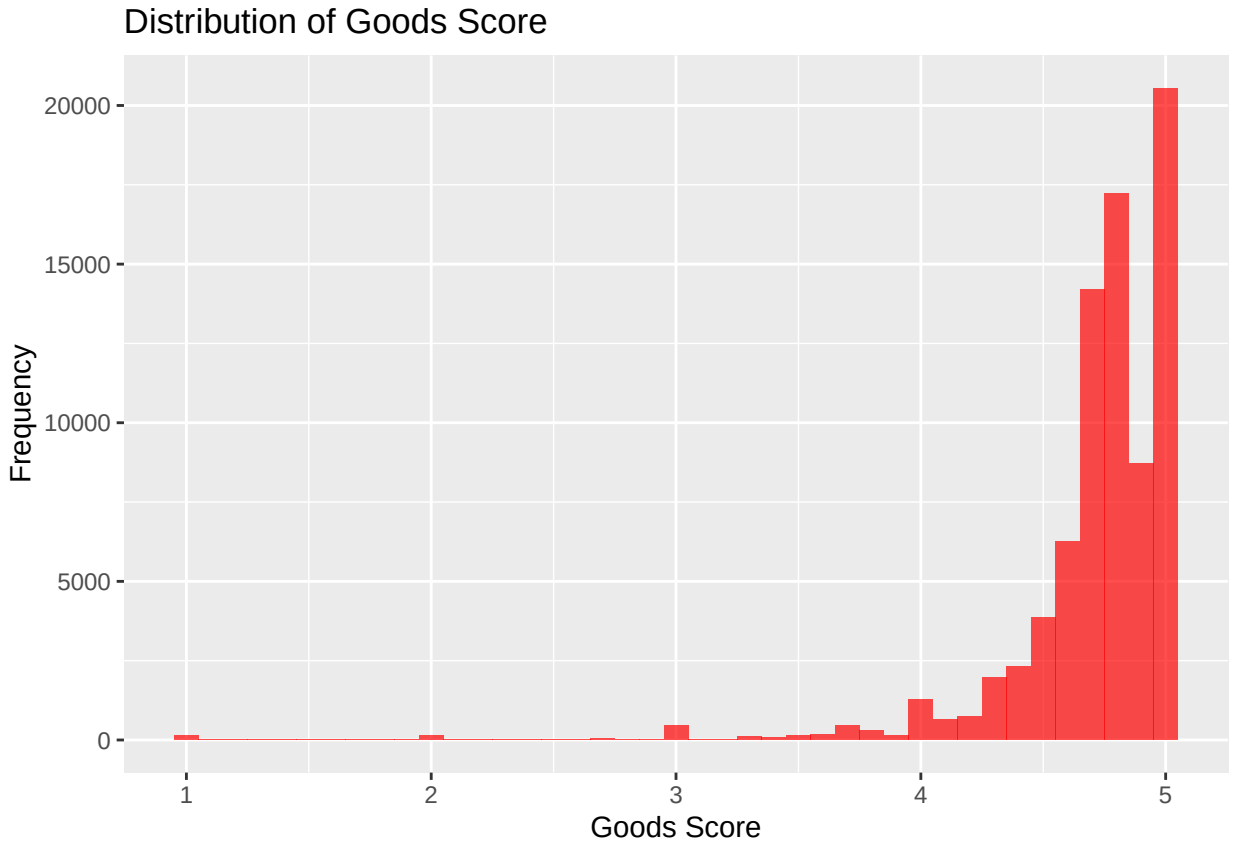
```
ggplot(data, aes(x = price)) +
  geom_histogram(binwidth = 5, fill = "red", alpha = 0.7) +
  labs(title = "Distribution of Price ", x = "Price", y = "Frequency")
```




```
ggplot(data, aes(x = comment_num)) +
  geom_histogram(binwidth = 50, fill = "red", alpha = 0.7) +
  labs(title = "Distribution of Comment Number ", x = "Comment Number", y = "Frequency")
```




```
ggplot(data, aes(x = goods_score)) +
  geom_histogram(binwidth = 0.1, fill = "red", alpha = 0.7) +
  labs(title = "Distribution of Goods Score ", x = "Goods Score", y = "Frequency")
```



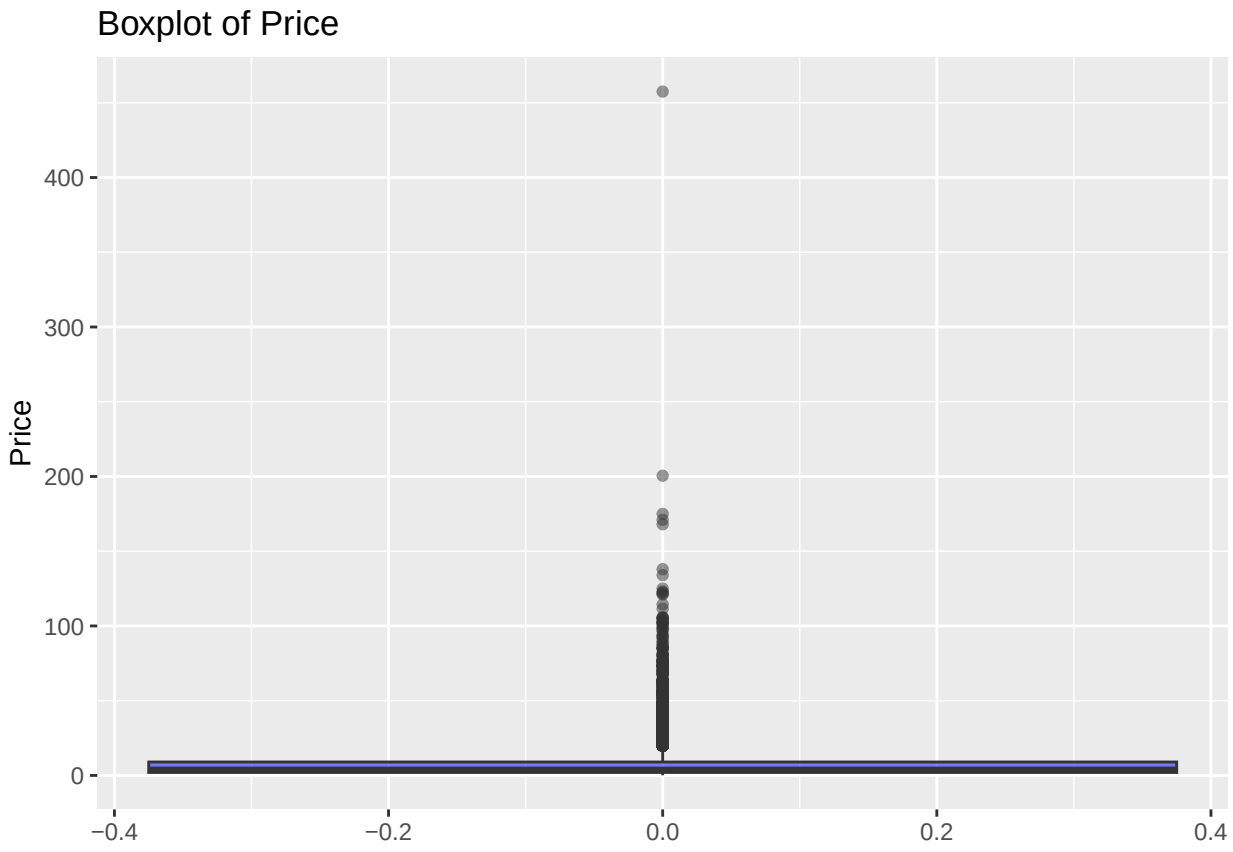
2.3 Boxplot

Boxplot of Price: The price distribution shows a concentration of values near the lower range with numerous outliers extending far beyond the upper whisker, indicating a heavily right-skewed distribution with a few very high-priced items.

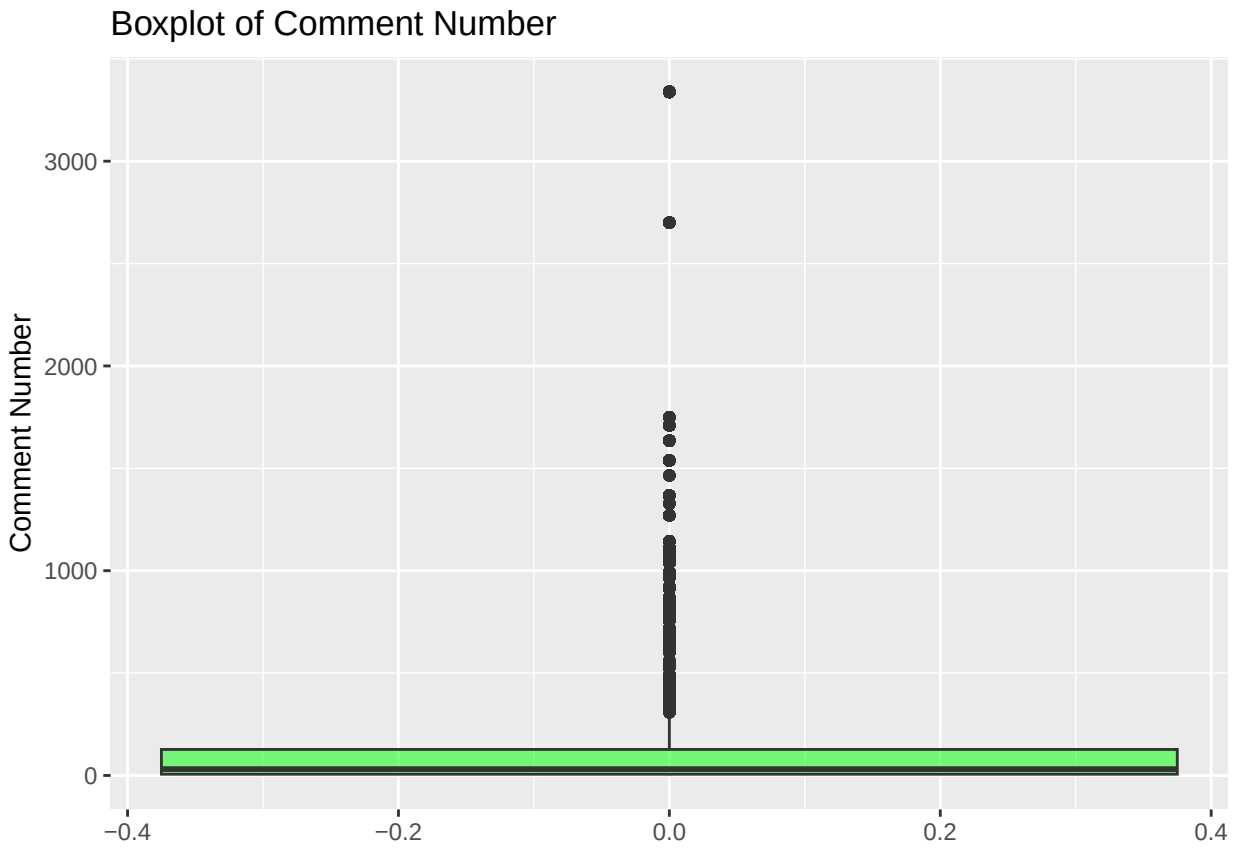
Boxplot of Comment Number: Most products have a low number of comments, with several extreme outliers indicating a small subset of products receiving a significantly higher number of comments compared to the majority.

Boxplot of Goods Score: The goods score distribution is tightly clustered near the higher range (around 5), with some lower outliers, suggesting most products are rated highly with only a few exceptions.

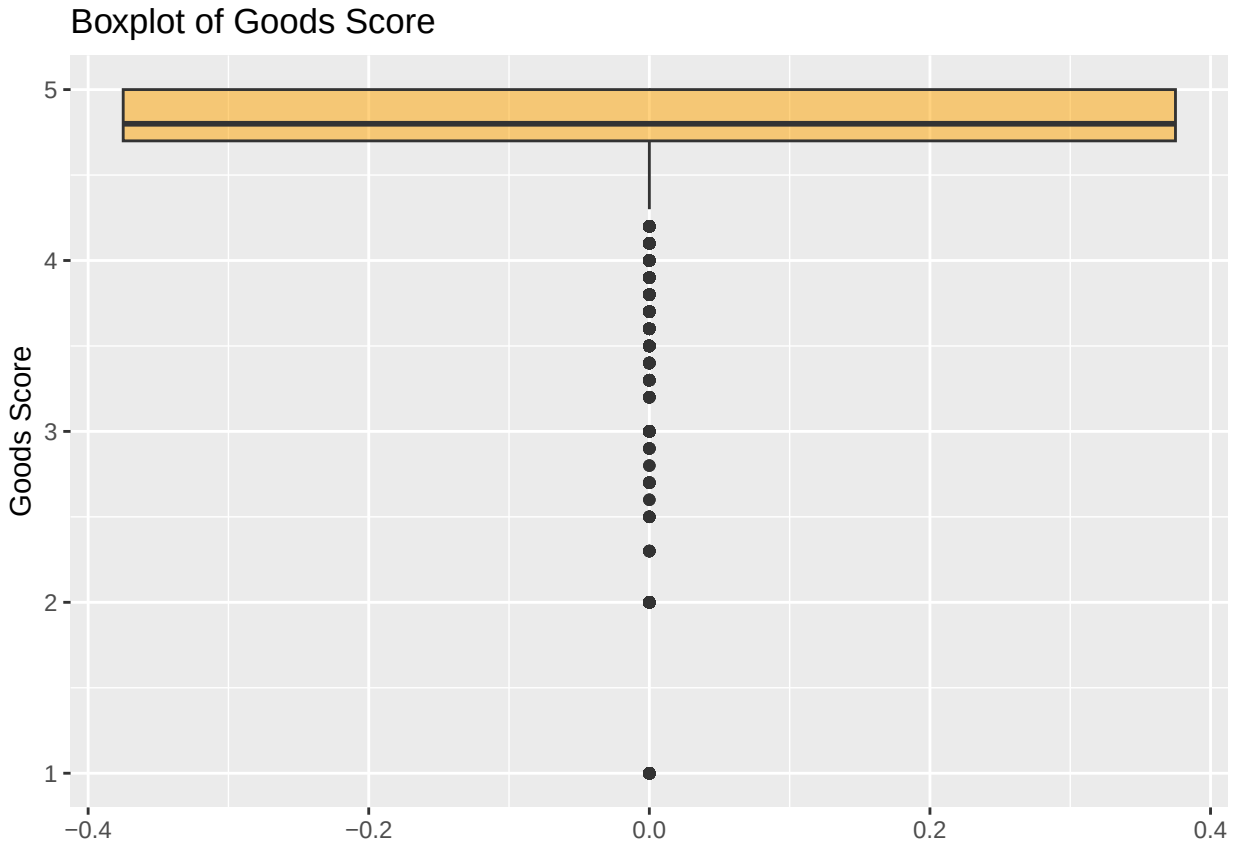
```
  # Boxplot of Price
ggplot(data, aes(y = price)) +
  geom_boxplot(fill = "blue", alpha = 0.5) +
  labs(title = "Boxplot of Price", y = "Price")
```




```
# Boxplot of Comment number
ggplot(data, aes(y = comment_num)) +
  geom_boxplot(fill = "green", alpha = 0.5) +
  labs(title = "Boxplot of Comment Number", y = "Comment Number")
```




```
# Boxplot of Score
ggplot(data, aes(y = goods_score)) +
  geom_boxplot(fill = "orange", alpha = 0.5) +
  labs(title = "Boxplot of Goods Score", y = "Goods Score")
```



2.3.1 Scatterplot

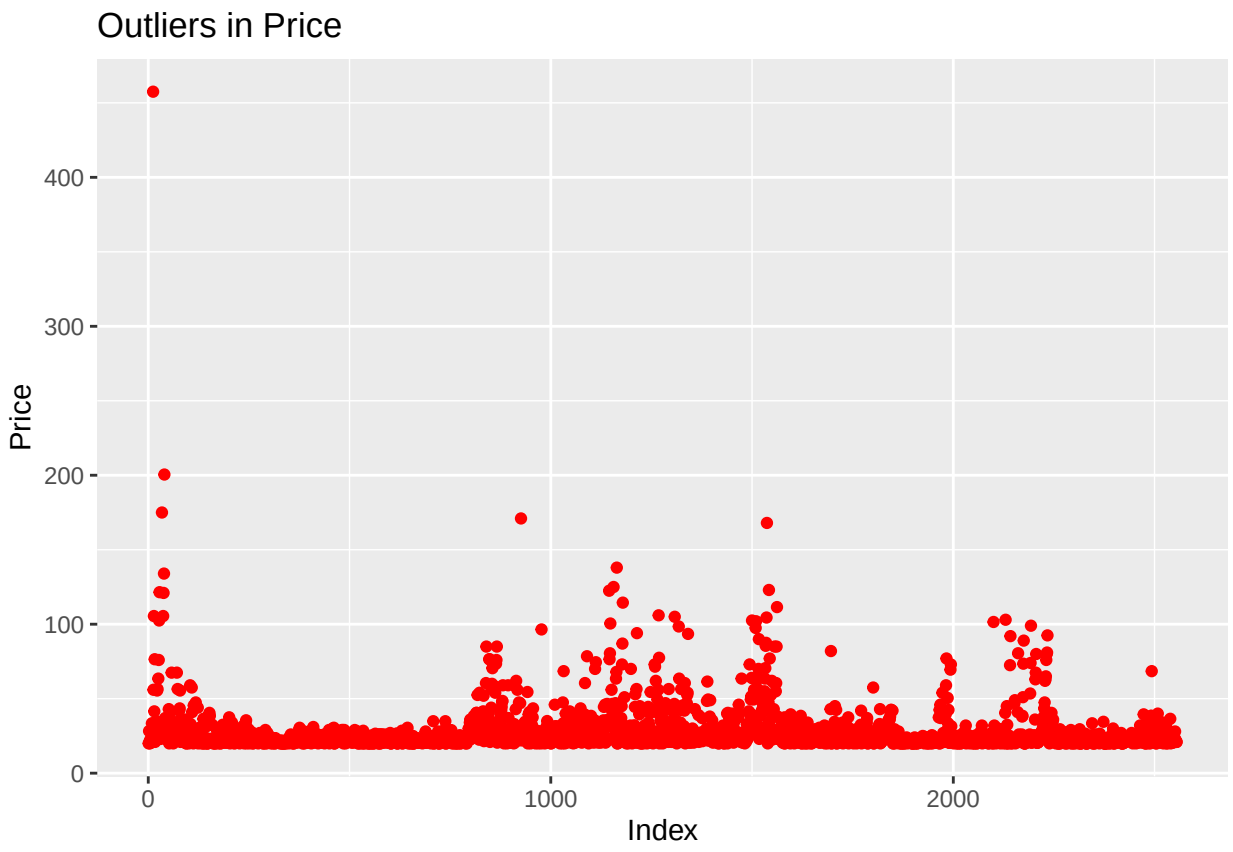
Outliers in Price: The scatterplot shows that most prices are clustered near the lower range, with several extreme outliers above 100, and a few even exceeding 400. These outliers could significantly influence model predictions and should be addressed, potentially through transformation or robust modeling techniques.

Outliers in Comment Number: Comment numbers are widely distributed, with most values under 1000, but several extreme outliers exceed 2000 and even 3000. These extreme values indicate high engagement for a few products, which may disproportionately impact models sensitive to outliers.

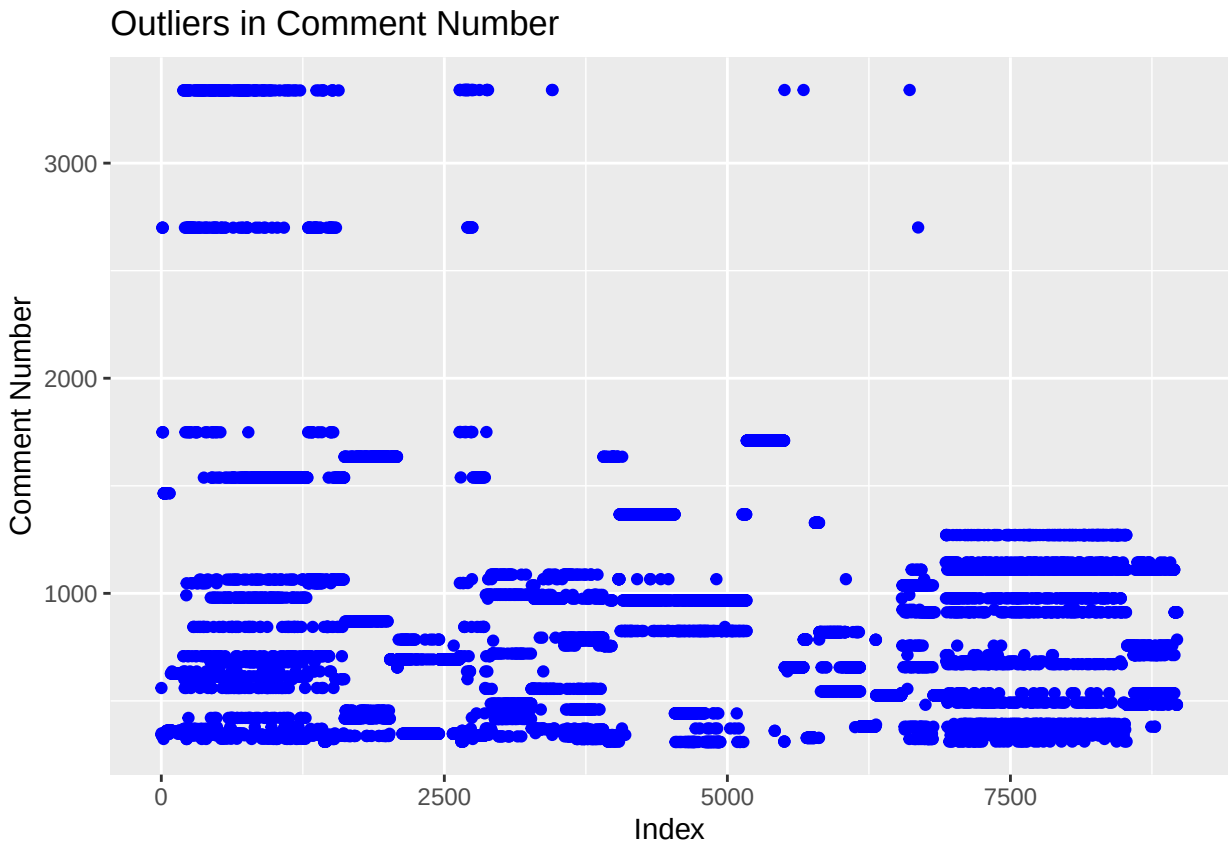
Outliers in Goods Score: Goods scores are concentrated around 4-5, with scattered outliers as low as 1. This indicates that most products are rated highly, with only a small proportion receiving low scores. The limited variability might reduce the ability to predict meaningful differences.

```
detect_outliers <- function(data, column) {
  Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
  Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
  IQR <- Q3 - Q1
  lower_bound <- Q1 - 1.5 * IQR
  upper_bound <- Q3 + 1.5 * IQR
  outliers <- data %>%
    filter(data[[column]] < lower_bound | data[[column]] > upper_bound)
  return(outliers)
}

# Extract price outliers
outliers_price <- detect_outliers(data, "price")

# Extract comment_num exceptions
```

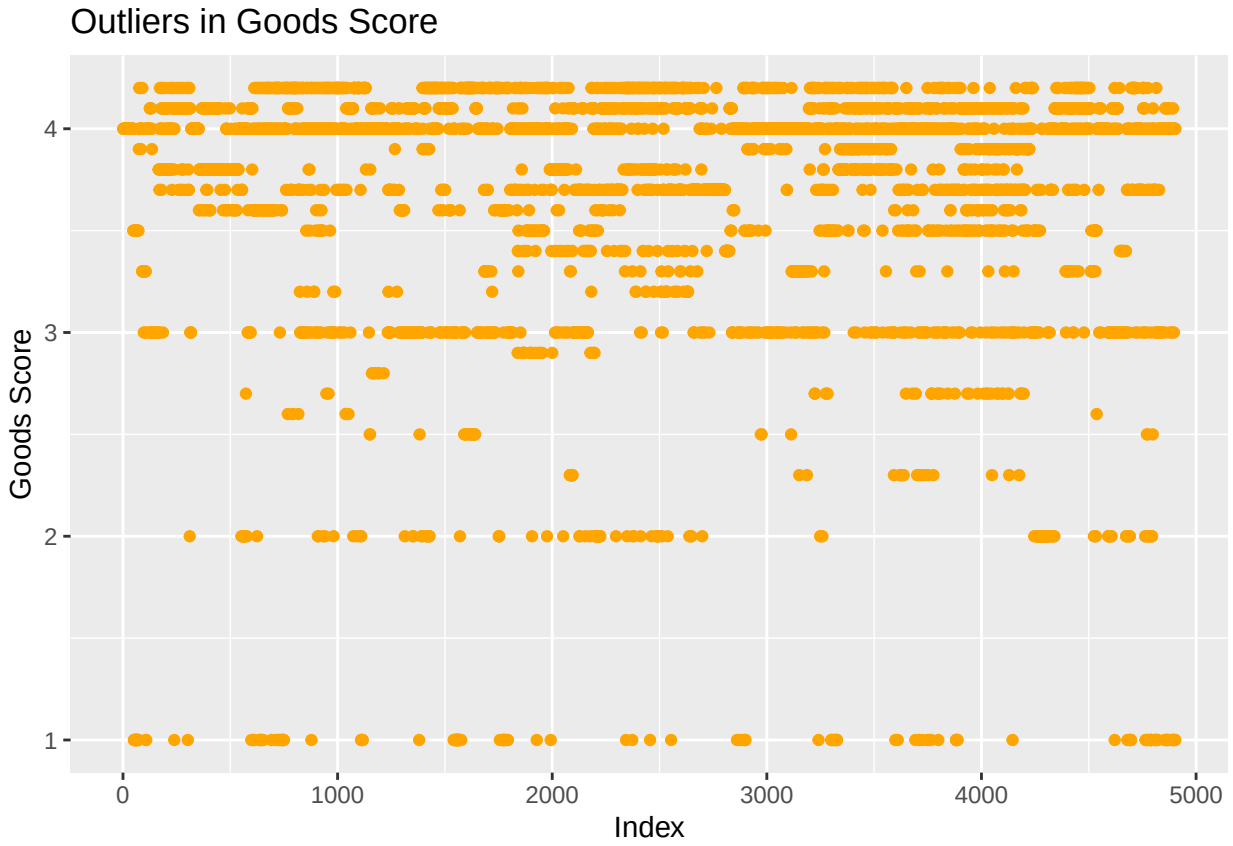
```r
outliers_comment <- detect_outliers(data, "comment_num")

# Extracting exceptions from goods_score
outliers_goods_score <- detect_outliers(data, "goods_score")


# Plotting a Scatter Plot of Price Outliers
ggplot(outliers_price, aes(x = seq_along(price), y = price)) +
  geom_point(color = "red") +
  labs(title = "Outliers in Price", x = "Index", y = "Price")
```




```r
# Plotting a Scatter Plot of Comment Number Anomalies
ggplot(outliers_comment, aes(x = seq_along(comment_num), y = comment_num)) +
  geom_point(color = "blue") +
  labs(title = "Outliers in Comment Number", x = "Index", y = "Comment Number")
```




```r
# Plotting a Scatter Plot of Goods Score Outliers
ggplot(outliers_goods_score, aes(x = seq_along(goods_score), y = goods_score)) +
  geom_point(color = "orange") +
  labs(title = "Outliers in Goods Score", x = "Index", y = "Goods Score")
```




```
cat("Number of outliers in price:", nrow(outliers_price), "\n")
```

```
## Number of outliers in price: 2555
```

```
cat("Number of outliers in comment_num:", nrow(outliers_comment), "\n")
```

```
## Number of outliers in comment_num: 8974
```

```
cat("Number of outliers in goods_score:", nrow(outliers_goods_score), "\n")
```

```
## Number of outliers in goods_score: 4904
```

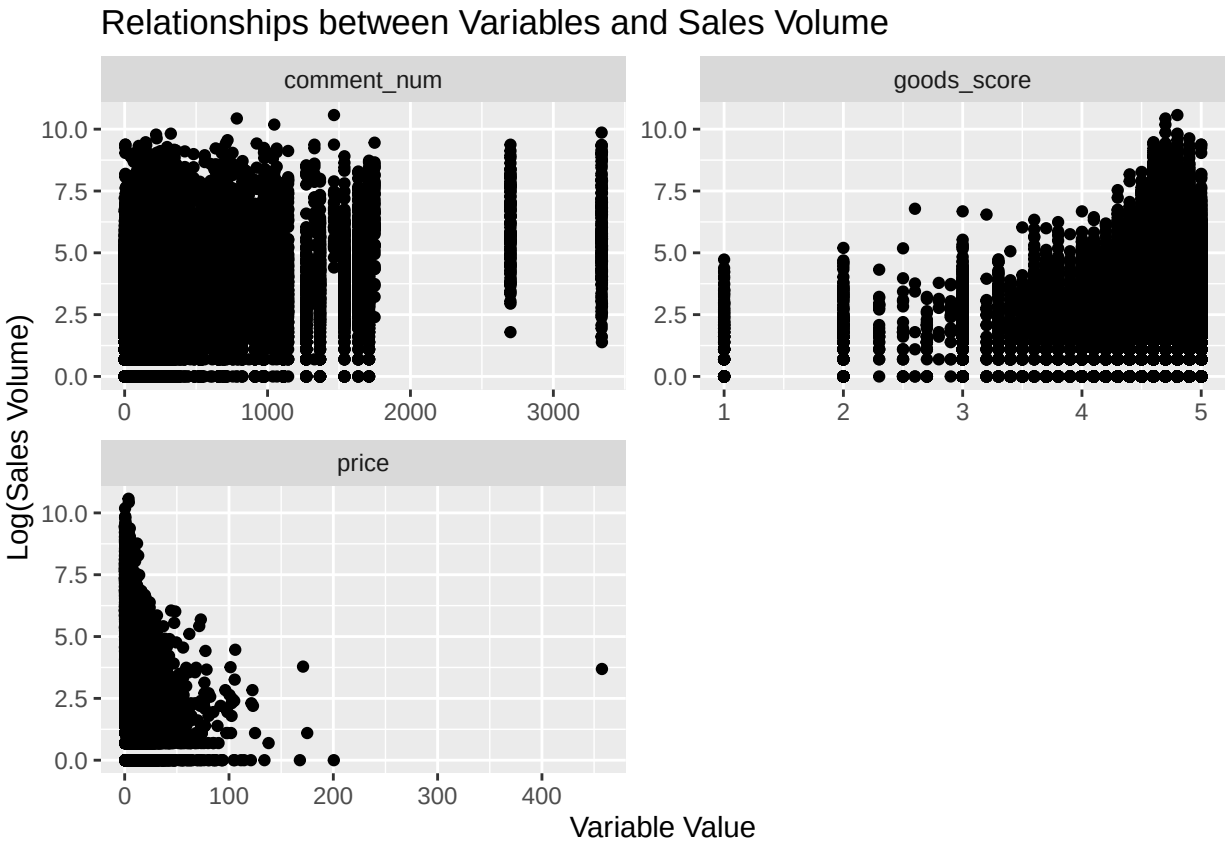
2.3.2 log transformation

comment_num vs. log(sales_volume): The plot indicates a positive association between the number of comments and the logarithm of sales volume. Products with more comments generally tend to have higher sales volume, but there is substantial variance at higher comment levels.

goods_score vs. log(sales_volume): A positive trend is evident, with higher goods scores associated with increased sales volume (in log scale). However, the variability suggests that other factors may also influence sales volume.

price vs. log(sales_volume): There is a noticeable negative association, with higher prices generally correlating with lower sales volume. This relationship weakens at higher price levels, possibly due to fewer data points or niche markets for expensive products.

```
data %>%
  pivot_longer(cols = c(price, comment_num, goods_score), names_to = "xname", values_to = "x") %>%
  ggplot(aes(x = x, y = log(sales_volume))) +
  geom_point() +
  facet_wrap(~xname, ncol = 2, scales = "free") +
  labs(title = "Relationships between Variables and Sales Volume",
       x = "Variable Value",
       y = "Log(Sales Volume)")
```

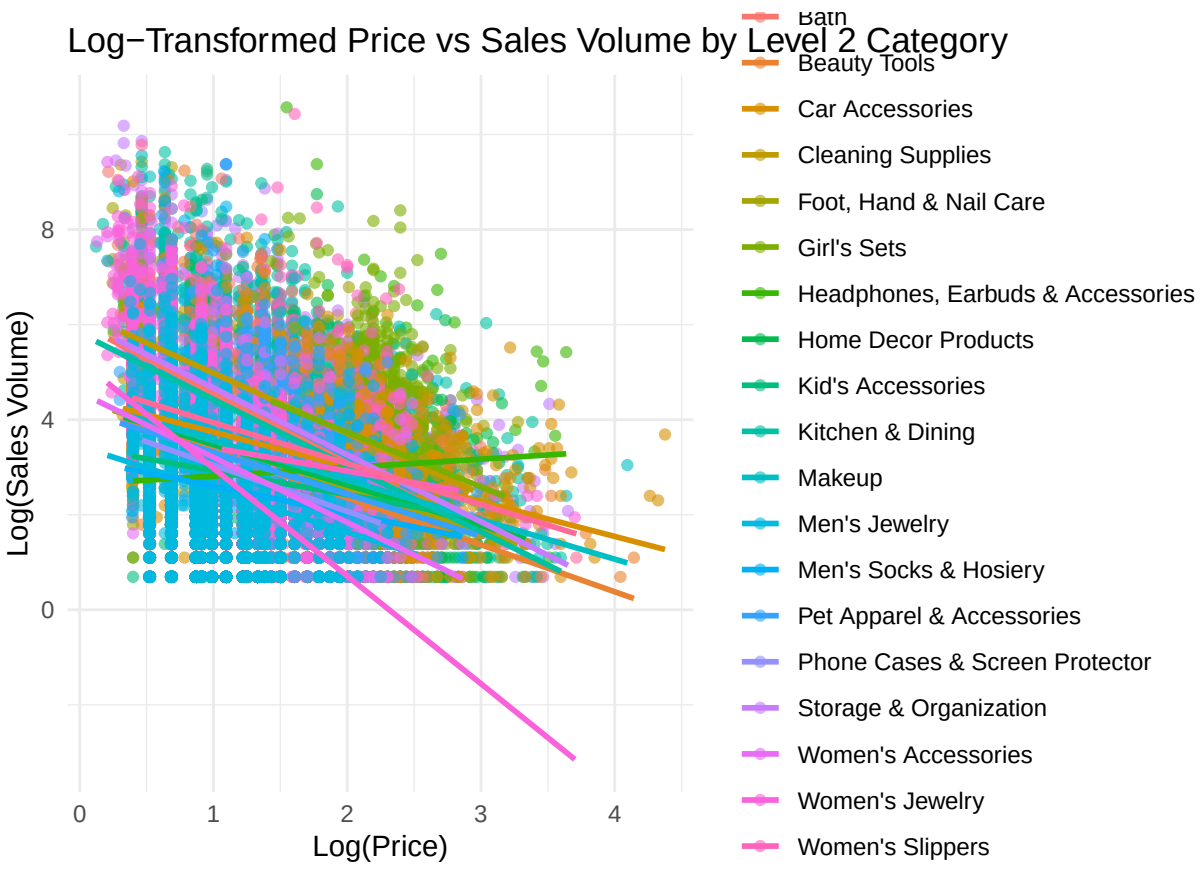


2.3.3 log transformed

The figure shows the difference in the relationship between Log(Price) and Log(Sales Volume) by Level 2 Category. The main findings are as follows:

```
ggplot(filtered_data, aes(x = log(price + 1), y = log(sales_volume + 1), color = leve_2_category_name)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Log-Transformed Price vs Sales Volume by Level 2 Category",
       x = "Log(Price)", y = "Log(Sales Volume)", color = "Level 2 Category") +
  theme_minimal()
```

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



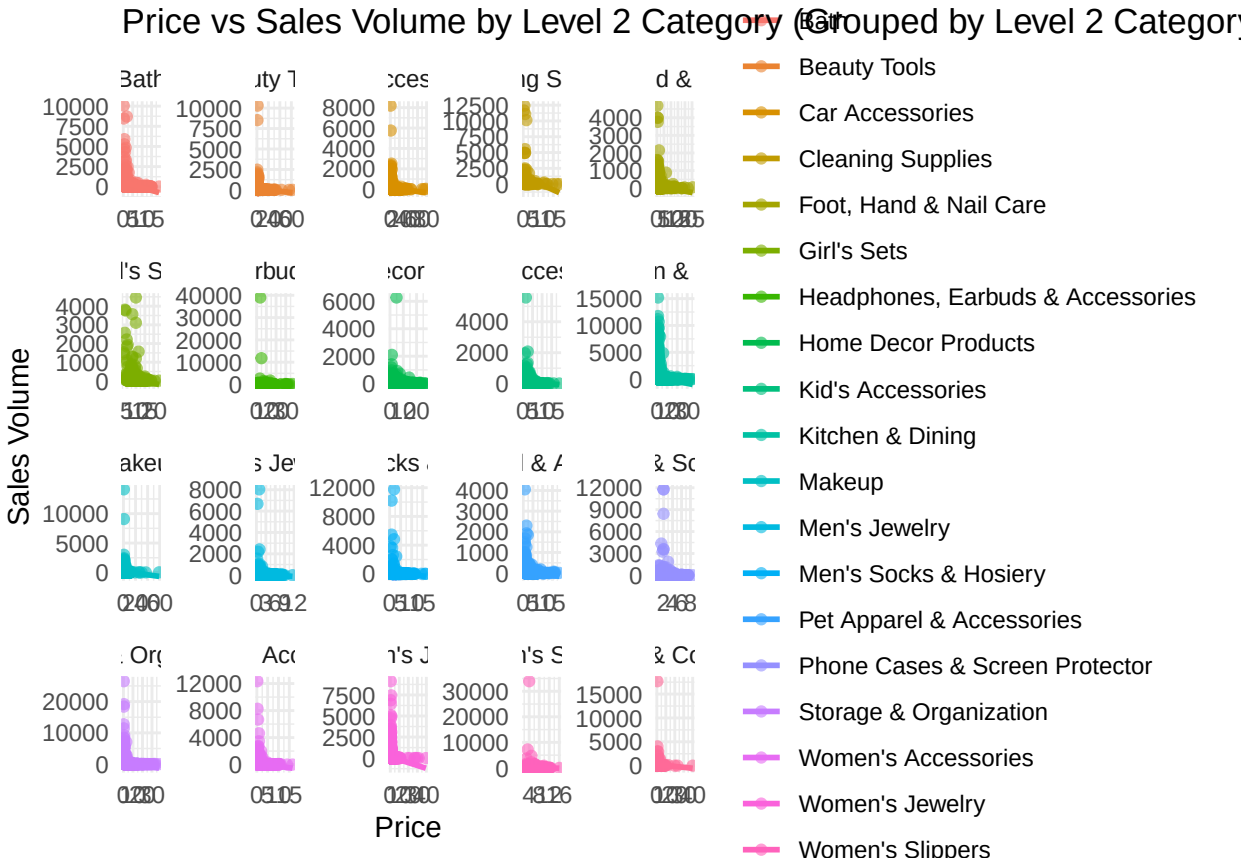
2.3.4 grouped

This figure shows the relationship between price and sales for the Level 2 category within the different Level 1 categories.

In most cases, price is negatively correlated with sales, i.e. the higher the price, the lower the sales. Each subfigure clearly demonstrates the difference in the distribution of secondary categories within the different primary categories.

```
ggplot(filtered_data, aes(x = price, y = sales_volume, color = leve_2_category_name)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(~ leve_2_category_name, scales = "free") +
  labs(title = "Price vs Sales Volume by Level 2 Category (Grouped by Level 2 Category)",
       x = "Price", y = "Sales Volume", color = "Level 2 Category") +
  theme_minimal()
```

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



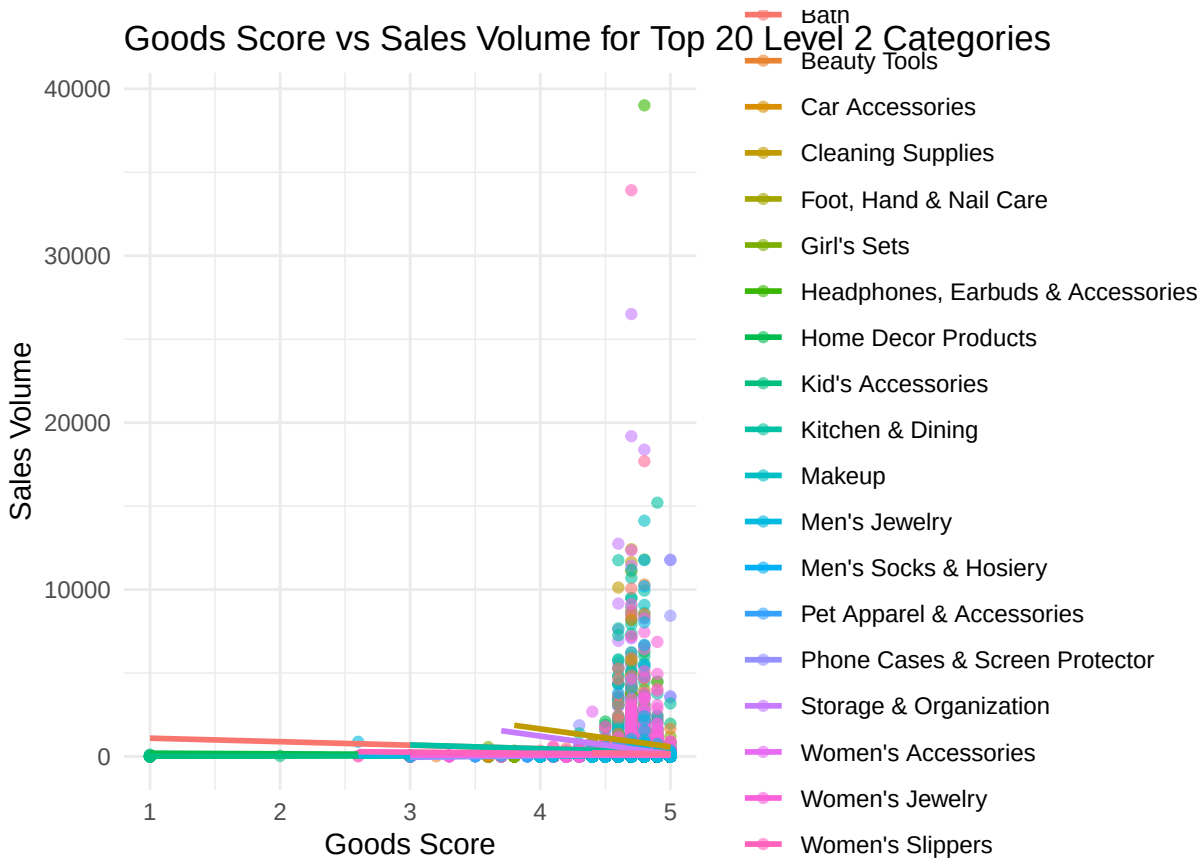
2.4 score rating

Some categories, such as "Cleaning Supplies" and "Phone Cases & Screen Protector," show sharper increases in sales volume at higher goods scores. In contrast, others, like "Foot, Hand & Nail Care," exhibit relatively flat trends, indicating less sensitivity of sales volume to goods scores in these categories.

Outliers: Certain categories, such as "Girl's Sets," show extreme sales volume for high goods scores, indicating popular products or high-demand items within that category. These outliers might skew trends.

```
ggplot(filtered_data, aes(x = goods_score, y = sales_volume, color = leve_2_category_name)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Goods Score vs Sales Volume for Top 20 Level 2 Categories",
       x = "Goods Score", y = "Sales Volume", color = "Level 2 Category") +
  theme_minimal()
```

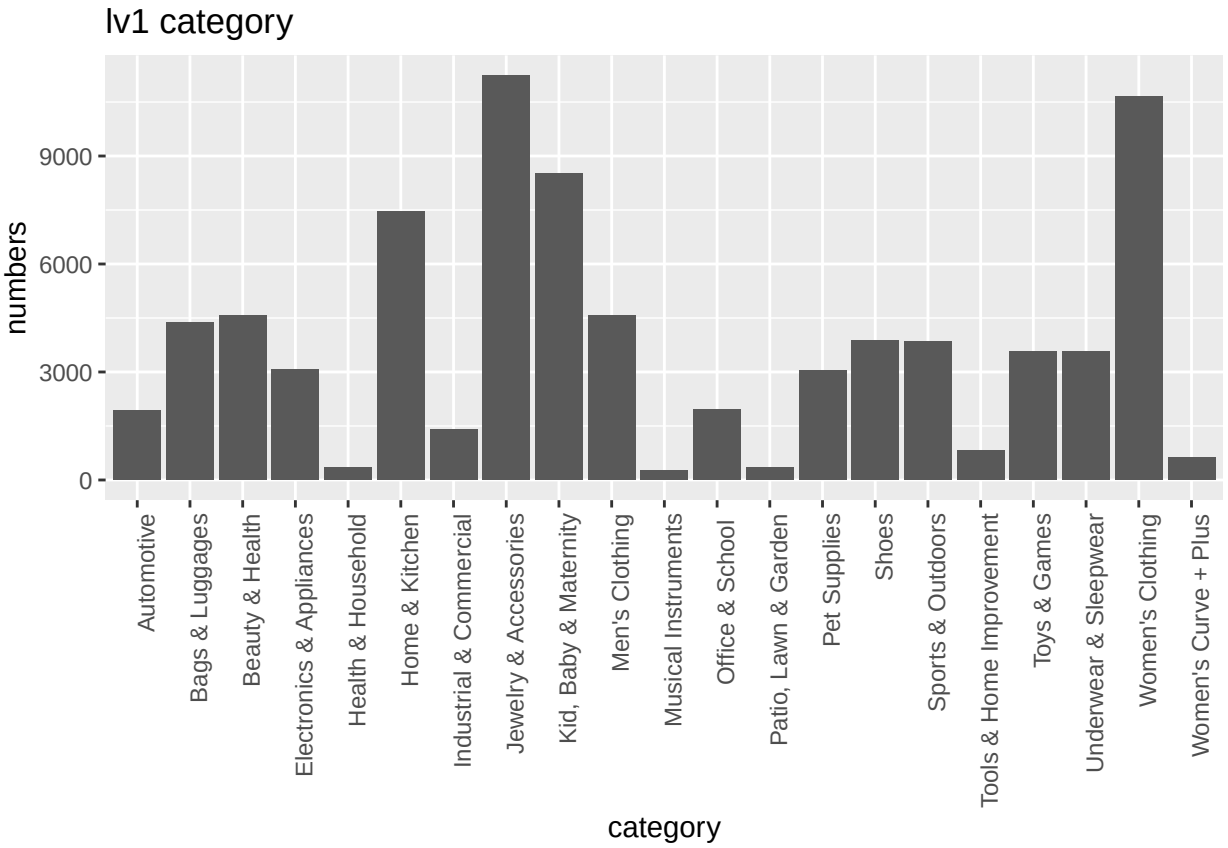
```
## `geom_smooth()` using formula = 'y ~ x'
```



2.5 lv1 category

The bar chart shows the distribution of the number of items in each level of category. The categories "Clothing" and "Accessories" have the highest number of items, which is significantly higher than the other categories, while categories such as "Appliances" and "Plus" have a lower number of items. This suggests that goods are not evenly distributed among the first-level categories, and that the impact of categories needs to be taken into account in the analysis.

```
# Level 2 categroy counts eda
category_counts <- data %>%
  count(leve_2_category_name, name = "category_count")
ggplot(data, aes(x=leve_1_category_name))+
  geom_bar()+
  labs(title= "lv1 category", x="category",y="numbers")+
    theme(axis.text.x= element_text(angle=90, hjust=1)
  )
```



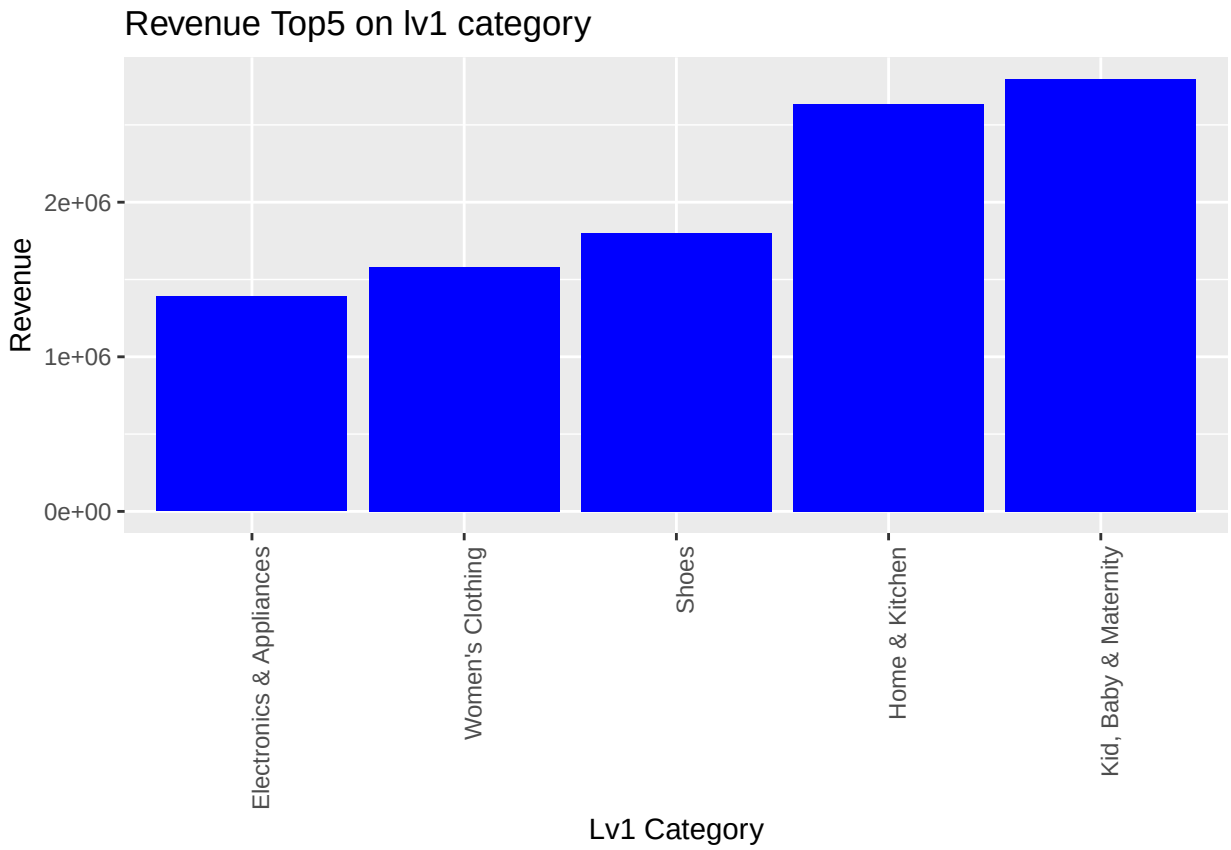
2.5 Sales Revenue

group by level1 and level 2 category

```
revenue_level1 <- data %>%
  group_by(leve_1_category_name) %>%
  summarise(total_revenue = sum(revenue, na.rm = TRUE)) %>%
  ungroup()
# level2 sales revenue
revenue_level2 <- data %>%
  group_by(leve_2_category_name) %>%
  summarise(total_revenue = sum(revenue, na.rm = TRUE)) %>%
  ungroup()
```

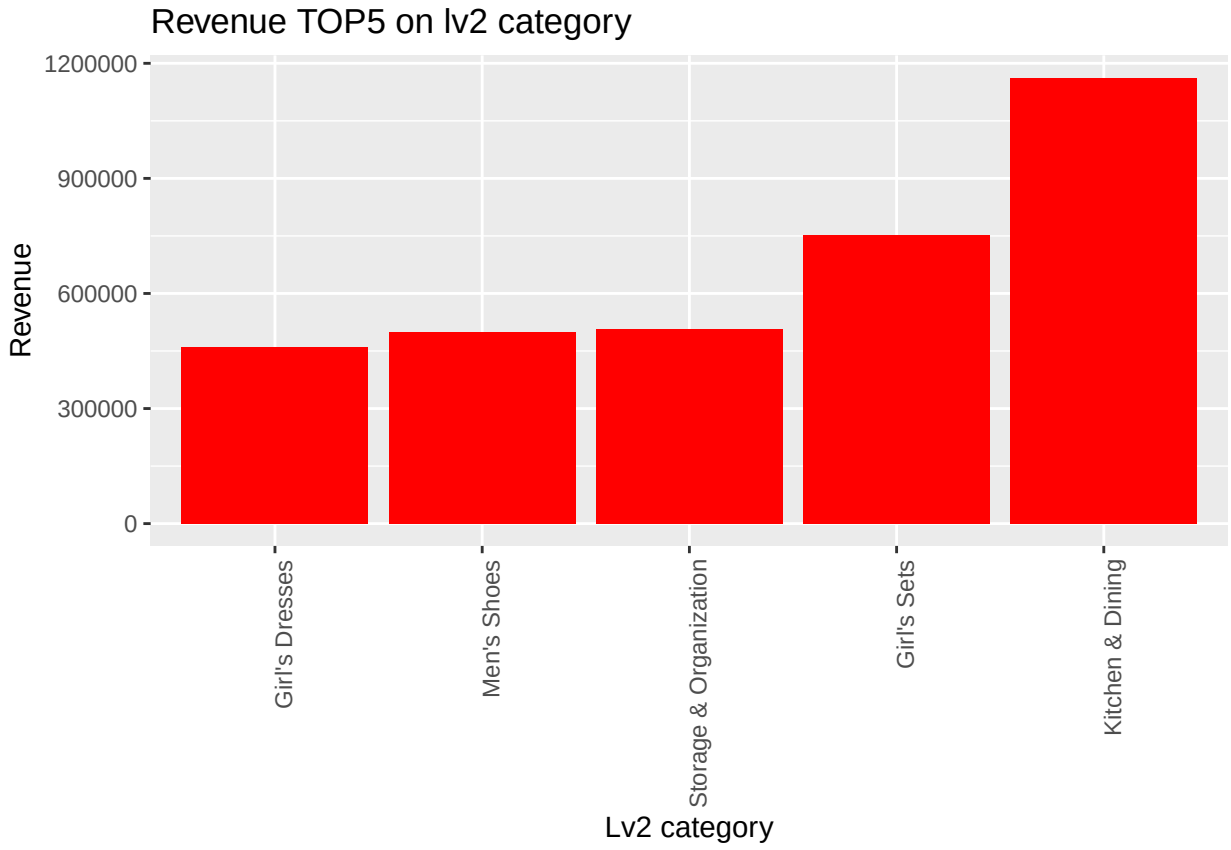
2.5.1. lv1 revenue distribution

```
top_revenue_level1 <- revenue_level1 %>%
  top_n(5, total_revenue)
# Income distribution by primary category
ggplot(top_revenue_level1, aes(x = reorder(leve_1_category_name, total_revenue), y = total_revenue)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Revenue Top5 on lv1 category", x = "Lv1 Category", y = "Revenue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



2.5.3 lv2 distribution

```
top_revenue_level2 <- revenue_level2 %>%
  top_n(5, total_revenue)
# Revenue Top5 on lv2 category
ggplot(top_revenue_level2, aes(x = reorder(leve_2_category_name, total_revenue), y = total_revenue)) +
  geom_bar(stat = "identity", fill = "red") +
  labs(title = "Revenue TOP5 on lv2 category ", x = "Lv2 category", y = "Revenue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



3 Model Building

```
# Extract names of Top5 categories from top_revenue_level2
top5_categories <- unlist(top_revenue_level2$leve_2_category_name, use.names = FALSE)

# Ensure that the raw data columns are character-based
data$leve_2_category_name <- as.character(data$leve_2_category_name)

# Extract data belonging to the Top5 categories from the raw data
top5_data <- data %>%
  filter(leve_2_category_name %in% top5_categories)

# Ensure that all variables (e.g., sales_volume and price) are non-missing and numeric
top5_data <- top5_data %>%
  filter(!is.na(sales_volume) & !is.na(price)) %>%
  mutate(
    sales_volume = as.numeric(sales_volume),
    price = as.numeric(price)
  )
```

Residual Distribution.

The large range of residuals (-1588 to 38475) indicates that the model underperforms on high-volume items. The low significance of the multinomial leve_2_category_name may indicate the presence of multicollinearity or noise interference in the category variable. Key Variables.

price: for each unit increase, sales decrease 4.39 4.39 units (significant, p-value $<$ 2e-16). high_score: not significant, indicating that whether an item is highly rated or not has limited contribution to sales prediction.

Significant categories: e.g. Adhesives & Sealants (positive effect), Cleaning Supplies (positive effect). Model Performance.

3.1 linear model

```
model_lm <- lm(sales_volume ~ price + high_score + leve_1_category_name + leve_2_category_name + commen

# Output Model Summary
summary(model_lm)
```

```
##
## Call:
## lm(formula = sales_volume ~ price + high_score + leve_1_category_name +
##     leve_2_category_name + comment_num, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -1588    -54    -16      8  38475
##
## Coefficients: (19 not defined because of singularities)
##                                                    Estimate
## (Intercept)                                       1.619e+02
## price                                            -4.392e+00
## high_score                                        5.126e+00
## leve_1_category_nameBags & Luggages              -1.409e+02
## leve_1_category_nameBeauty & Health              -8.690e+01
## leve_1_category_nameElectronics & Appliances     -1.026e+02
## leve_1_category_nameHealth & Household            1.049e+01
## leve_1_category_nameHome & Kitchen               -1.532e+02
## leve_1_category_nameIndustrial & Commercial      -1.380e+02
## leve_1_category_nameJewelry & Accessories        -1.493e+02
## leve_1_category_nameKid, Baby & Maternity        -1.510e+02
## leve_1_category_nameMen's Clothing               -1.032e+02
## leve_1_category_nameMusical Instruments          -2.247e+01
## leve_1_category_nameOffice & School              -4.708e+01
## leve_1_category_namePatio, Lawn & Garden         -3.330e+01
## leve_1_category_namePet Supplies                 -6.707e+01
## leve_1_category_nameShoes                        -5.929e+01
## leve_1_category_nameSports & Outdoors            -9.639e+01
## leve_1_category_nameTools & Home Improvement     -4.802e+01
## leve_1_category_nameToys & Games                 -2.598e+02
## leve_1_category_nameUnderwear & Sleepwear        -1.090e+02
## leve_1_category_nameWomen's Clothing             -1.269e+02
## leve_1_category_nameWomen's Curve + Plus         -1.242e+02
## leve_2_category_nameAccessories                  -1.172e+02
## leve_2_category_nameAdhesives & Sealants          3.470e+02
## leve_2_category_nameAir Quality & Irons           1.636e+02
## leve_2_category_nameArts & Crafts                 2.249e+02
## leve_2_category_nameBath                          2.139e+02
## leve_2_category_nameBeauty Tools                 -4.262e+01
## leve_2_category_nameBedding                       6.197e+01
## leve_2_category_nameBirds & Poultries            -1.113e+02
## leve_2_category_nameBoy's Bottoms                 3.521e+01
## leve_2_category_nameBoy's Coat                    9.194e+01
```

```
## leve_2_category_nameBoy's Sets                                  4.350e+01
## leve_2_category_nameBoy's Shoes                                 4.833e+01
## leve_2_category_nameBoy's Tops                                  3.379e+01
## leve_2_category_nameCar Accessories                            -1.379e+02
## leve_2_category_nameCar Storage & Organizers                    6.029e+01
## leve_2_category_nameCar Toys                                    2.752e+02
## leve_2_category_nameCleaning Supplies                           7.596e+02
## leve_2_category_nameCommercial Lighting                         1.175e+02
## leve_2_category_nameComputer Accessories                        5.313e+01
## leve_2_category_nameCurve + Plus Beachwear                      3.234e+00
## leve_2_category_nameCurve + Plus Bottoms                        1.742e+01
## leve_2_category_nameCurve + Plus Dresses & Jumpsuits            3.230e+01
## leve_2_category_nameCurve + Plus Tops & Outerwear               1.964e+01
## leve_2_category_nameCurve + Plus Two Pieces Set                 3.795e+01
## leve_2_category_nameCurve + Plus Underwear & Sleepwear                 NA
## leve_2_category_nameCutting Tools                               1.729e+01
## leve_2_category_nameCycling                                    -1.911e+01
## leve_2_category_nameDogs & Cats Supplies                       -7.278e+01
## leve_2_category_nameDolls & Plush Toys                          1.324e+02
## leve_2_category_nameEvent & Party Supplies                      2.614e+01
## leve_2_category_nameFish, Reptiles & Amphibians                -7.748e+01
## leve_2_category_nameFishing                                    -2.548e+01
## leve_2_category_nameFoot, Hand & Nail Care                     -2.133e+01
## leve_2_category_nameGardening & Urban Farming                  -4.754e-01
## leve_2_category_nameGirl's  Shoes                               6.308e+01
## leve_2_category_nameGirl's Bottoms                              9.027e+01
## leve_2_category_nameGirl's Coat                                 6.919e+01
## leve_2_category_nameGirl's Dresses                              9.533e+01
## leve_2_category_nameGirl's Sets                                 9.657e+01
## leve_2_category_nameGirl's Tops                                 2.766e+01
## leve_2_category_nameGreeting Cards & Postcards                 -1.537e+02
## leve_2_category_nameHair Care                                  -3.723e+01
## leve_2_category_nameHardware                                    1.966e+02
## leve_2_category_nameHeadphones, Earbuds & Accessories           9.527e+01
## leve_2_category_nameHealth Care Products                       -1.981e+02
## leve_2_category_nameHome Appliances                             1.714e+01
## leve_2_category_nameHome Decor Products                         3.018e+01
## leve_2_category_nameHome Use Medical Supplies & Equipment      -1.003e+02
## leve_2_category_nameHousehold Supplies                         -2.143e+01
## leve_2_category_nameIndustrial Materials                        1.511e+01
## leve_2_category_nameInstruments                                -8.991e+00
## leve_2_category_nameJewelry Accessories                         3.353e+01
## leve_2_category_nameKeychains&Key Shells                       -2.642e+02
## leve_2_category_nameKid's Accessories                           2.146e+01
## leve_2_category_nameKids' Electronics                           2.197e+02
## leve_2_category_nameKids' Home Store                            8.611e+01
## leve_2_category_nameKitchen & Dining                            1.704e+02
## leve_2_category_nameKitchen, Bath Fixtures & Appliances         1.456e+02
## leve_2_category_nameLearning & Education                        1.617e+02
## leve_2_category_nameLighting & Ceiling Fans                     5.477e+01
## leve_2_category_nameLights & Lighting Accessories              -1.038e+02
## leve_2_category_nameLuggage & Travel Gear                       1.638e+01
## leve_2_category_nameMakeup                                     -7.463e+00
## leve_2_category_nameMaternity Clothing                          2.667e+01
```

```
## leve_2_category_nameMaternity, Baby Gear & Nursery                NA
## leve_2_category_nameMen's Accessories                          8.961e+00
## leve_2_category_nameMen's Backbags                             5.579e+01
## leve_2_category_nameMen's Belts                                3.658e+01
## leve_2_category_nameMen's Boots                                1.348e+01
## leve_2_category_nameMen's Bottoms                              1.260e+01
## leve_2_category_nameMen's Clothing & Accessories              -3.369e+01
## leve_2_category_nameMen's Exotic & Novelty  Clothing          -3.189e+01
## leve_2_category_nameMen's Glasses                              4.124e+01
## leve_2_category_nameMen's Handbags                             4.210e+01
## leve_2_category_nameMen's Hats, Caps & Wigs                   -2.870e+01
## leve_2_category_nameMen's Jackets&Coats                        5.143e+01
## leve_2_category_nameMen's Jeans                                3.782e+01
## leve_2_category_nameMen's Jewelry                             -1.206e+01
## leve_2_category_nameMen's Loafers & Slip-Ons                  -4.561e+01
## leve_2_category_nameMen's Loungewear                           1.308e+01
## leve_2_category_nameMen's Mules&Clogs                         -5.234e+01
## leve_2_category_nameMen's Sandal & Slippers                   -3.996e+01
## leve_2_category_nameMen's Scarves & Gloves                     2.303e+01
## leve_2_category_nameMen's Shirts                              -1.833e+01
## leve_2_category_nameMen's Shoes                               -1.505e+01
## leve_2_category_nameMen's Shoulder Bags                        1.390e+01
## leve_2_category_nameMen's Socks & Hosiery                      2.010e+01
## leve_2_category_nameMen's Sports Shoes                         8.379e+01
## leve_2_category_nameMen's Suits                                1.692e+01
## leve_2_category_nameMen's Sweaters                             2.105e+01
## leve_2_category_nameMen's Sweatshirts                         -2.276e+00
## leve_2_category_nameMen's T-Shirts                            -2.414e+01
## leve_2_category_nameMen's Tops                                     NA
## leve_2_category_nameMen's Underwear                           -2.818e+00
## leve_2_category_nameMen's Waist & Chest bags                   2.325e+01
## leve_2_category_nameMen's Wallets & Card Cases                 2.537e+01
## leve_2_category_nameMen's Watches & Accessories                1.642e+01
## leve_2_category_nameMen's Work & Safety Shoes                 -6.732e+00
## leve_2_category_nameMotorcycles                               -1.620e+02
## leve_2_category_nameNovelty & Costumes                         1.615e+02
## leve_2_category_nameOccupational Health & Safety Products      3.567e+01
## leve_2_category_nameOffice & School Supplies                  -1.128e+02
## leve_2_category_nameOffice Electronics                        -3.303e+01
## leve_2_category_nameOrganization                              -7.303e+01
## leve_2_category_nameOther Animals                             -1.081e+02
## leve_2_category_nameOther Industrial Supplies                  1.865e+01
## leve_2_category_nameOther Toys & Games                         1.802e+02
## leve_2_category_nameOutdoor Lighting & Power Tools            -5.405e+01
## leve_2_category_nameOutdoor Lights                            -1.272e+01
## leve_2_category_nameOutdoor Recreation                        -1.685e+01
## leve_2_category_nameOutdoor Supplies & Patio Furniture             NA
## leve_2_category_namePaint, Wall Treatments & Supplies         -7.851e+01
## leve_2_category_namePapers, Labels & Indexes                  -7.553e+01
## leve_2_category_nameParty Supplies                             1.309e+02
## leve_2_category_namePersonal Care                              2.418e+00
## leve_2_category_namePersonal Care Electronics                 -5.540e+00
## leve_2_category_namePersonal Care Products                     1.886e+02
## leve_2_category_namePet Apparel & Accessories                 -7.426e+01
```

```
## leve_2_category_namePet Cleaning & Grooming                     NA
## leve_2_category_namePhone Accessories                        9.608e+00
## leve_2_category_namePhone Cables & Chargers                  1.744e+02
## leve_2_category_namePhone Cases & Screen Protector           2.482e+01
## leve_2_category_namePower & Hand Tools                       2.548e+01
## leve_2_category_namePower Banks & Batteries                  1.578e+02
## leve_2_category_namePower Tools & Safety                    -1.134e+01
## leve_2_category_namePretend Play                             1.683e+02
## leve_2_category_namePuzzles & Building Toys                  2.220e+02
## leve_2_category_nameQuadcopters & UAV                        2.489e+02
## leve_2_category_nameReplacement Parts                       -1.280e+02
## leve_2_category_nameSeasonal Decor                           4.309e+01
## leve_2_category_nameShoe Accessories                         5.204e-01
## leve_2_category_nameSports & Fitness Supplies                8.764e+00
## leve_2_category_nameSports Bags                             -1.627e+01
## leve_2_category_nameStationery & Gift Wrapping Supplies     -2.275e+02
## leve_2_category_nameStickers & Crafts Tape                       NA
## leve_2_category_nameStorage & Organization                   3.164e+02
## leve_2_category_nameStudio Recording & Stage Live                NA
## leve_2_category_nameTest, Measure & Inspect                  2.881e+01
## leve_2_category_nameTools & Equipment                            NA
## leve_2_category_nameWall Art                                     NA
## leve_2_category_nameWearable Technology                          NA
## leve_2_category_nameWellness & Relaxation Products               NA
## leve_2_category_nameWigs & Accessories                           NA
## leve_2_category_nameWinter Sports                           -2.430e+01
## leve_2_category_nameWomen's Accessories                      5.948e+01
## leve_2_category_nameWomen's Activewear                      -1.517e+01
## leve_2_category_nameWomen's Athleisure                       2.242e+01
## leve_2_category_nameWomen's Backbags                         3.422e+01
## leve_2_category_nameWomen's Beachwear                       -3.603e+00
## leve_2_category_nameWomen's Blazer                           3.440e+01
## leve_2_category_nameWomen's Bodysuits                       -6.879e+00
## leve_2_category_nameWomen's Boots                           -3.339e+00
## leve_2_category_nameWomen's Clutches & Evening Bags         -1.509e+01
## leve_2_category_nameWomen's Coat & Jacket                    6.510e+01
## leve_2_category_nameWomen's Cosplay Costume                  3.489e+01
## leve_2_category_nameWomen's Crossbody Bags                   2.349e+00
## leve_2_category_nameWomen's Denims                           4.510e+01
## leve_2_category_nameWomen's Dresses                          1.217e+01
## leve_2_category_nameWomen's Fashion Sneakers                 9.507e+00
## leve_2_category_nameWomen's Flats                            4.330e+00
## leve_2_category_nameWomen's Glasses                          3.032e+01
## leve_2_category_nameWomen's Handbags                         1.698e+01
## leve_2_category_nameWomen's Hats & Caps                     -1.845e+01
## leve_2_category_nameWomen's Jewelry                         -6.035e+00
## leve_2_category_nameWomen's Jumpsuits                        5.056e+00
## leve_2_category_nameWomen's Lingerie                         5.065e+00
## leve_2_category_nameWomen's Pants                            7.282e+00
## leve_2_category_nameWomen's Pumps                           -2.711e+01
## leve_2_category_nameWomen's Sandals                         -5.678e+00
## leve_2_category_nameWomen's Sexy Lingerie                   -2.293e+01
## leve_2_category_nameWomen's Shapewear & Others               4.940e+00
## leve_2_category_nameWomen's Shorts                           6.144e-01
```

```
## leve_2_category_nameWomen's Shoulder Bags                    4.281e+00
## leve_2_category_nameWomen's Skirts                           4.539e+00
## leve_2_category_nameWomen's Sleepwear                               NA
## leve_2_category_nameWomen's Slippers                                NA
## leve_2_category_nameWomen's Sports Shoes                     4.507e+01
## leve_2_category_nameWomen's Stockings & Hosiery              2.692e+01
## leve_2_category_nameWomen's Suits                            7.198e+00
## leve_2_category_nameWomen's Sweaters                         2.150e+01
## leve_2_category_nameWomen's Sweatshirts                      1.581e+01
## leve_2_category_nameWomen's Tops                                    NA
## leve_2_category_nameWomen's Tote Bags                        9.602e-01
## leve_2_category_nameWomen's Waist & Chest Bags               6.917e+00
## leve_2_category_nameWomen's Wallets & Card Cases                    NA
## leve_2_category_nameWomen's Watches                                 NA
## leve_2_category_nameWriting Supplies & Correction Supplies          NA
## leve_2_category_nameYoga&Studio                                     NA
## comment_num                                                  2.682e-01
##                                                             Std. Error t value
## (Intercept)                                                  2.912e+01   5.560
## price                                                        3.114e-01 -14.101
## high_score                                                   4.388e+00   1.168
## leve_1_category_nameBags & Luggages                          3.459e+01  -4.074
## leve_1_category_nameBeauty & Health                          3.557e+01  -2.443
## leve_1_category_nameElectronics & Appliances                 3.879e+01  -2.645
## leve_1_category_nameHealth & Household                       1.025e+02   0.102
## leve_1_category_nameHome & Kitchen                           4.004e+01  -3.826
## leve_1_category_nameIndustrial & Commercial                  6.012e+01  -2.295
## leve_1_category_nameJewelry & Accessories                    4.219e+01  -3.538
## leve_1_category_nameKid, Baby & Maternity                    3.353e+01  -4.503
## leve_1_category_nameMen's Clothing                           4.092e+01  -2.522
## leve_1_category_nameMusical Instruments                      6.389e+01  -0.352
## leve_1_category_nameOffice & School                          3.447e+01  -1.366
## leve_1_category_namePatio, Lawn & Garden                     4.521e+01  -0.737
## leve_1_category_namePet Supplies                             3.632e+01  -1.847
## leve_1_category_nameShoes                                    3.286e+01  -1.804
## leve_1_category_nameSports & Outdoors                        4.999e+01  -1.928
## leve_1_category_nameTools & Home Improvement                 7.754e+01  -0.619
## leve_1_category_nameToys & Games                             3.178e+01  -8.174
## leve_1_category_nameUnderwear & Sleepwear                    3.560e+01  -3.061
## leve_1_category_nameWomen's Clothing                         2.988e+01  -4.245
## leve_1_category_nameWomen's Curve + Plus                     4.969e+01  -2.499
## leve_2_category_nameAccessories                              7.130e+01  -1.643
## leve_2_category_nameAdhesives & Sealants                     7.911e+01   4.387
## leve_2_category_nameAir Quality & Irons                      1.023e+02   1.600
## leve_2_category_nameArts & Crafts                            2.905e+01   7.743
## leve_2_category_nameBath                                     3.431e+01   6.233
## leve_2_category_nameBeauty Tools                             2.558e+01  -1.666
## leve_2_category_nameBedding                                  5.573e+01   1.112
## leve_2_category_nameBirds & Poultries                        6.346e+01  -1.755
## leve_2_category_nameBoy's Bottoms                            3.937e+01   0.894
## leve_2_category_nameBoy's Coat                               3.847e+01   2.390
## leve_2_category_nameBoy's Sets                               2.385e+01   1.824
## leve_2_category_nameBoy's Shoes                              2.879e+01   1.679
## leve_2_category_nameBoy's Tops                               2.353e+01   1.436
```

```
## leve_2_category_nameCar Accessories                          3.133e+01  -4.400
## leve_2_category_nameCar Storage & Organizers                 5.880e+01   1.025
## leve_2_category_nameCar Toys                                 6.282e+01   4.380
## leve_2_category_nameCleaning Supplies                        4.902e+01  15.497
## leve_2_category_nameCommercial Lighting                      8.811e+01   1.334
## leve_2_category_nameComputer Accessories                     3.322e+01   1.599
## leve_2_category_nameCurve + Plus Beachwear                   1.486e+02   0.022
## leve_2_category_nameCurve + Plus Bottoms                     6.134e+01   0.284
## leve_2_category_nameCurve + Plus Dresses & Jumpsuits         5.124e+01   0.630
## leve_2_category_nameCurve + Plus Tops & Outerwear            5.013e+01   0.392
## leve_2_category_nameCurve + Plus Two Pieces Set              1.486e+02   0.255
## leve_2_category_nameCurve + Plus Underwear & Sleepwear              NA      NA
## leve_2_category_nameCutting Tools                            5.975e+01   0.289
## leve_2_category_nameCycling                                  4.608e+01  -0.415
## leve_2_category_nameDogs & Cats Supplies                     2.513e+01  -2.896
## leve_2_category_nameDolls & Plush Toys                       2.962e+01   4.469
## leve_2_category_nameEvent & Party Supplies                   3.214e+01   0.813
## leve_2_category_nameFish, Reptiles & Amphibians              7.133e+01  -1.086
## leve_2_category_nameFishing                                  4.397e+01  -0.580
## leve_2_category_nameFoot, Hand & Nail Care                   2.399e+01  -0.889
## leve_2_category_nameGardening & Urban Farming                5.673e+01  -0.008
## leve_2_category_nameGirl's  Shoes                            2.772e+01   2.276
## leve_2_category_nameGirl's Bottoms                           3.273e+01   2.758
## leve_2_category_nameGirl's Coat                              3.896e+01   1.776
## leve_2_category_nameGirl's Dresses                           2.481e+01   3.842
## leve_2_category_nameGirl's Sets                              2.202e+01   4.386
## leve_2_category_nameGirl's Tops                              2.337e+01   1.183
## leve_2_category_nameGreeting Cards & Postcards               3.684e+01  -4.172
## leve_2_category_nameHair Care                                3.488e+01  -1.068
## leve_2_category_nameHardware                                 9.171e+01   2.143
## leve_2_category_nameHeadphones, Earbuds & Accessories        3.262e+01   2.921
## leve_2_category_nameHealth Care Products                     1.374e+02  -1.441
## leve_2_category_nameHome Appliances                          4.460e+01   0.384
## leve_2_category_nameHome Decor Products                      2.986e+01   1.011
## leve_2_category_nameHome Use Medical Supplies & Equipment    1.205e+02  -0.832
## leve_2_category_nameHousehold Supplies                       1.141e+02  -0.188
## leve_2_category_nameIndustrial Materials                     6.112e+01   0.247
## leve_2_category_nameInstruments                              7.120e+01  -0.126
## leve_2_category_nameJewelry Accessories                      4.696e+01   0.714
## leve_2_category_nameKeychains&Key Shells                     4.295e+01  -6.152
## leve_2_category_nameKid's Accessories                        1.913e+01   1.122
## leve_2_category_nameKids' Electronics                        4.876e+01   4.506
## leve_2_category_nameKids' Home Store                         9.986e+01   0.862
## leve_2_category_nameKitchen & Dining                         2.926e+01   5.826
## leve_2_category_nameKitchen, Bath Fixtures & Appliances      9.417e+01   1.546
## leve_2_category_nameLearning & Education                     3.897e+01   4.149
## leve_2_category_nameLighting & Ceiling Fans                  6.620e+01   0.827
## leve_2_category_nameLights & Lighting Accessories            5.644e+01  -1.839
## leve_2_category_nameLuggage & Travel Gear                    2.672e+01   0.613
## leve_2_category_nameMakeup                                   2.484e+01  -0.300
## leve_2_category_nameMaternity Clothing                       1.054e+02   0.253
## leve_2_category_nameMaternity, Baby Gear & Nursery                  NA      NA
## leve_2_category_nameMen's Accessories                        3.582e+01   0.250
## leve_2_category_nameMen's Backbags                           3.789e+01   1.472
```

```
## leve_2_category_nameMen's Belts                                  5.655e+01   0.647
## leve_2_category_nameMen's Boots                                  3.426e+01   0.393
## leve_2_category_nameMen's Bottoms                                3.598e+01   0.350
## leve_2_category_nameMen's Clothing & Accessories                 4.732e+01  -0.712
## leve_2_category_nameMen's Exotic & Novelty  Clothing             4.290e+01  -0.743
## leve_2_category_nameMen's Glasses                                4.073e+01   1.012
## leve_2_category_nameMen's Handbags                               6.899e+01   0.610
## leve_2_category_nameMen's Hats, Caps & Wigs                      3.564e+01  -0.805
## leve_2_category_nameMen's Jackets&Coats                          3.484e+01   1.476
## leve_2_category_nameMen's Jeans                                  4.194e+01   0.902
## leve_2_category_nameMen's Jewelry                                3.194e+01  -0.377
## leve_2_category_nameMen's Loafers & Slip-Ons                     3.661e+01  -1.246
## leve_2_category_nameMen's Loungewear                             4.068e+01   0.322
## leve_2_category_nameMen's Mules&Clogs                            6.591e+01  -0.794
## leve_2_category_nameMen's Sandal & Slippers                      2.601e+01  -1.536
## leve_2_category_nameMen's Scarves & Gloves                       5.893e+01   0.391
## leve_2_category_nameMen's Shirts                                 3.194e+01  -0.574
## leve_2_category_nameMen's Shoes                                  2.129e+01  -0.707
## leve_2_category_nameMen's Shoulder Bags                          3.819e+01   0.364
## leve_2_category_nameMen's Socks & Hosiery                        2.558e+01   0.786
## leve_2_category_nameMen's Sports Shoes                           5.851e+01   1.432
## leve_2_category_nameMen's Suits                                  6.527e+01   0.259
## leve_2_category_nameMen's Sweaters                               4.549e+01   0.463
## leve_2_category_nameMen's Sweatshirts                            3.309e+01  -0.069
## leve_2_category_nameMen's T-Shirts                               3.169e+01  -0.762
## leve_2_category_nameMen's Tops                                          NA      NA
## leve_2_category_nameMen's Underwear                              3.048e+01  -0.092
## leve_2_category_nameMen's Waist & Chest bags                     3.355e+01   0.693
## leve_2_category_nameMen's Wallets & Card Cases                   3.562e+01   0.712
## leve_2_category_nameMen's Watches & Accessories                  3.800e+01   0.432
## leve_2_category_nameMen's Work & Safety Shoes                    7.070e+01  -0.095
## leve_2_category_nameMotorcycles                                  4.341e+01  -3.731
## leve_2_category_nameNovelty & Costumes                           3.765e+01   4.290
## leve_2_category_nameOccupational Health & Safety Products        7.153e+01   0.499
## leve_2_category_nameOffice & School Supplies                     2.606e+01  -4.331
## leve_2_category_nameOffice Electronics                           1.058e+02  -0.312
## leve_2_category_nameOrganization                                 3.115e+01  -2.344
## leve_2_category_nameOther Animals                                6.574e+01  -1.645
## leve_2_category_nameOther Industrial Supplies                    7.239e+01   0.258
## leve_2_category_nameOther Toys & Games                           2.064e+01   8.729
## leve_2_category_nameOutdoor Lighting & Power Tools               5.483e+01  -0.986
## leve_2_category_nameOutdoor Lights                               5.747e+01  -0.221
## leve_2_category_nameOutdoor Recreation                           4.463e+01  -0.377
## leve_2_category_nameOutdoor Supplies & Patio Furniture                  NA      NA
## leve_2_category_namePaint, Wall Treatments & Supplies            7.563e+01  -1.038
## leve_2_category_namePapers, Labels & Indexes                     2.951e+01  -2.559
## leve_2_category_nameParty Supplies                               2.734e+01   4.786
## leve_2_category_namePersonal Care                                3.031e+01   0.080
## leve_2_category_namePersonal Care Electronics                    5.687e+01  -0.097
## leve_2_category_namePersonal Care Products                       1.150e+02   1.640
## leve_2_category_namePet Apparel & Accessories                    2.530e+01  -2.935
## leve_2_category_namePet Cleaning & Grooming                             NA      NA
## leve_2_category_namePhone Accessories                            3.530e+01   0.272
## leve_2_category_namePhone Cables & Chargers                      3.810e+01   4.576
```

```
## leve_2_category_namePhone Cases & Screen Protector          2.937e+01   0.845
## leve_2_category_namePower & Hand Tools                      5.660e+01   0.450
## leve_2_category_namePower Banks & Batteries                 6.084e+01   2.593
## leve_2_category_namePower Tools & Safety                    8.080e+01  -0.140
## leve_2_category_namePretend Play                            5.796e+01   2.904
## leve_2_category_namePuzzles & Building Toys                 2.508e+01   8.853
## leve_2_category_nameQuadcopters & UAV                       8.049e+01   3.092
## leve_2_category_nameReplacement Parts                       1.051e+02  -1.218
## leve_2_category_nameSeasonal Decor                          3.101e+01   1.389
## leve_2_category_nameShoe Accessories                        4.061e+01   0.013
## leve_2_category_nameSports & Fitness Supplies               4.468e+01   0.196
## leve_2_category_nameSports Bags                             5.281e+01  -0.308
## leve_2_category_nameStationery & Gift Wrapping Supplies     1.041e+02  -2.186
## leve_2_category_nameStickers & Crafts Tape                         NA      NA
## leve_2_category_nameStorage & Organization                  3.136e+01  10.091
## leve_2_category_nameStudio Recording & Stage Live                  NA      NA
## leve_2_category_nameTest, Measure & Inspect                 5.982e+01   0.482
## leve_2_category_nameTools & Equipment                              NA      NA
## leve_2_category_nameWall Art                                       NA      NA
## leve_2_category_nameWearable Technology                            NA      NA
## leve_2_category_nameWellness & Relaxation Products                 NA      NA
## leve_2_category_nameWigs & Accessories                             NA      NA
## leve_2_category_nameWinter Sports                           7.848e+01  -0.310
## leve_2_category_nameWomen's Accessories                     3.344e+01   1.779
## leve_2_category_nameWomen's Activewear                      4.458e+01  -0.340
## leve_2_category_nameWomen's Athleisure                      3.068e+01   0.731
## leve_2_category_nameWomen's Backbags                        3.825e+01   0.895
## leve_2_category_nameWomen's Beachwear                       2.306e+01  -0.156
## leve_2_category_nameWomen's Blazer                          6.453e+01   0.533
## leve_2_category_nameWomen's Bodysuits                       7.874e+01  -0.087
## leve_2_category_nameWomen's Boots                           2.895e+01  -0.115
## leve_2_category_nameWomen's Clutches & Evening Bags         6.546e+01  -0.230
## leve_2_category_nameWomen's Coat & Jacket                   2.330e+01   2.793
## leve_2_category_nameWomen's Cosplay Costume                 1.241e+02   0.281
## leve_2_category_nameWomen's Crossbody Bags                  2.640e+01   0.089
## leve_2_category_nameWomen's Denims                          1.873e+01   2.408
## leve_2_category_nameWomen's Dresses                         1.236e+01   0.985
## leve_2_category_nameWomen's Fashion Sneakers                2.539e+01   0.374
## leve_2_category_nameWomen's Flats                           3.824e+01   0.113
## leve_2_category_nameWomen's Glasses                         5.013e+01   0.605
## leve_2_category_nameWomen's Handbags                        3.405e+01   0.499
## leve_2_category_nameWomen's Hats & Caps                     3.608e+01  -0.511
## leve_2_category_nameWomen's Jewelry                         3.158e+01  -0.191
## leve_2_category_nameWomen's Jumpsuits                       3.956e+01   0.128
## leve_2_category_nameWomen's Lingerie                        2.599e+01   0.195
## leve_2_category_nameWomen's Pants                           2.629e+01   0.277
## leve_2_category_nameWomen's Pumps                           5.377e+01  -0.504
## leve_2_category_nameWomen's Sandals                         4.309e+01  -0.132
## leve_2_category_nameWomen's Sexy Lingerie                   2.647e+01  -0.866
## leve_2_category_nameWomen's Shapewear & Others              4.899e+01   0.101
## leve_2_category_nameWomen's Shorts                          3.716e+01   0.017
## leve_2_category_nameWomen's Shoulder Bags                   2.840e+01   0.151
## leve_2_category_nameWomen's Skirts                          4.697e+01   0.097
## leve_2_category_nameWomen's Sleepwear                              NA      NA
```

```
## leve_2_category_nameWomen's Slippers                              NA      NA
## leve_2_category_nameWomen's Sports Shoes                       5.504e+01   0.819
## leve_2_category_nameWomen's Stockings & Hosiery                1.588e+01   1.696
## leve_2_category_nameWomen's Suits                              3.636e+01   0.198
## leve_2_category_nameWomen's Sweaters                           1.418e+01   1.516
## leve_2_category_nameWomen's Sweatshirts                        1.783e+01   0.887
## leve_2_category_nameWomen's Tops                                      NA      NA
## leve_2_category_nameWomen's Tote Bags                          2.363e+01   0.041
## leve_2_category_nameWomen's Waist & Chest Bags                 4.501e+01   0.154
## leve_2_category_nameWomen's Wallets & Card Cases                      NA      NA
## leve_2_category_nameWomen's Watches                                   NA      NA
## leve_2_category_nameWriting Supplies & Correction Supplies            NA      NA
## leve_2_category_nameYoga&Studio                                       NA      NA
## comment_num                                                    5.163e-03  51.958
##                                                               Pr(>|t|)
## (Intercept)                                                   2.71e-08 ***
## price                                                          < 2e-16 ***
## high_score                                                    0.242775
## leve_1_category_nameBags & Luggages                           4.63e-05 ***
## leve_1_category_nameBeauty & Health                           0.014553 *
## leve_1_category_nameElectronics & Appliances                  0.008163 **
## leve_1_category_nameHealth & Household                        0.918536
## leve_1_category_nameHome & Kitchen                            0.000131 ***
## leve_1_category_nameIndustrial & Commercial                   0.021722 *
## leve_1_category_nameJewelry & Accessories                     0.000403 ***
## leve_1_category_nameKid, Baby & Maternity                     6.72e-06 ***
## leve_1_category_nameMen's Clothing                            0.011683 *
## leve_1_category_nameMusical Instruments                       0.725107
## leve_1_category_nameOffice & School                           0.171998
## leve_1_category_namePatio, Lawn & Garden                      0.461391
## leve_1_category_namePet Supplies                              0.064818 .
## leve_1_category_nameShoes                                     0.071184 .
## leve_1_category_nameSports & Outdoors                         0.053857 .
## leve_1_category_nameTools & Home Improvement                  0.535734
## leve_1_category_nameToys & Games                              3.03e-16 ***
## leve_1_category_nameUnderwear & Sleepwear                     0.002204 **
## leve_1_category_nameWomen's Clothing                          2.19e-05 ***
## leve_1_category_nameWomen's Curve + Plus                      0.012439 *
## leve_2_category_nameAccessories                               0.100313
## leve_2_category_nameAdhesives & Sealants                      1.15e-05 ***
## leve_2_category_nameAir Quality & Irons                       0.109703
## leve_2_category_nameArts & Crafts                             9.86e-15 ***
## leve_2_category_nameBath                                      4.58e-10 ***
## leve_2_category_nameBeauty Tools                              0.095695 .
## leve_2_category_nameBedding                                   0.266126
## leve_2_category_nameBirds & Poultries                         0.079336 .
## leve_2_category_nameBoy's Bottoms                             0.371150
## leve_2_category_nameBoy's Coat                                0.016858 *
## leve_2_category_nameBoy's Sets                                0.068162 .
## leve_2_category_nameBoy's Shoes                               0.093208 .
## leve_2_category_nameBoy's Tops                                0.151053
## leve_2_category_nameCar Accessories                           1.08e-05 ***
## leve_2_category_nameCar Storage & Organizers                  0.305215
## leve_2_category_nameCar Toys                                  1.19e-05 ***
```

```
## leve_2_category_nameCleaning Supplies                        < 2e-16 ***
## leve_2_category_nameCommercial Lighting                     0.182246
## leve_2_category_nameComputer Accessories                    0.109739
## leve_2_category_nameCurve + Plus Beachwear                  0.982640
## leve_2_category_nameCurve + Plus Bottoms                    0.776366
## leve_2_category_nameCurve + Plus Dresses & Jumpsuits        0.528459
## leve_2_category_nameCurve + Plus Tops & Outerwear           0.695228
## leve_2_category_nameCurve + Plus Two Pieces Set             0.798493
## leve_2_category_nameCurve + Plus Underwear & Sleepwear            NA
## leve_2_category_nameCutting Tools                           0.772279
## leve_2_category_nameCycling                                 0.678376
## leve_2_category_nameDogs & Cats Supplies                    0.003783 **
## leve_2_category_nameDolls & Plush Toys                      7.87e-06 ***
## leve_2_category_nameEvent & Party Supplies                  0.416003
## leve_2_category_nameFish, Reptiles & Amphibians             0.277372
## leve_2_category_nameFishing                                 0.562176
## leve_2_category_nameFoot, Hand & Nail Care                  0.373949
## leve_2_category_nameGardening & Urban Farming               0.993314
## leve_2_category_nameGirl's  Shoes                           0.022863 *
## leve_2_category_nameGirl's Bottoms                          0.005824 **
## leve_2_category_nameGirl's Coat                             0.075769 .
## leve_2_category_nameGirl's Dresses                          0.000122 ***
## leve_2_category_nameGirl's Sets                             1.16e-05 ***
## leve_2_category_nameGirl's Tops                             0.236624
## leve_2_category_nameGreeting Cards & Postcards              3.02e-05 ***
## leve_2_category_nameHair Care                               0.285713
## leve_2_category_nameHardware                                0.032102 *
## leve_2_category_nameHeadphones, Earbuds & Accessories       0.003492 **
## leve_2_category_nameHealth Care Products                    0.149473
## leve_2_category_nameHome Appliances                         0.700798
## leve_2_category_nameHome Decor Products                     0.312104
## leve_2_category_nameHome Use Medical Supplies & Equipment   0.405340
## leve_2_category_nameHousehold Supplies                      0.851084
## leve_2_category_nameIndustrial Materials                    0.804681
## leve_2_category_nameInstruments                             0.899512
## leve_2_category_nameJewelry Accessories                     0.475192
## leve_2_category_nameKeychains&Key Shells                    7.70e-10 ***
## leve_2_category_nameKid's Accessories                       0.261883
## leve_2_category_nameKids' Electronics                       6.61e-06 ***
## leve_2_category_nameKids' Home Store                        0.388518
## leve_2_category_nameKitchen & Dining                        5.71e-09 ***
## leve_2_category_nameKitchen, Bath Fixtures & Appliances     0.122147
## leve_2_category_nameLearning & Education                    3.35e-05 ***
## leve_2_category_nameLighting & Ceiling Fans                 0.408043
## leve_2_category_nameLights & Lighting Accessories           0.065939 .
## leve_2_category_nameLuggage & Travel Gear                   0.539759
## leve_2_category_nameMakeup                                  0.763815
## leve_2_category_nameMaternity Clothing                      0.800320
## leve_2_category_nameMaternity, Baby Gear & Nursery                NA
## leve_2_category_nameMen's Accessories                       0.802442
## leve_2_category_nameMen's Backbags                          0.140913
## leve_2_category_nameMen's Belts                             0.517702
## leve_2_category_nameMen's Boots                             0.693981
## leve_2_category_nameMen's Bottoms                           0.726180
```

```
## leve_2_category_nameMen's Clothing & Accessories          0.476450
## leve_2_category_nameMen's Exotic & Novelty  Clothing      0.457261
## leve_2_category_nameMen's Glasses                         0.311305
## leve_2_category_nameMen's Handbags                        0.541760
## leve_2_category_nameMen's Hats, Caps & Wigs               0.420665
## leve_2_category_nameMen's Jackets&Coats                   0.139931
## leve_2_category_nameMen's Jeans                           0.367191
## leve_2_category_nameMen's Jewelry                         0.705813
## leve_2_category_nameMen's Loafers & Slip-Ons              0.212781
## leve_2_category_nameMen's Loungewear                      0.747781
## leve_2_category_nameMen's Mules&Clogs                     0.427096
## leve_2_category_nameMen's Sandal & Slippers               0.124523
## leve_2_category_nameMen's Scarves & Gloves                0.695927
## leve_2_category_nameMen's Shirts                          0.565929
## leve_2_category_nameMen's Shoes                           0.479684
## leve_2_category_nameMen's Shoulder Bags                   0.715915
## leve_2_category_nameMen's Socks & Hosiery                 0.432030
## leve_2_category_nameMen's Sports Shoes                    0.152120
## leve_2_category_nameMen's Suits                           0.795446
## leve_2_category_nameMen's Sweaters                        0.643589
## leve_2_category_nameMen's Sweatshirts                     0.945155
## leve_2_category_nameMen's T-Shirts                        0.446145
## leve_2_category_nameMen's Tops                                  NA
## leve_2_category_nameMen's Underwear                       0.926322
## leve_2_category_nameMen's Waist & Chest bags              0.488280
## leve_2_category_nameMen's Wallets & Card Cases            0.476410
## leve_2_category_nameMen's Watches & Accessories           0.665711
## leve_2_category_nameMen's Work & Safety Shoes             0.924142
## leve_2_category_nameMotorcycles                           0.000191 ***
## leve_2_category_nameNovelty & Costumes                    1.79e-05 ***
## leve_2_category_nameOccupational Health & Safety Products 0.617942
## leve_2_category_nameOffice & School Supplies              1.49e-05 ***
## leve_2_category_nameOffice Electronics                    0.754893
## leve_2_category_nameOrganization                          0.019072 *
## leve_2_category_nameOther Animals                         0.099999 .
## leve_2_category_nameOther Industrial Supplies             0.796661
## leve_2_category_nameOther Toys & Games                     < 2e-16 ***
## leve_2_category_nameOutdoor Lighting & Power Tools        0.324213
## leve_2_category_nameOutdoor Lights                        0.824891
## leve_2_category_nameOutdoor Recreation                    0.705820
## leve_2_category_nameOutdoor Supplies & Patio Furniture          NA
## leve_2_category_namePaint, Wall Treatments & Supplies     0.299242
## leve_2_category_namePapers, Labels & Indexes              0.010500 *
## leve_2_category_nameParty Supplies                        1.71e-06 ***
## leve_2_category_namePersonal Care                         0.936432
## leve_2_category_namePersonal Care Electronics             0.922394
## leve_2_category_namePersonal Care Products                0.101029
## leve_2_category_namePet Apparel & Accessories             0.003335 **
## leve_2_category_namePet Cleaning & Grooming                     NA
## leve_2_category_namePhone Accessories                     0.785470
## leve_2_category_namePhone Cables & Chargers               4.74e-06 ***
## leve_2_category_namePhone Cases & Screen Protector        0.398140
## leve_2_category_namePower & Hand Tools                    0.652613
## leve_2_category_namePower Banks & Batteries               0.009521 **
```

```
## leve_2_category_namePower Tools & Safety                    0.888405
## leve_2_category_namePretend Play                            0.003682 **
## leve_2_category_namePuzzles & Building Toys                  < 2e-16 ***
## leve_2_category_nameQuadcopters & UAV                       0.001991 **
## leve_2_category_nameReplacement Parts                       0.223309
## leve_2_category_nameSeasonal Decor                          0.164730
## leve_2_category_nameShoe Accessories                        0.989776
## leve_2_category_nameSports & Fitness Supplies               0.844497
## leve_2_category_nameSports Bags                             0.757960
## leve_2_category_nameStationery & Gift Wrapping Supplies     0.028810 *
## leve_2_category_nameStickers & Crafts Tape                        NA
## leve_2_category_nameStorage & Organization                   < 2e-16 ***
## leve_2_category_nameStudio Recording & Stage Live                 NA
## leve_2_category_nameTest, Measure & Inspect                 0.630095
## leve_2_category_nameTools & Equipment                             NA
## leve_2_category_nameWall Art                                      NA
## leve_2_category_nameWearable Technology                           NA
## leve_2_category_nameWellness & Relaxation Products                NA
## leve_2_category_nameWigs & Accessories                            NA
## leve_2_category_nameWinter Sports                           0.756865
## leve_2_category_nameWomen's Accessories                     0.075273 .
## leve_2_category_nameWomen's Activewear                      0.733684
## leve_2_category_nameWomen's Athleisure                      0.464818
## leve_2_category_nameWomen's Backbags                        0.370996
## leve_2_category_nameWomen's Beachwear                       0.875851
## leve_2_category_nameWomen's Blazer                          0.593985
## leve_2_category_nameWomen's Bodysuits                       0.930379
## leve_2_category_nameWomen's Boots                           0.908170
## leve_2_category_nameWomen's Clutches & Evening Bags         0.817731
## leve_2_category_nameWomen's Coat & Jacket                   0.005217 **
## leve_2_category_nameWomen's Cosplay Costume                 0.778621
## leve_2_category_nameWomen's Crossbody Bags                  0.929096
## leve_2_category_nameWomen's Denims                          0.016043 *
## leve_2_category_nameWomen's Dresses                         0.324728
## leve_2_category_nameWomen's Fashion Sneakers                0.708040
## leve_2_category_nameWomen's Flats                           0.909859
## leve_2_category_nameWomen's Glasses                         0.545258
## leve_2_category_nameWomen's Handbags                        0.618117
## leve_2_category_nameWomen's Hats & Caps                     0.609232
## leve_2_category_nameWomen's Jewelry                         0.848437
## leve_2_category_nameWomen's Jumpsuits                       0.898307
## leve_2_category_nameWomen's Lingerie                        0.845502
## leve_2_category_nameWomen's Pants                           0.781751
## leve_2_category_nameWomen's Pumps                           0.614071
## leve_2_category_nameWomen's Sandals                         0.895173
## leve_2_category_nameWomen's Sexy Lingerie                   0.386254
## leve_2_category_nameWomen's Shapewear & Others              0.919677
## leve_2_category_nameWomen's Shorts                          0.986810
## leve_2_category_nameWomen's Shoulder Bags                   0.880195
## leve_2_category_nameWomen's Skirts                          0.923026
## leve_2_category_nameWomen's Sleepwear                             NA
## leve_2_category_nameWomen's Slippers                              NA
## leve_2_category_nameWomen's Sports Shoes                    0.412898
## leve_2_category_nameWomen's Stockings & Hosiery             0.089899 .
```

```
## leve_2_category_nameWomen's Suits                              0.843061
## leve_2_category_nameWomen's Sweaters                           0.129450
## leve_2_category_nameWomen's Sweatshirts                        0.375267
## leve_2_category_nameWomen's Tops                                     NA
## leve_2_category_nameWomen's Tote Bags                          0.967582
## leve_2_category_nameWomen's Waist & Chest Bags                 0.877863
## leve_2_category_nameWomen's Wallets & Card Cases                     NA
## leve_2_category_nameWomen's Watches                                  NA
## leve_2_category_nameWriting Supplies & Correction Supplies           NA
## leve_2_category_nameYoga&Studio                                      NA
## comment_num                                                  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 428.9 on 79823 degrees of freedom
## Multiple R-squared:  0.06694,    Adjusted R-squared:  0.06469
## F-statistic: 29.83 on 192 and 79823 DF,  p-value: < 2.2e-16
```

Residual Analysis.

The smallest residual, -1623.4, indicates that the model's prediction of some data is highly biased. The wide range of residuals (-1623.4 to 25724.7) indicates that outliers of high volume items have a significant impact on the model fit. Interpretation of variables.

PRICE: Price significantly and negatively affects sales volume, with sales volume decreasing for each unit increase in price 25.43 units ($p < 2e\text{-}16$).

comment_num: number of comments has a significant positive effect on sales ($p < 2e\text{-}16$). Some categories (e.g. Men's Shoes and Storage & Organization) significantly and positively affect sales, but others (e.g. Girl's Sets and Kitchen & Dining) do not.

3.1.2

```r
model_lm1 <- lm(sales_volume ~ price + high_score + leve_2_category_name + comment_num , data = top5_da

# summary output
summary(model_lm1)
```

```
##
## Call:
## lm(formula = sales_volume ~ price + high_score + leve_2_category_name +
##     comment_num, data = top5_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1623.4  -182.3   -75.2    22.7 25724.7
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              284.91034   57.63167   4.944
## price                                    -25.42712    3.00379  -8.465
## high_score                                 3.60683   40.13908   0.090
## leve_2_category_nameGirl's Sets           -3.91343   46.98238  -0.083
## leve_2_category_nameKitchen & Dining      -72.63455   44.29808  -1.640
## leve_2_category_nameMen's Shoes           130.60856   51.30099   2.546
```

```
## leve_2_category_nameStorage & Organization 113.18993   49.32414   2.295
## comment_num                                   0.37392    0.02012  18.585
##                                                        Pr(>|t|)
## (Intercept)                                            7.88e-07 ***
## price                                                   < 2e-16 ***
## high_score                                               0.9284
## leve_2_category_nameGirl's Sets                          0.9336
## leve_2_category_nameKitchen & Dining                     0.1011
## leve_2_category_nameMen's Shoes                          0.0109 *
## leve_2_category_nameStorage & Organization               0.0218 *
## comment_num                                             < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 893.9 on 5767 degrees of freedom
## Multiple R-squared:  0.09732,    Adjusted R-squared:  0.09623
## F-statistic: 88.82 on 7 and 5767 DF,  p-value: < 2.2e-16
```

3.1.3 With high_score and comment num remove to compare to previous model

The r^2 explains less, indicating that high_score and comment num do influence users' purchase options

```
model_lm2 <- lm(sales_volume ~ price +  leve_2_category_name  , data = top5_data)

summary(model_lm2)
```

```
##
## Call:
## lm(formula = sales_volume ~ price + leve_2_category_name, data = top5_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -559.4  -276.7  -125.4     3.8 25935.0
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                 429.809     46.932   9.158  < 2e-16
## price                                       -36.379      3.026 -12.021  < 2e-16
## leve_2_category_nameGirl's Sets             -15.179     48.116  -0.315 0.752420
## leve_2_category_nameKitchen & Dining         15.424     45.274   0.341 0.733362
## leve_2_category_nameMen's Shoes             199.091     52.687   3.779 0.000159
## leve_2_category_nameStorage & Organization  157.425     50.519   3.116 0.001841
##
## (Intercept)                                ***
## price                                      ***
## leve_2_category_nameGirl's Sets
## leve_2_category_nameKitchen & Dining
## leve_2_category_nameMen's Shoes            ***
## leve_2_category_nameStorage & Organization **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 920.4 on 5769 degrees of freedom
## Multiple R-squared:  0.04267,    Adjusted R-squared:  0.04184
```

```
## F-statistic: 51.42 on 5 and 5769 DF,  p-value: < 2.2e-16
```

In the data, high volume items (outliers) have a significant impact on the model fit

Improve the distribution of the data to be closer to a normal distribution High-volume data tends to be right-skewed (positively skewed distribution), i.e., most items have low volume, but a few items have extremely high volume. A logarithmic transformation will flatten the right-skewed data and bring its distribution closer to normal. This is a basic assumption of linear regression: the residuals should be close to a normal distribution.

3.1.4

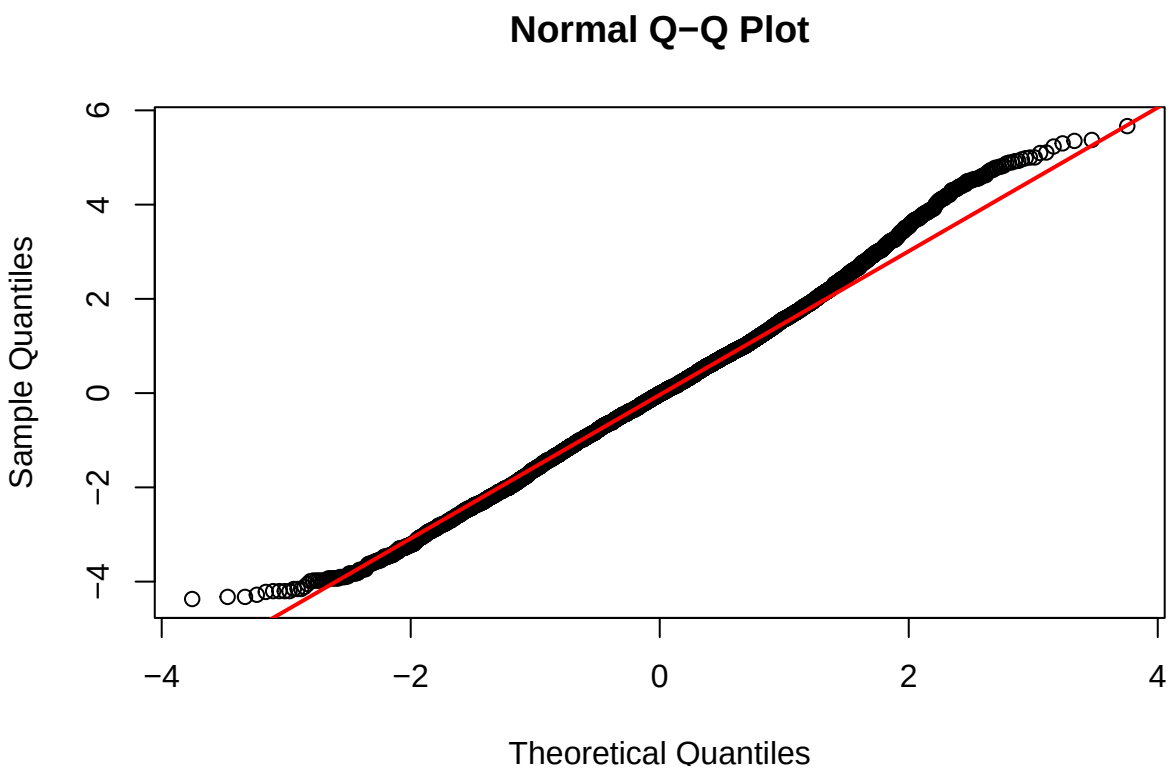
```
model_lm3 <- lm(log(sales_volume) ~ price +  leve_2_category_name  , data = top5_data)


summary(model_lm3)
```

```
##
## Call:
## lm(formula = log(sales_volume) ~ price + leve_2_category_name,
##     data = top5_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3694 -1.0655 -0.0194  0.9906  5.6655
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              4.801896   0.082746  58.032
## price                                   -0.150369   0.005335 -28.183
## leve_2_category_nameGirl's Sets         -0.128681   0.084833  -1.517
## leve_2_category_nameKitchen & Dining    -0.390842   0.079822  -4.896
## leve_2_category_nameMen's Shoes          0.002484   0.092891   0.027
## leve_2_category_nameStorage & Organization -0.223525   0.089070  -2.510
##                                         Pr(>|t|)
## (Intercept)                               <2e-16 ***
## price                                     <2e-16 ***
## leve_2_category_nameGirl's Sets           0.1294
## leve_2_category_nameKitchen & Dining       1e-06 ***
## leve_2_category_nameMen's Shoes           0.9787
## leve_2_category_nameStorage & Organization   0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.623 on 5769 degrees of freedom
## Multiple R-squared:  0.1912, Adjusted R-squared:  0.1905
## F-statistic: 272.7 on 5 and 5769 DF,  p-value: < 2.2e-16
```

3.1.5 qqplot

The QQ plot of the residuals shows that the distribution of the residuals of the original model is skewed away from normal, while the logarithmic transformation brings the residuals closer to the normal distribution.

```r
# QQ Plot
model_residuals <- residuals(model_lm3)
qqnorm(model_residuals)
qqline(model_residuals, col = "red", lwd = 2)
```



The model effectively captures the impact of price, category, ratings and number of reviews on item sales and reveals the following trends:

Price is a significant negative driver of sales volume. There is a significant difference in sales volume performance across merchandise categories. The number of reviews plays a limited role, while the direction of ratings' impact on sales deserves further exploration. However, the applicability of the Poisson regression model is limited by the excessive discretization of the sales data, and it is recommended that the model be adjusted to better cope with the characteristics of the data.

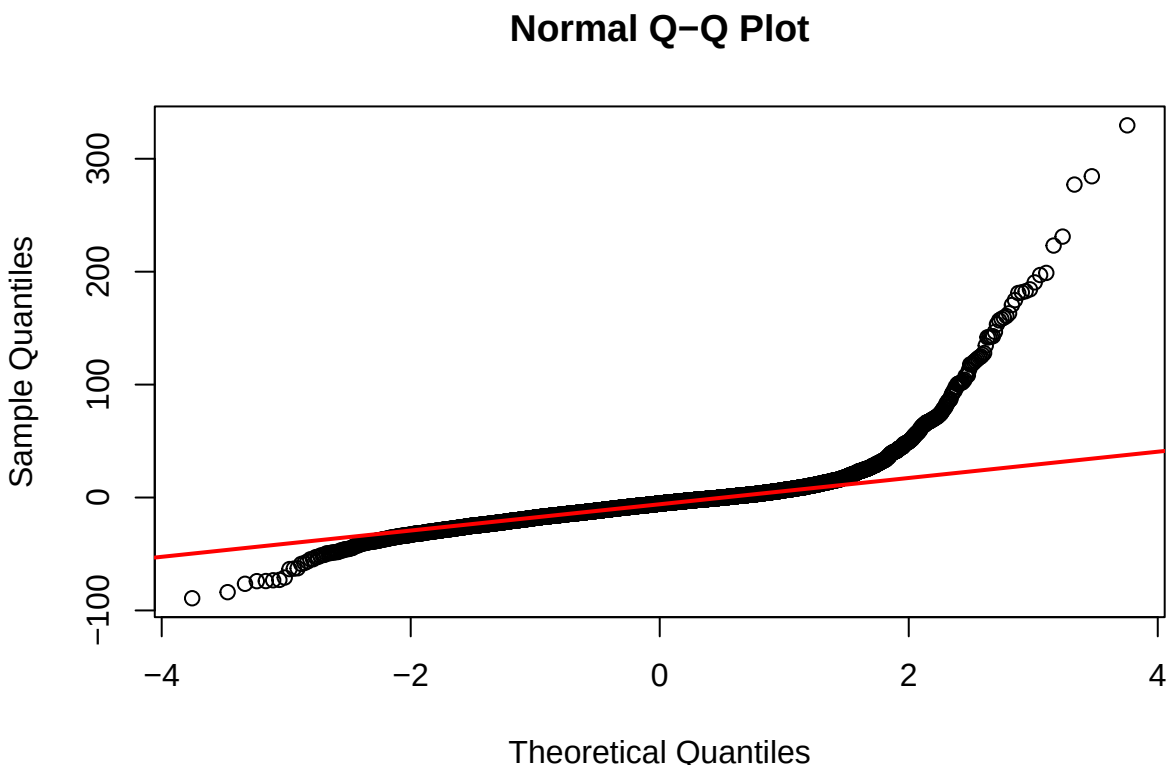
3.2 poisson

```r
# GLM Model: Poisson Regression Predicts Sales Volume
model_glm <- glm(sales_volume ~ price + leve_2_category_name + goods_score + comment_num , data = top5_

summary(model_glm)
```

```
##
## Call:
## glm(formula = sales_volume ~ price + leve_2_category_name + goods_score +
##     comment_num, family = poisson(), data = top5_data)
##
## Coefficients:
##                                        Estimate Std. Error z value
## (Intercept)                           9.578e+00  2.467e-02  388.18
```

```
## price                                       -4.269e-01  6.141e-04 -695.09
## leve_2_category_nameGirl's Sets              1.954e-01  5.216e-03   37.45
## leve_2_category_nameKitchen & Dining        -1.376e+00  5.149e-03 -267.23
## leve_2_category_nameMen's Shoes              1.447e+00  6.637e-03  218.03
## leve_2_category_nameStorage & Organization -8.971e-01  5.274e-03 -170.11
## goods_score                                 -3.962e-01  5.071e-03  -78.13
## comment_num                                  5.099e-04  7.930e-07  643.03
##                                            Pr(>|z|)
## (Intercept)                                  <2e-16 ***
## price                                        <2e-16 ***
## leve_2_category_nameGirl's Sets              <2e-16 ***
## leve_2_category_nameKitchen & Dining         <2e-16 ***
## leve_2_category_nameMen's Shoes              <2e-16 ***
## leve_2_category_nameStorage & Organization   <2e-16 ***
## goods_score                                  <2e-16 ***
## comment_num                                  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4878944  on 5774  degrees of freedom
## Residual deviance: 2935780  on 5767  degrees of freedom
## AIC: 2966492
##
## Number of Fisher Scoring iterations: 7
```

```r
model_residuals1 <- residuals(model_glm)
qqnorm(model_residuals1)
qqline(model_residuals1, col = "red", lwd = 2)
```



3.3 gamma

The glm.nb model has certain requirements for the value of the target variable (sales_volume). If sales_volume contains zero, negative or extreme outliers, the model may not converge.

However, extreme values are the main focus of our study, so we cannot use nb and will try to use a gamma model.

Gamma models have smaller residual variance, smoother fit and less sensitivity to outliers. Looking at the ratio of Residual deviance to Null deviance, the Gamma model seems to have more explanatory power. AIC Comparison.

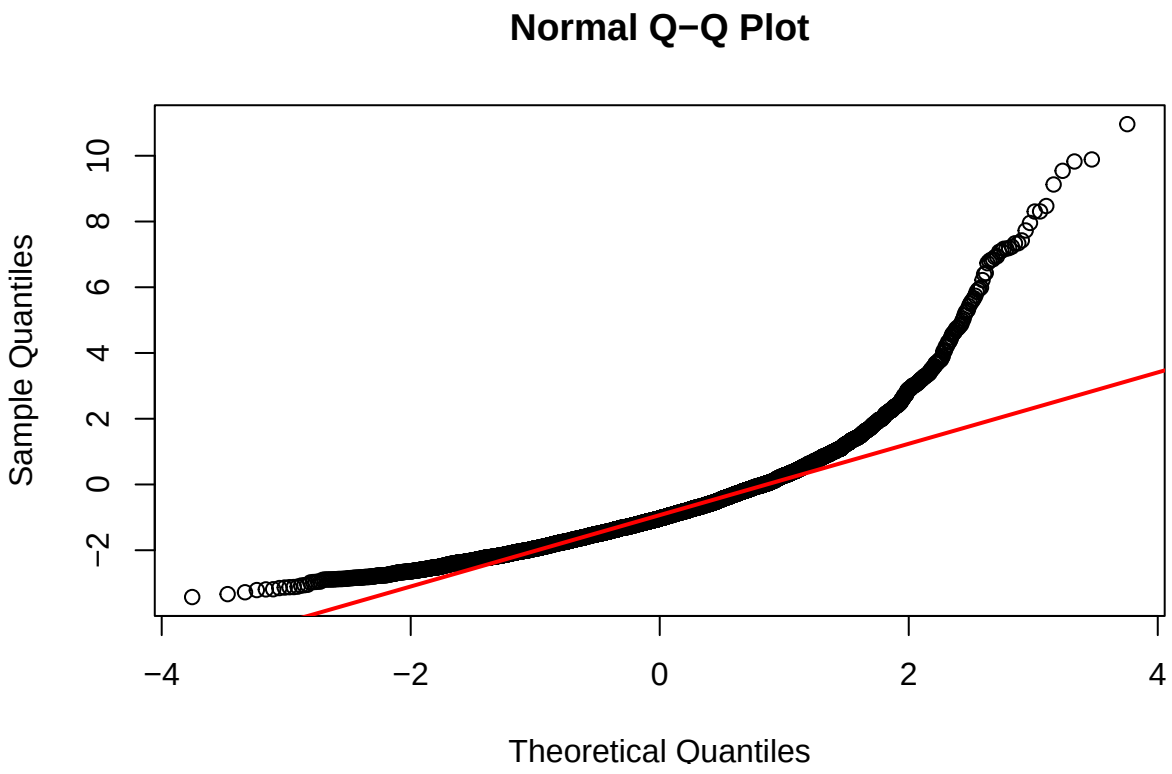
The AIC value of the Gamma model (63863) is higher than that of the Poisson model, but this may be due to a better match between the model assumptions and data characteristics. Lower AIC values indicate a better model fit, but AIC comparisons require that the model assumptions are identical; Poisson and Gamma are different distributions, so comparing AICs alone may not be sufficient to draw conclusions.

```
model_gamma <- glm(sales_volume ~ price + leve_2_category_name + goods_score + comment_num,
                   data = top5_data, family = Gamma(link = "log"))
summary(model_gamma)
```

```
##
## Call:
## glm(formula = sales_volume ~ price + leve_2_category_name + goods_score +
##     comment_num, family = Gamma(link = "log"), data = top5_data)
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             6.211e+00  7.265e-01   8.549
## price                                  -1.445e-01  1.023e-02 -14.132
## leve_2_category_nameGirl's Sets        -1.317e-01  1.596e-01  -0.825
```

```
## leve_2_category_nameKitchen & Dining       -3.093e-01  1.509e-01  -2.050
## leve_2_category_nameMen's Shoes            -5.374e-02  1.759e-01  -0.306
## leve_2_category_nameStorage & Organization -9.015e-02  1.676e-01  -0.538
## goods_score                                -1.076e-01  1.488e-01  -0.723
## comment_num                                 9.160e-04  6.854e-05  13.364
##                                            Pr(>|t|)
## (Intercept)                                  <2e-16 ***
## price                                        <2e-16 ***
## leve_2_category_nameGirl's Sets              0.4092
## leve_2_category_nameKitchen & Dining         0.0405 *
## leve_2_category_nameMen's Shoes              0.7599
## leve_2_category_nameStorage & Organization   0.5907
## goods_score                                  0.4697
## comment_num                                  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 9.305504)
##
##     Null deviance: 22251  on 5774  degrees of freedom
## Residual deviance: 14563  on 5767  degrees of freedom
## AIC: 63863
##
## Number of Fisher Scoring iterations: 17
```

```r
model_residuals2 <- residuals(model_gamma)
qqnorm(model_residuals2)
qqline(model_residuals2, col = "red", lwd = 2)
```



3.4 quasi poisson

Why choose the Quasi-Poisson model? The Poisson model assumes that the mean and variance of the data are equal (mean square equals variance). However, in practice, when the data exhibit **overdispersion**, the Poisson model underestimates the variance, leading to incorrect standard errors and hypothesis testing results. The Quasi-Poisson model relaxes this assumption by introducing a **Dispersion Parameter** that allows the model to accommodate variance greater than the mean.

The Dispersion Factor is 3480.231, which indicates that there is indeed a strong over-dispersion in the data. The Quasi-Poisson model is better able to cope with this problem, whereas the Poisson model may underestimate the variance, thus affecting the results of the significance test.

The Quasi-Poisson model is better at coping with overdispersion than the previous Poisson model, especially when dealing with the effect of high sales volume items on the overall data distribution. Based on the available data and analytical goals, Quasi-Poisson is a reasonable choice, but in the future it can be combined with outlier analysis or more sophisticated models to further optimize predictive performance.

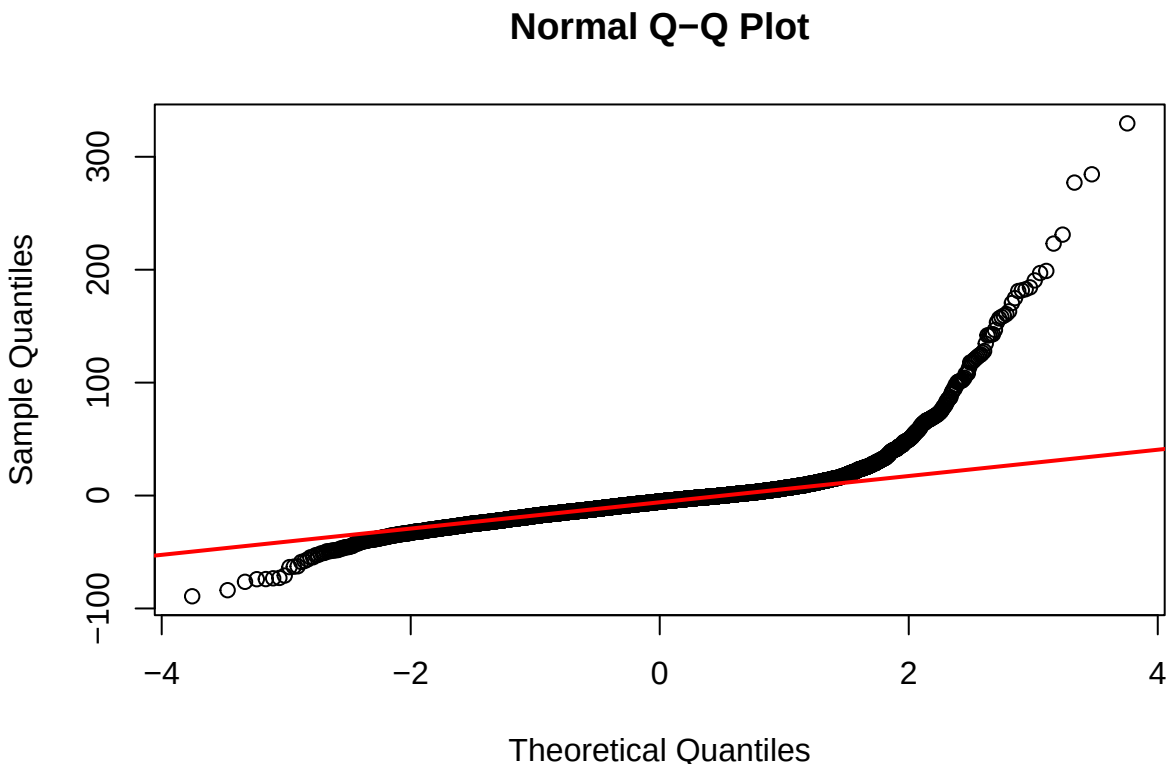
Quasi-Poisson is suitable for count data analysis, but is less capable of handling outliers and may not be the best choice for current data

```r
# Fitting the Quasi-Poisson Model
model_quasi <- glm(sales_volume ~ price + leve_2_category_name + goods_score + comment_num ,
                   family = quasipoisson(),
                   data = top5_data)

summary(model_quasi)
```

```
##
## Call:
## glm(formula = sales_volume ~ price + leve_2_category_name + goods_score +
##     comment_num, family = quasipoisson(), data = top5_data)
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                              9.578e+00  1.456e+00   6.580
## price                                   -4.269e-01  3.623e-02 -11.782
## leve_2_category_nameGirl's Sets          1.954e-01  3.077e-01   0.635
## leve_2_category_nameKitchen & Dining    -1.376e+00  3.038e-01  -4.530
## leve_2_category_nameMen's Shoes          1.447e+00  3.915e-01   3.696
## leve_2_category_nameStorage & Organization -8.971e-01  3.111e-01  -2.884
## goods_score                             -3.962e-01  2.992e-01  -1.324
## comment_num                              5.099e-04  4.678e-05  10.900
##                                         Pr(>|t|)
## (Intercept)                             5.12e-11 ***
## price                                    < 2e-16 ***
## leve_2_category_nameGirl's Sets         0.525533
## leve_2_category_nameKitchen & Dining    6.02e-06 ***
## leve_2_category_nameMen's Shoes         0.000221 ***
## leve_2_category_nameStorage & Organization 0.003946 **
## goods_score                             0.185436
## comment_num                              < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 3480.231)
##
##     Null deviance: 4878944  on 5774  degrees of freedom
```

```
## Residual deviance: 2935780  on 5767  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 7
```

```
model_residuals3 <- residuals(model_quasi)
qqnorm(model_residuals3)
qqline(model_residuals3, col = "red", lwd = 2)
```



3.5 linear mixed effecet

The inclusion of level_2_category_name as a random effect is justified by accounting for unobserved heterogeneity between categories

Even after accounting for fixed effects, the random intercept highlights differences in sales volume between categories. For example Categories such as Kitchen and Catering perform worse than average.

```
model_lmer <- lmer(sales_volume ~ price  + high_score + comment_num + (1 | leve_2_category_name), data =

# Extracting random effects
summary(model_lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## sales_volume ~ price + high_score + comment_num + (1 | leve_2_category_name)
##    Data: top5_data
##
## REML criterion at convergence: 94869.3
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -1.8000 -0.2038 -0.0858  0.0256 28.7935
##
## Random effects:
##  Groups               Name        Variance Std.Dev.
##  leve_2_category_name (Intercept)   6628    81.41
##  Residual                         799116   893.93
## Number of obs: 5775, groups:  leve_2_category_name, 5
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 309.11833   59.35148   5.208
## price       -24.42618    2.86823  -8.516
## high_score    3.05424   40.08089   0.076
## comment_num   0.37407    0.02009  18.620
##
## Correlation of Fixed Effects:
##             (Intr) price  hgh_sc
## price       -0.476
## high_score  -0.624  0.082
## comment_num -0.101  0.196 -0.099
```

3.5.1 randon intercet

High RI and high revenue categories (e.g., Men's Shoes and Storage & Organization) are stable, with high baseline sales and high revenue, and can maintain current strategies and enhance them further.

Low RI but high revenue categories (e.g., Kitchen & Dining), despite low baseline volume, are generating high revenue through high unit prices or other factors, and need to focus on optimizing pricing strategies and strengthening promotions to further increase volume.

Low random intercept and low-income categories (e.g., Girl's Dresses and Girl's Sets) are deficient in both baseline volume and market performance, and could be made more competitive through promotion or price optimization.

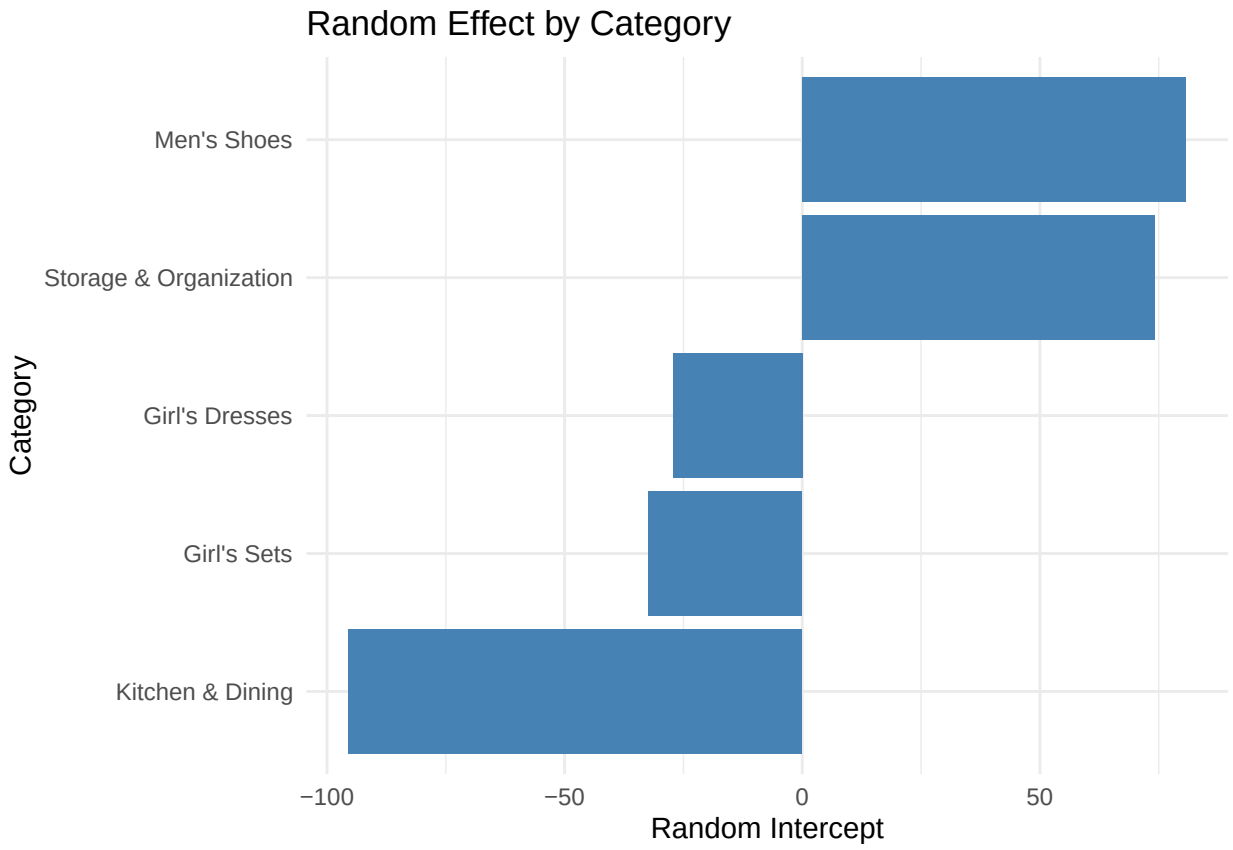
The overall recommendation is to develop targeted strategies to maximize market performance and revenue based on category differences in baseline sales and total revenue.

```
random_effects <- ranef(model_lmer)$leve_2_category_name %>%
  as.data.frame() %>%
  mutate(Category = rownames(ranef(model_lmer)$leve_2_category_name))

# Check the result
head(random_effects)
```

```
##                        (Intercept)              Category
## Girl's Dresses           -27.16198         Girl's Dresses
## Girl's Sets              -32.34419            Girl's Sets
## Kitchen & Dining         -95.50151       Kitchen & Dining
## Men's Shoes               80.78997            Men's Shoes
## Storage & Organization    74.21771 Storage & Organization
```

```
ggplot(random_effects, aes(x = reorder(Category, `(Intercept)`), y = `(Intercept)`)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
```

```
  labs(title = "Random Effect by Category",
       x = "Category",
       y = "Random Intercept") +
  theme_minimal()
```



3.5.2

```
model_lmer1 <- lmer(sales_volume ~ price  + high_score + (1 | comment_num) + (1 | leve_2_category_name)

# Extracting random effects
summary(model_lmer1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: sales_volume ~ price + high_score + (1 | comment_num) + (1 |
##     leve_2_category_name)
##    Data: top5_data
##
## REML criterion at convergence: 94810.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4199 -0.1798 -0.0722  0.0292 28.0843
##
## Random effects:
##  Groups               Name        Variance Std.Dev.
##  comment_num          (Intercept) 142148   377.03
```

```
##  leve_2_category_name (Intercept)   6911    83.13
##  Residual                          760248   871.92
## Number of obs: 5775, groups:  comment_num, 138; leve_2_category_name, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  478.278     72.419   6.604
## price        -24.268      3.106  -7.813
## high_score    12.214     45.932   0.266
##
## Correlation of Fixed Effects:
##            (Intr) price
## price      -0.369
## high_score -0.588  0.062
```

3.5.2

Loweast REML

Add (1 | Price) to explain for the fact that the variability in sales volume may vary significantly across price points

```
model_lmer2 <- lmer(sales_volume ~ (1 | price ) + high_score +  comment_num + (1 | leve_2_category_name)

# Extracting random effects
summary(model_lmer2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: sales_volume ~ (1 | price) + high_score + comment_num + (1 |
##     leve_2_category_name)
##    Data: top5_data
##
## REML criterion at convergence: 93854.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.4170 -0.0944 -0.0196  0.0504 27.9885
##
## Random effects:
##  Groups               Name        Variance Std.Dev.
##  price                (Intercept) 1091826  1044.90
##  leve_2_category_name (Intercept)    2907    53.92
##  Residual                          611640   782.07
## Number of obs: 5775, groups:  price, 200; leve_2_category_name, 5
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 443.86750   90.48299   4.906
## high_score  -13.53257   35.98940  -0.376
## comment_num   0.27985    0.01843  15.181
##
## Correlation of Fixed Effects:
##             (Intr) hgh_sc
## high_score  -0.346
## comment_num -0.052 -0.087
```