OmniMart Retailers

PRESENTED TO
Arijit Midya
(CEO: Manhattan Project)

PRESENTED BY
DS Explorers

Retail Store Data Analysis

Insights and trends for optimal performance

ROADMAP

Univariate Analysis of Numerical Variables

STEP 1

Univariate Analysis of Categorical Variables

STEP 2

Bivariate analysis - Numerical Vs Categorical variable

STEP 3

Bivariate Analysis: Categorical vs categorical

STEP 4

Multivariate Analysis

STEP 5

Customer Behavior Analysis and Key Insights

**STEP 6**

2
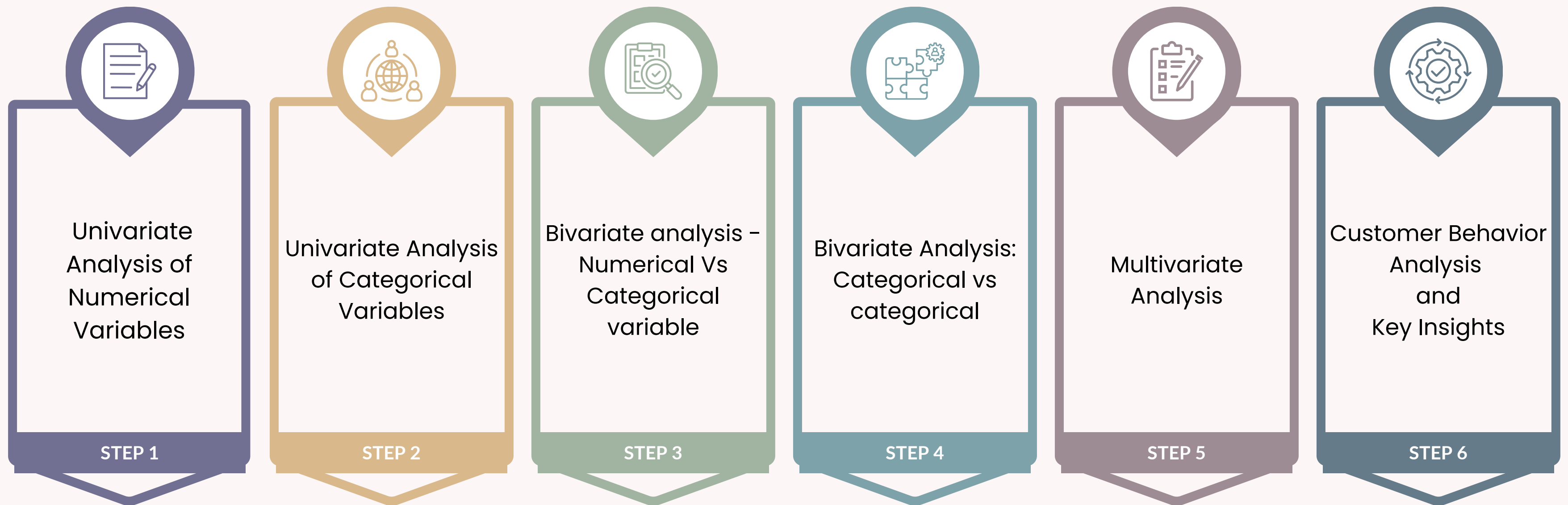
Total amount and Age Distribution

TOTAL AMOUNT DISTRIBUTION

- Total Amount has a mean of approximately 1367.65 and a median of 1041.12, ranging from 10.00 to 4999.63
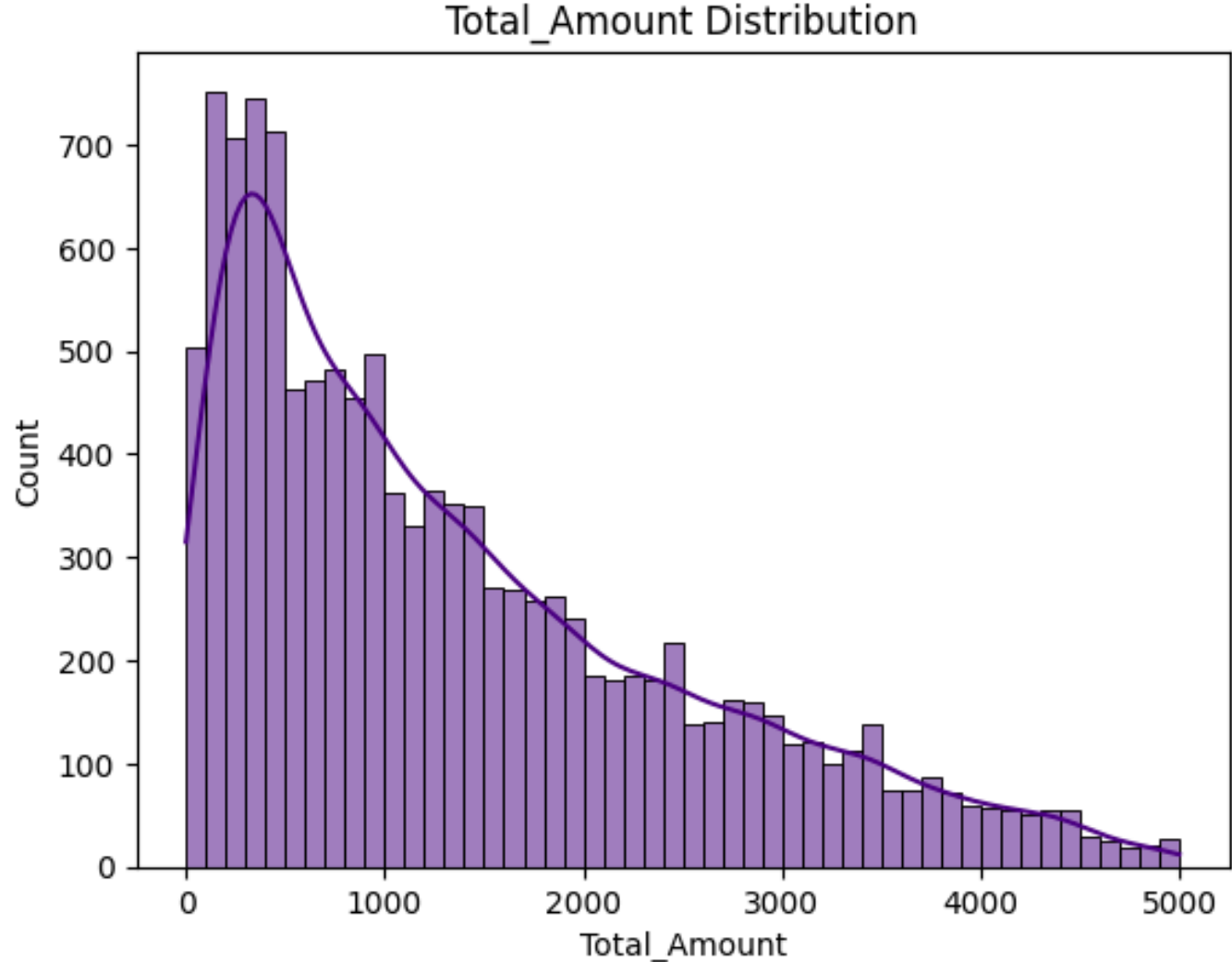
AGE DISTRIBUTION

.The mean Age is approximately 35.48, with a median of 32.00,ranging from 18.00 to 70.00.
.The most frequently occuring age is 20.
.Skewness is positive. Most individuals are younger, but there are some older individuals that stretch the distribution to the right.
.Kurtosis is negetive.The distribution is flatter than a normal distribution.

3
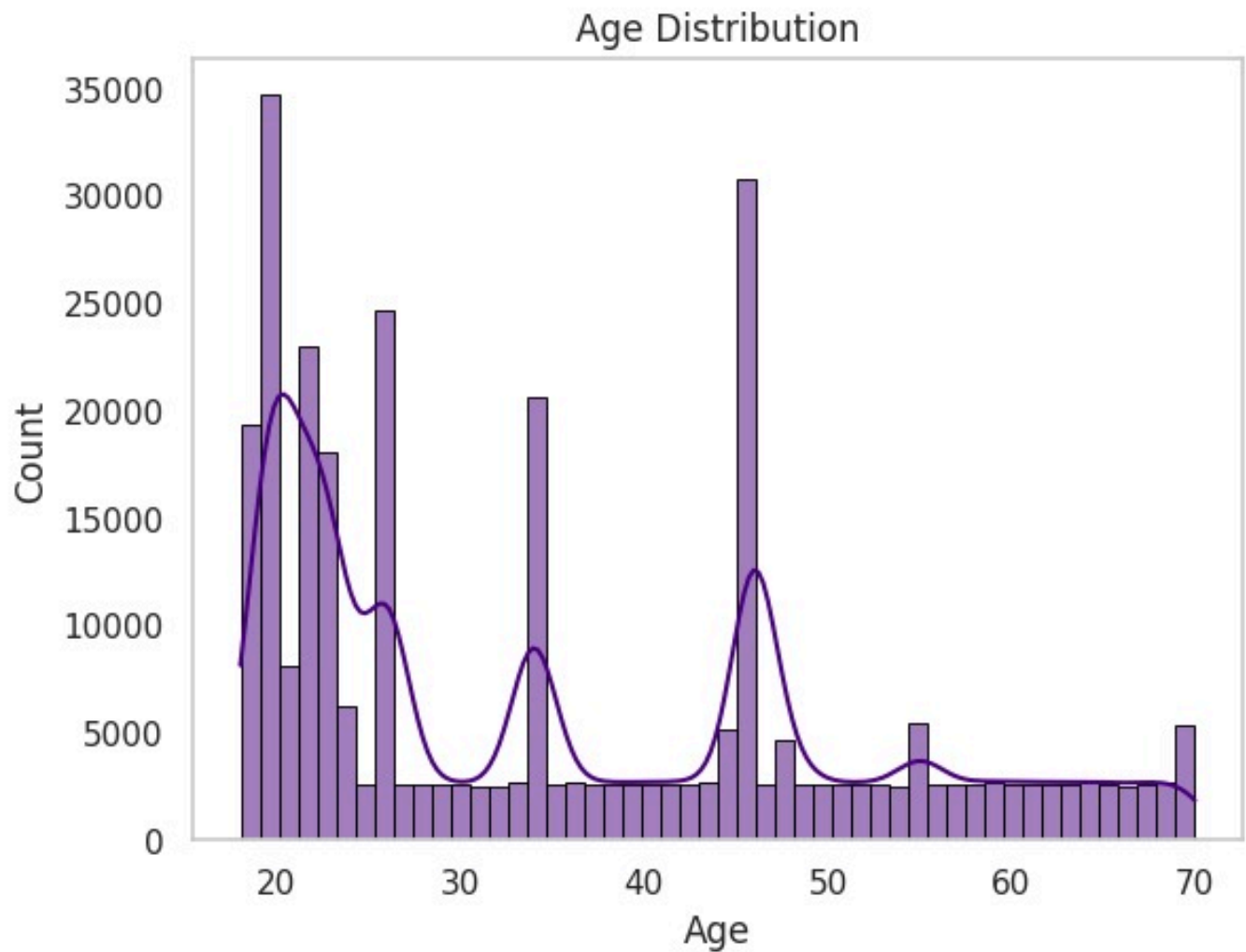
Box Plots of Key Numerical Variables (Outliers & Quartiles)
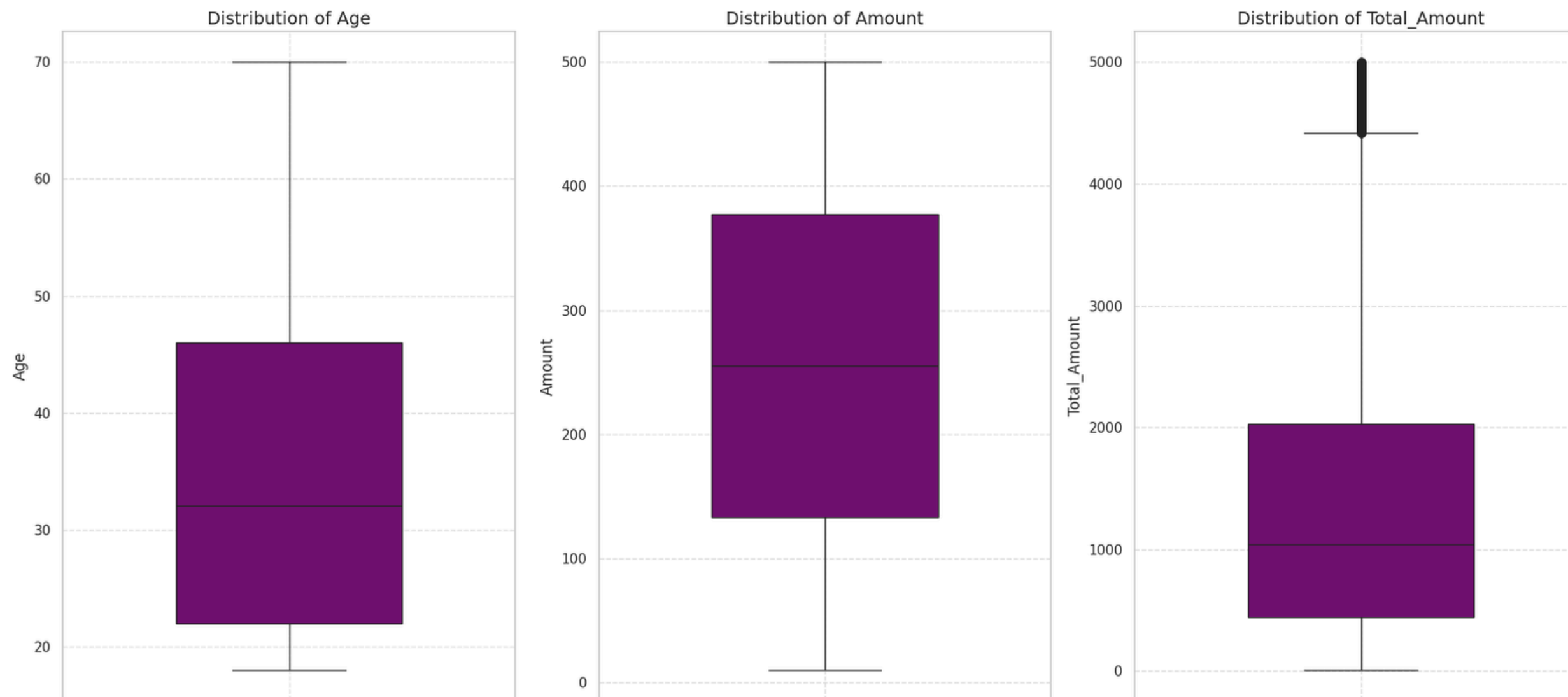
Interpretation

Age: The median customer age is around 32, with the interquartile range (IQR) between approximately 23 and 46. There are no noticeable outliers, suggesting the majority of customers fall within this age range.

Amount: The median amount per transaction is about 250, with the IQR ranging from roughly 120 to 380. The distribution is skewed slightly to the right, indicating a few larger transactions.

Total_Amount: The median total amount spent by a customer is around 1000. The IQR extends from about 500 to 2000. This plot shows a significant number of outliers, represented by the individual points above the upper whisker, indicating there are several customers with very high total spending.

Distribution of Ratings
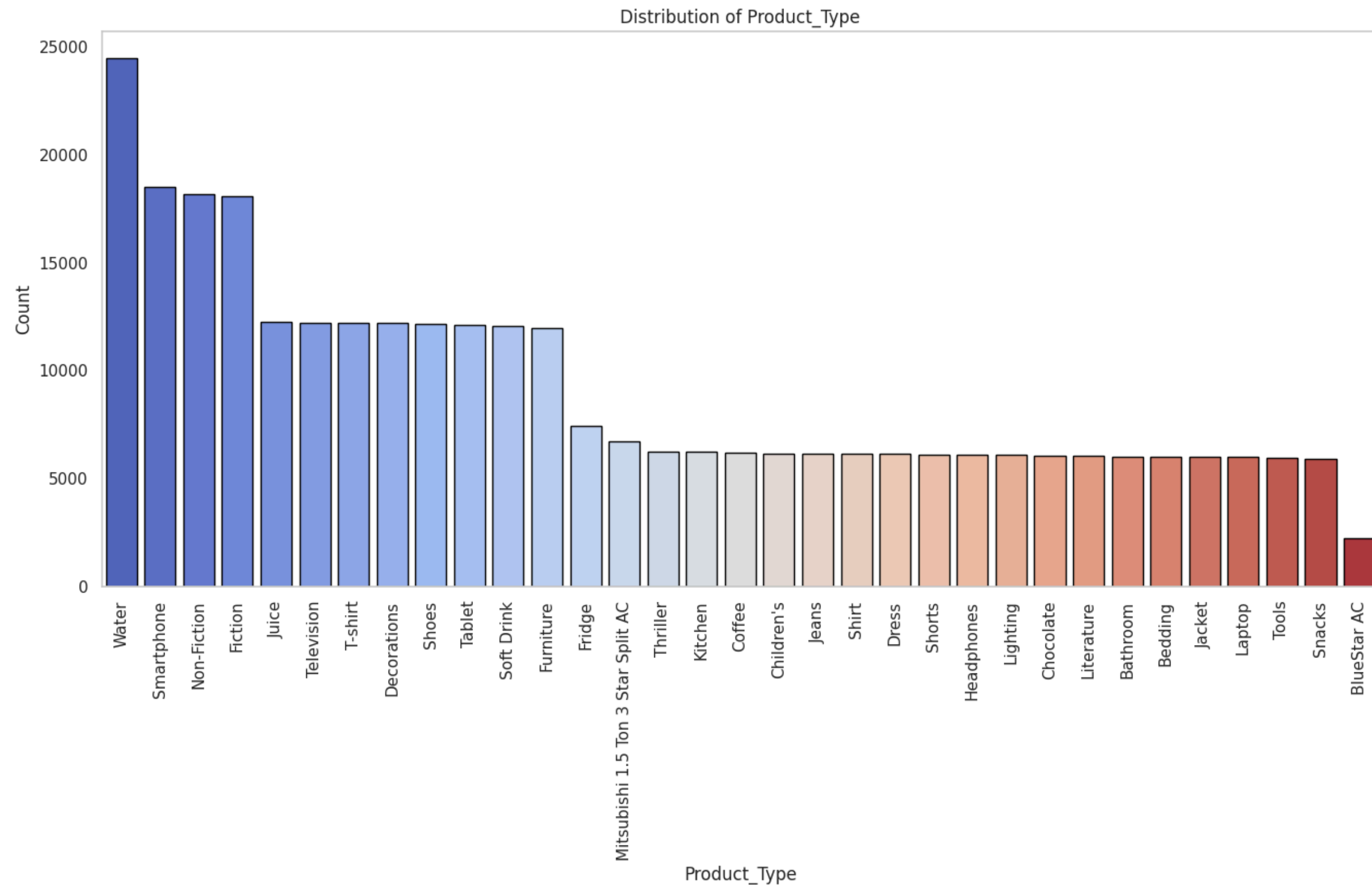

Distribution of Income
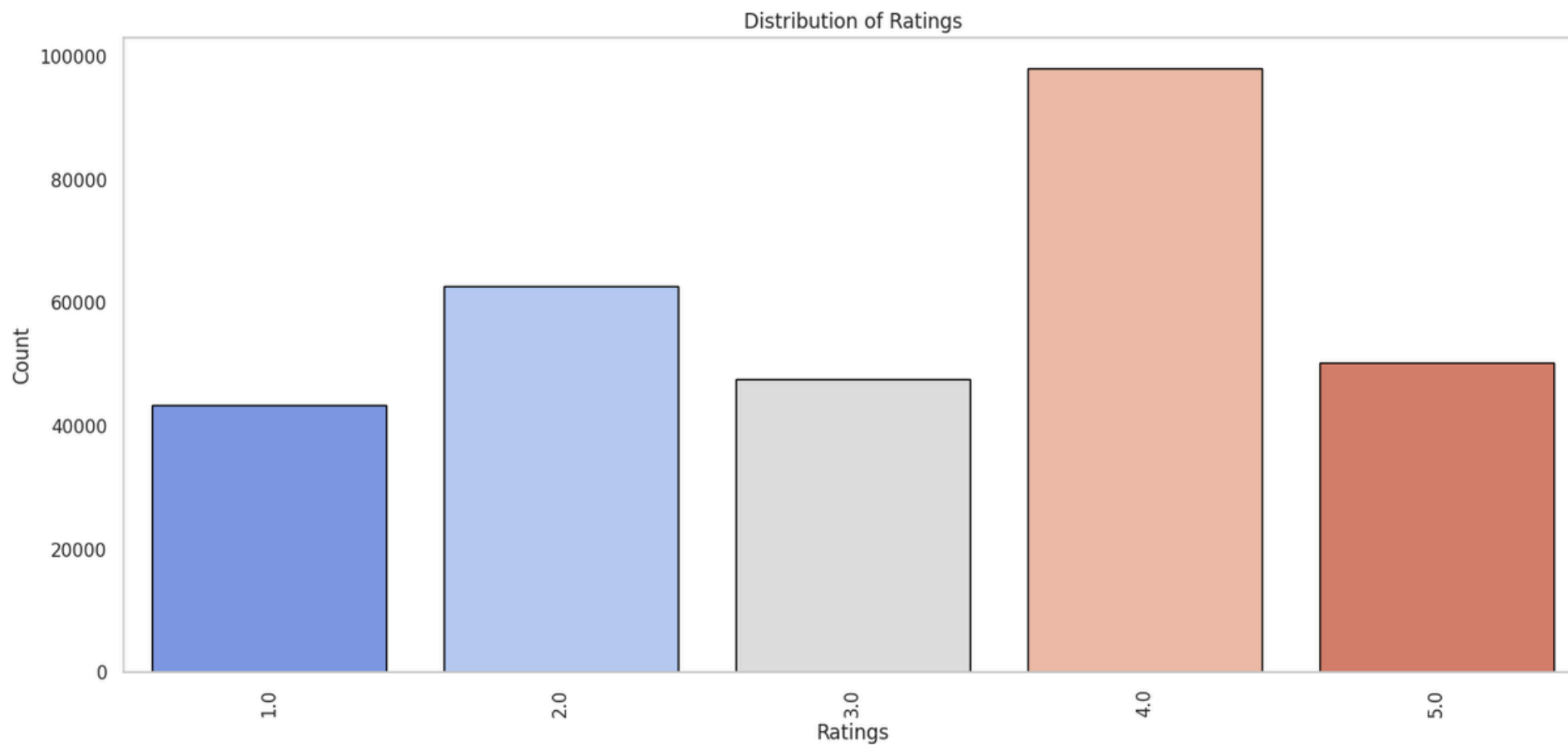
TOTAL RATING DISTRIBUTION

This plot reveals that the most common rating is a 4.0, followed by a 2.0, and then a 5.0. The number of ratings decreases for 1.0, and 3.0. This indicates a general trend towards positive ratings, with the majority of customers giving a good (4.0) or perfect (5.0) score, but a significant portion also providing lower ratings (2.0 and 1.0).
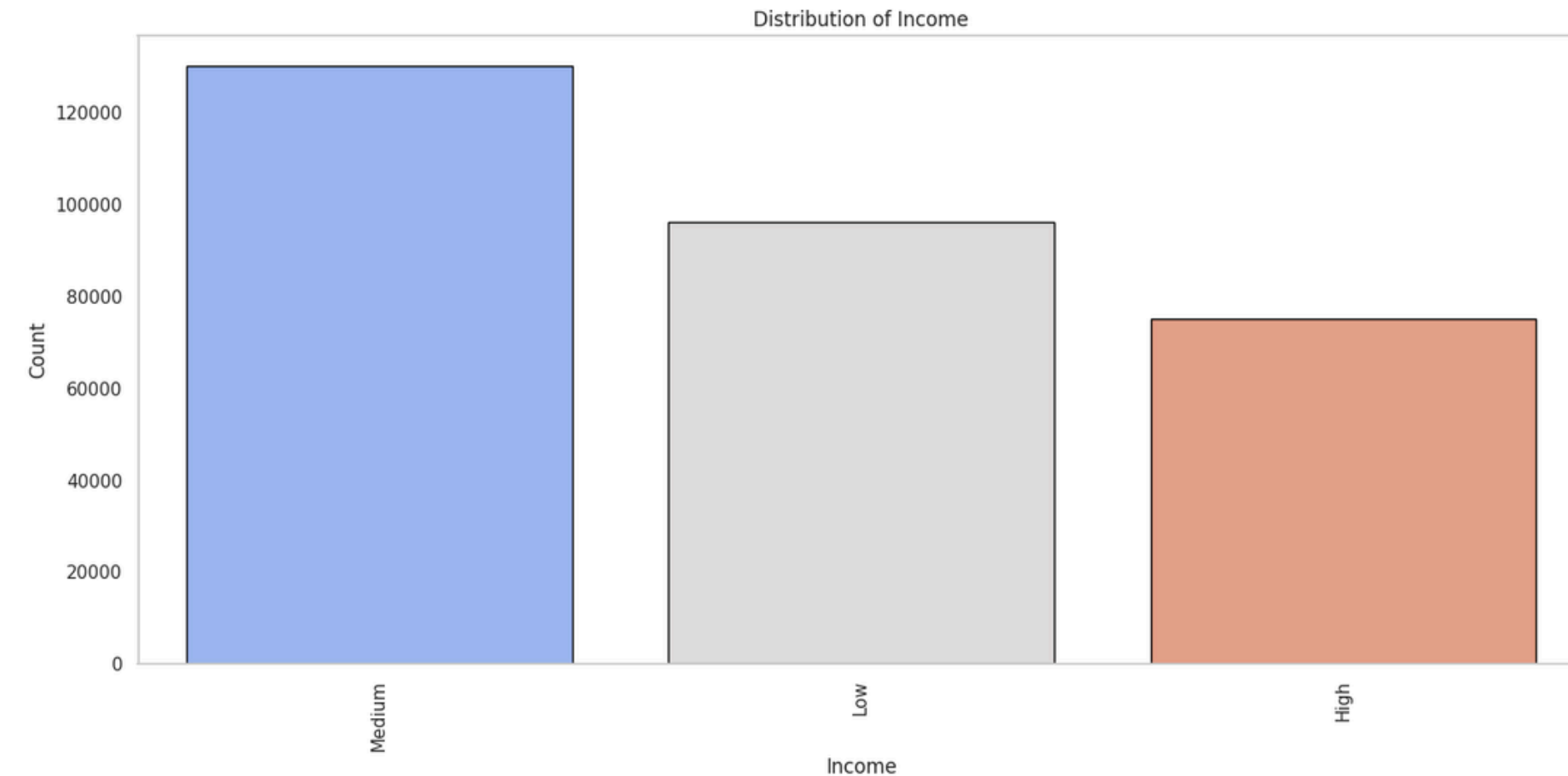
## DISTRIBUTION OF CUSTOMER INCOME

The majority of customers fall into the Medium income category, followed by Low and High income segments.

5

Distribution of Product_Type

From this distribution we can say customers have high demands on products
like - **Water, Fiction , Non-Fiction books and Smartphones**

Distribution of Product_Brand

Distribution of Product_Category
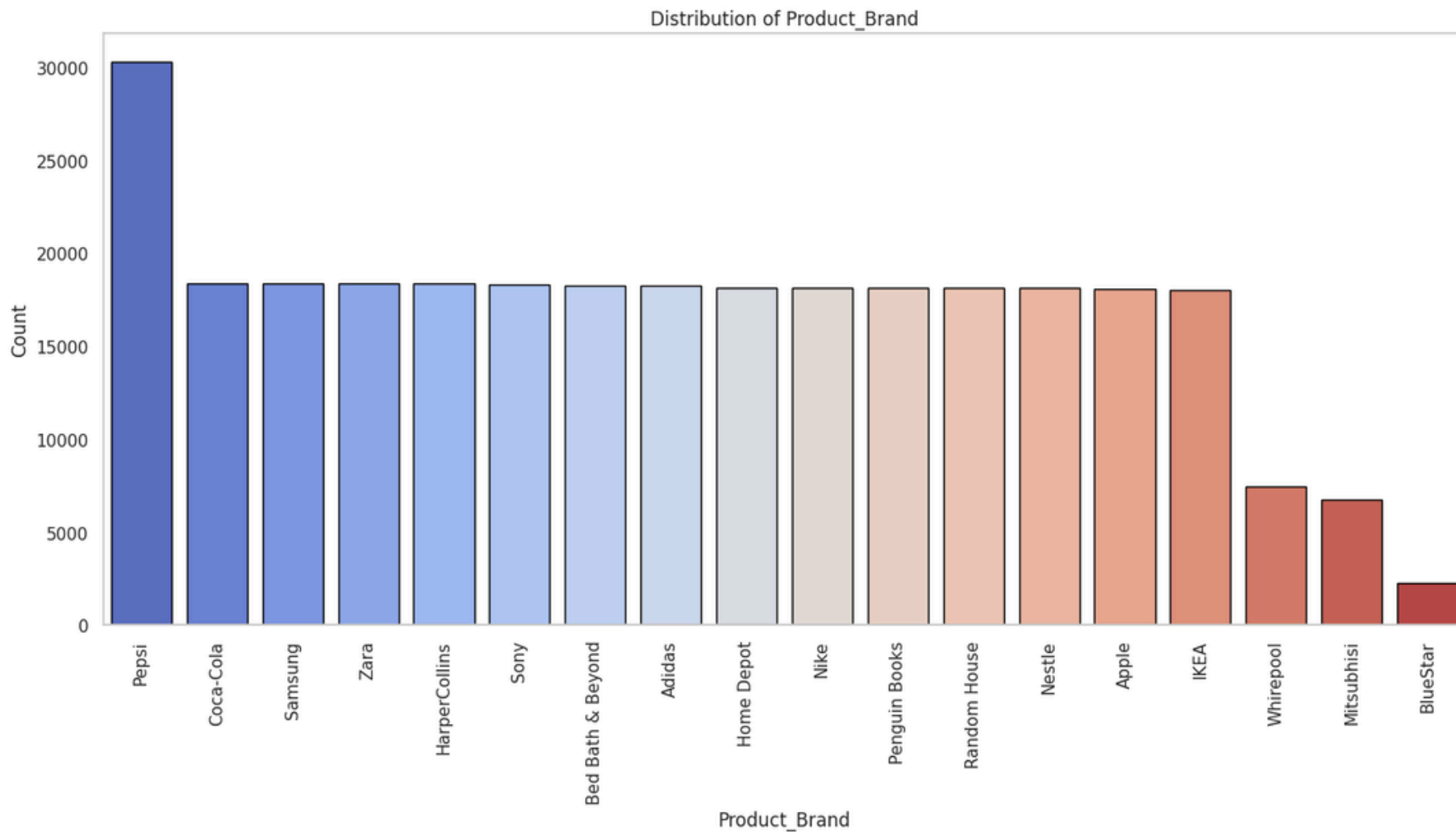
DISTRIBUTION OF PRODUCT BRANDS

Pepsi is by far the most frequently purchased brand, followed by brands like Coca-Cola, Samsung, and Zara. This indicates that drinks and electronics are dominant product areas.
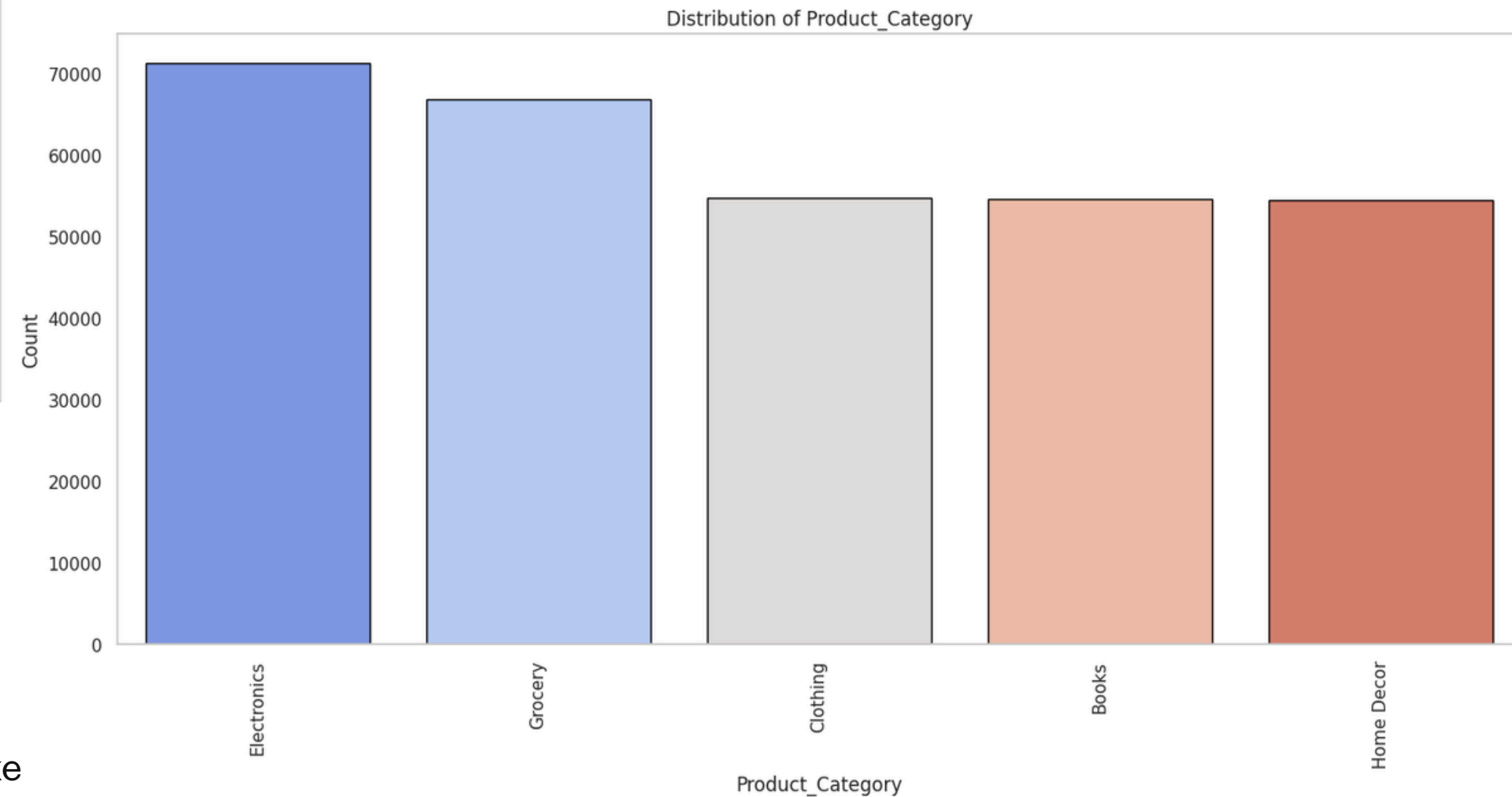
PRODUCT CATEGORY DISTRIBUTION

Electronics and Grocery are the most popular product categories, which aligns with the product brand distribution.

7

Donut Charts
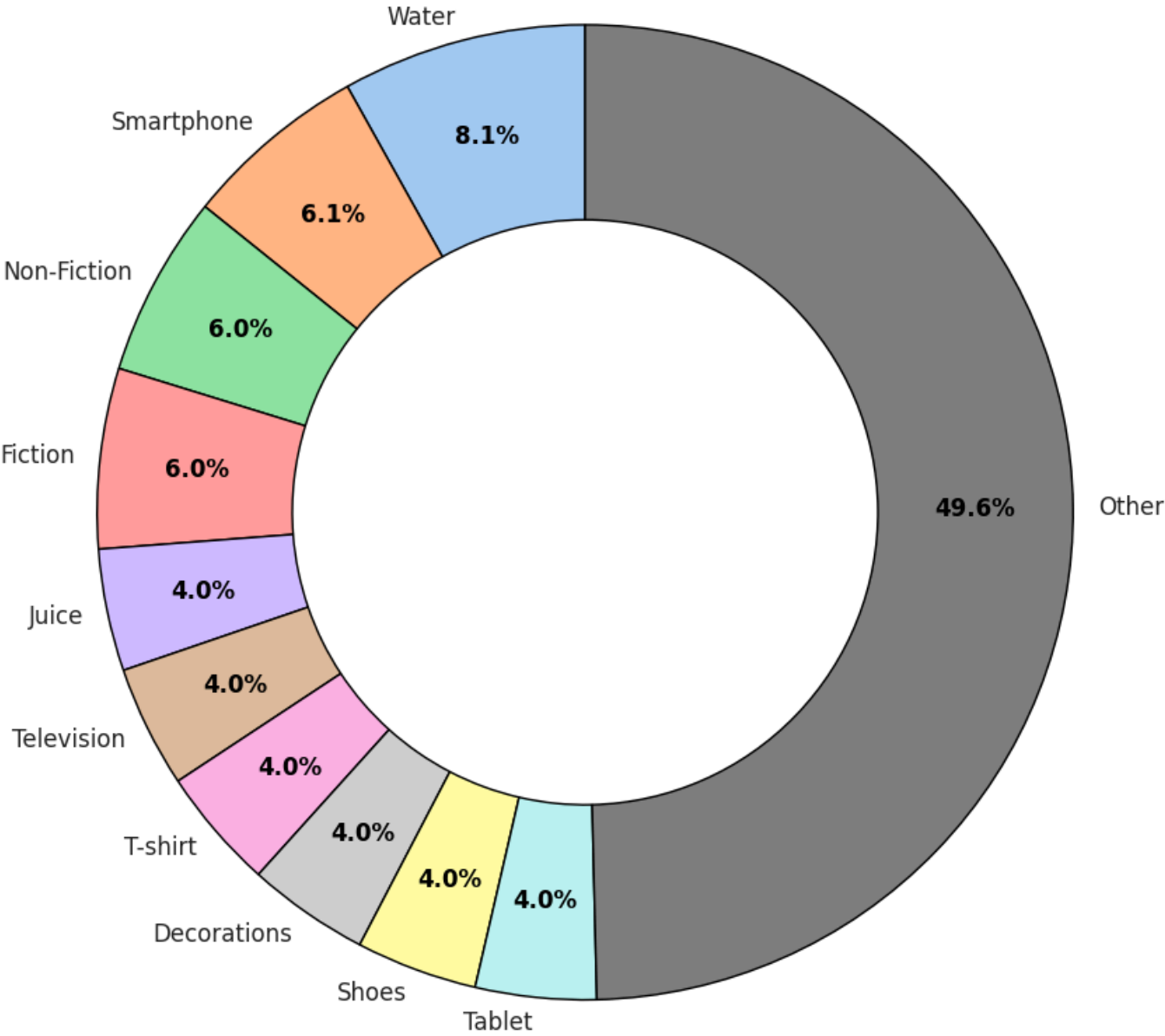
Proportional Distribution of Top 10 Product_Type Categories

- Water 8.1%
- Smartphone 6.1%
- Non-Fiction 6.0%
- Fiction 6.0%
- Juice 4.0%
- Television 4.0%
- T-shirt 4.0%
- Decorations 4.0%
- Shoes 4.0%
- Tablet 4.0%
- Other 49.6%

Proportional Distribution of Top 5 Product_Category Categories

- Electronics 23.6%
- Home Decor 18.0%
- Books 18.1%
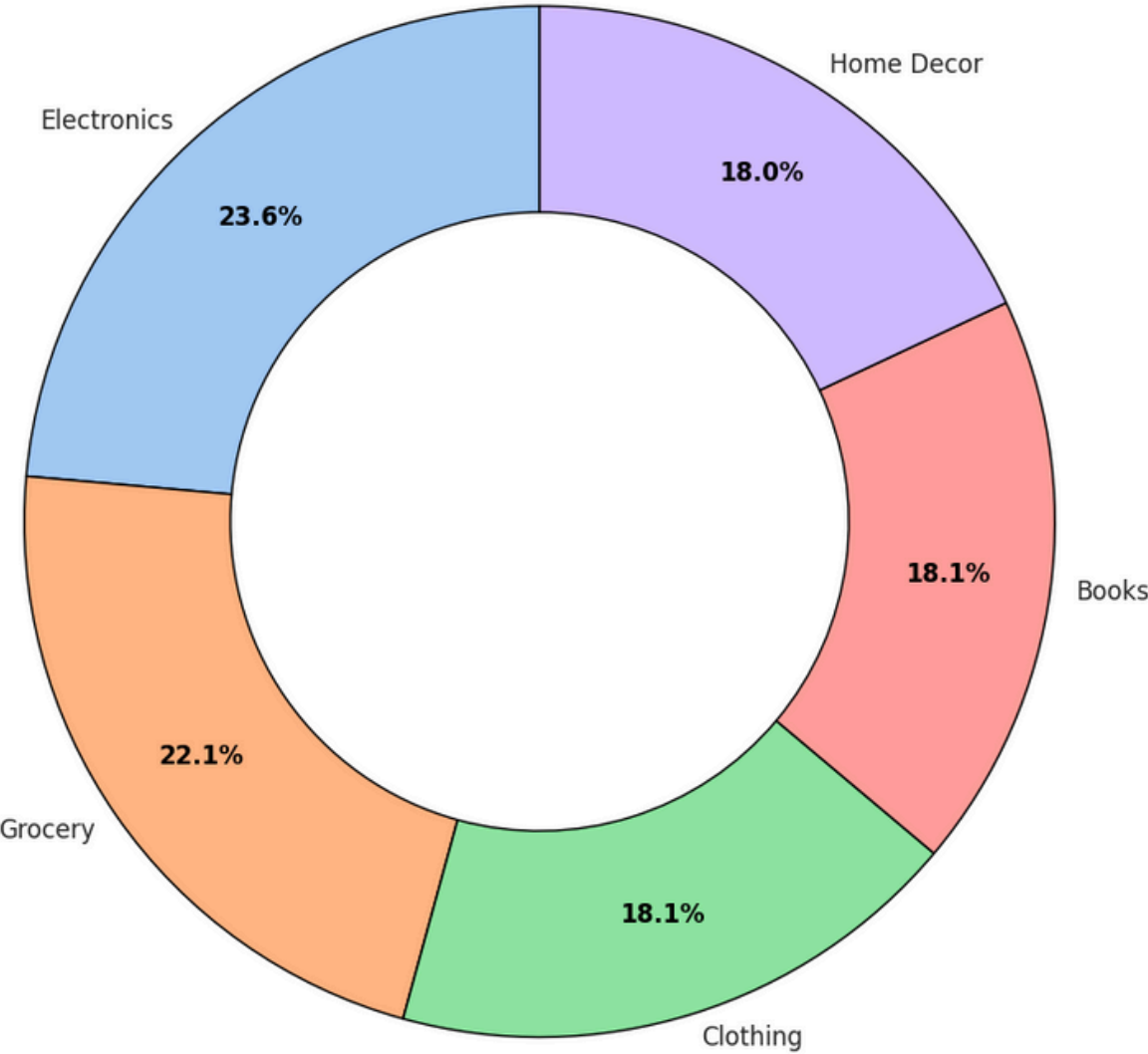- Clothing 18.1%
- Grocery 22.1%

TOP 10 PRODUCT TYPE DISTRIBUTION

The largest proportion of sales by product type comes from a wide variety of items grouped under "Other," which accounts for nearly half (49.6%) of the total. Among the top 10 specific categories, Water is the most sold product type (8.1%), followed by Smartphones (6.1%) and then Non-Fiction and Fiction books, both at 6.0%. This suggests that while there are a few leading product types, the retailer's sales are highly diversified.
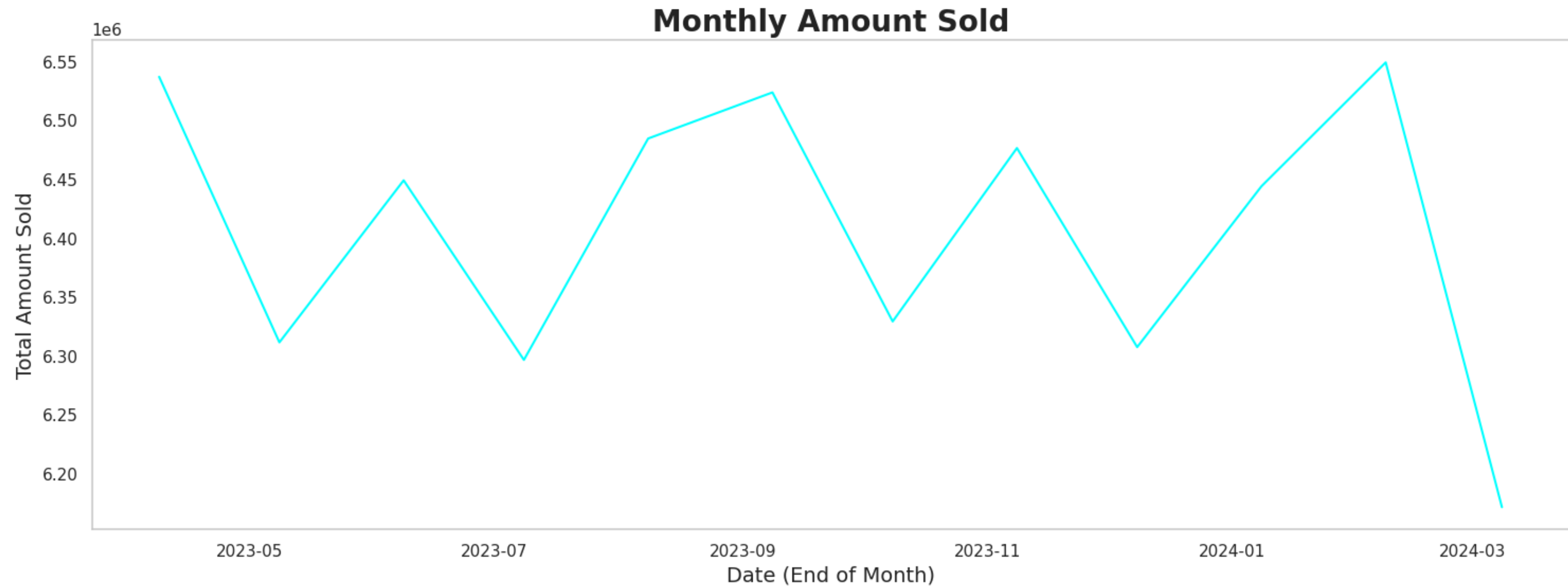
TOP 5 PRODUCT CATEGORY DISTRIBUTION

We can see that electronic products are in high demand among customers
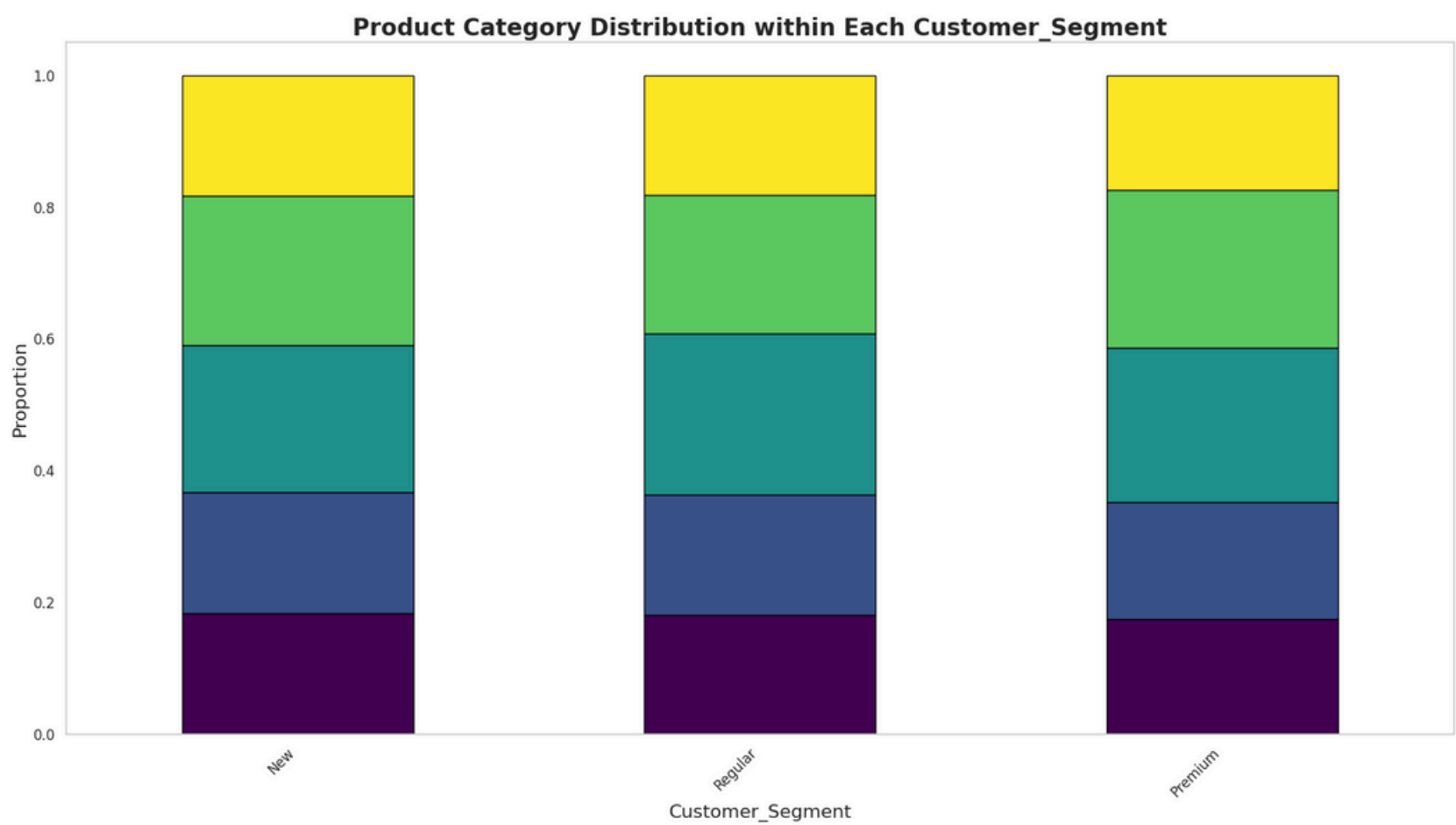
8

Line plot

Monthly Amount Sold

MONTHLY AMOUNT SOLD

We can see sharp peaks during the ending of 2023 (9th and 11th month) and the starting of 2024 (Jan to Feb)

9

Bivariate Analysis: Categorical vs categorical

Product Category Distribution within Each Customer_Segment

The distribution of product categories is also consistent across the New, Regular, and Premium customer segments. This suggests that as **customers become more loyal, they don't necessarily shift their purchasing focus to different product types.**

✅ The overall finding from these charts is that sales are not being driven by a specific demographic or segment's preference for a particular product category. Instead, **the general purchasing pattern is stable and consistent across the entire customer base.** This highlights the broad appeal of product mix.

Customer Segment Distribution within Each Country
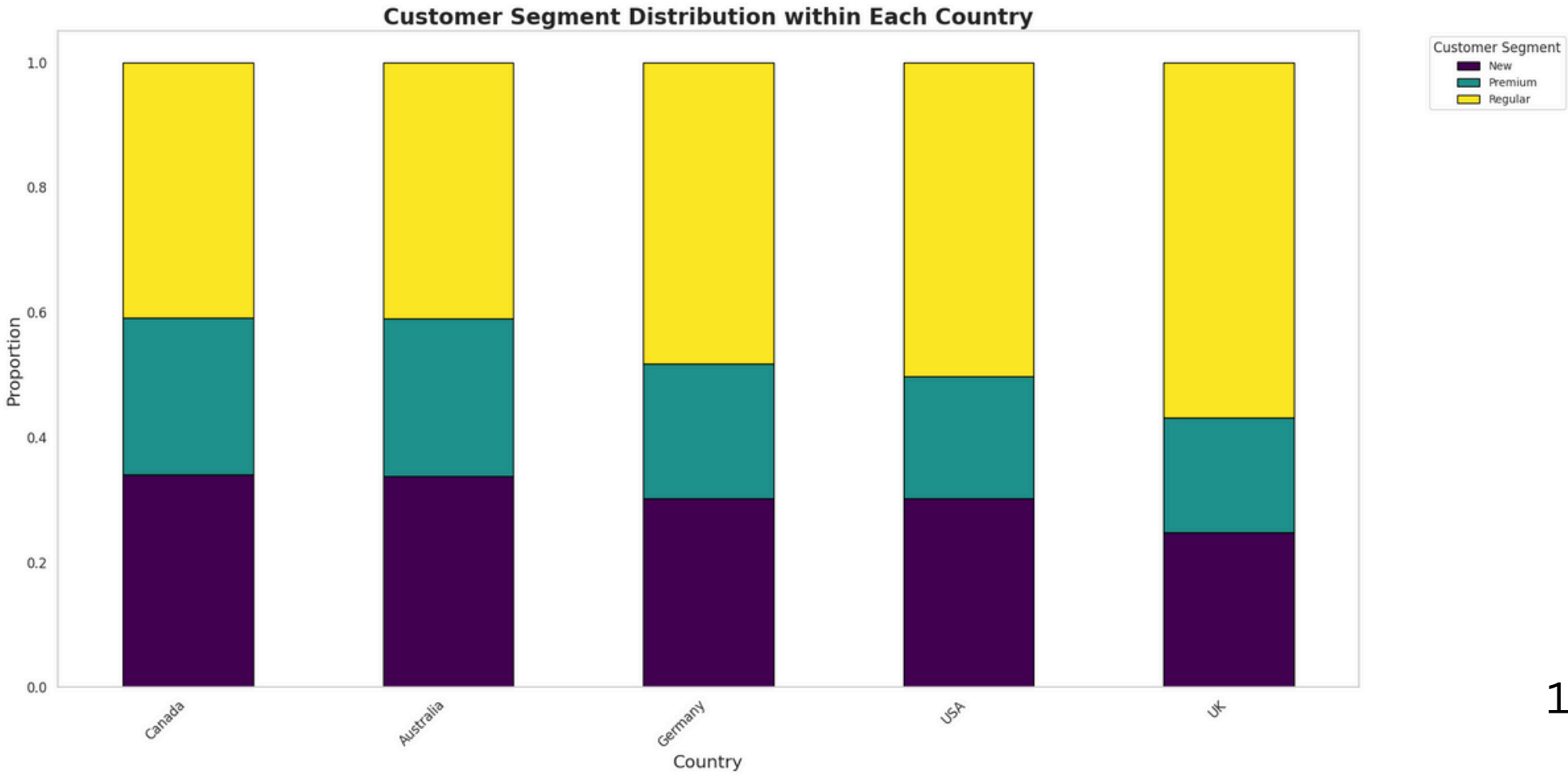
Bivariate Analysis

✓ These stacked bar charts illustrate the distribution of customer segments within different demographic groups.

Customer Segment Distribution within Each Gender: **The distribution of customer segments is nearly identical for both females and males.** This indicates that a customer's gender does not influence whether they fall into the New, Regular, or Premium segments..

Customer Segment Distribution within Each Gender

Customer Segment Distribution within Each Income
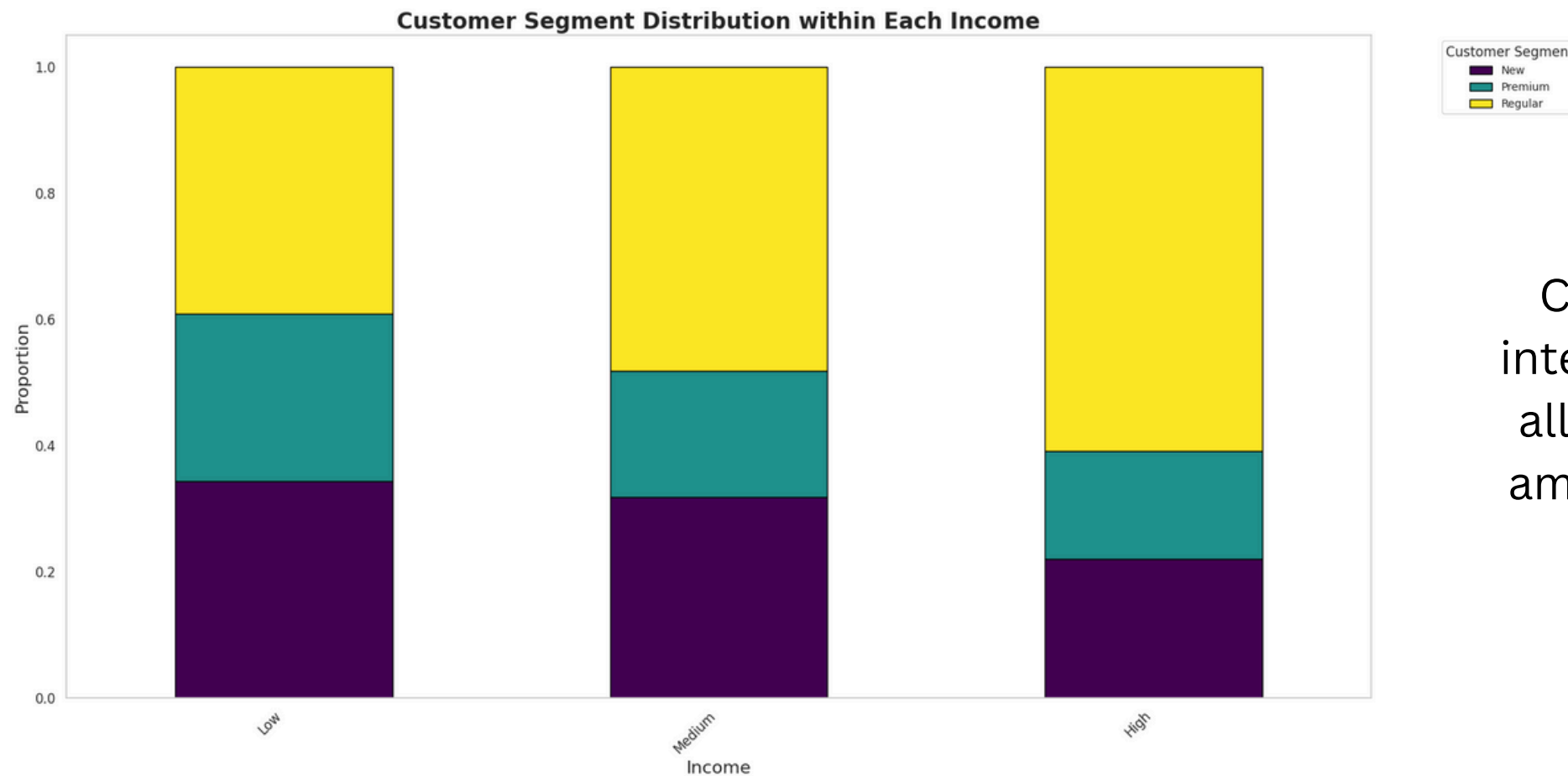
Customer Segment Distribution within Each Income: This chart shows an interesting insight. While the Regular customer segment is the largest across all income levels, the proportion of Premium customers is relatively higher among Low income individuals. The proportion of New customers is slightly higher among Low income individuals.

11

Multivariate Analysis

Average Amount Spent by Age and Income Group
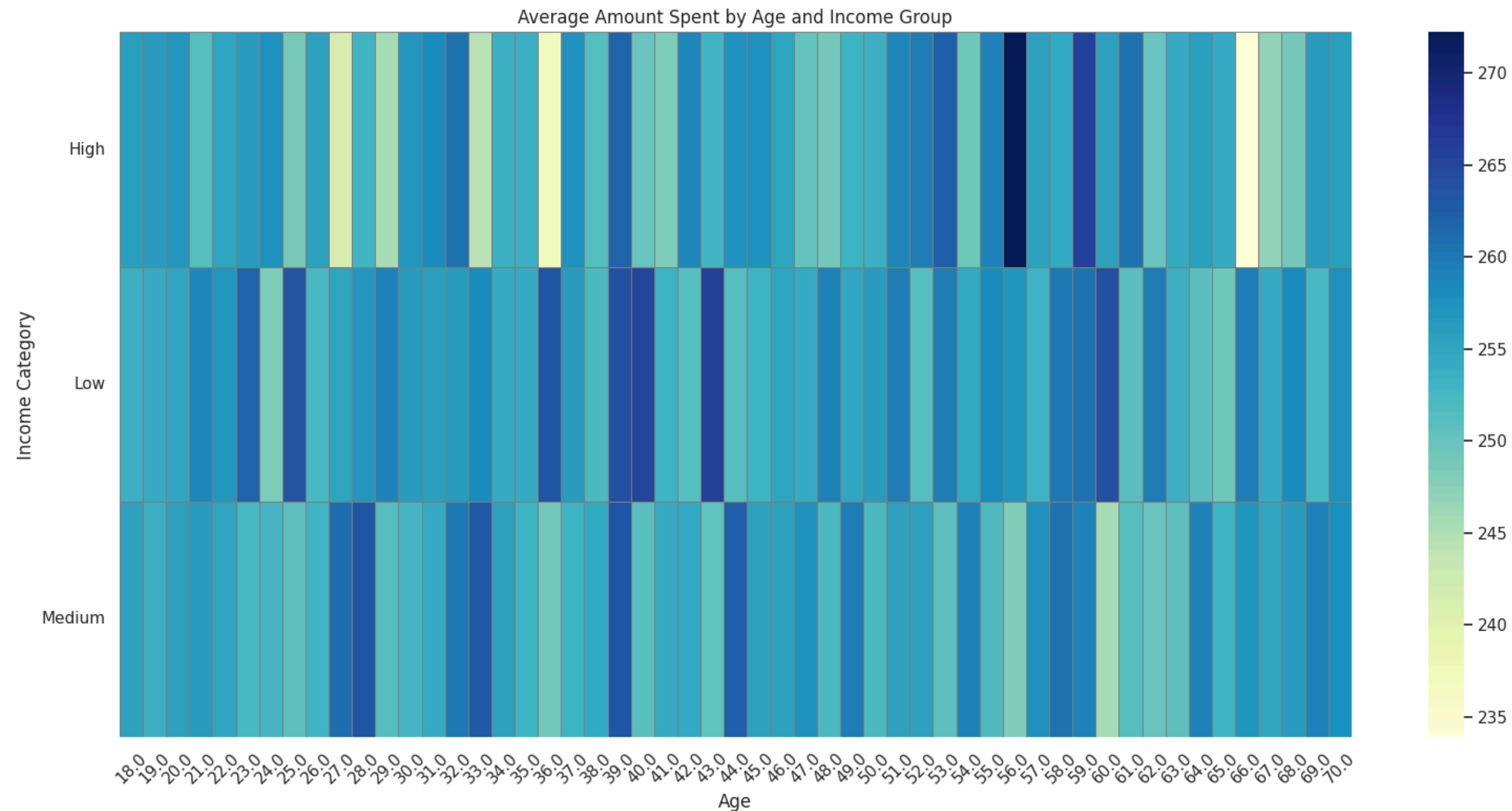
Average Amount Spent by Age and Income Group
This heat map visualizes the average amount spent by customer age and income category. **The darker colors represent higher average spending**. It's clear that average spending is generally consistent across all income groups and ages, with most values clustering around the $250-260 range. However, there are a few notable instances of higher spending, particularly for High income customers around the age of 57 and 39.

Multivariate Analysis

Total Purchases by Product Category and Customer Segment
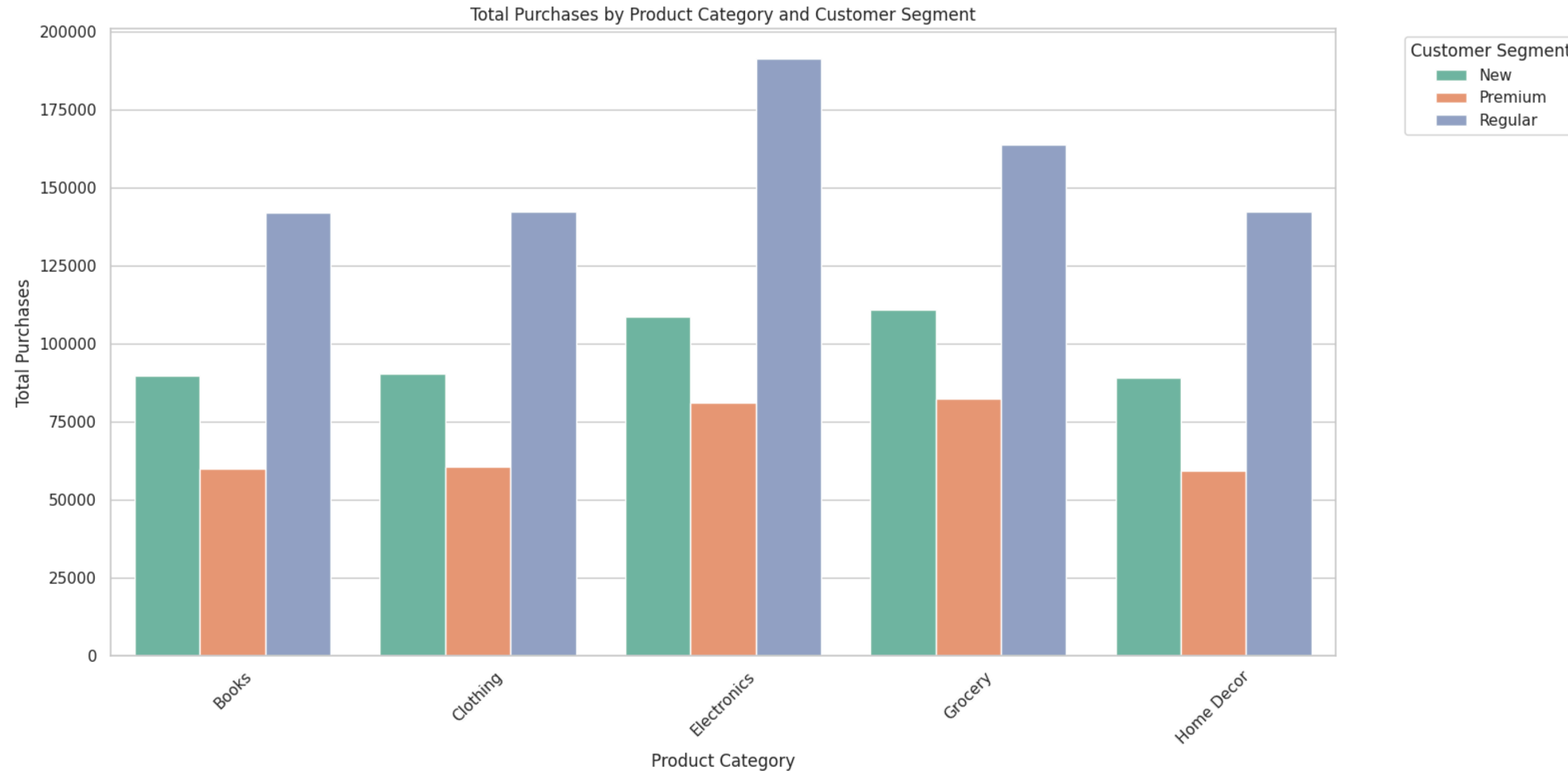
Total Purchases by Product Category and Customer Segment

This grouped bar chart displays the total number of purchases for each product category, broken down by customer segment. The Regular customer segment consistently makes the highest number of purchases across all product categories, followed by the New segment. The Premium segment makes the fewest purchases. This indicates that the business relies heavily on its regular customer base for sales volume, while there's a smaller, but still valuable, group of new and premium customers.

13

Multivariate Analysis

Monthly Sales Trends by Product Type
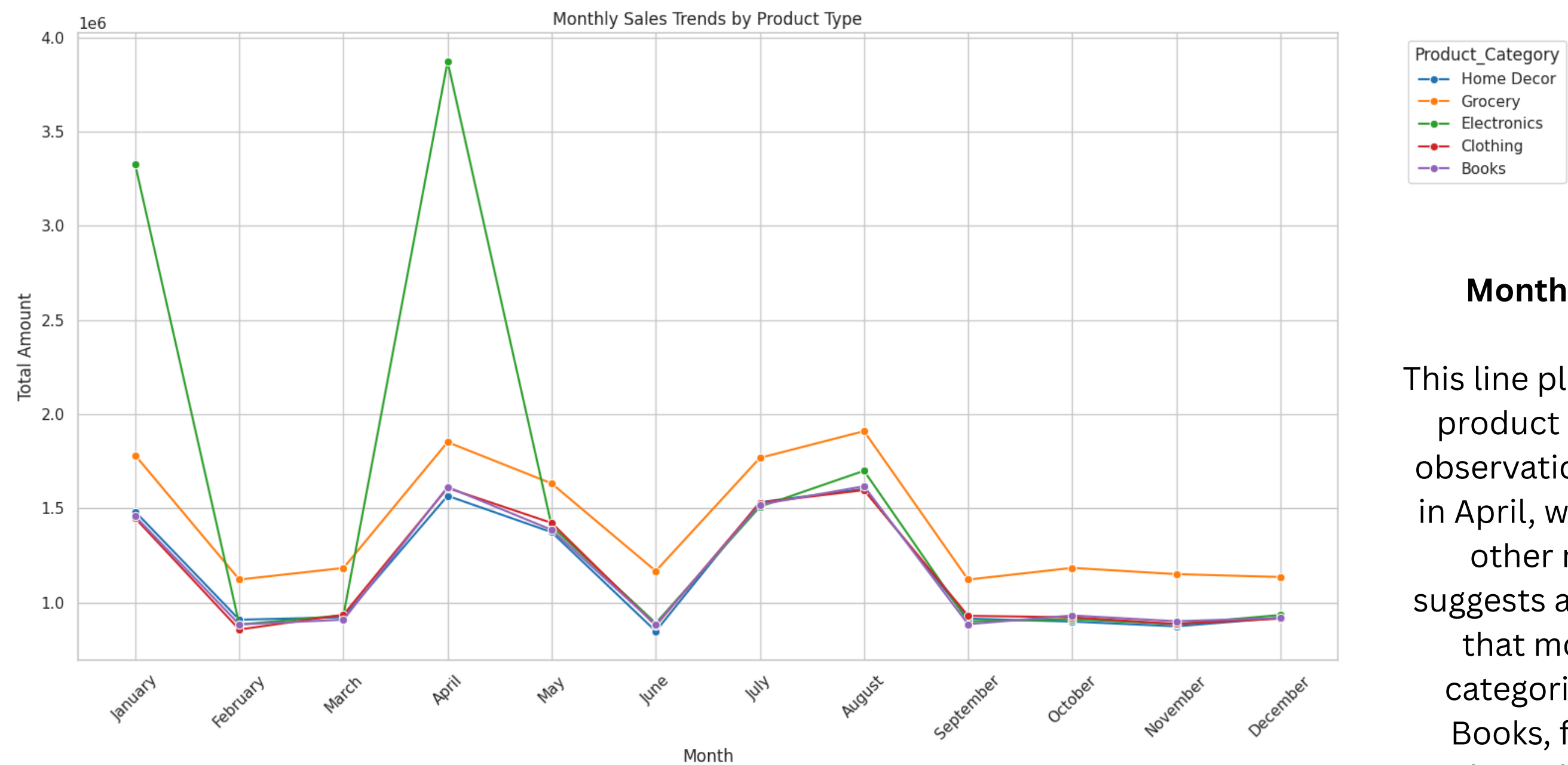
Monthly Sales Trends by Product Category

This line plot shows the total sales amount for each product category over a year. The most striking observation is the large sales spike for Electronics in April, which is a significant outlier compared to other months and product categories. This suggests a seasonal event or a major promotion in that month. Otherwise, sales for all product categories, particularly Grocery, Clothing, and Books, follow a similar, relatively stable trend throughout the year, with a dip around June.

Outlier Analysis

Outlier Density by Product Type and Customer Segment

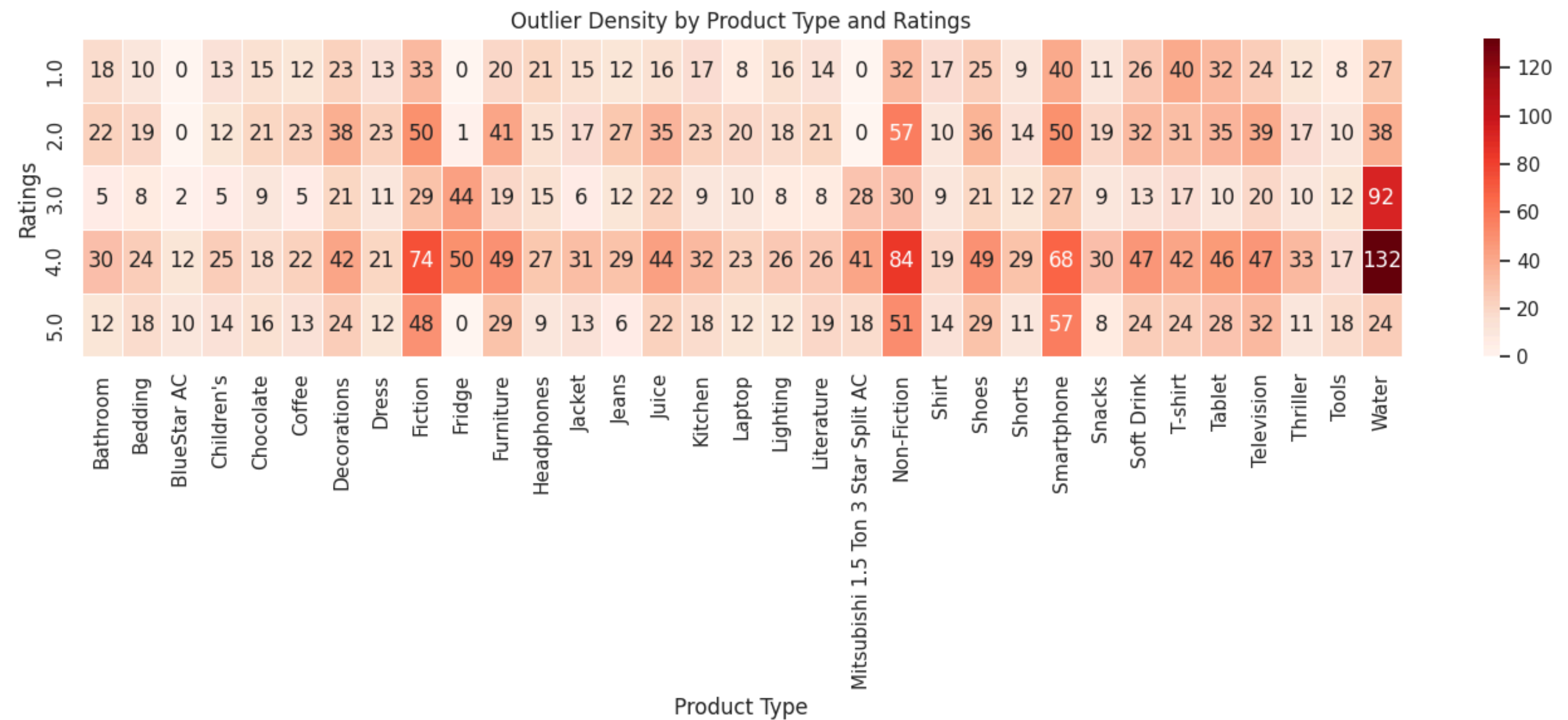| Customer Segment	Bathroom	Bedding	BlueStar AC	Children's	Chocolate	Coffee	Decorations	Dress	Fiction	Fridge	Furniture	Headphones	Jacket	Jeans	Juice	Kitchen	Laptop	Lighting	Literature	Mitsubishi 1.5 Ton 3 Star Split AC	Non-Fiction	Shirt	Shoes	Shorts	Smartphone	Snacks	Soft Drink	T-shirt	Tablet	Television	Thriller	Tools	Water
New	25	18	9	21	20	22	50	22	84	5	39	31	27	29	49	29	21	27	31	38	73	23	40	26	72	19	38	47	59	37	33	19	96
Premium	8	19	4	12	16	14	22	18	37	15	31	18	16	16	29	22	15	14	19	24	62	10	37	17	53	12	37	33	32	40	14	18	78
Regular	54	42	11	36	43	39	76	40	113	75	88	38	39	41	61	48	37	39	38	25	119	36	83	32	117	46	67	74	60	86	36	28	138

Product Type

## OUTLIER DENSITY BY PRODUCT TYPE AND CUSTOMER SEGMENT

The Regular customer segment has the highest number of outliers for almost every product type. This reinforces the finding that most loyal customers are the ones making the largest, most unusual purchases.

15

Outlier Analysis

Outlier Density by Product Type and Ratings

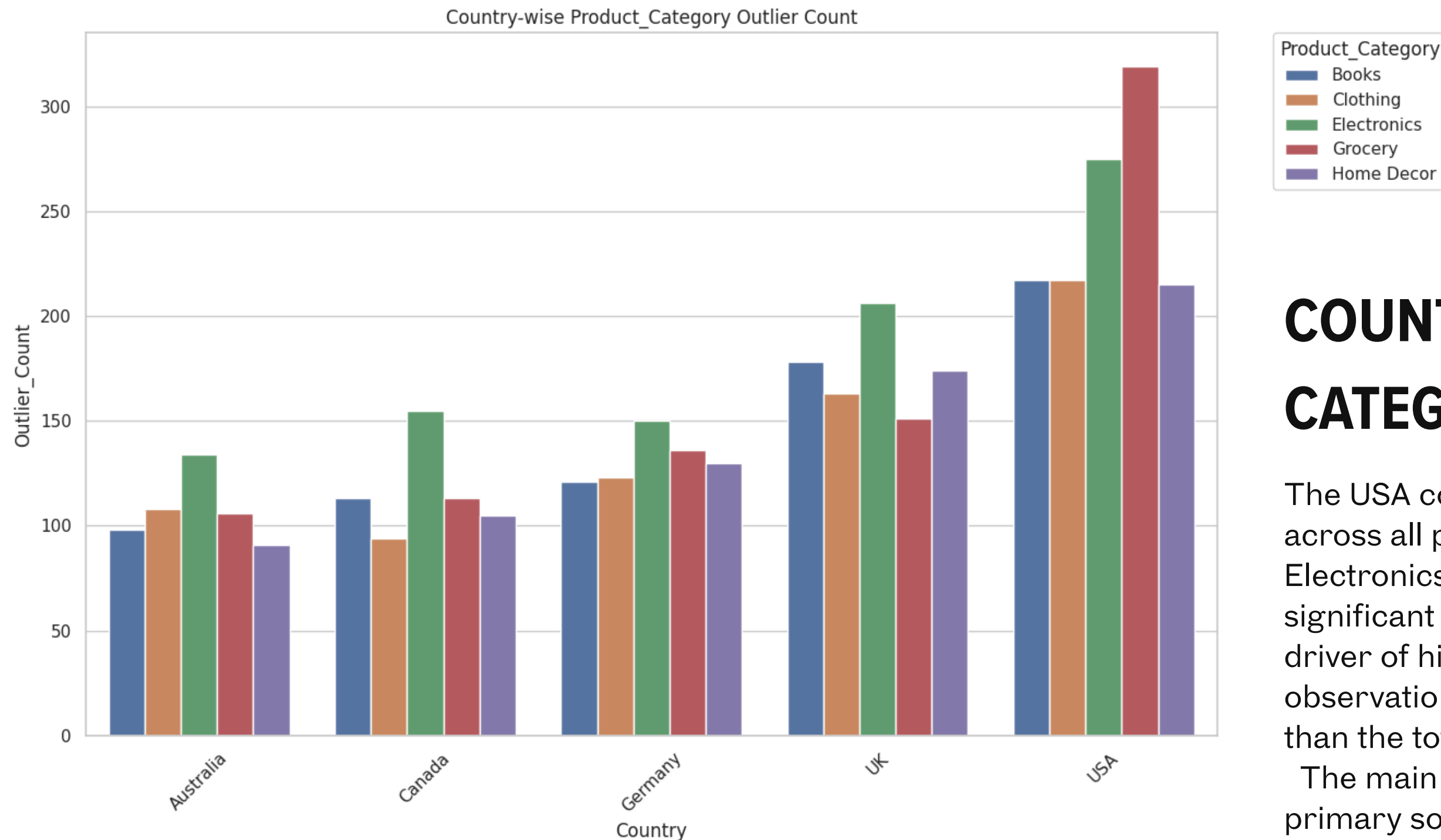| Ratings	Bathroom	Bedding	BlueStar AC	Children's	Chocolate	Coffee	Decorations	Dress	Fiction	Fridge	Furniture	Headphones	Jacket	Jeans	Juice	Kitchen	Laptop	Lighting	Literature	Mitsubishi 1.5 Ton 3 Star Split AC	Non-Fiction	Shirt	Shoes	Shorts	Smartphone	Snacks	Soft Drink	T-shirt	Tablet	Television	Thriller	Tools	Water
1.0	18	10	0	13	15	12	23	13	33	0	20	21	15	12	16	17	8	16	14	0	32	17	25	9	40	11	26	40	32	24	12	8	27
2.0	22	19	0	12	21	23	38	23	50	1	41	15	17	27	35	23	20	18	21	0	57	10	36	14	50	19	32	31	35	39	17	10	38
3.0	5	8	2	5	9	5	21	11	29	44	19	15	6	12	22	9	10	8	8	28	30	9	21	12	27	9	13	17	10	20	10	12	92
4.0	30	24	12	25	18	22	42	21	74	50	49	27	31	29	44	32	23	26	26	41	84	19	49	29	68	30	47	42	46	47	33	17	132
5.0	12	18	10	14	16	13	24	12	48	0	29	9	13	6	22	18	12	12	19	18	51	14	29	11	57	8	24	24	28	32	11	18	24

Product Type

OUTLIER DENSITY BY PRODUCT TYPE AND RATINGS

A deeper dive into product types shows that the highest concentration of outliers is found in Fiction, Non-Fiction, smartphones, water particularly for those with a 4.0 rating.

Country-wise Product Brand Outlier Count
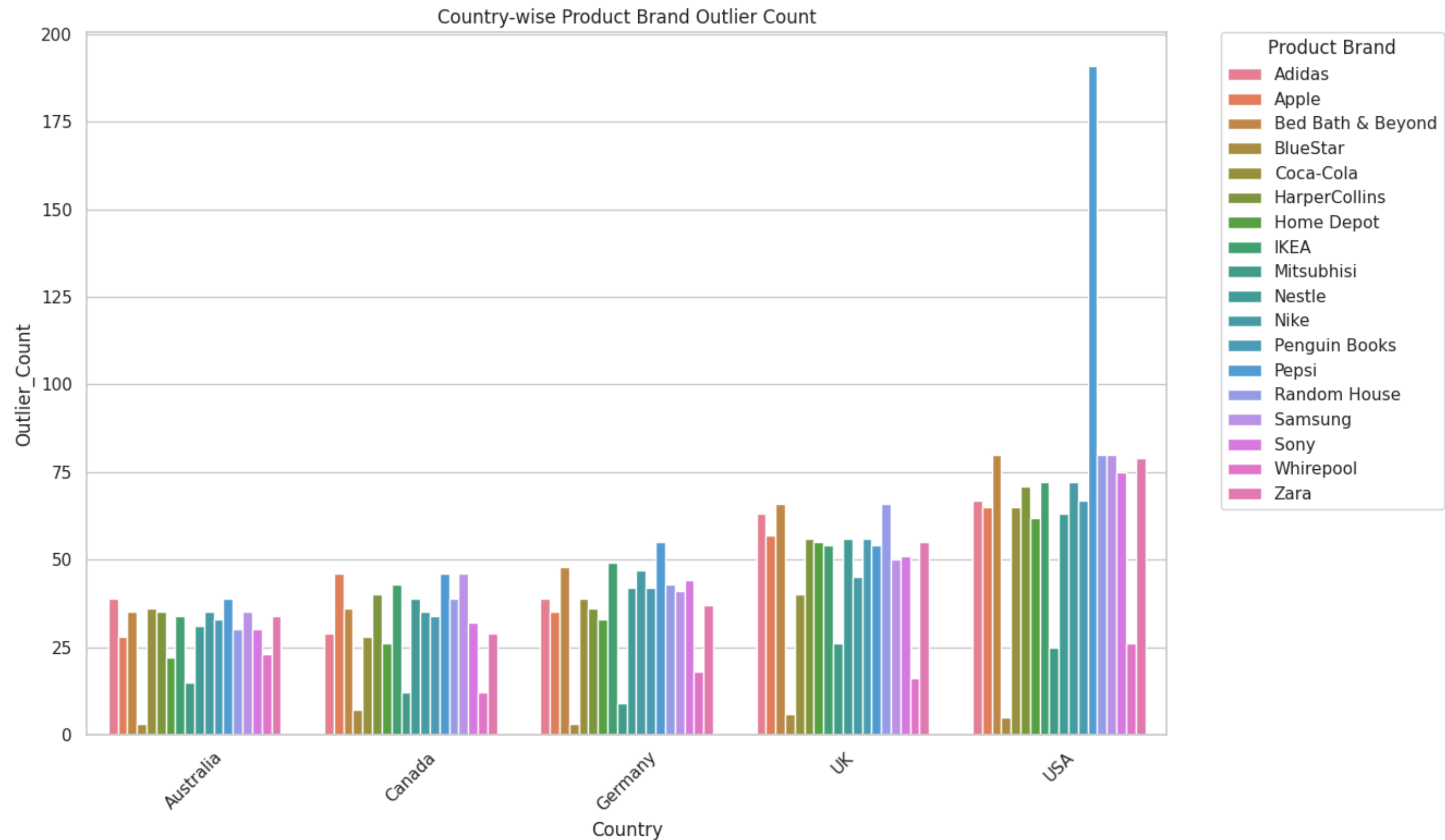
COUNTRY-WISE PRODUCT BRAND OUTLIER COUNT

The chart shows that the USA has a significantly higher number of outliers for almost every product brand compared to all other countries. Within the USA, the Pepsi brand has the most outliers, indicating that high-value purchases are most frequent for this brand in the USA.
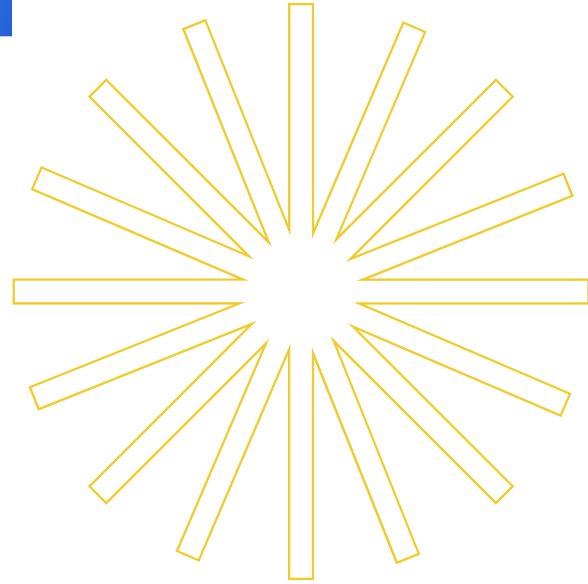
Country-wise Product_Category Outlier Count

The main takeaway from these plots is that the USA is the primary source of high-value outliers in the dataset, especially within the electronics and grocery categories, and for the Pepsi brand.

COUNTRY-WISE PRODUCT CATEGORY OUTLIER COUNT

The USA consistently shows the highest count of outliers across all product categories. This is particularly true for Electronics, which has the most outliers in the USA by a significant margin. This suggests that the US market is a key driver of high-value transactions. But here a significant observation is, in USA Total amount of Groceries sold is more than the total amount of electronics sold.

The main takeaway from these plots is that the USA is the primary source of high-value outliers in the dataset, especially within the electronics and grocery categories, and for the Samsung brand.

Performance Indicators: Customer Behavior Analysis

Key metrics for retail performance evaluation

Distribution of the most loyal customers across different categorical variables.

- Regular segment is the largest
- Medium income customers have high frequencies

Distribution of Purchase Frequency

- Identifies Loyal vs One-time customer
- Reveals customer retention and repeat purchase trends
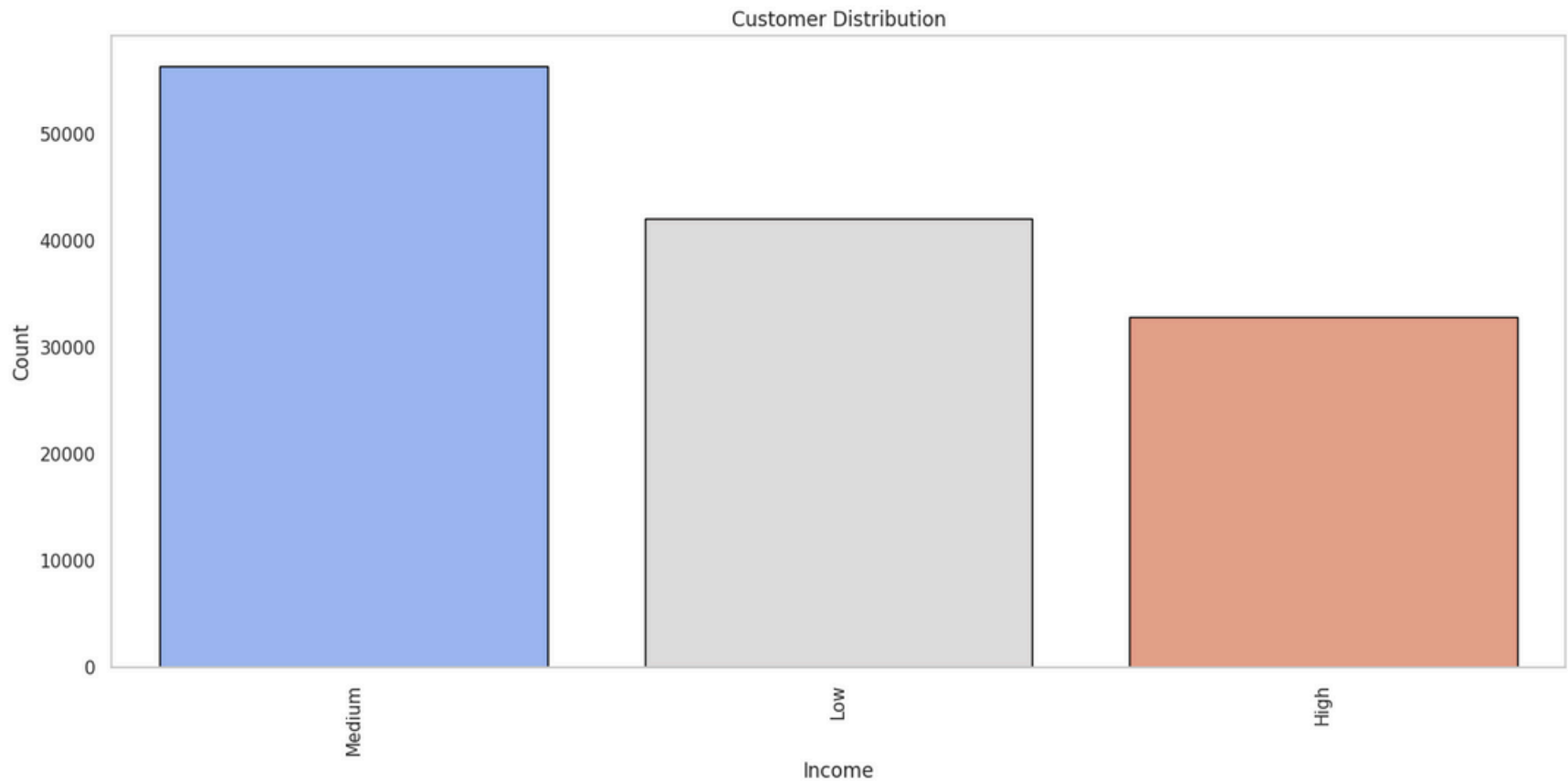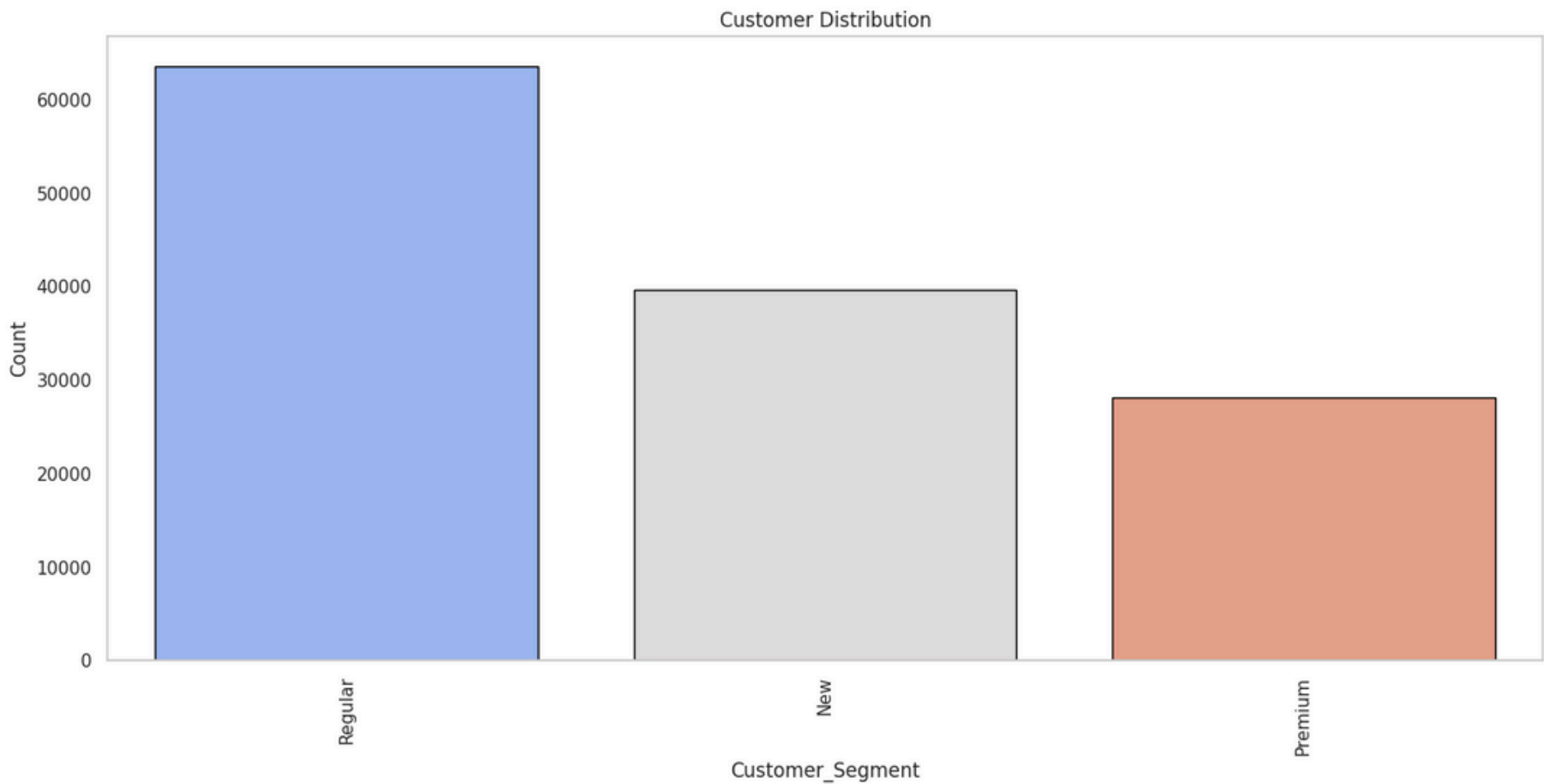- Helps segment customers by buying behavior

Customer Behavior Analysis

The chart is right-skewed, meaning most customers have a low purchase frequency, with the highest number of customers making only two, three or four purchases. The number of customers drops off significantly as the purchase frequency increases, with very few customers making more than 10 purchases. This suggests that while the customer base is large, a small, highly loyal segment of repeat buyers drives a significant portion of total sales.

Customer Behavior Analysis

Customer Distribution

Customer Distribution

Income Distribution:
Customers with a Medium income are the most frequent purchasers, followed by Low and then High income customers.
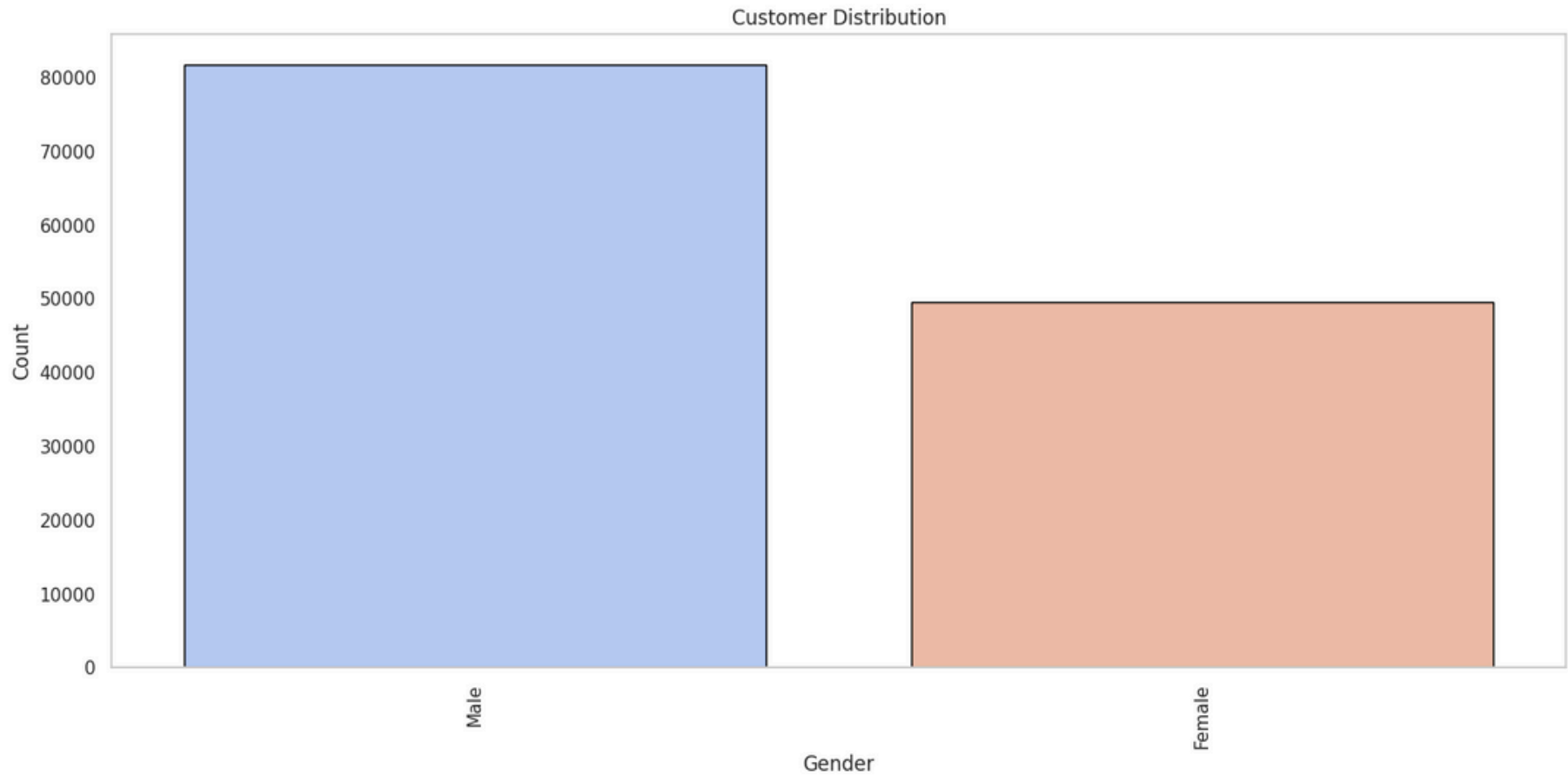
Customer Segment Distribution:
Within the high-frequency customer base, the Regular segment is the largest, followed by New and then Premium customers. This confirms that repeat purchases are primarily driven by the regular customer base, not just by high-spenders (Premium) or new customers.

21

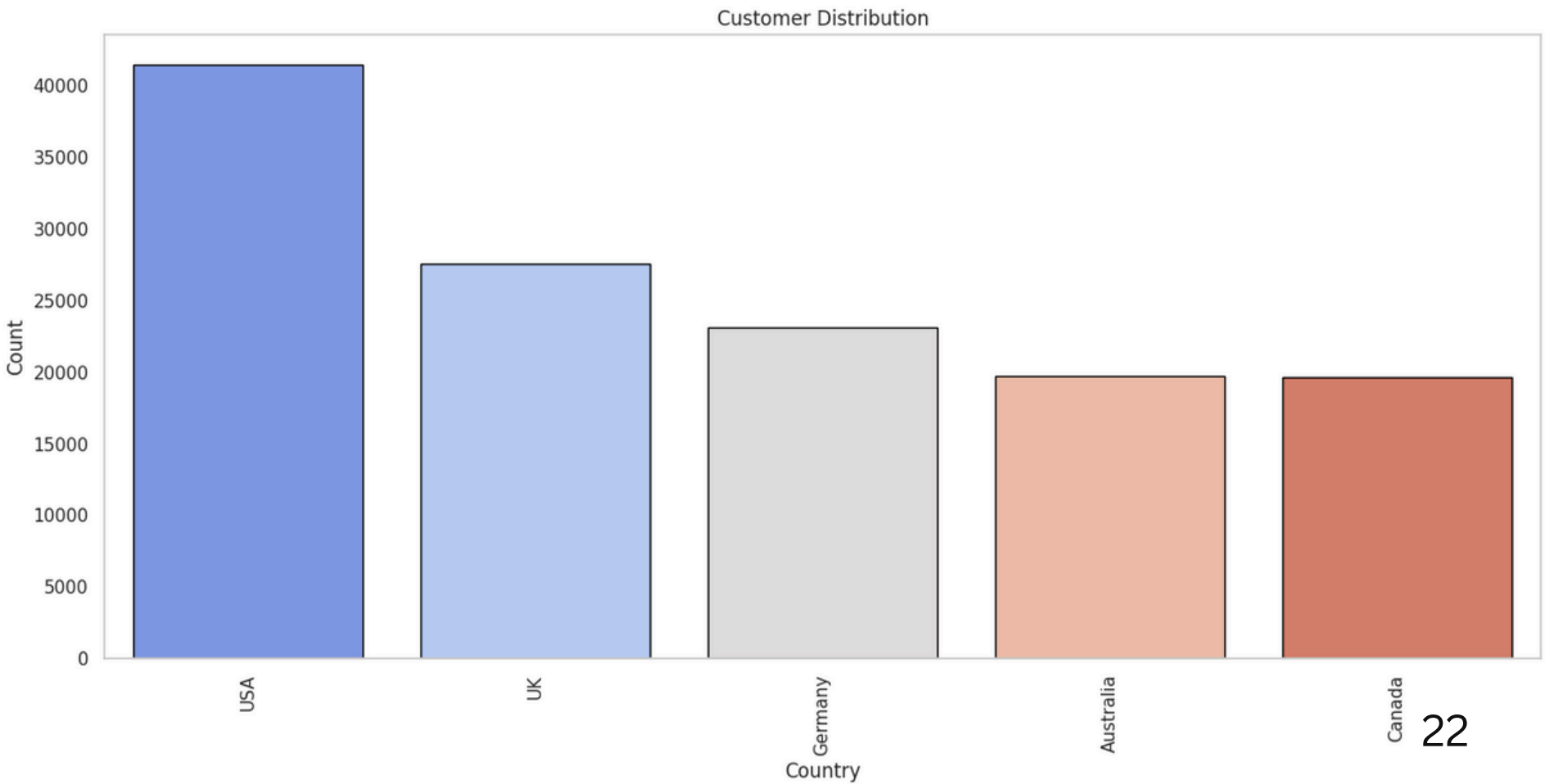# Customer Behavior Analysis

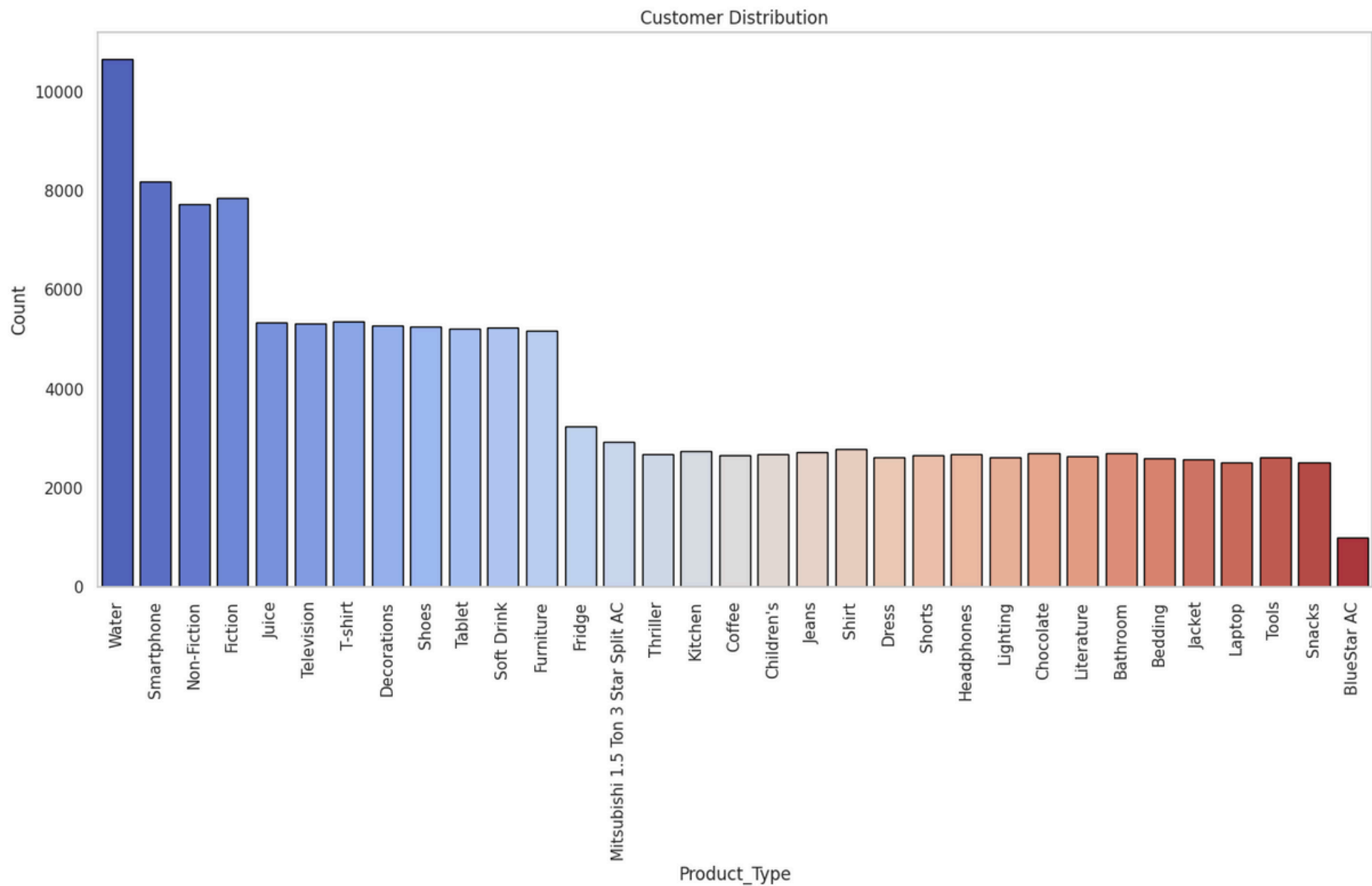Customer Distribution (Gender)

Gender Distribution:
The distribution of the high-frequency customers shows that there are more male customers; which aligns with the observation that in the distribution of total customers, males occupy larger chunk.

Country Distribution:
A significant majority of the most loyal customers are located in the USA, followed by the UK and Germany. This indicates that the USA is the most important market for repeat business.
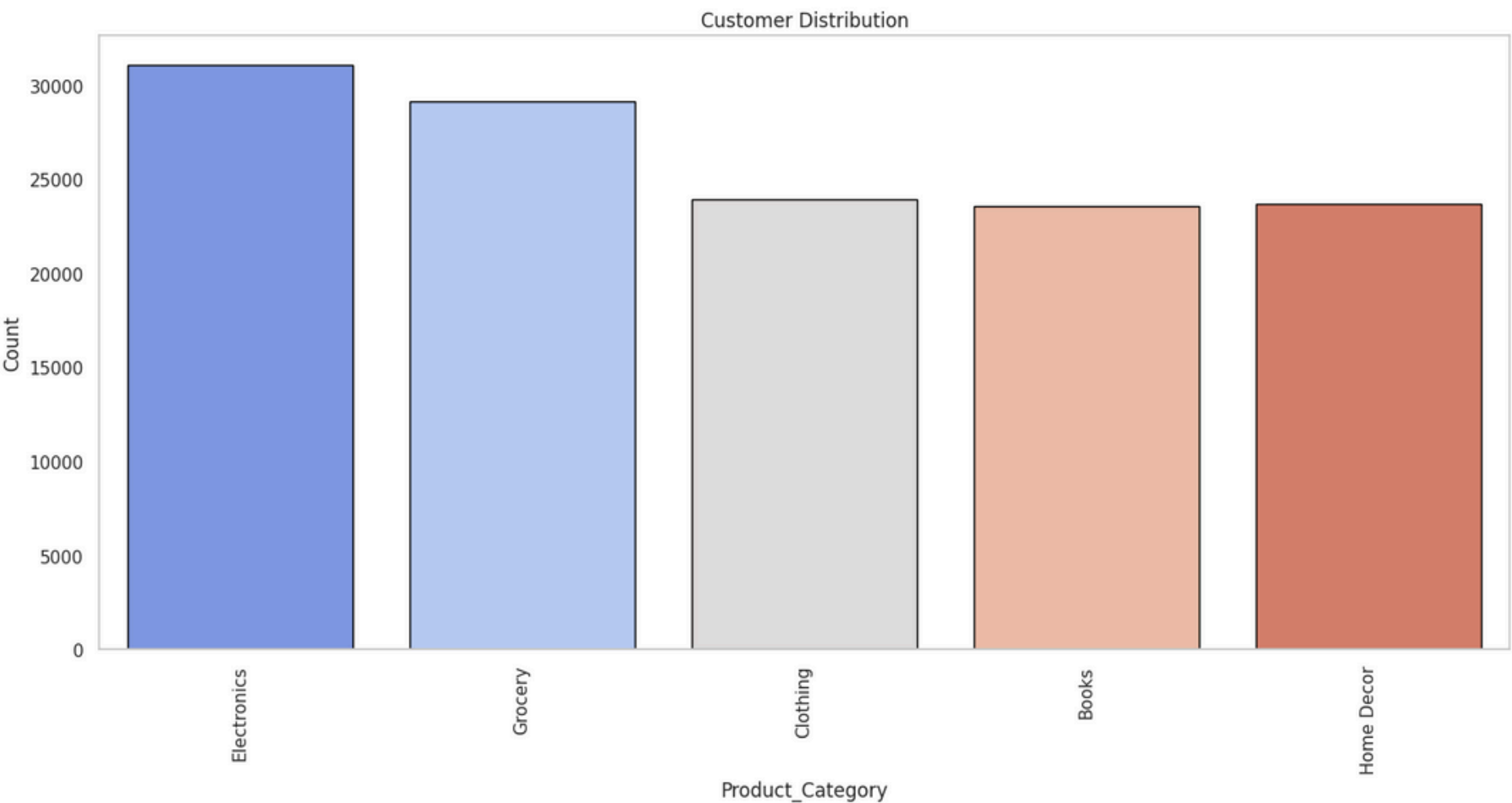

Customer Distribution (Country)

Customer Behavior Analysis
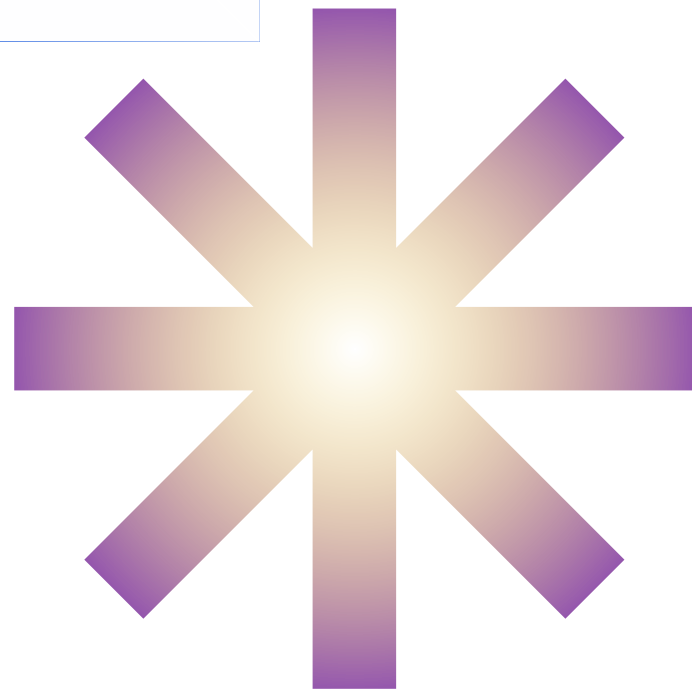

Customer Distribution

Customer Distribution

Product Type Distribution:
The most popular product types among high-frequency customers are Water, Smartphones, Non-Fiction & Fiction books.

Product Category Distribution:
Among the most loyal customers, Electronics is the most purchased product category, followed by Grocery, then Clothing, Books, and Home Decor.

23

Essential findings for boosting performance

- The USA is the core market for loyal customers and repeat business
- Regular customers drive more repeat purchases than Premium or New customers
- Medium-income groups are the most frequent buyers
- Electronics and Grocery dominate among the most loyal buyers
- Water, Smartphones, and Non-Fiction books are top product types for repeat buyers
- Most customers purchase only a few times, so loyalty programs should focus on increasing purchase frequency.

Key Takeaways: Insights for Retail Success
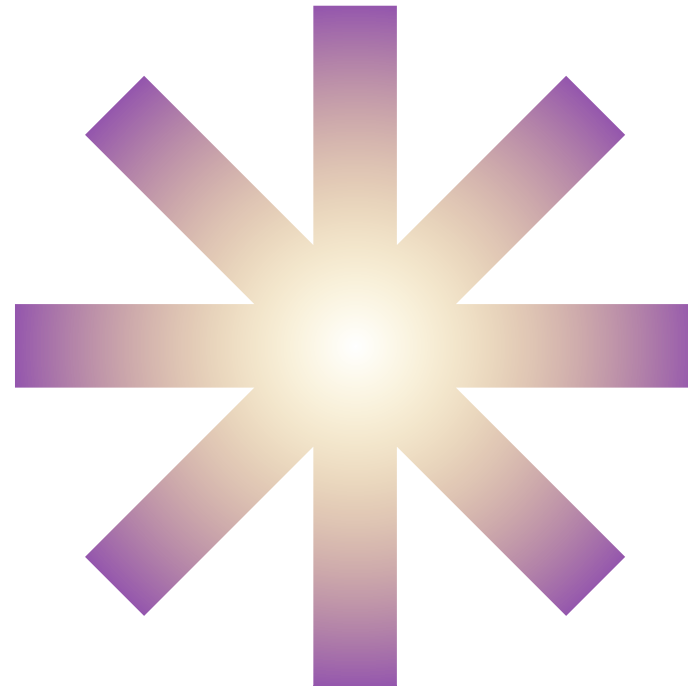
Key Takeaways: Insights for Retail Success

Essential findings for boosting performance

- A small group of highly loyal customers contributes disproportionately to sales.
- The USA leads in high-value outlier transactions across many product categories.
- Despite Electronics being strong, Groceries generate higher total sales in the USA
- Non-Fiction books are the top outlier product type in the UK
- The Pepsi brand drives the most high-value purchases in the USA
- Outliers are concentrated in Fiction, Non-Fiction, Smartphones, and Water with strong ratings.

25

Thank You