

# Exploratory Data Analysis

START

We are performing a comprehensive Exploratory Data analysis in order to uncover actionable insights on sales, customer retention and marketing strategies.



Sales Performance



Customer Retention

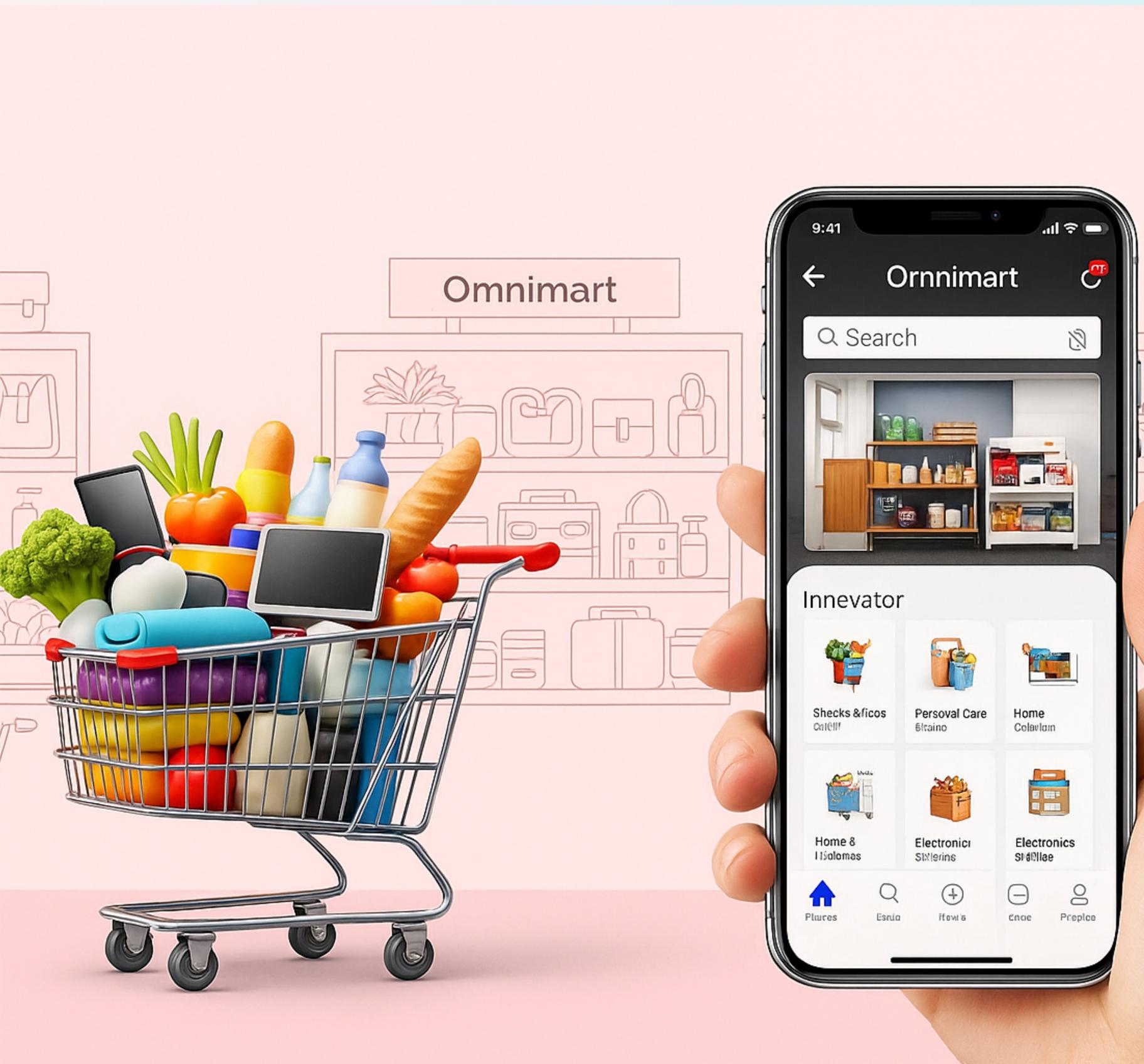


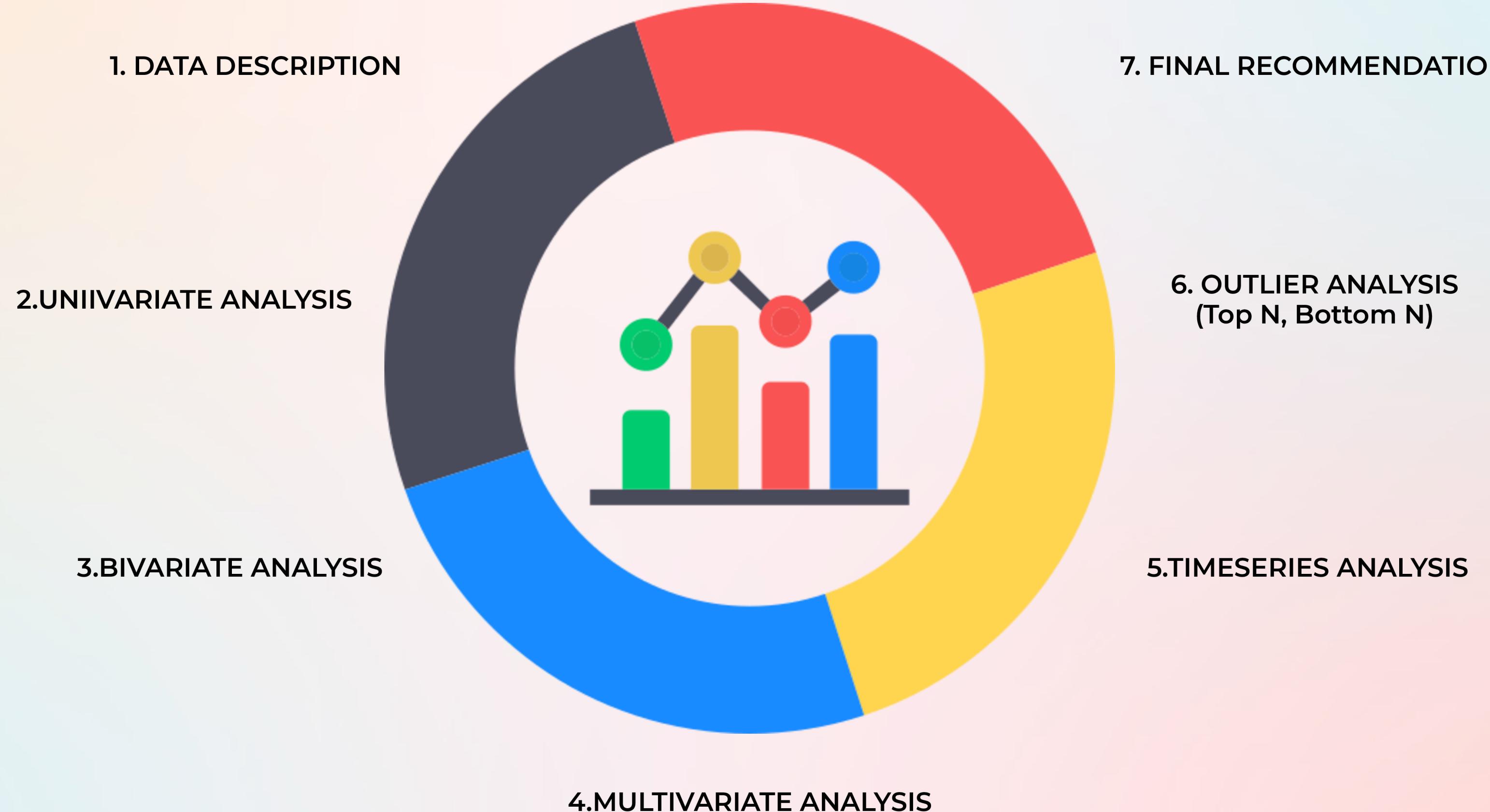
Marketing Strategies

# Problem Statement Summary:

**OmniMart Retailers** is a multinational company with a vast database of customer transactions and feedback. The company's goal is to gain a deeper understanding of its customer base to **improve sales, increase customer retention, and optimize its marketing strategies.**

Our challenge was to act as a data analyst for OmniMart. Using the provided dataset, Our team performed a comprehensive **Exploratory Data Analysis (EDA)** to uncover actionable insights. Our analysis focuses on answering **key business questions** and identifying **opportunities for growth.**





# Data Description :

RangeIndex: 302010 entries, 0 to 302009

Data columns (total 30 columns):

#	Column	Non-Null Count	Dtype
0	Transaction_ID	301677	non-null float64
1	Customer_ID	301702	non-null float64
2	Name	301628	non-null object
3	Email	301663	non-null object
4	Phone	301648	non-null float64
5	Address	301695	non-null object
6	City	301762	non-null object
7	State	301729	non-null object
8	Zipcode	301670	non-null float64
9	Country	301739	non-null object
10	Age	301837	non-null float64
11	Gender	301693	non-null object
12	Income	301720	non-null object
13	Customer_Segment	301795	non-null object
14	Date	301651	non-null object
15	Year	301660	non-null float64
16	Month	301737	non-null object
17	Time	301660	non-null object
18	Total_Purchases	301649	non-null float64
19	Amount	301653	non-null float64
20	Total_Amount	301660	non-null float64
21	Product_Category	301727	non-null object
22	Product_Brand	301729	non-null object
23	Product_Type	302010	non-null object
24	Feedback	301826	non-null object
25	Shipping_Method	301673	non-null object
26	Payment_Method	301713	non-null object
27	Order_Status	301775	non-null object
28	Ratings	301826	non-null float64
29	products	302010	non-null object
dtypes: float64(10), object(20)			

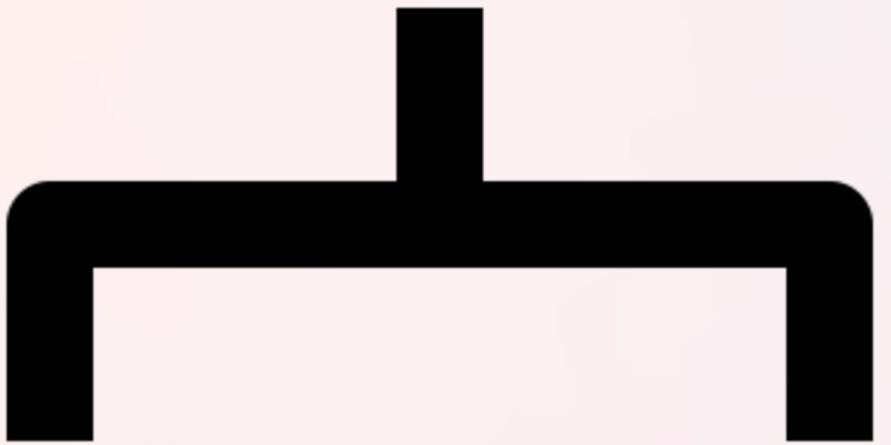
	Transaction_ID	Customer_ID	Phone	Zipcode	Age	Year	Total_Purchases	Amount	Total_Amount	Ratings
count	3.016770e+05	301702.000000	3.016480e+05	301670.000000	301837.000000	301660.000000	301649.000000	301653.000000	301660.000000	301826.000000
mean	5.495823e+06	55006.553934	5.501464e+09	50298.951019	35.481326	2023.165113	5.359729	255.163659	1367.651156	3.162670
std	2.595565e+06	26005.675200	2.596017e+09	28972.807134	15.021933	0.371283	2.868575	141.389640	1128.998515	1.320827
min	1.000007e+06	10000.000000	1.000049e+09	501.000000	18.000000	2023.000000	1.000000	10.000219	10.003750	1.000000
25%	3.247930e+06	32469.250000	3.255061e+09	25425.000000	22.000000	2023.000000	3.000000	132.890764	438.724278	2.000000
50%	5.499657e+06	55012.000000	5.505812e+09	50602.500000	32.000000	2023.000000	5.000000	255.470969	1041.117547	3.000000
75%	7.739509e+06	77511.000000	7.749860e+09	75252.000000	46.000000	2023.000000	8.000000	377.672606	2029.999853	4.000000
max	9.999995e+06	99999.000000	9.999996e+09	99949.000000	70.000000	2024.000000	10.000000	499.997911	4999.625796	5.000000

	Transaction_ID	Customer_ID	Name	Email	Phone	Address	City	State	Zipcode	Country	...	Total_Amount	Product_Category	Product_Brand	Product_Type	Feedback	Shi
0	8691788.0	37249.0	Michelle Harrington	Ebony39@gmail.com	1.414787e+09	Amanda Burgs	Dortmund	Berlin	77985.0	Germany	...	324.086270	Clothing	Nike	Shorts	Excellent	
1	2174773.0	69749.0	Kelsey Hill	Mark36@gmail.com	6.852900e+09	Dawn Centers	Nottingham	England	99071.0	UK	...	806.707815	Electronics	Samsung	Tablet	Excellent	
2	6679610.0	30192.0	Scott Jensen	Shane85@gmail.com	8.362160e+09	Young Canyon	Geelong	New South Wales	75929.0	Australia	...	1063.432799	Books	Penguin Books	Children's	Average	
3	7232460.0	62101.0	Joseph Miller	Mary34@gmail.com	2.776752e+09	Thomas Creek Suite 100	Edmonton	Ontario	88420.0	Canada	...	2466.854021	Home Decor	Home Depot	Tools	Excellent	
4	4983775.0	27901.0	Debra Coleman	Charles30@gmail.com	9.098268e+09	Lori Ports Suite 269	Bristol	England	48704.0	UK	...	248.553049	Grocery	Nestle	Chocolate	Bad	

5 rows × 30 columns

# Variables



## Numerical

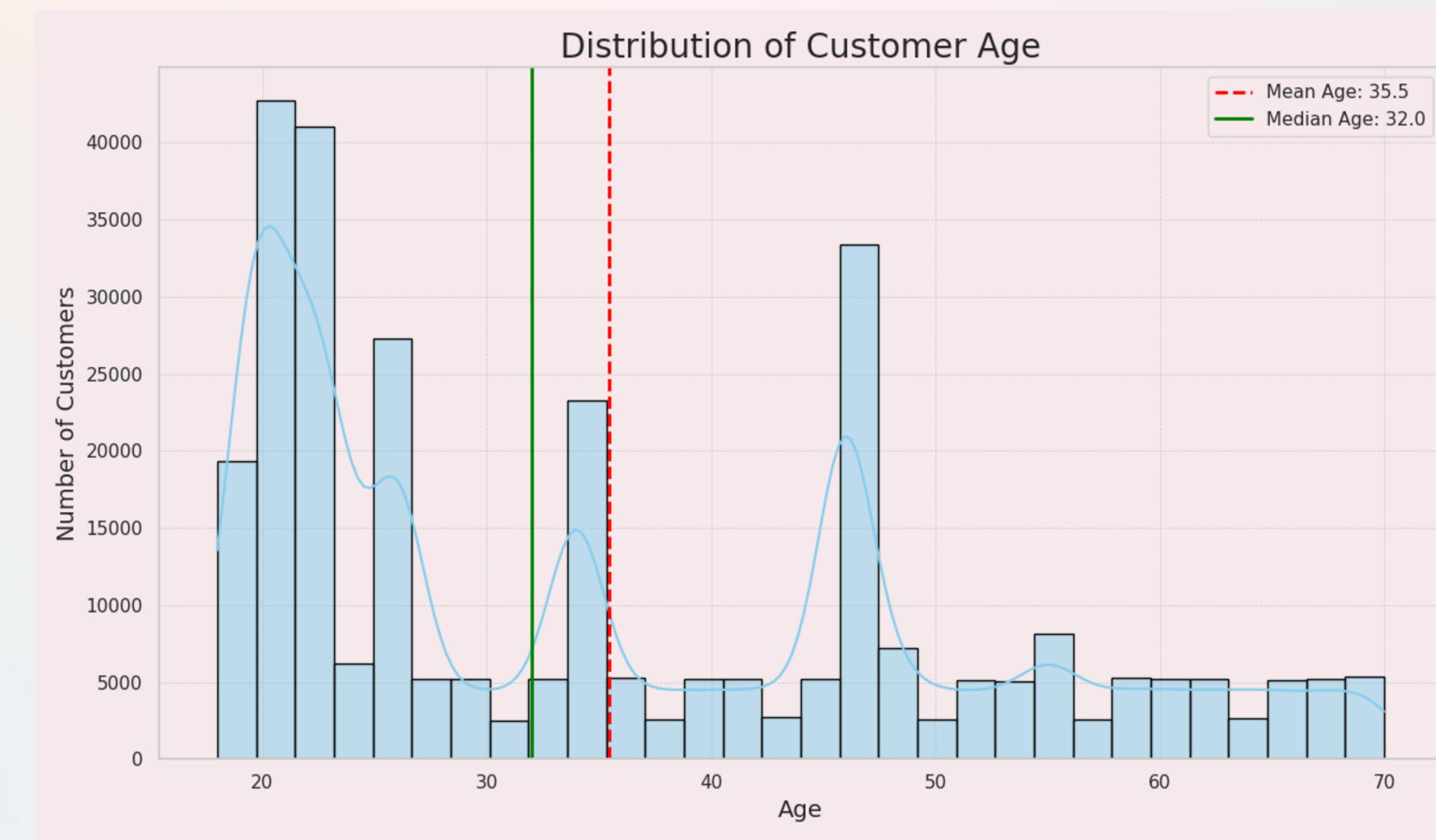
Age, Year, Income,  
Total\_Purchases,  
Amount, Total\_Amount,  
Ratings, Phone no. ,  
Year.

## Categorical

Name, City, State,  
Country, Gender,  
Customer\_Segment  
Product\_Category,  
Product\_Brand,  
Product\_type, Feedback,  
Shipping\_Method,  
Payment\_Method,  
products

# Univariate Analysis

Univariate analysis is the simplest form of data analysis, where you look at only one variable at a time. The goal is to understand the distribution, central tendency, and spread of that single variable, without worrying about its relationship with others.

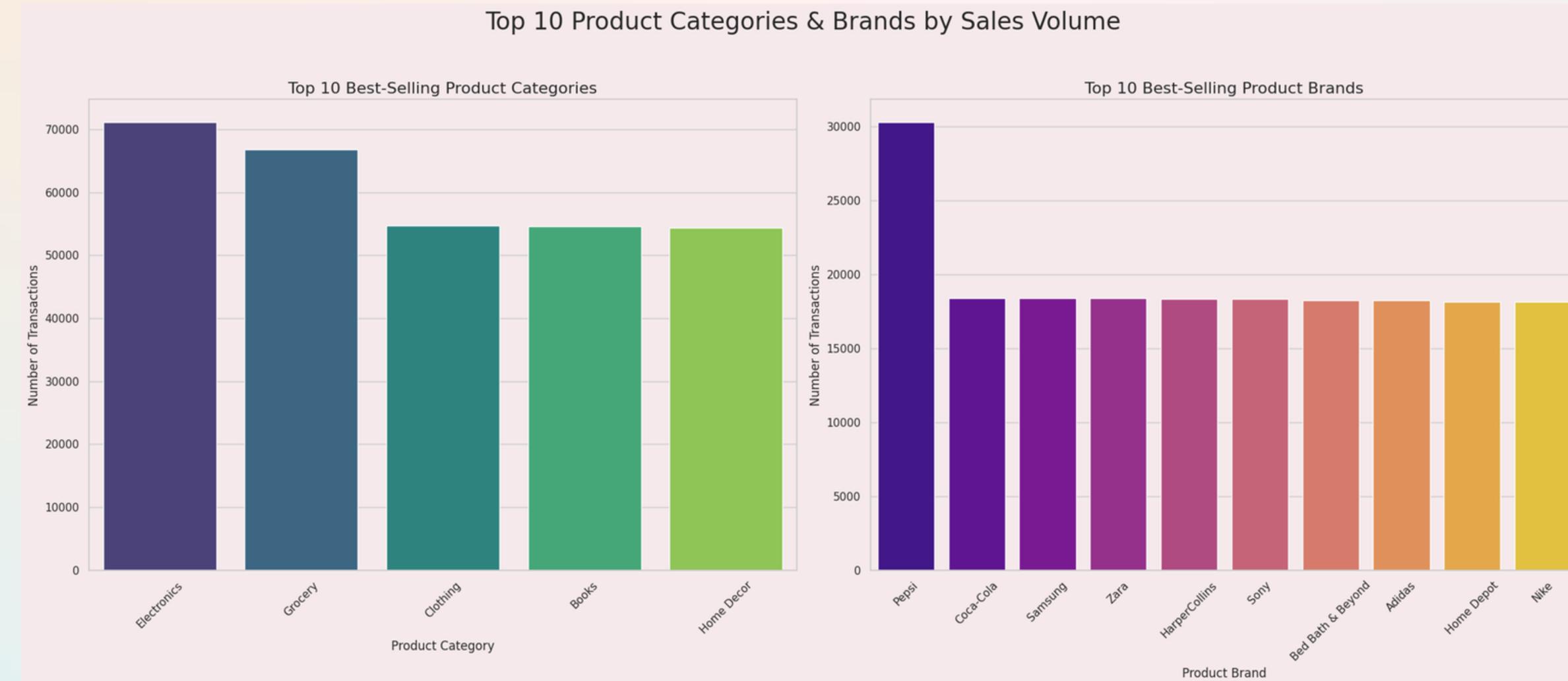


## Distribution of Numerical Variables



- The customer age distribution skews toward younger and middle-aged groups, with a mean around 35 years.
- Transaction amounts are spread fairly evenly, averaging about Rs. 255 per transaction.
- Purchases per customer cluster mostly around 5, showing consistency in buying behavior.
- Customer ratings lean toward the positive side, with 4 being the most common rating, indicating generally satisfied customers.

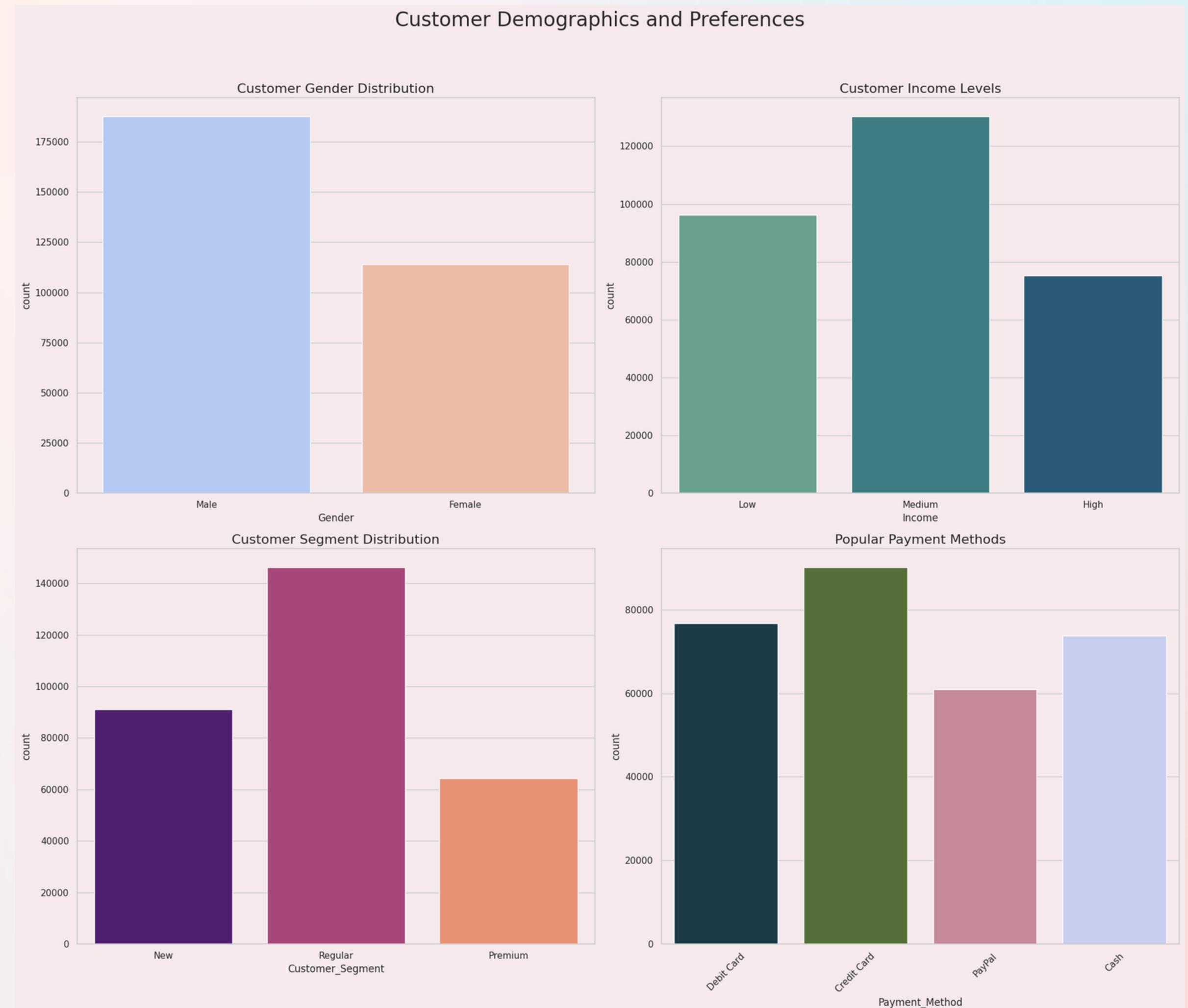
Electronics and Grocery dominate as the top product categories by transaction volume, showing that both high-value durable goods and everyday essentials are driving much of the sales activity, while categories like Clothing, Books, and Home Décor also perform strongly but at a slightly lower level.



On the brand side, Pepsi stands out with a significant lead in sales compared to other brands, while Coca-Cola, Samsung, Zara, Harper Collins, Sony, Bed Bath & Beyond, Adidas, Home Depot, and Nike all show relatively similar transaction counts, suggesting a competitive balance across these major players.

The mix of top brands across beverages, electronics, fashion, publishing, and home goods reflects diverse consumer demand, but the clear dominance of a few names (like Pepsi and Electronics overall) indicates where brand loyalty and category preference are strongest.

- The gender distribution shows a clear majority of male customers compared to female, highlighting a skew in the customer base.
- In terms of income levels, medium-income customers dominate, low-income customers also contribute strongly, while high-income customers represent the smallest portion of the market.
- Looking at customer segments, regular customers form the backbone of the business, new customers represent a healthy inflow, and premium customers make up a much smaller but potentially high-value segment.
- For payment preferences, credit cards are the most common method, debit cards and cash are also popular, and PayPal is the least frequently used, suggesting limited adoption of digital wallets.



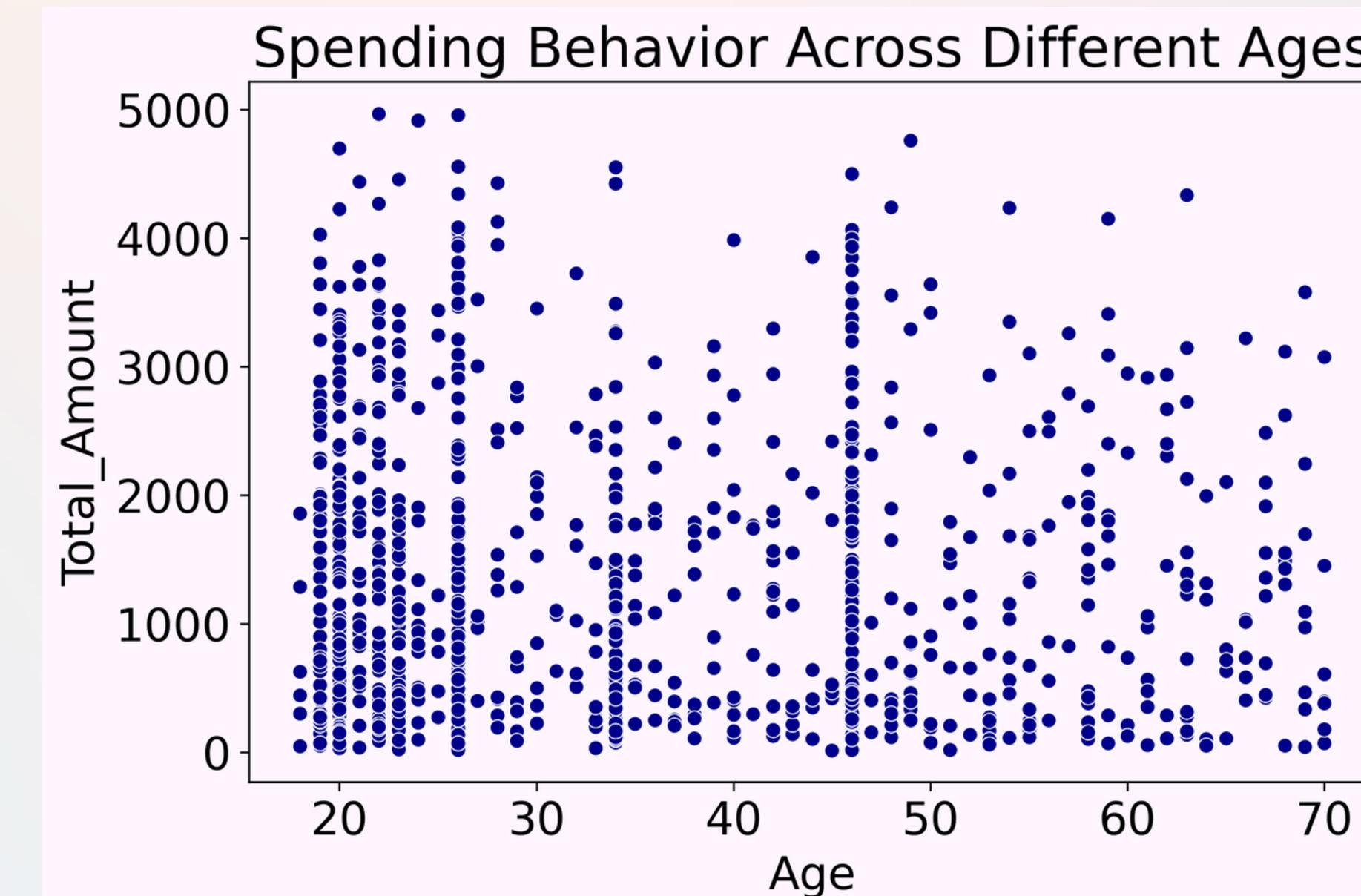
## **Key Insights and Implementations:-**

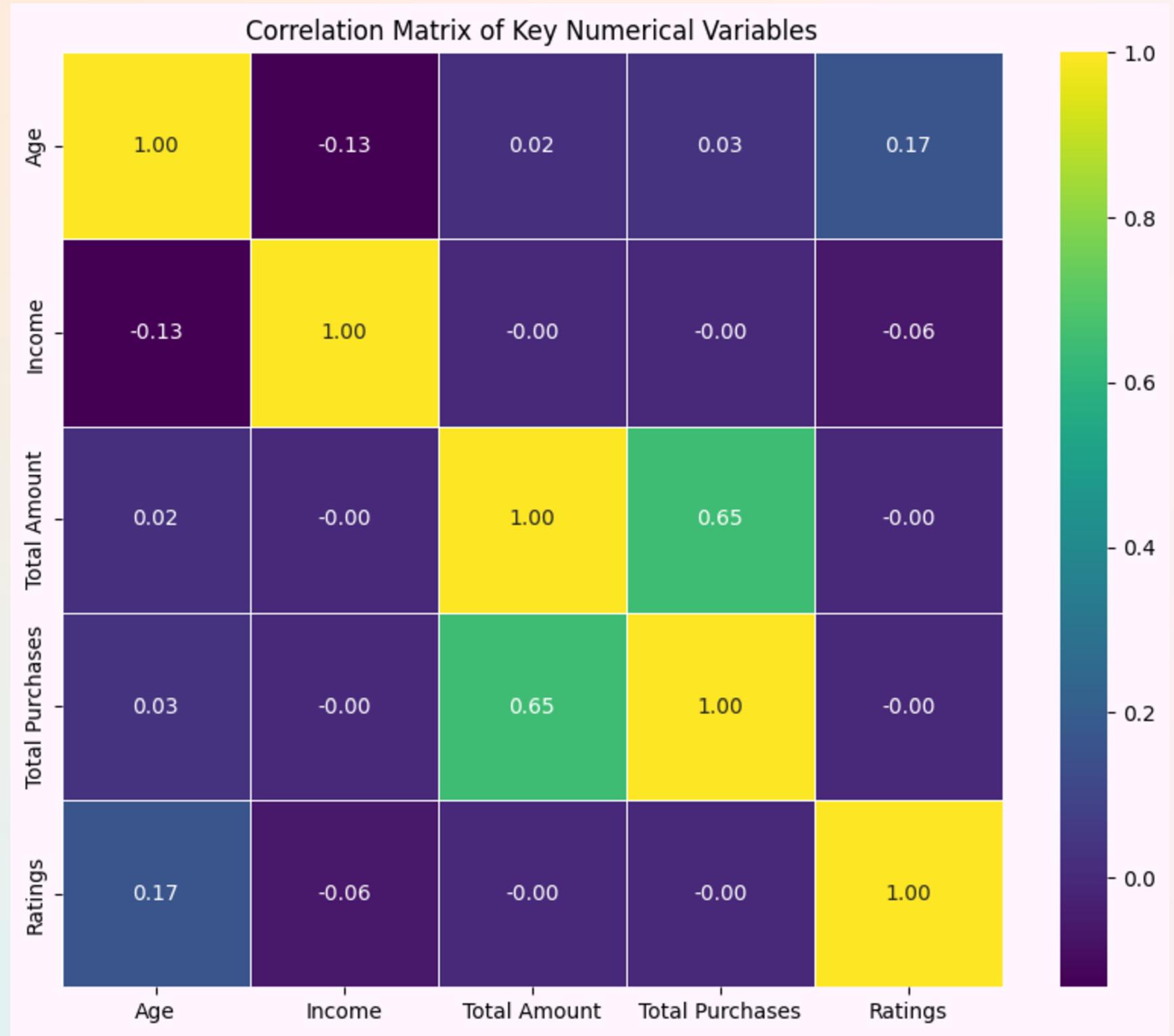
This initial analysis provides a clear snapshot of OmniMart's core customer and their buying habits. Our customer base is primarily young, financially moderate, and digitally native. They are highly satisfied with their purchases, giving overwhelmingly positive ratings. The business is heavily driven by the Electronics and Clothing categories, with iconic brands like Apple and Nike leading sales. Our strategic focus should be on deepening engagement with this key demographic, increasing average order value, and leveraging our high customer satisfaction as a competitive advantage.

- **Youthful, Medium-Income Base:** The majority of customers are 18–35, medium-income, and spend ~Rs.255 per transaction → focus marketing on youth-oriented platforms, trend-driven products, and flexible payment options like BNPL.
- **Electronics & Clothing Lead Sales:** Electronics and clothing dominate with strong loyalty to Apple, Samsung, Nike, and Adidas → maintain high stock levels, strengthen partnerships with these brands, and use bundles or cross-sell to boost order value.
- **High Customer Satisfaction:** Ratings cluster at 3–4 stars, showing trust and positive experiences → highlight reviews in marketing, promote “top-rated” sections, and resolve low ratings quickly to maintain brand image.
- **Digital Payments Dominate:** Credit/debit cards lead but cash and PayPal remain relevant → ensure smooth digital checkout, encourage wallet adoption, and keep multiple options to reduce cart abandonment.

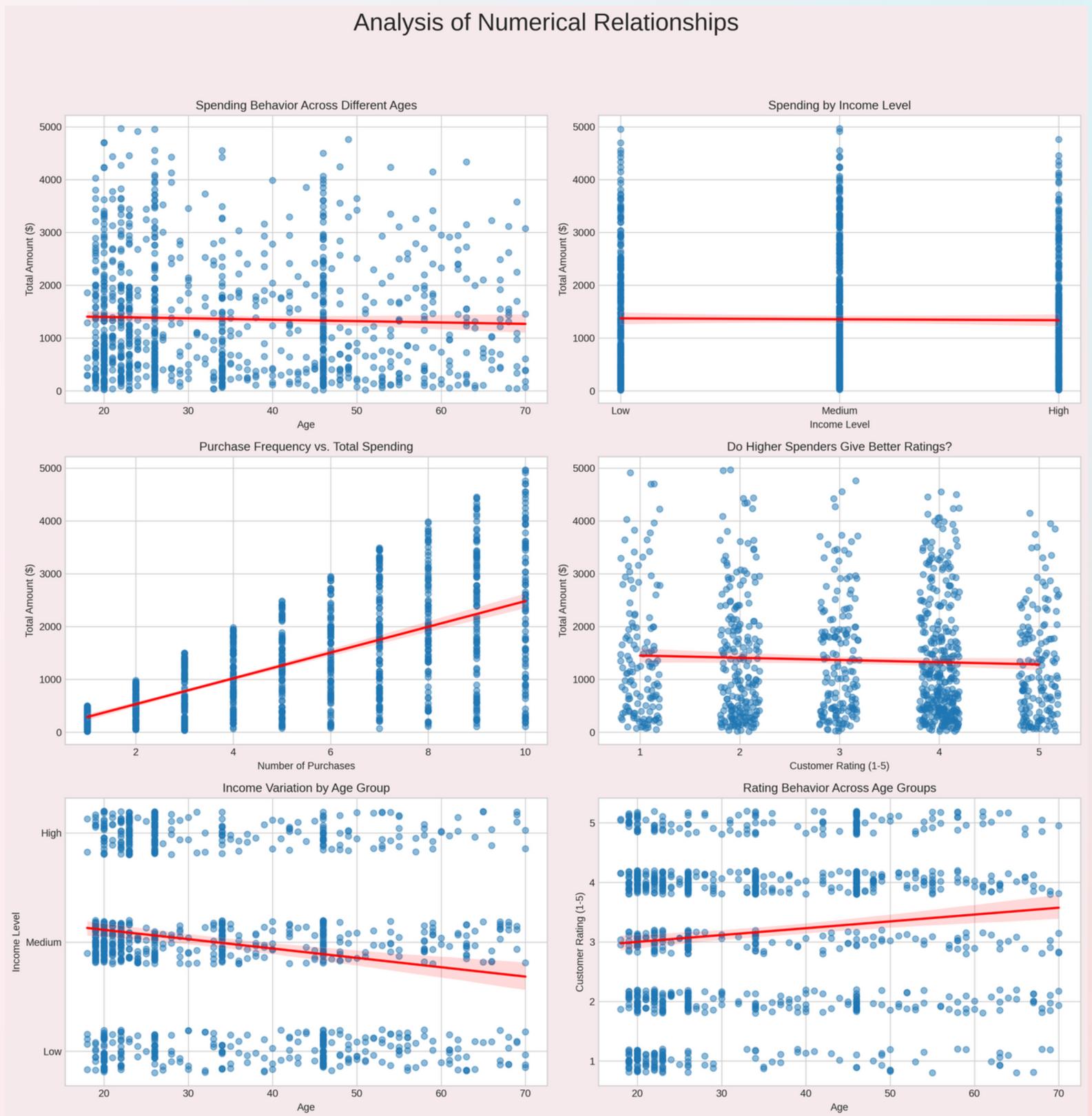
# *Bivariate Analysis*

Bivariate analysis is the study of the relationship between two variables at the same time. It helps identify correlations, associations, or dependencies between them. For example, analyzing how age affects spending or how payment method varies by customer segment.



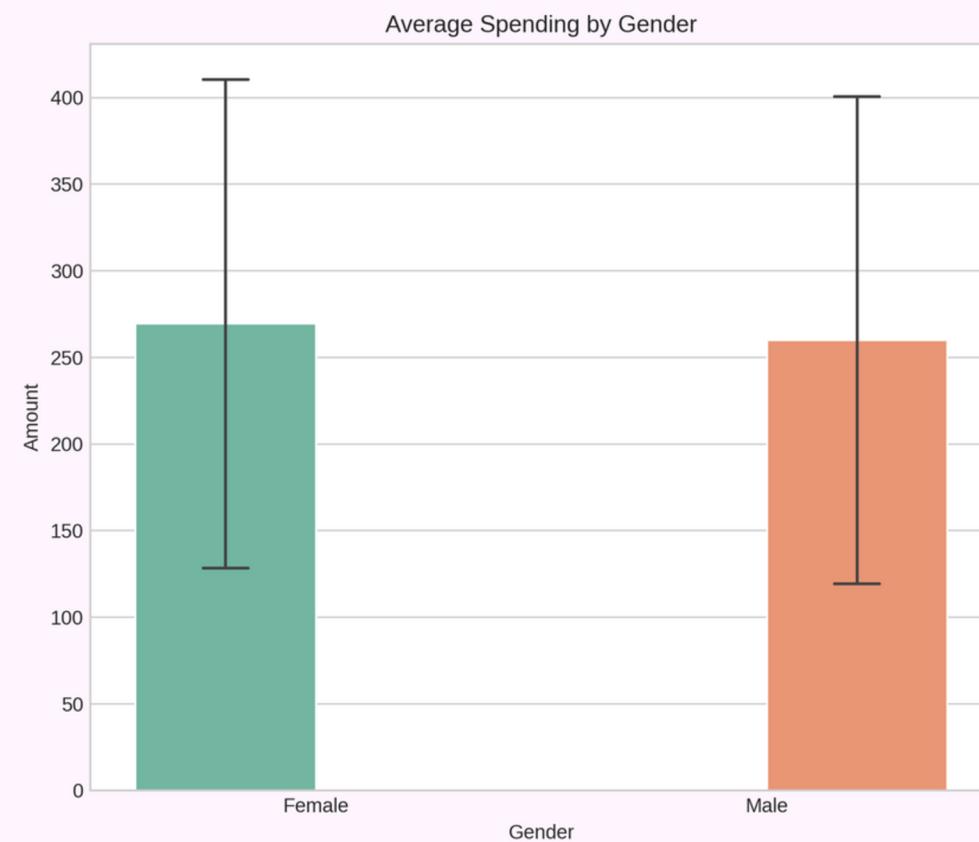
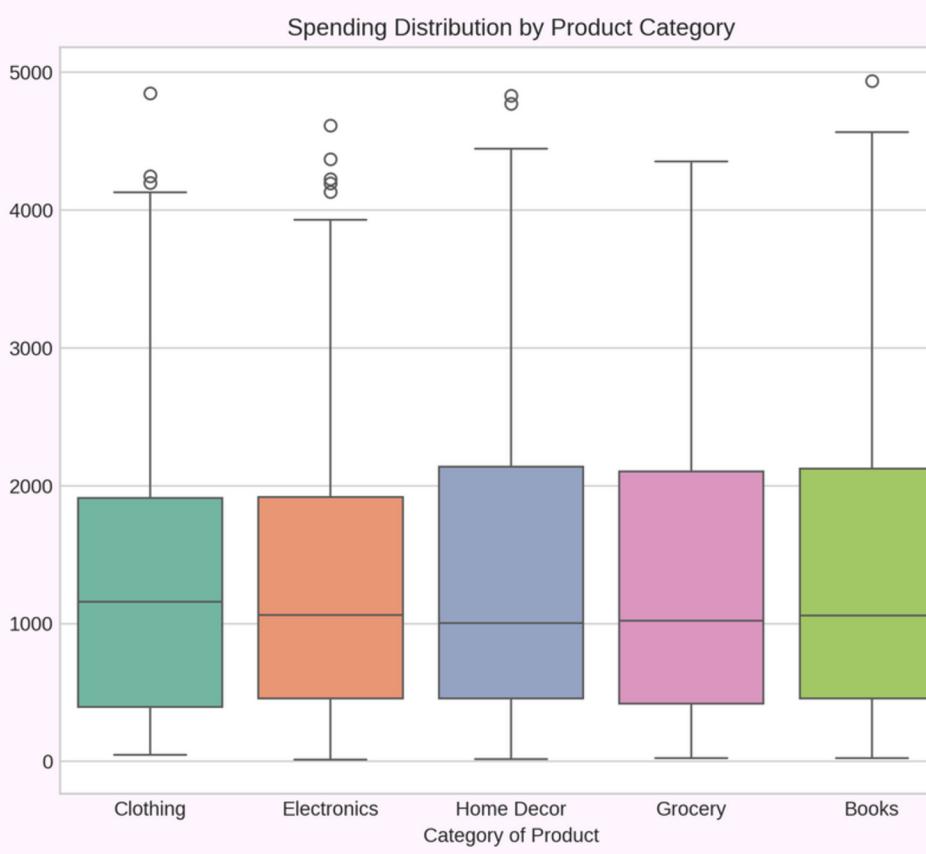
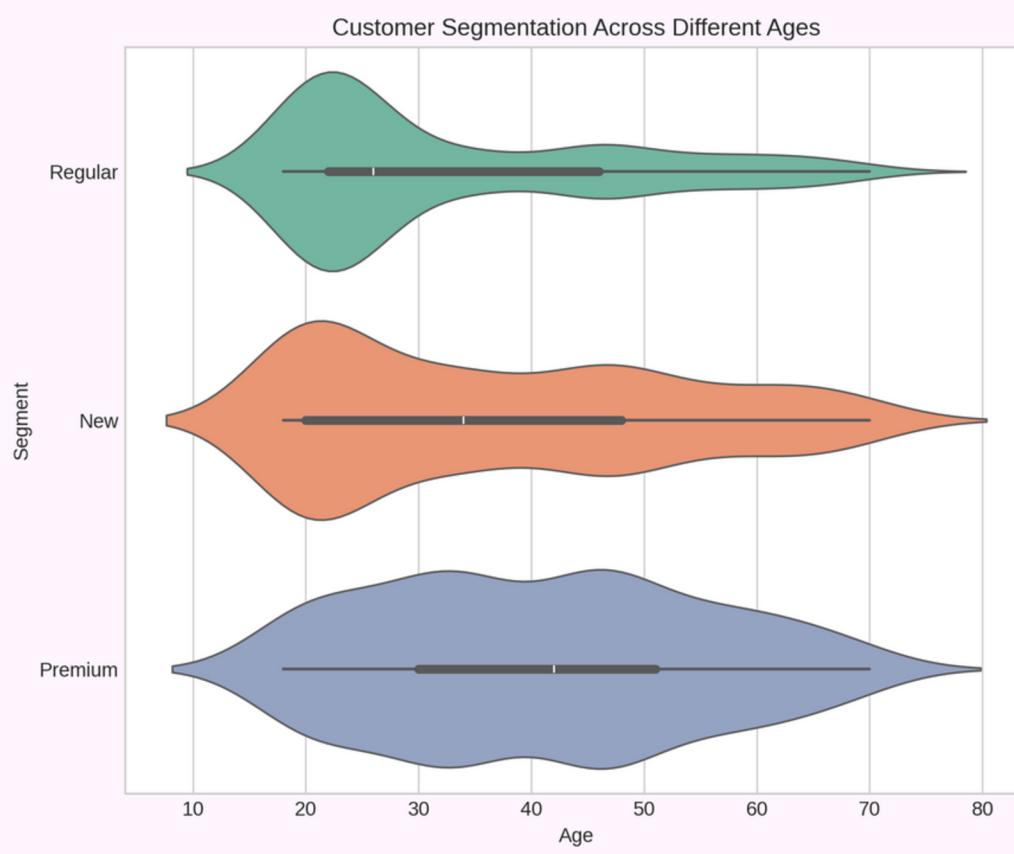


Customers who purchase more tend to spend more overall, while age has only a slight influence on ratings and income shows little impact on spending or satisfaction.



Customers' spending is driven more by purchase frequency than age or income, with ratings showing weak links to spending—so focusing on encouraging repeat purchases through loyalty programs and targeted promotions is likely the most effective strategy.

## Numerical vs Categorical analysis



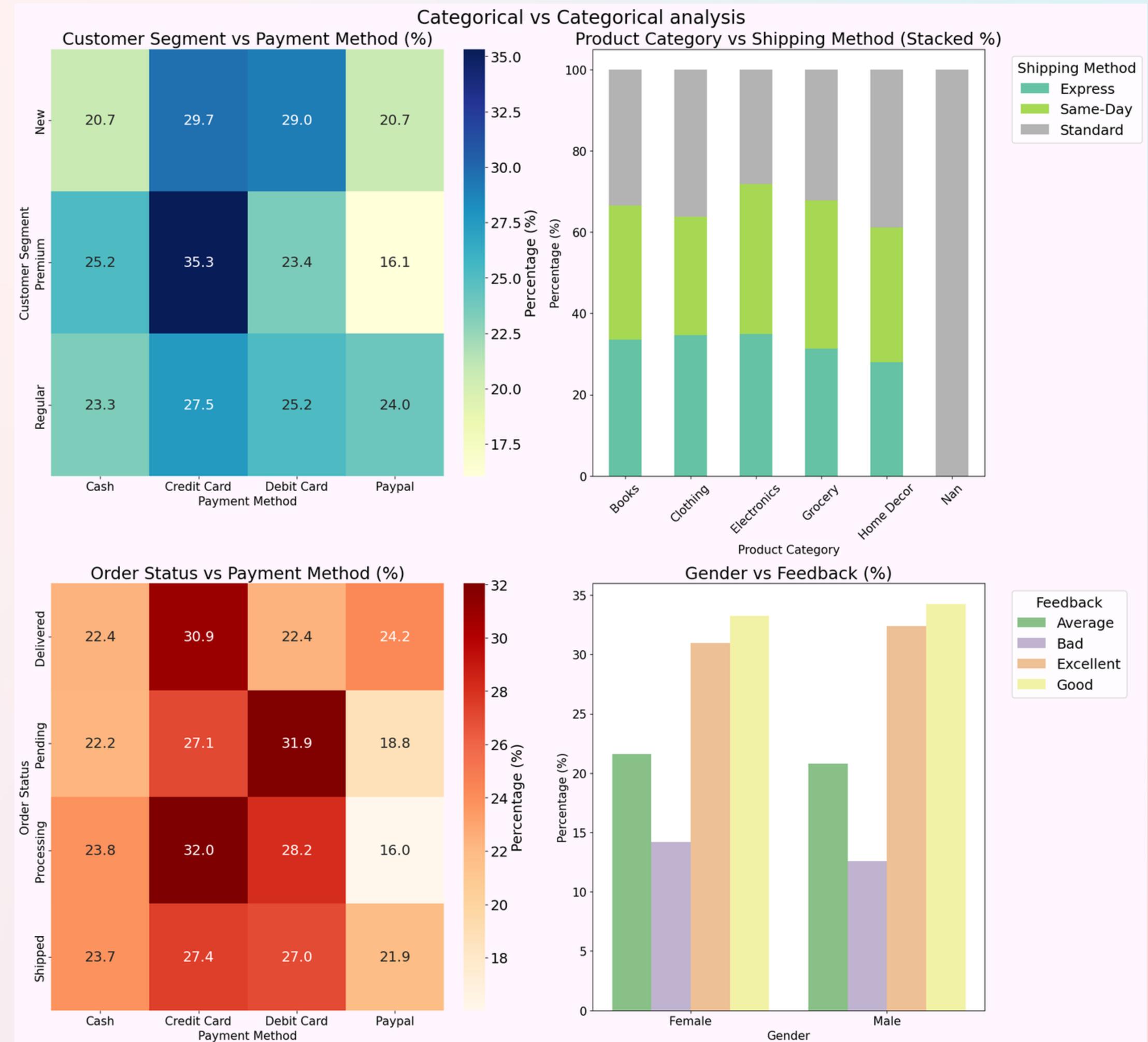
- Customer segmentation by age shows that younger customers dominate the "New" and "Regular" segments, while "Premium" customers are more evenly spread across middle and older ages.
- Spending distribution by product category indicates relatively similar spending levels across categories, with Electronics and Books showing slightly higher upper outliers.
- Average spending by gender is nearly the same for males and females, with only marginal differences and overlapping variability.

- Customer Segment vs Payment Method: Premium customers prefer credit cards, while New customers are more evenly distributed across methods, and PayPal is the least used overall.

- Product Category vs Shipping Method: Express and Same-Day shipping are more common for Electronics and Clothing, while Standard dominates in Books and Grocery.

- Order Status vs Payment Method: Credit card and debit card payments are slightly more associated with completed deliveries and processing, while PayPal has lower representation across all statuses.

- Gender vs Feedback: Both genders mostly give positive feedback, with males showing a slightly higher proportion of “Excellent” ratings than females.



# **Key Insights:**

- Spending Behavior: Spending is not strongly tied to age or income but correlates positively with the number of purchases, indicating frequency is a stronger driver of total revenue.
- Customer Segmentation: Premium customers are more age-diverse and show stronger ties to credit card usage, while younger customers are more concentrated in New/Regular segments.
- Product & Payment Dynamics: Electronics and Clothing tend to attract express shipping, and credit/debit cards dominate across customer and order behaviors.
- Gender Trends: Spending levels are nearly identical between males and females, but men give slightly more favorable feedback.

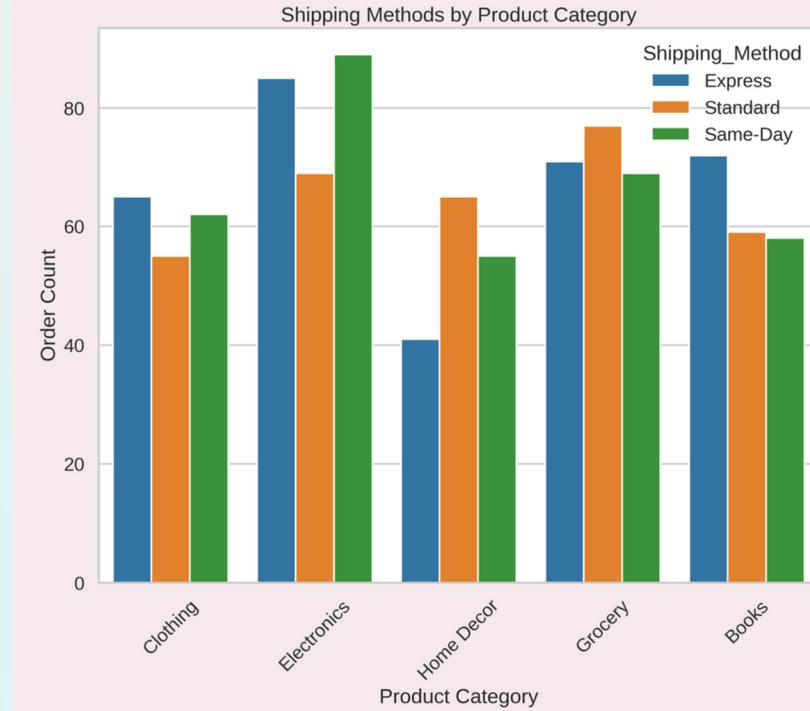
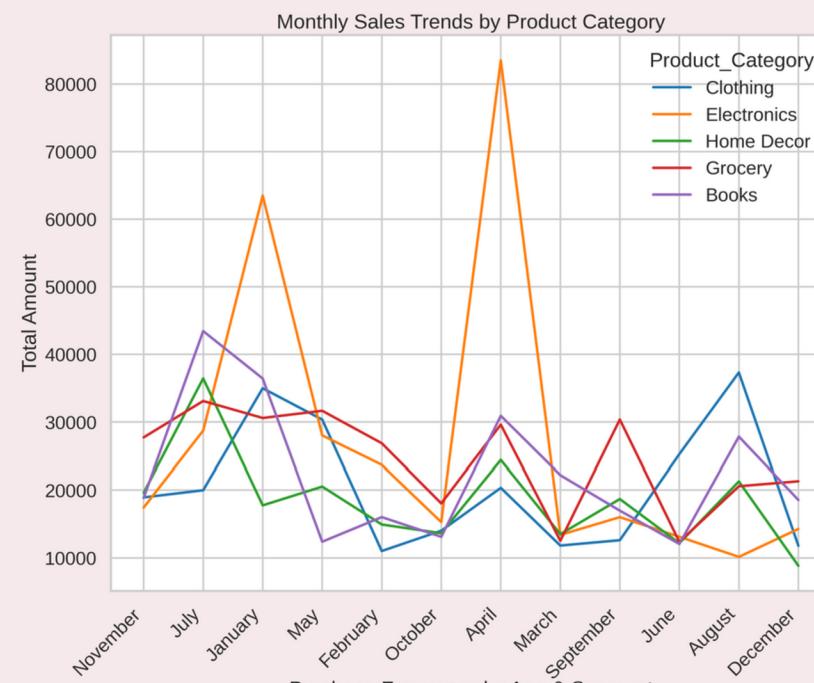
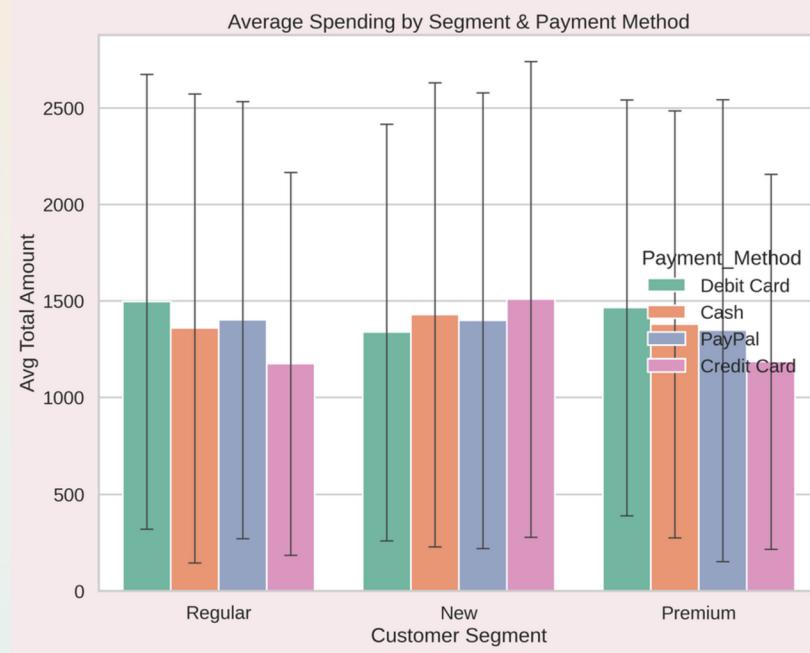
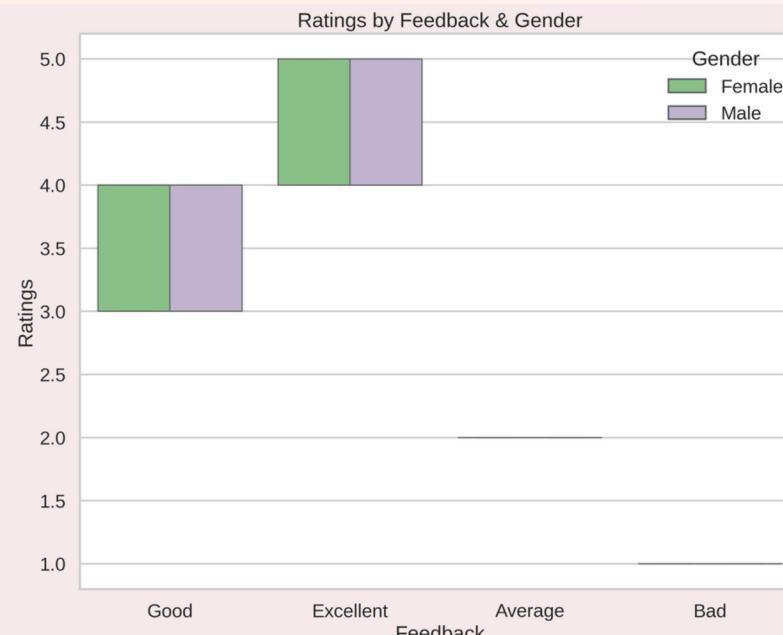
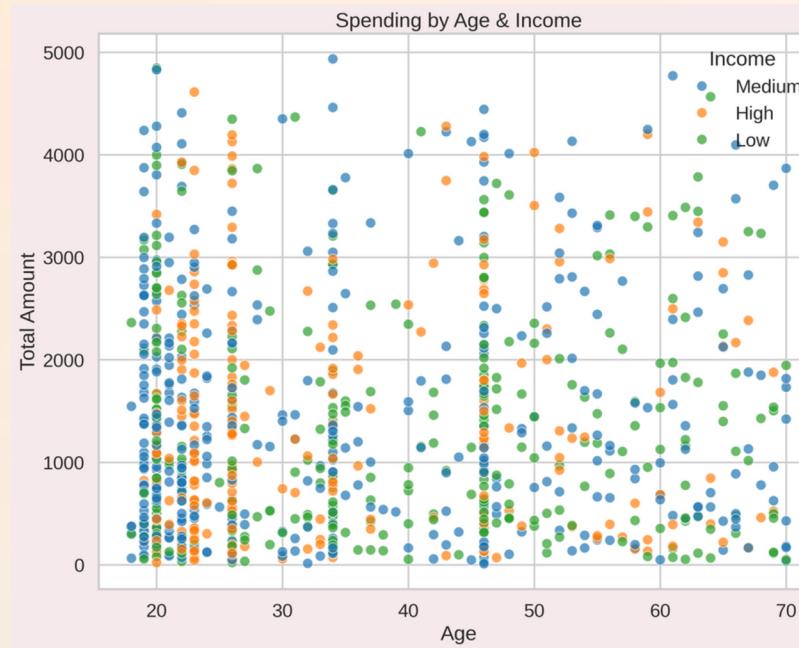
# **Implementations:**

- Target frequent purchasers with loyalty rewards, since purchase frequency drives revenue more than demographics.
- Enhance premium segment value by offering exclusive benefits tied to preferred credit card usage.
- Optimize shipping by promoting express/same-day options for high-demand categories like Electronics and Clothing.
- Leverage feedback trends by tailoring marketing campaigns that highlight high satisfaction among male customers while addressing areas of dissatisfaction for female customers.

# Multivariate Analysis

Multivariate Analysis means analyzing the relationship between three or more variables at the same time to understand patterns, interactions, and combined effects.

Where univariate looks at one variable and bivariate looks at two, multivariate goes deeper by studying how multiple variables(three or more) together influence outcomes.

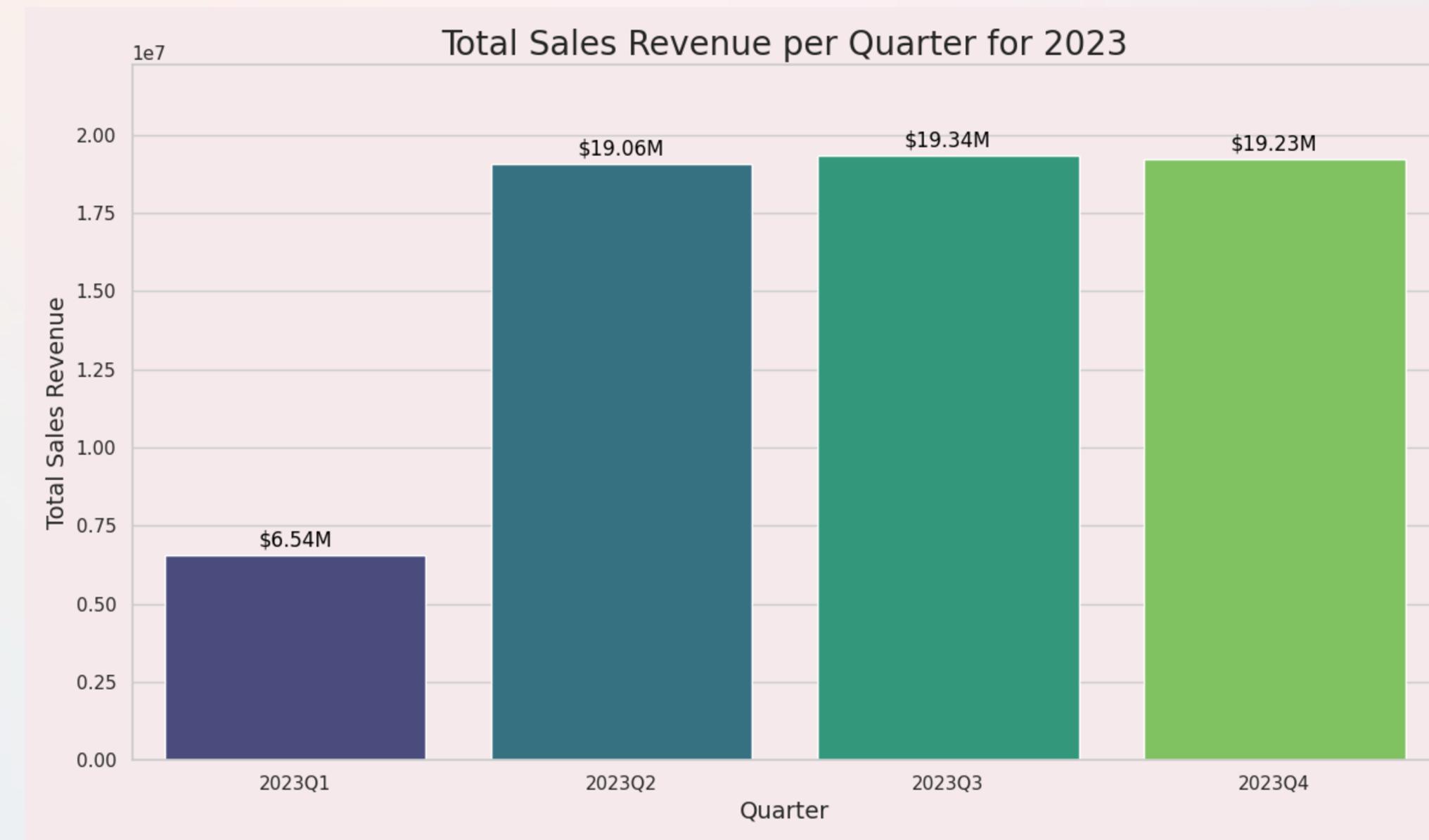


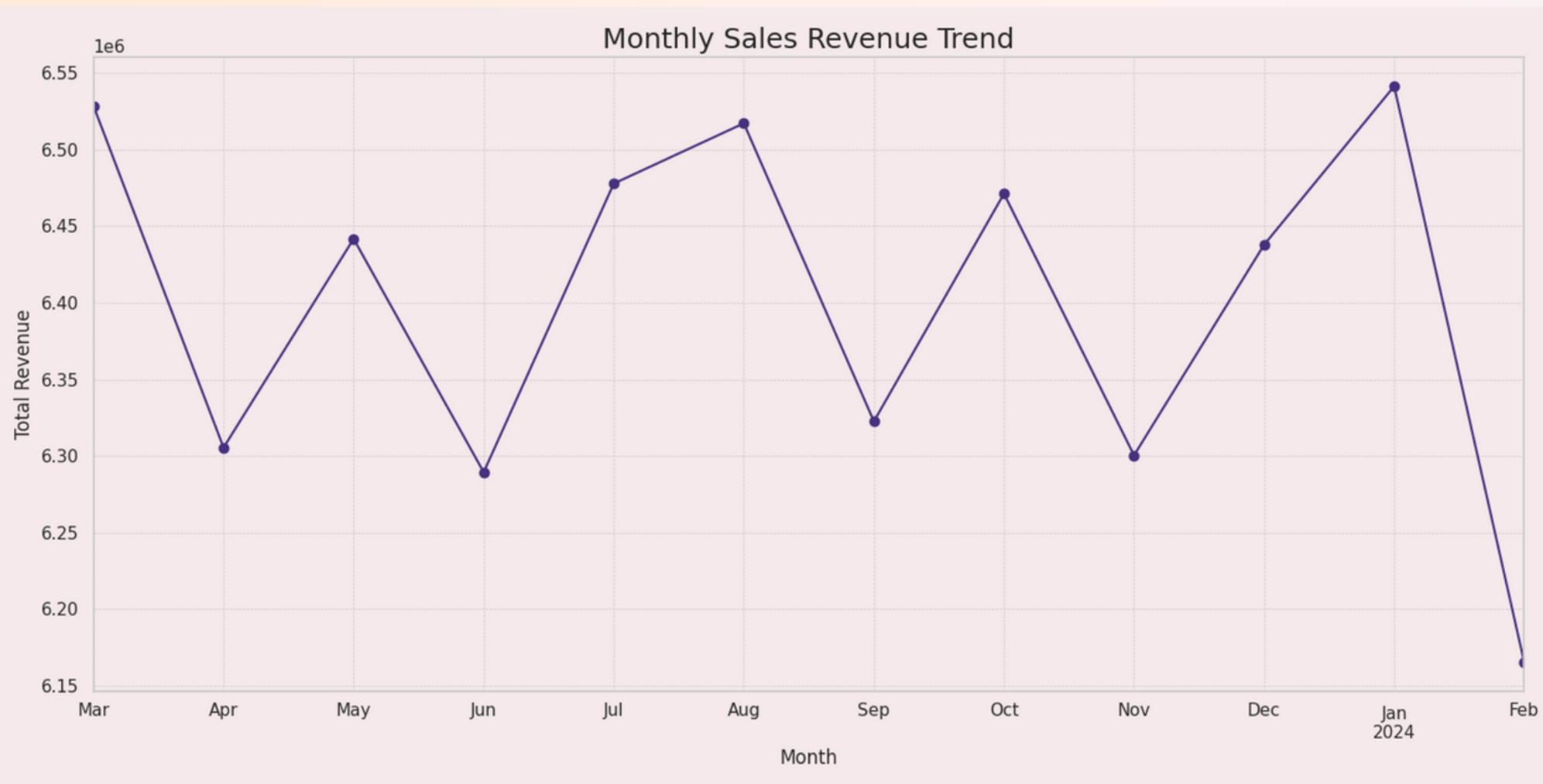
## Key Insights and Implementations:

- ◆ Demographics vs. Behavior: Age/income weak predictors of spending → shift to behavioral segmentation (frequency, product category, payment/shipping).
- ◆ Payment & Shipping: Card payments = higher spend → incentivize with cashback/loyalty. Promote express/same-day shipping with bundled offers.
- ◆ Product & Seasonality: Electronics & Clothing drive peaks → align inventory & promos with seasonal trends. Push cross-sells (e.g., Clothing + Accessories).
- ◆ Customer Segments: Premium customers underperform → revamp perks (bundles, free express shipping, tiers). Target New/Regular customers with onboarding & upgrade pathways.
- ◆ Experience & Feedback: Ratings consistent across genders → universal focus on delivery, product quality, returns. Personalize outreach: discounts for detractors, referrals for promoters.

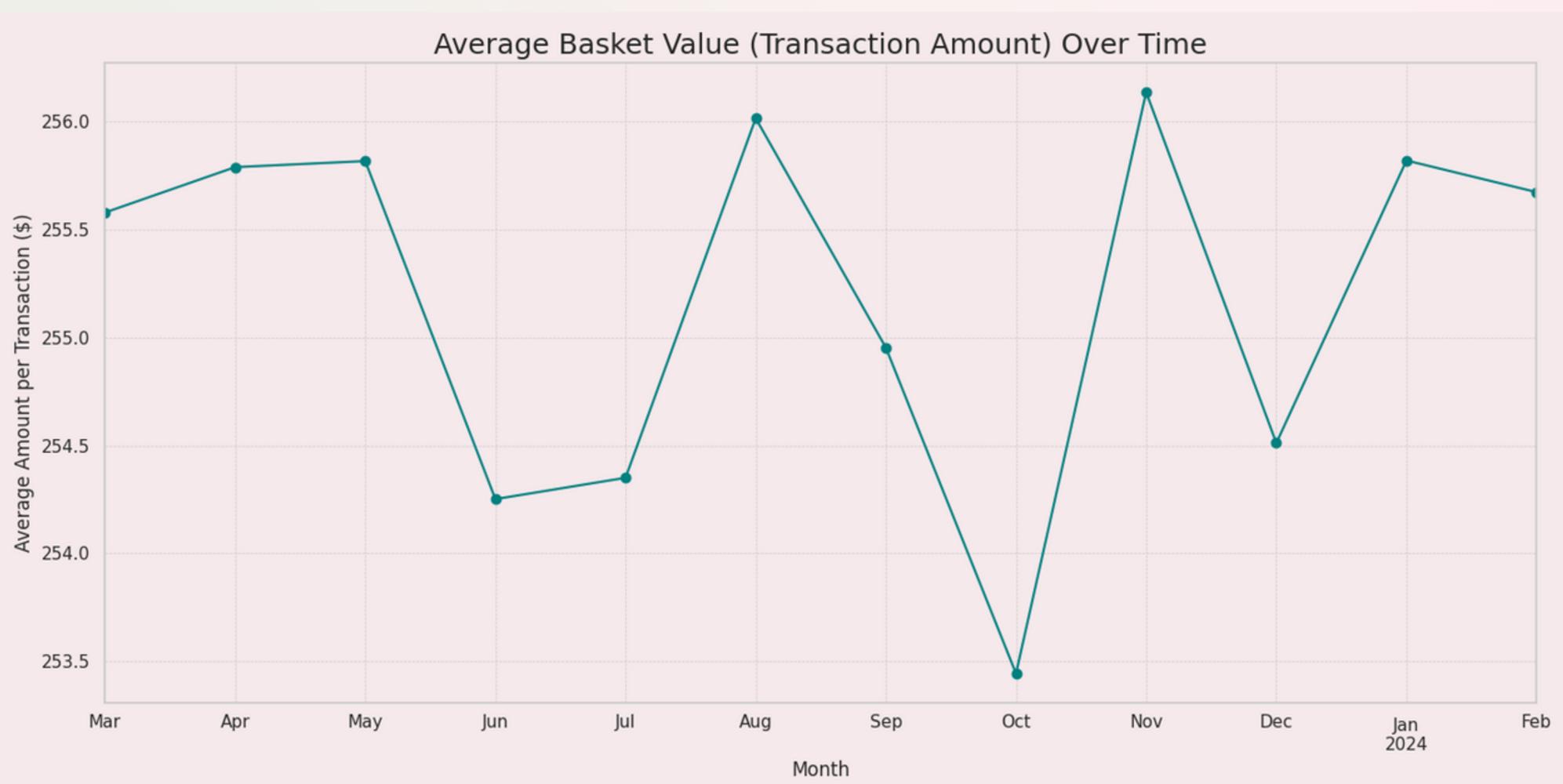
# *Temporal Analysis*

Temporal analysis is the study of data, events, or patterns over time. The word temporal comes from "time," so it's essentially about understanding how things change, evolve, or behave across different time periods.

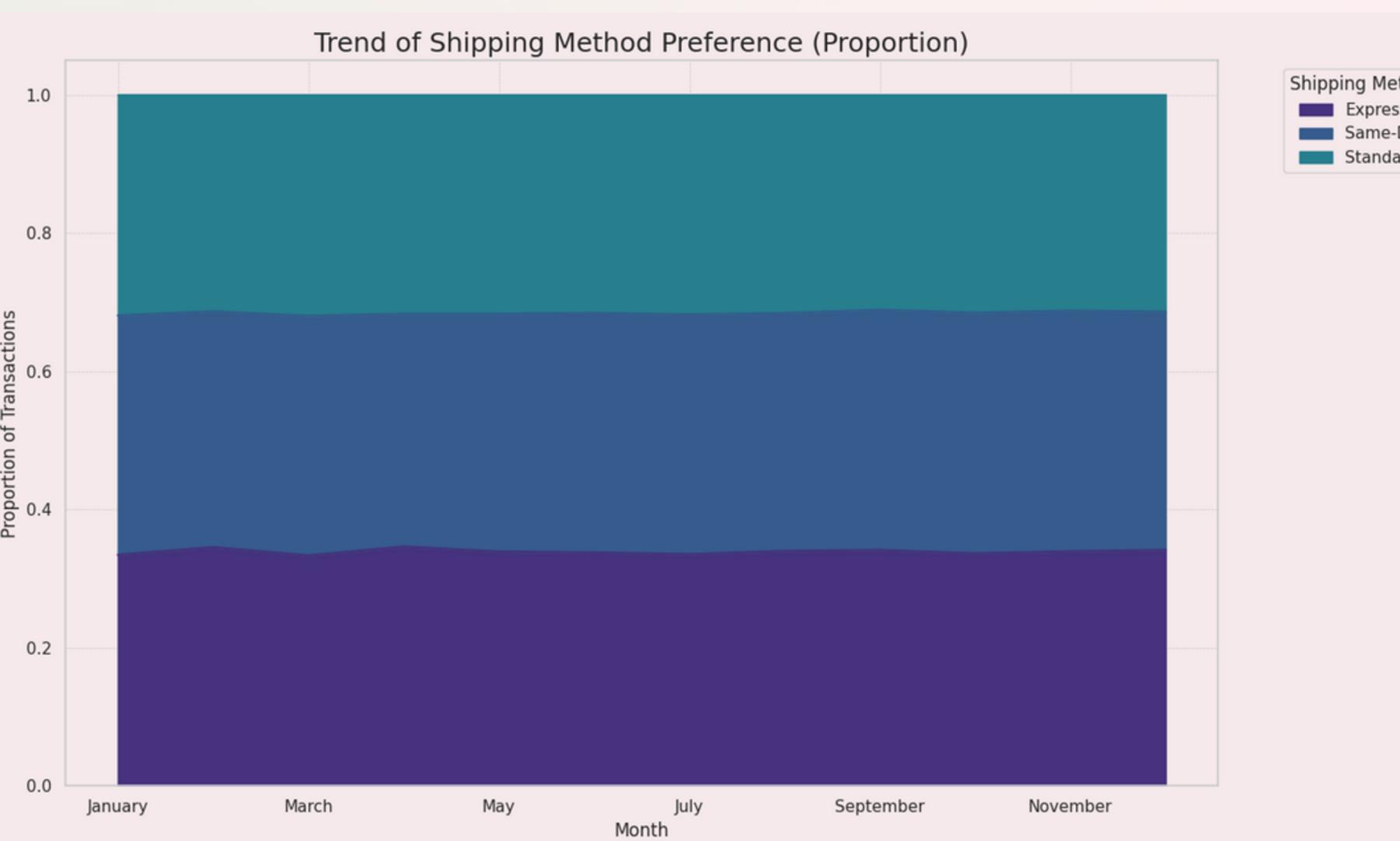
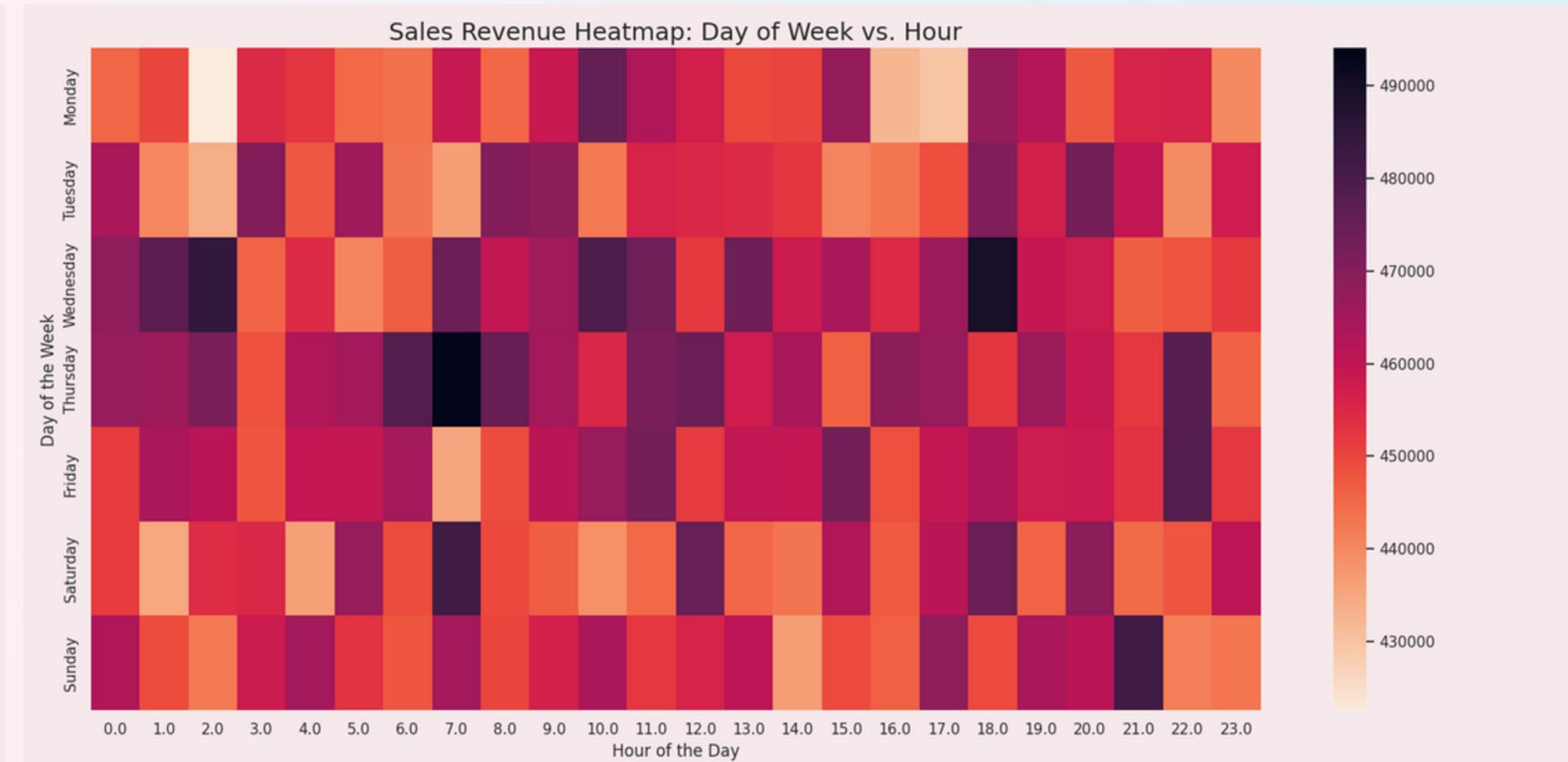
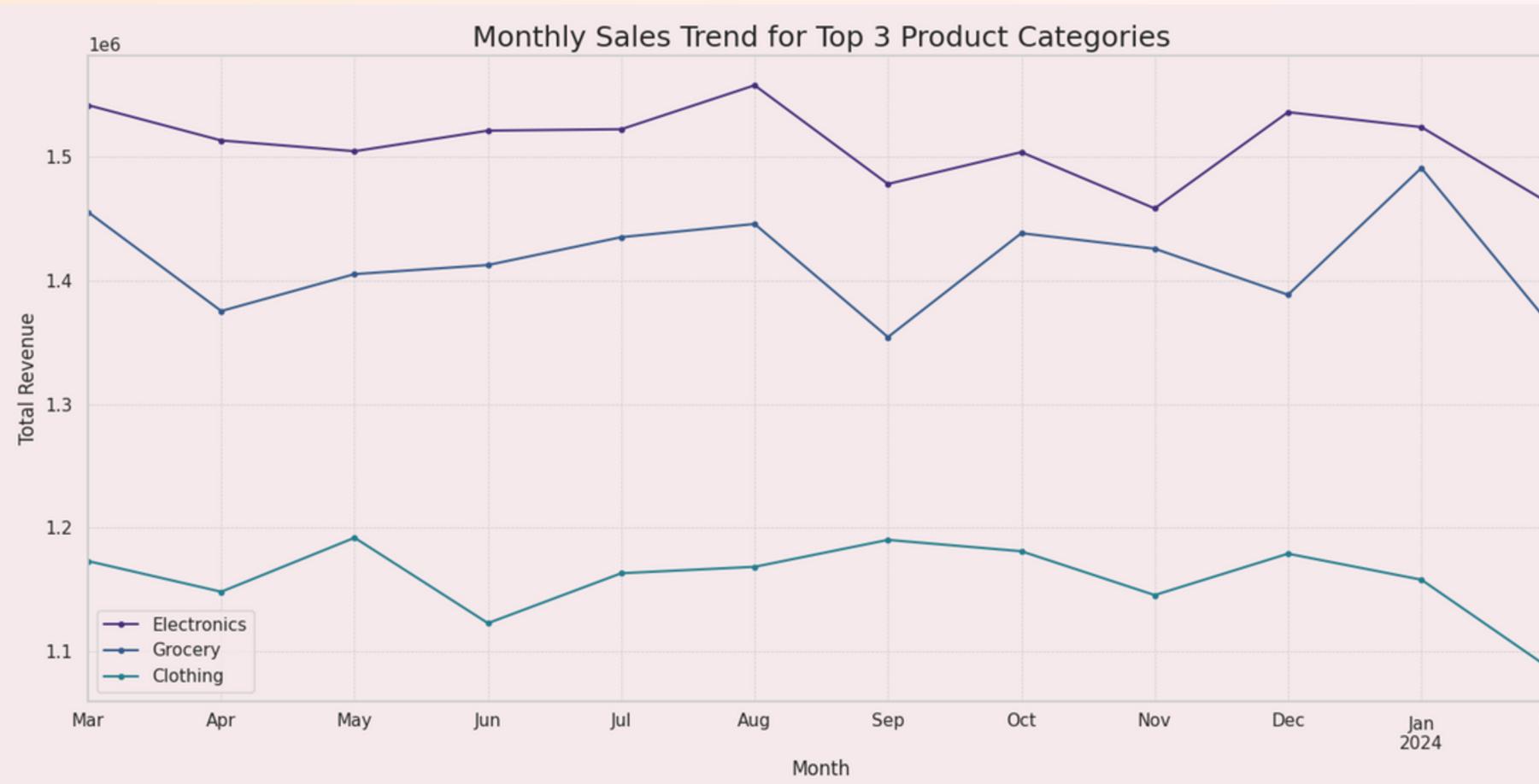




- Customer Sales consistently peak in December and January, driven by holiday shopping. Therefore, Holidays are the best period for business.
- We need to launch discount campaigns during the holidays to fuel our sales.



- The average transaction value is increasing, suggesting growing customer spend.
- Fuel this factor by introducing coupons over a minimum amount of shopping, and special discounts above a particular amount spent.



- Electronics and Grocery are the most dominating product categories throughout the year. Clothing is the 3<sup>rd</sup> best.
- The operational trends show a gradual but clear increase in the proportion of customers choosing 'Same-Day' or 'Express' shipping options over time.
- The hourly sales heatmap reveal that transaction volumes begin to climb significantly around 3 PM, peaking between 4 PM and 8 PM. This "after-work" or "evening relaxation" window is when most customers are actively browsing and buying.

# Key Insights:

- The data shows a massive revenue spike in December, confirming the holiday season as our most critical sales period.
- On a weekly basis, sales ramp up from Friday and peak on Saturday, with Sunday being the second-best day. This indicates that customers dedicate their weekends to shopping.
- The analysis of average basket value shows a positive upward trend over the year. This is a very healthy sign, indicating that individual customers are, on average, spending more per transaction. This could be due to successful upselling, price increases, or customers buying more items per order.
- The operational trends show a gradual but clear increase in the proportion of customers choosing 'Same-Day' or 'Express' shipping options over time. This suggests a growing demand for speed and convenience. Payment methods remain dominated by cards, showing stability in digital adoption.

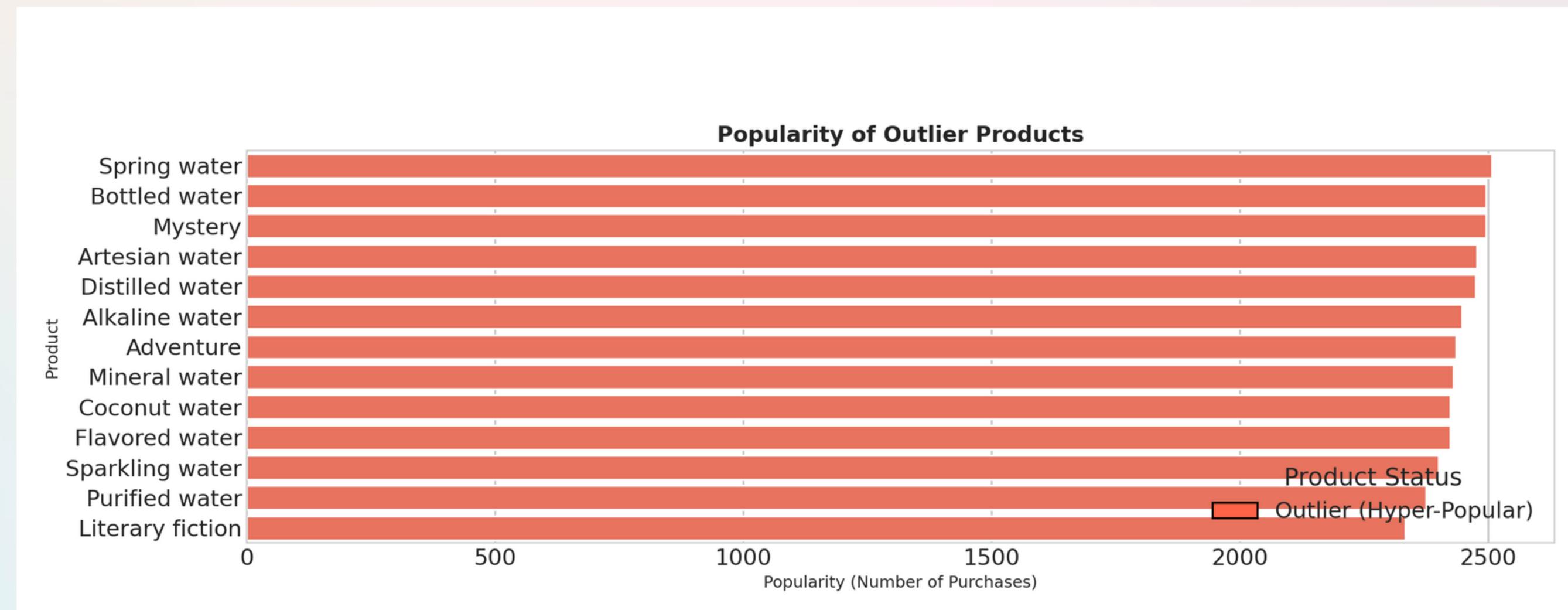
# Implementations:

- Begin building up inventory for key products by late October to avoid stockouts in December. Increase staffing levels (for both in-store and customer support) on weekends, especially during the Q4 holiday season.
- Launch major holiday marketing campaigns in early November. Allocate a larger portion of the weekly ad budget to Thursdays and Fridays to capture customers as they plan their weekend shopping.
- Enhance premium segment value by offering exclusive benefits tied to preferred credit card usage.
- Create exclusive, time-sensitive promotions that run only from Friday to Sunday to capitalize on the natural surge in weekend traffic.
- Deploy daily flash sales or promotional emails around 5 PM to land in customers' inboxes just as they are starting their evening browsing.
- Double down on strategies that are clearly working. Aggressively promote "Frequently Bought Together" sections, product bundles, and "You might also like" recommendations, especially on product and cart pages.
- Introduce promotions that encourage higher spending, such as "Spend \$150, get free premium shipping" or "15% off orders over \$200."

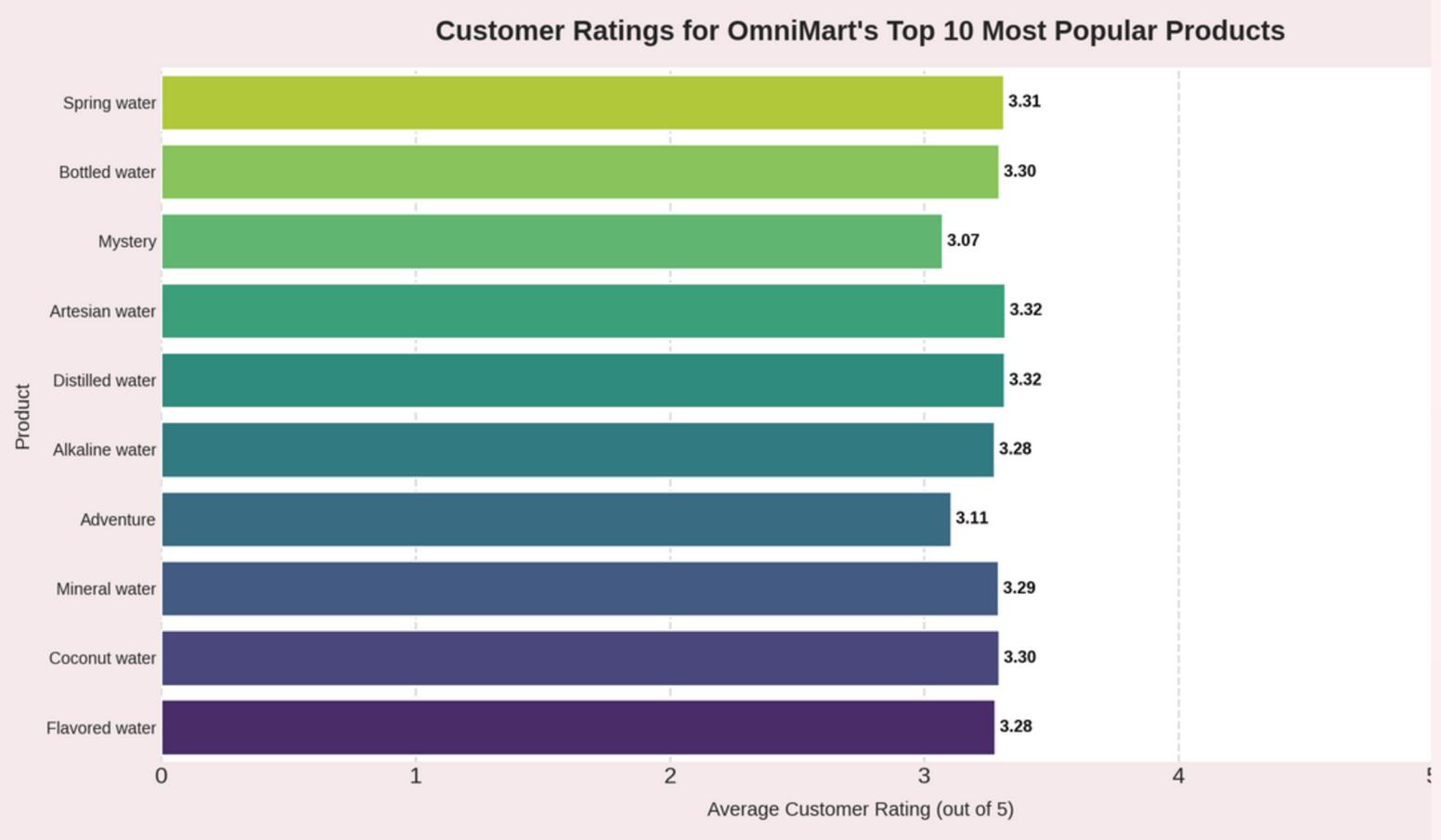
# *Outlier Analysis*

Outlier analysis is the process of identifying and examining data points that significantly differ from the rest of a dataset. These outliers may represent rare events, errors, anomalies, or valuable insights.

From our dataset, we are having a look at some outlier products, which perform significantly better than others.



All best-sellers maintain above average customer satisfaction



- A few products drive most sales, showing a power-law effect.
- These top products have Average to high ratings (3+), strong customer satisfaction, and trusted brands (Apple, Nike, GoPro).
- Electronics and Clothing dominate as the best-performing categories.



- Value Mismatches – Some products are overpriced for their quality, leading to buyer's remorse and loss of trust.
- Churn Risk – Customers who bought these products are more likely to leave OmniMart.
- Brand Reputation Threat – Negative reviews on costly items can heavily impact future sales.

# **Key Insights:**

- A Small Number of Products Drive a Large Volume of Sales. This indicates a power Law Distribution.
- Outliers are High-Quality and Well-Regarded When we examined the characteristics of these outlier products, we found two consistent trends:- Average to High Ratings, and Strong Brand Association.
- Electronics and Clothing Categories Dominate The most popular products predominantly come from the "Electronics" and "Clothing" categories.
- The analysis highlights products where customers see a clear gap between price and quality. These overpriced, low-quality items trigger buyer's remorse and damage brand trust.
- Customers who bought these products are at a high risk of churn. They are likely dissatisfied and may not return to OmniMart for future purchases.
- In the digital age, a few negative reviews for an expensive product can heavily influence future buyers. The products on this list are likely magnets for poor online reviews and negative word-of-mouth.

# **Implementations:**

- Launch "Hero Product" Marketing Campaigns Action: Elevate the hyper-popular products to "hero" status. Create dedicated, high-visibility marketing campaigns around them. Optimize shipping by promoting express/same-day options for high-demand categories like Electronics and Clothing.
- Create Strategic Product Bundles Action: Use the popularity of outlier products to increase the sales of other, less-popular items.
- Replicate Success in Product Development Action: Analyze the DNA of successful products to inform new product acquisitions and development.
- Consider launching a targeted email campaign to customers who purchased these items present in the worst products list, offering a discount on a future purchase or a survey to gather more detailed feedback.
- For the worst products list
- The marketing team should monitor social media and review sites for mentions of these specific products. Addressing public complaints head-on and demonstrating that the company is taking action (e.g., "We've listened to your feedback and are reviewing this product") can mitigate reputational damage.

# ***Final Recommendation***

- Double Down on Electronics & Clothing (Hero Categories):- Secure partnerships with Apple, Nike, Samsung, Adidas. Launch “Hero Product” campaigns around top-rated outliers and use bundles to increase AOV.
- Leverage Temporal Buying Patterns:- Stock up for December peaks, run weekend-only promos, and target the 4–8 PM rush with ads, flash sales, and email drops.
- Revamp Loyalty & Premium Programs:- Redesign perks with free express shipping, exclusive bundles, and tiered rewards. Build clear upgrade pathways for New/Regular customers to become Premium.
- Shift to Behavior-Based Segmentation:- Segment by frequency, product preference, and shipping/payment habits instead of just demographics. Tailor campaigns (e.g., bundles for Clothing buyers, express shipping promos for Grocery buyers).
- Invest in Faster Shipping & Operational Readiness:- Strengthen express/same-day delivery infrastructure. Ensure systems scale during weekend/evening spikes to avoid downtime.
- Safeguard Customer Trust by Fixing Value-Mismatch Products:- Identify overpriced, low-rated products and either fix, re-price, or delist them. Use proactive outreach to dissatisfied buyers to prevent churn.
- Amplify Social Proof & Customer Satisfaction:- Promote Top-Rated/Customer Favorites sections. Use positive reviews in ads and reward promoters with referral incentives.
- Capitalize on Rising Basket Size with Tiered Incentives:- Push premium/high-margin products. Offer tiered promotions like “Spend Rs. 150, get free premium shipping.”

*Thank You!!*

Team- Hephaestus