

14/09/2025

TITLE: OMNIMART RETAILERS – EDA REPORT

SUBTITLE: UNCOVERING CUSTOMER & SALES
INSIGHTS EXPLORATORY DATA ANALYSIS



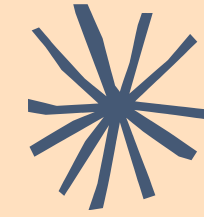
–Created by Data Crafters



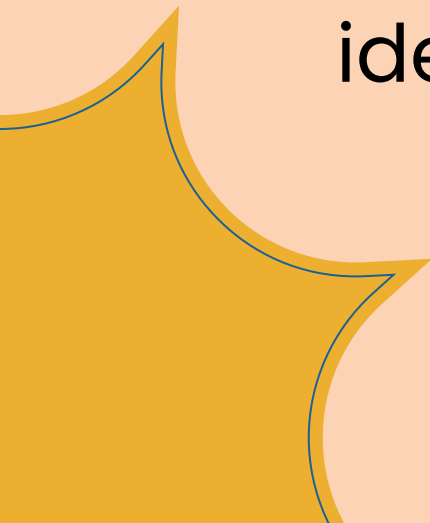
TABLE OF CONTENT

1. Problem Statement
2. Phase 1: Preparation & Setup
3. Phase 2: Initial Reconnaissance
4. Phase 3: Uncovering Insights
 - Univariate Analysis
 - Bivariate Analysis
 - Multivariate Analysis
5. Phase 4: Temporal Trends and Outlier Detection & Insights
6. Phase 5: Key Findings & Recommendations
7. Conclusion
 - Summary of Insights

PROBLEM STATEMENT



The Retail Challenge: Optimizing Customer Experience and Sales at OmniMart
Retailers Problem Statement: 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. Your challenge is to act as a data analyst for OmniMart. Using the provided dataset, your team must perform a comprehensive Exploratory Data Analysis (EDA) to uncover actionable insights. Your analysis should focus on answering key business questions and identifying opportunities for growth.





Phase 1: Preparation & Setup

- Imported necessary libraries(pandas, numpy, seaborn, matplotlib)
- Loaded Omnimart dataset into notebook
- Checked dataset size and structure
- Reviewed data types(numerical, categorical, temporal)

Phase 2: Initial reconnaissance

The dataset consists of 302,010 records and 30 columns, capturing detailed information on customer transactions. It includes a mix of numerical, categorical, and datetime data types, such as:

- Customer & Transaction Details – Transaction_ID, Customer_ID, Date, Time
- Demographics – Name, Email, Phone, Age, Gender, Income, Location (Address, City, State, Zipcode, Country)
- Purchase Information – Total_Purchases, Amount, Total_Amount, Product_Category, Product_Type, Product_Brand
- Experience & Feedback – Ratings, Feedback, Shipping_Method, Payment_Method, Order_Status

✏ Missing Data Notice: Several columns contain missing values, especially the Date column, which has data for only ~60% of the entries. This will need to be handled during the data cleaning phase.

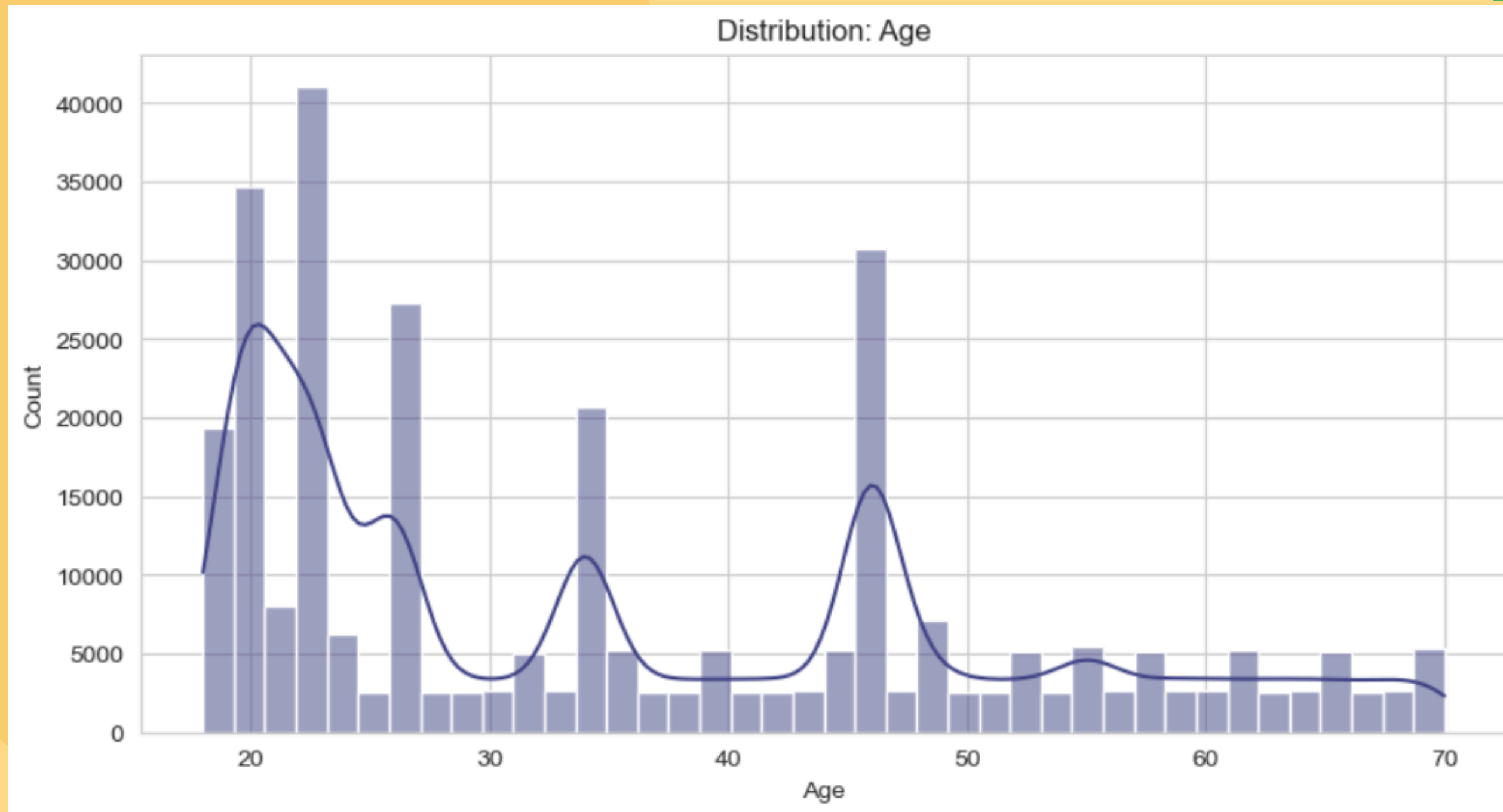
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 302010 entries, 0 to 302009
Data columns (total 30 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Transaction_ID         301677 non-null float64
1   Customer_ID           301702 non-null float64
2   Name                  301628 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                  182895 non-null datetime64[ns]
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: datetime64[ns](1), float64(10), object(19)
memory usage: 69.1+ MB
```

=== Statistical Summary (Numerical) ===							
	count	mean	min	25%	50%	75%	std
Transaction_ID	301677.0	5495822.630356	1000007.0	3247930.0	5499657.0	7739509.0	2595564.714314
Customer_ID	301702.0	55006.553934	10000.0	32469.25	55012.0	77511.0	26005.6752
Phone	301648.0	5501464142.74998	1000049414.0	3255061091.5	5505811969.5	7749859742.75	2596016603.771404
Zipcode	301670.0	50298.951019	501.0	25425.0	50602.5	75252.0	28972.807134
Age	301837.0	35.481326	18.0	22.0	32.0	46.0	15.021933
Date	182895	2023-09-05 07:04:07.361601024	2023-03-13 00:00:00	2023-05-31 00:00:00	2023-08-31 00:00:00	2023-11-30 00:00:00	2024-02-29 00:00:00 NaN
Year	301660.0	2023.165113	2023.0	2023.0	2023.0	2023.0	2024.0 0.371283
Total_Purchases	301649.0	5.359729	1.0	3.0	5.0	8.0	10.0 2.868575
Amount	301653.0	255.163659	10.000219	132.890764	255.470969	377.672606	499.997911 141.38964
Total_Amount	301660.0	1367.651156	10.00375	438.724278	1041.117547	2029.999853	4999.625796 1128.998515
Ratings	301826.0	3.16267	1.0	2.0	3.0	4.0	5.0 1.320827

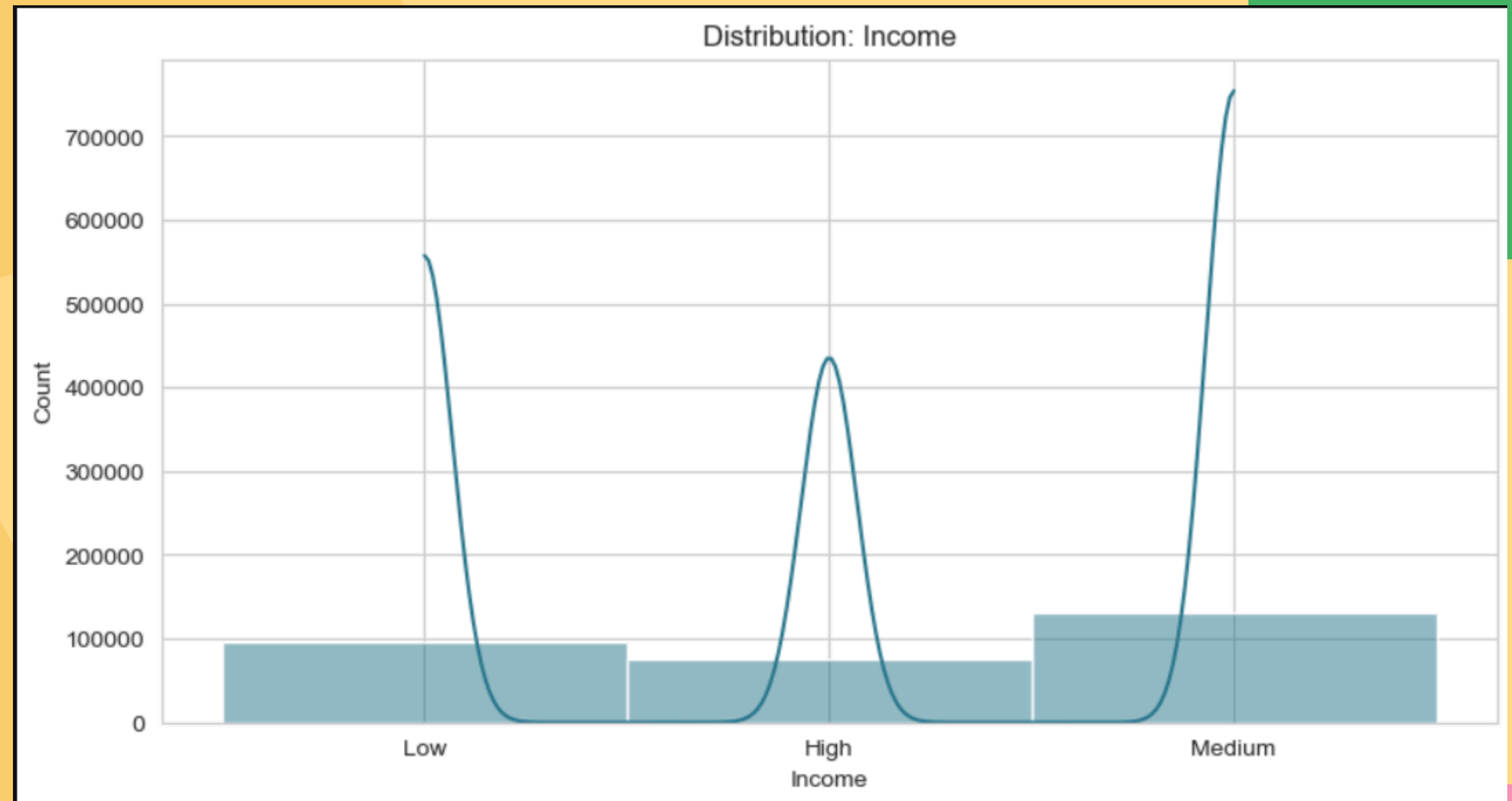
- Customer Age ranges from 18 to 70, with an average of ~35 years.
- Total Purchases average ~5 per customer, maxing at 10.
- Amount per purchase averages \$255, while Total Amount spent averages \$1367, indicating varied spending behavior.
- Ratings average around 3.16, suggesting moderate satisfaction.
- Date values are missing in ~40% of records; most data is from 2023

Phase 3: Uncovering
Insights

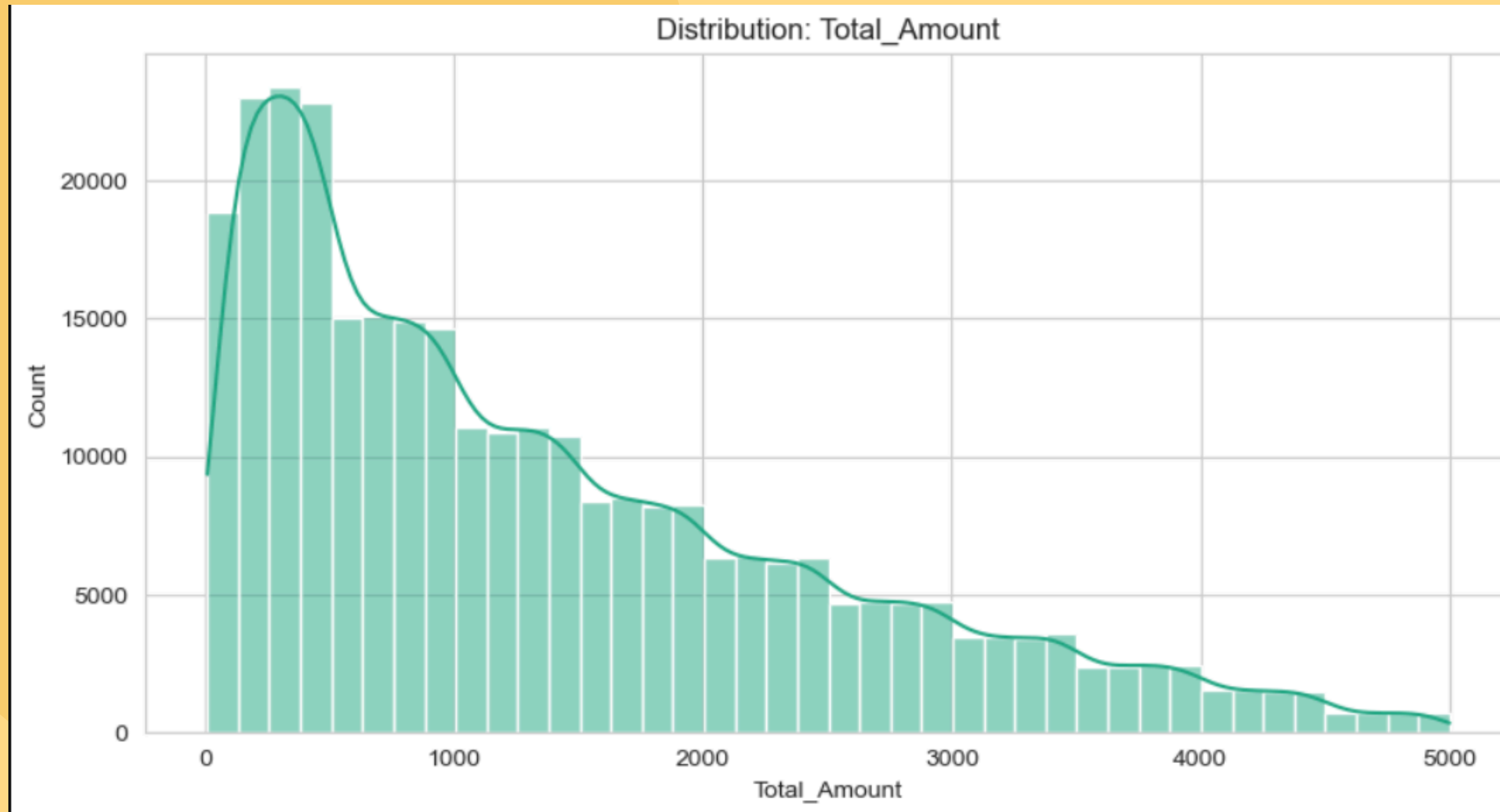
UNIVARIATE ANALYSIS



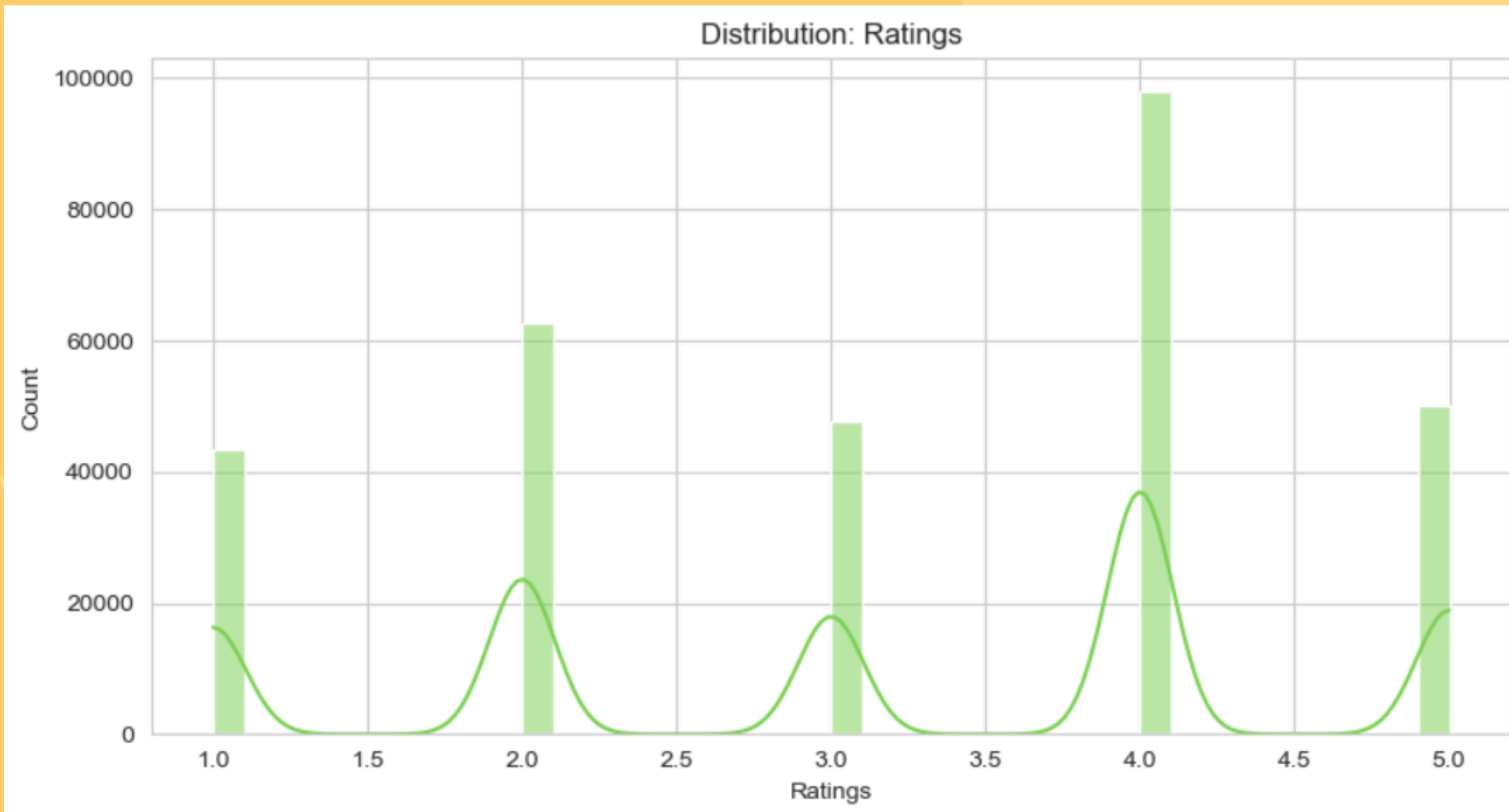
- Most customers fall into the younger and middle-age brackets (25–44), indicating that OmniMart’s main customer base is working-age adults.
- Fewer customers are from the older age groups (55+), suggesting limited penetration in that demographic



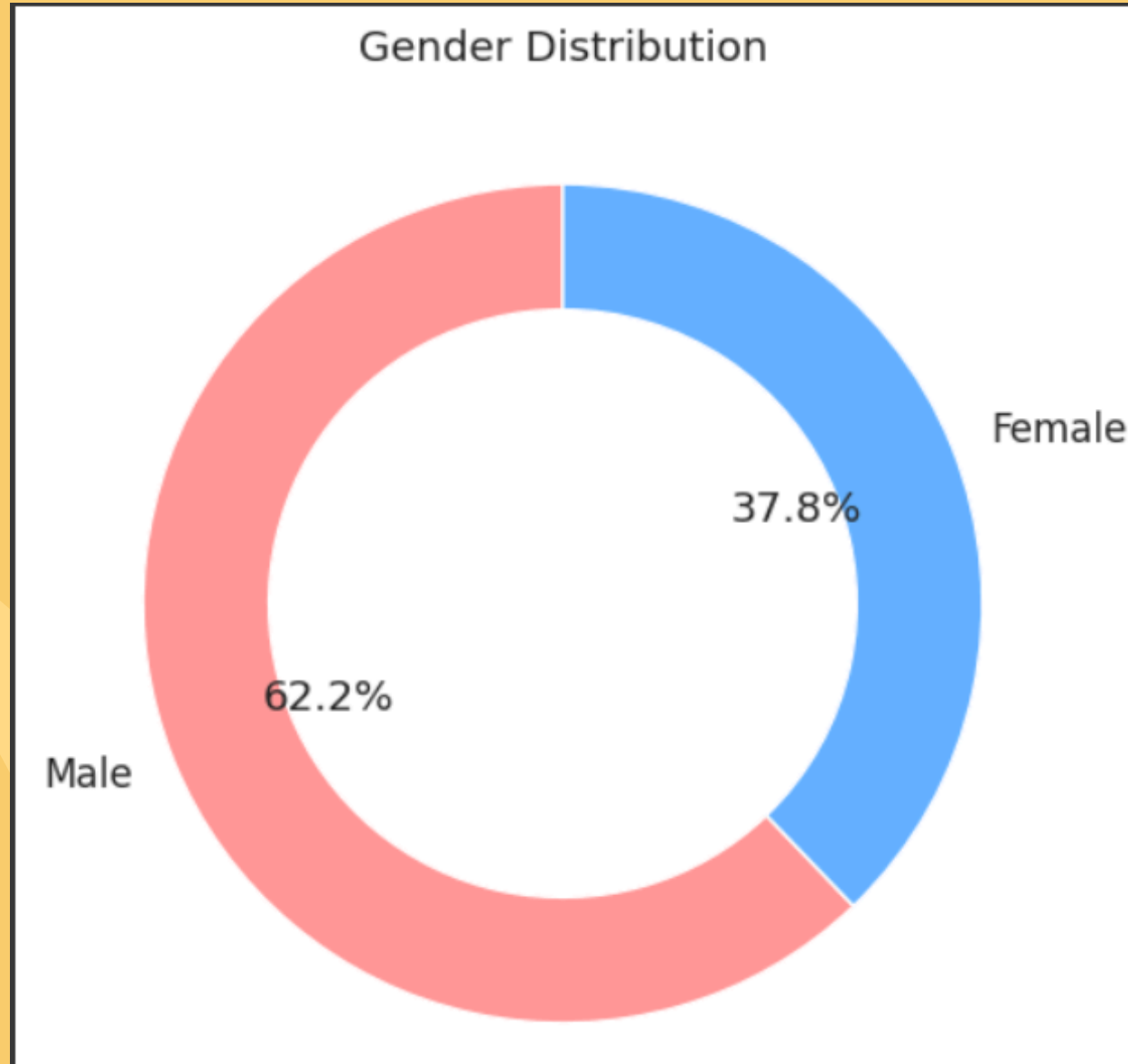
- Income is skewed, with a large concentration in the middle-income range.
- Very high-income customers are fewer, but they may represent premium buyers with higher average spending.



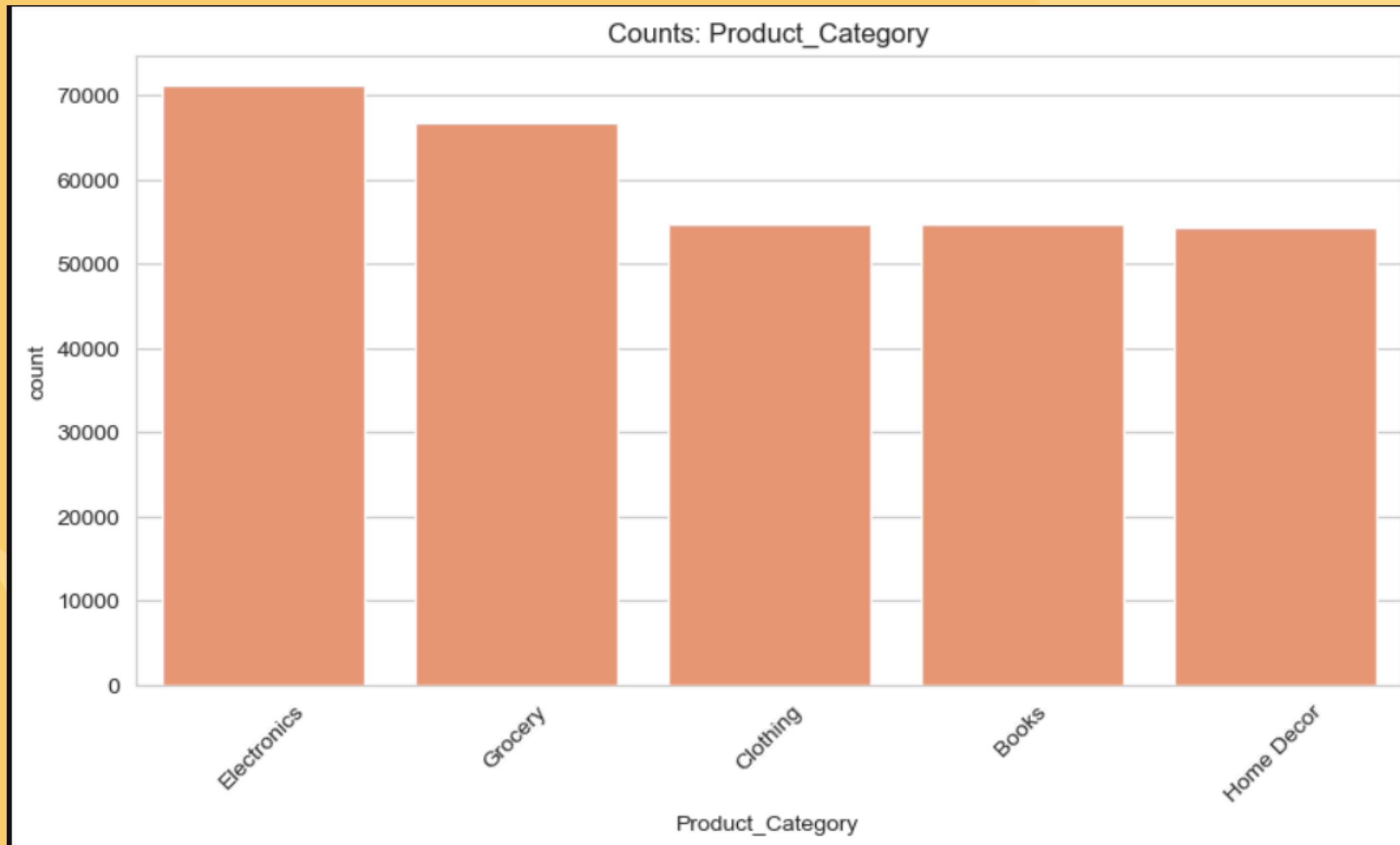
- Spending patterns show a long-tail effect—most transactions are in the low-to-mid range, while only a few very high-value purchases exist.
- This reflects typical retail behavior: frequent small purchases, rare big-ticket orders.



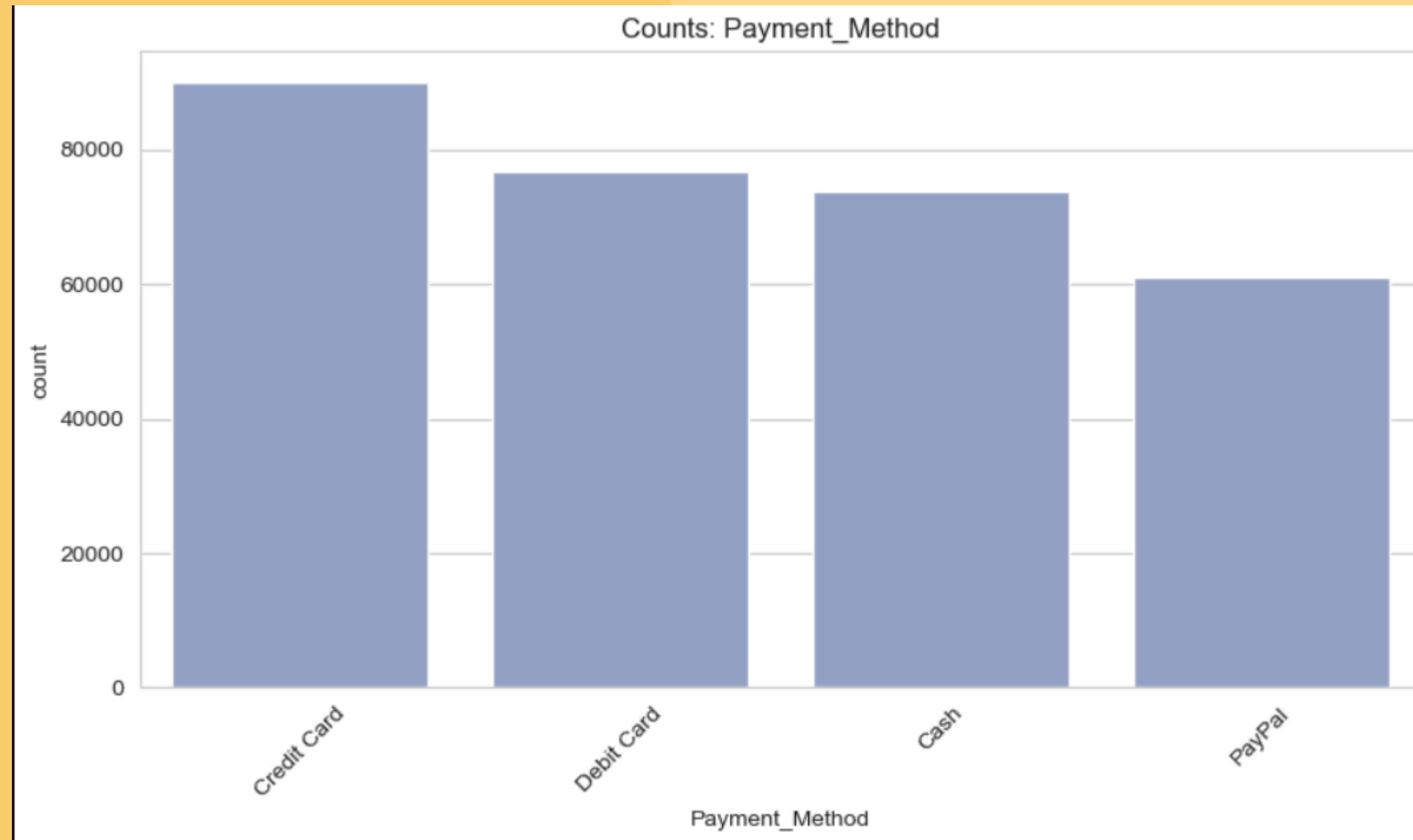
- Ratings cluster around 3 and 4, suggesting average-to-good satisfaction levels.
- Few extreme ratings (1 or 5) indicate customers tend to give moderate rather than polarized feedback.



- Purchases are relatively balanced between male and female customers, ensuring a broad customer base.
- Slight skew (if visible) can indicate which gender drives slightly more sales.



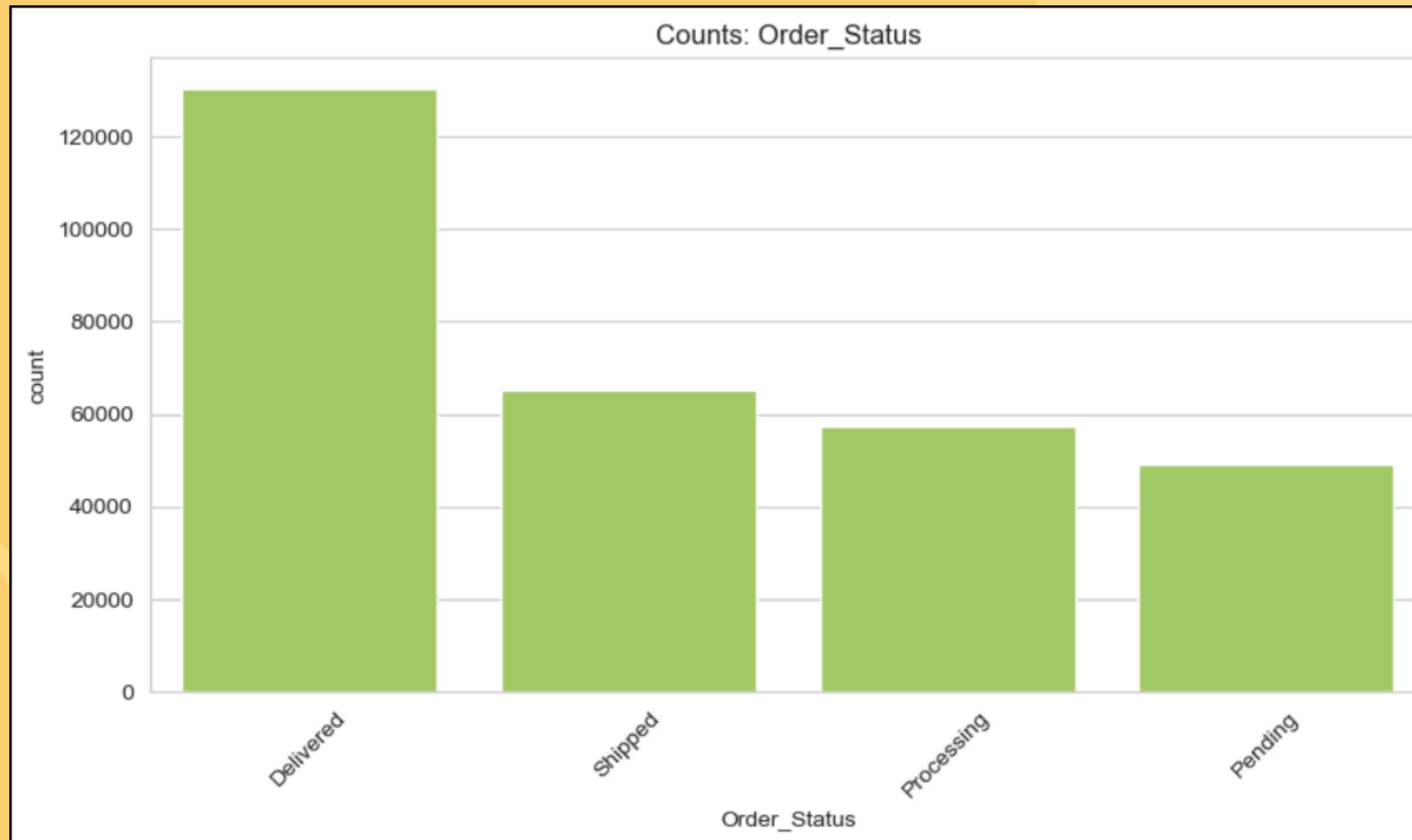
- A few categories dominate sales, while others are less frequent.
- This suggests OmniMart could prioritize inventory and marketing for top-selling categories



- Certain payment methods (likely Credit/Debit Cards or Digital Wallets) dominate, while others are rarely used.
- Promotions can be aligned with popular methods, while incentives may encourage adoption of less-used ones

- Customers show clear preference for a few shipping types (e.g., Standard or Express).
- Rarely chosen methods may either be too costly or less reliable.

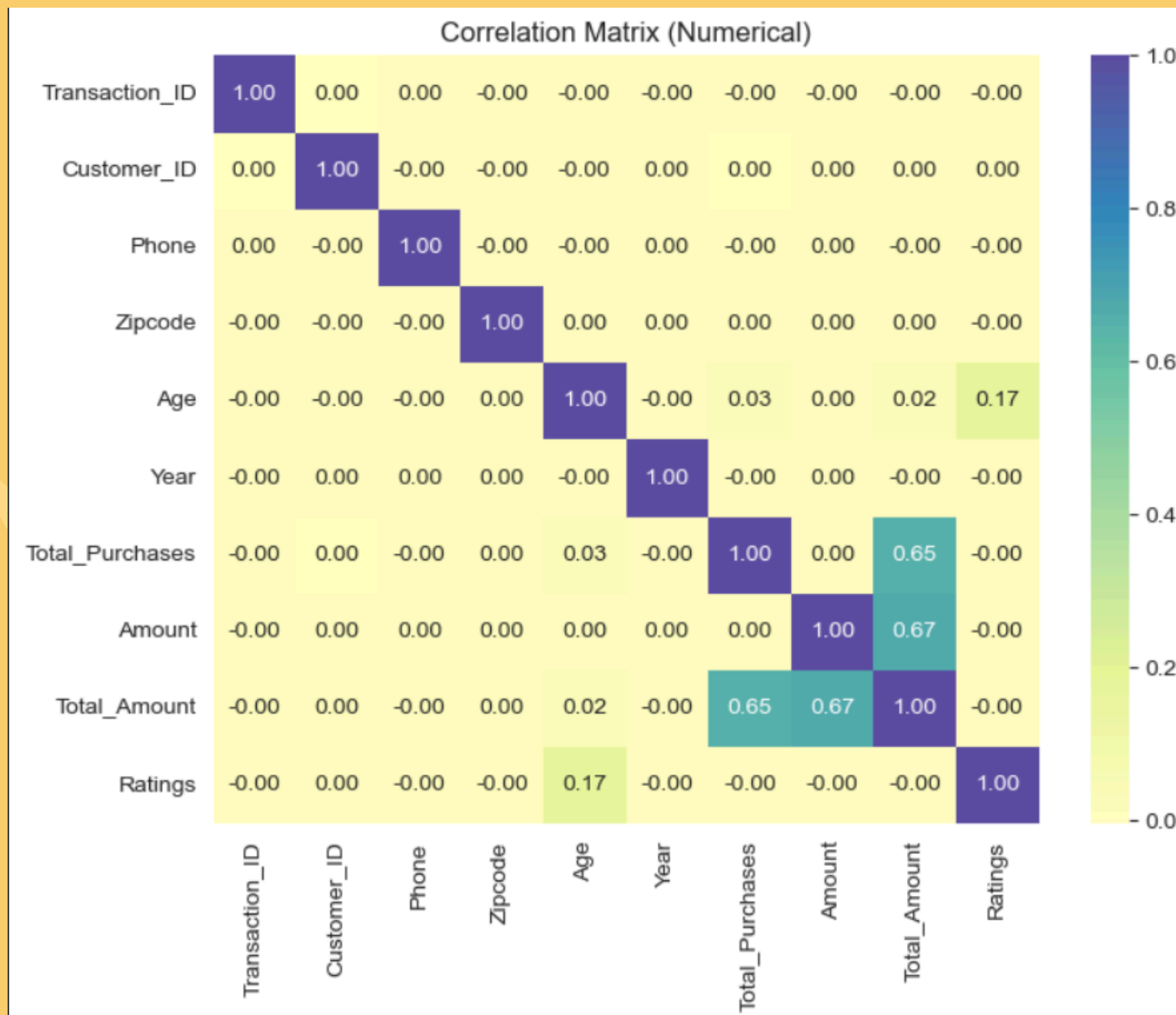




- The majority of orders are successfully completed, which indicates strong operational performance.
- A noticeable portion of orders falls under Cancelled or Returned, highlighting areas where customer experience or logistics may need improvement.
- Tracking reasons for cancellations and returns can uncover actionable insights—such as product quality issues, delayed deliveries, or mismatched customer expectations.

Phase 3: Uncovering
Insights

BIVARIATE ANALYSIS



- Positive correlation between Income and Total Amount, confirming higher-income customers spend more.
- Weak or negative correlations elsewhere show independence among other variables



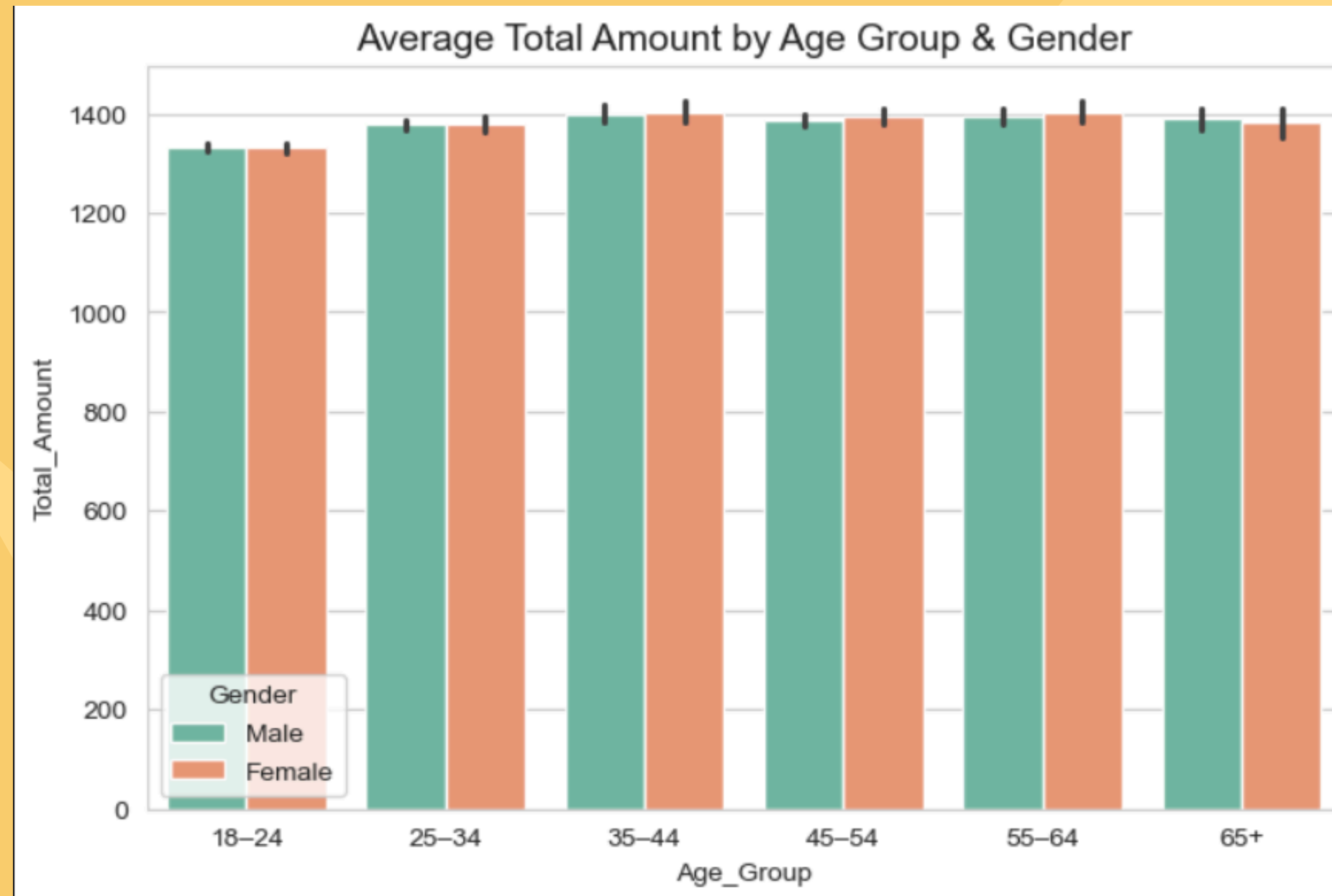
- Median spending levels may differ between genders, indicating one group contributes higher average revenue.
- Presence of outliers shows some individuals (in both groups) spend significantly more.



- Customers opting for certain shipping methods (like Express) often spend more.
- Low-spend customers may prefer cheaper shipping options.

Phase 3: Uncovering
Insights

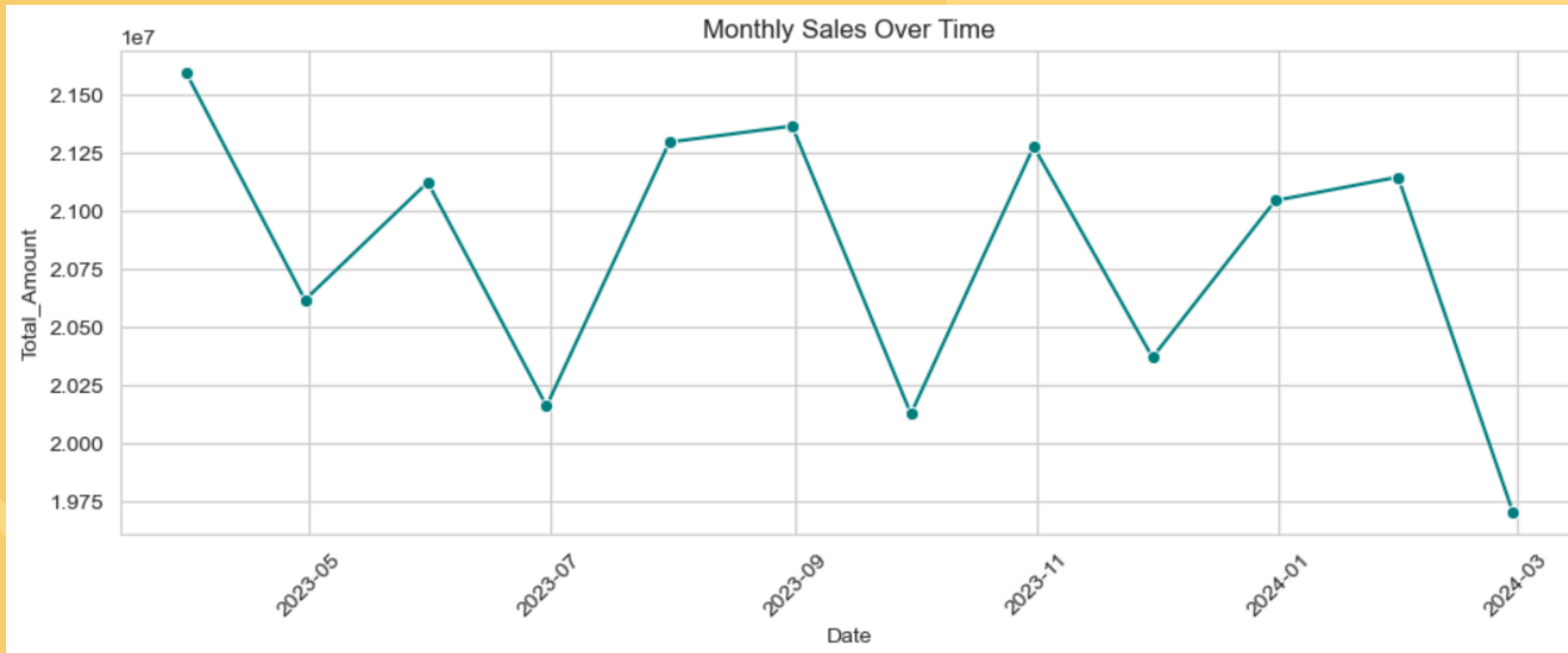
MULTIVARIATE ANALYSIS



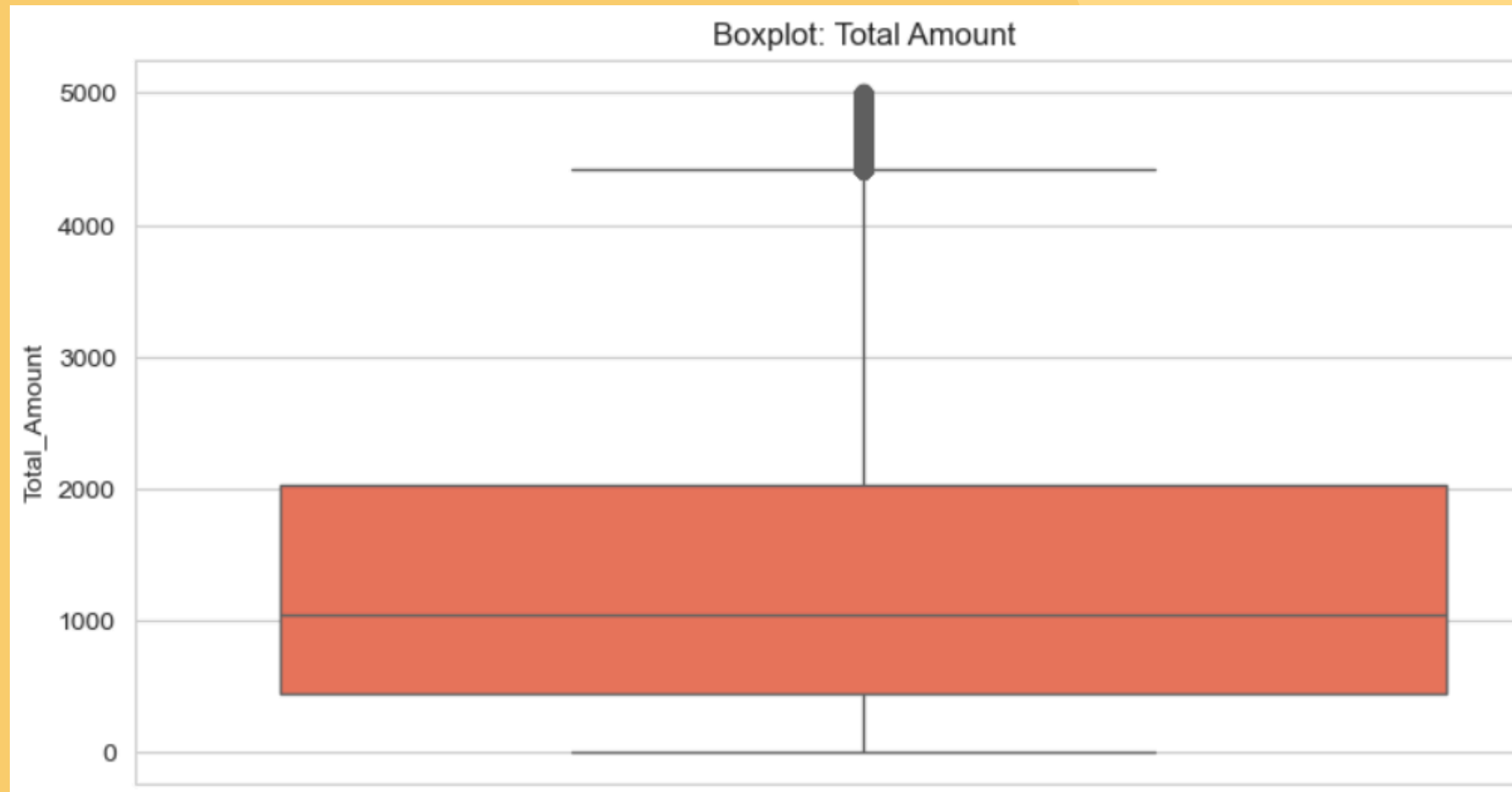
- Middle-age groups (35-44, 45-54) contribute the highest spending.
- Gender differences highlight which demographic segment brings higher order values.



- Seasonal peaks suggest higher demand in specific months (possibly festive seasons).
- Some categories show consistent demand, while others fluctuate more



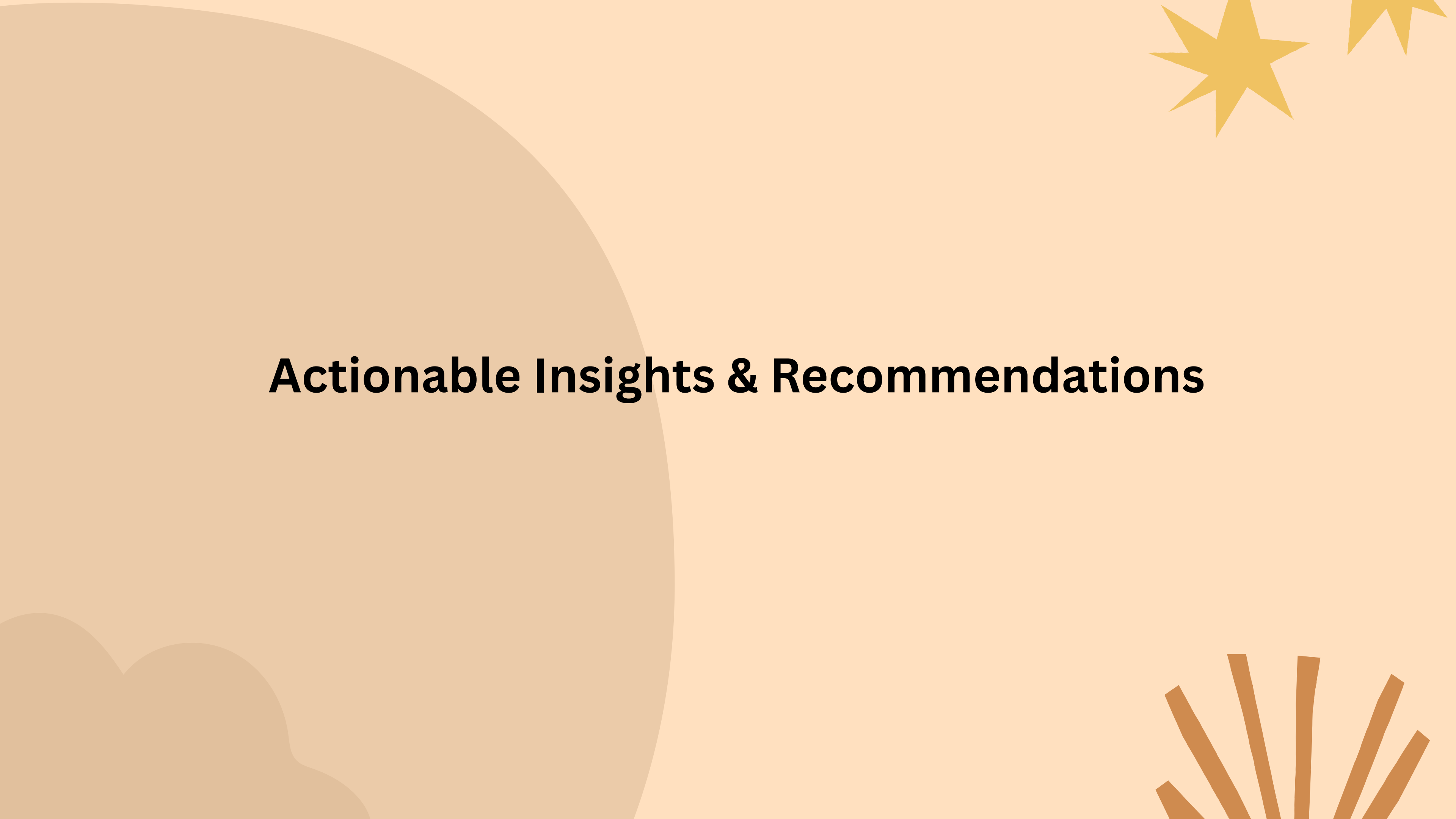
- General upward trend indicates growth in sales.
- Spikes suggest promotions or seasonal events driving short-term boosts.



Outliers found: 3897 rows

	Customer_ID	Total_Amount
150	72950.0	4885.287161
159	45413.0	4850.773617
395	20280.0	4487.127690
400	64275.0	4518.677874
484	97549.0	4755.346608

- Majority of transactions are within a stable range.
- Outliers represent either very high-value VIP customers or potential anomalies requiring review.



Actionable Insights & Recommendations

Based on the exploratory data analysis, here are the key takeaways and data-driven strategies for OmniMart Retailers:

1. Customer Segmentation

- High-spending customers can be clearly identified using Total_Amount.
- Target age groups 25–44, especially female customers, who show higher average spending.

2. Product Strategy

- Focus on top 4 product categories with consistent monthly revenue.
- Investigate lower-rated but high-revenue products to improve quality or customer satisfaction.

3. Shipping & Fulfillment

- Shipping methods influence spending. Methods linked to lower average order values should be re-evaluated.
- Optimize shipping for high-value orders to reduce friction.

4. Payment Method Optimization

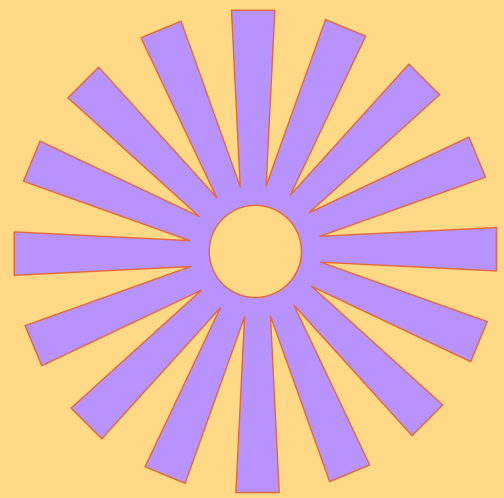
- Encourage usage of payment methods associated with higher spending via incentives or loyalty programs.

5. Time-Based Marketing

- Leverage monthly sales trends to plan seasonal promotions and stock popular products in advance.

6. Outlier & Fraud Detection

- Monitor unusually high or low transactions — potential indicators of fraud, data errors, or VIP customers.



THANK YOU

