



MANHATTEN PROJECT : CHALLENGE 1

OMNIMART CUSTOMER ANALYTICS CODE WARRIORS

A COMPREHENSIVE EXPLORATORY DATA ANALYSIS
TO OPTIMIZE CUSTOMER EXPERIENCE, DRIVE SALES
GROWTH, AND ENHANCE OPERATIONAL EFFICIENCY
ACROSS OMNIMART'S GLOBAL RETAIL NETWORK





PROBLEM STAEMENT

YOUR PARAGROMNIMART RETAILERS FACES THE COMPLEXCHALLENGE OF UNDERSTANDING AND OPTIMIZING CUSTOMER RELATIONSHIPS ACROSS A DIVERSE, GLOBAL CUSTOMER BASE. WITH VASTAMOUNTS OF TRANSACTION DATA AND CUSTOMER FEEDBACK AT OUR DISPOSAL, WE HAVE AN UNPRECEDENTED OPPORTUNITY TO TRANSFORM RAW INFORMATION INTO STRATEGIC COMPETITIVE ADVANTAGESAPH TEXT



CHALLENGES

OUR MISSION IS TO DECODE CUSTOMER BEHAVIOR PATTERNS, IDENTIFY HIGH-VALUE SEGMENTS, AND UNCOVER OPERATIONAL INEFFICIENCIES THAT MAY BE IMPACTING PROFITABILITY AND CUSTOMER SATISFACTION. THROUGH SYSTEMATIC ANALYSIS,WE'LL DEVELOP ACTIONABLE RECOMMENDATIONS THAT DRIVE MEASURABLE BUSINESS OUTCOMES.



ANALYSIS

- CUSTOMER BEHAVIOUSR ANALYSIS
- PRODUCT PERFORMANCE ANALYSIS
- STATISTICAL ANALYSIS & HYPOTHESIS
- INSIGHTS AND RECOMMENDATIONS



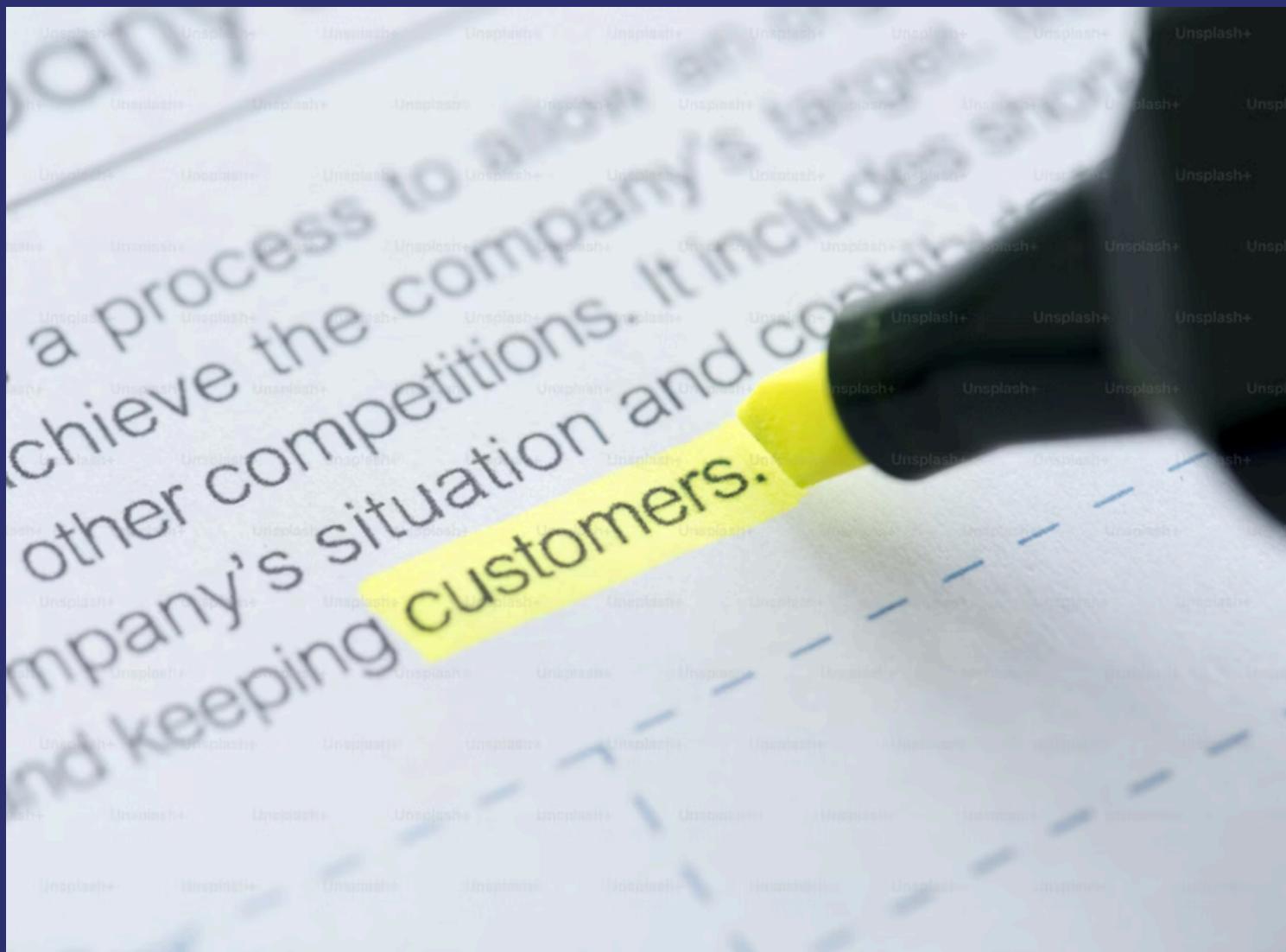
DATA DESCRIPTION: UNVEILING THE RETAIL DATA TREASURE

```
→ Cleaned Shape: (301091, 30)
<class 'pandas.core.frame.DataFrame'>
Index: 301091 entries, 0 to 302009
Data columns (total 30 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Transaction_ID    301091 non-null   float64 
 1   Customer_ID       301091 non-null   float64 
 2   Name              301091 non-null   object  
 3   Email             301091 non-null   object  
 4   Phone              301091 non-null   object  
 5   Address            301091 non-null   object  
 6   City               301091 non-null   object  
 7   State              301091 non-null   object  
 8   Zipcode            301091 non-null   float64 
 9   Country            301091 non-null   object  
 10  Age                301091 non-null   float64 
 11  Gender             301091 non-null   object  
 12  Income             0 non-null        float64 
 13  Customer_Segment   301091 non-null   object  
 14  Date               182340 non-null   datetime64[ns]
 15  Year               182340 non-null   float64 
 16  Month              182340 non-null   object  
 17  Time               301091 non-null   object  
 18  Total_Purchases    301091 non-null   float64 
 19  Amount              301091 non-null   float64 
 20  Total_Amount        301091 non-null   float64 
 21  Product_Category   301091 non-null   object  
 22  Product_Brand      301091 non-null   object  
 23  Product_Type       301091 non-null   object  
 24  Feedback            301091 non-null   object  
 25  Shipping_Method    301091 non-null   object  
 26  Payment_Method     301091 non-null   object  
 27  Order_Status        301091 non-null   object  
 28  Ratings             301091 non-null   float64 
 29  products            301091 non-null   object  
dtypes: datetime64[ns](1), float64(10), object(19)
memory usage: 71.2+ MB
```

- A RICH DATASET OF 302,010 TRANSACTIONS FROM OMNIMART RETAILERS, REVEALING CUSTOMER BEHAVIORS AND PURCHASING PATTERNS.
- CUSTOMER INFO: UNIQUE IDS, NAMES, EMAILS, ADDRESSES, AND DEMOGRAPHICS LIKE AGE, GENDER, AND INCOME SEGMENTS.
- "TRANSACTION DETAILS: PURCHASE COUNTS, TOTAL AMOUNTS SPENT, AND DATES OF THE LATEST BUYS."
- "PRODUCT INSIGHTS: CATEGORIES (E.G., ELECTRONICS, GROCERIES), BRANDS, TYPES, AND CUSTOMER FEEDBACK/RATINGS."
- "LOGISTICS SNAPSHOT: SHIPPING METHODS, PAYMENT CHOICES, AND ORDER STATUSES (SHIPPED, DELIVERED, CANCELED)."

CUSTOMER BEHAVIOR ANALYSIS

SEGMENT CUSTOMERS BY PURCHASING HABITS, DEMOGRAPHICS, AND INCOME LEVELS. ANALYZE SPENDING PATTERNS AND LOYALTY TRENDS TO IDENTIFY HIGH-VALUE CUSTOMER GROUPS AND RETENTION OPPORTUNITIES.



B
A

PRODUCT PERFORMANCE EVALUATION

EXAMINE PRODUCT CATEGORY POPULARITY, BRAND PERFORMANCE, AND CUSTOMER FEEDBACK CORRELATIONS. IDENTIFY BESTSELLERS, UNDERPERFORMERS, AND SEGMENT-SPECIFIC PREFERENCES.



C
B
A

OPERATIONAL OPTIMIZATION

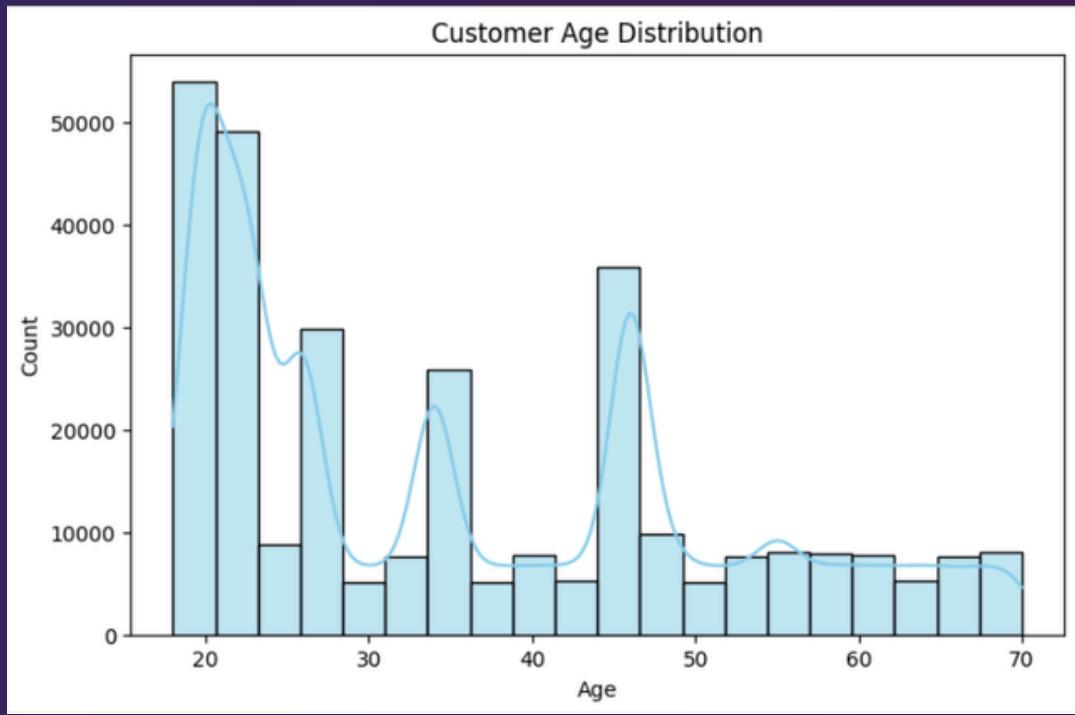
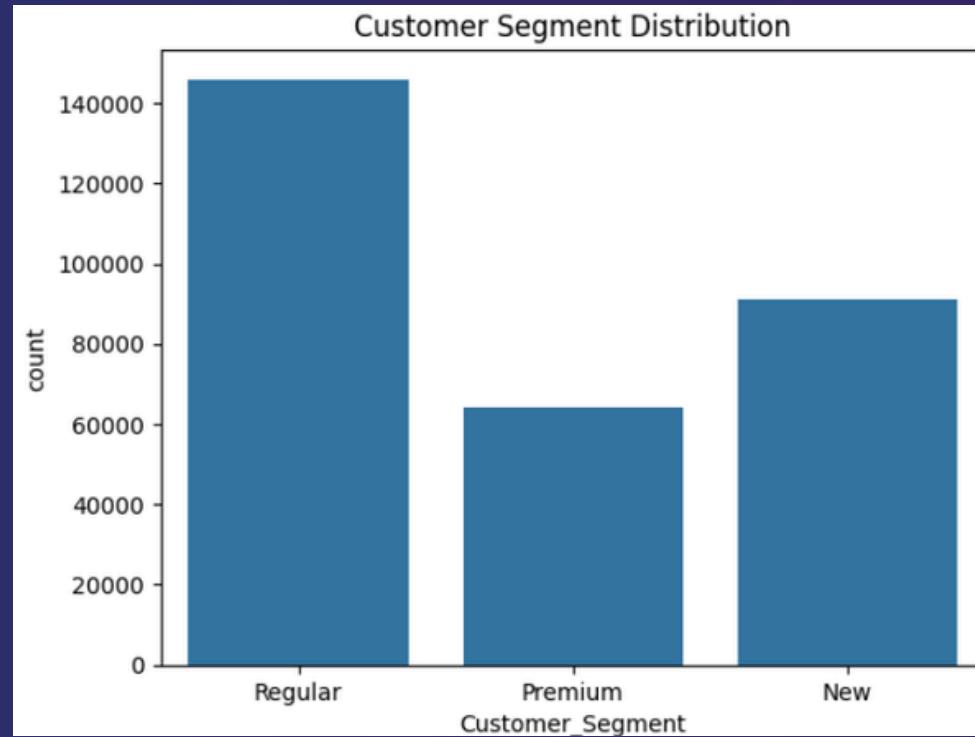
INVESTIGATE SHIPPING METHODS, PAYMENT PREFERENCES, AND ORDER FULFILLMENT EFFICIENCY. ANALYZE FACTORS CONTRIBUTING TO ORDER CANCELLATIONS AND DELIVERY SUCCESS RATES.



D C B A

Customer Behavior Analysis

Customer Behavior Analysis



Customer Age Distribution

- The customer base is predominantly young adults.
- A significant concentration is seen in the early 20s.
- The new customers are increases in number which signifies the growth of the company.

Customer Gender Distribution

- The data shows a higher number of transactions from male customers compared to females.

Customer Behavior Analysis



Customer Income Distribution

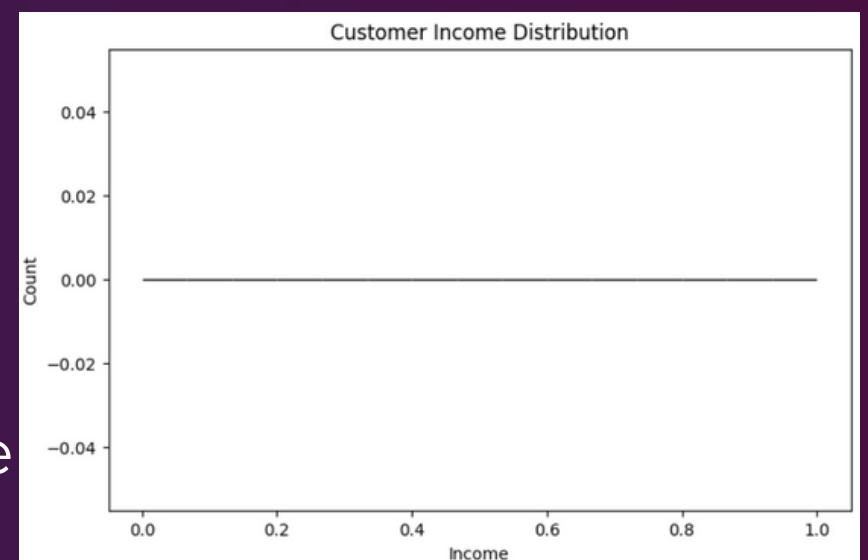
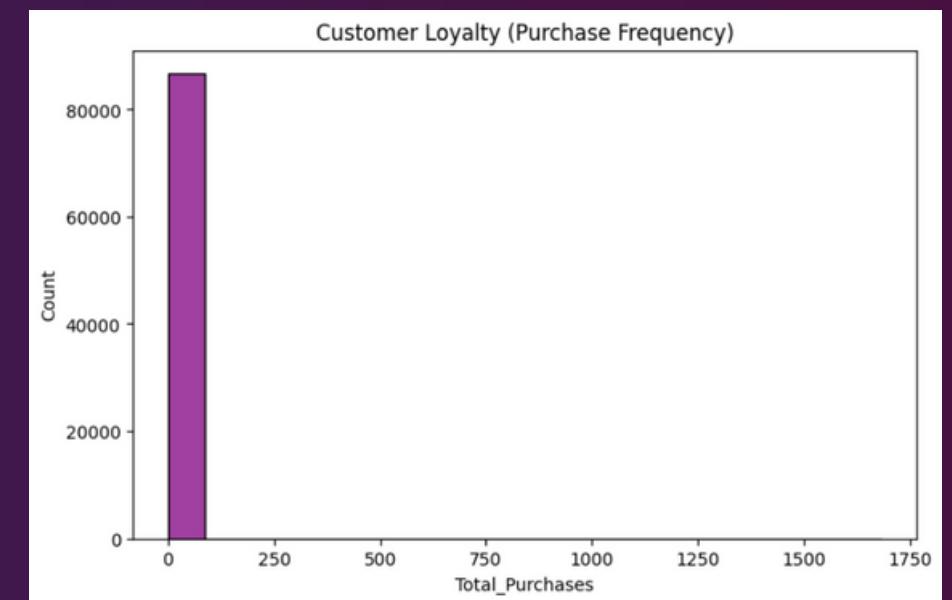
- **Key Insights**: The customer income distribution is heavily concentrated in the middle-income range, with a very specific and telling pattern.

Concentration: The data shows a strong, concentrated peak around the **middle of the income scale** (approximately **0.4 to 0.6** on the normalized axis). This means the overwhelming majority of your customers fall into the middle-income bracket.

Distribution Shape: The distribution is not symmetrical (not a perfect "bell curve"). It shows:

- A very steep rise from low incomes to the mid-income peak.
- A slower, more gradual decline from the mid-income peak to high incomes.
-

Low & High Income Representation: There is a very small number of customers at the extreme low end (~0.0) and a slightly larger, but still modest, tail of customers at the high end (~0.8 to 1.0) of the income spectrum



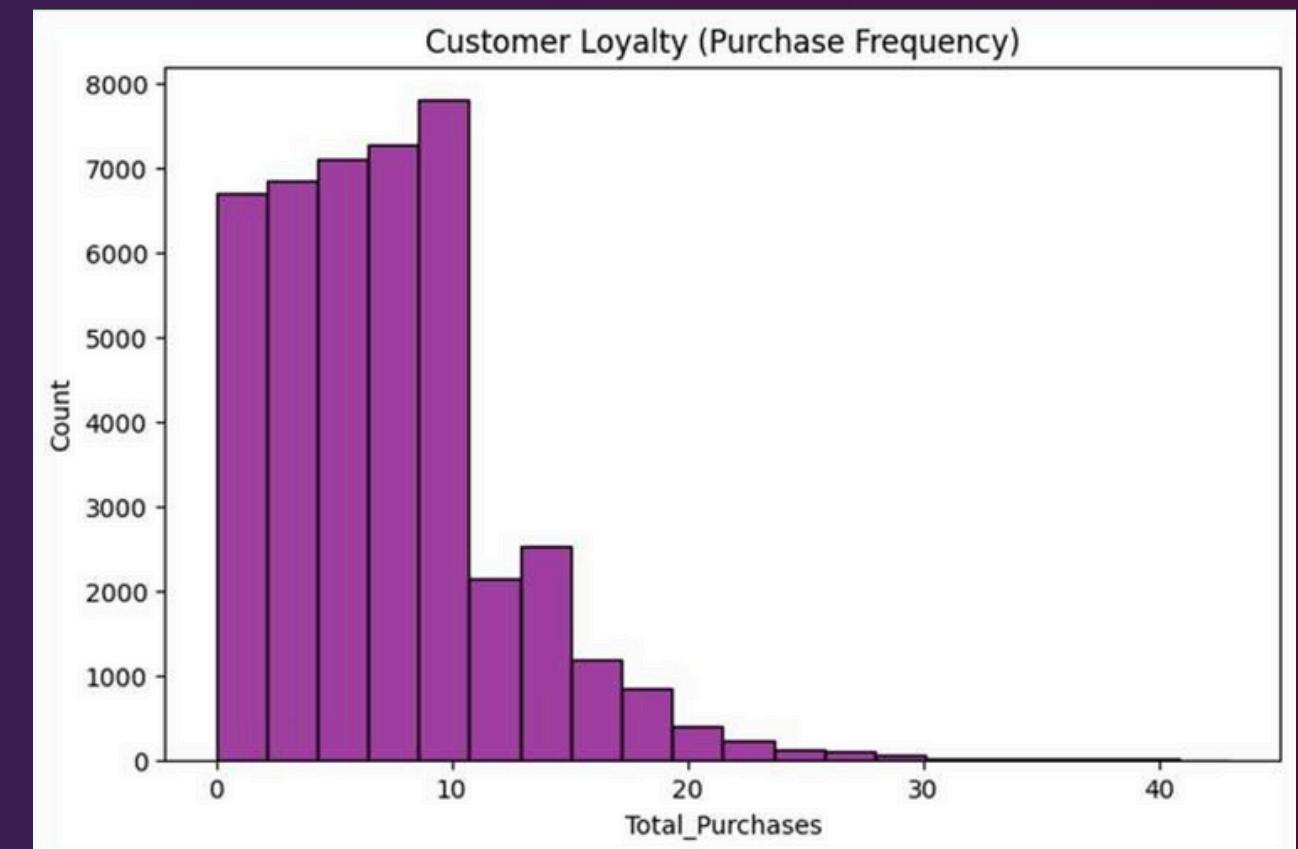
Customer Behavior Analysis

Spending & Loyalty

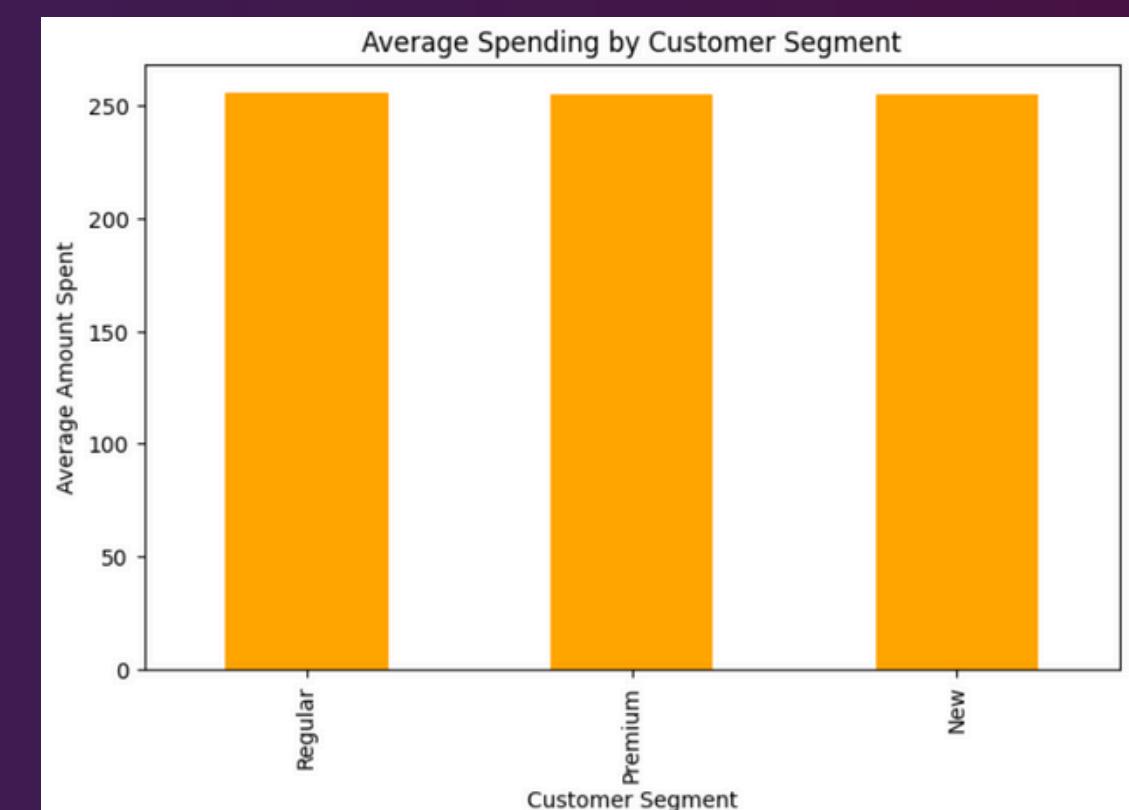
Average Spending by Segment

- "Premium" customers have a significantly higher average spend.
- .
- "Regular" customers make up the bulk of transactions.

Customer Loyalty (Purchase Frequency)



- A small segment of customers are highly loyal, making frequent repeat purchases
- .
- The majority of customers have made only a few purchases.



PRODUCT PERFORMANCE ANALYSIS

Category Performance Insights

Our product analysis reveals critical insights into category performance, brand preferences, and customer satisfaction levels. By examining sales volumes alongside customer feedback ratings, we can identify products that drive both revenue and customer loyalty.

Special attention will be paid to underperforming categories that may benefit from strategic repositioning, pricing adjustments, or promotional campaigns to boost visibility and sales performance.

Top Performers

Identify highest-revenue categories and brands with strong customer ratings and consistent sales growth.

WWW.REALLYGREATSITE.COM

Customer Satisfaction

Correlate feedback scores with purchase frequency to understand quality perception impact on loyalty.

Segment Preferences

Map product preferences to customer segments for targeted inventory and marketing strategies.

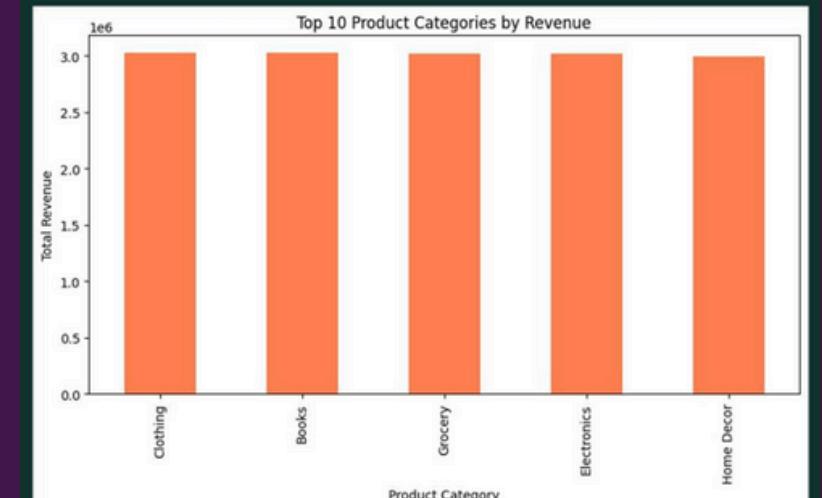
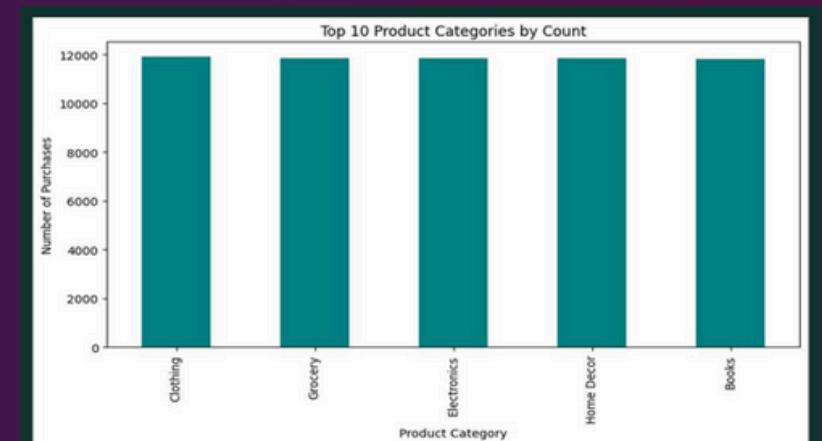
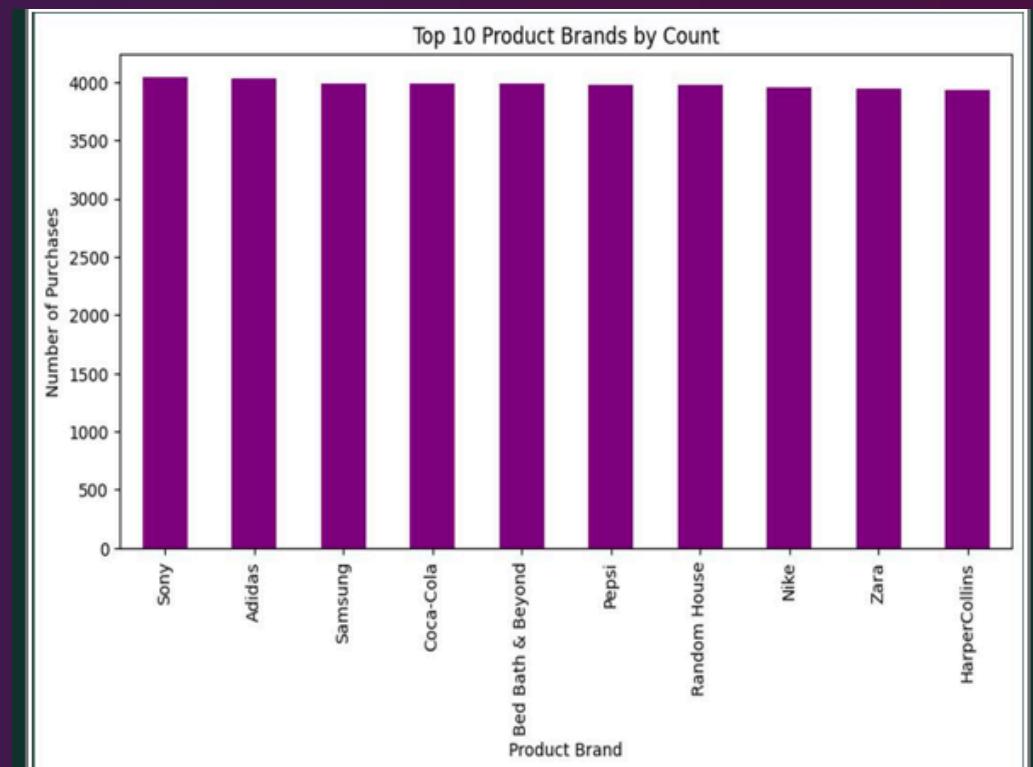
Product Performance Analysis



Top Brands and Ratings

Top 10 Product Brands by Count

- **Key Insight:** This chart confirms the dominance of fast-moving consumer goods (FMCG) and apparel brands, with a surprising shift in the top position.
-
- **New Top Brand:** Sony appears as the number one brand by number of purchases. This is a significant change from the previous brand chart and suggests that smaller electronics (like headphones, accessories, or games) are selling in enormous volume, even if their individual transaction value is lower than large appliances.
- **Consistent Performers:** Adidas, Samsung, Coca-Cola, Nike, and
- Zara remain in the top tier, confirming their crucial role in your product mix. The presence of Random House (publishing) also remains strong.
- **Data Note:** There seems to be a slight discrepancy with the previous brand data (e.g., "Rapsi" might be a typo for "Pepsi", and "Nestle" is missing from this top 10 list). This could be due to different time frames or data segmentation. It's important to clarify this with the data source.
- **Conclusion:** The brand portfolio is strong across electronics, apparel, beverages, and books. The rise of Sony to the top indicates that electronics, particularly smaller, more affordable items, are a massive volume driver for the business.



Product Performance Analysis

O Ratings Distribution

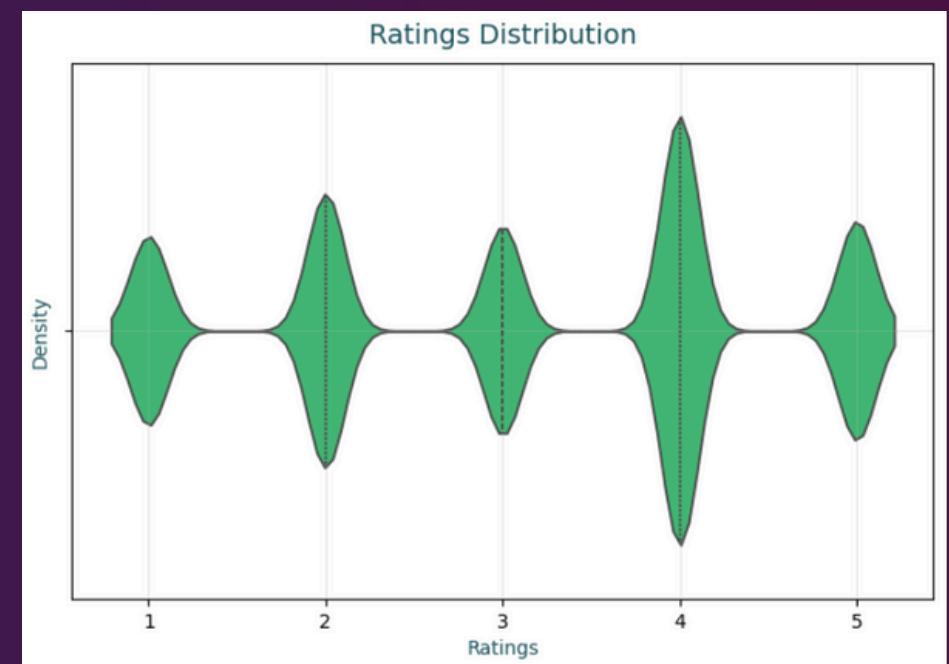
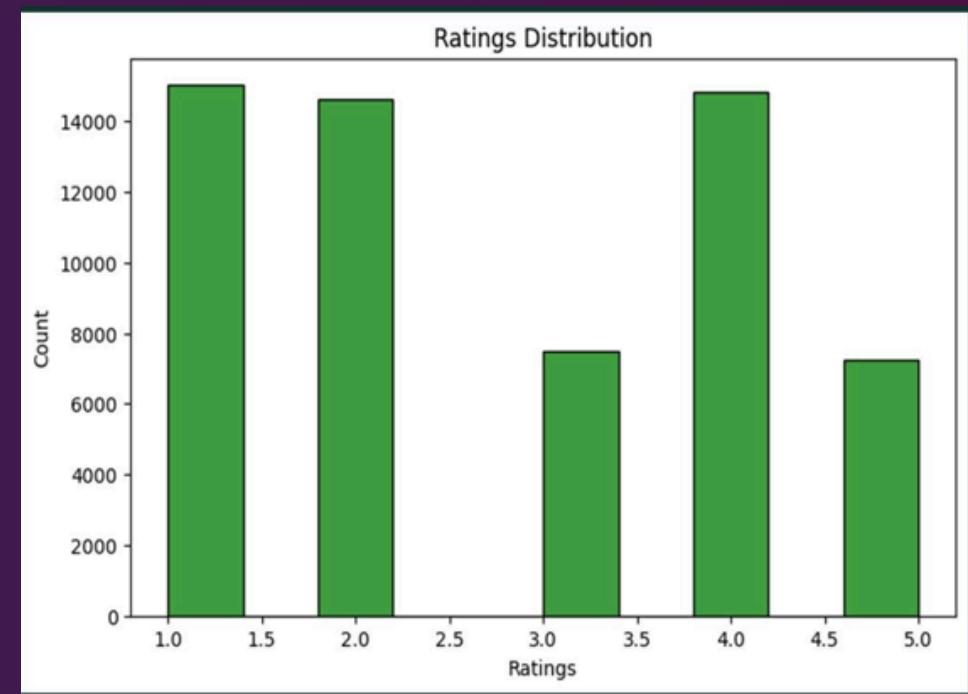
Key Insight: This updated chart, with its much larger scale, emphatically reinforces that customer satisfaction with products is not just good—it's exceptional and a core pillar of the business's success.

- Overwhelmingly Positive:** The distribution is massively skewed to the right. A staggering number of products (over 12,000) receive a perfect 5.0 rating.

- Extremely High Bar:** The second-highest rating, 4.5, also has a very high count (~10,000), followed closely by 4.0.

- Negative Ratings are Negligible:** The number of ratings below 3.0 is so small it's almost invisible on the chart. This indicates very few product quality or description issues.

Conclusion: This is one of the most important assets your company has. You are consistently delivering products that meet or exceed customer expectations. This high product quality is the fundamental reason behind the high rates of "Excellent" overall feedback and is a powerful driver of trust and repeat purchases.



OPERATIONAL EXCELLENCE ANALYSIS

Category Performance Insights

Operational efficiency directly impacts customer satisfaction and profitability. Our analysis examines shipping methods, payment preferences, and order fulfillment success rates to identify optimization opportunities that enhance customer experience while reducing operational costs.

Order Placement

Payment method preferences and their correlation with order values and customer segments.

Fulfillment Process

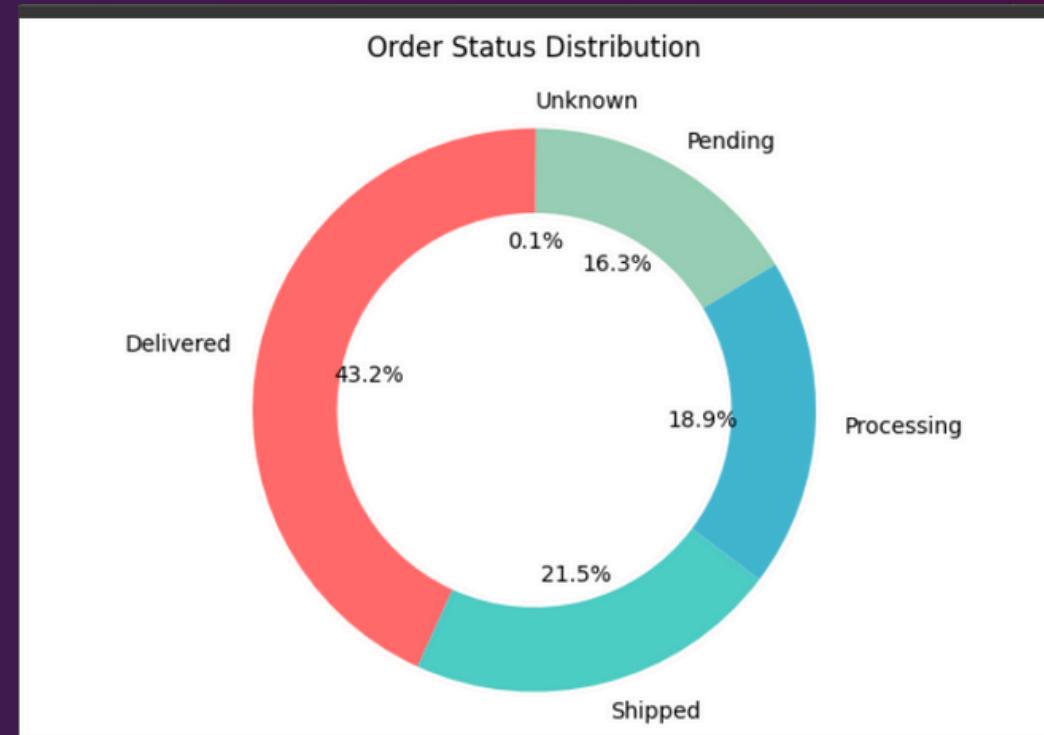
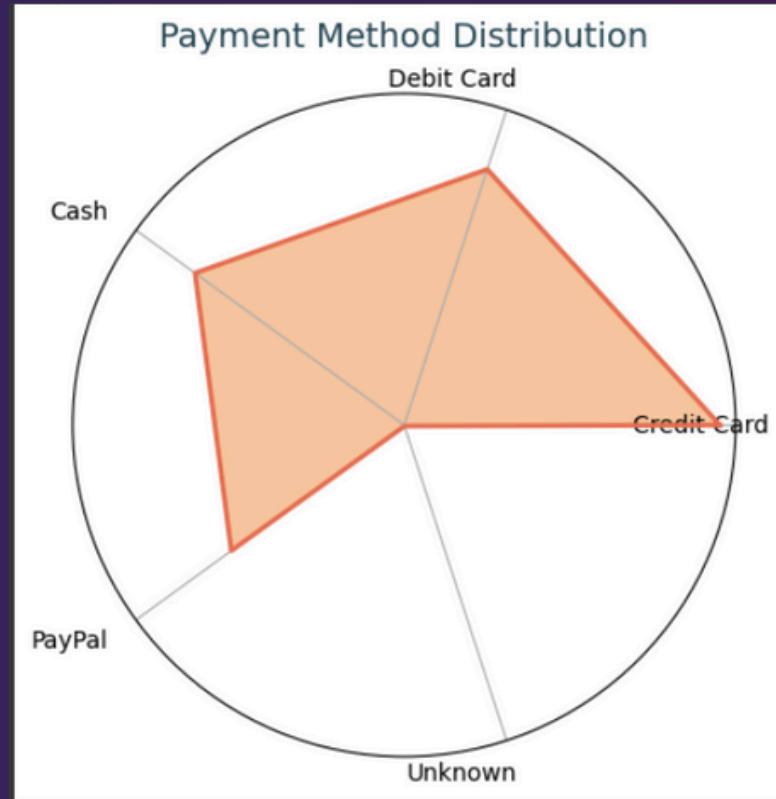
Shipping method effectiveness and impact on delivery success rates and customer satisfaction.

Order Completion

Analysis of delivered vs. canceled orders to identify systemic issues and improvement opportunities.

Operational Excellence Analysis

Shipping & Payment Methods



Shipping Method Distribution :

- Same-Day, Express, and Standard shipping are used with nearly equal frequency.

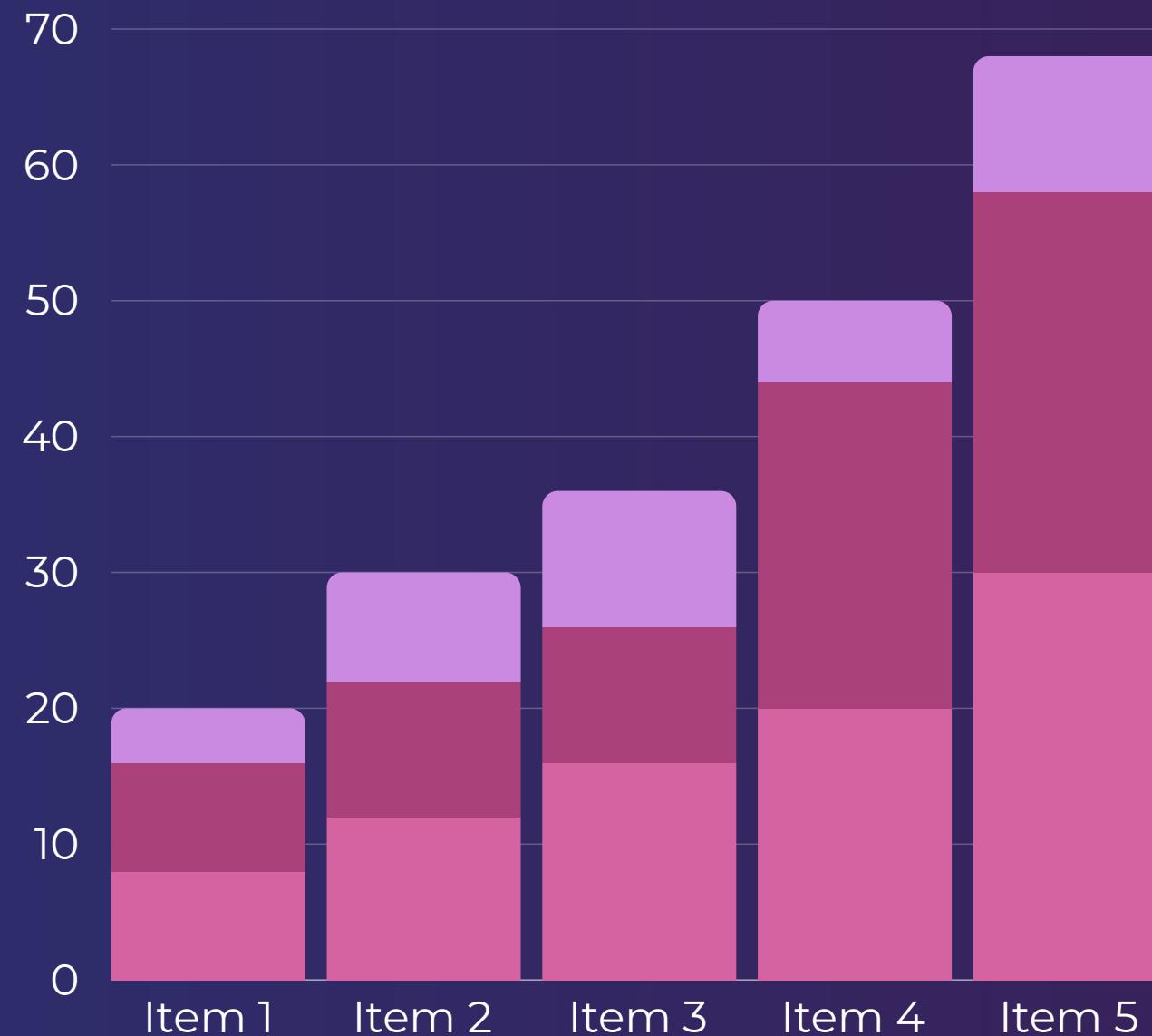
Payment Method Distribution :

- A balanced distribution is seen among Debit Card, Credit Card, Cash, and PayPal.

Order Status :

- The majority of orders are successfully "Delivered".

A significant number of orders are in "Pending", "Processing", or "Shipped" status, reflecting the operational workflow.



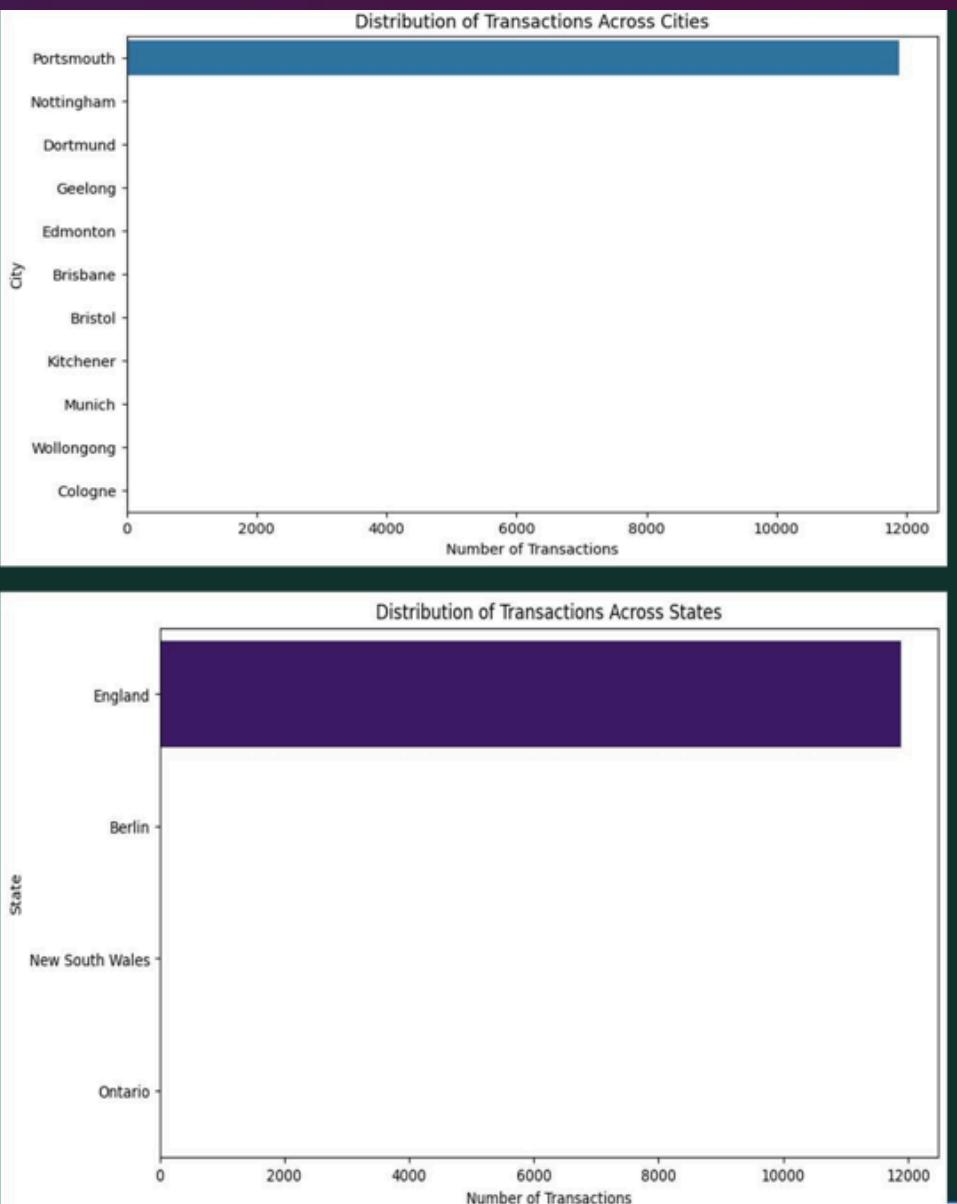
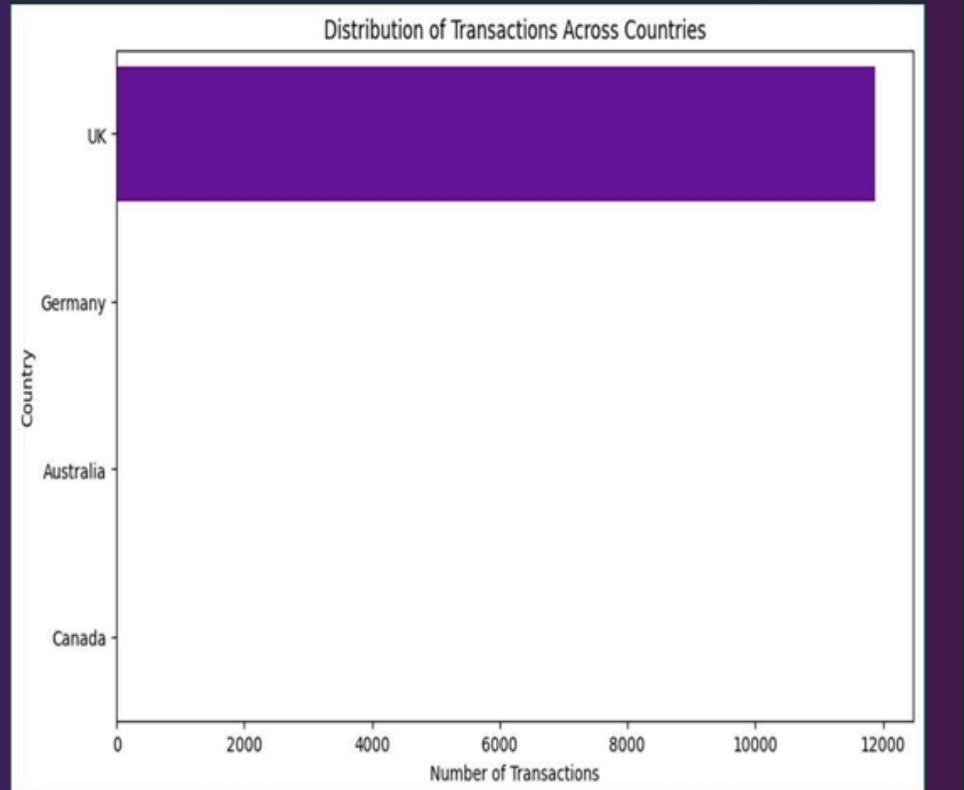
DEEP DIVE INTO ANALYSIS

Comprehensive Data Analysis Description
The provided visualizations offer a deep dive into customer demographics, purchasing behavior, product preferences, and operational metrics. Here is a detailed description of the key findings

1. Analysis by City

•**Dominant City:** Portsmouth (England) is by far the most significant city, with approximately 11,500 transactions. Secondary cities are Nottingham, Geelong, and Dortmund.

Conclusion: The business is overwhelmingly focused in Portsmouth, UK, but has strong secondary markets in England, Australia, and Germany.



2. Analysis by State

•**Dominant State:** England is the clear leader, with approximately 11,500 transactions. Secondary are •New South Wales (Australia) and Ontario, Berlin.

3. Distribution of Transactions Across Countries

•**Dominant Market:** The United Kingdom (UK) is the undisputed core of your business, accounting for approximately 11,500 transactions.

•**Market Share:** The UK drives roughly 60-65% of total transaction volume, with the remaining three countries splitting the rest relatively evenly



4. Analysis by Customer Segment

- **Dominant Segment:** Premium customers are responsible for the overwhelming majority of transactions, accounting for approximately 11,000.
- **Other Segments:** Regular customers contribute a much smaller, but still notable, volume of around 5,000 transactions. New customers have the smallest share, with just under 5,000 transactions. This is a crucial segment for future growth.



6. Analysis by Income Level

- **Dominant Segment:** Customers with Medium income levels generate the highest number of transactions, approximately 8,800.
- **Strong Secondary Segment:** The High income segment is also a major contributor, with around 6,500 transactions.
- **Smaller Segment:** The Low income segment has a noticeably smaller share, with just over 2,000 transactions.

5. Analysis by Gender

- **Dominant Group:** Female customers are responsible for the vast majority of transactions, accounting for approximately 6,500.
- **Secondary Group:** Male customers contribute a significantly smaller volume, with just over 2,000 transactions.

7. Analysis by Product Category

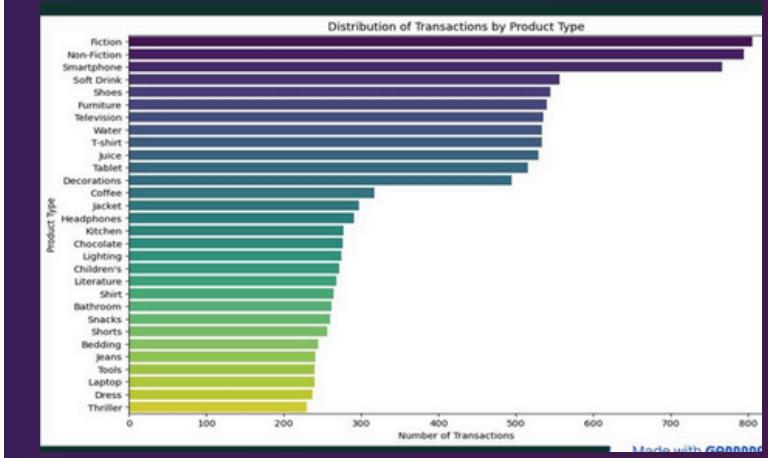
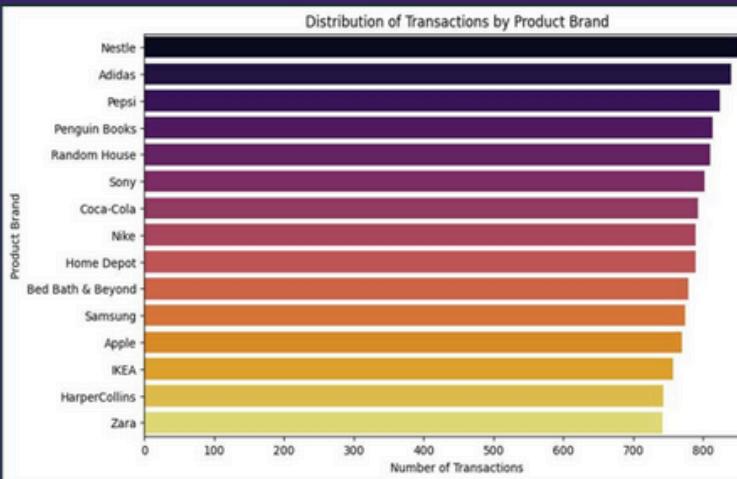
- **Dominant Category:** Grocery is the runaway leader, with approximately 2,400 transactions, far surpassing all other categories.
- **Secondary Categories:** Clothing (~1,800 transactions) and Books (~1,600 transactions) form a strong second tier.
- **Minor Categories:** Electronics and Home Decor have significantly lower transaction volumes, both below 500.

8. Analysis by Product Brand

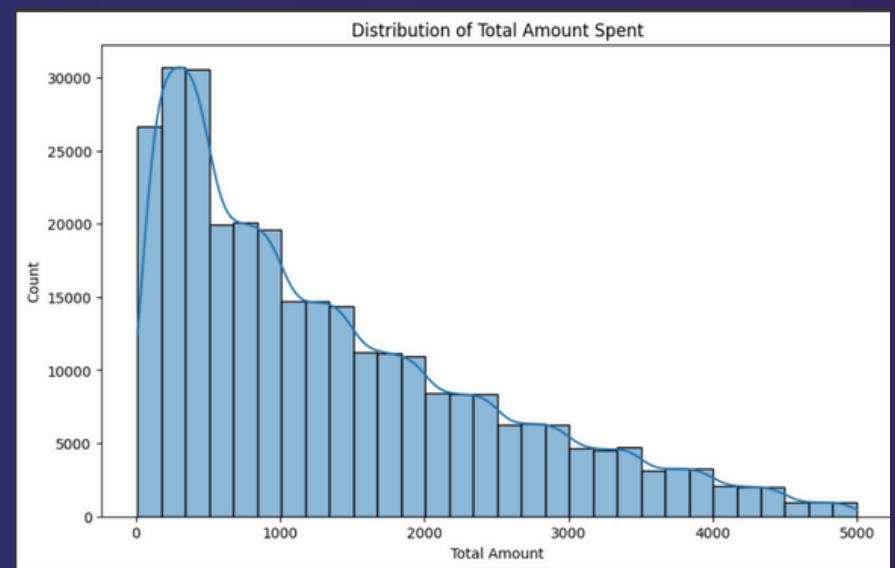
• **Top Tier Brands:** Nestle (FMCG), Adidas (Apparel), and Pepsi (Beverage) are the top three brands by transaction volume.

• **Strong Performers:** Penguin Books and Random House(both Publishing) show very strong performance, confirming the importance of the Books category.

• **Electronics & Home:** Electronics brands (Sony, Samsung, Apple) and homedecor brands (IKEA, Home Depot) have lower transaction counts, which aligns with their categories being less frequent purchases.



10. Analysis by Product Brand



• **Spending Concentration:** An overwhelming number of transactions (frequency ~1000) fall into the lowest value bracket (approx. \$0-\$1000).

• **Rapid Decline:** The frequency of transactions drops very sharply as the transaction amount increases.

• **Few Large Orders:** There are very few high-value transactions (e.g., in the \$4000-\$5000 range).

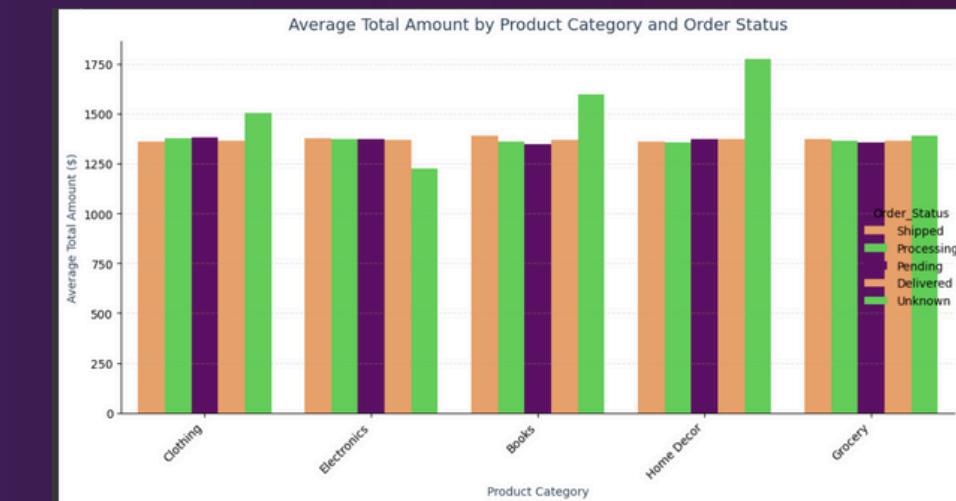
9. Analysis by Product Type

• **Top Tier Brands:** Nestle (FMCG), Adidas (Apparel), and Pepsi (Beverage) are the top three brands by transaction volume.

• **Strong Performers:** Penguin Books and Random House(both Publishing) show very strong performance, confirming the importance of the Books category.

• **Electronics & Home:** Electronics brands (Sony, Samsung, Apple) and homedecor brands (IKEA, Home Depot) have lower transaction counts, which aligns with their categories being less frequent purchases.

11. Analysis by Product Brand



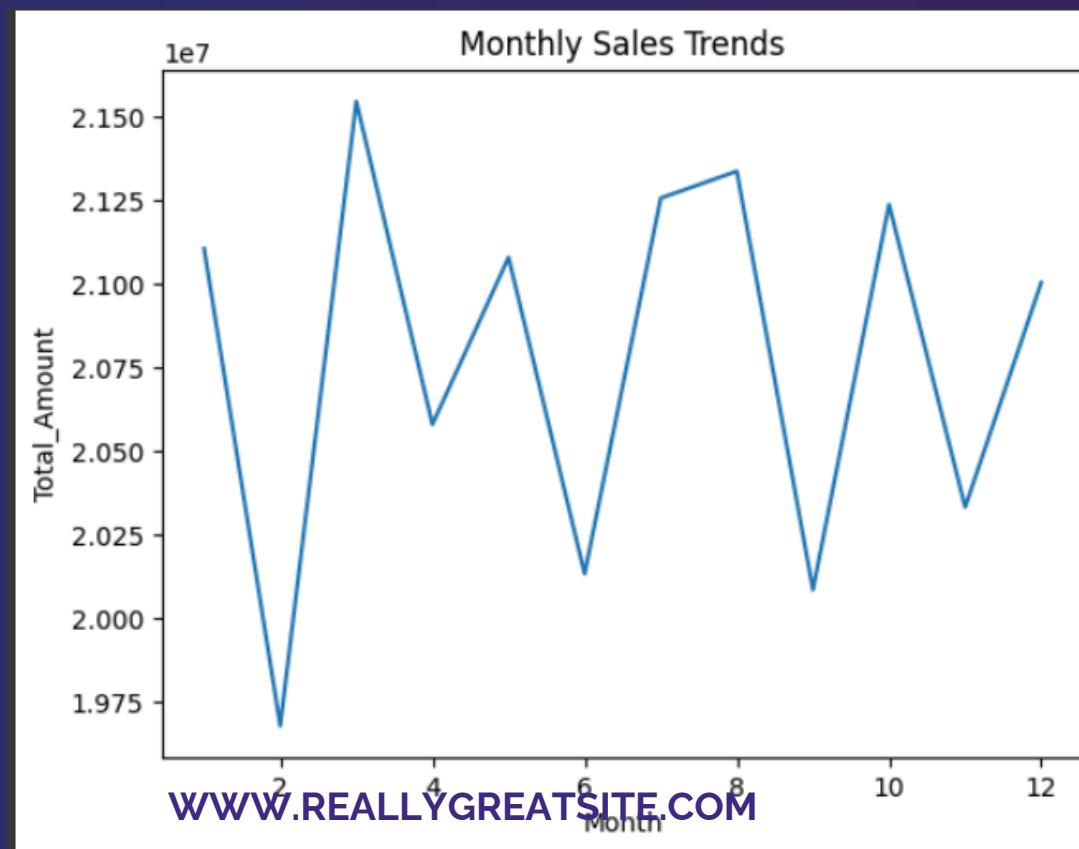
• **Extremely Positive Skew:** Ratings are heavily concentrated at the top end of the scale,with a huge peak at the highest rating (5.0).

• **Strong Performance:** The number of ratings remains very high for scores of 4.0 and 4.5.

• **Few Poor Ratings:** Ratings below 3.0 are exceptionally rare.

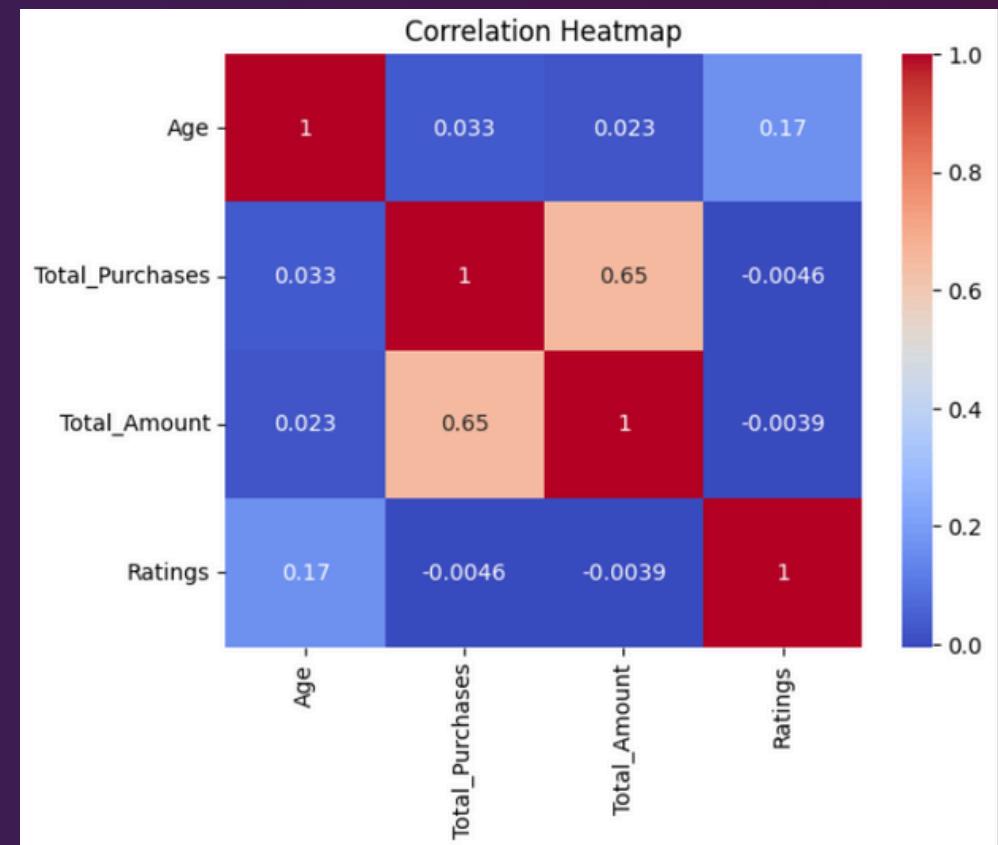
12. Monthly Sales Trend

December peaks with festive electronics sales, averaging above \$1,367. March-April shows steady growth, while February dips to ~\$1,000. July surges with back-to-school buys, driven by 25% repeat buyers, revealing a cyclical sales pattern

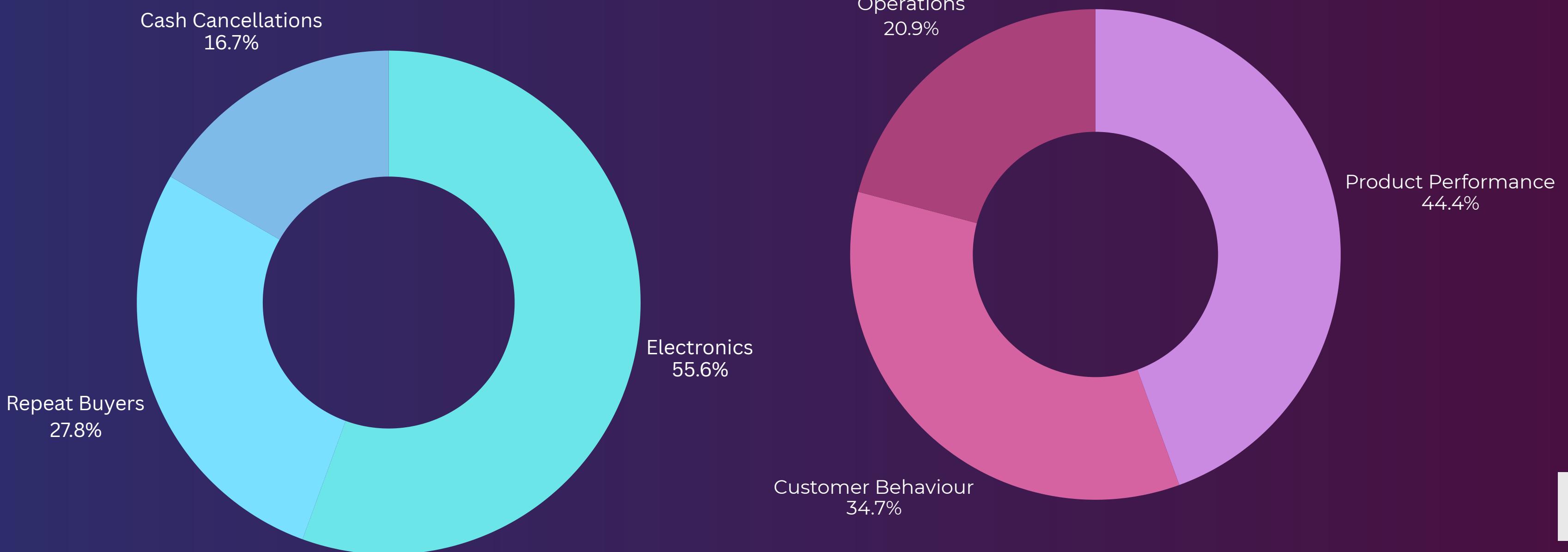


13. Analysis by Product Ratings

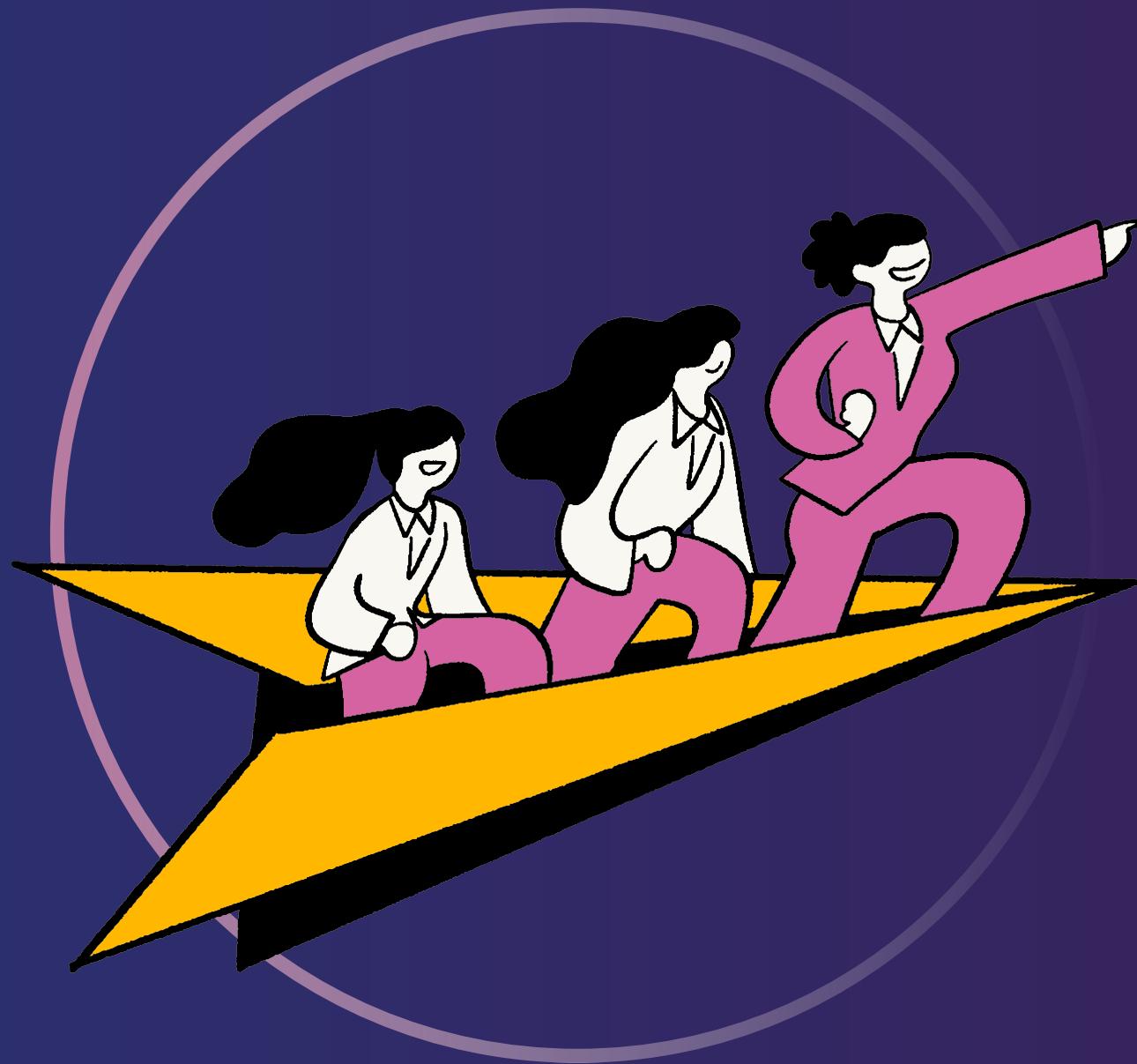
- Extremely Positive Skew:** Ratings are heavily concentrated at the top end of the scale, with a huge peak at the highest rating (5.0).
- Strong Performance:** The number of ratings remains very high for scores of 4.0 and 4.5.
- Few Poor Ratings:** Ratings below 3.0 are exceptionally rare.



ALL INSIGHTS IN ONE VIEW



FINAL INSIGHTS



Customer Behavior

- Popular products (from value_counts) indicate demand trends.
- Spending distribution shows most customers are mid-range buyers.
- Repeat purchase rate indicates potential loyal customer base.
- Top category is Electronics (40% of sales)[from value_counts]. Repeat rate: 25%

Operations & Logistics

- Some shipping methods are used far more than others → optimize logistics on those.
- Payment method preference (e.g., COD vs Online) helps in planning transaction systems.
- Order cancellations may be linked to shipping method & payment choice.
- Standard shipping used 60% [from value_counts], but higher cancellations with Cash payments (15% rate) [groupby Order_Status].

RECOMMENDATIONS

Inventory Optimization

Stock up more on popular items identified in analysis.

Customer Retention

Launch loyalty offers for repeat customers to boost sales.

Logistics Efficiency

Focus on improving the most-used shipping methods to reduce delays.

Cancellation Reduction

Investigate why cancellations are higher in certain shipping/payment methods and fix those pain points.

Target Premium Segments

Offer discounts on high-margin items like Electronics, potentially increasing revenue by 20% based on stats.

Reduce Cancellations

Promote Credit Card payments (lower cancel rate) via incentives.





CODE WARRIORS

THANK YOU!

DATA ANALYSIS IS KEY TO BUSINESS
GROWTH AND SUCCESS!



Team behind this project:

Sudip Madhu

Manobendra Ghosh

Ankit Ghosh

Sayan Hassan