

Customer Behavior and Shopping Habits - Project Report

Analyst: Shijin Ramesh

Dataset Source: Kaggle - Consumer Behavior and Shopping Habits Dataset

Tools Used: Python, R, Jupyter Notebook, RPubS, SQL

Objective:

To analyze consumer shopping behavior and identify demographic and behavioral patterns that impact high-value transactions and payment method preferences. The insights aim to assist retail stakeholders in optimizing their marketing strategies and personalizing offers based on consumer profiles.

Key Analyses Conducted & Observations:

Overall Data Reliability & Customer Demographics

- The dataset is **clean and reliable for analysis**, with no null values, appropriate data types, and no inconsistencies or duplicates found.
- **Males constitute the majority (68%) of the customer base**¹. This demographic insight is crucial for tailoring marketing and product strategies.
- The **mid-life and pre-retirement age groups (25–64) are the dominant purchasing segments, accounting for 76.7% of the total shopper base**². Younger consumers (18–24, 12.5%) and older consumers (65–70, 10.9%) represent smaller segments, possibly due to limited disposable income or more cautious spending habits.

Product & Category Performance

- **Clothing is by far the most preferred category, accounting for 44.5% of all purchased items**, significantly outpacing outerwear (8.3%). This highlights clothing as a core revenue driver.
- **The top five purchased items (Blouse, Jewelry, Pants, Shirt, Dress) and least five (Sneakers, Boots, Backpack, Gloves, Jeans) show relatively small variance in purchase counts**, indicating a balanced demand across the product range rather than extreme polarization in popularity.
- **Medium (M) and Large (L) sizes are consistently the most purchased, with high demand for colors like Olive, Violet, Yellow, and Green**. Conversely, XL sizes show the

least demand overall. This insight can optimize inventory management and promotional strategies.

Customer Behavior & Engagement

- **Customer purchases remain steady throughout the year, with no significant seasonal fluctuations.** This indicates a consistent, year-round demand for products.
- **PayPal, Credit Card, and Cash are the most commonly used payment methods,** while Bank Transfer is the least preferred⁶. Marketing efforts could promote preferred methods like PayPal for enhanced convenience.
- **There is a highly engaged customer segment, with 13.8% buying weekly and 27.9% purchasing bi-weekly,** alongside a substantial portion (29.4%) who shop quarterly⁷. This indicates potential for high purchasing power or business-related needs (e.g., resellers, influencers).
- **The customer base is diverse, with a healthy mix of customers across all levels of past purchases,** suggesting no over-reliance on a single type of shopper.
- **A significant opportunity exists in converting the 73% of customers who have not opted for a subscription,** by effectively showcasing its value and benefits.
- **Free shipping is the most preferred shipping method (17.3%),** emphasizing its importance as a competitive advantage.
- **Approximately 57% of customers do not apply discounts or promo codes,** suggesting a segment with strong purchasing intent even without incentives. This also presents an opportunity to increase engagement and sales through more visible, personalized, or accessible promotions.
- **Promo codes, when used, show only a marginal difference in average purchase amount** (Yes: \$59.28 vs No: \$60.13), suggesting they are not currently driving significantly higher spend per transaction.

High-Value Customers & Customer Experience

- **A small group of 50 top customers (representing only 1.28% of the customer base) generates a disproportionate 2.5% of total Customer Lifetime Value (CLV).** This segment is critical for long-term revenue.

- A key missed opportunity is that **only 13 of these top 50 customers are subscribed to the platform**, highlighting potential for targeted efforts to encourage subscription among the remaining 37 high-value non-subscribers.
- A significant concern is that **29 out of these 50 most valuable customers gave a review rating below 4.11**. This indicates a critical need for the Customer Relationship team to proactively engage with these high-value customers to address concerns and improve their experience.
- Overall customer experience requires substantial improvement, as a majority of users (both male and female) across all categories, particularly Clothing and Accessories, provided ratings below 4 (negative sentiment).
- While customer review ratings are mostly in the moderate-to-good range (2.75–4.75), **there is clear room for improvement to consistently achieve higher ratings (4.5 and above)**.
- Immediate actions to improve customer satisfaction include conducting product quality checks, gathering qualitative feedback (e.g., via surveys), improving product descriptions, imagery, and sizing guides to manage expectations, and offering personalized follow-ups or offers to those who left negative feedback.
- Given that **male customers contribute a higher volume of ratings, loyalty programs or satisfaction follow-ups could be slightly tailored towards them**, while also implementing targeted campaigns for female customers to enhance long-term retention.

Strategic Implications

- Underperforming geographical regions like Rhode Island, Kansas, Hawaii, Arizona, and New Jersey represent significant growth opportunities for marketing and sales teams to boost engagement and sales through tailored strategies.
- There is no meaningful correlation between age, review rating, or previous purchases and the purchase amount. This suggests that purchase amount is not strongly predictable by these individual factors alone, and other variables or a more complex model might be needed for prediction.
- Strategic marketing opportunities exist to retain and reward high-frequency shoppers through loyalty programs, exclusive discounts, or referral campaigns, and by exploring collaboration opportunities with influencers for wider reach.

R Statistical Analysis

- ANOVA and Chi-Square test was conducted to analyze categorical data.
 - A ANOVA result (p-value = 0.225), suggests that there is no statistically significant difference in average purchase amount across categories.
 - Chi-square test (with p-value less than 0.05) showed that there is a statistically significant relationship between subscription status and whether the customer used a promo code. Subscribers and non-subscribers behave differently when it comes to using promo codes.
-

Recommendations / Action Plan for Stakeholders:

1. **Subscription Focus:**
 - Launch targeted campaigns for high-value, non-subscribed users to boost long-term retention. Promoting the benefit of subscriptions more clearly to non-subscribers would be recommended.
2. **Customer Recovery Plan:**
 - Prioritize engagement with low-rating high-value customers through personalized outreach.
3. **Product & Experience Enhancement:**
 - Improve quality checks, visuals and sizing guidance to boost review scores.
4. **Promotions Strategy:**
 - Increase visibility of promo codes and tailor them to customer segments for better adoption.
5. **Loyalty & Engagement:**
 - Introduce tier-based loyalty programs for frequent buyers and exclusive perks for top spenders.
6. **Geographic Expansion:**
 - Tailor marketing efforts to low-performing regions like Kansas, Rhode Island and Hawaii.

Deliverables:

- Python notebook (.ipynb): All cleaning, EDA and visualizations
 - R markdown/report: Chi-square tests and additional plots (RPubs link)
 - Project report (.pdf/Word): Summary of business insights and action plan
 - GitHub README file
-

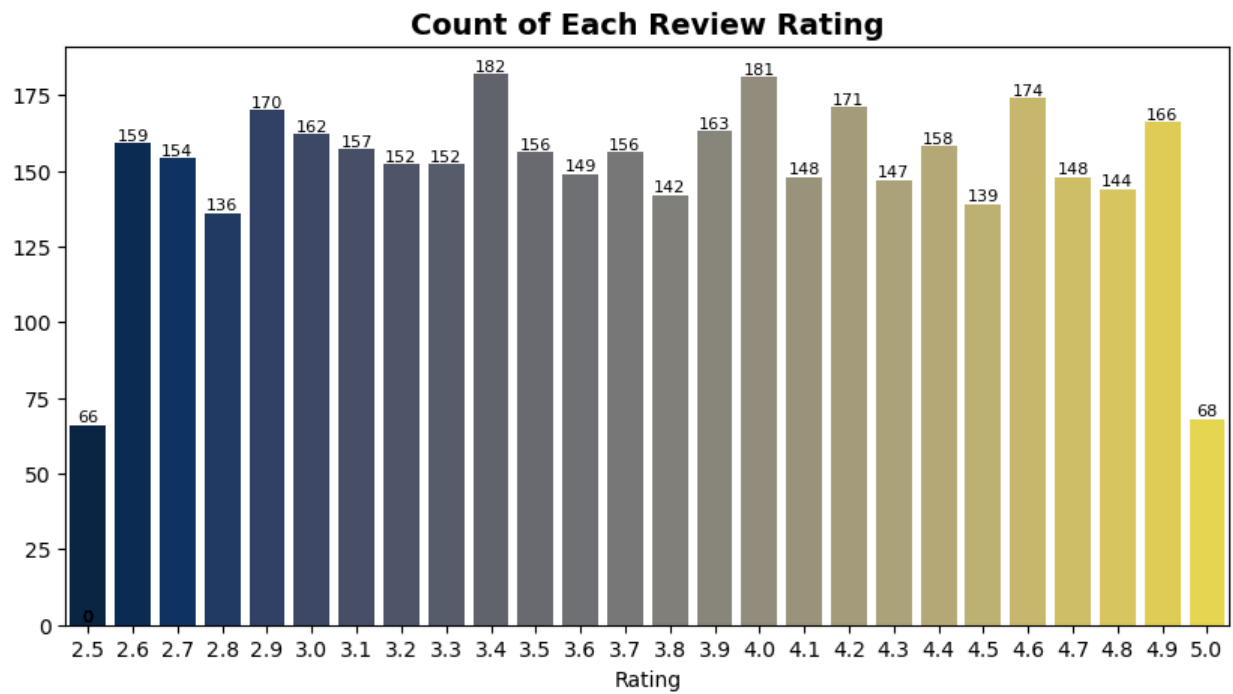
Live Report Link (R):

View Report on RPub: http://rpubs.com/shijinramesh/customerbehavior_statisticalanalysisi

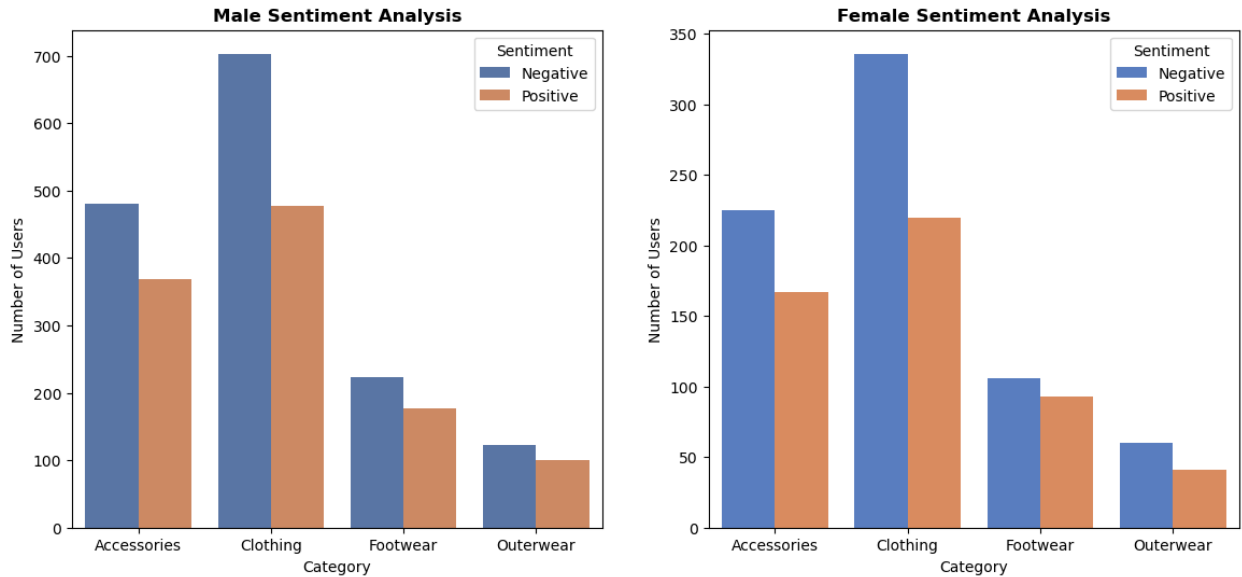
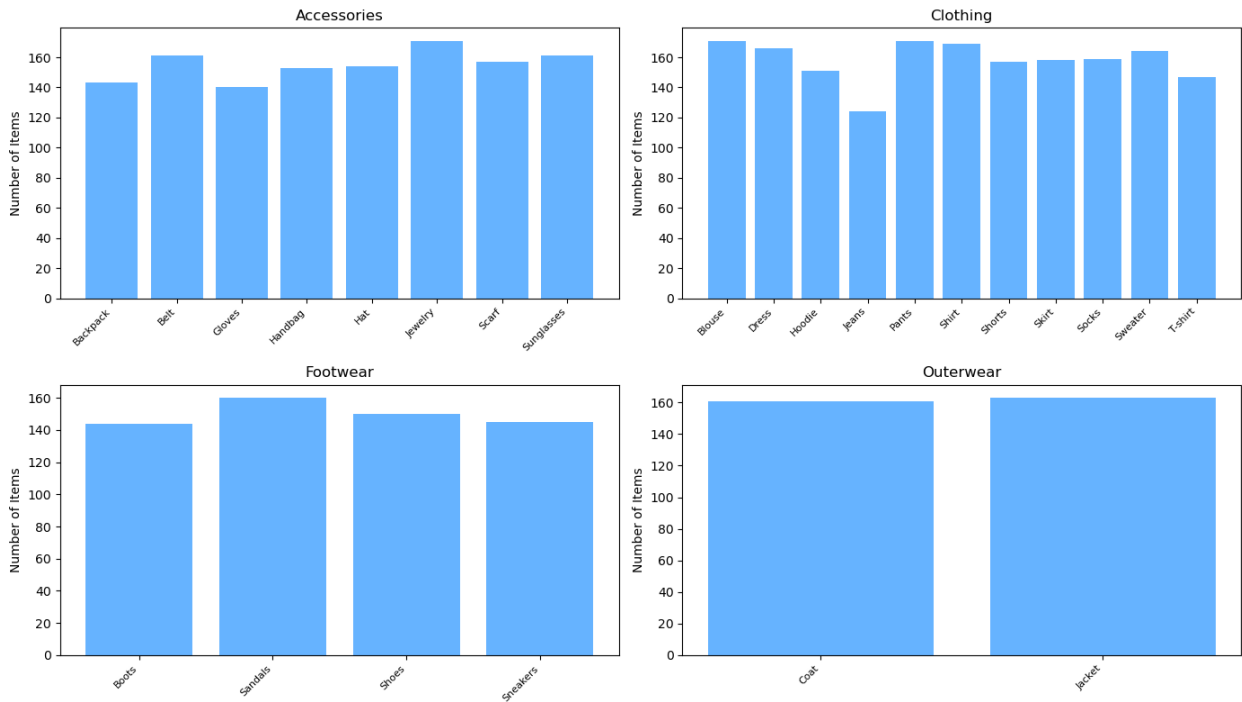
Estimated Completion Time: 4 days

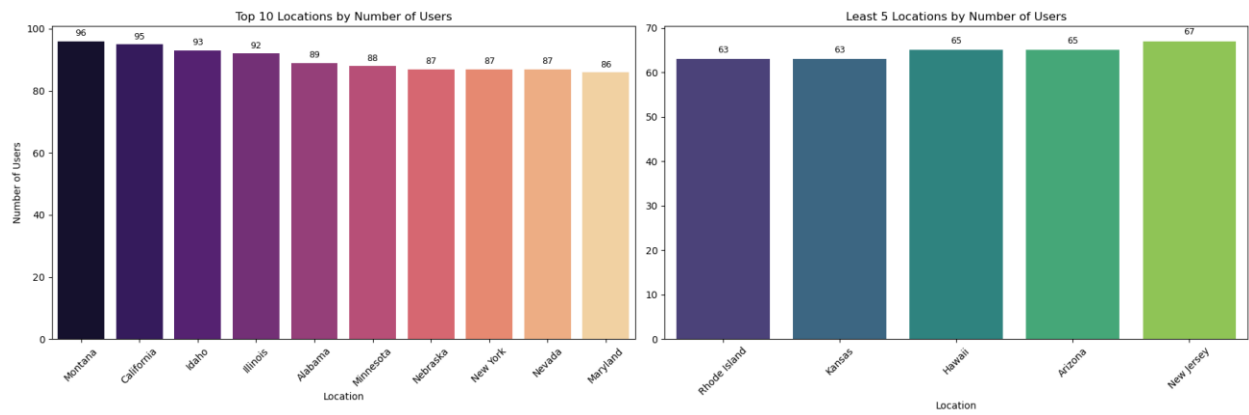
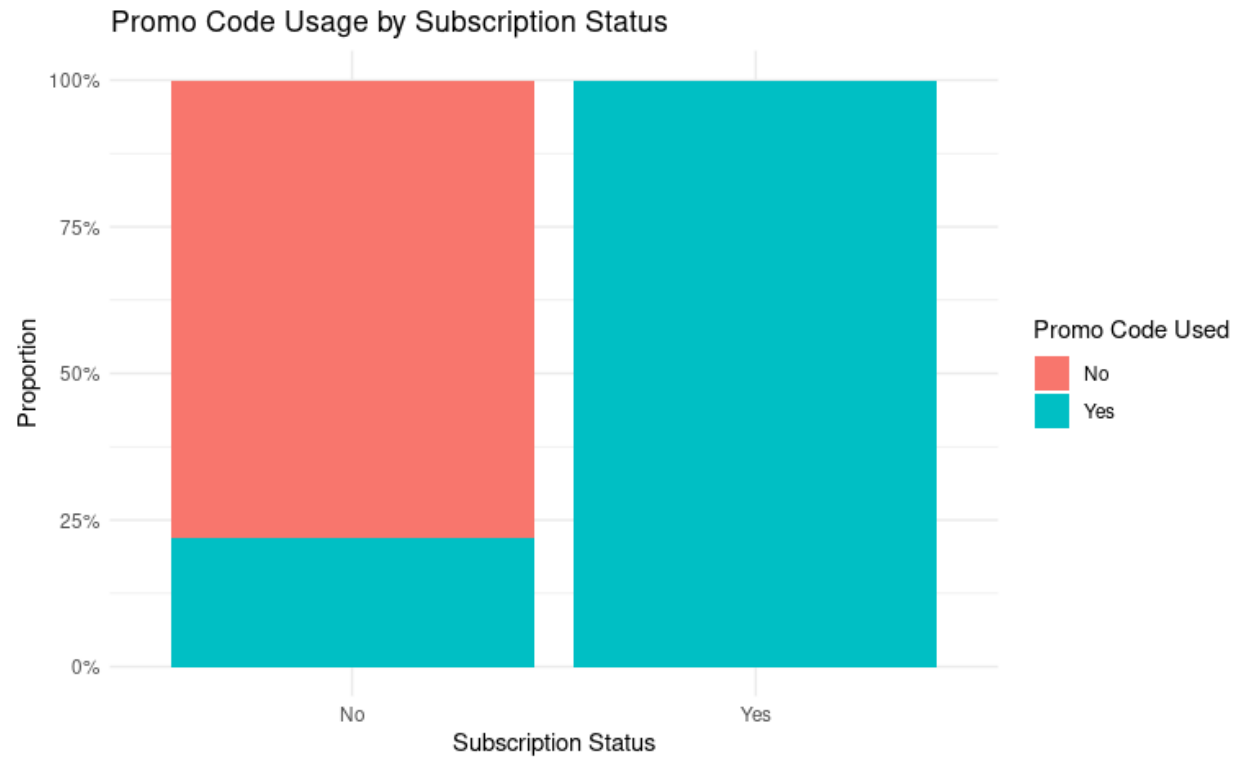
Date: 4th July, 2025

Sample Visuals



Number of Items Purchased per Category





Category Distribution

