Customer Shopping Behavior Analysis

1. Project Overview

This project examines customer purchasing patterns across 3,900 transactions spanning multiple product categories. The objective is to identify meaningful insights regarding customer spending habits, segment characterization, product preferences, and subscription trends to inform strategic business decisions.

2. Dataset Summary

- **Total Records:** 3,900 transactions
- **Feature Count:** 18 attributes
- **Customer Attributes:** Age, Gender, Location, Subscription Status
- **Transaction Details:** Item Purchased, Category, Purchase Amount, Season, Size, Color
- **Behavioral Metrics:** Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type
- **Data Quality:** 37 missing values in Review Rating column

3. Data Preparation with Python

The initial phase involved data cleaning and transformation using Python:

- **Data Import:** Loaded the dataset using **pandas** DataFrame
- **Structural Assessment:** Examined data types and statistical summaries using **df.info()** and **.describe()**
- **Missing Value Treatment:** Identified 37 null entries in the **Review Rating** column and imputed using median values by product category
- **Naming Convention:** Standardized all column names to **snake_case** format for consistency
- **Feature Engineering:** Created **age_group** variable through binning of customer ages and derived **purchase_frequency_days** from transaction timestamps
- **Data Redundancy Check:** Evaluated if **discount_applied** and **promo_code_used** contained overlapping information; removed **promo_code_used** to eliminate redundancy
- **BigQuery Integration:** Configured Python connection to BigQuery and migrated the cleaned DataFrame to the cloud database for advanced analysis

4. Business Intelligence Analysis with BigQuery

Structured SQL queries were executed in BigQuery to extract actionable business insights:

1. **Revenue by Gender:** Comparison of total revenue contributions between male and female customer segments

Row	gender	revenue
1	Male	157890
2	Female	75191

2. **High-Value Discount Users:** Identification of customers who leveraged discounts while maintaining above-average spending levels

Row	customer_id	purchase_amount
1	106	96
2	319	67
3	1005	79
4	1170	94
5	1428	90
6	1006	94
7	1090	61
8	1530	63

3. **Top-Performing Products:** Discovery of 5 products with the highest customer review ratings

Row	item_purchased	avg_product_rating
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.8
5	Handbag	3.78

4. **Shipping Method Comparison:** Analysis of average purchase amounts across Standard and Express shipping options

Row	shipping_type	avg_purchase_a...
1	Express	60.48
2	Standard	58.46

5. **Subscription Impact:** Evaluation of average spend and aggregate revenue between subscribers and non-subscribers

Row	subscription_status	total_customers	avg_spend	total_revenue
1	Yes	1053	59.49	62645.0
2	No	2847	59.87	170436.0

6. **Discount-Reliant Products:** Ranking of 5 products with the greatest proportion of discounted sales

Row	item_purchased	discount_rate
1	Hat	50.0
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. **Customer Lifecycle Segmentation:** Classification into New, Returning, and Loyal customer groups based on purchase history

Row	customer_segment	number_of_customers
1	Returning	701
2	Loyal	3116
3	New	83

8. **Category Leaders:** Identification of top 3 most-purchased items within each product category

Row	item_rank	category	item_purchased	total_orders
1	1	Accessories	Jewelry	171
2	2	Accessories	Belt	161
3	3	Accessories	Sunglasses	161
4	1	Clothing	Pants	171
5	2	Clothing	Blouse	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150

9. **Subscription Propensity:** Examination of whether repeat customers (5+ purchases) show higher subscription rates

Row	subscription_status	repeat_buyers
1	Yes	958
2	No	2518

10. **Age Group Revenue Contribution:** Breakdown of total revenue by age cohort

Row	age_group	total_revenue
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763

11. **Revenue by Category-Size Mix:** Average purchase and total revenue by category–size, highlighting the top 3 revenue-driving combinations.

Row	avg_purchase_amount	total_revenue	category	size
1	60.46	47041	Clothing	M
2	59.25	33299	Accessories	M
3	57.93	27864	Clothing	L

12. **Location Revenue Efficiency:** Total revenue and revenue per customer by location, surfacing locations with the highest revenue per customer.

Row	location	unique_customers	total_revenue	avg_revenue_per_customer
1	Alaska	72	4867	67.6
2	Pennsylvania	74	4926	66.57
3	Arizona	65	4326	66.55
4	West Virginia	81	5174	63.88
5	Nevada	87	5514	63.38
6	Washington	73	4623	63.33
7	North Dakota	83	5220	62.89
8	Virginia	77	4842	62.88
9	Utah	71	4443	62.58
10	Michigan	73	4533	62.1
11	Tennessee	77	4772	61.97
12	New Mexico	81	5014	61.9
13	Rhode Island	63	3871	61.44
14	Texas	77	4712	61.19
15	Arkansas	79	4828	61.11

13. **Payment Method High-Value Share:** Revenue share from top 20% high-value customers versus others for each payment method.

Row	payment_method	customer_segment	segment_revenue	total_revenue_payment_method	revenue_share_within_payment_method
1	Bank Transfer	High Value	11258	36544	0.3080669877408056
2	Bank Transfer	Other	25286	36544	0.69193301225919435
3	Cash	High Value	11892	40002	0.29728513574321286
4	Cash	Other	28110	40002	0.7027148642567872
5	Credit Card	High Value	12690	40310	0.31481022078888615
6	Credit Card	Other	27620	40310	0.6851897792111139
7	Debit Card	High Value	14359	38742	0.37063135615094728
8	Debit Card	Other	24383	38742	0.62936864384905267
9	PayPal	High Value	13110	40109	0.32685930838465183
10	PayPal	Other	26999	40109	0.6731406916153481 7
11	Venmo	High Value	12545	37374	0.33566115481350672
12	Venmo	Other	24829	37374	0.66433884518649333

5. Interactive Dashboard in Power BI

6. Strategic Recommendations

1. Unlock Subscription Value Through Loyalty

Current Gap: Repeat buyers (5+ purchases) remain largely unsubscribed (73% non-subscribers)

Recommended Actions:
• Target high-frequency Loyal (3,116) and Returning (701) customers with subscription campaigns emphasizing frequency benefits, not just per-order discounts
• Design tiered subscription offerings with exclusive benefits tied to purchase frequency milestones and member-only events

Expected Business Impact: Increase subscription conversion by 25–40% among existing high-value segments, driving incremental recurring revenue and improving customer lifetime value (CLV)

2. Protect Margins on Discount-Heavy Products

Current Issue: Hat, Sneakers, Coat, Sweater, Pants all carry 47–50% discount rates, eroding profitability

Recommended Actions:

• Gradually test reduced discount depth or more targeted promotions on highly discount-dependent items
• Shift promotional budget from broad price cuts to value-added offers (bundles, loyalty points, free shipping) for products with strong demand or high ratings
• Implement selective, strategic price increases on key SKUs to normalize margin expectations

Expected Business Impact: Improve gross margin by 3–5 percentage points without sacrificing sales volume; reduce customer dependency on discounts

3. Hero Products Drive Campaign & Merchandising

Current State: Highest-rated products (Gloves 3.86, Sandals 3.84, Boots 3.82) and category leaders (Jewelry, Pants, Blouse, Sandals) demonstrate strong customer preference

Recommended Actions:
• Feature top-performing SKUs prominently in homepage banners, email campaigns, and recommendation carousels
• Create bundled offerings pairing hero products with complementary items (e.g., Clothing M/L combinations, Accessories pairings)
• Use hero products as anchor items in cross-sell and upsell strategies to drive higher basket sizes

Expected Business Impact: 15–20% uplift in featured product sales; increased average order value (AOV) and improved category mix

4. Geographic & Demographic Precision Marketing

Current State: Revenue efficiency varies by location (Alaska, Pennsylvania, Arizona: $64–68 per customer); Young Adults lead revenue ($62,143) but all age groups contribute meaningfully

Recommended Actions:
• Pilot localized campaigns in high-efficiency states with tailored messaging and promotional strategies
• Develop age-cohort-specific creative and messaging for Young Adults, Middle-aged, Adult, and Senior segments
• Allocate incremental marketing budget to high-ROI geographies and test expansion in adjacent regions

Expected Business Impact: Reduce customer acquisition cost (CAC) by 20%; increase return on ad spend (ROAS) by 25–30%; improve marketing efficiency and campaign relevance

5. Monetize Express Shipping Users

Current State: Express shipping users demonstrate 3.5% higher average order value ($60.48 vs $58.46), indicating premium customer segment

Recommended Actions:
• Bundle premium and convenience-focused products with Express delivery options
• Create a "fast-track" loyalty tier offering expedited shipping, early access to new products, and exclusive deals
• Use personalized messaging emphasizing convenience, urgency, and time-sensitive offers for this segment

Expected Business Impact: 2–3% incremental AOV from Express segment; increase subscription adoption among fast-track members

6. Rebalance Gender Revenue Mix

Current Imbalance: Male customers generate 2.1x female revenue ($157,890 vs $75,191), indicating significant untapped growth opportunity

Recommended Actions:
• Conduct assortment audit of female-focused products and identify gaps or under-represented categories
• Launch targeted female-focused campaigns with influencer partnerships and lifestyle-aligned messaging
• Test female-specific promotions, colors, sizing options, and product recommendations

Expected Business Impact: Narrow male-to-female revenue gap to 1.5x; grow female customer segment revenue by $30–40k annually

7. Leverage Payment Method Affinity for High-Value Acquisition

Current State: High-value customers comprise 30–37% of revenue across payment methods; Debit Card (37%), PayPal (33%), and Venmo (34%) show highest concentration

Recommended Actions:
• Introduce method-specific incentive programs (cashback, bonus loyalty points, exclusive discounts) for high-concentration payment types
• Create payment-method-specific offers in checkout experience to reinforce high-value behavior
• Monitor payment method trends and revenue shifts via dashboard to identify emerging opportunities

Expected Business Impact: Increase high-value segment revenue share by 5–10% year-over-year; strengthen customer retention among premium segments

8. Inventory & Pricing Strategy

Current State: Top revenue-driving category-size combinations are Clothing M ($47k), Accessories M ($33k), and Clothing L ($28k)

Recommended Actions:
• Prioritize inventory allocation and stock availability for high-revenue category-size mixes
• Test dynamic pricing strategies on leading SKU combinations to optimize revenue without suppressing demand
• Create bundled offers around category-size combinations (e.g., "Complete Your Wardrobe" bundles)
• Implement real-time stock monitoring and automated reorder triggers for key SKUs

Expected Business Impact: Reduce stockouts by 50% on top revenue drivers; improve sell-through rates by 10–15%; optimize inventory turns and working capital