Customer Shopping Behavior

1. Project overview

This project analyzes customer shopping behavior using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

Dataset size:

- **Number of Rows**: 3,900
- **Columns**: 18

Key Features:

- **Customer demographics** (Age, Gender, Location, Subscription Status)
- **Purchase details** (Item Purchased, Category, Purchase Amount, Season, Size, Color)
- **Shopping Behavior** (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- **Missing Data**: 37 values in Review Rating column

```
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             3900 non-null   int64
 1   Age                     3900 non-null   int64
 2   Gender                  3900 non-null   object
 3   Item Purchased          3900 non-null   object
 4   Category                3900 non-null   object
 5   Purchase Amount (USD)   3900 non-null   int64
 6   Location                3900 non-null   object
 7   Size                    3900 non-null   object
 8   Color                   3900 non-null   object
 9   Season                  3900 non-null   object
 10  Review Rating           3863 non-null   float64
 11  Subscription Status     3900 non-null   object
 12  Shipping Type           3900 non-null   object
 13  Discount Applied        3900 non-null   object
 14  Promo Code Used         3900 non-null   object
 15  Previous Purchases      3900 non-null   int64
 16  Payment Method          3900 non-null   object
 17  Frequency of Purchases  3900 non-null   object
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

Dataset Before

3. Exploratory Data Analysis Using Python

We began with data preparation and cleaning in Python:

- **Missing Data Handling**: Checked for null values and imputed missing values in the Review Rating column using the **median** rating of each product category.
- **Column Standardization**: Renamed columns to snake case for better readability and documentation.
- **Feature Engineering**:
  - Created age_group column by binning customer ages.
  - Created purchase_frequency_days column from purchase data.
- **Data Consistency Check**:
  - Verified if discount_applied and promo_code_used were redundant; dropped promo_code_used.
- **Database Integration**: Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis.

```
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 19 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   customer_id             3900 non-null   int64
 1   age                     3900 non-null   int64
 2   gender                  3900 non-null   object
 3   item_purchased          3900 non-null   object
 4   category                3900 non-null   object
 5   purchase_amount         3900 non-null   int64
 6   location                3900 non-null   object
 7   size                    3900 non-null   object
 8   color                   3900 non-null   object
 9   season                  3900 non-null   object
 10  review_rating           3900 non-null   float64
 11  subscription_status     3900 non-null   object
 12  shipping_type           3900 non-null   object
 13  discount_applied        3900 non-null   object
 14  previous_purchases      3900 non-null   int64
 15  payment_method          3900 non-null   object
 16  frequency_of_purchases  3900 non-null   object
 17  age_group               3900 non-null   category
 18  purchase_freqency_days  3900 non-null   int64
dtypes: category(1), float64(1), int64(5), object(12)
memory usage: 552.6+ KB
```

Dataset after

4. Data Analysis Using SQL

performed structured analysis in PostgreSQL to answer key business questions:

1. **Revenue by Gender:** Compared total revenue generated by male vs. female customers.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

Recommendation:

- **Targeted Marketing**: Since the male segment shows a high conversion or high spending power, consider increasing the ad spend on male-oriented channels to maximize ROI.
- **Market Expansion**: The female segment represents a significant growth opportunity. You should investigate if this gap is due to product-market fit, pricing, or a lack of targeted marketing toward women

2. **High-Spending Discount Users:** Identified customers who used discounts but still spent above the average purchase amount.

	customer_id bigint	purchase_amount bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90

Total rows: 839 | Query complete 00:0

Recommendation:

- **Upselling Opportunity:** Since these users already spend above average, they are the ideal targets for premium product bundles or "buy more, save more" offers.
- **Loyalty Retention:** This group is highly valuable. Implement a Tiered Loyalty Program to ensure they don't switch to competitors when discounts are not available.
- **Margin Optimization:** Test reducing the discount depth for this specific group (e.g., from 20% to 15%) to see if their high-spending behavior persists, thereby increasing the net profit margin

3. **Top 5 Products by Rating:** Found products with the highest average review ratings.

	item_purchased text	average_review_rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

Recommendation:

- **Top-Rated Badging:** Use the "Customer Favorite" or "Top-Rated" badge on marketing materials for Gloves and Sandals to drive higher conversion rates.
- **Inventory Prioritization:** Since Boots and Gloves are high-rated, ensure stock levels are optimized for peak seasons, as high ratings usually correlate with lower return rates.
- **Room for Improvement:** While these are the "top" products, none have reached the 4.0 threshold. This suggests a need for a Sentiment Analysis on written reviews to identify minor pain points (e.g., sizing issues or material durability) that could push these ratings from "Good" to "Excellent."

4. **Shipping Type Comparison:** Compared average purchase amounts between Standard and Express shipping.

	shipping_type text	avg_spend numeric
1	Standard	58.46
2	Express	60.48

Recommendation:

- **Threshold-Based Incentives:** Implement a "Free Express Shipping" offer for orders exceeding a certain amount (e.g., $75). This leverages the existing behavior of high-spenders to encourage Standard users to increase their basket size.
- **Operational Efficiency:** Since the spending difference is marginal but consistent, the business should focus on optimizing the logistics of Express shipping, as these customers represent a higher-value segment that expects premium service.
- **Checkout Optimization:** During the checkout process, highlight the "Express" option for carts that are already above the average ($60+) to reinforce the premium experience for high-value shoppers.

5. **Subscribers vs. Non-Subscribers:** Compared average spend and total revenue across subscription status.

	subscription_status text	total_customer bigint	average_spend numeric	total_revenue numeric
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

Recommendation:

- **Revise Subscription Benefits:** Since subscribers aren't spending more on average, the business should consider introducing exclusive subscriber-only deals or loyalty multipliers to increase their average order value (AOV).
- **Conversion Opportunity:** With 2,847 non-subscribers, there is a massive pool of potential members. A targeted "First-Month Free" campaign could convert these high-revenue-generating non-subscribers into loyal members.
- **Retention Focus:** Analyze why subscribers spend slightly less (59.49) than non-subscribers. It's possible that subscribers are making smaller, more frequent purchases, but the goal should be to push their average above the 60.00 mark.

6. **Discount-Dependent Products:** Identified 5 products with the highest percentage of discounted purchases.

	item_purchased text	dicount_rate numeric
1	Hat	50.00
2	Sneakers	49.00
3	Coat	49.00
4	Sweater	48.00
5	Pants	47.00

Recommendation:

- **Pricing Strategy Review:** For these "Discount-Heavy" items, consider a permanent price reduction (Everyday Low Price) instead of constant flash sales. This could stabilize demand and improve inventory management.
- **Bundle Offers:** Instead of a direct percentage discount, try "Buy One Get One" (BOGO) or bundling a high-discount item (like a Hat) with a low-discount item to protect overall profit margins.
- **Psychological Pricing:** Since these customers wait for discounts, use "Limited-Time Offers" or "Countdown Timers" specifically for these 5 categories to create urgency and drive faster conversions.

7. **Customer Segmentation:** Classified customers into New, Returning, and Loyal segments based on purchase history.

	customer_segment text	Number of customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

Recommendation:

- **Focus on Acquisition:** The "New" segment (83) is quite small relative to the "Loyal" group (3,116). The business should prioritize new customer acquisition campaigns (Top-of-Funnel marketing) to ensure future growth.
- **Upselling the "Returning" Group:** The 701 "Returning" customers are at a critical junction. Target them with personalized "Welcome Back" offers or bundles to transition them into the "Loyal" segment.
- **Reward Loyalty:** Since the business relies heavily on the 3,116 Loyal customers, implement an exclusive VIP program to prevent "churn" and ensure these high-value individuals don't move to competitors.

8. **Top 3 Products per Category:** Listed the most purchased products within each category.

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessori...	Jewelry	171
2	2	Accessori...	Sunglasses	161
3	3	Accessori...	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

Recommendation:

- **Inventory Synergy:** Since Blouses, Pants, and Jewelry are all top-sellers at the same volume (171), create "Complete Look" bundles featuring these three items to increase the total basket value.
- **Footwear Opportunity:** There is a noticeable drop-off in Footwear compared to Clothing. Investigate if adding more "Sandals" variations could capitalize on its current lead within that category.
- **Cross-Category Marketing:** Use the high traffic from Clothing (Blouses/Pants) to promote the top-ranked Accessories (Jewelry) through "Frequently Bought Together" recommendations.

9. **Repeat Buyers & Subscriptions:** Checked whether customers with >5 purchases are more likely to subscribe.

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

Recommendation:

- **Targeted Content:** Since Young Adults are the primary revenue drivers, focus digital marketing efforts (like social media ads) on this group to maintain their lead.
- **Retention for Seniors:** While the Senior group spends the least, the gap is small. Consider "Senior-Specific" loyalty perks or simplified subscription models to boost their retention and spending.
- **Engagement Strategy:** Use the data from the "Repeat Buyers & Subscriptions" check mentioned in the header to see if a specific age group is more likely to subscribe after 5 purchases, then tailor the subscription pitch accordingly.

10. **Revenue by Age Group**: Calculated total revenue contribution of each age group.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

Recommendation:

- **Conversion Campaign:** Target the 2,518 non-subscribed repeat buyers with a specialized "Loyalty Upgrade" offer. Since they are already buying repeatedly, they are the most likely group to see value in a subscription if given the right incentive (e.g., free shipping or exclusive discounts).
- **Analyze Barriers:** Investigate why 72% of loyal customers choose not to subscribe. Is the subscription too expensive, or are the benefits not clear enough for frequent shoppers?
- **Reward Subscribers:** To ensure the 958 current subscribers remain loyal, introduce "Surprise & Delight" rewards that are exclusive to them, further differentiating the subscriber experience from the standard repeat buyer experience.

5. Dashboard in Power BI

6. Business recommendations summary

1. **Bridge the Gender Revenue Gap:** Double down on high-performing male marketing while investigating the female segment to fix potential product-market fit or pricing barriers.
2. **Convert Loyal Non-Subscribers:** Target the 2,518 repeat buyers who aren't yet members with a "Loyalty Upgrade" campaign to turn existing trust into recurring subscription revenue.
3. **Incentivize Subscriber Spending:** Revise the subscription model to include exclusive multipliers or member-only deals, as subscribers currently don't spend more than non-members.
4. **Prioritize New Customer Acquisition:** Shift marketing focus toward "Top-of-Funnel" strategies to grow the small "New Customer" segment (83) and ensure long-term business scaling.
5. **Optimize "Discount-Heavy" Pricing:** Move high-sensitivity items like Hats and Sneakers toward an Everyday Low Price (EDLP) or "BOGO" model to stabilize margins and reduce reliance on flash sales.
6. **Maximize Basket Value through Bundling:** Create "Complete Look" bundles by pairing top-sellers across categories (e.g., Blouses with Jewelry) to naturally increase average order values.
7. **Leverage Shipping for Upselling:** Implement a Free Express Shipping threshold (e.g., $75) to encourage standard users to add more items to their carts to match the high-spend behavior of express users.