

Customer Shopping Analysis

1. Executive Summary

This report presents an analysis of customer shopping behavior for a retail company using a dataset of approximately **3,900 transactions**. The goal of the analysis was to uncover spending trends, customer segments, and purchase drivers that can improve engagement and boost profitability.

The total estimated revenue across all transactions was **\$233,081**, with an **average purchase amount of \$59.76**.

Gender-based analysis revealed that **male customers generated \$157,890**, accounting for roughly **68% of total revenue**, while **female customers contributed \$75,191**, representing about **32%**.

The study identified **young adults** as the most valuable demographic segment, with this age group contributing the largest share of total revenue and demonstrating strong purchase frequency. This insight highlights the importance of tailoring marketing strategies and promotional offers to appeal to younger, more active customers.

The project leveraged **Python** for data cleaning and transformation, **SQL** for analytical queries, and **Power BI** for interactive visualizations, forming a complete end-to-end data pipeline. Together, these tools revealed actionable insights into spending behavior, discount impact, and subscription patterns that can guide targeted marketing and loyalty initiatives.

2. Business Problem

The company is a retail business seeking to understand its customers' shopping behavior to improve overall sales, customer satisfaction, and long-term loyalty. Management has observed variations in purchasing patterns across demographics, product categories, and sales channels (online and offline).

They want to know what drives customers to spend more — is it discounts, product quality, subscription benefits, or convenience factors such as shipping type?

By identifying these behavioral patterns, the company aims to make informed decisions about pricing, promotions, and customer engagement strategies.

Business Objective:

To analyze customer transaction data and uncover insights that explain what influences purchasing decisions, how different customer segments contribute to revenue, and what strategies can increase retention and sales growth.

Analytical Focus:

- Evaluate how demographic factors (gender, age group) affect spending.
- Measure the effect of discounts and subscriptions on total revenue.
- Identify high-value customer segments and their characteristics.
- Compare shopping preferences such as product types and shipping methods.

3. Dataset Summary

- 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

4. Methodology

The project followed a structured, three-phase analytical workflow designed to transform raw transactional data into actionable insights.

Phase 1 — Data Preparation (Python)

- Imported and cleaned the dataset of 3,900 customer records.
- Removed duplicates, handled missing values, and standardized field formats (e.g., gender, discount flags, review ratings).

	purchase_frequency_days	frequency_of_purchases
0	14	Fortnightly
1	14	Fortnightly
2	7	Weekly
3	7	Weekly
4	365	Annually
5	7	Weekly
6	90	Quarterly
7	7	Weekly
8	365	Annually
9	90	Quarterly

- Created new calculated fields such as age_group, purchase_frequency_days and discount_applied for segmentation and trend analysis.

```
[22]: # create a new column age_group
labels = ['Young Adult', 'Adult', 'Middle-aged', 'Senior']
df['age_group'] = pd.qcut(df['age'], q=4, labels = labels)
```

```
[23]: df[['age', 'age_group']].head(10)
```

	age	age_group
0	55	Middle-aged
1	19	Young Adult
2	50	Middle-aged
3	21	Young Adult
4	45	Middle-aged
5	46	Middle-aged
6	63	Senior
7	27	Young Adult
8	26	Young Adult
9	57	Middle-aged

- Verified data consistency through exploratory data analysis (EDA) using Pandas and Matplotlib.

	customer_id	age	gender	item_purchased	category	purchase_amount	location	size	color	season	review_rating	subscription_status	shipping_type
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping

Phase 2 — Data Analysis (SQL)

- Developed ten focused SQL queries to extract measurable business insights.
- Each query targeted a specific question, including revenue by gender, the effect of discounts, and customer loyalty patterns.

```
##Q4. Compare the average Purchase Amounts between Standard and Express Shipping.
```

```
select shipping_type,  
ROUND(AVG(purchase_amount),2)  
from customer  
where shipping_type in ('Standard','Express')  
group by shipping_type;
```

```
##Q5. Do subscribed customers spend more? Compare average spend and total revenue  
##between subscribers and non-subscribers.
```

```
SELECT subscription_status,  
       COUNT(customer_id) AS total_customers,  
       ROUND(AVG(purchase_amount),2) AS avg_spend,  
       ROUND(SUM(purchase_amount),2) AS total_revenue  
FROM customer  
GROUP BY subscription_status  
ORDER BY total_revenue,avg_spend DESC;
```

- Aggregated purchase behavior by demographic factors, subscription status, and shipping type.
- Identified high-value segments and top-performing products.

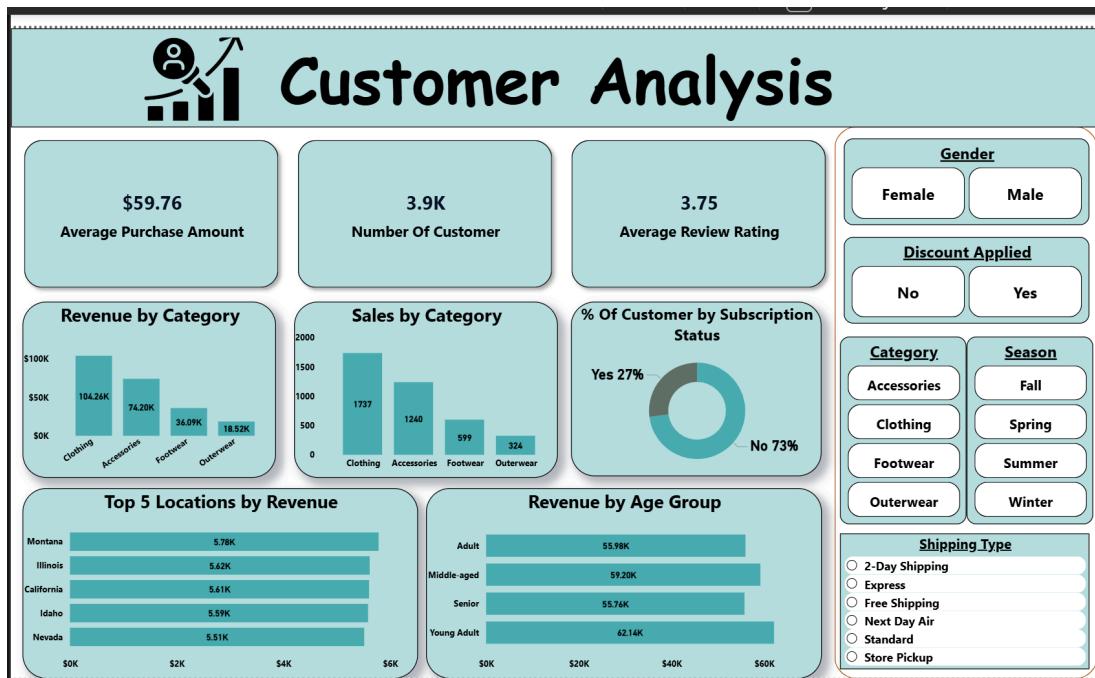
	age_group	total_revenue
▶	Young Adult	62143
	Middle-aged	59197
	Adult	55978
	Senior	55763

	subscription_status	repeat_buyers
▶	Yes	958
	No	2518

	customer_segment	Number of Customers
▶	Loyal	3116
	Returning	701
	New	83

Phase 3 — Visualization & Reporting (Power BI)

- Imported processed data into Power BI to build interactive dashboards.
- Created visuals for revenue distribution, discount performance, and product trends.
- Designed filters for gender, age group, and subscription status to support dynamic analysis.
- Combined results from Python and SQL into a single, business-focused dashboard for stakeholders.



This methodology ensured a complete data pipeline — from cleaning to analysis and visualization — enabling the company to make informed, evidence-based decisions.

5. Results & Insights

Analysis of 3,900 customer transactions revealed clear spending patterns and customer behaviors:

Total Revenue: \$233,081

Average Purchase Amount: \$59.76

Gender Revenue Split: Males generated \$157,890 (68%), while females contributed \$75,191 (32%).

Shipping Preference: Customers using Express shipping spent an average of \$60.48, slightly higher than \$58.46 for Standard shipping—showing that customers who value convenience also tend to spend more.

High-Value Discount Users: Top discounted products included Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), and Pants (47.37%). These discounts attract profitable purchases.

Loyalty Segment: Loyal customers contributed the highest share of total revenue (3,116 customers), followed by Returning (701) and New (83) customers, highlighting the importance of retention strategies.

Repeat Buyers and Subscription Status: Among repeat buyers, 958 were subscribed while 2,518 were non-subscribed, suggesting opportunities to convert repeat buyers into subscribers.

Age Group Contribution: The Young Adult segment (18–35) produced the most revenue (\$62,143), followed by Middle-aged (\$59,197), Adult (\$55,978), and Senior (\$55,763), indicating Young Adults are the core spending demographic.

Overall Insight: Convenience, loyalty, and targeted incentives drive the strongest financial outcomes, and understanding customer segments can help optimize promotions, retention, and subscription strategies

6. Business Recommendations

1. Targeted Promotions for High-Value Customers:

Focus discount campaigns on loyal customers and high-spending segments to maximize revenue impact while maintaining profitability.

2. Enhance Subscription Uptake Among Repeat Buyers:

Offer tailored incentives, such as exclusive discounts or loyalty rewards, to convert repeat buyers into subscribers, increasing recurring revenue.

3. Optimize Shipping Options:

Promote Express shipping to convenience-seeking customers, as they tend to spend more per transaction. Consider bundling fast shipping with high-margin products.

4. Leverage Age-Specific Marketing:

Prioritize marketing campaigns towards Young Adults (18–35), the segment generating the highest revenue. Use age-appropriate messaging and channels to boost engagement.

5. Monitor and Adjust Discount Strategies:

Regularly review the impact of high-discount items like Hats, Sneakers, and Coats to ensure promotions are driving revenue without eroding margins.

6. Customer Retention Programs:

Strengthen loyalty programs to reward returning and loyal customers, maintaining their engagement and spending while encouraging new customers to return.

7. Data-Driven Product Placement:

Highlight top-performing and high-rated products in campaigns, emails, and website features to drive higher sales and improve customer satisfaction.

Overall Strategy: By combining targeted discounts, subscription incentives, age-focused marketing, and loyalty programs, the business can increase revenue, customer retention, and long-term profitability.

7. Next Steps

- Conduct deeper analysis on customer lifetime value and churn.
- Use predictive models to forecast sales and subscription trends.
- Monitor KPIs with dynamic dashboards for real-time insights.
- Refine marketing strategies using additional demographics and purchase behavior.
- Regularly review the impact of implemented recommendations to drive sustained growth.