

# Customer Behavior Analysis

## Data-Driven Insights for E-Commerce Strategy



# 1. Project Overview

**Project Title:** Customer Behavior Analysis

**Objective:** Analyze customer purchasing patterns, demographics, and satisfaction metrics to identify growth opportunities and optimize marketing strategies.

**Business Context:** Understanding customer behavior is critical for e-commerce success. This analysis examines 3,900 customer transactions to uncover actionable insights that drive revenue growth, improve customer retention, and inform targeted marketing campaigns. The insights support strategic decisions in inventory management, pricing, and customer engagement.

# 2. Problem Statement & 3. Dataset Description

## Problem Statement

- Male-dominated customer base (68%) indicates untapped female market potential
- 27% subscription rate suggests opportunity to increase customer lifetime value
- Review ratings average 3.75/5 — room for improvement in customer satisfaction
- Need to identify high-value customer segments for targeted retention strategies

## Dataset Description

**Data Source:** E-commerce transaction database

**Records:** 3,900 customer transactions

**Columns:** 18 variables (after cleaning)

**Key Variables:** Customer ID, Age, Gender, Category, Purchase Amount, Review Rating, Subscription Status, Discount Applied, Frequency of Purchases, Previous Purchases

**Time Period:** Current customer base snapshot

**Data Quality:** No duplicates, minimal missing values (37 missing Review Ratings, 0.95%)

# 4. Business Questions & KPIs

## Key Questions Answered:

- What are the demographic characteristics of our customer base?
- Which product categories drive the most revenue?
- How do customer age groups differ in purchasing behavior?
- What is the relationship between purchase frequency and customer satisfaction?
- How effective are discounts and promotions in driving sales?

## Key Metrics & KPIs Analyzed:

- **Average Purchase Amount:** \$59.76 per transaction
- **Total Revenue:** \$233,081.00
- **Average Customer Age:** 44.1 years
- **Average Review Rating:** 3.75/5.0
- **Subscription Rate:** 27.0%
- **Discount/Promo Usage:** 43.0%
- **Average Previous Purchases:** 25.4 transactions per customer
- **Top Category:** Clothing (44.5% of sales)

# 5. Data Cleaning & Preparation

## Missing Value Handling:

- Review Rating: 37 missing values (0.95%) — imputed using category-wise mean to preserve distribution

## Duplicate Removal:

- No duplicate records found in dataset
- All 3,900 records are unique customer transactions

## Data Transformations:

- Created Age Groups: Segmented customers into 4 age brackets (Young Adult, Adult, Middle-aged, Senior) for demographic analysis
- Frequency Mapping: Converted purchase frequency text (Weekly, Monthly, etc.) to numeric days for correlation analysis
- Standardized categorical variables for consistent analysis

## Assumptions Made:

- Each row represents a unique customer transaction
- Missing Review Ratings are missing at random (MAR)
- Purchase amounts are in USD and accurately recorded
- Customer age is accurate at time of transaction

# 6. Exploratory Data Analysis (EDA)

## Key Patterns & Trends Discovered:

### Demographics

- Age distribution is broad and uniform (18-70 years), indicating diverse customer base with no dominant age group
- Male-dominated: 68% male vs. 32% female — significant opportunity for gender-targeted marketing
- Age groups evenly distributed: Young Adult, Adult, Middle-aged, and Senior segments all represent ~25% each

### Purchase Behavior

- Purchase amounts evenly spread (\$20-\$100 range) — consistent spending across customer segments
- Clothing and Accessories are top performers, accounting for majority of sales volume
- Footwear and Outerwear contribute smaller but stable portions

### Customer Satisfaction

- Review ratings concentrated between 3.0-4.5 (positive sentiment)
- Average rating of 3.75/5 indicates overall satisfaction but room for improvement
- No strong correlation between age, purchase amount, and review ratings (all correlations  $< 0.05$ )

### Seasonal Patterns

- Sales evenly distributed across all seasons (Spring, Summer, Fall, Winter)
- No seasonal dependency — stable year-round demand
- Business demonstrates consistent performance regardless of season



# 7. Key Insights

## Stable Revenue Base

\$233,081 total revenue from 3,900 customers averaging \$59.76 per transaction indicates consistent, predictable business performance

## High Repeat Purchase Rate

Average of 25.4 previous purchases per customer demonstrates strong customer loyalty and lifetime value potential

## Clothing Dominance

Clothing category leads with 44.5% of sales, followed by Accessories — these two categories should be inventory and marketing priorities

## Gender Imbalance Opportunity

68% male customer base reveals untapped female market segment; targeted campaigns could significantly expand customer base

## Moderate Satisfaction Levels

3.75/5 average rating suggests customers are satisfied but not delighted — improvement in product quality or service could boost ratings and retention

## Effective Discount Strategy

43% discount/promo usage rate indicates customers respond well to incentives; this is a proven lever for driving conversions

## Low Subscription Penetration

Only 27% subscription rate suggests opportunity to convert more customers to recurring revenue model

## No Seasonal Volatility

Even distribution across seasons reduces forecasting complexity and allows for consistent operational planning

# 8. Recommendations



## Expand Female Customer Segment

- Launch targeted marketing campaigns aimed at female customers (currently 32% of base)
- Analyze female purchasing preferences and tailor product offerings accordingly
- Expected impact: 10-15% increase in customer base within 6 months



## Increase Subscription Adoption

- Develop subscription incentive program (e.g., 15-20% discount for annual subscribers)
- Target high-frequency purchasers (weekly/fortnightly buyers) for conversion
- Expected impact: Increase subscription rate from 27% to 40%+ within 12 months



## Optimize Clothing & Accessories Inventory

- Allocate 60-70% of inventory budget to Clothing and Accessories categories
- Reduce slow-moving categories (Outerwear, Footwear) to free up capital
- Expected impact: Improved inventory turnover and reduced carrying costs



## Improve Customer Satisfaction

- Implement quality assurance program to boost average rating from 3.75 to 4.2+
- Gather customer feedback on product quality and service delivery
- Expected impact: Higher retention rates and positive word-of-mouth



## Leverage Discount Strategy

- Continue promotional campaigns — 43% engagement rate proves effectiveness
- Test tiered discount structures to maximize profit margins
- Expected impact: Maintain conversion rates while improving margins by 5-10%



## Develop Age-Specific Marketing

- Create tailored messaging for each age group (Young Adult, Adult, Middle-aged, Senior)
- Test different product recommendations and messaging by segment
- Expected impact: Improved conversion rates and customer satisfaction

# 9. Expected Business Impact

## Revenue Impact:

- Female segment expansion:  
+\$23,000-\$35,000 annually (10-15% growth)
- Subscription adoption: +\$50,000-\$75,000 annually (recurring revenue from 40% subscriber base)
- Improved satisfaction → retention:  
+\$15,000-\$25,000 annually (reduced churn)

### Total Potential Revenue Increase:

\$88,000-\$135,000 (38-58% growth)

## Cost Efficiency:

- Inventory optimization: \$10,000-\$15,000 annual savings (reduced carrying costs)
- Targeted marketing: 20-30% improvement in marketing ROI through segmentation
- Operational efficiency: Stable seasonal demand reduces forecasting errors and overstock

## Customer Metrics:

- Customer base growth: 3,900 → 4,300-4,500 customers (10-15% increase)
- Subscription rate: 27% → 40%+ (recurring revenue stability)
- Average review rating: 3.75 → 4.2+ (improved satisfaction and retention)
- Repeat purchase rate: Maintain 25.4+ average (strong loyalty)

## Strategic Value:

- Diversified customer base reduces dependency on male segment
- Recurring revenue model improves business predictability and valuation
- Enhanced customer satisfaction drives organic growth through referrals
- Data-driven approach enables continuous optimization and scaling

# 10. Tools & Technologies Used

## Programming & Data Analysis

- **Python** — Primary language for data analysis and statistical computations
- **Pandas** — Data manipulation, cleaning, and transformation
- **NumPy** — Numerical computations and array operations
- **Matplotlib & Seaborn** — Data visualization and exploratory analysis

## Database & Storage

- **PostgreSQL** — Relational database for storing and querying customer data
- **SQLAlchemy** — Python ORM for database connectivity and operations

## Development Environment

- **Jupyter Notebook** — Interactive analysis and documentation
- **Python 3.13** — Latest Python version for optimal performance

## Analysis Techniques

- Descriptive statistics (mean, median, standard deviation, quartiles)
- Univariate analysis (distributions, frequency counts)
- Correlation analysis (relationship between numerical variables)
- Categorical analysis (gender, category, subscription status)
- Segmentation analysis (age groups, purchase frequency)

# 11. Limitations & Future Scope

## Current Limitations:

- **Temporal Data:** Analysis is a snapshot of current customer base; lacks historical trends to identify growth patterns or seasonal cycles
- **External Factors:** Does not account for market conditions, competitor activity, or macroeconomic factors affecting purchasing behavior
- **Limited Behavioral Data:** No information on browsing behavior, cart abandonment, or customer journey touchpoints
- **Geographic Insights:** Location data available but not deeply analyzed for regional preferences or market penetration
- **Product Details:** Limited product-level analysis; cannot identify specific items driving satisfaction or dissatisfaction
- **Customer Acquisition Cost:** No data on marketing spend or customer acquisition costs to calculate ROI

## Future Scope & Recommendations:

- **Time Series Analysis:** Collect historical data to identify trends, seasonality, and growth patterns over 12-24 months
- **Customer Journey Mapping:** Integrate web analytics to understand browsing behavior and conversion funnel optimization
- **Predictive Modeling:** Build machine learning models to predict customer churn, lifetime value, and purchase propensity
- **Sentiment Analysis:** Analyze review text comments to identify specific pain points and satisfaction drivers
- **A/B Testing:** Implement controlled experiments to validate recommendations and measure impact of changes
- **Regional Analysis:** Deep dive into geographic performance to identify high-potential markets for expansion
- **Cohort Analysis:** Track customer cohorts over time to measure retention and lifetime value by acquisition period



# 12. Conclusion & Decision

## Recommendation

### Summary:

This analysis of 3,900 customer transactions reveals a stable, profitable e-commerce business with significant growth potential. The customer base demonstrates strong loyalty (25.4 average previous purchases) and consistent spending patterns (\$59.76 average transaction). However, the business faces untapped opportunities in female customer acquisition, subscription adoption, and customer satisfaction improvement.

### Key Findings:

- ✓ Stable revenue base of \$233,081 with predictable purchasing patterns
- ✓ Strong repeat purchase behavior indicating high customer lifetime value
- ✓ Clothing and Accessories categories drive 60%+ of revenue
- ✓ Effective discount/promo strategy with 43% engagement rate
- ✗ Gender imbalance (68% male) limits market reach
- ✗ Low subscription rate (27%) leaves recurring revenue on the table
- ✗ Moderate satisfaction (3.75/5) indicates service improvement opportunity

### Recommended Actions for Stakeholders:

01

#### Immediate (0-3 months):

Launch female-targeted marketing campaign and develop subscription incentive program

02

#### Short-term (3-6 months):

Implement quality assurance initiatives to improve customer satisfaction ratings

03

#### Medium-term (6-12 months):

Optimize inventory allocation and test age-specific marketing strategies

04

#### Long-term (12+ months):

Develop predictive models and expand geographic reach

### Expected Outcome:

Implementation of these recommendations could drive 38-58% revenue growth (\$88,000-\$135,000 annually) while improving customer satisfaction and operational efficiency. The data-driven approach positions the business for sustainable, scalable growth.

**Confidence Level:** High — Recommendations are based on clear patterns in customer behavior and industry best practices for e-commerce optimization.

