

Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping behavior using transactional data to identify **revenue drivers, discount usage, subscription impact, and operational preferences**. The objective is to support **data-driven business decisions** related to marketing strategy and customer engagement.

The project demonstrates an **end-to-end data analyst workflow** using:

- **Python** for data preparation and exploratory analysis
- **SQL** for business-oriented querying
- **Power BI** for visualization and insight communication

2. Business Objective

Primary business question:

How can customer shopping data be used to understand revenue patterns, discount effectiveness, and the role of subscriptions in customer behavior?

Key focus areas:

- Revenue contribution by customer attributes
- Discount usage across transactions
- Subscription penetration
- Category-level performance
- Shipping and payment preferences

3. Dataset Overview

- **Total transactions:** 3,900
- **Total columns:** 18

Key fields:

- Customer attributes: Age, Gender, Location
- Purchase details: Category, Item Purchased, Purchase Amount
- Engagement indicators: Previous Purchases, Subscription Status
- Promotions: Discount Applied
- Operations: Shipping Type, Payment Method
- Feedback: Review Rating

Data Quality Notes:

- Minor missing values in review ratings (~1%)
- No duplicate records
- Transaction-level dataset

4. Data Preparation & Exploratory Data Analysis (Python)

Tools: pandas, matplotlib, seaborn

Steps Performed:

- Loaded and inspected dataset structure
- Handled missing review ratings using median imputation
- Standardized column names
- Created age group bins for better interpretability

Key EDA Findings:

- No strong **linear relationship** between purchase amount and numerical variables
- Customer spending varies significantly across **categories and transaction attributes**
- Discounts are frequently used but do not significantly increase transaction value

- Subscription status shows minimal impact on average transaction amount

These findings indicate that customer behavior is **categorical and segment-driven**, which is typical in real-world retail data.

5. Business Analysis Using SQL

Database: PostgreSQL

SQL queries were written to answer **specific business questions**, including:

- Total revenue by gender
- Identification of high-spending customers using discounts
- Top-rated products based on reviews
- Comparison of spending across shipping types
- Subscription vs non-subscription spending
- Products with highest discount dependency
- Top products within each category
- Revenue contribution by age group

The focus was on **clear aggregation logic** and **business interpretation**, rather than complex SQL constructs.

6. Power BI Dashboard

Dashboard Components:

KPI Cards

- Total Revenue
- Total Customers
- Average Spend per Transaction
- % Subscribers
- % Discounted Transactions

Visual Analysis

- Revenue by product category
- Revenue by age group
- Average spend by subscription status
- Purchase amount by discount application
- Average spend by shipping type

Filters

- Category
- Age Group
- Subscription Status
- Shipping Type
- Season

All KPIs and visuals respond dynamically to filters.

7. Key Insights

- Revenue is driven primarily by **product category**, not customer age
- Discounts increase transaction frequency but do not significantly increase purchase value
- Subscribed customers do not spend more per transaction, but contribute to consistent engagement
- Faster shipping options show slightly higher average transaction values
- Customer behavior is better explained through **grouped analysis** rather than linear trends

8. Business Recommendations

1. **Optimize Discount Usage**
Reduce blanket discounts and focus promotions on price-sensitive products.
2. **Improve Subscription Strategy**
Position subscriptions as engagement tools rather than price incentives.
3. **Category-Focused Marketing**
Prioritize high-revenue categories for campaigns and inventory planning.
4. **Operational Enhancements**
Promote faster shipping options to higher-value customer segments.

9. Limitations & Future Scope

Limitations:

- No time-based analysis
- No causal inference
- Descriptive analysis only

Future Enhancements:

- Time-series trend analysis
- Customer lifetime value (CLV)
- Churn or repeat purchase prediction
- A/B testing of discount strategies

10. Tools & Technologies

- **Python:** pandas, matplotlib, seaborn
- **SQL:** PostgreSQL
- **Visualization:** Power BI
- **Version Control:** GitHub