Analyzing Customer Shopping Behavior

Driving Strategic Decisions Through Data Insights

Ruby Ineba Briggs

INTRODUCTION AND PROJECT SETUP

Project Overview

- Data Scope: Analysis of 3,900 customer transactions across diverse product categories.
- Core Goal: To reveal key behavioral insights regarding purchasing frequency, demographics, product preferences, and subscription influence.
- Purpose: Equip the business with actionable data for strategic decision-making and optimizing customer engagement.

Business Problem & Objective

- Problem: A major retailer seeks to boost sales, enhance satisfaction, and cultivate long-term loyalty amidst shifting purchase trends across demographics and channels.
- Key Interest: Uncovering which variables (discounts, reviews, seasonality, payment methods) drive immediate and sustained repurchase behavior.
- Central Question: "How can the organization utilize consumer shopping insights to pinpoint emerging trends, strengthen customer relationships, and refine its marketing and product development strategies?"

DATA SUMMARY AND STRUCTURE

Data Summary

- Structure: 3,900 transactional records and 18 distinct features.
- Key Feature Groups:
- Customer Demographics: Age, Gender, Location, Subscription Status.
- Purchase Specifics: Item, Category, Purchase Amount, Season, Size, Color.
- Behavioral Indicators: Discount/Promo Used, Previous Purchases, Review Rating, Shipping Type, Frequency.

Data Quality Note

- Missing Data: 37 missing entries were identified in the Review Rating column.

DATA CLEANING AND PREPROCESSING

Data Cleaning and Preprocessing Steps
Missing Value Imputation:

- Missing Review Rating values were filled using the median rating specific to the product category. (Median was chosen for its immunity to outliers).

Data Type Conversion:

- Verified and converted key columns (e.g., age, amount) to appropriate numerical types and textual columns to categorical types.

Standardization & Feature Reduction:

- Renamed and converted all column names to lower and snake casings for readability.

- Ensured discount_applied and promo_code_used were consistent; the redundant promo_code_used was dropped.

EXPLORATORY DATA ANALYSIS (EDA)

Process and Tools

- Data Loading: Performed in a Jupyter Notebook using the Pandas library (pandas as pd).

- Initial Check: Used df.info() and df.describe() for descriptive statistics and data structure insight.

Feature Engineering

- Age Grouping: Created a new categorical variable, age_group, from the numerical customer_age column for easier segmentation.

- Frequency Calculation: Created purchase_frequency_days from the purchase data to quantify loyalty.

Tools Used for Analysis

- Python (Pandas): Utilized for statistical tests, cleaning, and initial data manipulation.

- MSSQL: Used for complex aggregation queries to identify specific behavioral patterns.

KEY FINDINGS AND INSIGHTS (PART 1)

Insights driven by Python and MSSQL aggregation queries:

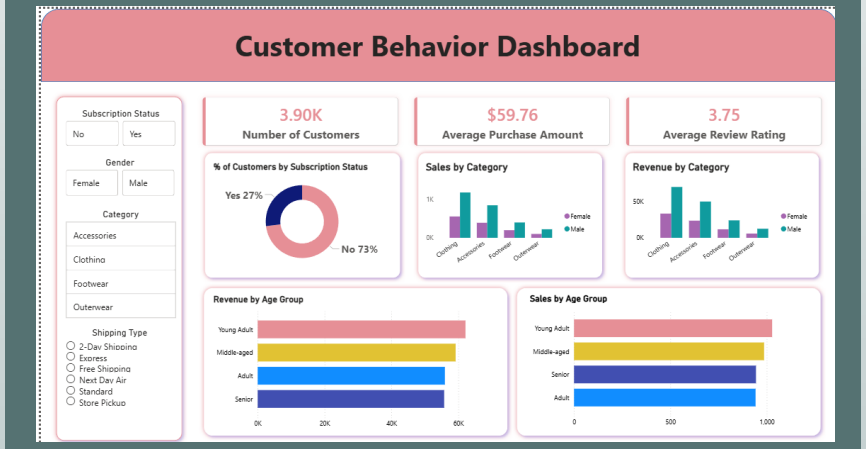
Analysis Point	Key Focus	Business Relevance
Revenue by Gender	Comparing total revenue generated by female vs. male customers.	Targeted marketing spend.
High Spending Discount Users	Identifying customers who use discounts but still spend above the average.	Retention strategy for high-value shoppers.
Shipping vs. Purchase Amount	Comparing average purchase amounts between Standard and Express shipping types.	Optimizing shipping strategy/pricing.
Top 5 Products by Rating	Identifying products with the highest average customer review ratings.	Product development and inventory focus.
Subscribers vs. Non-Subscribers	Comparing average spend and total revenue between the two groups.	Value proposition of the subscription program.
Top 5 Discounted Products	Identifying products with the highest percentage of discount application.	Promotion effectiveness/discount dependency.

KEY FINDINGS AND INSIGHTS (PART 2)

Insights driven by Python and MSSQL aggregation queries:(Cont.)

Analysis Point	Key Focus	Business Relevance
Customer Segmentation	Quantifying Returning, Loyal, and New customer segments.	Tailored loyalty programs.
Top Products per Category	Identifying the most purchased items within specific categories.	Assortment planning and cross-selling.
Repeat Buyers & Subscribers	Determining if customers with >5 purchases are more likely to subscribe.	Loyalty program enrollment targeting.
Revenue by Age Group	Calculating revenue contribution from different customer age groups.	Demographic-specific campaign design.
Peak Seasons & Location	Analyzing clothing purchases in Kentucky across different seasons.	Localized inventory and promotional timing.
Subscription Value Analysis	Calculating and comparing the average purchase amount for subscribed vs. non-subscribed customers.	Confirming the financial value of subscribers.
Payment Methods by Demographics	Identifying preferred payment methods by age groups and gender.	Payment gateway prioritization.

VISUALIZATION VIA POWERBI



Customer Behavior Dashboard Overview

Total Customers: The dataset analyzes 3.90K (3,900) customer transactions.

Key Metrics: The average purchase amount is $59.76, and the overall average customer review rating is 3.75.

Subscription Status: Only 27% of customers are currently subscribed, while the majority (73%) are not.

Revenue by Age Group: Young Adults generate the highest total revenue, followed closely by Middle-aged and Senior groups.

Sales by Age Group: Young Adults also account for the highest total volume of sales (units), followed by Adult and Senior groups.

Sales/Revenue by Category: The Clothing category dominates both the total Sales (units) and total Revenue, followed by Accessories and Footwear.

Gender Split in Categories: Female customers generally contribute more to sales/revenue across most product categories compared to Male customers.

Filtering Options: The dashboard allows for filtering the data by Subscription Status, Gender, Category, and Shipping Type.

CONCLUSION AND RECOMMENDATIONS

Conclusion

- The project successfully utilized Python and SQL to transition from data ingestion to in-depth analysis and Power BI visualization, achieving all set objectives and yielding relevant business insights.

Business Recommendations

Based on the analysis, I propose the following actionable steps:

- Loyalty & Discount Strategy: Offer more targeted discount rates and loyalty points specifically to returning and repeat customers to solidify loyalty.
- Express Shipping Focus: Given the potential insight into high purchase value, focus efforts on the express shipping strategy and conduct marketing campaigns to promote its benefits among customers.
- Targeted Marketing: Design campaigns to attract and retain the high revenue-generating age groups identified in the analysis.
- Continuous Improvement: Explore more metrics and advanced modeling (e.g., predictive analytics) to further enhance business value and forecasting accuracy.

THANK YOU!


Online Sales Tracker
Product Profit Per Item
$16.00
$14.00
$12.00
$10.00
$8.00
$6.00
$4.00
$2.00
$0.00
[Item 1]
[Item 2]
[Item 3]
[Item 4]
[Item 5]
% Income Per Product
[Item 1]
[Item 2]
[Item 3]
[Item 4]
[Item 5]
Shipping
Shipping Cost/Item
Profit per Item (incl. shipping)
Returns
$14.25