Customer Behavior Analysis Report

End-to-end retail data analytics — Python · MySQL · Power BI

A leading retail company · Consumer Shopping Dataset · Sakshi Sharma · 2025

10
SQL Queries

3
Project Phases

Business Problem

A leading retail company noticed shifts in purchasing patterns across demographics, product categories, and sales channels. The core challenge: **how can consumer shopping data be leveraged to identify trends, improve engagement, and optimize marketing and product strategies?**

Project Workflow

PHASE	TOOL	ACTIVITY	OUTPUT
1 · Data Prep	Python / Pandas	Clean, impute, engineer features	final_dataset.csv
2 · Analysis	MySQL	10 business queries & segmentation	Structured insights
3 · Dashboard	Power BI	Interactive charts & filters	Customer_Behavior.pbix

Phase 1 — Data Preparation (Python)

Missing Values	Imputed review_rating using per-category median to preserve category-level accuracy.
Column Names	Standardized all headers to snake_case for MySQL compatibility.
Age Groups	Created age_group via pd.qcut() → Young Adult, Adult, Middle-aged, Senior.
Frequency Days	Mapped text frequency (Weekly, Monthly…) to numeric day values for quantitative analysis.
Redundancy	Dropped promo_code_used — confirmed 100% identical to discount_applied.

Phase 2 — SQL Business Queries

Ten targeted queries were written to answer the company's core business questions, covering revenue, discounts, loyalty, subscriptions, and product performance.

#	Business Question	Key Technique	Insight
Q1	Revenue by gender	SUM / GROUP BY	Gender-based revenue comparison
Q2	Discount users above avg spend	Subquery + WHERE	High-value deal-seekers identified
Q3	Top 5 products by review rating	AVG, ORDER BY, LIMIT	Best-rated items for promotion
Q4	Standard vs Express shipping spend	Conditional GROUP BY	Express users spend more
Q5	Subscriber vs non-subscriber spend	COUNT, AVG, SUM	Subscribers drive more revenue
Q6	Top 5 most-discounted products	CASE WHEN + %	Discount-heavy SKUs flagged
Q7	New / Returning / Loyal segments	CTE + CASE WHEN	3-tier loyalty model built
Q8	Top 3 products per category	ROW_NUMBER() window fn	Category leaders ranked
Q9	Repeat buyers & subscriptions	WHERE filter, GROUP BY	Repeat buyers likely to subscribe
Q10	Revenue by age group	Engineered col, GROUP BY	Age-segment revenue ranking

Phase 3 — Power BI Dashboard

Revenue by Gender & Age Demographic revenue breakdown	**Loyalty Segments** New / Returning / Loyal distribution
Subscription Spend Subscriber vs non-subscriber comparison	**Top Products** By purchase volume and review rating
Discount Impact Discount rate per product vs revenue	**Shipping Preferences** Type distribution & avg spend

Key Findings

Subscribers Generate Disproportionate Revenue

Subscribed customers spend significantly more on average and contribute a higher cumulative revenue total. The subscription model is a proven, scalable revenue lever.

Loyal Customers Are Small but Critical

Customers with 11+ purchases are a minority yet account for an outsized revenue share. Even modest retention improvements in this group have high LTV impact.

Discounts Don't Always Create Value

Several of the most-discounted products carry thin margins. Blanket discounting on these SKUs may be eroding profitability without acquiring high-value customers.

Repeat Buyers Are Prime Subscription Candidates

Customers with 5+ previous purchases show strong subscription affinity — they're the most efficient group to target with conversion campaigns.

Top-Rated Products Are Under-Promoted

Highest-rated products don't always align with highest-revenue products, revealing an opportunity to better market well-reviewed items to boost conversion.

Revenue Is Age-Group Concentrated

Revenue is not evenly distributed across age segments. Marketing budgets are not currently weighted toward the highest-contributing demographic groups.

Business Recommendations

HIGH — **Expand Subscription Program**

Introduce tiered subscription perks — free express shipping, early access, member discounts. Target high-frequency non-subscribers with conversion campaigns. This is the single highest-ROI lever.

HIGH — **Launch a Loyalty / VIP Program**

Introduce points, tiered status, and exclusive rewards for the Loyal segment (11+ purchases). Even small retention gains here have large LTV impact.

MED — **Amplify Top-Rated Products**

Feature the 5 highest-rated products in ads, emails, and homepage placements. High ratings are proven conversion drivers and build trust with first-time buyers.

HIGH — **Trigger Sub Prompt at 5th Purchase**

Automate an in-app or email prompt at the 5th purchase milestone offering a discounted first month. Repeat buyers are the most subscription-ready segment.

MED — **Audit Discount Strategy**

Replace blanket discounts with personalized, behavior-triggered offers. High-discount products should be cross-checked against margin data to protect profitability.

MED — **Optimize Ad Spend by Age Group**

Shift budget toward the highest-revenue age segment for premium pushes, while running lower-cost experiments on underperforming segments before scaling.

Conclusion

This project demonstrates how combining Python data engineering, structured SQL analysis, and Power BI visualization delivers a complete, stakeholder-ready picture of consumer behavior. The analysis confirms that a small subset of customers — loyal, subscribed, and repeat buyers — contribute disproportionately to revenue. Focusing retention and conversion efforts on this core is the company's clearest data-backed path to sustainable growth.

Implementing the top three recommendations (subscription expansion, loyalty program, and repeat-buyer conversion triggers) within the next business quarter is expected to measurably improve customer retention, average order value, and subscription cohort size.

10	4	3	6	3
SQL Queries	Age Segments	Loyalty Tiers	Recommendations	Project Phases

GitHub Repository

github.com/SAKSHISHR2703/customer-behavior-analysis

All source files — Python notebook, SQL queries, and Power BI dashboard are available in the repository. See README for setup instructions.

Tools & Stack

Python (Pandas, NumPy) · MySQL · Power BI

Google Colab for notebook execution · MySQL Workbench for query development

PREPARED BY

Data Analyst

EMAIL

sakshishr2703@gmail.com

LINKEDIN

linkedin.com/in/sakshi2703

DATE

STATUS

Completed