Customer Shopping Behavior Analysis

End-to-End Data Analytics Project

---

1. Executive Summary

Retail businesses generate vast amounts of customer transaction data, but without structured analysis, this data does not translate into actionable insights. This project focuses on analyzing customer shopping behavior using transactional data from **3,900 purchases** across multiple product categories.

The analysis was performed using **Python for data preparation and exploratory analysis**, **PostgreSQL for structured business querying**, and **Power BI for interactive visualization**. The objective was to uncover insights related to customer purchasing behavior, product performance, discount dependency, customer loyalty, and revenue contribution across different customer segments.

The final outcome includes a cleaned dataset, SQL-driven business insights, an interactive dashboard, a detailed report, and a professional presentation suitable for stakeholders.

---

2. Business Problem Statement

Retail organizations often struggle to understand why customers purchase certain products, how discounts influence buying behavior, and which customer segments contribute most to revenue. Without clear insights into customer behavior and loyalty patterns, businesses risk inefficient marketing spend, poor product positioning, and missed revenue opportunities.

The goal of this project is to analyze customer shopping behavior data to identify purchasing trends, segment customers based on loyalty, evaluate discount effectiveness, and highlight high-performing products and customer segments. The insights generated aim to support data-driven decision-making for marketing, pricing, and customer retention strategies.

---

3. Dataset Overview

- **Total Records:** 3,900 customer transactions

- **Total Columns:** 18
- **Data Type:** Retail transactional data

Key Data Categories

- **Customer Demographics:** Age, Gender, Location, Subscription Status
- **Purchase Details:** Item Purchased, Category, Purchase Amount, Season, Size, Color
- **Behavioral Attributes:** Discount Applied, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type

Data Quality Notes

- Missing values were found only in the **Review Rating** column (37 records).
- No duplicate customer transaction records were identified.

---

4. Data Preparation & Feature Engineering (Python)

4.1 Data Loading & Initial Exploration

- Dataset imported using pandas
- Structure validated using df.info()
- Summary statistics reviewed using df.describe()

4.2 Data Cleaning

- Missing review ratings were imputed using **category-wise median values**
- Column names standardized to **snake_case**
- Data types validated to ensure consistency

4.3 Feature Engineering

To enhance analytical depth, additional features were created:

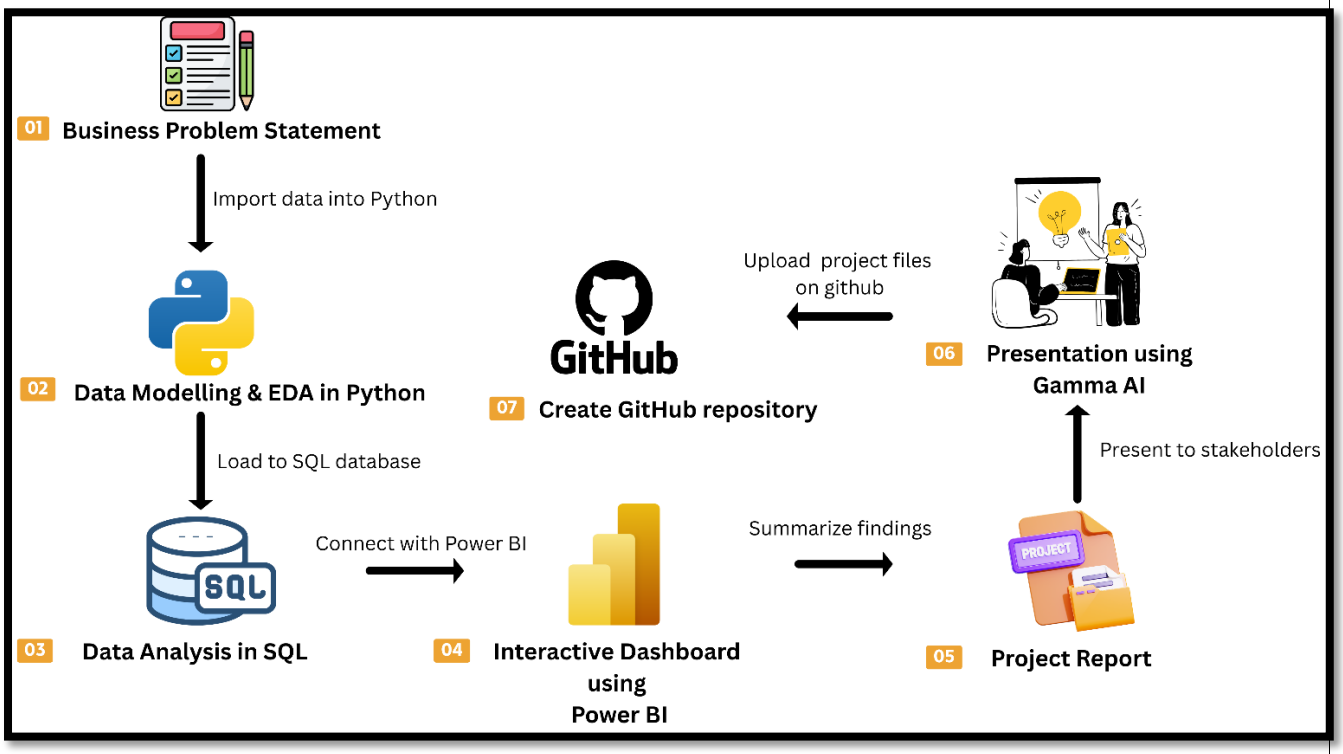
- **Age Group:** Customers grouped into Young Adult, Adult, Middle-Aged, and Senior
- **Purchase Frequency (Days):** Derived to analyze buying behavior over time

The cleaned and enriched dataset was then exported and used for SQL analysis.

---

5. Database Integration

The prepared dataset was loaded into a **PostgreSQL** database using SQLAlchemy. This enabled structured querying and efficient business analysis using SQL, simulating a real-world production analytics environment.



---

6. 📊 Numerical Analysis of SQL Query Results

1️⃣ Revenue by Gender

Gender	Total Revenue
Female	75,191
Male	157,890

📌 Analysis

- Male customers contribute **more than double** the revenue compared to female customers.
- Male revenue ≈ **67.7%** of total gender-based revenue.

💡 Business Meaning

- Male customers are currently the **primary revenue drivers**.
- Female segment may be **under-targeted or underserved**.

✅ Business Action

- Maintain strong campaigns for male customers.
- Introduce **targeted promotions** to increase female customer spending.

2️⃣ High-Spending Discount Users

- **839 customers** used discounts **and still spent above the average purchase amount**.

📌 Analysis

- Discounts do **not always reduce revenue**.
- A significant group of customers uses discounts strategically but still spends more.

□ Business Meaning

- These customers are **value-seeking but high-value**.
- They respond well to offers without reducing basket size.

✅ **Business Action**

- Offer **personalized premium discounts** to these customers.
- Avoid removing discounts blindly.

3️⃣ Top 5 Products by Average Review Rating

Product	Avg Rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.80
Skirt	3.78

📌 **Analysis**

- Ratings are consistently **high (>3.7)**.
- Indicates **strong customer satisfaction**.

□ Business Meaning

- These products are **quality leaders**.
- High satisfaction increases **repeat purchases**.

✅ **Business Action**

- Highlight these products in marketing campaigns.
- Bundle them with lower-rated products.

4️⃣ Shipping Type vs Average Purchase Amount

Shipping Type	Avg Purchase
Standard	58.46
Express	60.48

📌 Analysis

- Express shipping customers spend **~3.5% more** per transaction.
- Indicates higher purchase intent.

🧠 Business Meaning

- Express users value **speed and convenience**.
- They are willing to pay more.

✅ Business Action

- Promote express shipping for high-value customers.
- Offer free express shipping above a threshold.

5️⃣ Subscribers vs Non-Subscribers

Segment	Customers	Avg Spend	Total Revenue
Subscribers	1,053	59.49	62,645
Non-Subscribers	2,847	59.87	170,436

📌 Analysis

- Non-subscribers generate **~73% of revenue** due to volume.
- Average spend per transaction is **similar**.

🧠 Business Meaning

- Subscription is **underutilized**.
- Subscriptions are not yet driving higher spend.

✅ Business Action

- Improve subscription benefits (exclusive offers, faster shipping).
- Convert high-frequency buyers to subscribers.

6️⃣ Discount-Dependent Products (Top 5)

Product	Discount Usage %
Hat	50.00%
Sneakers	49.66%
Coat	49.07%
Sweater	48.17%
Pants	47.37%

📌 Analysis

- Nearly **half of purchases** depend on discounts.
- These products are **price-sensitive**.

⚠️ Business Meaning

- Customers wait for discounts before buying.
- Profit margins may be affected.

✅ Business Action

- Reduce discounts gradually.
- Introduce **value bundles** instead of direct discounts.

7️⃣ Customer Segmentation (Based on Previous Purchases)

Segment	Customers
Loyal	3,116

Returning	701
New	83

📌 **Analysis**

- **~80%** of customers are Loyal.
- Very few new customers entering the system.

 Business Meaning

- Strong retention.
- Weak customer acquisition.

✅ **Business Action**

- Invest in new customer acquisition campaigns.
- Retain loyal customers through rewards.

8️⃣ Top 3 Products per Category (Sales Volume)

Example Highlights

- **Accessories:** Jewelry (171), Sunglasses (161), Belt (161)
- **Clothing:** Blouse (171), Pants (171), Shirt (169)
- **Footwear:** Sandals (160), Shoes (150), Sneakers (145)
- **Outerwear:** Jacket (163), Coat (161)

📌 **Analysis**

- Sales are **concentrated** among top products.
- Product diversity exists, but leaders dominate.

 Business Meaning

- Inventory planning should focus on these products.
- Low performers may be optimized or removed.

9️⃣ Repeat Buyers & Subscription Status

Subscription	Repeat Buyers
No	2,518
Yes	958

📌 Analysis

- Majority of repeat buyers are **not subscribed**.
- Indicates missed retention opportunity.

💡 Business Meaning

- Subscriptions are not positioned as a loyalty tool.

✅ Business Action

- Convert repeat buyers into subscribers using targeted offers.

🔟 Revenue by Age Group

Age Group	Revenue
Young Adult	62,143
Middle-aged	59,197
Adult	55,978
Senior	55,763

📌 Analysis

- Young Adults generate the **highest revenue**.
- Revenue decreases gradually with age.

💡 Business Meaning

- Younger customers are more active buyers.

✅ **Business Action**

- Focus marketing on young and middle-aged segments.
- Use tailored messaging for seniors.

□ Overall Business Summary (Executive View)

- Revenue is driven by **male, loyal, young-adult customers**.
- Discounts work but create **dependency** for certain products.
- Subscriptions have **high potential but low conversion**.
- Express shipping correlates with **higher spend**.

---

7. Power BI Dashboard

An interactive **Power BI dashboard** was developed to present insights visually for stakeholders.

Dashboard Highlights

- Total customers, average purchase amount, and average review rating
- Revenue and sales distribution by product category
- Revenue contribution by age group
- Subscription status distribution
- Interactive filters for gender, category, shipping type, and subscription status

The dashboard enables non-technical users to explore trends and make informed decisions quickly

Customer Behavior Analysis

.

---

8. Key Insights

- **Loyal customers** represent the largest segment and drive repeat purchases
- **Express shipping users** spend more per transaction compared to standard shipping users
- Certain products show **high dependency on discounts**, indicating price sensitivity
- **Subscription customers** contribute a significant share of total revenue
- Specific **age groups** generate higher overall revenue and should be targeted strategically

---

9. Business Recommendations

- **Boost Subscriptions:** Promote exclusive benefits to convert returning customers into subscribers

- **Customer Loyalty Programs:** Reward repeat buyers to strengthen long-term retention
- **Optimize Discount Strategy:** Balance discount usage to protect margins while driving volume
- **Product Positioning:** Highlight top-rated and best-selling products in marketing campaigns
- **Targeted Marketing:** Focus efforts on high-revenue age groups and express shipping users

---

10. Project Deliverables

- Cleaned dataset (.csv)
- EDA notebook (.ipynb)
- SQL query file (.sql)
- Power BI dashboard (.pbix)
- Detailed project report (.pdf)
- Stakeholder presentation (.pptx)

---

11. Tools & Technologies

- **Python:** Pandas, NumPy, Matplotlib, Seaborn
- **PostgreSQL:** SQL-based business analysis
- **Power BI:** Interactive data visualization
- **Jupyter Notebook**
- **GitHub:** Version control and documentation

---

12. Conclusion

This project demonstrates a complete end-to-end data analytics workflow, from business problem definition and data preparation to SQL-based analysis, visualization, and reporting. The insights generated provide actionable

guidance for improving customer retention, optimizing pricing strategies, and enhancing overall business performance through data-driven decision-making.