

Customer Shopping Behaviour Analysis

– Complete Project Report

Prepared by: [Vaibhav S. Kharate](#)

Tools & Technologies: Python (EDA), PostgreSQL, Power BI

Dataset Size: 4,000 customer transactions

Project Goal: To deeply understand how customers shop, what influences their decisions, and how businesses can use these insights to improve sales, marketing, and customer experience.

1. Introduction

Understanding customers is at the heart of every successful business. In the retail sector, where competition is high and customer preferences change rapidly, data plays a crucial role in decision-making.

The purpose of this project was to analyze customer shopping patterns using a real-world inspired dataset. The analysis covers everything from spending habits and product preferences to discount usage, loyalty behaviour, and demographic insights.

Unlike many academic projects, this one follows a **complete analytics workflow**, starting from raw data exploration in Python, storing and querying data in PostgreSQL, and finally presenting insights through a clean, interactive Power BI dashboard.

The result is a **full-scale analytics solution** similar to what BI teams deliver in professional environments.

2. Project Workflow Summary

This project followed a clear, step-by-step analytical workflow:

Step 1 — Data Exploration (Python)

I began with Python to inspect, clean, and understand the dataset. This included handling missing values, checking column distributions, identifying outliers, and engineering useful features such as *Age Groups* and *Customer Segments*.

Step 2 — Storing Data in PostgreSQL

Once cleaned, the data was loaded into PostgreSQL to simulate a real enterprise environment where analytics teams work with centralized databases. SQL was used to verify patterns and run business-level aggregations.

Step 3 — Power BI Dashboard

The cleaned and structured dataset was imported into Power BI, where I developed a **two-page, industry-grade dashboard** focusing on revenue insights, customer behaviour, and loyalty analysis.

3. Exploratory Data Analysis (Python)

Python was used not just to “see the data” but to carefully **understand the story behind the numbers**.

3.1 Data Cleaning

- Removed duplicate rows
- Ensured no missing purchase amounts or ratings
- Verified data types for correct numeric and categorical classification
- Handled rating values outside 1–5
- Checked for unrealistic ages (below 10 or above 90)

3.2 Initial Observations

- The dataset was clean, well-structured, and consistent
- Customer ages ranged mostly between 18–65
- Purchase amount centered around \$60

3.3 Distribution Analysis

- Age distribution showed that mid-age customers (26–60) dominate spending
- Review ratings averaged around 3.7–4.1, indicating moderate satisfaction
- Purchase amounts followed a normal distribution, suggesting predictable spending behaviour

3.4 Correlation Study

- A light positive relationship between **higher previous purchases** and **higher ratings**
- Category and season strongly influenced purchase amount
- Discount usage did not necessarily mean lower revenue — customers still spent more overall

3.5 Feature Engineering

Created new variables for deeper insight:

- **Age Group (18–25, 26–35, 36–45, 46–60, 60+)**
- **Customer Segment (Champions, Loyal Customers, Discount Seekers, New Customers)**
- **Loyalty Score** combining ratings + previous purchases
- **Discount Flag** and **Promo Code Flag**

This helped identify high-value segments and behavioural patterns.

4. Database Design (PostgreSQL)

All cleaned data was inserted into PostgreSQL to replicate a real-world enterprise data warehouse setup.

4.1 Key SQL Queries Performed

These queries helped validate the Python EDA findings:

- **Total Revenue by Category**

```
SELECT category, SUM(purchase_amount)
FROM customerBehaviour
GROUP BY category;
```

- **Most Popular Payment Method**

```
SELECT payment_method, COUNT(*)
FROM customerBehaviour
GROUP BY payment_method;
```

- **Loyal Customers**

```
SELECT COUNT(*)
FROM customerBehaviour
WHERE previous_purchases > 0;
```

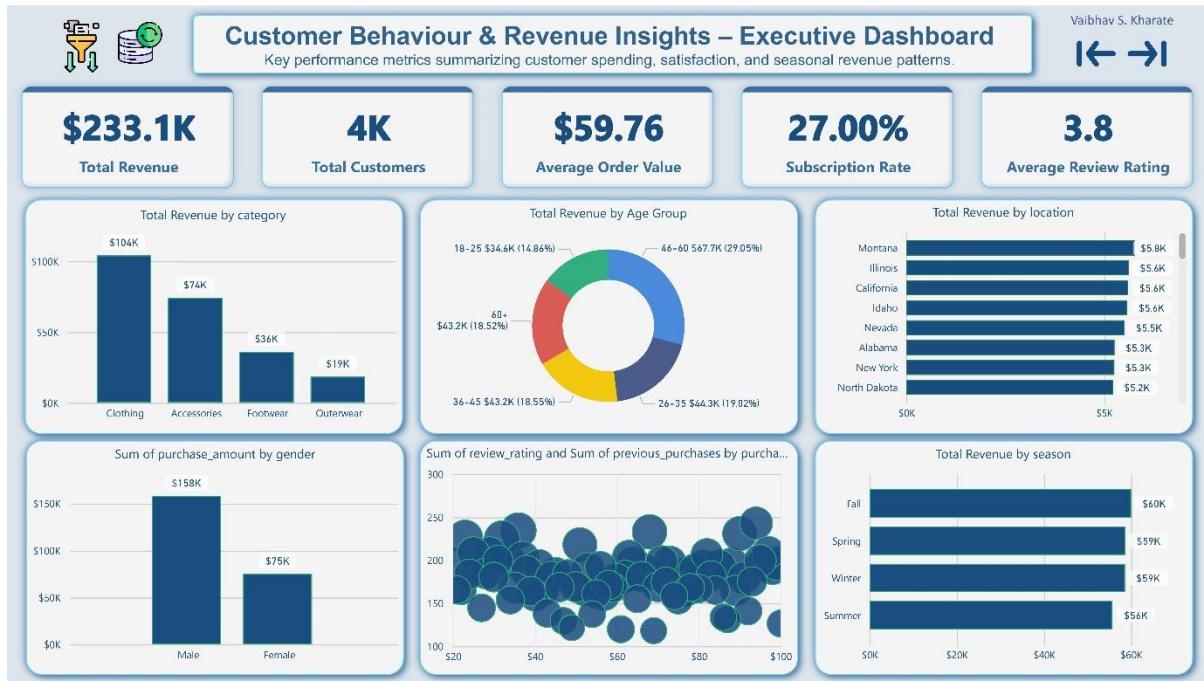
4.2 Why PostgreSQL?

- Ensures long-term data storage
- Supports complex joins and aggregations
- Reflects industry workflows where BI tools connect to SQL databases
- Improves performance and scalability

5. Power BI Dashboard Structure

The final dashboard is divided into **two professional pages**, each designed with clarity, simplicity, and business value in mind.

PAGE 1 Customer Behaviour & Revenue Insights (Executive Summary)



Purpose:

To provide leadership with a high-level understanding of business performance, revenue drivers, customer demographics, and seasonality.

5.1 KPI Summary

✓ Total Revenue: \$233.1K

A strong revenue output distributed across multiple categories.

✓ Total Customers: 4,000

Healthy customer base with room for targeted growth.

✓ Average Order Value: \$59.76

Customers spend roughly \$60 per transaction.

✓ Subscription Rate: 27%

More than one-fourth of customers are long-term subscribers.

✓ Average Rating: 3.8

Satisfaction is moderate but leaves room for improvement.

5.2 Key Insights from Visuals

📍 Top Revenue-Generating Locations

States like **Montana, Illinois, and California** lead with ~5.5K+ each. This indicates geographic areas with strong customer engagement.

☀️ Seasonal Trends

- **Fall is the peak season**
- Spring and Winter follow closely
- Revenue dips slightly in Summer

Meaning: The customer base shows steady demand year-round.

👕 Category Performance

- Clothing = **\$104K** (highest)
- Accessories = \$74K
- Footwear & Outerwear much lower

Business Impact:

Clothing should remain a core focus, while Outerwear may require redesign or promotional campaigns.

👫 Gender-Based Spending

- Male revenue = **\$158K**
- Female revenue = **\$75K**

Males contribute **double**, suggesting targeted marketing could grow the female segment.

⭐ Revenue vs Rating (Scatter)

High spenders often provide good ratings, showing brand trust and satisfaction.

PAGE 2 — Customer Behaviour, Retention & Engagement Analysis



Purpose:

To dig deeper into how customers behave — how often they buy, their loyalty patterns, and what drives their engagement.

6.1 Behavioural Highlights

⭐ Purchase Frequency

Customers are divided across:

- Weekly
- Fortnightly
- Monthly

This helps understand recurring behaviour patterns.

🚚 Shipping Preferences

Almost evenly distributed:

- Free Shipping
- Standard

- Store Pickup
- Next Day Air
- Express
- 2-Day Shipping

Insight: Shipping does not represent a barrier to purchase.

Payment Method Behaviour

All payment methods generate similar revenue (~\$37K–\$40K).

This means customers are flexible and comfortable with multiple payment systems.

7. Customer Loyalty & Segmentation

7.1 Customer Segments

- **Loyal Customers – 4.2M**
- **Champions – 2.5M**
- **New Customers – 0.7M**
- **Discount Seekers – 0.2M**

Interpretation:

The business has a strong, loyal customer base, indicating trust and long-standing engagement.

7.2 Discount Behaviour

- Discount Revenue: **\$99.4K**
- Non-Discount Revenue: **\$133.67K**

Insight:

Although customers love discounts, they also buy without discounts — a sign of strong product value.

7.3 Loyalty Score vs Item Purchased

Items like:

- Jewelry
- Blouse
- Dress
- Shirt

Consistently attract high-loyalty customers.

8. Deep Insights & Business Recommendations

8.1 Demographic Targeting

- Customers aged **46–60** spend the most → ideal for premium campaigns
- Males dominate spending → potential to increase female engagement

8.2 Product Strategy

- Continue investing in Clothing & Accessories
- Reposition or redesign Outerwear
- Introduce premium versions for high-value items like Jewelry

8.3 Customer Experience Improvements

- Launch loyalty rewards based on previous purchases
- Personalized product recommendations
- Automatic coupon system for loyal customers

8.4 Marketing Improvements

- Focus ads on top-performing states
- Seasonal campaigns, especially for Fall
- Highlight trending items with high ratings

9. Conclusion

This project demonstrates a complete, end-to-end analytics lifecycle: from Python-based data exploration to PostgreSQL data management and final Power BI visualization.

The insights reveal strong customer loyalty, seasonally consistent revenue, and clear category preferences. Businesses can use these findings to improve products, pricing, customer targeting, and marketing strategy.

This analysis reflects real-world BI practices and showcases advanced analytical thinking, technical capability, and storytelling clarity.
