

# Customer Behaviour Analysis Report

A Data-Driven Analysis of Purchase Patterns



Figure 1: Power BI Dashboard Overview

**Author:** Md. Sahid

**Date:** November 17, 2025

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Dataset Description</b>	<b>2</b>
<b>3</b>	<b>Data Engineering Pipeline</b>	<b>3</b>
3.1	Python Preprocessing . . . . .	3
3.2	Database Loading . . . . .	3
<b>4</b>	<b>SQL Querying for Business Insights</b>	<b>3</b>
<b>5</b>	<b>Detailed Dashboard Analysis</b>	<b>4</b>
5.1	Key Performance Indicators (KPIs) . . . . .	4
5.2	Revenue and Sales by Category . . . . .	4
5.3	Subscriber Analysis . . . . .	4
5.4	Revenue and Sales by Age Group . . . . .	5
5.5	Dynamic Filtering . . . . .	5
<b>6</b>	<b>Additional SQL Insights</b>	<b>5</b>
6.1	Top Rated Products (SQL Q3) . . . . .	5
6.2	Discount & Promo Code Impact (SQL Q6) . . . . .	5
6.3	Shipping Type vs. Spend (SQL Q4) . . . . .	5
<b>7</b>	<b>Business Impact &amp; Outcomes</b>	<b>5</b>
<b>8</b>	<b>Recommendations &amp; Future Actions</b>	<b>6</b>
<b>9</b>	<b>Conclusion</b>	<b>6</b>

## Abstract

This report presents a comprehensive analysis of customer purchasing behaviour utilizing a combination of Python, SQL, and Power BI. By examining demographic attributes, transaction details, and key behavioural metrics from 3,900 customer records, the goal is to uncover actionable insights to support strategic business decisions. The analysis covers data preprocessing, exploratory SQL querying, and a detailed breakdown of an interactive Power BI dashboard. The findings reveal key customer segments, product performance, and the significant impact of subscription status and shipping preferences on revenue.

## 1 Introduction

This report provides a comprehensive analysis of customer purchasing behaviour using a multi-tool data stack: Python for data engineering, SQL for business querying, and Power BI for interactive visualization. By examining demographic attributes (Age, Gender), purchase details (Category, Amount, Season), and behavioural metrics (Subscription Status, Review Rating), the goal is to uncover actionable insights. These insights are intended to support strategic business decisions in marketing, inventory management, and customer relationship management, ultimately driving revenue and enhancing customer satisfaction.

## 2 Dataset Description

The analysis is based on the `customer_shopping_behavior.csv` dataset, containing 3,900 unique customer transactions. The dataset consists of 18 columns, providing a rich, multi-dimensional view of each purchase event. The detailed schema is provided in Table 1.

Table 1: Dataset Schema Description

Column Name	Description
Customer ID	Unique identifier for each customer.
Age	Customer's age in years.
Gender	Customer's gender (e.g., Male, Female).
Item Purchased	Name of the specific product.
Category	Product category (e.g., Clothing, Accessories).
Purchase Amount (USD)	The total value of the purchase in USD.
Location	Customer's location (State or Region).
Size	Size of the purchased item (e.g., S, M, L).
Color	Color of the purchased item.
Season	Season of the purchase (e.g., Winter, Summer).
Review Rating	Rating given by the customer (1-5).
Subscription Status	Indicates if the customer is a subscriber (Yes/No).
Shipping Type	Selected shipping method (e.g., 2-Day, Standard).
Discount Applied	Boolean indicating if a discount was used.
Promo Code Used	Boolean indicating if a promo code was applied.
Previous Purchases	Count of all prior purchases by the customer.
Payment Method	Method of payment (e.g., Credit Card, PayPal).
Frequency of Purchases	Categorical frequency (e.g., Weekly, Monthly).

## 3 Data Engineering Pipeline

### 3.1 Python Preprocessing

The raw dataset was rigorously cleaned and preprocessed using Python libraries, primarily Pandas and NumPy, as documented in the `Customer_Shopping_Behaviour_Analysis.ipynb` notebook. Key steps included:

- Identifying and handling missing values.
- Encoding categorical variables for analysis.
- Engineering new features, such as 'Age Group' (e.g., Young Adult, Middle-aged, Senior), to facilitate segmentation.
- Removing duplicate entries to ensure data accuracy.

### 3.2 Database Loading

Following preprocessing, the cleaned Pandas DataFrame was loaded into a PostgreSQL database named `customer_behavior` using the `sqlalchemy` library. This created a persistent, queryable table named `customer`, which serves as the single source of truth for all subsequent SQL analysis and Power BI visualization.

## 4 SQL Querying for Business Insights

The `customer_beaviour_analysis.sql` file contains numerous queries used to extract insights. These queries validate dashboard findings and uncover deeper trends. Below are key examples.

### Query 1: Subscriber Value Analysis (SQL Q5)

To quantify the impact of the subscription program, we compared the Average Order Value (AOV) and total revenue between subscribers and non-subscribers.

```
SELECT subscription_status,
       COUNT(customer_id) AS total_customers,
       ROUND(AVG(purchase_amount),2) AS average_spend,
       SUM(purchase_amount) AS total_revenue
  FROM customer
 GROUP BY subscription_status
 ORDER BY total_revenue, average_spend DESC;
```

### Query 2: Customer Segmentation (SQL Q7)

We segmented customers based on their purchase history to identify high-value "Loyal" customers versus "New" and "Returning" ones.

```
WITH customer_type AS (
    SELECT customer_id, previous_purchases,
    CASE
        WHEN previous_purchases = 1 THEN 'New'
        WHEN previous_purchases BETWEEN 2 AND 10 THEN 'Returning'
        ELSE 'Loyal'
    END AS customer_segment
   FROM customer
```

```

)
SELECT customer_segment, COUNT (*) AS "Number of customers"
FROM customer_type
GROUP BY customer_segment;

```

### Query 3: Top Products per Category (SQL Q8)

This query was used to identify the best-selling items within each category, allowing for granular inventory and marketing decisions.

```

WITH item_counts AS (
    SELECT category, item_purchased,
    COUNT(customer_id) AS total_orders,
    ROW_NUMBER() OVER(PARTITION BY category
        ORDER BY COUNT(customer_id) DESC) AS item_rank
    FROM customer
    GROUP BY category, item_purchased
)
SELECT item_rank, category, item_purchased, total_orders
FROM item_counts
WHERE item_rank <= 3;

```

## 5 Detailed Dashboard Analysis

The Power BI dashboard (represented on the title page, Figure ) aggregates all findings into an interactive narrative.

### 5.1 Key Performance Indicators (KPIs)

The dashboard highlights four crucial top-line metrics for at-a-glance understanding:

- **Total Customers:** 3.90K
- **Total Revenue:** \$233K
- **Average Purchase Amount:** \$59.76
- **Average Review Rating:** 3.75

### 5.2 Revenue and Sales by Category

The horizontal bar charts for "Revenue By Category" and "Sales By Category" both confirm that **Clothing** is the dominant category, generating the most revenue and the highest number of sales. **Accessories** is the clear second-place performer, followed by **Footwear** and **Outerwear**.

### 5.3 Subscriber Analysis

The "% of Subscribers" donut chart reveals that **27%** of the customer base are subscribers. While a minority, the SQL analysis (Section 4.1) confirms this group is highly valuable, spending more on average and contributing significantly to total revenue.

## 5.4 Revenue and Sales by Age Group

The "Revenue By Age Group" and "Sales By Age Group" charts clearly identify **Young Adults** as the most valuable customer segment, leading in both total spending and number of transactions. This group is followed by "Middle-aged" customers. Interestingly, "Seniors" account for more sales than "Adults," but "Adults" generate slightly more revenue, indicating a higher AOV for the "Adult" segment.

## 5.5 Dynamic Filtering

The filters on the left-hand side of the dashboard (Gender, Subscription, Category, Shipping Type) allow for dynamic segmentation. For example, a user can drill down to see that "Female" customers dominate spending in "Clothing" or that "Free Shipping" is the most popular shipping type.

# 6 Additional SQL Insights

The SQL analysis provided deeper insights not immediately visible on the main dashboard.

## 6.1 Top Rated Products (SQL Q3)

The top 5 products by average review rating were identified, providing a "best-of" list that can be used for marketing campaigns (e.g., "Customer Favorites").

## 6.2 Discount & Promo Code Impact (SQL Q6)

Analysis showed which products have the highest purchase rate using discounts. This helps determine if discounts are successfully driving sales for specific items or if margin is being unnecessarily eroded.

## 6.3 Shipping Type vs. Spend (SQL Q4)

A comparison between "Standard" and "Express" shipping showed a negligible difference in average purchase amount, suggesting that customers' spending habits are not strongly influenced by the choice between these two paid shipping tiers. This finding, combined with the popularity of "Free Shipping," emphasizes the importance of delivery cost over speed.

# 7 Business Impact & Outcomes

Based on the analysis, several key business outcomes have been achieved:

- Defined clear high-value customer segments (e.g., Young Adult Subscribers) for precision promotional targeting.
- Quantified the value of the subscription program, identifying subscribers as the most profitable customer group.
- Uncovered low-performing categories ('Outerwear') and seasonal lulls where strategic improvements can be applied.
- Highlighted the importance of discounting and promo codes in influencing purchase behaviour, especially for new customers.
- Enabled the business to optimize inventory planning and marketing spend by aligning with seasonal and category-specific demand.

## 8 Recommendations & Future Actions

To maximize business performance based on these findings, the following actions are recommended:

- **Enhance subscription benefits** to encourage more sign-ups, possibly by offering exclusive access or deeper shipping discounts.
- **Strengthen inventory allocation** for 'Clothing' and 'Accessories' to prevent stockouts during peak demand.
- **Improve marketing focus on Young Adults** with personalized offers and campaigns on platforms this demographic frequents.
- **Optimize shipping services** by promoting fast and free delivery options as a key value proposition.
- **Launch "Top Rated" campaigns** using the insights from SQL Query 3 to leverage social proof.
- **Introduce loyalty rewards** based on purchase frequency (from SQL Query 7) and review ratings to further engage high-value customers.

## 9 Conclusion

This customer behaviour analysis demonstrates the power of a data-driven, multi-tool approach. By combining Python for data analytics, SQL for deep business querying, and Power BI for interactive visualization, we have gained a 360-degree understanding of customer preferences. These insights enable the business to move from reactive to proactive decision-making, optimizing marketing, inventory, and customer retention strategies for sustained growth.