Customer Behavioural Segmentation & Revenue Analysis

Project Overview

This project delivers a structured analysis of customer transactional data to uncover revenue patterns and behavioural segments. Leveraging SQL-based aggregation and quantile-driven segmentation logic, customers were classified into strategic segments based on purchasing frequency and total spending contribution.

The resulting framework enables data driven identification of high value customers, retention priorities, and growth opportunities. Insights were consolidated into an executive level Power BI dashboard designed to support strategic business decision making.

1. Business Objective

The goal of this project is to move beyond descriptive sales reporting and develop a behavioural customer segmentation framework that enables identification of high value customers, potential growth segments, and at-risk groups contributing to revenue leakage.

2. Dataset Summary

- Rows: 3,900
- Columns: 18
- Customer demographics (Age, Gender, Location, Subscription Status)
- Purchase details (Item Purchased, Category, Purchase Amount, Season, Size,Color)
- Shopping behaviour (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 values in Review Rating column

3. Data Preparation

Raw transactional data was first processed using Python (Pandas) to ensure consistency, accuracy, and analytical readiness prior to SQL-based analysis.

The following preparation steps were performed:

- **Data Inspection:** Evaluated dataset structure, column types, and summary statistics to understand distribution and identify inconsistencies.

- **Missing Value Handling:** Identified missing values in the *Review Rating* column and applied median imputation at the product category level to preserve rating integrity without skewing distribution.
- **Column Standardisation:** Renamed columns to snake_case format to improve readability and ensure SQL compatibility.
- **Feature Engineering:**
    - Created an `age_group` column by binning customer ages for demographic analysis.
    - Derived structured customer-level metrics for segmentation.
- **Data Consistency Validation:** Evaluated overlap between discount-related variables and removed redundant fields to avoid duplication in analysis.
- **Data Export:** Cleaned dataset was exported and loaded into PostgreSQL for structured SQL-based business analysis.

4. SQL Analysis

Following data preparation, structured SQL analysis was performed in PostgreSQL to answer key business questions and uncover revenue driving patterns.

The analysis focused on the following areas:

Revenue & Category Performance

- Calculated total revenue by product category.
- Identified top performing categories contributing to overall sales.
- Compared revenue distribution across demographic segments.

Subscription Behaviour Analysis

- Evaluated revenue contribution from subscribers vs non-subscribers.
- Measured average purchase value differences across subscription status.
- Assessed the relationship between purchase frequency and subscription likelihood.

Discount & Purchase Behaviour

- Identified products with high discount dependency.
- Analysed spending behaviour among discount users.
- Evaluated impact of promotional activity on purchase amounts.

Customer-Level Aggregation

Customer-level metrics were derived using aggregation functions to compute:

- Total orders per customer
- Total spending contribution
- Average order value
- Lifetime purchase frequency

- Common Table Expressions (CTEs) were used to structure intermediate transformations and improve query readability.

Behavioural Segmentation (FM Model)

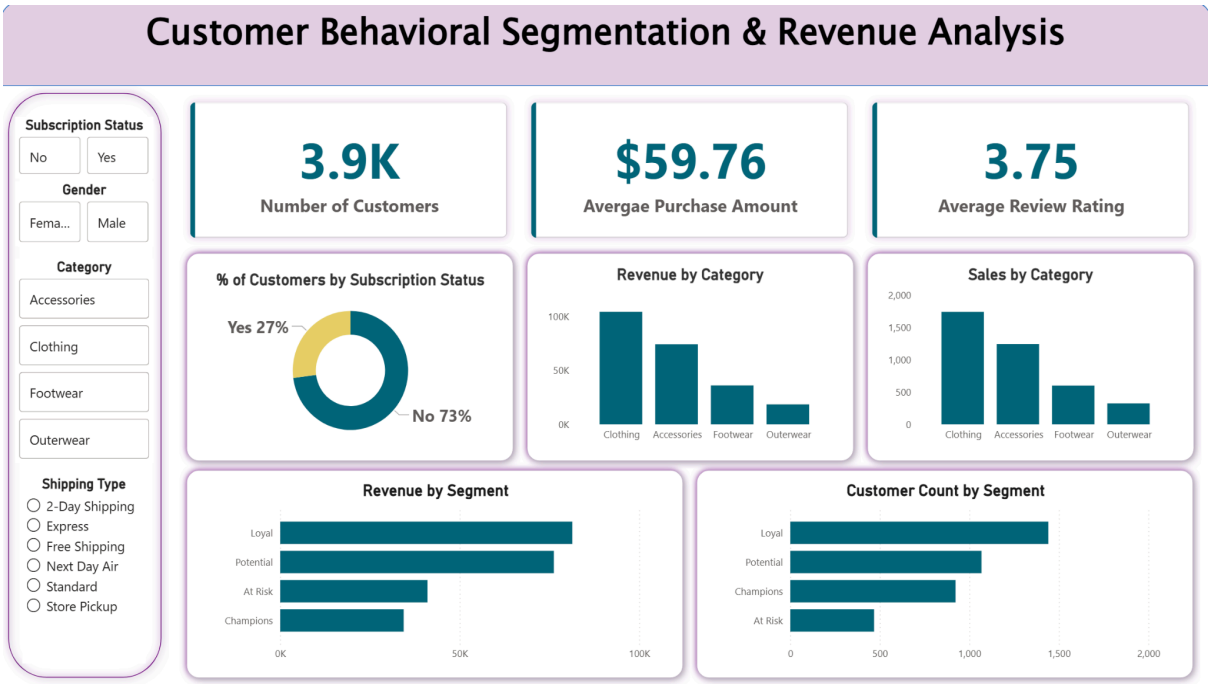
Due to the absence of timestamp-based recency data, a Frequency–Monetary (FM) segmentation model inspired by RFM was implemented.

- Customers were ranked using **NTILE window functions** to create quantile-based frequency and monetary scores.
- A combined score was computed to classify customers into behavioural segments:
    - Champions
    - Loyal
    - Potential
    - At Risk

This segmentation framework enabled identification of high-value customers and retention-priority groups.

5. Dashboard Overview

An interactive Power BI dashboard was developed to translate analytical findings into a clear and decision-oriented visual framework. The dashboard consolidates revenue metrics, behavioural insights, and customer segmentation results into a structured layout designed for executive-level consumption.



6. Key Insights & Business Implications

The analysis revealed several strategic insights relevant to revenue optimization and customer management:

- **Loyal customers drive the highest total revenue**, primarily due to their large population size and consistent purchasing behaviour. Retention strategies should focus on maintaining engagement within this segment.
- **Champions represent high-engagement customers**, scoring strongly on both frequency and spending metrics. Although smaller in number, they are strategically valuable and suitable for premium loyalty initiatives.
- The **Potential segment presents growth opportunities**, with strong behavioral indicators suggesting conversion into higher-value segments through targeted campaigns.
- **At Risk customers contribute lower revenue and weaker engagement**, indicating the need for retention-focused interventions to reduce churn risk.