Customer Shopping Behavior Analysis

# Project Overview

This project analyzes customer shopping behavior using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

# Dataset Summary

* Rows: 3,900 - Columns: 18 - Key Features:
* Customer demographics (Age, Gender, Location, Subscription Status)
* Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
* Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of

Purchases, Review Rating, Shipping Type)

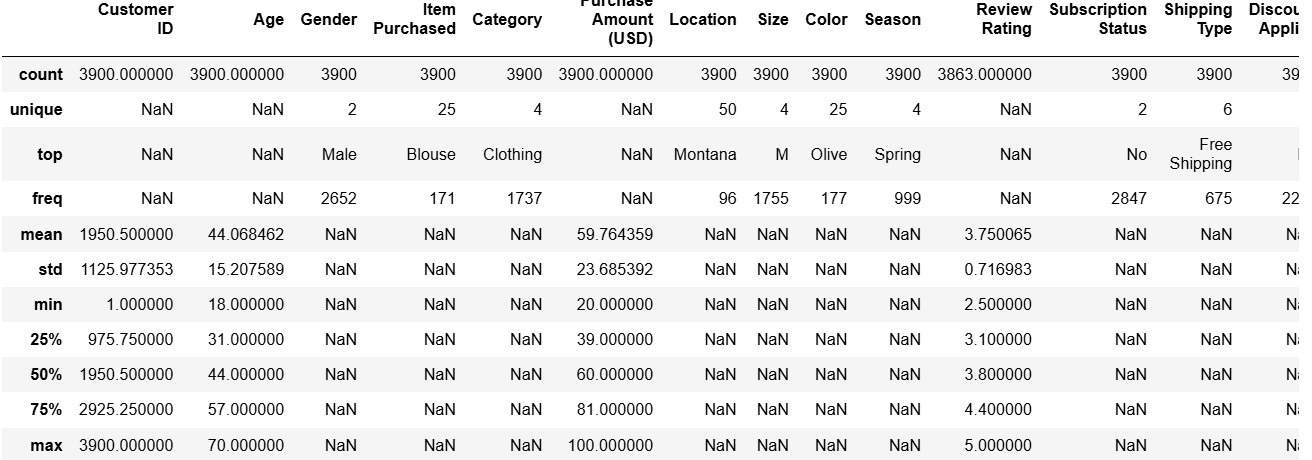
* Missing Data: 37 values in Review Rating column

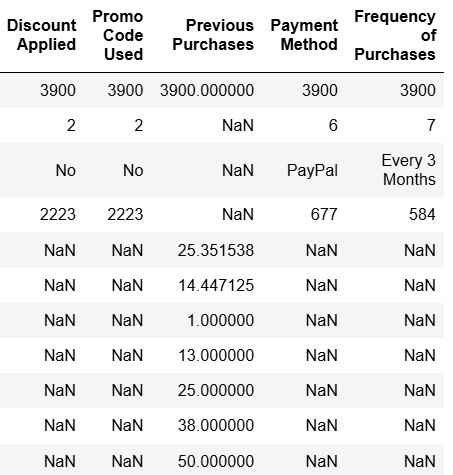
# Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

* **Data Loading:** Imported the dataset using pandas.

* **Initial Exploration:** Used df.info() to check structure and .describe() for summary statistics.





* **Missing Data Handling:** Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

* **Column Standardization:** Renamed columns to **snake case** for better readability and documentation.

* **Feature Engineering:**

○ Created **age\_group** column by binning customer ages.

○ Created **purchase\_frequency\_days** column from purchase data.

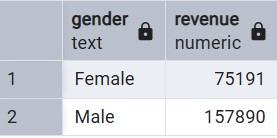
* **Data Consistency Check:** Verified if discount\_applied and promo\_code\_used were redundant; dropped promo\_code\_used.

* **Database Integration:** Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis.

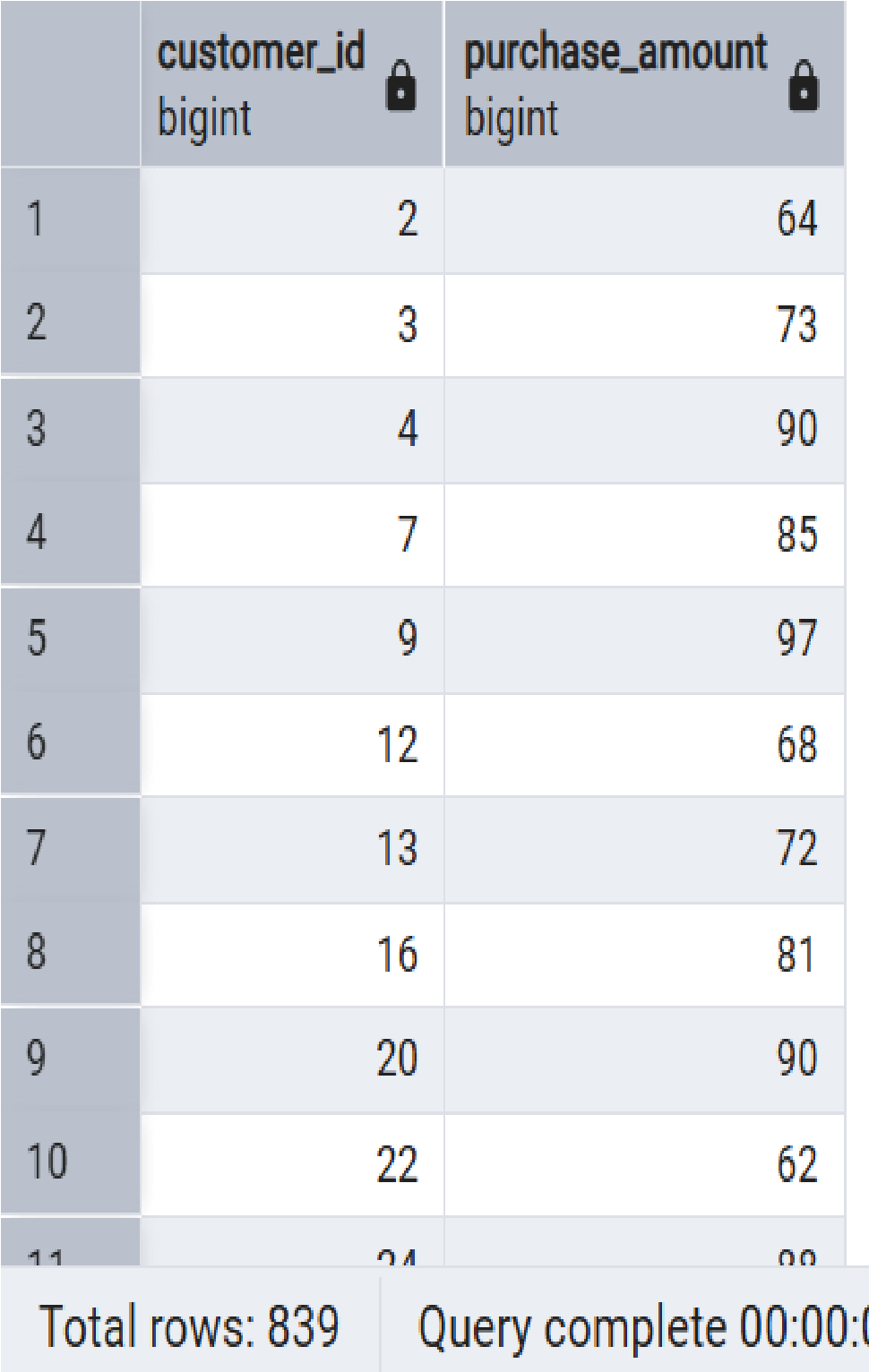
# Data Analysis using SQL (Business Transactions)

We performed structured analysis in PostgreSQL to answer key business questions:

1. **Revenue by Gender** – Compared total revenue generated by male vs. female customers.



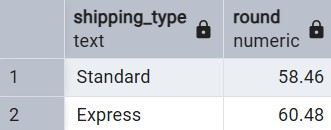
1. **High-Spending Discount Users** – Identified customers who used discounts but still spent above the average purchase amount.



1. **Top 5 Products by Rating** – Found products with the highest average review ratings.



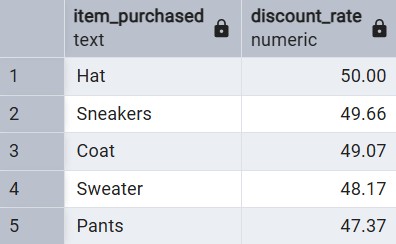
1. **Shipping Type Comparison** – Compared average purchase amounts between Standard and Express shipping.



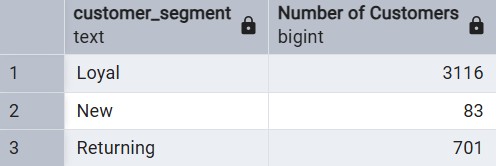
1. **Subscribers vs. Non-Subscribers** – Compared average spend and total revenue across subscription status.



1. **Discount-Dependent Products** – Identified 5 products with the highest percentage of discounted purchases.



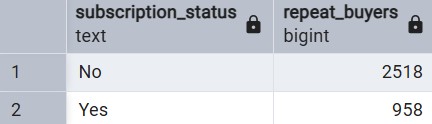
1. **Customer Segmentation** – Classified customers into New, Returning, and Loyal segments based on purchase history.



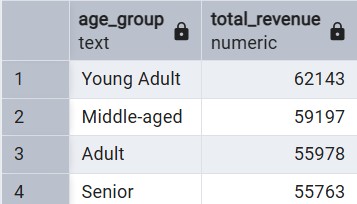
1. **Top 3 Products per Category** – Listed the most purchased products within each category.



1. **Repeat Buyers & Subscriptions** – Checked whether customers with >5 purchases are more likely to subscribe.



1. **Revenue by Age Group** – Calculated total revenue contribution of each age group.



# Dashboard in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.



# Business Recommendations

**1. Boost Subscription Engagement**

* Problem: Low number of subscribed customers despite higher spending among them.
* Recommendation: Launch membership tiers with exclusive benefits like free express shipping, early access, and loyalty bonuses to increase retention and recurring revenue.

**2. Strengthen Customer Loyalty**

* Problem: The Majority of customers are “New” or “Returning” with limited repeat purchases.
* Recommendation: Introduce a loyalty rewards program with points, cashback, and personalized offers to encourage frequent buying and long-term engagement.

**3. Optimize Discount Strategy**

* Problem: Heavy reliance on discounts reduces profit margins.
* Recommendation: Implement targeted discounts for loyal or high-value customers and promote product quality to reduce overdependence on price cuts.

**4. Enhance Shipping Efficiency**

* Problem: Express shipping users generate more revenue but face limited incentives.
* Recommendation: Offer free or discounted express shipping for high-value orders or subscribers to improve satisfaction and increase order value.

**5. Focus on High-Performing Categories**

* Problem: Uneven performance across product categories.
* Recommendation: Priorities inventory and marketing for top-performing segments like Clothing and Accessories while reviewing low-performing ones for improvement.

**6. Improve Customer Experience & Reviews**

* Problem: Average review rating of 3.75 indicates moderate satisfaction.
* Recommendation: Collect post-purchase feedback, incentivize honest reviews, and resolve common complaints to enhance service quality and brand trust.