Customer Shopping Behavior Analysis – End-to-End Data Project

1. Introduction

This project analyzes customer shopping behavior using Python, SQL, and Power BI. The goal is to understand revenue patterns, customer segmentation, product performance, and the impact of discounts and subscriptions on purchasing behavior.

2. Tools & Technologies

• Python – Data cleaning & feature engineering
• Pandas – Data manipulation
• MySQL – Data storage & analytical queries
• SQLAlchemy – Python–MySQL connection
• Power BI – Data visualization
• GitHub – Version control & documentation

3. Dataset Overview

Rows: 3900
Columns: 18

Key fields:
customer_id, age, gender, category, purchase_amount, location, review_rating, subscription_status, shipping_type, discount_applied, previous_purchases, payment_method, frequency_of_purchases

4. Python Data Cleaning & Feature Engineering

Step 1: Load Data
Customer_data = pd.read_csv('customer_shopping_behavior.csv')

Step 2: Exploratory Analysis
Used .shape(), .describe()
Checked missing values using .isnull().sum()

Step 3: Handle Missing Values
37 null values found in review_rating
Filled using median

Step 4: Standardize Column Names
Converted to lowercase and replaced spaces with underscores

Step 5: Feature Engineering

Created age_group using quartiles
Mapped purchase frequency into numeric days

Step 6: Dimensional Reduction
Removed promo_code_used column as it matched discount_applied

5. Load Data into MySQL

Used SQLAlchemy engine to load cleaned data into MySQL database 'CD' and created table cust_data

6. SQL Analysis

Key business queries executed:
• Revenue by gender
• High spenders using discount
• Top rated products
• Shipping comparison
• Subscription impact
• Discount usage percentage
• Customer segmentation
• Top products per category

7. Power BI Dashboard

Page 1 – Business Overview
• Total Revenue
• Revenue by Gender
• Revenue by Location
• Seasonal Trend

Page 2 – Customer Behavior
• Age group vs spend
• New vs Returning customers
• Payment preferences
• Subscription impact

Page 3 – Product Insights
• Top & bottom products
• Discount effectiveness
• Shipping preference
• Size-color heatmap

8. Key Insights

• Male customers contribute slightly higher revenue
• Subscribed users spend more on average
• Discounts increase purchase volume
• Weekly buyers generate highest revenue
• Clothing category dominates sales

9. Business Recommendations

• Expand loyalty programs
• Personalized campaigns for high-value age groups
• Optimize discount strategy
• Focus inventory on top-selling products

10. Conclusion

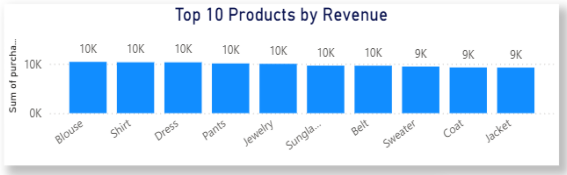
This project demonstrates my ability to clean data using Python, perform advanced SQL analysis, build dashboards in Power BI, and translate data into actionable insights.

11. Skills Demonstrated

✔ Python (Pandas, SQLAlchemy)
✔ SQL (CTE, Window Functions)
✔ Data Modeling
✔ Power BI Visualization
✔ Business Analysis

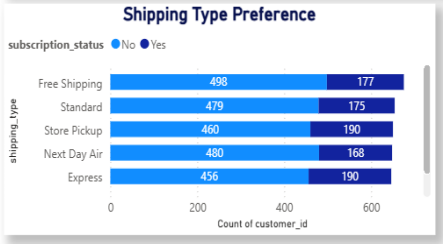
Power BI Dashboard Screenshots

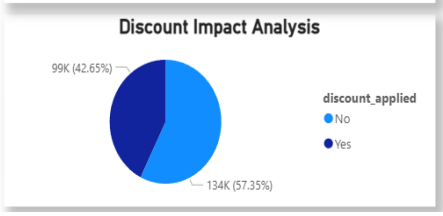
Product & Promotion Insights



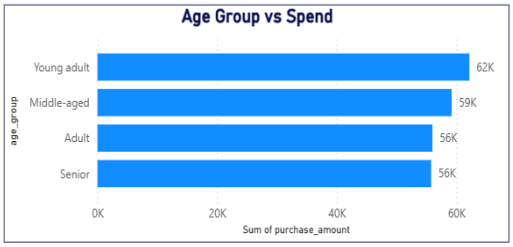
Category Performance Matrix

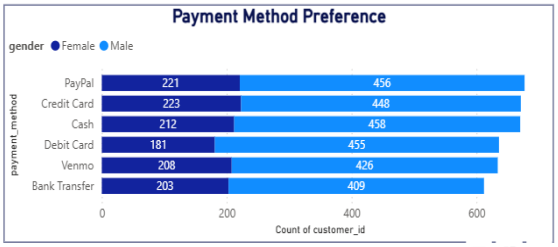
category	Sum of purchase_amount	Average of purchase_amount	Average of review_rating
Clothing	104264	60.03	3.72
Accessories	74200	59.84	3.77
Footwear	36093	60.26	3.79
Outerwear	18524	57.17	3.75
Total	**233081**	**59.76**	**3.75**

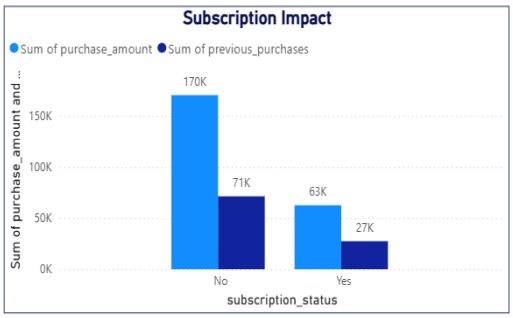


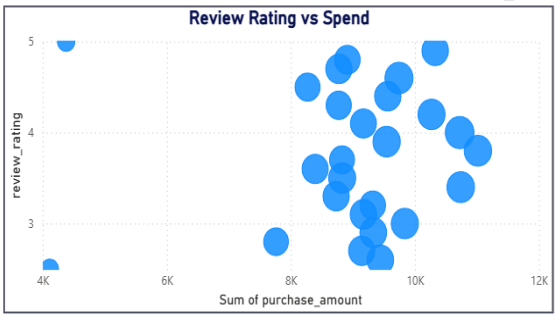


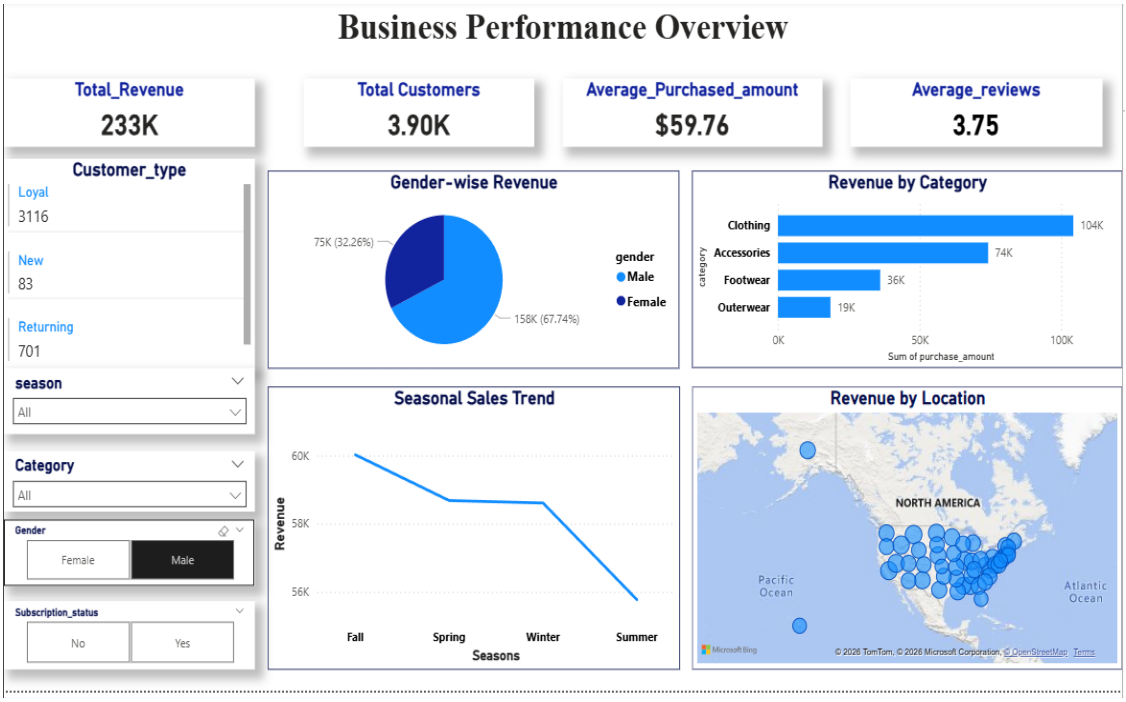
Customer Behavior Analysis











Project by: Daksh Makhija