Customer Shopping Behavior Analysis Project Report

I. Project Overview

Goal: To analyze a retail company's customer shopping behavior data to identify key trends, understand customer segments, and provide data-driven recommendations to improve sales, customer engagement, and long-term loyalty.

Business Problem Statement: A leading retail company has noticed changes in purchasing patterns and is particularly interested in uncovering which factors (such as discounts, reviews, seasons, or payment preferences) drive consumer decisions. The objective is to leverage the customer shopping data to identify trends, improve customer engagement, and optimize marketing and product strategies.

Dataset: The analysis is based on a customer shopping behavior dataset, where each row represents one customer and their latest purchase.

- **Number of Records (Rows): 3,900**
- **Number of Features (Columns): 18 (e.g., Gender, Item Purchased, Purchase Amount (USD), Review Rating, Subscription Status, etc.)**

II. Tools and Technologies

Phase	Tool	Purpose
Data Cleaning/EDA	Python (Pandas)	Initial data loading, cleaning (handling missing values in review_rating column), feature engineering, and data formatting.
Advanced Analysis	PostgreSQL / SQL	Deep-dive analysis and answering specific business questions using advanced SQL queries, subqueries, and window functions.
Visualization/Reporting	Power BI	Creation of an interactive dashboard to visualize key performance indicators (KPIs) and insights.
Documentation	GitHub	Version control and public portfolio showcase of the entire project.

III. Data Preparation & Setup

The raw data was processed in a Python Jupyter Notebook (customer_behaivor_analysis.ipynb) for cleaning and transformation before being loaded into the database.

Key Data Cleaning and Feature Engineering Steps:

1. **Column Formatting:** All column names were converted to **snake_case** for consistency and ease of use in SQL (e.g., Customer ID became customer_id, Purchase Amount (USD) became purchase_amount).
2. **Missing Value Imputation:** Missing values in the review_rating column were imputed using the **median rating for each product category**, a method more accurate than using the overall median.
3. **Feature Engineering:**
    - A new column, age_group, was created to segment customers into Young Adult, Adult, Middle-Aged, and Senior categories.
    - The categorical frequency_of_purchases column was mapped to a numerical purchase_frequency_days column for easier quantitative analysis.
4. **Redundant Column Removal:** The promo_code_used column was dropped as it contained identical information to discount_applied.
5. **Database Loading:** The cleaned and transformed data was loaded into a PostgreSQL database named customer_behavior into a table named customer.

IV. SQL Analysis and Key Findings

The initial query to view the table structure:

```
select * from customer limit 20
```

Q1. What is the total revenue generated by male vs female customers?

SQL Command:

```
select gender, SUM(purchase_amount) as revenue
from customer
group by gender
```

	gender text	revenue numeric
1	Female	75191
2	Male	157890

Insight: Total revenue is almost equally split between genders, indicating both are critical segments for the business. Marketing campaigns should continue to target both groups with relevant products.

Q2. Which customer used a discount but still spent more than the avg purchase amount?

SQL Command:

```
select customer_id, purchase_amount

from customer

where discount_applied = 'Yes' and purchase_amount >= (select AVG(purchase_amount)
from customer)
```

customer_id bigint	purchase_amount bigint
2	64
3	73
4	90
7	85
9	97
12	68
13	72
16	81
20	90

ows: 839 Query complete 00:00:00.182

Insight: This segment represents high-value, discount-sensitive customers. Discounts act as a powerful final incentive for them to complete a large purchase. These customers should be targeted with premium product offers with high-value discounts.

Q3. Which are the top 5 products with the highest avg review rating?

SQL Command:

```
select item_purchased, ROUND(AVG(review_rating::numeric),2) as "Average Product Rating"

from customer
```

group by item_purchased

order by avg(review_rating) desc

limit 5;

item_purchased text	Average Product Rating numeric
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.80
Skirt	3.78

Insight: These top-rated products should be highlighted in marketing campaigns and used to build brand reputation. Products with consistently high ratings can also command a premium price or be bundled with newer items.

Q4. Compare the avg purchase amounts b/w standard and express shipping.

SQL Command:

select shipping_type,

ROUND(AVG(purchase_amount),2)

from customer

where shipping_type in ('Standard', 'Express')

group by shipping_type

shipping_type text	round numeric
Standard	58.46
Express	60.48

Insight: Customers opting for **Express** shipping have a slightly higher average purchase amount. This indicates that customers prioritizing speed are often willing to spend more. The business should consider promoting and optimizing the Express shipping option to increase the overall Average Order Value (AOV).

Q5. Do subscribed customers spend more?

SQL Command:

select subscription_status,

COUNT(customer_id) as total_customers,

ROUND(AVG(purchase_amount),2) as avg_spend,

ROUND(SUM(purchase_amount),2) as total_revenue

from customer

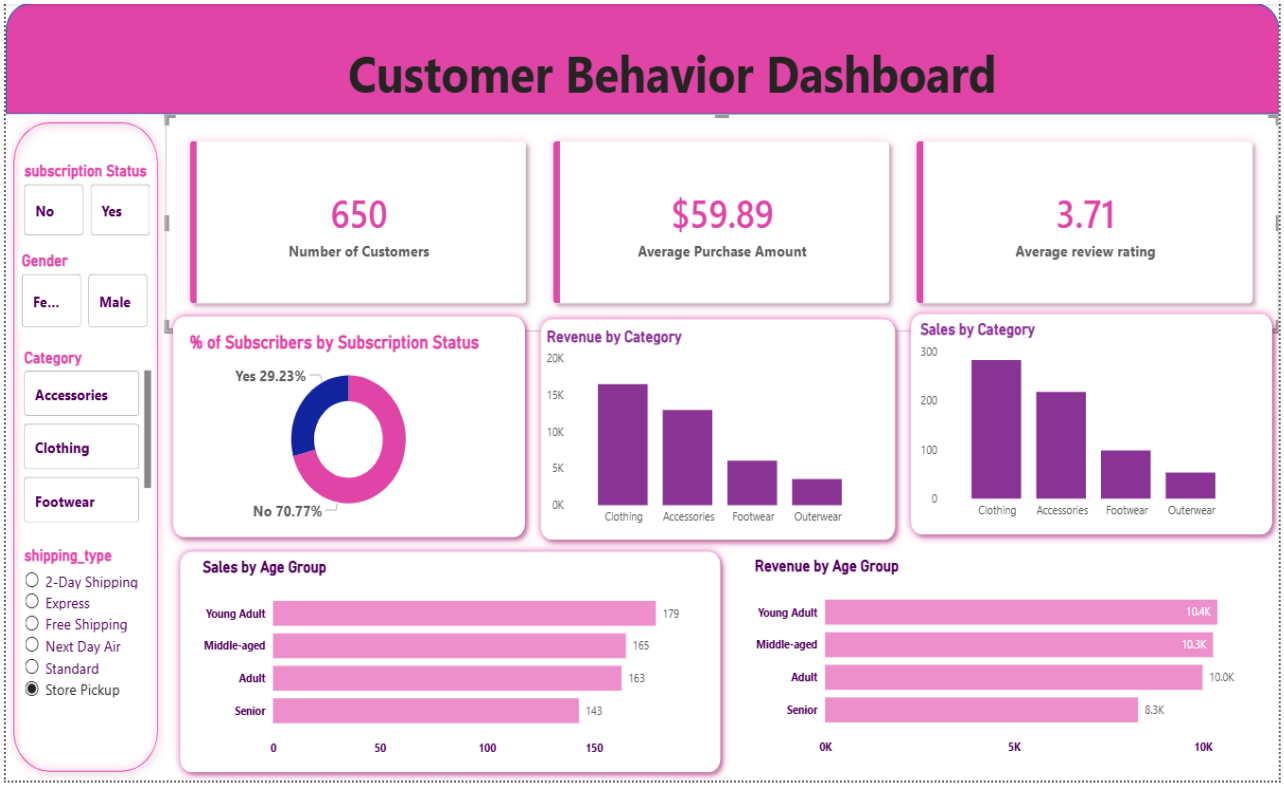
group by subscription_status

order by total_revenue, avg_spend desc;

subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
Yes	1053	59.49	62645.00
No	2847	59.87	170436.00

Insight: Non-subscribed customers currently generate higher overall revenue and have a slightly higher average spend than subscribed customers. This suggests the current subscription benefits may not be compelling enough to drive a significant increase in transaction value. The subscription program should be immediately reviewed and revised to incentivize larger purchases.

CUSTOMER BEHAVIOUR DASHBOARD



V. Power BI Dashboard

The dashboard visually summarizes the findings, including:

- **KPI Cards:** Total Customers, Average Purchase Amount, Average Review Rating.

- **Charts:** Revenue by Category, Sales by Category, Revenue by Age Group.
- **Segmentation:** Distribution of Customers by Subscription Status and Gender.
- **Interactivity:** Slicers for Gender, Subscription Status, Shipping Type, and Category.

VI. Business Recommendations

Based on the comprehensive data analysis, the following strategic recommendations are proposed:

1. **Re-evaluate Subscription Program:** The current subscription model is not translating into higher average spend. **Recommendation:** Introduce **tiered benefits** based on purchase volume or exclusive access to premium, high-margin products to boost the average transaction value for subscribers.
2. **Target High-Value Discount Users:** The segment of customers who spend above average even when using a discount should be treated as a high-potential loyalty group. **Recommendation:** Offer highly personalized, **time-sensitive promotions** on premium or full-price items, framing the discount as a reward rather than a necessity.
3. **Capitalize on Express Shipping Preference:** Customers using Express Shipping have a higher AOV. **Recommendation:** Promote and prioritize the Express Shipping option more aggressively, perhaps by offering it free for orders above a high-value threshold to further increase overall sales.
4. **Promote Top-Rated Products:** Use the top 5 highest-rated products (Gloves, Sandals, Boots, Hat, Skirt) as **brand pillars** in all marketing materials. **Recommendation:** Leverage customer testimonials and ratings directly in advertising to build trust and brand quality perception.
5. **Develop Age-Specific Marketing:** The analysis of revenue by age group should be leveraged to tailor product offerings and marketing channels to the most valuable demographic (e.g., Young Adults often drive the highest revenue).