Customer Shopping Behavior Analysis

1. Project Overview

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

2. Dataset Summary

- Rows: 3,900
- Columns: 18
- Key Features include customer demographics, purchase details, and shopping behavior indicators.
- Missing Data: 37 values identified in the Review Rating column.

3. Exploratory Data Analysis using Python

- Data preprocessing and exploratory analysis were performed using Python in VS Code.
- Data Loading: Dataset imported using pandas.
- Initial Exploration: Structure and statistics reviewed using df.info() and df.describe().
- Missing Value Treatment: Review Rating null values imputed using the median rating per product category.
- Column Standardization: Column names converted to snake_case format.
- Feature Engineering: – Created age_group column by binning customer ages. – Derived purchase_frequency_days based on historical purchase behavior.
- Data Validation: Identified redundancy between discount_applied and promo_code_used and removed promo_code_used.
- Database Integration: Cleaned dataset loaded into a MySQL database using MySQL Workbench, with all SQL queries executed directly from VS Code.

4. Data Analysis using SQL (MySQL)

SQL-based business analysis was conducted using MySQL Workbench, with queries executed in VS Code to extract insights:

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

gender	revenue
Male	157890
Female	75191

2. High-Spending Discount Users – Identified customers who used discounts but spent above the average purchase value.

customer_id	purshase_amount
2	64
3	73
4	90
7	85
9	97

3. Top 5 Products by Rating – Retrieved products with the highest average customer ratings.

item_purchased	Average Product Rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Skirt	3.78

4. Shipping Type Comparison – Compared average purchase amounts between standard and express shipping.

shipping_type	avg_purchase
Express	60.48
Standard	58.46

5. Subscribers vs Non-Subscribers – Analyzed average spend and revenue contribution based on subscription status.

subscription_status	total_customers	avg_spend	total_revenue
No	2847	59.87	170436
Yes	1053	59.49	62645

6. Discount-Dependent Products – Identified products with the highest percentage of discounted purchases.

item_purchased	discount_rate
Hat	50
Sneakers	49.66
Coat	49.07
Sweater	48.17
Pants	47.37

7. Customer Segmentation – Segmented customers into New, Returning, and Loyal groups based on purchase count.

customer_segment	Number of Customers
Loyal	3116
Returning	701
New	83

8. Top 3 Products per Category – Determined most frequently purchased products in each category.

item_rank	category	item_purchased	total_orders
1	Accessories	Jewelry	171
2	Accessories	Sunglasses	161
3	Accessories	Belt	161
1	Clothing	Blouse	171
2	Clothing	Pants	171
3	Clothing	Shirt	169
1	Footwear	Sandals	160
2	Footwear	Shoes	150
3	Footwear	Sneakers	145
1	Outerwear	Jacket	163
2	Outerwear	Coat	161

9. Repeat Buyers & Subscriptions – Evaluated subscription likelihood among customers with more than five purchases.

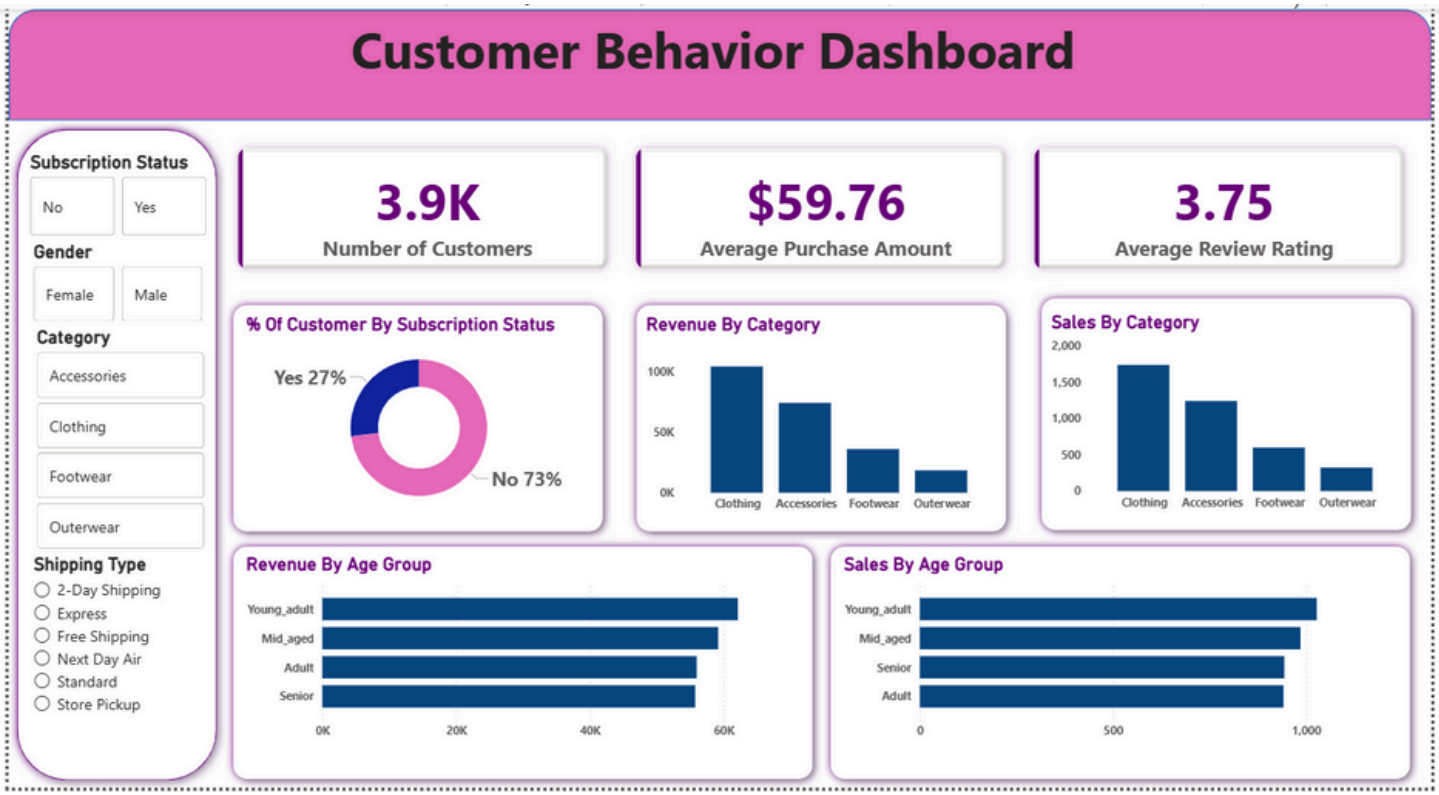
subscription_status	repeat_buyers
Yes	958
No	2518

10. Revenue by Age Group – Calculated revenue contribution across defined age groups.

age_group	total_revenue
Young_adult	62143
Mid_aged	59197
Adult	55978
Senior	55763

5. Dashboard in Power BI

An interactive Power BI dashboard was created to visualize key insights, trends, and customer segments for business stakeholders.



6. Business Recommendations

- Promote subscription plans with exclusive benefits to increase recurring revenue.
- Implement customer loyalty programs targeting repeat buyers.
- Optimize discount strategies to balance revenue growth and profit margins.
- Highlight top-rated and best-selling products in marketing campaigns.
- Focus targeted marketing on high-revenue age groups and express shipping users.