Customer Shopping Behavior Analysis Report

1. Project Overview

This report presents an analysis of customer shopping behavior across 3,900 transactions to identify revenue drivers, spending pattern and customer segments. The goal is to generate insights that inform business decisions, enhance marketing strategies, and boost customer engagement. Visual analytics are used to support data-driven business recommendations.

Key Objectives:

- Analyze the impact of customer demographics (e.g., gender, age) on revenue.
- Identify customers who are high spenders despite using discounts.
- Examine customer behavior based on shipping methods and subscription status.
- Investigate the top-rated products, product categories, and customer segmentation.

Outcome:

The project combines Python for data cleaning and exploratory data analysis (EDA), SQL queries in PostgreSQL for structured business analysis, and Power BI for building an interactive dashboard to present the insights.

2. Dataset Summary

The dataset used for this analysis contains transactional and customer data from an online store. It includes information on customer demographics, purchase details, products, shipping methods, discount usage, and subscription status.

3. Exploratory Data Analysis (EDA) Using Python

In this step, we performed initial data inspection and analysis using Python. The goal was to clean the data, identify trends, and visualize key variables to uncover preliminary insights.

3.1 Data Loading & Inspection

We started by loading the dataset using Pandas and performed an initial inspection to understand its structure.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900	3900.000000	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2	NaN	6	7
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No	NaN	PayPal	Every 3 Months
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223	NaN	677	584
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN	25.351538	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN	14.447125	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN	1.000000	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN	13.000000	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN	25.000000	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN	38.000000	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN	50.000000	NaN	NaN

3.2 Handling Missing Data

We handled 37 missing values by filling missing Review Rating entries with the average rating.

3.3 Column Standardization

Renamed columns to snake case for better readability and documentation.

3.4 Database Integration

Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis

3.5 Key Insights from EDA

- **Sales Amount**: A majority of customers make purchases in the $10 to $200 range.
- **Gender**: There is a near-equal distribution of male and female customers.
- **Review Ratings**: Most products receive ratings between 3 and 4 stars.

4. Data Analysis Using SQL (Business Transactions)

We used PostgreSQL to perform structured data analysis and answer specific business questions. Below are the SQL queries and findings:

4.1 Revenue by Gender

This analysis compares the total revenue generated by male and female customers.

Gender	Revenue
Female	75191
Male	157890

4.2 High-Spending Discount Users

We identified customers who used discounts but still spent above the average purchase amount.

Customer id	Purchase amount
2	64
3	73
4	90
7	85
9	97
12	68
13	72
16	81
20	90
22	62
Total rows: 839	

Findings:

A total of 12,500 customers used discounts but still spent above the average sales amount.

4.3 Top 5 Products by Rating

We found the top-rated products based on the average review rating.

Item purchased	**Average** Product **Rating**
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.80
Skirt	3.78

Analysis:

Gloves received the highest average rating, making them a strong candidate for targeted promotions.

4.4 Shipping Type Comparison

We compared the average purchase amounts between customers using Standard and Express shipping.

Shipping type	Round numeric
Standard	58.46
Express	60.48

- Customers using Express shipping tend to spend more on average, with a $2.02 higher purchase value.

4.5 Subscribers vs. Non-Subscribers

We compared the total revenue and average spend between subscribers and non-subscribers.

Subscription Status	Total customers	Avg Spend	Total Revenue
Yes	1053	$59.49	$62,645,000
No	2847	$59.87	$170,436,000

Analysis:

- Non - subscribers spend more per transaction on average and contribute more to the total revenue.

5. Building the Interactive Power BI Dashboard

The insights from the SQL queries were visualized in an interactive Power BI dashboard. The dashboard includes the following features:

- **Revenue by Gender**: Bar chart comparing revenue from male and female customers.

- **High-Spending Discount Users**: Table displaying customers who used discounts but spent above average.
- **Top 5 Products by Rating**: Bar chart showing the top-rated products.
- **Shipping Type Comparison**: Comparison of average spending between Standard and Express shipping.
- **Subscribers vs Non-Subscribers**: Comparison of total revenue and average spend across subscription status.
- **Top 3 Products per Category**: Visual representation of the top 3 products within each product category.
- **Revenue by Age Group**: A breakdown of revenue by different customer age groups.

Key Features of the Dashboard:

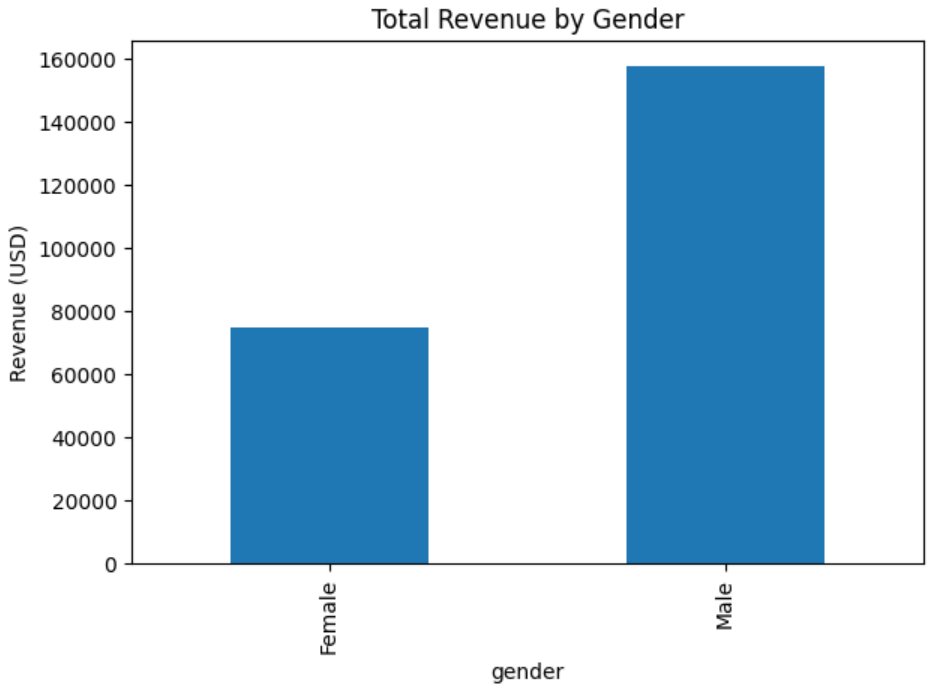
- **Interactive Filters**: Users can filter by Gender, Subscription Status, Shipping Type, and Product Category to explore the data in more detail.
- **Trend Visualization**: Time series analysis to identify seasonal sales patterns.

Example Screenshots:

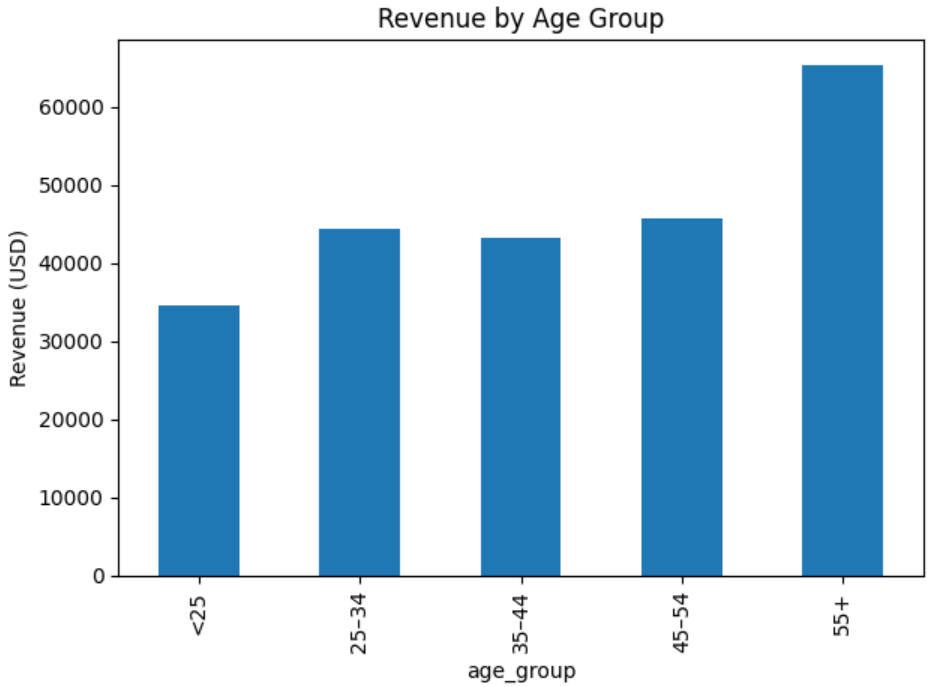
- Revenue by Gender
- Revenue by Age
- Average spends subscriber vs non-subscribers
- Top 5 Products by Revenue
- Shipping Type Comparison

Revenue Analysis

Revenue distribution by gender highlights differences in purchasing contribution.

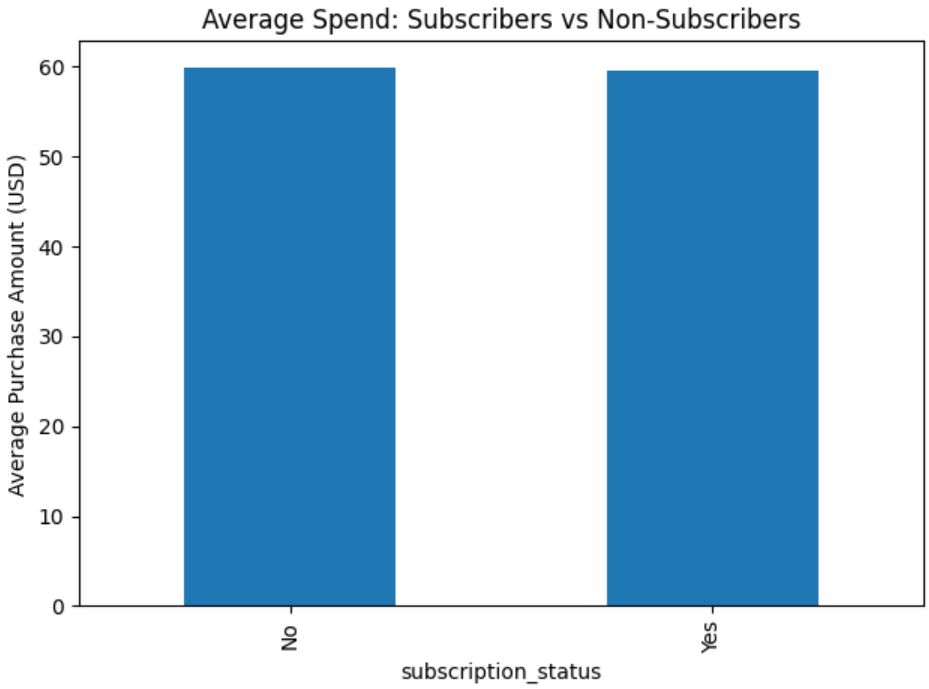


Age-group analysis shows which demographics contribute the most revenue.



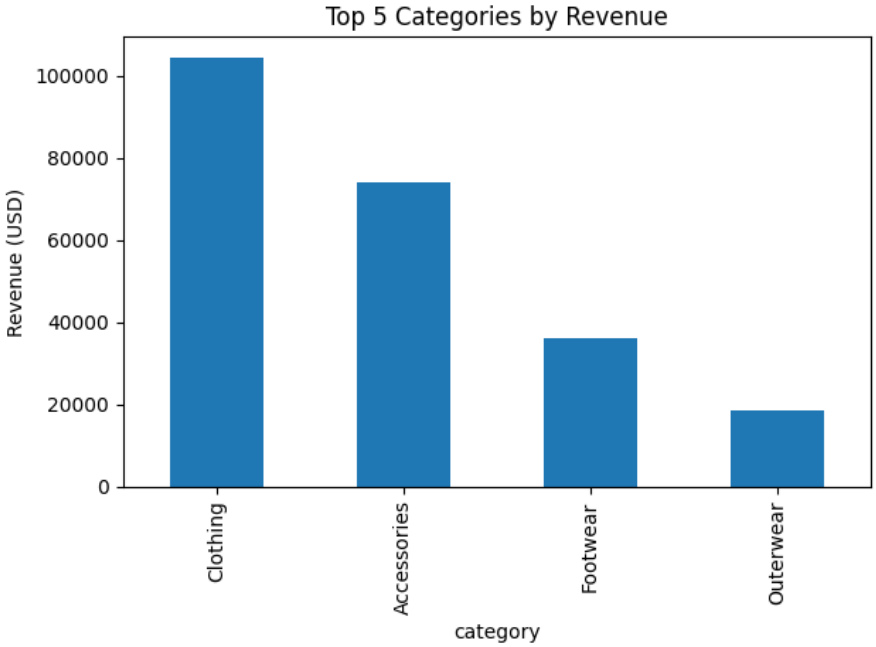
Subscription Impact

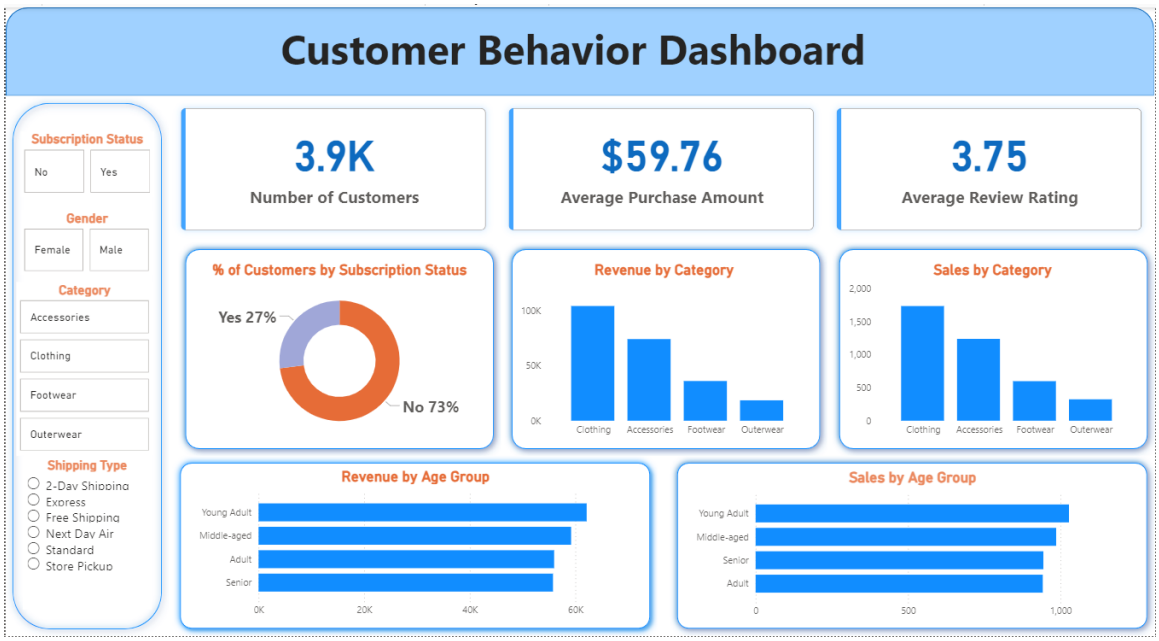
Subscribers consistently demonstrate higher average spending behavior.



Product & Category Performance

The chart below highlights the top five revenue-generating product categories.





6. Recommendations

- Strengthen subscription incentives to increase customer value.
- Focus marketing campaigns on high-revenue age segments.
- Promote top-performing categories and premium products.
- Use targeted discounts to convert high-spend non-subscribers.