



Customer Shopping Behavior Analysis

This project delves into customer shopping behavior, leveraging transactional data from 3,900 purchases across diverse product categories. Our primary objective is to unearth critical insights into spending patterns, identify distinct customer segments, understand product preferences, and analyze subscription behaviors. These findings are crucial for guiding strategic business decisions and optimizing marketing efforts.

Project Overview & Dataset Summary

Project Overview

Our analysis aims to provide a comprehensive understanding of customer interactions with our products. By examining transactional data, we seek to:

- Identify key spending trends.
- Segment customers based on their purchasing habits.
- Pinpoint popular product categories and items.
- Evaluate the impact of subscription models.

These insights will inform targeted marketing campaigns, product development, and overall business strategy.

Dataset Summary

The dataset comprises extensive transactional information:

- **Rows:** 3,900 individual purchases
- **Columns:** 18 distinct features
- **Key Features:** Customer demographics (Age, Gender, Location, Subscription Status), Purchase details (Item, Category, Amount, Season, Size, Color), Shopping behavior (Discount, Promo Code, Previous Purchases, Frequency, Review Rating, Shipping Type).
- **Missing Data:** 37 values in the 'Review Rating' column, which were carefully handled during data preparation.

Exploratory Data Analysis: Preparation & Cleaning

Data Loading

The initial step involved importing the dataset into Excel for preliminary review and understanding its structure.

Missing Data Handling

Identified and addressed null values in the `Review Rating` column. Missing ratings were imputed using the median rating specific to each product category to maintain data integrity.

Column Standardization

Renamed all columns to **snake_case** for enhanced readability, consistency, and improved documentation across the project.

Feature Engineering

Created new, insightful features: `age_group` by categorizing customer ages and `purchase_frequency_days` derived from purchase dates.

Data Consistency Check

Verified the relationship between `discount_applied` and `promo_code_used`. As they were found to be redundant, `promo_code_used` was subsequently dropped to streamline the dataset.

SQL Analysis: Revenue by Gender & High Spenders

Revenue by Gender

A crucial aspect of our analysis involved comparing the total revenue generated by male versus female customers. This provides insights into gender-specific purchasing power and preferences.

Female	75191
Male	157890

The data indicates a significant difference in revenue contribution between genders, with male customers generating substantially higher revenue.

High-Spending Discount Users

We identified customers who utilized discounts but still made purchases above the average amount. This segment is valuable for understanding the effectiveness of promotional strategies on high-value customers.

2	64
3	73
4	90
7	85
9	97

A total of 839 customers fell into this category, highlighting a segment that responds to discounts while maintaining high spending.

SQL Analysis: Top Products & Shipping Insights

Top 5 Products by Rating

Understanding customer satisfaction is key. We identified products with the highest average review ratings, indicating strong customer approval and potential for marketing focus.

Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.80
Skirt	3.78

These top-rated products represent opportunities for increased promotion and potential cross-selling.

Shipping Type Comparison

Analyzing average purchase amounts across different shipping types helps evaluate the impact of shipping speed on customer spending habits.

Standard	58.46
Express	60.48

Customers opting for Express shipping tend to have slightly higher average purchase amounts, suggesting a correlation between perceived value and willingness to spend more.

SQL Analysis: Subscription & Discount Insights

Subscribers vs. Non-Subscribers

Comparing average spend and total revenue between subscribers and non-subscribers reveals the financial impact of subscription models.

Yes	1053	59.49	62645.00
No	2847	59.87	170436.00

While non-subscribers contribute more to total revenue due to their larger number, the average spend per subscriber is comparable, indicating a stable customer base.

Discount-Dependent Products

Identifying products with a high percentage of discounted purchases helps in optimizing pricing strategies and understanding customer sensitivity to promotions.

Hat	50.00
Sneakers	49.66
Coat	49.07
Sweater	48.17
Pants	47.37

These products may benefit from strategic discounting or bundling to maximize sales while managing margins.

SQL Analysis: Customer Segmentation & Repeat Buyers

Customer Segmentation

Classifying customers into segments (New, Returning, Loyal) based on their purchase history allows for tailored engagement strategies.

Loyal	3116
New	83
Returning	701

The majority of customers are loyal, indicating strong retention, but there's an opportunity to nurture new and returning customers.

Repeat Buyers & Subscriptions

We investigated whether customers with more than 5 purchases are more inclined to subscribe, linking purchase frequency to subscription potential.

No	2518
Yes	958

While a significant number of repeat buyers are not subscribed, nearly a thousand are, suggesting a strong correlation that can be leveraged.

SQL Analysis: Top Products by Category & Revenue by Age Group

Top 3 Products per Category

Identifying the most purchased items within each category helps in inventory management, merchandising, and targeted promotions.

Accessories	Jewelry	171
Accessories	Sunglasses	161
Accessories	Belt	161
Clothing	Blouse	171
Clothing	Pants	171
Clothing	Shirt	169
Footwear	Sandals	160
Footwear	Shoes	150
Footwear	Sneakers	145
Outerwear	Jacket	163
Outerwear	Coat	161

This detailed breakdown highlights popular items across various categories, enabling focused marketing and inventory strategies.

Revenue by Age Group

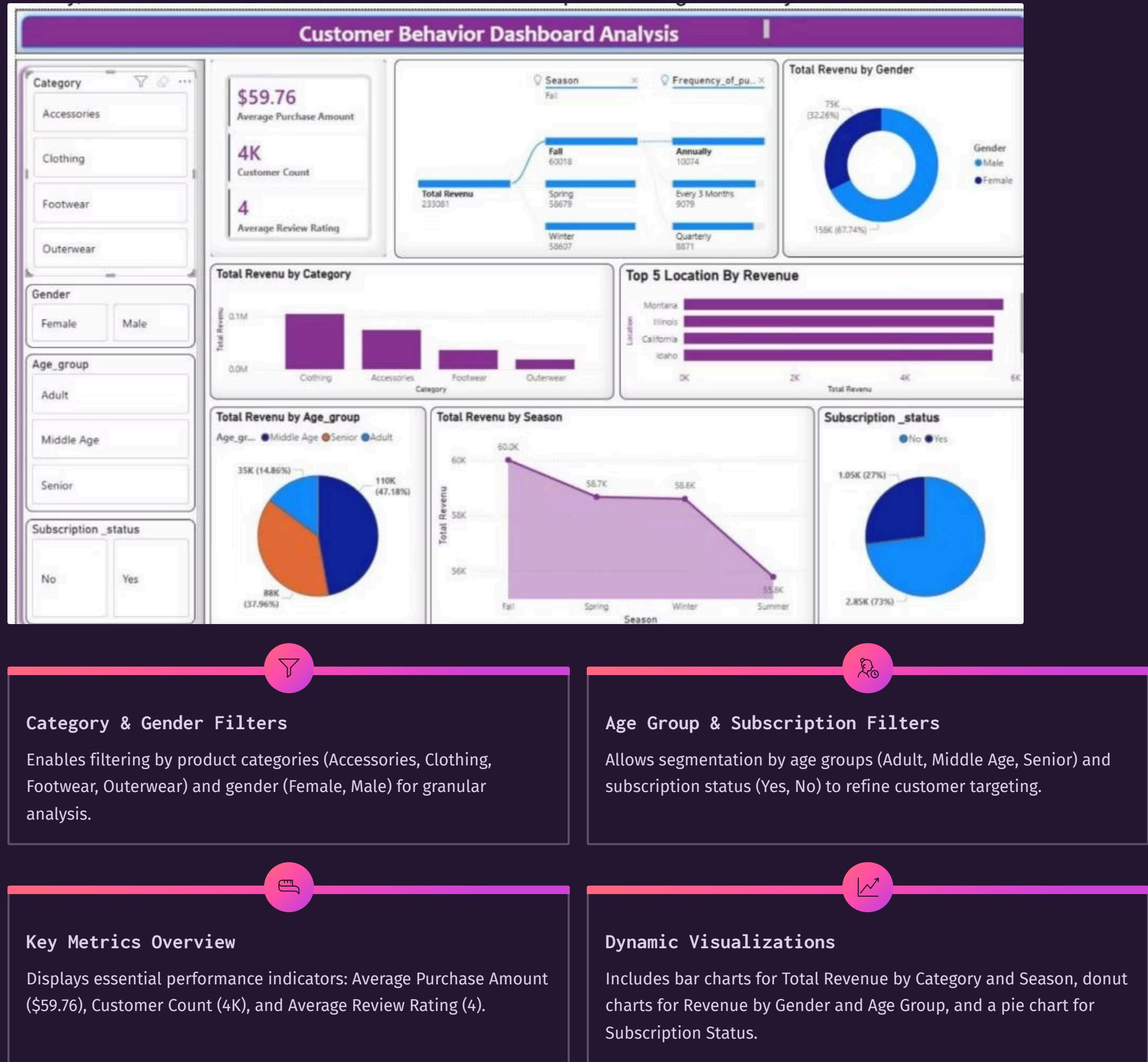
Understanding the revenue contribution from different age groups is vital for demographic-specific marketing and product targeting.

Young Adult	62143
Middle-aged	59197
Adult	55978
Senior	55763

Young Adults and Middle-aged groups are the highest revenue contributors, suggesting these segments should be prioritized in marketing efforts.

Interactive Dashboard in Power BI

To visualize and interact with these insights, a comprehensive dashboard was developed in Power BI. This dashboard allows for dynamic exploration of customer behavior data.



The dashboard provides an intuitive interface to explore customer behavior, identify trends, and support data-driven decision-making.

Strategic Business Recommendations



Boost Subscriptions

Implement targeted campaigns to promote exclusive benefits for subscribers, converting more repeat buyers into loyal members.



Customer Loyalty Programs

Develop and enhance loyalty programs to reward repeat buyers, encouraging them to transition into the "Loyal" customer segment.



Review Discount Policy

Strategically re-evaluate discount policies to balance sales boosts with margin control, ensuring profitability while attracting customers.



Product Positioning

Highlight top-rated and best-selling products in marketing campaigns to capitalize on proven customer preferences and drive sales.



Targeted Marketing

Focus marketing efforts on high-revenue age groups and express-shipping users, tailoring messages to resonate with these valuable segments.

These recommendations aim to optimize customer engagement, enhance revenue streams, and foster sustainable business growth.