Customer Analytics: Trends, Preferences, and Revenue

PROJECT OVERVIEW

This project analyzes customer shopping behaviour using transaction data to uncover trends in spending, product popularity, and customer engagement. The insights are presented through a Power BI dashboard to support business decision-making and marketing strategies.

DATASET SUMMARY

The dataset consists of 3,900 records across 18 columns, encompassing demographic, transactional, and behavioural data. Demographic details include age, gender, location, and subscription status, while purchase information captures items bought, their category, purchase amount, season, size, and colour. Behavioural aspects of the dataset record whether discounts or promo codes were used, previous purchase history, buying frequency, review ratings, and shipping preferences.

DATA COLLECTION & CLEANING IN PYTHON

The raw dataset (customer_shopping_behavior.csv) underwent preprocessing in Python before analysis:

- **Loading:** Dataset read into a pandas DataFrame.
- **Handling Null Values:** Missing entries in Review Rating were filled with the median rating per category.
- **Column Formatting:** Renamed columns to snake_case to maintain consistency and readability.
- **Feature Creation:**
    - age_group assigned by segmenting customer ages.
    - spending_segment derived from purchase_amount bins.
    - purchase_date generated where missing for temporal analysis.
    - Extracted month and weekday from purchase_date to facilitate trend exploration.
- **Data Consistency:** Checked for overlap between discount_applied and promo_code_used; redundant columns were removed.
- **Database Loading:** Cleaned data uploaded to PostgreSQL to allow SQL-based exploration.
- **Cleaned Dataset:** customer_shopping_behavior_cleaned.csv.

SQL ANALYSIS

Structured queries in PostgreSQL provided actionable insights:

- Top Products and Revenue by Gender: Identified items most frequently purchased and highest revenue contributors for male and female customers.

- Average Ratings Across Ages: Binned ages into groups to examine review patterns among different demographic segments.

	age_group text	avg_rating numeric
1	18-25	3.82
2	60+	3.76
3	36-45	3.75
4	26-35	3.74
5	46-60	3.72

- Discount Patterns by Category: Measured percentage of purchases with discounts applied for each product category.

	category text	discount_rate numeric
1	Outerwear	44.44
2	Accessori...	43.79
3	Footwear	43.24
4	Clothing	42.08

- Customer Loyalty by Age: Summed previous purchases to determine the most loyal age segments.

	age_group text	total_previous_purchases numeric
1	46-60	29246
2	60+	18838
3	26-35	18372
4	36-45	18349
5	18-25	14066

- Spending Patterns by Gender and Category: Compared average purchase amounts to find where men and women spend differently.

	gender text	category text	avg_spend numeric
1	Female	Accessori...	60.76
2	Female	Clothing	60.50
3	Female	Footwear	59.47
4	Female	Outerwear	58.43
5	Male	Footwear	60.65
6	Male	Clothing	59.80
7	Male	Accessori...	59.41
8	Male	Outerwear	56.61

- Top 3 Products per Age Group: Used window functions to highlight leading products for each age range.

	age_group text	item_purchased text	total_orders bigint
1	18-25	Sweater	31
2	18-25	Dress	29
3	18-25	Coat	29
4	26-35	Shirt	44
5	26-35	Backpack	39
6	26-35	Jewelry	38
7	36-45	Scarf	45
8	36-45	Pants	44
9	36-45	Jacket	39
10	46-60	Shoes	60
11	46-60	Sandals	55
12	46-60	Shorts	54
13	60+	Jewelry	45
14	60+	Hoodie	37
15	60+	Dress	36

- Monthly Revenue Leaders: Calculated which products generate the highest revenue per month to uncover seasonal trends.

	category text	total_revenue numeric	avg_purchase numeric
1	Clothing	104264	60.03
2	Accessori...	74200	59.84
3	Footwear	36093	60.26
4	Outerwear	18524	57.17

CTEs and views, such as customer_with_age_group, streamlined repeated queries and improved query readability.

DASHBOARD DESIGN IN POWER BI

The dashboard shown below was designed to provide a comprehensive view of customer behaviour and sales performance using interactive visualizations:

- **KPI Cards:** Three cards highlight the total number of customers, average review rating, and total revenue at a glance, giving a high-level snapshot of performance.

- **Clustered Bar Chart:** Displays revenue by product category, showing how different categories contribute to overall sales.

- **Pie Chart:** Illustrates the distribution of customers by subscription status (Yes vs No) as percentages, helping to quickly assess subscription engagement.

- **Column Chart:** Breaks down customers by age groups, showing the number of customers in each segment for demographic insights.

- **100% Stacked Column Chart:** Shows discount usage over months, comparing the number of customers who used a discount versus those who did not, allowing easy comparison of seasonal discount trends.

- **Line Chart:** Displays monthly trends in purchase amounts to identify peaks and dips throughout the year.

- **Slicer Buttons:** Interactive buttons allow filtering by gender (Female vs Male), shipping type (2-Day, Free, Express, Next Day Air, Standard, Store Pickup), and product category (Accessories, Clothing, Footwear, Outerwear), giving users the ability to explore data by specific segments.

This combination of cards, charts, and interactive buttons provides both high-level KPIs and detailed trend analysis, making the dashboard intuitive, flexible, and suitable for business decision-making.

INSIGHTS & ANALYSIS

The Power BI dashboard provides a detailed view of customer behaviour, spending patterns, and product performance. The dataset shows 3.9K customers generating total revenue of $233K, with an average review rating of 3.75, indicating generally positive customer satisfaction. Revenue is highest in the clothing and accessories categories, suggesting strong demand; these categories could be expanded further through new product lines, targeted promotions, or bundling strategies. Footwear and outerwear generate lower revenue, which could indicate untapped potential; strategic discount campaigns or seasonal promotions could be applied to these categories to stimulate sales.

Discount usage trends vary by month, with certain periods showing higher uptake, signalling when promotions are most effective. This suggests that timing discounts strategically could boost revenue in lower-performing months. The customer age breakdown highlights that the 46–60 age group is the largest, while younger age groups like 18–25 are smaller segments; marketing campaigns could be tailored to engage younger audiences to expand the customer base. Subscription status shows that a small portion of customers are subscribed, presenting opportunities to increase engagement through loyalty programs or subscription incentives.

Interactive filters for gender, shipping type, and product category reveal distinct purchasing patterns: for instance, males and females prefer different products, which can guide targeted marketing and inventory allocation. The monthly purchase amount and review ratings indicate slight seasonal trends, suggesting that promotional efforts and product launches could be aligned with peak periods for maximum impact. Overall, the dashboard provides actionable insights to optimize product strategy, pricing, promotions, and customer engagement, helping to drive revenue growth across underperforming categories while capitalizing on high-performing segments.