Project

Customer Shopping Trend Analysis

Problem Statement

The retail business wants to better understand **customer purchasing behavior** to improve sales, customer retention, and marketing effectiveness.
Currently, decisions related to promotions, subscriptions, and product offerings are made without a clear, data-driven understanding of:

- Who the customers are
- What they buy
- How often they purchase
- Which factors influence purchase value and loyalty

Using the provided customer transaction dataset, this project aims to analyze **demographics, purchasing patterns, payment preferences, subscription impact, and promotional effectiveness** to generate actionable business insights that can support strategic decision-making.

Project Objectives:

- Analyze customer demographics and buying behavior
- Identify high-value customers and frequently purchased products
- Evaluate the impact of discounts, promo codes, and subscriptions
- Understand payment and shipping preferences
- Provide insights through interactive dashboards for stakeholders

Dataset Summary:

- The dataset contains customer-level retail transaction data, capturing demographics, purchase behavior, and payment details.
- It includes attributes such as age, gender, location, and subscription status to analyze customer segments.
- Product-related fields cover item purchased, category, size, color, and season, enabling product and seasonal trend analysis.

- Transaction details include purchase amount, discounts, promo code usage, payment method, and shipping type.
- Customer engagement is measured through review ratings, purchase frequency, and previous purchase history.
- The dataset supports analysis of revenue trends, customer loyalty, promotional effectiveness, and regional performance.
- Rows : 3,900, Columns: 18
- Key Features: - Customer demographics (Age, Gender, Location, Subscription Status) - Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color) - Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of

Data Cleaning & Preparation

Data Loading: Loaded data using pandas and visualized initial rows

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Payment Method	Sh
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Credit Card	
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Bank Transfer	
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Cash	
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	PayPal	
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Cash	

Data Handling: Checked for missing data values (if any) and wrong data types using df.info()

Column Standardization: Renamed columns to snake case for better understanding.

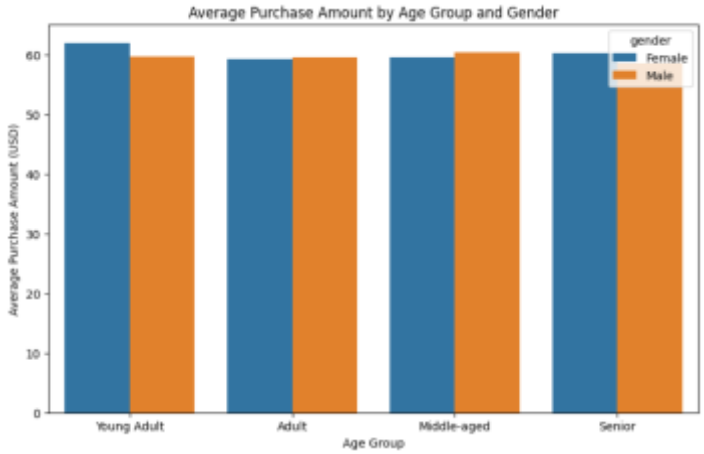
Feature Engineering:

- Created age_group column by categorizing customer ages into quartiles.
- Created purchase_frequency_days column from purchase data.

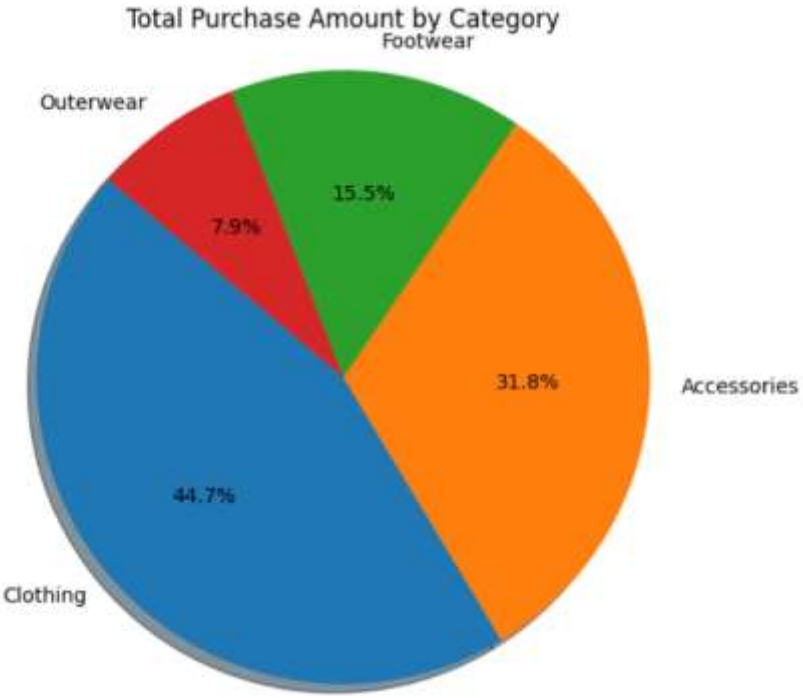
Exploratory Data Analysis (EDA)

Used python to perfom EDA and answer some research questions.

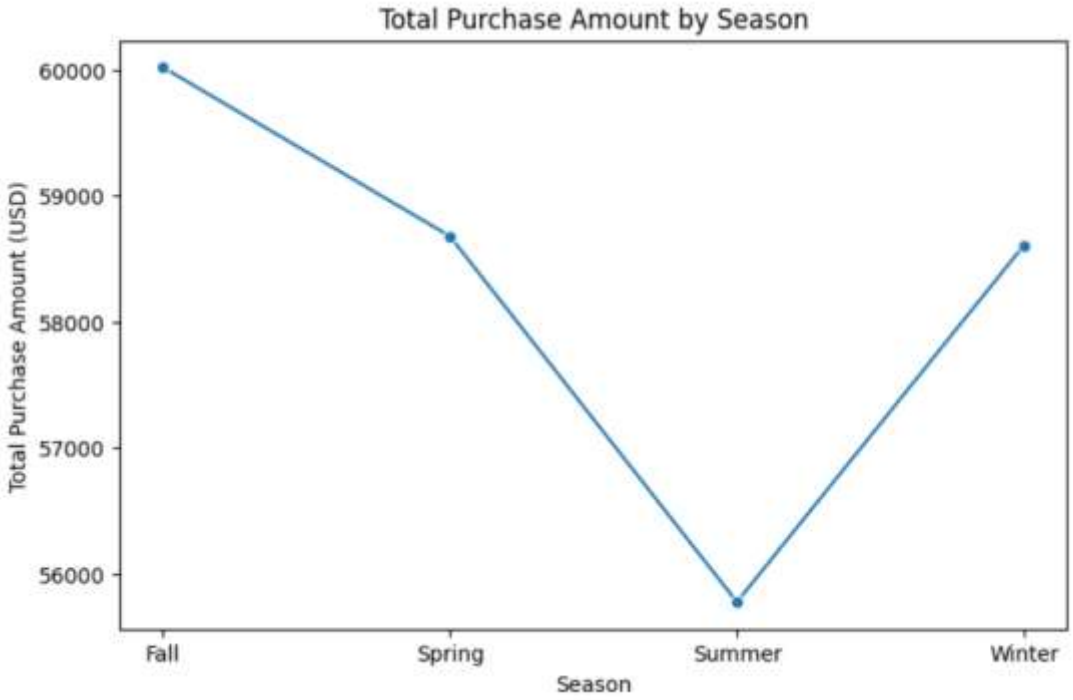
How do customer demographics influence purchasing behavior?



Which product categories and items drive the most revenue and engagement?



Are there seasonal patterns in customer purchases and preferences?



Business Questions to Answer Using SQL

1. **What is the total and average purchase amount by product category?**

	category	total_revenue	avg_purchase_amount
▶	Clothing	104264	60.03
	Accessories	74200	59.84
	Footwear	36093	60.26
	Outerwear	18524	57.17

2. **Which locations generate the highest total revenue?**

location	total_revenue
Montana	5784
Illinois	5617
California	5605
Idaho	5587
Nevada	5514
Alabama	5261
New York	5257
North Dakota	5220
West Virginia	5174
Nebraska	5172
New Mexico	5014
Minnesota	4977
Pennsylvania	4926
Mississippi	4883
Alaska	4867

3. **How does purchase behavior differ between subscribed and non-subscribed customers?**

subscription_status	total_purchases	avg_purchase_amount	avg_previous_purchases
Yes	1053	59.49	26.08
No	2847	59.87	25.08

4. **Which items are most frequently purchased?**

item_purchased	purchase_count
Blouse	171
Pants	171
Jewelry	171
Shirt	169
Dress	166
Sweater	164
Jacket	163
Coat	161
Sunglasses	161
Belt	161

5. **Does applying a discount or promo code increase the average purchase amount?**

	discount_applied	avg_purchase_amount
▶	Yes	59.28
	No	60.13

6. **Which season records the highest number of purchases and revenue?**

	season	total_purchases	total_revenue
▶	Fall	975	60018
	Spring	999	58679
	Winter	971	58607
	Summer	955	55777

7. **Segment customers into New, Returning, and Loyal based on their total number of previous purchases, and show the count of each segment.**

	customer_segment	Number of Customers
▶	Loyal	3116
	Returning	701
	New	83

8. **What are the top 3 most purchased products within 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

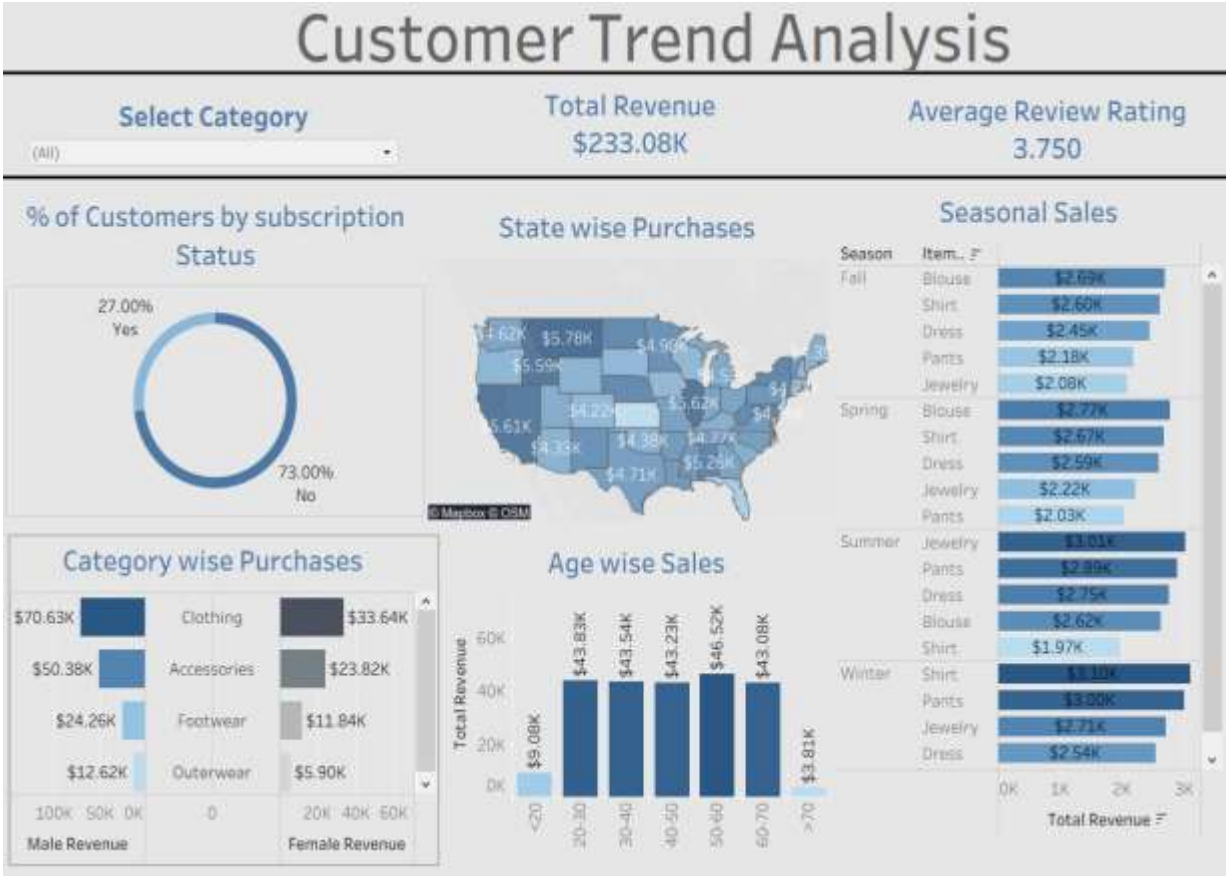
9. **Are customers who are repeat buyers (more than 5 previous purchases) also likely to subscribe?**

	subscription_status	repeat_buyers
▶	Yes	958
	No	2518

10. **Is there a relationship between review ratings and repeat purchases (Previous Purchases)?**

	review_rating	avg_previous_purchases	total_customers
▶	2.5	24.80	66
	2.6	25.38	159
	2.7	25.23	154
	2.8	25.21	136
	2.9	24.45	170
	3	25.89	162
	3.1	25.51	157
	3.2	24.36	152
	3.3	26.63	152
	3.4	23.86	182
	3.5	27.36	156
	3.6	26.58	149
	3.7	23.93	156
	3.8	23.61	142
	3.9	26.44	163
	4	26.12	181
	4.1	25.78	148

Tableau Dashboard:



Business Recommendations:

Subscription Growth & Customer Retention:

Expand the subscription program to convert the 73% non-subscribed customers by offering exclusive benefits like free shipping and loyalty rewards.

Targeted Marketing by Age Segment:

Focus marketing efforts on customers aged 20–60, who contribute the highest revenue, and tailor messaging by life stage.

Product Category Optimization

Prioritize high-performing categories such as Clothing and Accessories, while reassessing pricing and assortment for lower-performing categories.

Seasonal Inventory & Demand Planning

Optimize seasonal inventory planning, especially for high-revenue Summer and Winter items, to reduce stockouts and excess inventory.

Regional Performance Strategy

Adopt a region-specific sales strategy by increasing marketing spend in high-revenue states and investigating underperforming regions.

Customer Experience & Review Improvement

Improve customer review ratings by addressing product quality and delivery issues, and incentivizing customer feedback.

Personalized Gender-Based Marketing

Leverage gender-based purchasing patterns to personalize recommendations and run targeted promotional campaigns.