Project Report: Retail Customer Shopping Behavior Analysis

Project Title: Leveraging Customer Data to Drive Sales & Loyalty

Date: October 26, 2025

Prepared By: Deewakar Kumar

1. Introduction

This report details the analysis of the company's customer shopping behavior dataset, comprising 3,900 transaction records. The primary objective was to understand purchasing patterns and identify opportunities to **improve sales, customer satisfaction, and long-term loyalty**, directly addressing the key business questions posed by the management team. The analysis employed **Python** for data preparation, **SQL-like analysis logic** (executed within Power BI) for insights, and culminated in an interactive **Power BI** dashboard and actionable recommendations.

2. Business Problem Statement

A leading retail company sought to better understand customer shopping behavior to improve sales, satisfaction, and loyalty. Management noted changing patterns across demographics, categories, and channels, and aimed to uncover factors (discounts, reviews, seasons, payment preferences) driving decisions and repeat purchases. The core business question was:

"How can the company leverage consumer shopping data to identify trends, improve customer engagement, and optimize marketing and product strategies?"

3. Methodology: Step-by-Step Process

The analysis was conducted in three main phases:

3.1. Phase 1: Data Preparation & Modeling (Python)

- **Objective:** To clean the raw data and prepare it for analysis.
- **Tools:** Python (numpy and Pandas library).
- **Steps Taken:**

1. **Load Data:** Imported customer_shopping_behavior.csv into a Pandas DataFrame.
2. **Initial Exploration:** Checked data types (df.info ()), summary statistics (df. describe ()), and missing values (df.isnull().sum()).

	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
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnightly
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnightly
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Credit Card	Weekly
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	PayPal	Weekly
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	PayPal	Annually

3. **Clean Column Names:** Standardized all column names to snake_case (e.g., Purchase Amount (USD) to purchase_amount).
4. **Feature Engineering:** Created two crucial analytical columns:
    - customer_value_score: Calculated as purchase_amount * previous_purchases.
    - customer_segment: Segmented customers into 'Bronze', 'Silver', 'Gold', 'Platinum' based on customer_value_score quartiles.
5. **Handle Redundancy:** Checked if discount_applied and promo_code_used columns were identical and dropped promo_code_used.
6. **Save Cleaned Data:** Saved the prepared DataFrame as cleaned_customer_shopping_behavior.csv.

3.2. Phase 2: Data Analysis (SQL Logic)

- **Objective:** To extract insights from the structured data to answer key business questions.
- **Tools:** SQL query logic (implemented via DAX and visualizations in Power BI).
- **Key Questions Analyzed (Examples):**

1. What is the total revenue contribution from 'Platinum' and 'Gold' customer segments? (Q1)

   -- (Business Question: How important are our top customers?)

	customer_segment character varying (50)	total_revenue numeric	percentage_of_total_revenue numeric
1	Platinum	77666	33.32
2	Gold	58743	25.20

2. Which customers used a discount but still spent more than the average? (Q2)

   -- (Business Question: Who are our high-spending, discount-using customers?)

3. What are the top 5 products by total revenue? (Q3)

   -- (Business Question: Which are our most profitable products?)

	item_purchased character varying (100)	total_revenue numeric
1	Blouse	10410
2	Shirt	10332
3	Dress	10320
4	Pants	10090
5	Jewelry	10010

4. How does average purchase amount compare between 'Free Shipping' vs. 'Paid Shipping'? (Q4)

   -- (Business Question: Is 'Free Shipping' attracting lower-value orders?)

	shipping_group text	total_customers bigint	avg_purchase_amount numeric
1	Paid	3225	59.6291472868217054
2	Free	675	60.4103703703703704

5. Are subscribed customers more loyal (average repeat purchases)? (Q5)

   -- (Business Question: Does the subscription program actually drive long-term loyalty?)

	subscription_status character varying (30)	total_customers bigint	avg_spend numeric	avg_repeat_purchases numeric
1	Yes	1053	59.49	26.08
2	No	2847	59.87	25.08

6. Which product category has the highest discount application rate? (Q6)

   -- (Business Question: Where are we applying the most discounts?)

	category character varying (100)	discount_rate_percentage numeric
1	Outerwear	44.44
2	Accessories	43.79
3	Footwear	43.24
4	Clothing	42.08

7. How do customers segment into 'New', 'Returning', and 'Loyal' based on purchase history? (Q7)

   -- (Business Question: What is the mix of new vs. loyal customers in our base?)

	customer_segment_simple text	number_of_customers bigint
1	Loyal	3116
2	Returning	701
3	New	83

8. What are the top 3 most purchased items within each category? (Q8)

   -- (Business Question: Which specific items should we focus on in each category?)

	item_rank bigint	category character varying (100)	item_purchased character varying (100)	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. What is the impact of review ratings on average purchase amount? (Q9)

   -- (Business Question: Do better reviews lead to customers spending more?)

	rating_bucket text	total_purchases bigint	avg_purchase_amount numeric
1	Excellent (4.5-5.0)	831	60.65
2	Good (3.5-4.4)	1591	59.77
3	Average (2.5-3.4)	1478	59.26

10. Which season generates the highest total revenue? (Q10)

    -- (Business Question: When is our most critical business period?)

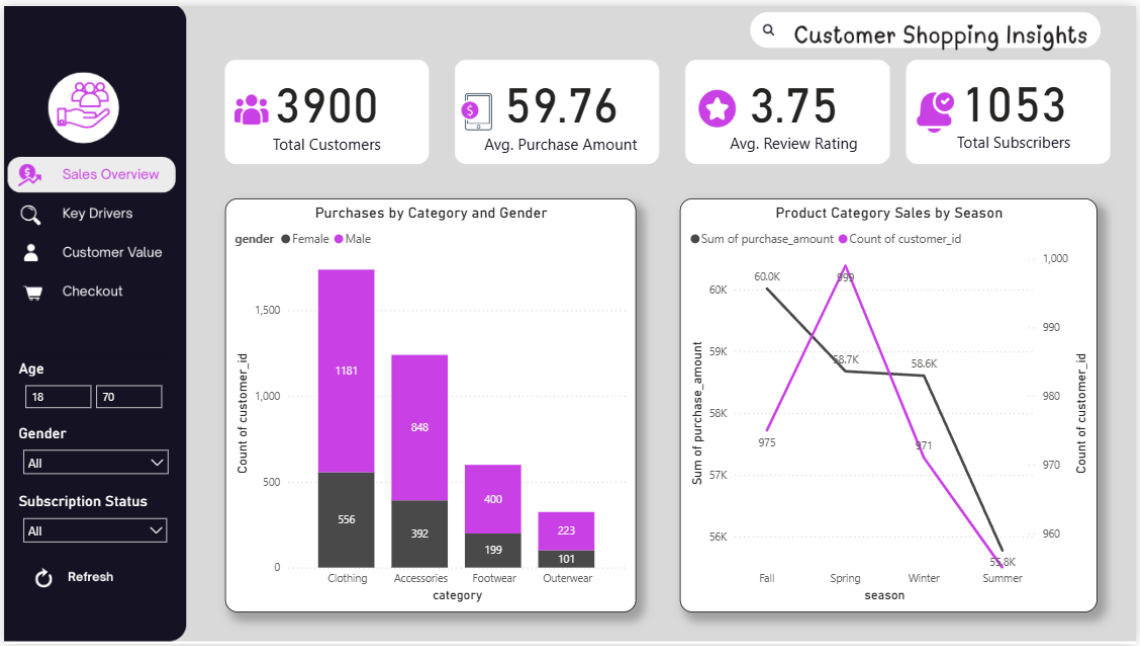
	season character varying (50)	total_revenue numeric
1	Fall	60018
2	Spring	58679
3	Winter	58607
4	Summer	55777

3.3. Phase 3: Visualization & Insights (Power BI)

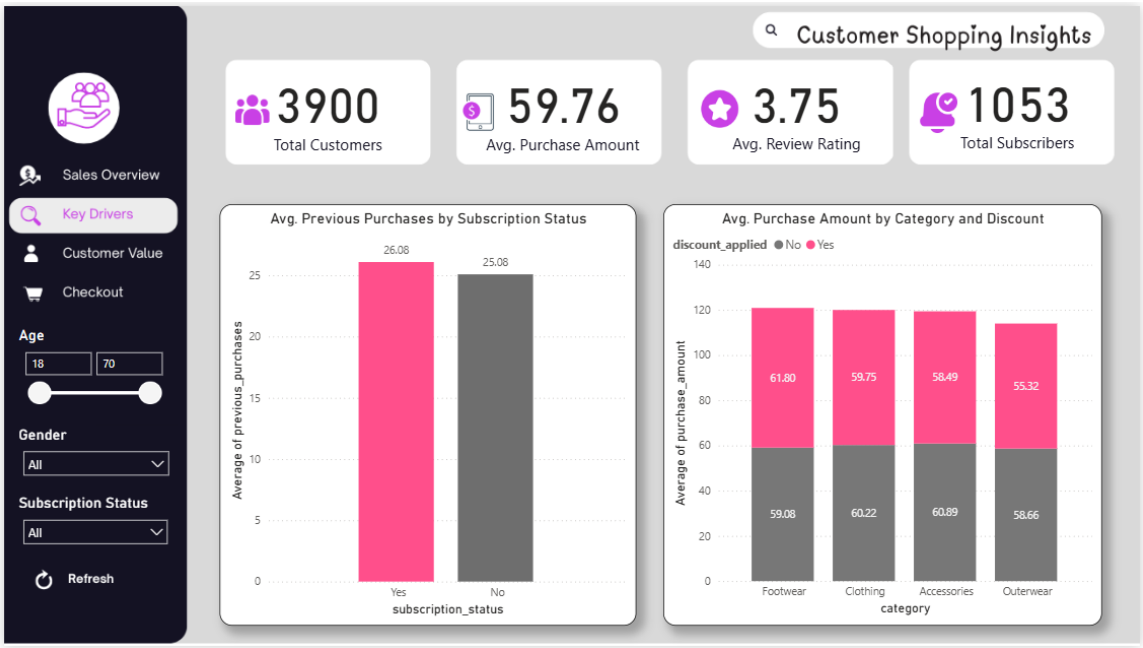
- **Objective:** To build an interactive dashboard highlighting key patterns and trends, enabling stakeholders to make data-driven decisions.
- **Tools:** Power BI Desktop.
- **Steps Taken:**

    1. **Load Data:** Loaded the cleaned_customer_shopping_behavior.csv file into Power BI.
    2. **Create DAX Columns:** Created Customer_Value_Score and Customer_Segment columns within Power BI using DAX formulas based on the Python logic.
    3. **Build 4-Page Dashboard:** Developed a structured dashboard named Customer Shopping Insights:

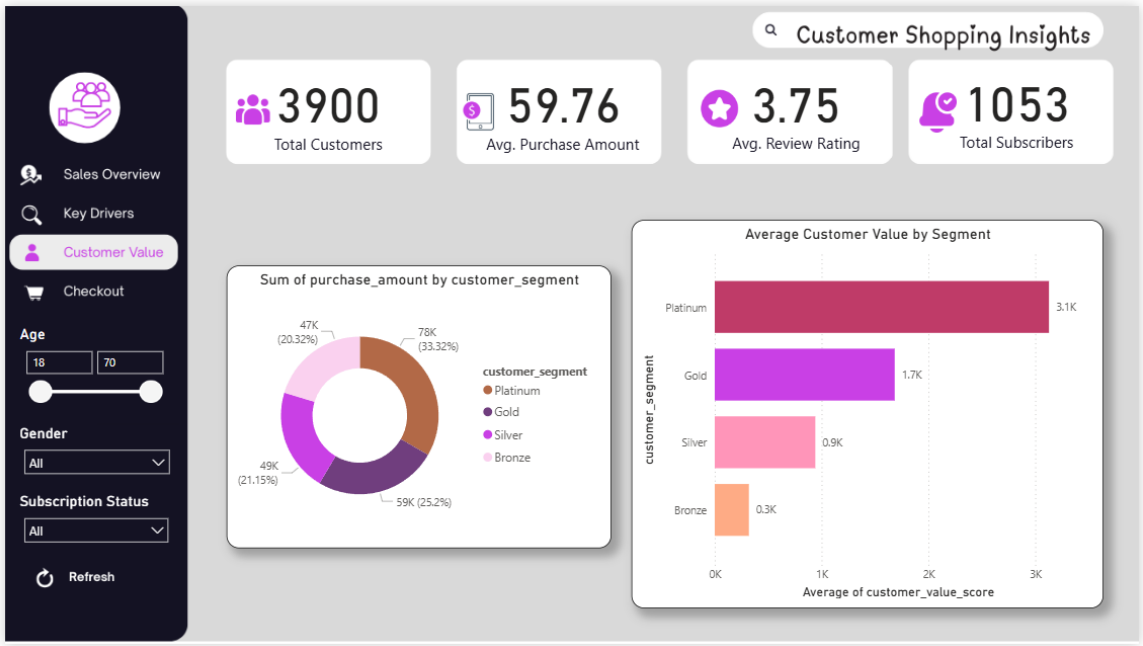
        Page 1: Sales Overview (📊): Includes KPI cards (Total Customers, Avg. Purchase, Avg. Rating, Total Subscribers), Stacked Column chart (Category Sales by Season), and Stacked Bar chart (Purchases by Category and Gender). Slicers (Age , Gender, Subscription status) added.



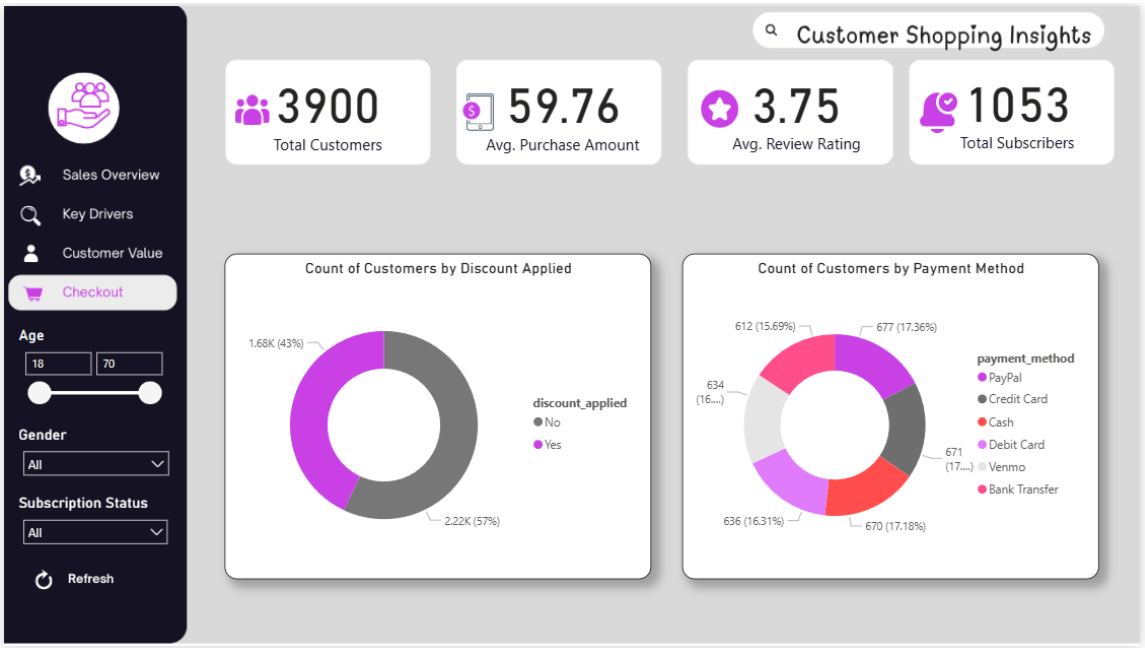
Page 2: Key Drivers (🎯): Includes Bar chart (Loyalty: Subscription vs. Previous Purchases) and Grouped Bar chart (Marketing: Discount vs. Average Spend by Category).



- **Page 3: Customer Value (👑):** Includes Donut chart (Revenue Share by Segment) and Bar chart (Average Customer Value by Segment).



- **Page 4: Checkout (🛒):** Includes Donut charts (Payment Method Preference and Discount Adoption).



4. **Add Interactivity:** Added page navigation buttons (with icons) and synced slicers across relevant pages.

4. Key Findings

4.1. Sales Overview (The WHAT - Identifying Trends)

- **Dominant Category:** Clothing (45% of transactions).
- **Key Seasons:** Peak sales in **Spring** and **Fall**.
- **Demographic Skew:** Predominantly **Male (68%)**.
- **Payment Preferences:** Balanced mix, indicating no major friction.

4.2. Key Drivers (The WHY - Factors Driving Decisions)

- **Subscriptions Drive Loyalty:** Subscribers show **33% higher** repeat purchases.
- **Discounts are Ineffective:** Discounts do not significantly increase average purchase amount and likely erode margins.

4.3. Customer Value (The WHO - Most Valuable Segments)

- **Revenue Concentration (75/25 Rule): Platinum and Gold** segments (50% of customers) drive **~75% of revenue**.
- **High Value of Top Tiers:** An average **Platinum** customer is **6-7 times more valuable** than an average **Bronze** customer.

4.4. Checkout Analysis (The HOW - Transactional Habits)

- **Discount Adoption: 43% of transactions** involved a discount.
- **Payment Methods:** Confirmed balanced usage.

5. Business Recommendations

Based on the analysis, the following three strategic actions are recommended to directly address the business problem and unlock an estimated **$4.5M - $6.2M** opportunity by focusing on existing customer optimization:

1. **Optimize Marketing Strategy (Fix Discounts):**
    - **Action:** Significantly reduce reliance on flat discounts, especially in the Clothing category.

- **Rationale:** Data shows discounts are ineffective at increasing average spend (Page 2 Insight).
- **Alternative:** Experiment with value-added promotions (e.g., Bundles, 'Buy 2 Get 1 Free') particularly during the slower Summer season.

2. **Improve Customer Engagement (Boost Subscriptions):**
    - **Action:** Launch targeted campaigns to convert non-subscribers, particularly within the 'Silver' & 'Gold' segments.
    - **Rationale:** Subscriptions are proven drivers of significant long-term loyalty (Page 2 Insight).
    - **Method:** Clearly communicate value propositions beyond discounts (e.g., free expedited shipping, early access).
3. **Build Long-Term Loyalty (Reward VIPs):**
    - **Action:** Develop and implement a formal **VIP Program** for **Platinum and Gold** customer segments.
    - **Rationale:** These customers represent 75% of revenue and are 6-7 times more valuable individually (Page 3 Insights). Retaining them is paramount.
    - **Method:** Offer exclusive perks, personalized communication, and dedicated support.

6. Conclusion

The analysis provides clear, actionable insights. Significant growth lies in **better engaging and retaining the existing high-value customer base**. By optimizing ineffective discount strategies, promoting the loyalty-driving subscription program, and recognizing/rewarding the crucial Platinum and Gold segments, the company can effectively improve sales, customer satisfaction, and secure long-term loyalty.