Customer Shopping Behavior Analysis Report

Business Problem & Objective

Retail organizations generate large volumes of customer transaction data, but without structured analysis, this data often remains underutilized. The business faces challenges in understanding customer purchasing behavior, identifying high-value customer segments, evaluating product performance, and assessing the effectiveness of discounts and subscription programs.

The objective of this analysis is to transform raw customer shopping data into actionable insights that support revenue growth, customer engagement, and long-term loyalty through data-driven decision-making.

Executive Summary

This report analyzes 3,900 customer transactions to uncover insights related to spending behavior, customer segmentation, product performance, and subscription trends. Python was used for data preparation, PostgreSQL for querying and analysis, and Power BI for visualization.

Dataset Overview

- Total Records: 3,900
- Total Columns: 18

Key Features:

- Demographics: Age, Gender, Location, Subscription Status
- Purchase Details: Item Purchased, Category, Purchase Amount, Season, Size, Color
- Behavioral Attributes: Discount Applied, Previous Purchases, Purchase Frequency, Review Rating, Shipping Type

Data Quality:

- 37 missing values in the Review Rating column

Data Preparation & Exploratory Analysis (Python)

- Imported and explored the dataset using pandas (df.info(), df.describe()).
- Imputed missing review ratings using median values at the product-category level.
- Standardized column names to snake_case.
- Engineered features such as age_group and purchase_frequency_days.
- Validated data consistency and removed redundant attributes.
- Loaded cleaned data into PostgreSQL for analysis.

SQL-Based Business Analysis

SQL queries were used to analyze revenue distribution, customer segmentation, discount usage, subscription behavior, repeat purchases, and product performance across categories.

Power BI Dashboard – Visual Insights

An interactive Power BI dashboard was developed to visualize key performance indicators, customer demographics, product category performance, and subscription trends.

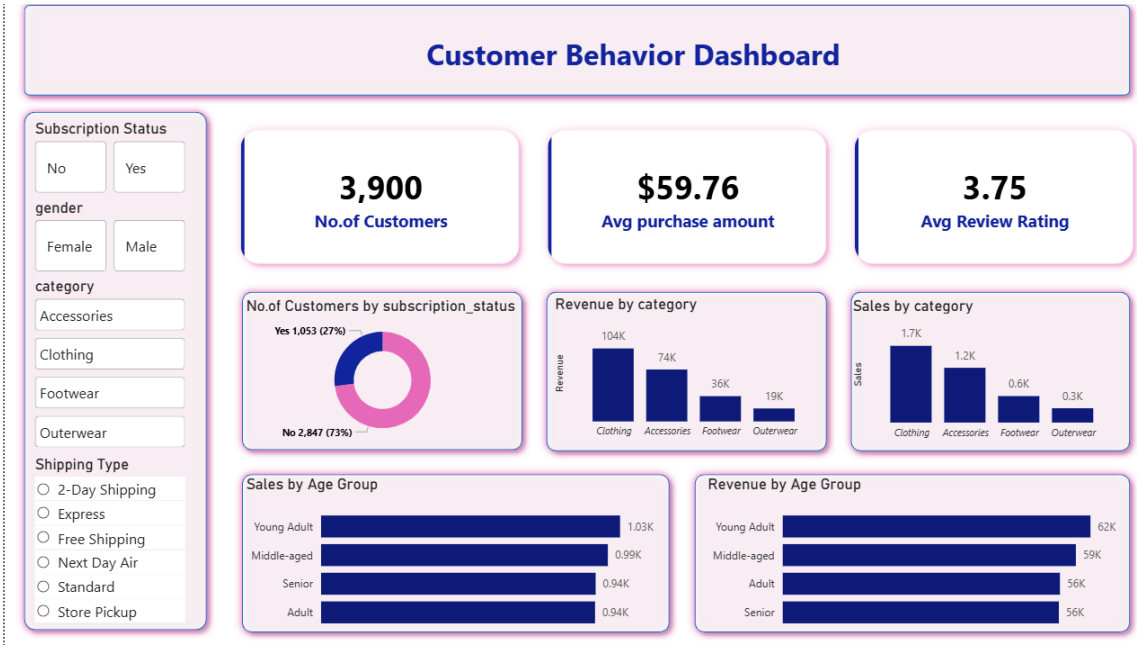

Figure: Customer Behavior Dashboard Overview

- Displays total customers, average purchase amount, and average review rating.
- Clothing is the highest revenue-generating category.
- Young Adult and Middle-aged customers contribute the highest revenue.
- Majority of customers are non-subscribers, highlighting subscription growth opportunities.

Business Recommendations

- Promote subscription plans through exclusive benefits and personalized offers.
- Implement loyalty programs to convert returning customers into loyal customers.
- Optimize discount strategies to balance revenue growth and profitability.
- Focus marketing efforts on high-performing categories such as Clothing and Accessories.
- Target high-revenue age groups with tailored campaigns and express shipping incentives.

Appendix A: SQL Queries

```
-- Q1. Total revenue by gender
SELECT
   gender,
   SUM(purchase_amount) AS revenue
FROM customer
GROUP BY gender;
```

	gender text	revenue numeric
1	Female	75191
2	Male	157890

```
-- Q2. Customers who used discounts but spent more than average
SELECT
   customer_id,
   purchase_amount
FROM customer
WHERE discount_applied = 'Yes'
  AND purchase_amount > (
     SELECT AVG(purchase_amount)
     FROM customer
  );
```

	customer_id bigint	purchase_amount bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	88
12	29	94
13	32	79
14	33	67
15	35	91

Total rows: 839 Query complete 00:00:

-- Q3. Top 5 products by highest average review rating
SELECT
  item_purchased,
  ROUND(AVG(review_rating::NUMERIC), 2) AS average_rating
FROM customer
GROUP BY item_purchased
ORDER BY average_rating DESC
LIMIT 5;

	item_purchased text	average_product_rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

Total rows: 5 Query complete 00:00:00.569

	item_purchased text	average_product_rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

Total rows: 5 Query complete 00:00:00.182

```
-- Q4. Average purchase amount by shipping type
SELECT
   shipping_type,
   ROUND(AVG(purchase_amount), 2) AS avg_purchase_amount
FROM customer
WHERE shipping_type IN ('Standard', 'Express')
GROUP BY shipping_type;
```

	shipping_type text	avg_purchase_amount numeric
1	Standard	58.46
2	Express	60.48

```
-- Q5. Subscriber vs non-subscriber spending behavior
SELECT
   subscription_status,
   COUNT(customer_id) AS total_customers,
   ROUND(AVG(purchase_amount), 2) AS avg_spend,
   ROUND(SUM(purchase_amount), 2) AS total_revenue
FROM customer
GROUP BY subscription_status;
```

	subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
1	No	2847	59.87	170436.00
2	Yes	1053	59.49	62645.00

```
-- Q6. Products with highest percentage of discounted purchases
SELECT
   item_purchased,
```

```
  ROUND(
    100.0 * SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END) / COUNT(*),
    2
  ) AS discount_rate
FROM customer
GROUP BY item_purchased
ORDER BY discount_rate DESC
LIMIT 5;
```

	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

Total rows: 5 | Query complete 00:00:00.1

```
-- Q7. Customer segmentation based on previous purchases
WITH customer_type AS (
  SELECT
    customer_id,
    CASE
      WHEN previous_purchases = 1 THEN 'New'
      WHEN previous_purchases BETWEEN 2 AND 10 THEN 'Returning'
      ELSE 'Loyal'
    END AS customer_segment
  FROM customer
)
SELECT
  customer_segment,
  COUNT(*) AS number_of_customers
FROM customer_type
GROUP BY customer_segment;
```

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

```
-- Q8. Top 3 most purchased products within each category
WITH item_counts AS (
   SELECT
      category,
      item_purchased,
      COUNT(customer_id) AS total_orders,
      ROW_NUMBER() OVER (
         PARTITION BY category
         ORDER BY COUNT(customer_id) DESC
      ) AS item_rank
   FROM customer
   GROUP BY category, item_purchased
)
SELECT
   category,
   item_purchased,
   total_orders
FROM item_counts
WHERE item_rank <= 3;
```

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessori...	Jewelry	171
2	2	Accessori...	Sunglasses	161
3	3	Accessori...	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

Total rows: 11 Query complete 00:00:00.130

```
-- Q9. Subscription status of repeat buyers (>5 purchases)
SELECT
   subscription_status,
   COUNT(customer_id) AS repeat_buyers
FROM customer
WHERE previous_purchases > 5
GROUP BY subscription_status;
```

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

```
-- Q10. Revenue contribution by age group
SELECT
   age_group,
   SUM(purchase_amount) AS total_revenue
FROM customer
GROUP BY age_group
ORDER BY total_revenue DESC;
```

	age_group text	total_revenue numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763