

Customer Shopping Behavior Analysis - Project Report

Table of Contents

Customer Shopping Behavior Analysis - Project Report.....	1
1. Executive Summary.....	2
2. Dataset Overview	2
3. Data Cleaning & Preparation	2
4. Exploratory Data Analysis (Python)	2
5. SQL Analysis & Insights	2
6. Power BI Dashboard.....	9
7. Seasonality Insights.....	11
8. Strategic Recommendations.....	11
8.1 Strengthen Marketing Toward High-Value Segments	11
8.2 Revamp the Subscription Program.....	11
8.3 Optimize Discount Strategy	11
8.4 Capitalize on Repeat-Driving Products.....	12
8.5 Improve Customer Experience Without Increasing Logistics Cost	12
8.6 Leverage Seasonality While Maintaining Steady Inventory	12

1. Executive Summary

This project analyzes customer shopping behavior using Python, SQL, and Power BI. The goal is to uncover trends related to customer demographics, purchase preferences, discount behavior, loyalty, and seasonal patterns.

2. Dataset Overview

The dataset comprises 3,900 customer transactions, featuring 18 columns with fields such as customer ID, age, gender, category, purchase amount, review ratings, subscription status, shipping type, and previous purchases.

3. Data Cleaning & Preparation

- Removed unnecessary columns
- Ensured consistent lowercase column names
- Created age groups using quantiles
- Converted categorical frequency-of-purchase text into numeric days
- The verified dataset contains no missing values
- Created PostgreSQL Database Using SQL.

4. Exploratory Data Analysis (Python)

Basic profiling was performed using pandas (df.describe). Key statistics include:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Review Rating
count	3900	3900	3900	3900	3900	3900	3900
unique	-	-	2	25	4	-	-
top	-	-	Male	Blouse	Clothing	-	-
freq	-	-	2652	171	1737	-	-
mean	1950.50	44.07	-	-	-	59.76	3.75
std	1125.98	15.21	-	-	-	23.69	0.72
min	1	18	-	-	-	20	2.5
25%	975.75	31	-	-	-	39	3.1
50%	1950.50	44	-	-	-	60	3.7
75%	2925.25	57	-	-	-	81	4.4
max	3900	70	-	-	-	100	5.0

5. SQL Analysis & Insights

Multiple SQL queries were executed to analyze revenue drivers, customer loyalty, discount sensitivity, and product performance. Key insights include:

Q1: What is the total revenue generated by male vs female customers?

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
1 -- Customer Profile & Revenue Drivers--
2
3 -- Q1: What is the total revenue generated by male vs female customers?
4 SELECT gender, SUM(purchase_amount) AS revenue
5 FROM customer_behavior
6 GROUP BY gender;
7
```

Data Output Messages Notifications

	gender	revenue
	text	numeric
1	Female	75191
2	Male	157890

Showing rows: 1 to 2 Page No: 1 of 1

Total rows: 2 Query complete 00:00:00.348 CRLF Ln 8, Col 39

Type here to search

-- Q2: What is the revenue contribution of each age group?

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
8 -- Q2: What is the revenue contribution of each age group?
9 SELECT
10 age_group,
11 SUM(purchase_amount) AS total_revenue
12 FROM customer_behavior
13 GROUP BY age_group
14 ORDER BY total_revenue DESC;
15
```

Data Output Messages Notifications

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

Showing rows: 1 to 4 Page No: 1 of 1

Total rows: 4 Query complete 00:00:00.235 CRLF Ln 9

Type here to search

Q3: Do subscribed customers spend more? Compare avg spend & total revenue.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```

17 SELECT subscription_status,
18       COUNT(customer_id) AS total_customers,
19       ROUND(AVG(purchase_amount),2) AS avg_spend,
20       ROUND(SUM(purchase_amount),2) AS total_revenue
21 FROM customer_behavior
22 GROUP BY subscription_status
23 ORDER BY total_revenue DESC, avg_spend DESC;
24

```

Data Output Messages Notifications

Showing rows: 1 to 2 Page No: 1 of 1

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

Total rows: 2 Query complete 00:00:00.239

CRLF

Type here to search

11/22/2025

Q4: Which age groups are most likely to subscribe?

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```

21 FROM customer_behavior
22 GROUP BY subscription_status
23 ORDER BY total_revenue DESC, avg_spend DESC;
24
25 -- Q4: Which age groups are most likely to subscribe?
26 SELECT
27     age_group,
28     subscription_status,

```

Data Output Messages Notifications

Showing rows: 1 to 8 Page No: 1 of 1

	age_group	subscription_status	customers
1	Adult	No	688
2	Adult	Yes	254
3	Middle Aged	No	702
4	Middle Aged	Yes	284
5	Senior	No	696
6	Senior	Yes	248
7	Young Adult	No	761
8	Young Adult	Yes	267

Total rows: 8 Query complete 00:00:00.225

CRLF Ln 25

Type here to search

11/22/2025

Q5: Segment customers as New, Returning, Loyal based on previous purchases.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
-- Customer Loyalty & Repeat Behavior--
-- Q5: Segment customers as New, Returning, Loyal based on previous purchases.
WITH customer_type AS (
  SELECT customer_id, previous_purchases,
  CASE
    WHEN previous_purchases = 1 THEN 'New'
    WHEN previous_purchases BETWEEN 2 AND 10 THEN 'Returning'
    ELSE 'Loyal'
  END AS customer_segment
  FROM customer_behavior
)
SELECT customer_segment, COUNT(*) AS number_of_customers
FROM customer_type
GROUP BY customer_segment;
```

Data Output Messages Notifications

customer_segment	number_of_customers
Loyal	3116
New	83
Returning	701

Total rows: 3 Query complete 00:00:00.235

CRLF

Type here to search

11/22/2025

Q6: Are repeat buyers (>5 purchases) more likely to subscribe?

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
SELECT customer_segment, COUNT(*) AS number_of_customers
FROM customer_type
GROUP BY customer_segment;

-- Q6: Are repeat buyers (>5 purchases) more likely to subscribe?
SELECT subscription_status,
  COUNT(customer_id) AS repeat_buyers
FROM customer_behavior
WHERE previous_purchases > 5
GROUP BY subscription_status;
```

Data Output Messages Notifications

subscription_status	repeat_buyers
No	2518
Yes	958

Total rows: 2 Query complete 00:00:00.350

CRLF Ln 52, 9

Type here to search

11/22/2025

Q7: Which products drive repeat customers (>3 previous purchases)?

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
-- Q7: Which products drive repeat customers (>3 previous purchases)?
SELECT
    item_purchased,
    COUNT(*) AS orders_from_repeat_customers,
    ROUND(SUM(purchase_amount), 2) AS revenue_from_repeat_customers
FROM customer_behavior
WHERE previous_purchases > 3
```

Data Output Messages Notifications

Showing rows: 1 to 10 Page No: 1 of 1

	item_purchased text	orders_from_repeat_customers bigint	revenue_from_repeat_customers numeric
1	Blouse	164	9877.00
2	Jewelry	164	9660.00
3	Dress	160	9851.00
4	Shirt	157	9644.00
5	Pants	155	9212.00
6	Belt	154	9388.00
7	Shorts	153	9187.00
8	Sweater	152	8847.00
9	Socks	151	8804.00
10	Scarf	151	9170.00

Total rows: 10 Query complete 00:00:00.239

CRLF Ln 6 11/22/2025

Q8: Top 3 most purchased products within each category.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
FROM customer_behavior
GROUP BY category, item_purchased
)
SELECT item_rank, category, item_purchased, total_orders
FROM item_counts
WHERE item_rank <= 3;
```

Data Output Messages Notifications

Showing rows: 1 to 11 Page No: 1 of 1

	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 00:00:00.269

CRLF Ln 84 11/22/2025

Q9: Top 5 products with the highest average review rating

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```

83 WHERE item_rank <= 3;
84
85 -- Q9: Top 5 products with the highest average review rating.
86 SELECT item_purchased,
87        ROUND(AVG(review_rating::numeric),2) AS avg_product_rating
88 FROM customer_behavior
89 GROUP BY item_purchased
90 ORDER BY AVG(review_rating) DESC
91 LIMIT 5;

```

Data Output Messages Notifications

Showing rows: 1 to 5 Page No: 1 of 1

item_purchased	avg_product_rating
1 Gloves	3.86
2 Sandals	3.84
3 Boots	3.81
4 Hat	3.81
5 T-shirt	3.78

Total rows: 5 Query complete 00:00:00.216

CRLF Ln 11/22/2025

Q10: Top 5 products with highest percentage of discounted purchases.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql* x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```

95 -- Q10: Top 5 products with highest percentage of discounted purchases.
96 SELECT item_purchased,
97        ROUND(
98          100.0 * SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END) / COUNT(*),
99          2) AS discount_rate
100 FROM customer_behavior
101 GROUP BY item_purchased
102 ORDER BY discount_rate DESC
103 LIMIT 5;

```

Data Output Messages Notifications

Showing rows: 1 to 5 Page No: 1 of 1

item_purchased	discount_rate
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 00:00:00.250

CRLF Ln 11/22/2025

Q11: Categories most discount-sensitive (highest discount %).

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql* x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
-- Q11: Categories most discount-sensitive (highest discount %).
SELECT
  category,
  ROUND(
    100.0 * SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END) / COUNT(*),
    2) AS discount_share_pct,
    COUNT(*) AS total_orders
FROM customer_behavior
GROUP BY category
```

Data Output Messages Notifications

category	discount_share_pct	totalOrders
Outerwear	44.44	324
Accessori...	43.79	1240
Footwear	43.24	599
Clothing	42.08	1737

Showing rows: 1 to 4 Page No: 1 of 1

Total rows: 4 Query complete 00:00:00.229 CRLF Ln 105, Col 4

Type here to search 3:57 AM 11/22/2025

Q12: Compare average purchase amounts for Standard vs Express shipping.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql* x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
--Shipping Choice & Customer Experience
-- Q12: Compare average purchase amounts for Standard vs Express shipping.
SELECT shipping_type,
  ROUND(AVG(purchase_amount),2) AS avg_purchase_amount
FROM customer_behavior
WHERE shipping_type IN ('Standard','Express')
GROUP BY shipping_type;
```

Data Output Messages Notifications

shipping_type	avg_purchase_amount
Standard	58.46
Express	60.48

Showing rows: 1 to 2 Page No: 1 of 1

Total rows: 2 Query complete 00:00:00.219 CRLF Ln 121, Col 9

Type here to search 3:58 AM 11/22/2025

Q13: Does Express shipping improve customer satisfaction (avg review rating)?

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql* x

customer_behavior/postgres@PostgreSQL 18

Query Query History

```
-- Q13: Does Express shipping improve customer satisfaction (avg review rating)?
SELECT
  shipping_type,
  ROUND(AVG(review_rating::numeric), 2) AS avg_rating,
  COUNT(*) AS total_orders
FROM customer_behavior
GROUP BY shipping_type
ORDER BY avg_rating DESC;
```

Data Output Messages Notifications

Showing rows: 1 to 6 Page No: 1 of 1

	shipping_type	avg_rating	totalOrders
1	Standard	3.82	654
2	Express	3.78	646
3	2-Day Shipping	3.76	627
4	Next Day Air	3.72	648
5	Free Shipping	3.72	675
6	Store Pickup	3.71	650

Total rows: 6 Query complete 00:00:00.225

CRLF Ln 1

Type here to search

11/22/2023

Q14: Seasonal demand – revenue by season and category.

pgAdmin 4

File Object Tools Edit View Window Help

Object Explorer Servers

Dashboard x customer_behavior.sql* x

customer_behavior/postgres@PostgreSQL 18

Query Query History

Data Output Messages Notifications

Showing rows: 1 to 16 Page No: 1 of 1

	season	category	totalOrders	totalRevenue
1	Fall	Clothing	427	26220.00
2	Fall	Accessori...	324	19874.00
3	Fall	Footwear	136	8665.00
4	Fall	Outerwear	88	5259.00
5	Spring	Clothing	454	27692.00
6	Spring	Accessori...	301	17007.00
7	Spring	Footwear	163	9555.00
8	Spring	Outerwear	81	4425.00
9	Summer	Clothing	408	23078.00
10	Summer	Accessori...	312	19028.00
11	Summer	Footwear	160	9393.00
12	Summer	Outerwear	75	4278.00
13	Winter	Clothing	448	27274.00
14	Winter	Accessori...	303	18291.00
15	Winter	Footwear	140	8480.00
16	Winter	Outerwear	80	4562.00

Total rows: 16 Query complete 00:00:00.200

CRLF Ln 135, Col 4

Type here to search

4:01 AM 11/22/2023

6. Power BI Dashboard

An interactive dashboard was built in Power BI, including metrics such as:

- Total customers
- Total revenue
- Average purchase amount

- Revenue by category
- Sales by age group
- Subscription-based segmentation



7. Seasonality Insights

Across all seasons, **Clothing** consistently generates strong revenue, with peaks in **Spring and Winter**.

Insight: The business exhibits stable year-round performance with moderate seasonal variation.

8. Strategic Recommendations

Based on the findings from the SQL analysis, the following targeted recommendations are proposed:

8.1 Strengthen Marketing Toward High-Value Segments

- Prioritize campaigns targeting **Male** and **Young Adult** customers, the highest revenue contributors.
- Tailor product recommendations and email campaigns around these demographics.

8.2 Revamp the Subscription Program

Because non-subscribers currently:

- Spend slightly more,
- Generate significantly higher revenue,
- And include the majority of repeat buyers,

Implement improvements such as:

- Exclusive member-only offers
- Loyalty points multiplied
- Early access to new collections
- Personalized incentives for returning and Loyal customers (the 3,800+ most valuable users)

8.3 Optimize Discount Strategy

Since discount rates hover around 42–44% across all categories:

- Reduce blanket discounts that erode margins.
- Shift to **targeted** discounts for new or dormant customers only.
- Re-price products with ~50% discount reliance (e.g., Hat, Sneakers, Coat).

8.4 Capitalize on Repeat-Driving Products

- Promote **Blouse, Jewelry, Dress, Shirt, Pants** as anchor products in advertising.
- Bundle high-performing items (e.g., Blouse + Jewelry) to increase average order value.
- Feature top-rated products (Gloves, Sandals, Boots) on the homepage.

8.5 Improve Customer Experience Without Increasing Logistics Cost

- Keep **Standard Shipping** as the default since it has the highest satisfaction.
- Offer **Express** selectively as an upsell on high-value carts.

8.6 Leverage Seasonality While Maintaining Steady Inventory

- Prepare seasonal campaigns for **Spring and Winter**, when Clothing performs best.
- Maintain consistent inventory year-round to support stable demand.