

CUSTOMER PURCHASE PATTERN ANALYSIS

OBJECTIVE:

The objective of this project was to identify key revenue drivers, customer segments, and growth opportunities using customer purchase data. This project was designed to support data-driven business decision making by identifying revenue drivers and customer growth opportunities.

Project Title: Customer Purchase Pattern Analysis using Python, SQL, and Power BI.

1. Project Overview:

This capstone project demonstrates an end-to-end data analytics workflow. The objective was to analyze customer purchase patterns and generate actionable business insights. The project was implemented using Python for data cleaning, PostgreSQL for data analysis, and Power BI for data visualization.

2. Tools and Technologies Used:

- Python (Pandas) – Data Cleaning & Preparation.
- PostgreSQL – SQL Data Analysis.
- Power BI – Dashboard & Insights Generation.

3. Dataset Description:

- Customer demographics (Age group, Gender).
- Purchase behaviour (Purchase amount, Frequency, Previous purchases).
- Product details (Category, Item purchased).
- Discount & Subscription information.
- Shipping type.

4. Data Cleaning & Preparation (Python – Pandas):

- ✓ Reading CSV file from local system using “df = pd.read_csv (“file path”)” command in Pandas.
- ✓ Null values handling:
Finding null values using “df.isnull().sum()”
Handling missing values using ‘Median’ to reduce outlier impact
- ✓ Categorical values standardization:
Purchase frequency mapping (Weekly, Monthly → Days).
- ✓ New derived columns:
Purchase Frequency Days.
Customer segment (New / Returning / Loyal).
Cleaned numeric metrics.
- ✓ Data consistency check:
Verified if ‘discount applied’ and ‘Promo code Used’ were redundant and therefore removed. ‘Promo code Used’.
- ✓ The cleaned Python dataset was loaded into a PostgreSQL database for further analysis.

These steps ensured data accuracy, consistency, and readiness for SQL analysis and dashboard reporting.

SQL Analysis – Key Business Insights

Insight 1: Gender-wise Revenue:

Male customers generate higher total revenue than female customers. Indicates higher average spend or purchase volume by males.

```
4 --Q1. What is the total revenue generated by male vs. female customers?
5 select gender, SUM(purchase_amount) as revenue
6 from customer
7 group by gender
```

Data Output Messages Notifications

| | gender | revenue |
|---|--------|---------|
| 1 | Female | 75191 |
| 2 | Male | 157890 |

Showing rows: 1 to 2 | Page No: 1 of 1 | < << > >>

Insight 2: Discount Users Behaviour:

Customers who used discounts still spent more than the average purchase amount, indicating they are high-value customers.

```
9
10 --Q2. Which customers used a discount but still spent more than the average purchase amount?
11 select customer_id, purchase_amount
12 from customer
13 where discount_applied = 'Yes' and purchase_amount >= (select AVG(purchase_amount) from customer)
```

Data Output Messages Notifications

| | customer_id | purchase_amount |
|---|-------------|-----------------|
| 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 |

Showing rows: 1 to 839 | Page No: 1

Insight 3: Top Rated Products:

Top 5 highest rated products:

- Gloves
- Sandals
- Boots
- Hat
- Skirt

```
17 -- Q3. Which are the top 5 products with the highest average review rating?
18 select item_purchased, round(avg(review_rating::numeric),2) as "Average Product Rating"
19 from customer
20 group by item_purchased
21 order by avg(review_rating) desc
22 limit 5
23
```

Data Output Messages Notifications

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

| | item_purchased | Average Product Rating |
|---|----------------|------------------------|
| 1 | Gloves | 3.86 |
| 2 | Sandals | 3.84 |
| 3 | Boots | 3.82 |
| 4 | Hat | 3.80 |
| 5 | Skirt | 3.78 |

Insight 4: Shipping Type Impact:

Express shipping customers spend slightly more than Standard.

Faster delivery = higher willingness to pay

```
25 --Q4. Compare the average Purchase Amounts between Standard and Express Shipping.
26 select shipping_type,
27 ROUND(AVG(purchase_amount),2)
28 from customer
29 where shipping_type in ('Standard','Express')
30 group by shipping_type;
```

Data Output Messages Notifications

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

| | shipping_type | round |
|---|---------------|-------|
| 1 | Standard | 58.46 |
| 2 | Express | 60.48 |

Insight 5: Subscription Impact:

- Non-subscribers generate higher total revenue.
- Subscribers have slightly higher average spend per order.

Subscription helps retention, but volume still comes from non-subscribers so

The business should encourage more customers to adopt subscriptions to improve retention and build predictable revenue.

```
34 --Q5. Do subscribed customers spend more? Compare average spend and total revenue
35 --between subscribers and non-subscribers.
36 SELECT subscription_status,
37     COUNT(customer_id) AS total_customers,
38     ROUND(AVG(purchase_amount),2) AS avg_spend,
39     ROUND(SUM(purchase_amount),2) AS total_revenue
40 FROM customer
41 GROUP BY subscription_status
42 ORDER BY total_revenue,avg_spend DESC;
```

Data Output Messages Notifications

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

Showing rows: 1 to 2 | Page No: 1

Insight 6: Discount by Product:

Highest discount usage:

- Hat
- Sneakers
- Coat

Discounts work best for fashion & seasonal items.

```
45 --Q6. Which 5 products have the highest percentage of purchases with discounts applied?
46 SELECT item_purchased,
47     ROUND(100.0 * SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END)/COUNT(*),2) AS discount_rate
48 FROM customer
49 GROUP BY item_purchased
50 ORDER BY discount_rate DESC
51 LIMIT 5;
```

Data Output Messages Notifications

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

Showing rows: 1 to 5 | Page No: 1

Insight 7: Customer Segmentation:

Customer distribution:

- Loyal customers – highest
- Returning customers – medium
- New customers – lowest

Strong retention base exists.

```
78 --Q7. Segment customers into New, Returning, and Loyal based on their total
79 -- number of previous purchases, and show the count of each segment.
80 with customer_type as (
81   SELECT customer_id, previous_purchases,
82   CASE
83     WHEN previous_purchases = 1 THEN 'New'
84     WHEN previous_purchases BETWEEN 2 AND 10 THEN 'Returning'
85     ELSE 'Loyal'
86   END AS customer_segment
87   FROM customer)
88
89 select customer_segment, count(*) AS "Number of Customers"
90 from customer_type
91 group by customer_segment;
92
93 --Q8. What are the top 3 most purchased products within each category?
```

Data Output Messages Notifications

| | customer_segment | Number of Customers |
|---|------------------|---------------------|
| 1 | Loyal | 3116 |
| 2 | New | 83 |
| 3 | Returning | 701 |

Showing rows: 1 to 3 Page No: 1 of 1

Insight 8: Category-wise Top Products:

Each category has clear best-selling items, e.g.:

- Accessories → Jewellery, Sunglasses
- Clothing → Pants, Shirts
- Footwear → Sandals, Shoes

Inventory optimization possible

```
70 --Q8. What are the top 3 most purchased products within each category?
71 WITH item_counts AS (
72   SELECT category,
73         item_purchased,
74         COUNT(customer_id) AS total_orders,
75         ROW_NUMBER() OVER (PARTITION BY category ORDER BY COUNT(customer_id) DESC) AS item_rank
76   FROM customer
77   GROUP BY category, item_purchased
78 )
79 SELECT item_rank, category, item_purchased, total_orders
80 FROM item_counts
81 WHERE item_rank <=3;
```

Data Output Messages Notifications

| | item_rank | category | item_purchased | total_orders |
|---|-----------|--------------|----------------|--------------|
| 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 |

Showing rows: 1 to 11 Page No: 1

Insight 9: Repeat Buyers & Subscription:

Majority repeat buyers are **not subscribed**. Missed opportunity for subscription upsell

“Therefore, the business should run a dedicated subscription campaign specifically targeted at repeat buyers.”

```
85 --Q9. Are customers who are repeat buyers (more than 5 previous purchases) also likely to subscribe?
86 SELECT subscription_status,
87     COUNT(customer_id) AS repeat_buyers
88 FROM customer
89 WHERE previous_purchases > 5
90 GROUP BY subscription_status;
```

Data Output Messages Notifications

Showing rows: 1 to 2 | Page No: 1

| | subscription_status | repeat_buyers |
|---|---------------------|---------------|
| 1 | No | 2518 |
| 2 | Yes | 958 |

Insight 10: Age Group Revenue:

Revenue contribution:

1. Young Adults – highest
2. Middle-aged
3. Adults
4. Seniors

Young adults are primary revenue drivers

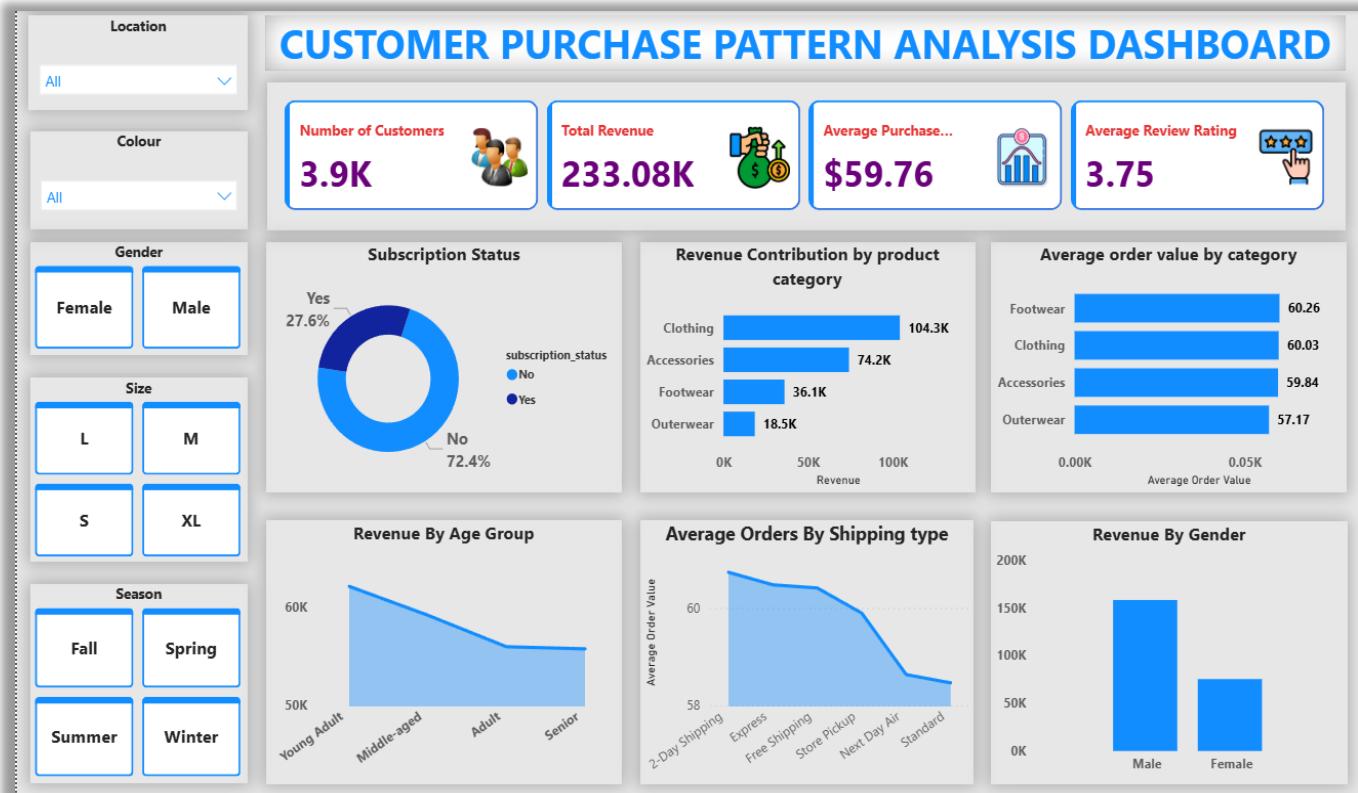
```
93 --Q10. What is the revenue contribution of each age group?
94 SELECT
95     age_group,
96     SUM(purchase_amount) AS total_revenue
97 FROM customer
98 GROUP BY age_group
99 ORDER BY total_revenue desc;
```

Data Output Messages Notifications

Showing rows: 1 to 4 | Page No: 1

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

CUSTOMER PURCHASE PATTERN ANALYSIS DASHBOARD



1. Overall Business Performance:

- The dashboard analyses data from approximately 3,900 customers.
- The business has generated around 233K in total revenue.
- The Average Order Value (AOV) is about 59.76, indicating that customers spend nearly 60 per order on average.
- The average customer rating is 3.75, showing moderate to good customer satisfaction.

Insight:

- ✓ The business is performing steadily, with healthy customer spending and acceptable satisfaction levels.

2. Subscription Behaviour:

- Only about 27.6% of customers are subscribed, while 72.4% are non-subscribers.

Insight:

- ✓ A large majority of customers are not subscribed. This represents a strong subscription upselling opportunity to increase customer lifetime value and retention.

3. Revenue Contribution by Product Category:

- Clothing is the highest revenue-generating category (~104K).
- Accessories is the second highest (~74K).
- Footwear contributes moderately.
- Outerwear generates the lowest revenue.

Insight:

- ✓ Clothing and Accessories are the core revenue drivers of the business.

Action:

- ✓ Marketing, promotions, and inventory planning should primarily focus on these categories.

4. Average Order Value by Category (High-Value Insight):

- Footwear has the highest Average Order Value (~60.26).
- Clothing and Accessories are close to the overall average.
- Outerwear has the lowest AOV (~57.17).

Insight:

- ✓ Although Footwear has fewer orders, it generates higher value per transaction, indicating a premium customer segment. Footwear can be positioned as a premium category.

Action:

- Introduce premium bundles, limited editions, and targeted upselling strategies for this category.

5. Revenue by Age Group:

- Young Adults contribute the highest share of revenue.
- This is followed by Middle-aged customers, Adults, and Seniors.

Insight:

- ✓ Young adults are the primary revenue-driving customer segment.

Action:

- Marketing campaigns, product positioning, and user experience design should strongly target young adults.

6. Shipping Type vs Spending Behaviour:

- Customers choosing 2-Day or Express shipping show a higher average order value compared to standard shipping users.

Insight:

- ✓ Customers who prefer faster delivery tend to be high-value buyers.

Action:

- Express and faster shipping options should be highlighted at checkout as an upselling strategy.

7. Revenue by Gender:

- Male customers generate higher total revenue than female customers.

Insight:

- ✓ Male customers currently contribute more to overall revenue.

Action:

- Male-focused campaigns and product strategies can be explored for further growth.

KEY BUSINESS OPPORTUNITIES AND RECOMMENDATIONS:

- Increase subscription adoption among existing customers.
- Promote Footwear as a premium, high-value category.
- Focus marketing strategies on young adults.
- Upsell express and faster shipping options.
- Optimize inventory and promotions around Clothing and Accessories.

CONCLUSION:

This Customer Purchase Pattern Analysis project demonstrates a complete end-to-end data analytics workflow using Python, SQL, and Power BI. The analysis identified key revenue drivers, high-value customer segments, and important purchasing patterns. Results show that young adult customers and fashion-related categories are the primary contributors to revenue, while the low subscription adoption rate highlights a major growth opportunity. Overall, the project successfully converts raw customer data into actionable business insights to support data-driven decision-making.