

Customer Shopping Behaviour Analysis

Understanding Consumer Patterns to Improve Engagement, Sales, and Retention

Prepared by: Abdou Salam Sisawo

Domain: Data Science / Business Analytics

Tool Stack: Python (Pandas, NumPy), MySQL, Power BI



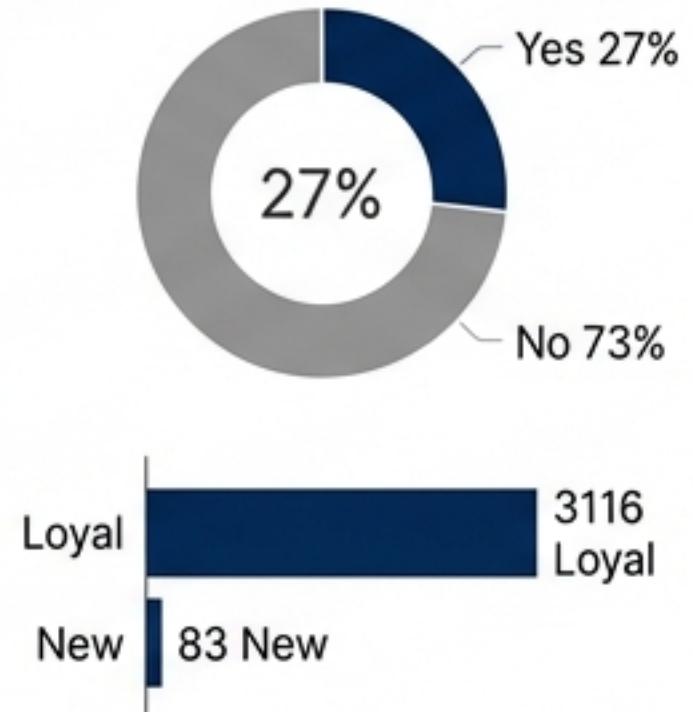
Executive Summary: Data-Driven Strategies to Optimise Retention

The Challenge

The retail environment is shifting, requiring a move from intuition-based decisions to evidence-based strategies regarding subscriptions and discounting. Management currently lacks visibility into how these levers impact loyalty.

Key Findings

- Subscription Power:** 27% of customers are subscribed; this segment is critical for consistent engagement.
- Revenue Drivers:** 'Clothing' is the dominant category. 'Young Adults' are the highest-grossing age demographic.
- Customer Base:** The customer base is highly mature, with 3,116 'Loyal' customers compared to only 83 'New' ones.



The Approach

A full-stack analysis pipeline utilising Python for cleaning, SQL for aggregation, and Power BI for visualisation to transform raw transactions into intelligence.

Strategic Recommendation

Focus resources on retaining the high-volume 'Loyal' segment through personalised incentives and increasing subscription adoption among the 73% non-subscriber base.

Shifting Consumer Patterns Demand Enhanced Visibility

Situation

A leading retail company has observed notable changes in purchasing patterns across demographics and sales channels (online and offline).



Complication

Management lacks a data-driven understanding of how specific levers—specifically discounts, subscription status, and product ratings—impact overall sales performance and loyalty.



Key Business Question

How can consumer shopping data be leveraged to identify trends, improve customer engagement, and optimise marketing and product strategies?



Objectives: Identify purchasing patterns, determine factors influencing repeat purchases, and optimise marketing strategies.

From Raw Transactions to Strategic Intelligence

Data Preparation (Python)

Cleaning, handling missing values, standardising variables.

```
# Final check of the cleaned dataset
df.head()
```

	customer_id	age	gender	Item_purchased	category	purchase
0	1	35	Male	Blouse	Clothing	
1	2	19	Male	Sweater	Clothing	
2	3	50	Male	Jeans	Clothing	
3	4	21	Male	Sandals	Footwear	
4	5	45	Male	Blouse	Clothing	
5	6	45	Male	Blouse	Clothing	
6	7	75	Male	Blouse	Clothing	

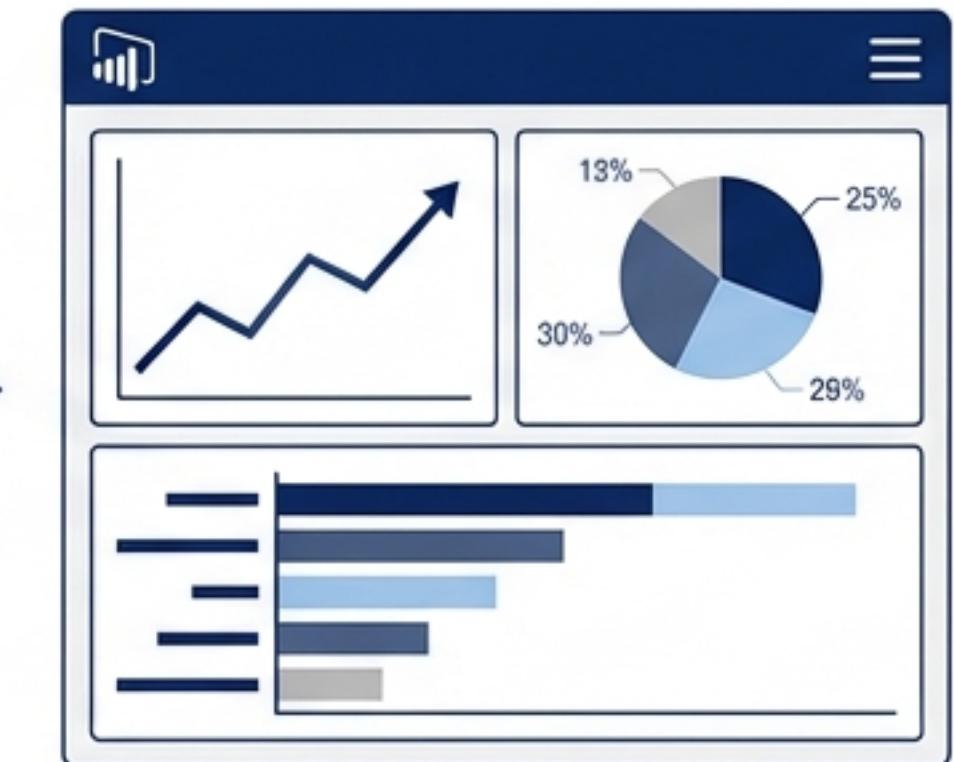
Database Design (MySQL)

Structured queries for aggregation, segmentation, and KPI computation.

customer_id	age	gender	Item_purchased	category	purchase_amount	location
1	55	Male	Blouse	Clothing	53	Rensselaer
2	19	Male	Sweater	Clothing	64	Manhattan
2	30	Male	Jeans	Clothing	73	Boston
4	21	Male	Sandals	Footwear	80	Shore Island
2	45	Male	Beuse	Clothing	49	Gregory
6	45	Male	Sneakers	Footwear	20	Wyoming
7	62	Male	Shirt	Clothing	89	Montana
9	27	Male	Shirts	Clothing	55	Montana
9	25	Male	Coat	Outerwear	97	West Virginia
10	57	Male	Handbag	Accessories	31	Pittsburgh
11	53	Male	Shoes	Footwear	32	Arkansas
12	30	Male	Shirts	Clothing	68	Hawaii
12	61	Male	Coat	Outerwear	72	Delaware
14	65	Male	Dress	Clothing	51	New Hampshire
24	64	Male	Coat	Outerwear	53	New York
16	64	Male	Skirt	Clothing	53	Rhode Island
17	15	Male	Sunglasses	Accessories	36	Alabama
28	65	Male	Coat	Clothing	81	Rhode Island
no	no	no	no	no	no	Carolina

Visualisation (Power BI)

Interactive dashboards for stakeholder monitoring.



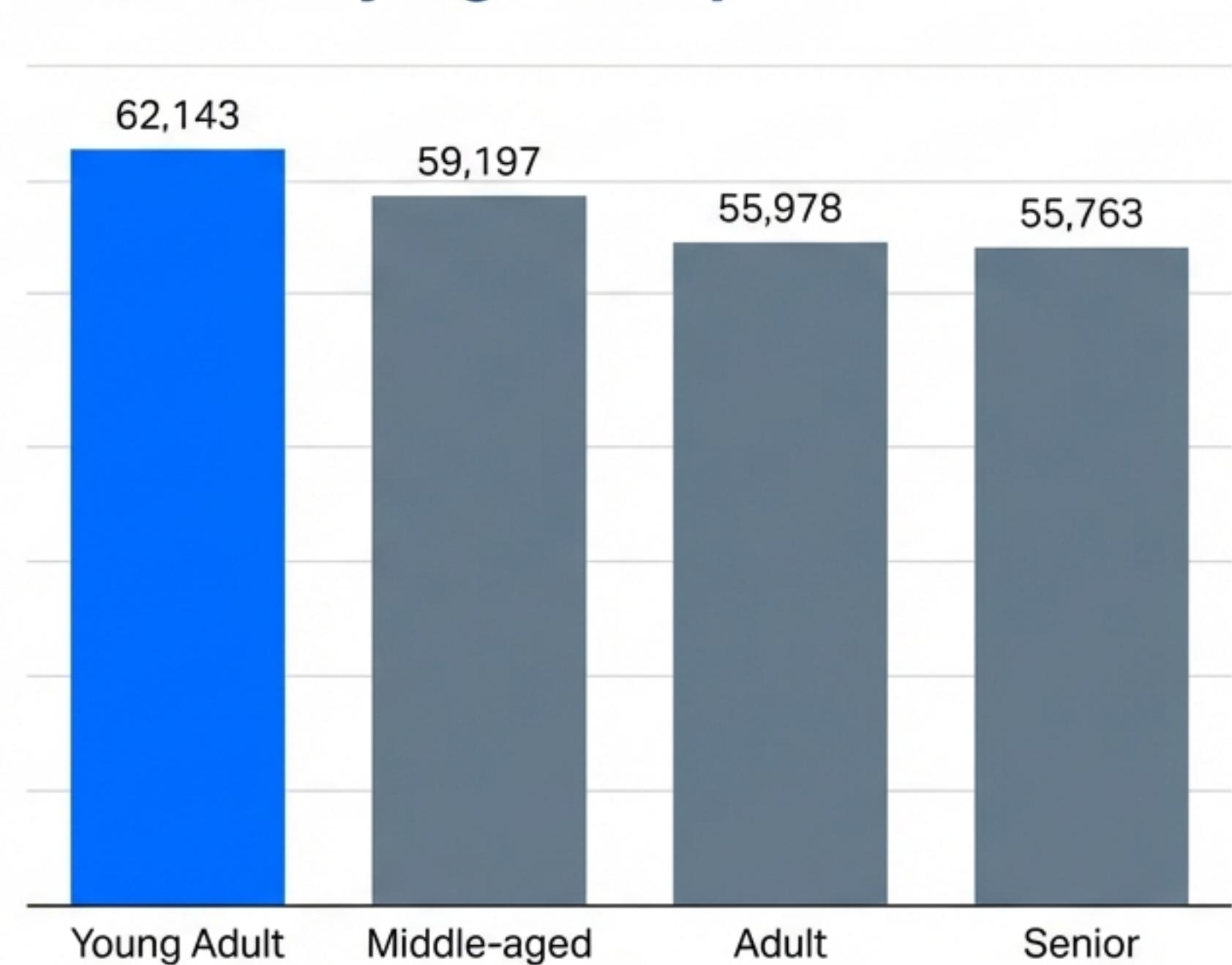
Dataset Scope: 3,900 Total Customers / 3,900 Total Orders tracking Age, Gender, Purchase Amount, and Subscription Status.

Young Adults and Male Shoppers Dominate Revenue Streams

Gender Revenue Disparity

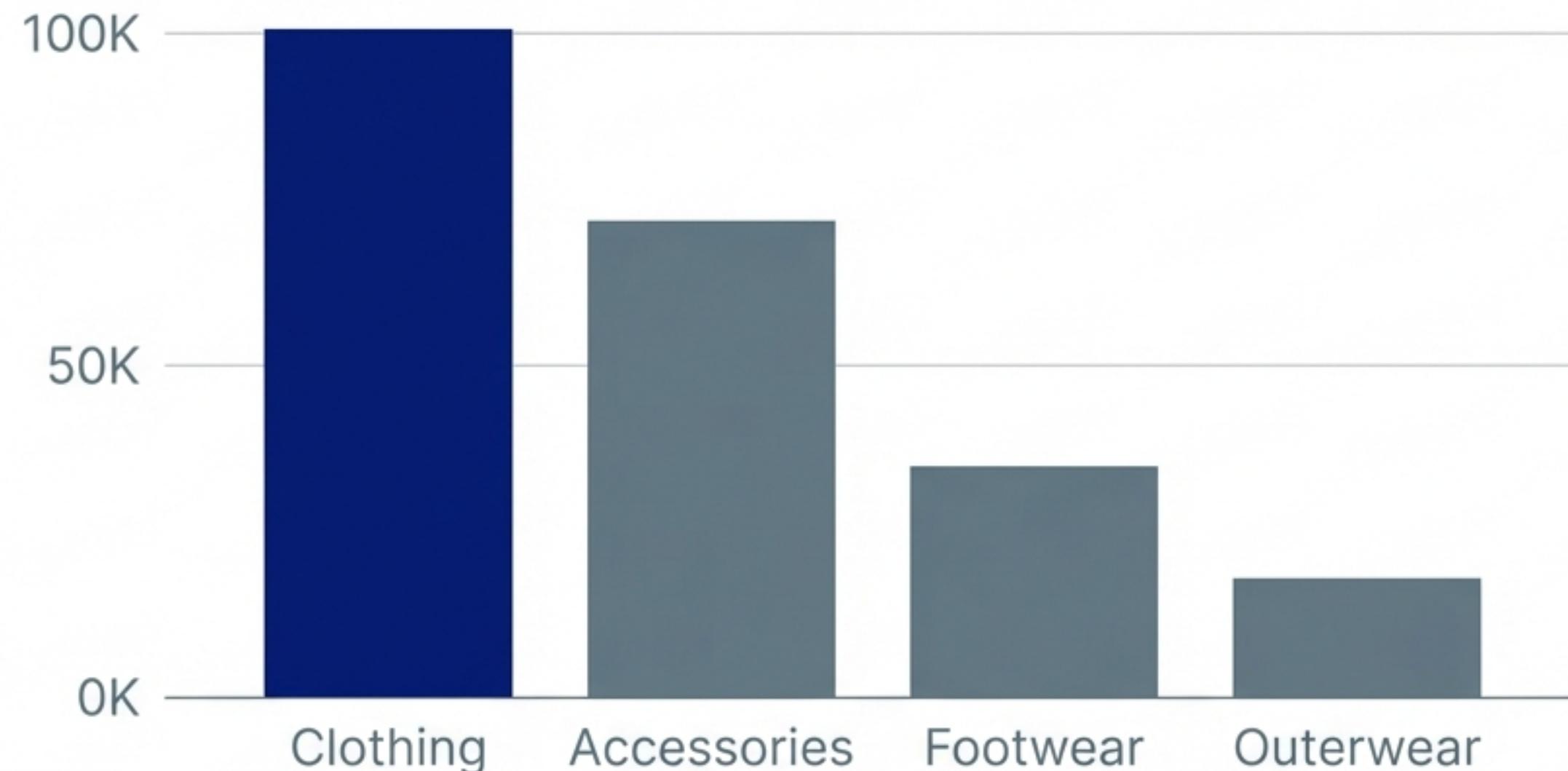


Revenue by Age Group



Clothing and Accessories Drive the Majority of Turnover

Revenue by Category



Top Sellers per Category

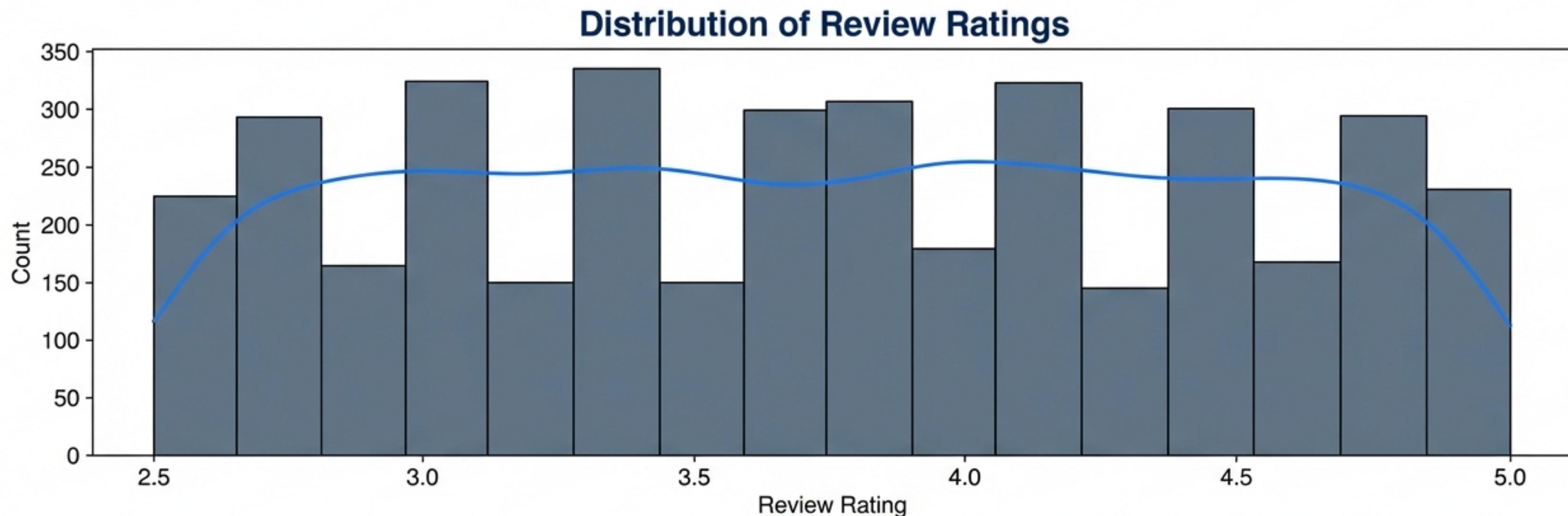
Accessories: Jewelry, Sunglasses, Belts

Clothing: Blouse, Pants, Shirt

Footwear: Sandals, Shoes, Sneakers

Outerwear: Jacket, Coat

Strong Customer Sentiment Across Key Products



Top Rated Products (Avg / 5.0)

Gloves: 3.86

Sandals: 3.84

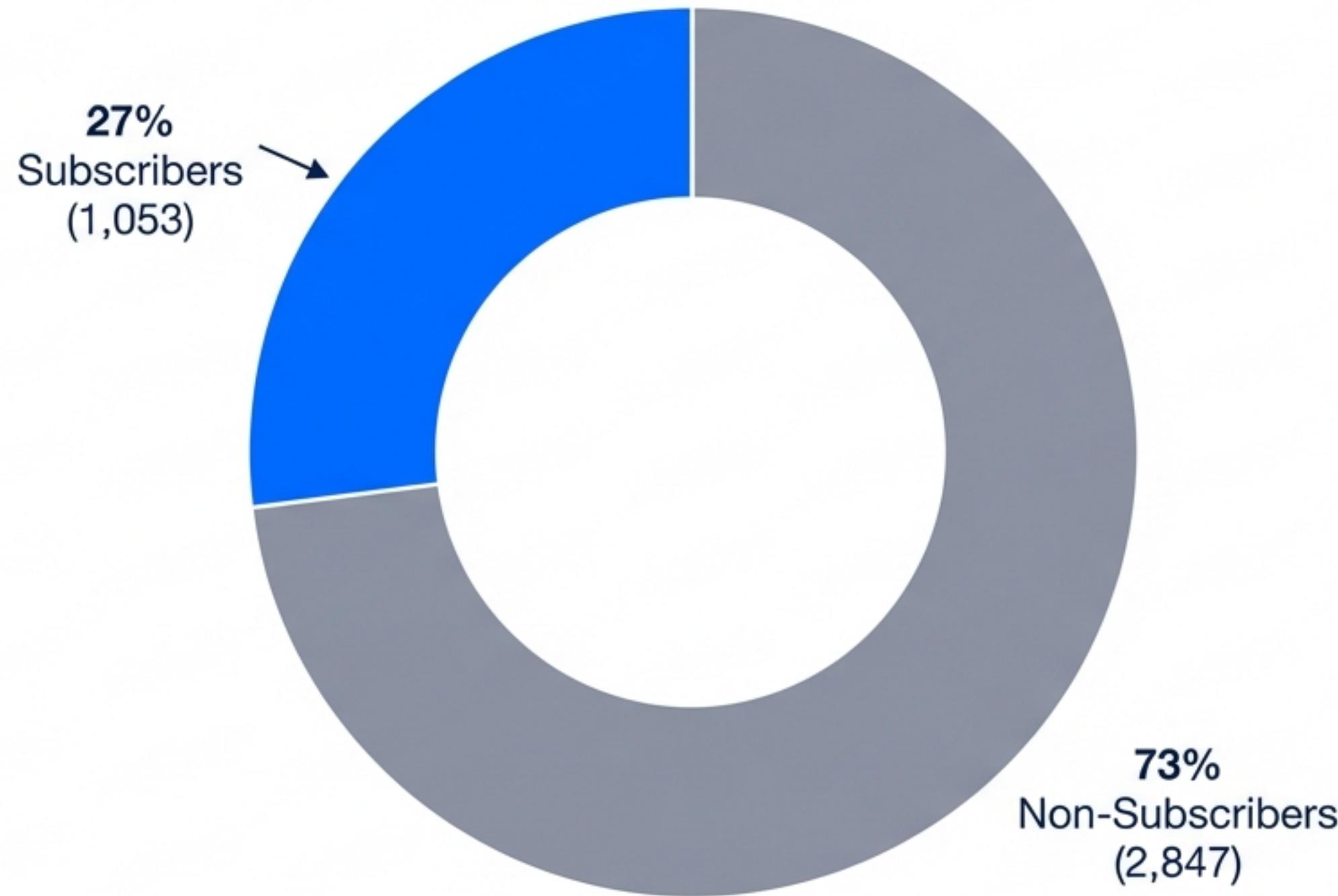
Boots: 3.82

Hat: 3.80

Skirt: 3.78

Average Overall Rating: 3.75

The Subscription Economy: An Untapped Growth Lever



Opportunity: Nearly 3/4 of the customer base operates transactionally. Converting even a fraction of the 2,847 non-subscribers represents the largest lever for CLV growth.

Subscriptions Drive Retention and Consistent Engagement

Basket Size (Immediate Value)

59.86 59.49

(Non-Subscribers)

(Subscribers)

Average Purchase Amount is nearly identical.
Subscription status does not increase
single-transaction value.

Retention (Long-Term Value)

Observation:

While non-subscribers have higher volume due to population size (2,518 repeat buyers), the subscription model ensures a guaranteed engagement channel.

Strategic Takeaway: The value of a subscriber is not in the single transaction, but in the predictability of revenue.

Identifying Discount-Dependent Product Lines

Top Products by Discount Rate

Item	Discount Rate	Insight
Hat	50.00%	Highest dependency
Sneakers	49.66%	Margin risk
Coat	49.07%	High value item
Sweater	48.17%	Seasonal
Pants	47.37%	Volume driver

Insight: 839 "High-Spending Discount Users" used discounts but spent above average. This suggests discounts can be used to upsell rather than just clear stock.

Express Shipping Correlates with Higher Basket Value

Standard Shipping

\$58.46

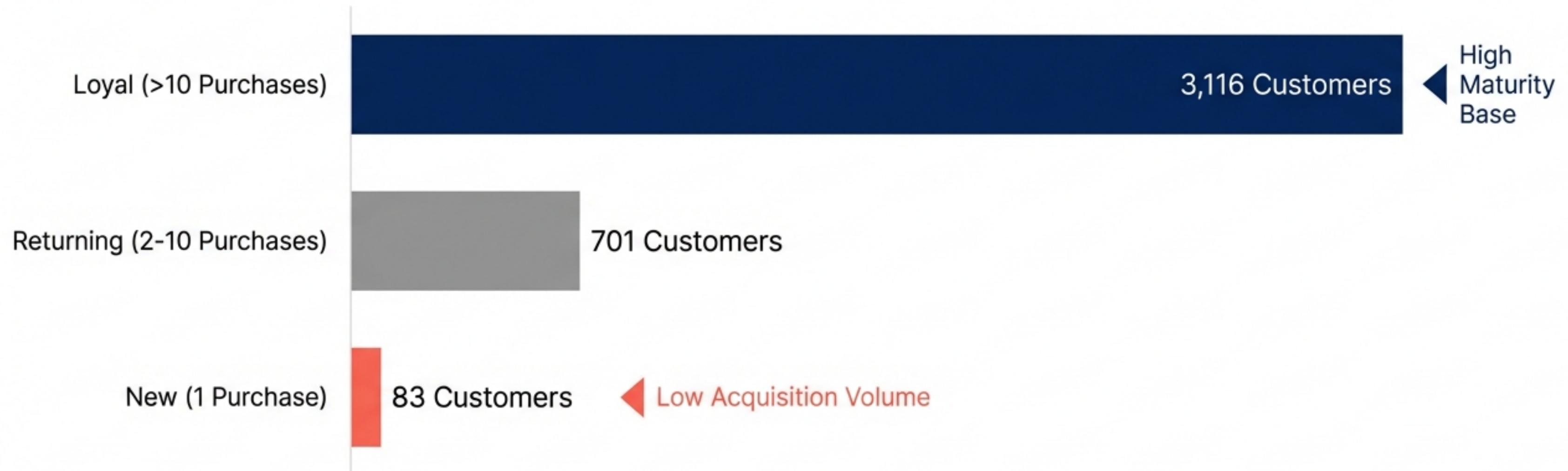
Express Shipping

\$60.48

↗ +\$2.02

Customers opting for speed are willing to spend marginally more per order.
Prioritising Express options at checkout could incrementally lift overall revenue.

A Highly Mature Customer Base Requires Retention Focus



Segmentation Logic: New = 1 purchase, Returning = 2-10 purchases, Loyal = >10 purchases.

Continuous Monitoring via Interactive Dashboards



Strategic Recommendations

1. Incentivise Subscriptions

Convert the 73% non-subscriber base. While basket size is similar, subscribers offer higher retention stability.

2. Optimise Discount Strategy

Move away from blanket discounts. Focus on 'High-Spending Discount Users' to protect margins on products like Hats and Sneakers.

3. Leverage Social Proof

Utilise the strong 3.75 average rating in marketing campaigns, specifically highlighting top-rated items like Gloves and Sandals.

4. Personalise for Loyalty

The customer base is overwhelmingly 'Loyal' (3,116 users). Shift budget from broad acquisition to personalised retention campaigns.

Conclusion: Turning Transactional Data into Business Growth

```
SELECT gender, SUM(purchase_amount) AS revenue FROM customer GROUP BY gender;
```

--Q2. Which customers used a discount but still spent more than the average purchase amount?

```
SELECT customer_id, purchase_amount, discount_applied, purchase_amount - (purchase_amount * discount_applied) AS net_spent FROM customer;
```

--Q3. Which are the top 5 products with the highest average review rating?

```
SELECT item_purchased, ROUND(AVG(review_rating), 2) AS "Average Product Rating" FROM purchases GROUP BY item_purchased ORDER BY AVG(review_rating) LIMIT 5;
```

Identifying Subscription Status, Discount Strategy, and Customer Sentiment as the three pillars of purchasing behaviour, the company can transition from reactive sales

--Q4. Compare the average Purchase Amounts between Standard and Express Shipping.

```
SELECT shipping_type, ROUND(AVG(purchase_amount), 2) AS "Average Purchase Amount" FROM purchases WHERE shipping_type IN ('Standard', 'Express')
```

```
GROUP BY shipping_type;
```

--Q5. Do subscribed customers spend more? Compare average spend and total revenue between subscribers and non-subscribers.

```
SELECT Subscription_status, COUNT(customer_id) AS Total_Customers, SUM(purchase_amount) AS revenue, AVG(purchase_amount) AS 'Average Purchase Amount' FROM purchases GROUP BY Subscription_status;
```