

CUSTOMER PURCHASE INSIGHTS

1. Project Overview

This project analyzes customer shopping behaviour using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behaviour to guide strategic business decisions.

2. Business Problem Statement

A leading retail company wants to better understand its customers' shopping behavior in order to improve sales, customer satisfaction, and long-term loyalty. The management team has noticed changes in purchasing patterns across demographics, product categories, and sales channels (online vs. offline). They are particularly interested in uncovering which factors, such as discounts, reviews, seasons, or payment preferences, drive consumer decisions and repeat purchases.

Tasked with analyzing the company's consumer behavior dataset to answer the following overarching business question:

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

3. Dataset Summary

- Rows: 3,900 - Columns: 18

-Key Features:

- Customer demographics (Age, Gender, Location, Subscription Status)
- Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
- Shopping behaviour (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 values in Review Rating column

4. Tools and Technologies

- Python (Jupyter Notebook)
- SQL (MySQL Workbench)
- Power BI

5. Methods and Insights

a. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- Data Loading: Imported the dataset using pandas.
- Initial Exploration: Used `df.info()` to check structure and `.describe()` for summary statistics.

	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
count	3900	3900	3900	3900	3900	3900	3900	3900	3900	3900	3863	3900	3900	3900	3900	3900	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2	NaN	6	7
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No	NaN	PayPal	Every 3 Months
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223	NaN	677	584
mean	1950.5	44.06846	NaN	NaN	NaN	59.76436	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN	25.35154	NaN	NaN
std	1125.97735	15.20759	NaN	NaN	NaN	23.68539	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN	14.44713	NaN	NaN
min	1	18	NaN	NaN	NaN	20	NaN	NaN	NaN	NaN	2.5	NaN	NaN	NaN	NaN	1	NaN	NaN
25%	975.75	31	NaN	NaN	NaN	39	NaN	NaN	NaN	NaN	3.1	NaN	NaN	NaN	NaN	13	NaN	NaN
50%	1950.5	44	NaN	NaN	NaN	60	NaN	NaN	NaN	NaN	3.8	NaN	NaN	NaN	NaN	25	NaN	NaN
75%	2925.25	57	NaN	NaN	NaN	81	NaN	NaN	NaN	NaN	4.4	NaN	NaN	NaN	NaN	38	NaN	NaN
max	3900	70	NaN	NaN	NaN	100	NaN	NaN	NaN	NaN	5	NaN	NaN	NaN	NaN	50	NaN	NaN

- Missing Data Handling: Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.
- Column Standardization: Renamed columns to snake case for better readability and documentation.
- Feature Engineering:
 - Created age_group column by binning customer ages.
 - Created purchase_frequency_days column from purchase data.
- Data Consistency Check: Verified if discount_applied and promo_code_used were redundant; dropped promo_code_used.
- Database Integration: Connected Python script to MySQL Workbench and loaded the cleaned DataFrame into the database for SQL analysis.

b. Data Analysis using SQL (Business Transactions)

We performed structured analysis in PostgreSQL to answer key business questions:

- Revenue by Gender** – Compared total revenue generated by male vs. female customers.

gender	revenue
Male	157890
Female	75191

- High-Spending Discount Users** – Identified customers who used discounts but still spent above the average purchase amount.

discount_applied	customer_id	purchase_amount
Yes	2	64
Yes	3	73
Yes	4	90
Yes	7	85
Yes	9	97
Yes	12	68
Yes	13	72
Yes	16	81
Yes	20	90
Yes	22	62
Yes	24	88
Yes	29	94
Yes	32	79
Yes	33	67
Yes	35	91
Yes	37	69
Yes	40	60
Yes	41	76

iii. **Top 5 Products by Rating** – Found products with the highest average review ratings.

item_purchased	Avg Product Rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Skirt	3.78

iv. **Shipping Type Comparison** – Compared average purchase amounts between Standard and Express shipping.

shipping_type	round(avg(purchase_amount),2)	
Express	60.48	60.48
Standard	58.46	

v. **Subscribers vs. Non-Subscribers** – Compared average spend and total revenue across subscription status.

avg_spend	subscription_status	total_revenue	total_customers
59.49	Yes	62645	1053
59.87	No	170436	2847

vi. **Discount-Dependent Products** – Identified 5 products with the highest percentage of discounted purchases.

item_purchased	discount_rate
Hat	50.00
Sneakers	49.66
Coat	49.07
Sweater	48.17
Pants	47.37

vii. **Customer Segmentation** – Classified customers into New, Returning, and Loyal segments based on purchase history.

customer_segment	Number_of_Customers
Loyal	3116
Returning	701
New	83

viii. **Top 3 Products per Category** – Listed the most purchased products within each category.

item_rank	category	item_purchased	total_orders
1	Accessories	Jewelry	171
2	Accessories	Sunglasses	161
3	Accessories	Belt	161
1	Clothing	Blouse	171
2	Clothing	Pants	171
3	Clothing	Shirt	169
1	Footwear	Sandals	160
2	Footwear	Shoes	150
3	Footwear	Sneakers	145
1	Outerwear	Jacket	163
2	Outerwear	Coat	161

- ix. **Repeat Buyers & Subscriptions** – Checked whether customers with >5 purchases are more likely to subscribe.

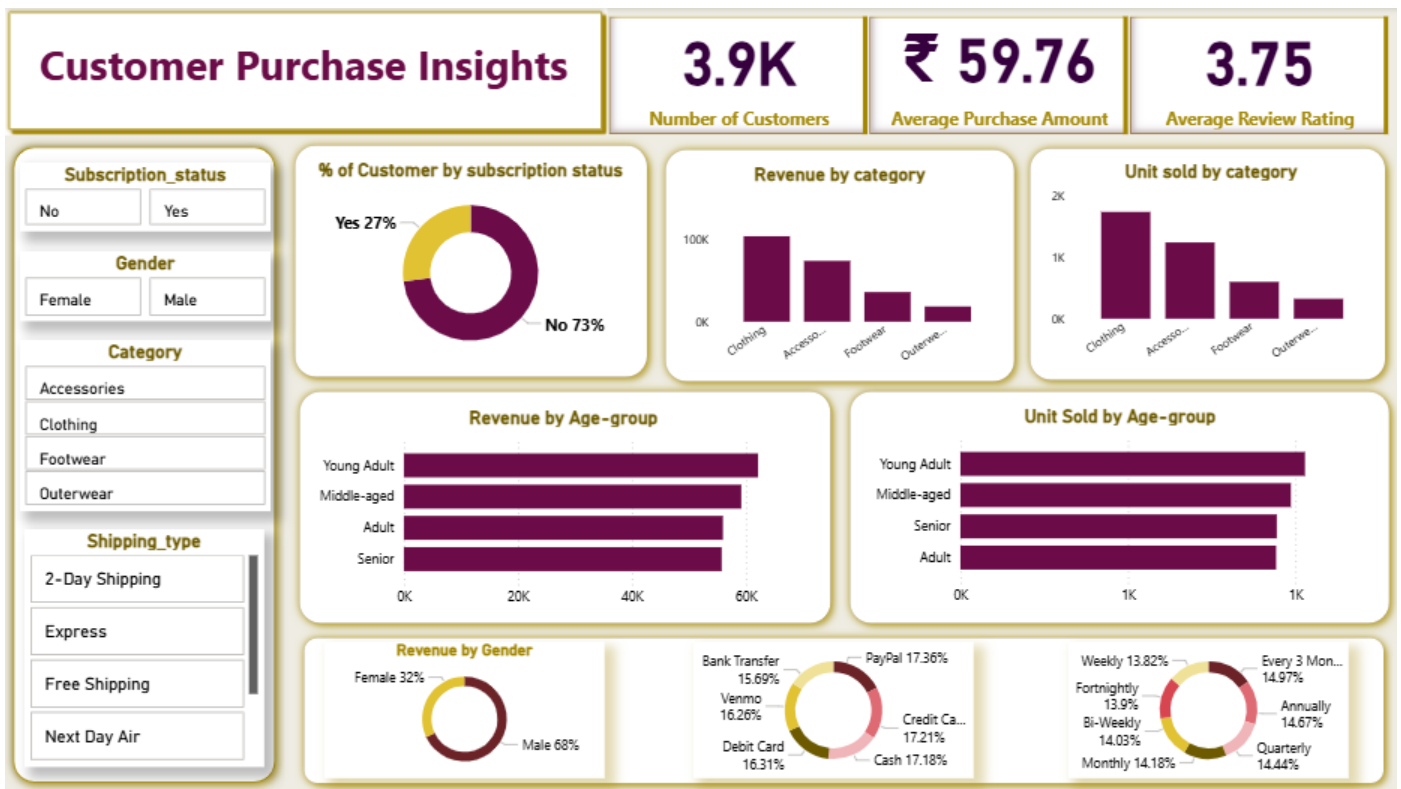
subscription_status	repeat_buyers
Yes	958
No	2518

- x. **Revenue by Age Group** – Calculated total revenue contribution of each age group.

age_group	total_revenue
Young Adult	62143
Middle-aged	59197
Adult	55978
Senior	55763

c. Dashboard in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.



6. Business Recommendations

5 data-backed recommendations that your company can use to increase sales and customer engagement.

- a. **Strengthen Loyalty & Subscription Programs** - Only 27% of customers are subscribed, while 73% are not.

Recommendation:

- Introduce *loyalty rewards* or *exclusive discounts* for subscribers.
- Offer a “first-month free” or points-based reward for every purchase.
- Send personalized product recommendations to subscribers.

Expected Impact:

Increasing the subscription base by even 10–15% could significantly lift recurring sales and customer retention.

- b. **Focus Marketing on “Young Adult” Segment** - The Young Adult group drives the highest revenue and sales volume across all age groups.

Recommendation:

- Prioritize digital marketing (Instagram, YouTube, influencer tie-ups).
- Launch trendy product lines and limited-time offers tailored to this age group.
- Optimize product recommendations and email campaigns to highlight items popular with young adults.

Expected Impact:

Higher engagement from the segment that already shows the strongest purchase intent → faster conversion and higher average order value.

- c. **Expand High-Performing Categories (Clothing & Accessories)** - Clothing and accessories are the top-performing categories, contributing the most revenue and units sold.

Recommendation:

- Introduce new styles, seasonal collections, or bundles in these categories.
- Run category-specific discounts to drive repeat purchases.
- Feature top-rated products on the homepage and newsletters.

Expected Impact:

Leveraging top categories can bring 20–30% more incremental sales with minimal marketing cost.

- d. **Optimize Payment & Checkout Options** - Sales are distributed across many payment methods — with high use of PayPal, credit/debit cards, and bank transfer.

Recommendation:

- Simplify the checkout flow for popular methods (like PayPal or UPI).
- Offer one-click checkout or “Buy Now, Pay Later” options.
- Incentivize faster payments (e.g., 5% off on prepaid orders).

Expected Impact:

Reduced cart abandonment and smoother checkout → improved conversion rate and customer satisfaction.

- e. **Promote Fast & Free Shipping Options** - “2-Day Shipping” and “Express Shipping” are preferred by a significant number of customers.

Recommendation:

- Offer free express delivery for orders above a threshold (₹1000+).
- Highlight estimated delivery times on the product page.
- Partner with logistics providers to reduce shipping cost per order.

Expected Impact:

Faster delivery increases repeat purchases and trust — especially important for first-time customers.

Author

Sakshi Awasthi

9026362501