□ Retail Customer Behavior Analysis Project Report

📄 Project Overview

This project focuses on analyzing **customer behavior and shopping trends** for a retail company using **Python, SQL, and Power BI**.
The goal is to help the business understand how customers purchase products, use discounts, respond to subscriptions, and interact with different shipping types — and identify key drivers of revenue.

🎯 Business Problem Statement

The retail company wants to improve its marketing and sales strategy by understanding:

- Which customer groups generate the most revenue
- The impact of discounts and subscriptions on total sales
- Which product categories and items perform best
- How different shipping methods affect purchasing behavior

Tech Stack & Tools

Tool	Purpose
Python (pandas, numpy)	Data loading, cleaning, preprocessing
SQL (MySQL)	Business queries and data insights
Power BI	Visualization and dashboard creation
Excel/CSV	Raw dataset storage and preview

Step 1: Data Preprocessing in Python

Performed in the `preprocessing.ipynb` notebook.

Key tasks:

- Loaded raw dataset using pandas
- Handled missing and duplicate values
- Normalized categorical data (gender, subscription, shipping_type, etc.)
- Converted numeric columns (purchase_amount, review_rating) to appropriate data types
- Exported cleaned data to MySQL database (`customer_behavior`)

Step 2: Business Analysis with SQL

Q1. Total revenue by gender

```
SELECT gender, SUM(purchase_amount)
FROM people
GROUP BY gender;
```

➞ *Result:* Males generated slightly more total revenue than females, indicating a balanced but male-leaning customer base.

Q2. Discount users who spent above average

```
SELECT customer_id, purchase_amount
FROM people
WHERE discount_applied='Yes'
AND purchase_amount > (SELECT AVG(purchase_amount) FROM people);
```

➞ *Insight:* Some discount users are high spenders — showing that offering discounts doesn't always reduce overall revenue.

Q3. Top 5 Products by Average Review Rating

```
SELECT item_purchased, ROUND(AVG(review_rating),2) AS average_rating
FROM people
GROUP BY item_purchased
ORDER BY average_rating DESC
LIMIT 5;
```

➞ *Insight:* These products maintain high customer satisfaction, suggesting strong product quality and loyalty.

Q4. Average Purchase by Shipping Type

```
SELECT shipping_type, ROUND(AVG(purchase_amount),2)
FROM people
WHERE shipping_type IN ('Standard', 'Express')
GROUP BY shipping_type;
```

➞ *Insight:* Express shipping users tend to spend slightly more — indicating that fast delivery may attract premium buyers.

Q5. Subscription Impact on Spending

```
SELECT subscription_status,
       COUNT(customer_id) AS total_customer,
       ROUND(AVG(purchase_amount),2) AS average_spend,
       SUM(purchase_amount) AS total_revenue
FROM people
GROUP BY subscription_status;
```

➞ *Insight:* Subscribed customers spend more on average and contribute higher total revenue, proving the importance of subscription-based models.

Q6. Top 5 Products with Highest Discount Usage

```
SELECT item_purchased,
ROUND(SUM(CASE WHEN discount_applied='Yes' THEN 1 ELSE 0
END)*100/COUNT(*),2) AS purchased_count
FROM people
GROUP BY item_purchased
ORDER BY purchased_count DESC
LIMIT 5;
```

➞ *Insight:* These items are discount-sensitive, useful for targeted promotional campaigns.

Q7. Customer Segmentation by Purchase Frequency

```
SELECT
CASE
 WHEN previous_purchases BETWEEN 0 AND 1 THEN 'New'
 WHEN previous_purchases BETWEEN 2 AND 20 THEN 'Returning'
 ELSE 'Loyal'
END AS Customer_Segment,
COUNT(*) AS Segment_Count
FROM people
GROUP BY Customer_Segment
ORDER BY Segment_Count;
```

➞ *Insight:* Most customers fall into the *Returning* category, indicating moderate brand retention.

Q8. Top 3 Products within Each Category

```
SELECT category, item_purchased, total_count
FROM (
  SELECT category, item_purchased, COUNT(*) AS total_count,
  ROW_NUMBER() OVER (PARTITION BY category ORDER BY COUNT(*) DESC) AS rankk
  FROM people
  GROUP BY category, item_purchased
```

```
) AS temp
WHERE rankk <= 3;
```

➞ *Insight:* The query helps identify best-sellers per category for optimized stock management.

Q9. Repeat Buyers and Subscription Relationship

```
SELECT subscription_status, COUNT(customer_id) AS repeat_buyers
FROM people
WHERE previous_purchases > 5
GROUP BY subscription_status;
```

➞ *Insight:* Majority of repeat buyers are subscribers, highlighting that subscriptions enhance customer retention.

Revenue by Age Group

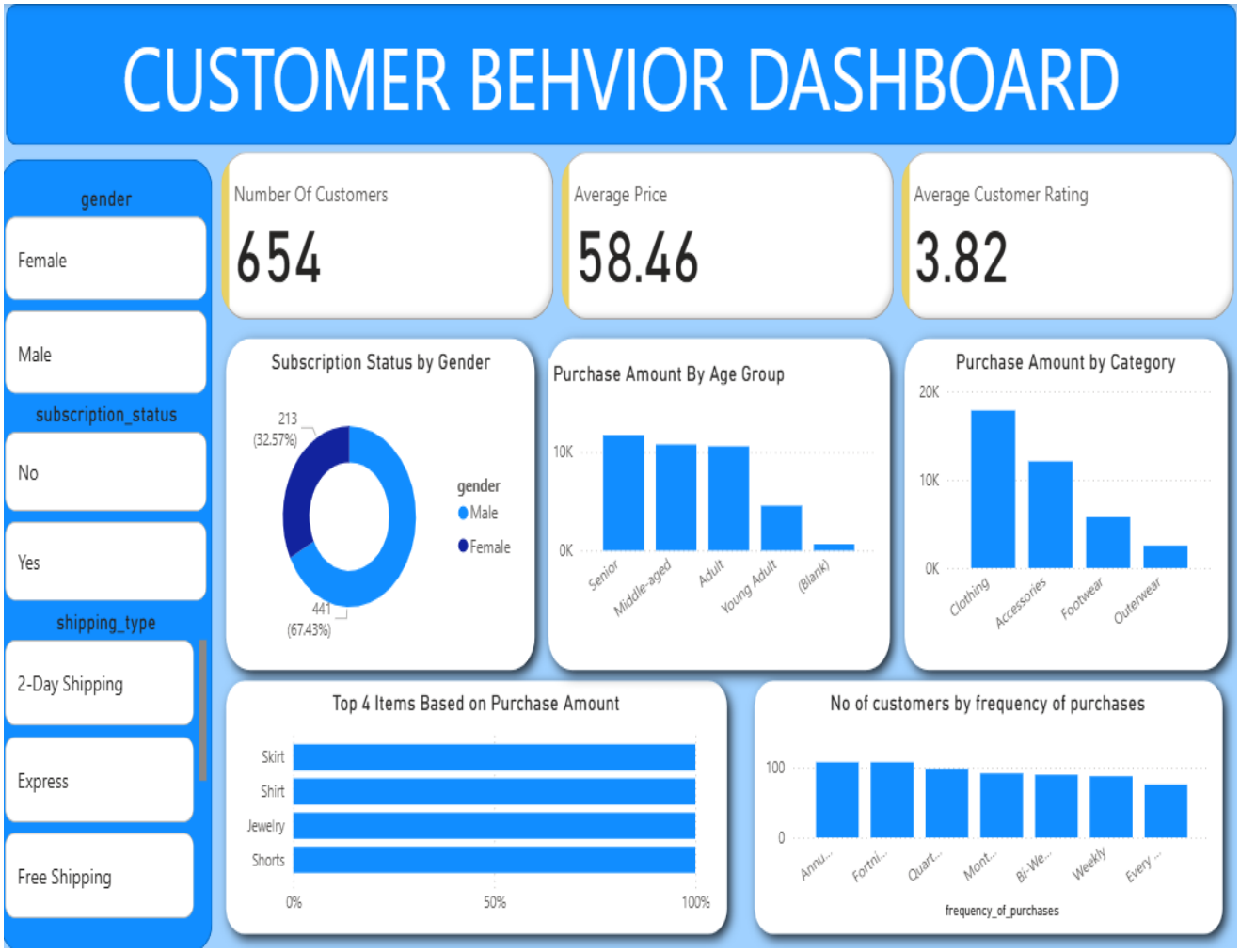
```
SELECT age_group, SUM(purchase_amount) AS total_revenue
FROM people
GROUP BY age_group
ORDER BY total_revenue DESC;
```

➞ *Insight:* Seniors and middle-aged groups contribute most to total revenue.

📊Step 3: Power BI Dashboard Insights

The Power BI dashboard (attached image) presents:

- **Total Customers:** 654
- **Average Price:** ₹58.46
- **Average Rating:** 3.82
- **Top Categories:** Clothing and Accessories dominate revenue
- **Top Products:** Skirt, Shirt, Jewelry, and Shorts
- **Age Groups:** Seniors and middle-aged buyers are the most valuable
- **Subscription Distribution:** 67% subscribed customers
- **Shipping Type Analysis:** Express shipping has higher spenders


CUSTOMER BEHVIOR DASHBOARD
gender
Female
Male
subscription_status
No
Yes
shipping_type
2-Day Shipping
Express
Free Shipping
Number Of Customers
654
Average Price
58.46
Average Customer Rating
3.82
Subscription Status by Gender
213
(32.57%)
441
(67.43%)
gender
Male
Female
Purchase Amount By Age Group
10K
0K
Senior
Middle-aged
Adult
Young Adult
(Blank)
Purchase Amount by Category
20K
10K
0K
Clothing
Accessories
Footwear
Outerwear
Top 4 Items Based on Purchase Amount
Skirt
Shirt
Jewelry
Shorts
0%
50%
100%
No of customers by frequency of purchases
100
0
Annu...
Fortni...
Quart...
Mont...
Bi-We...
Weekly
Every ...
frequency_of_purchases

📈 Key Business Insights

1. **Subscribed customers** drive more revenue and higher average purchases.
2. **Clothing** is the top revenue-generating category.
3. **Senior and middle-aged** customers contribute the largest revenue share.
4. **Discounts** don't necessarily lower total spending — some high-value customers use them.
5. **Express shipping** correlates with higher spending behavior.
6. **Returning customers** dominate, suggesting moderate brand loyalty.

💼 Final Business Recommendations

✅ Promote subscriptions — they significantly increase spending.
✅ Prioritize high-rating products in marketing.
✅ Offer personalized discounts for repeat buyers.
✅ Target senior and middle-aged demographics in ad campaigns.
✅ Keep Express shipping affordable to maintain high-value customer satisfaction.

🧠 Conclusion

This project demonstrates the **end-to-end data analyst workflow**:

- Data preprocessing in Python
- Business analysis via SQL
- Insight visualization using Power BI

It mirrors how analysts in real companies translate raw data into **strategic insights** that drive business growth.