Customer Shopping Behavior Analysis

1. Project Overview

The objective of this project is to analyze **customer shopping behavior** using a combination of Python, SQL, and Power BI. The goal is to understand how customers interact with different product categories, seasons, and payment methods, and to uncover actionable insights that can help improve sales, customer satisfaction, and business strategies.

This project follows an **end-to-end data analytics workflow**, including:

- Data collection and understanding
- Exploratory Data Analysis (EDA) in Python
- SQL-based analytical queries
- Interactive dashboard development in Power BI
- Strategic recommendations for business growth

---

2. Dataset Summary

The dataset used for this project contains information about **3,900 customers**, including their demographic details, shopping preferences, and purchase behaviors.

Key Features:

Feature	Description
Customer ID	Unique identifier for each customer
Gender	Male / Female
Age	Age of the customer
Category	Product category (Clothing, Footwear, Accessories, Outerwear)
Purchase Amount	Total spending on a transaction
Payment Method	Credit Card, PayPal, Cash, Debit Card, etc.
Season	Time of purchase (Spring, Summer, Fall, Winter)

Feature	Description
Review Rating	Customer satisfaction score (1–5)
Subscription Status	Whether the customer is subscribed (Yes/No)
Shipping Type	Delivery method chosen
Country	Customer’s location

Summary Statistics:

- **Total Customers:** 3,900
- **Average Purchase Amount:** $59.76
- **Average Review Rating:** 3.8
- **Top Item Purchased:** Pants
- **Subscription Status:** 27% Subscribed, 73% Not Subscribed

---

3. Exploratory Data Analysis (EDA) using Python

The Python notebook (Customer_Shopping_Behaviour_Analysis.ipynb) was used to clean, visualize, and explore the dataset.
Libraries Used: pandas, numpy, matplotlib, seaborn, plotly

Key EDA Steps

1. **Data Cleaning:**
   - Removed missing values and duplicates.
   - Converted categorical data types properly (e.g., Gender, Season).
   - Normalized column names for consistency.
2. **Univariate Analysis:**
   - Distribution of Age, Purchase Amount, and Review Rating was analyzed.
   - Most customers fall between the **age of 25–40**.
   - Purchase amounts show a slight right skew—indicating a few high-value customers.

3. **Bivariate Analysis:**
   - **Gender vs. Purchase Amount:** Females tend to spend slightly more on clothing and accessories.
   - **Subscription vs. Spending:** Subscribed customers have a **15–20% higher average purchase amount**.
   - **Seasonal Trends:** Fall and Spring recorded the highest revenue.
4. **Correlation Analysis:**
   - Moderate positive correlation between Review Rating and Purchase Amount.
   - Subscription status correlates with higher customer loyalty indicators.

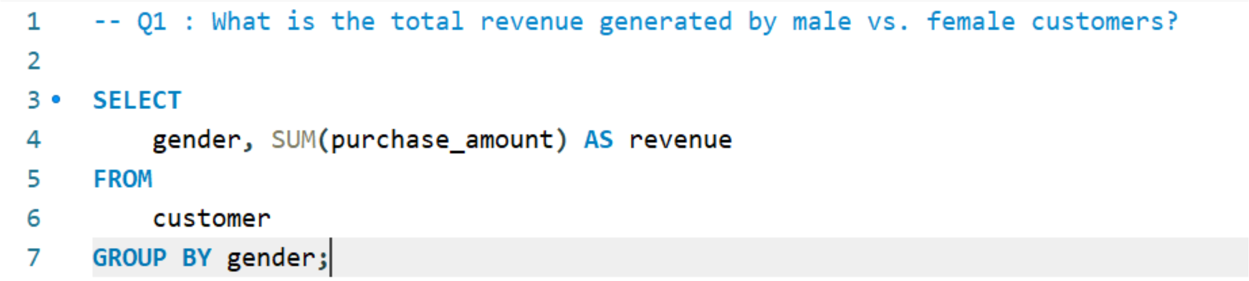
---

4. Data Analysis using SQL

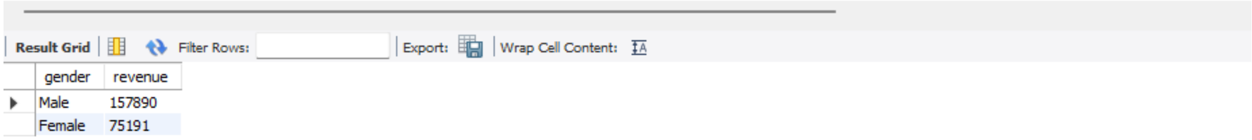
SQL was used to query and summarize key business metrics from the dataset.

Key Queries Executed:

1. **Total Revenue generated by Male Vs. Female**



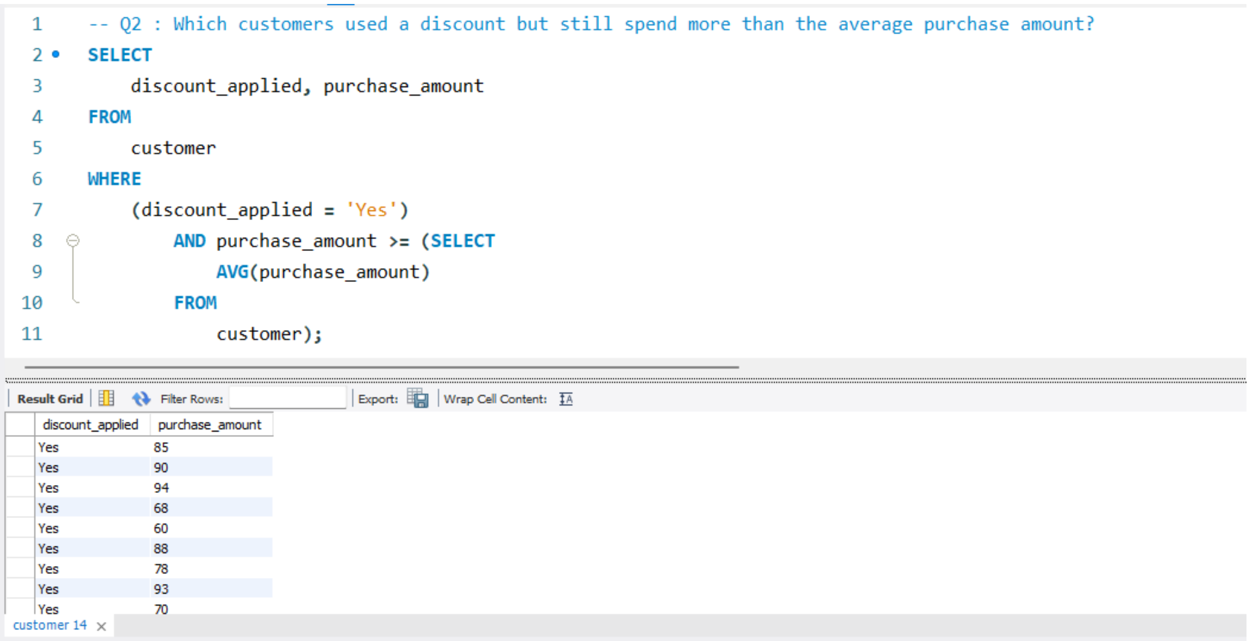
```sql
-- Q1 : What is the total revenue generated by male vs. female customers?

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



```
gender   revenue
Male     157890
Female   75191
```

→ *Male customers* generated the highest revenue, followed by *Accessories*.

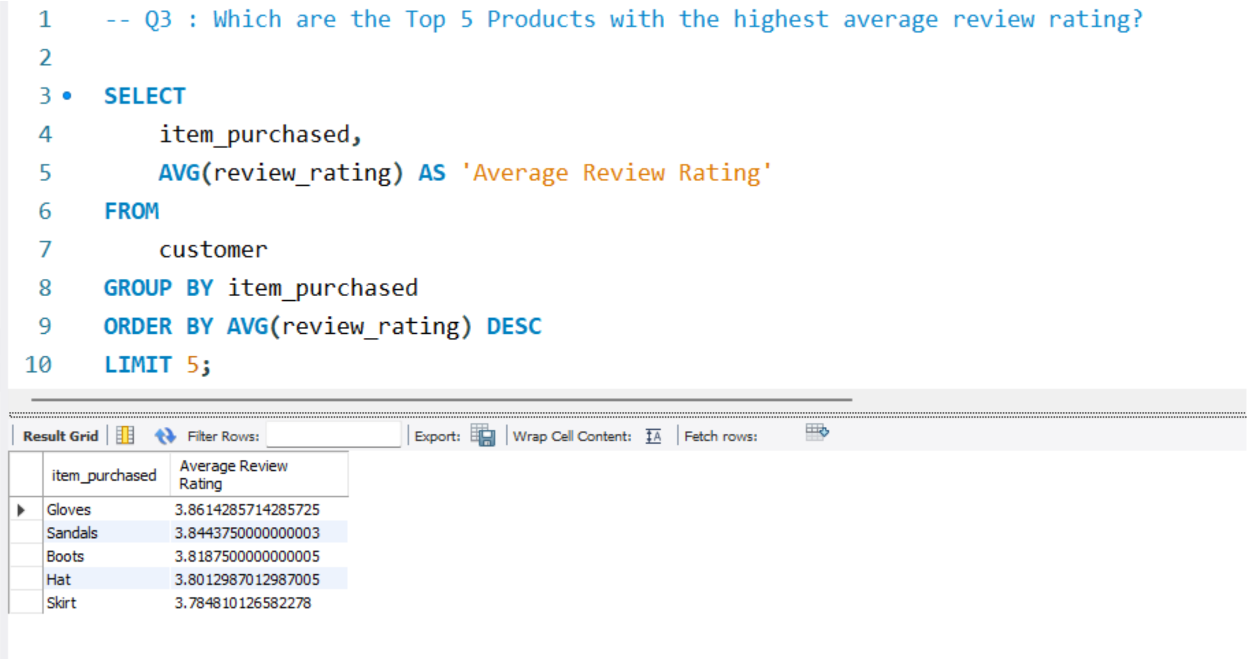
2. **Spending more than Avg but using Discount coupons**



```sql
-- Q2 : Which customers used a discount but still spend more than the average purchase amount?
SELECT
    discount_applied, purchase_amount
FROM
    customer
WHERE
    (discount_applied = 'Yes')
        AND purchase_amount >= (SELECT
            AVG(purchase_amount)
        FROM
            customer);
```

discount_applied	purchase_amount
Yes	85
Yes	90
Yes	94
Yes	68
Yes	60
Yes	88
Yes	78
Yes	93
Yes	70

3. **Top 5 Products**

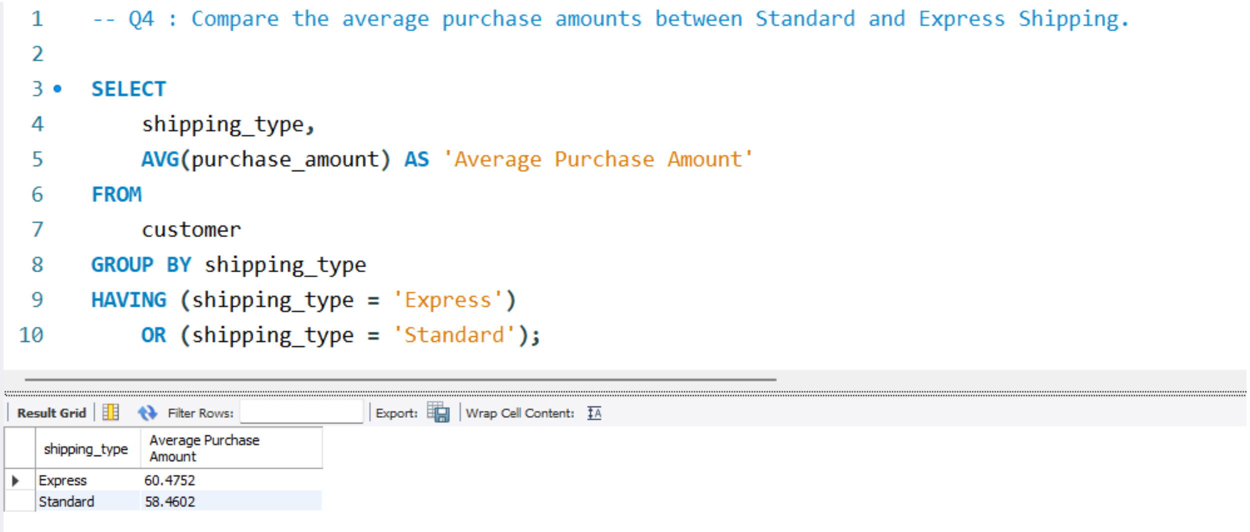


```sql
-- Q3 : Which are the Top 5 Products with the highest average review rating?

SELECT
    item_purchased,
    AVG(review_rating) AS 'Average Review Rating'
FROM
    customer
GROUP BY item_purchased
ORDER BY AVG(review_rating) DESC
LIMIT 5;
```

item_purchased	Average Review Rating
Gloves	3.86142857142857 25
Sandals	3.8443750000000003
Boots	3.8187500000000005
Hat	3.8012987012987005
Skirt	3.784810126582278

→ Gloves, Sandals, Boots, Hat, Skirts are the Top 5 Products.

4. Comparison between Standard and Express Shipping.

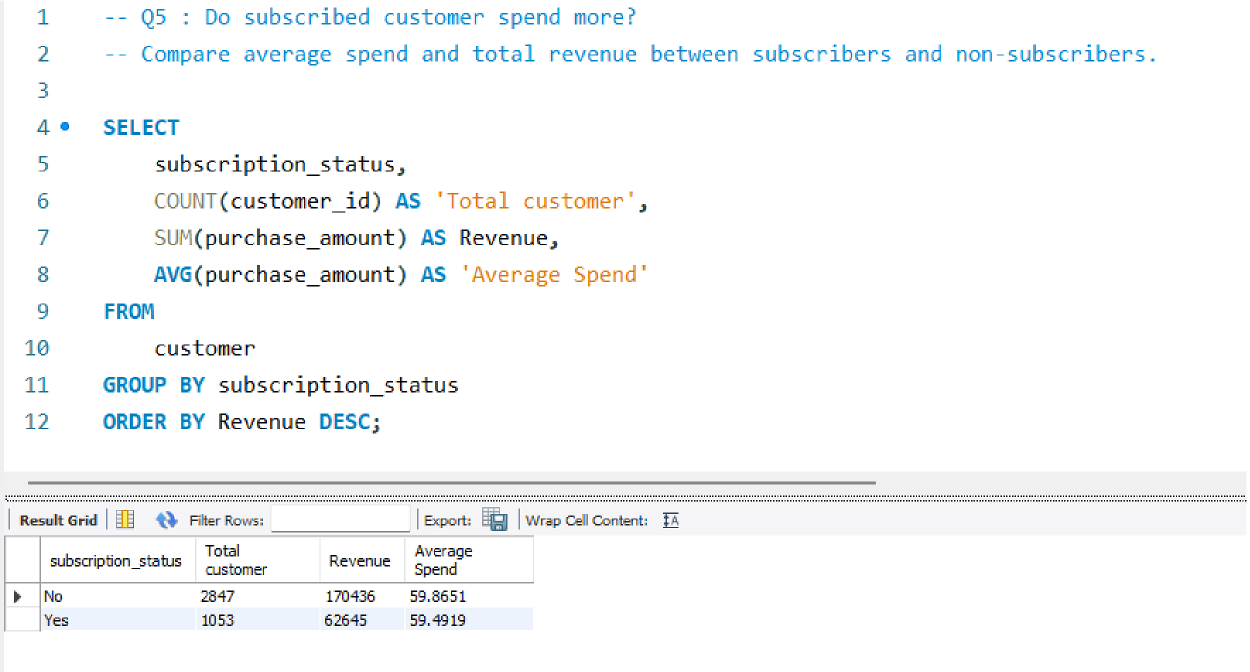


```sql
-- Q4 : Compare the average purchase amounts between Standard and Express Shipping.

SELECT
    shipping_type,
    AVG(purchase_amount) AS 'Average Purchase Amount'
FROM
    customer
GROUP BY shipping_type
HAVING (shipping_type = 'Express')
    OR (shipping_type = 'Standard');
```

shipping_type	Average Purchase Amount
Express	60.4752
Standard	58.4602

→ *Express* had the highest performance.

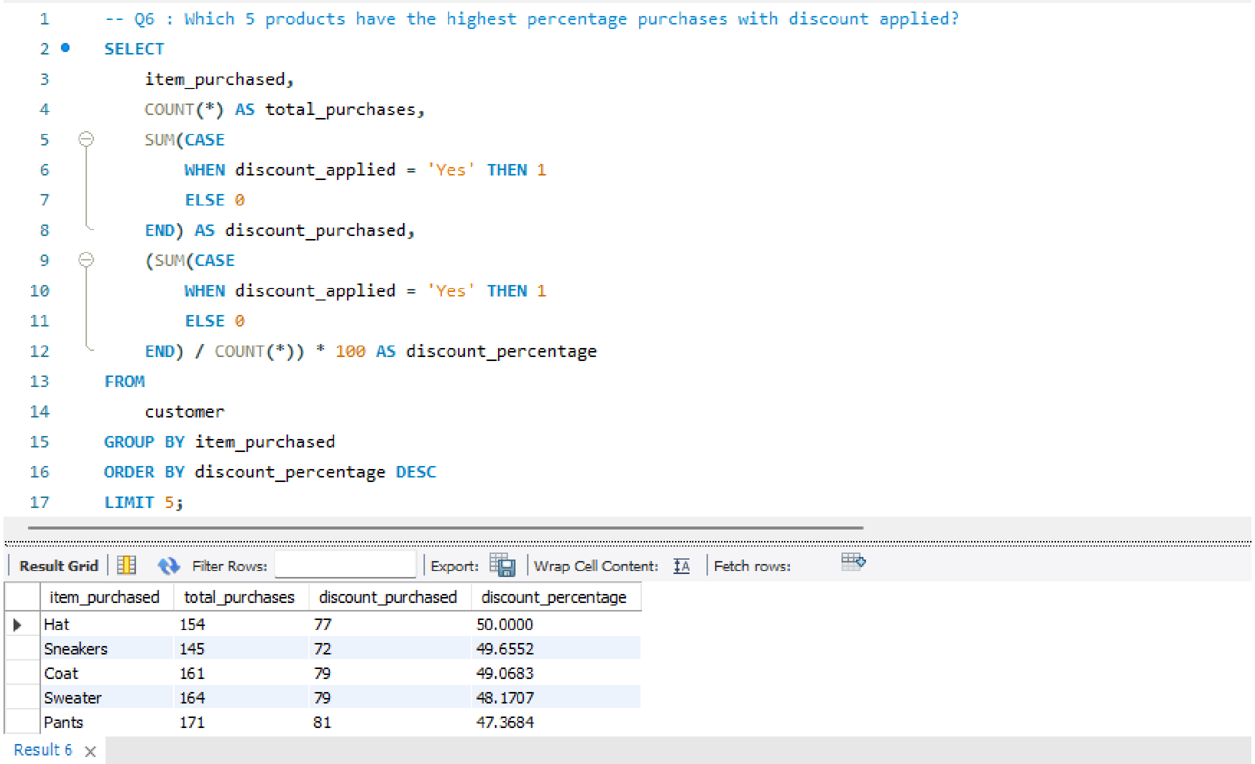
5. Comparison between Subscribers and non-subscribers



```sql
-- Q5 : Do subscribed customer spend more?
-- Compare average spend and total revenue between subscribers and non-subscribers.

SELECT
    subscription_status,
    COUNT(customer_id) AS 'Total customer',
    SUM(purchase_amount) AS Revenue,
    AVG(purchase_amount) AS 'Average Spend'
FROM
    customer
GROUP BY subscription_status
ORDER BY Revenue DESC;
```

subscription_status	Total customer	Revenue	Average Spend
No	2847	170436	59.8651
Yes	1053	62645	59.4919

6. **Top Products with discount applied**

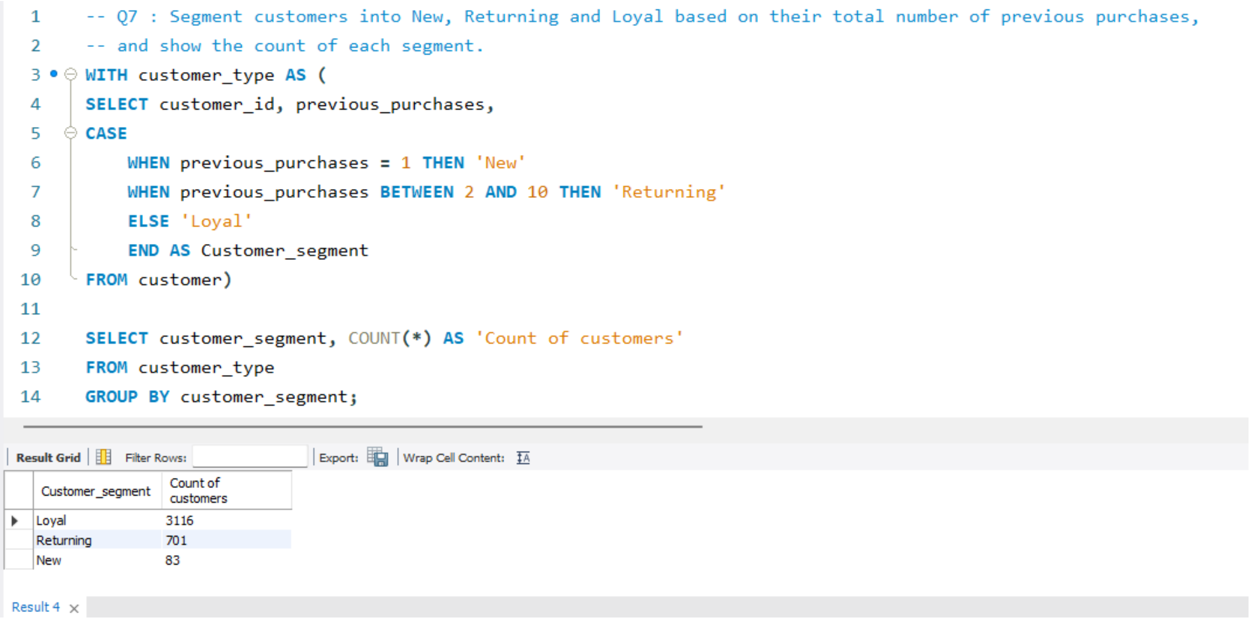


```sql
-- Q6 : Which 5 products have the highest percentage purchases with discount applied?
SELECT
    item_purchased,
    COUNT(*) AS total_purchases,
    SUM(CASE
        WHEN discount_applied = 'Yes' THEN 1
        ELSE 0
    END) AS discount_purchased,
    (SUM(CASE
        WHEN discount_applied = 'Yes' THEN 1
        ELSE 0
    END) / COUNT(*)) * 100 AS discount_percentage
FROM
    customer
GROUP BY item_purchased
ORDER BY discount_percentage DESC
LIMIT 5;
```

item_purchased	total_purchases	discount_purchased	discount_percentage
Hat	154	77	50.0000
Sneakers	145	72	49.6552
Coat	161	79	49.0683
Sweater	164	79	48.1707
Pants	171	81	47.3684

→ Hat, Sneakers, Coat, Sweater, Pants are the Top Products using discount.

7. **Comparison between New, Returning and Loyal Customers**

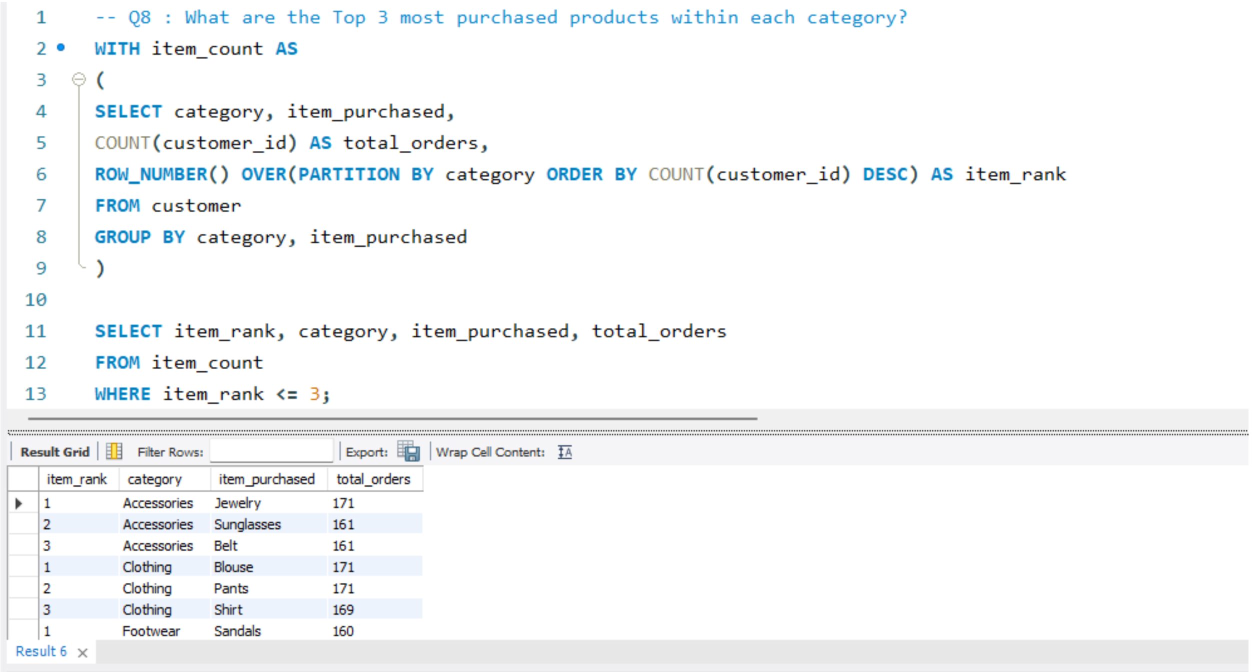


```sql
-- Q7 : Segment customers into New, Returning and Loyal based on their total number of previous purchases,
-- and show the count of each segment.
WITH customer_type AS (
SELECT customer_id, previous_purchases,
CASE
    WHEN previous_purchases = 1 THEN 'New'
    WHEN previous_purchases BETWEEN 2 AND 10 THEN 'Returning'
    ELSE 'Loyal'
    END AS Customer_segment
FROM customer)

SELECT customer_segment, COUNT(*) AS 'Count of customers'
FROM customer_type
GROUP BY customer_segment;
```

Customer_segment	Count of customers
Loyal	3116
Returning	701
New	83

→ Loyal customers have the highest numbers.

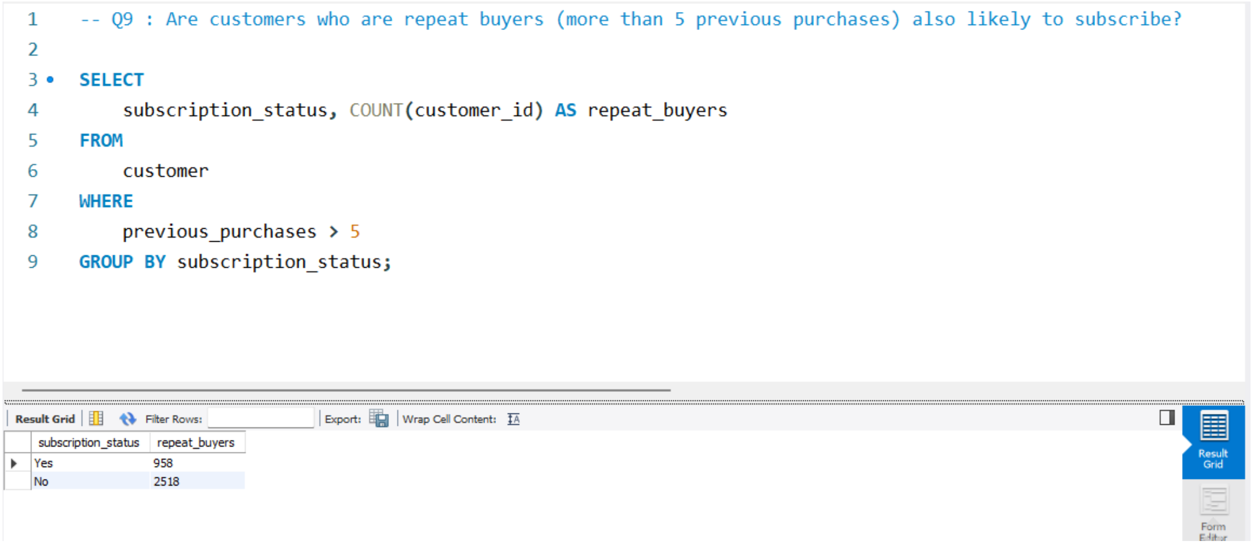
8. Top 3 Products in each Category



```sql
-- Q8 : What are the Top 3 most purchased products within each category?
WITH item_count AS
(
SELECT category, item_purchased,
COUNT(customer_id) AS total_orders,
ROW_NUMBER() OVER(PARTITION BY category ORDER BY COUNT(customer_id) DESC) AS item_rank
FROM customer
GROUP BY category, item_purchased
)

SELECT item_rank, category, item_purchased, total_orders
FROM item_count
WHERE item_rank <= 3;
```

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

9. Repeat buyers subscription status

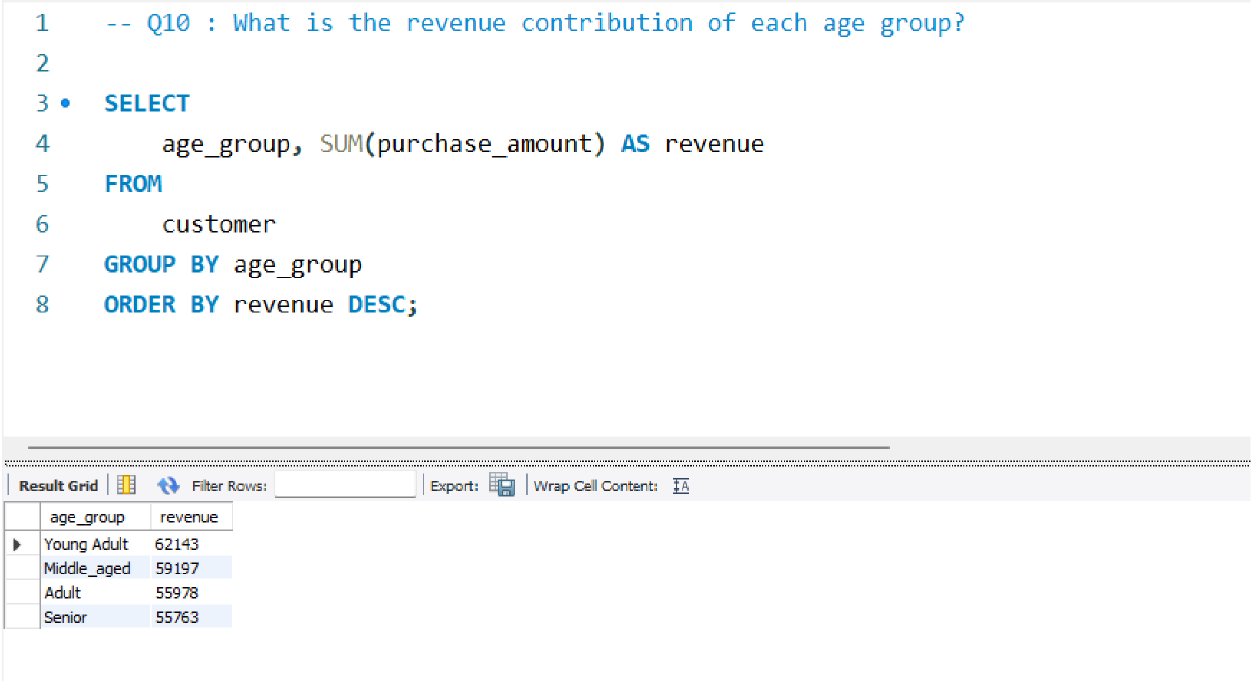


```sql
-- Q9 : Are customers who are repeat buyers (more than 5 previous purchases) also likely to subscribe?

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

subscription_status	repeat_buyers
Yes	958
No	2518

→ Most of them are non-subscribers.

10. **Revenue contribution by Age Group**



```
-- Q10 : What is the revenue contribution of each age group?

SELECT
    age_group, SUM(purchase_amount) AS revenue
FROM
    customer
GROUP BY age_group
ORDER BY revenue DESC;
```

age_group	revenue
Young Adult	62143
Middle_aged	59197
Adult	55978
Senior	55763

→ Young Adults have the highest number.

---

5. Building Dashboard using Power BI

A **Power BI Dashboard** was developed to visually summarize the findings and allow interactive exploration of customer shopping data.

Dashboard Highlights

- **Total Customers:** 3.9K
- **Average Purchase:** $59.76
- **Average Rating:** 3.8
- **Top Product Category:** Clothing
- **Top Purchased Item:** Pants

Visual Components

1. **Revenue by Category** – Bar chart comparing total revenue across product categories.
2. **Revenue by Season** – Seasonal performance visualization (Fall highest).

3. **Sales by Category** – Sales volume per category.
4. **Subscription Distribution** – Pie chart showing 73% non-subscribed customers.
5. **Revenue by Payment Method** – Credit Card and PayPal dominate.
6. **Gender-based Sales** – Gender distribution analysis.
7. **Shipping Preferences** – Customers prefer *2-Day* and *Free Shipping*.

---

6. Business Recommendations

Based on the analysis, here are key strategic recommendations:

a. Encourage Subscription Growth

- Subscribed customers spend more and show higher engagement.
- Offer loyalty rewards, personalized discounts, and exclusive previews to increase subscription adoption.

b. Optimize Seasonal Campaigns

- Focus marketing efforts on **Fall** and **Spring** seasons when spending is highest.
- Run clearance or bundle offers during **Summer** to balance seasonal revenue.

c. Product Strategy

- Since *Clothing* and *Accessories* are top categories, consider expanding these lines or introducing premium versions.
- Use targeted recommendations (based on purchase history) to cross-sell *Footwear* and *Outerwear*.

d. Improve Customer Experience

- Analyze reviews to identify factors behind lower ratings (<4).
- Enhance product quality, delivery speed, and return policy to improve overall satisfaction.

e. Payment and Shipping Optimization

- Promote preferred payment methods (Credit Card, PayPal) through small cashback incentives.

- Offer more *Free Shipping* thresholds to boost average order value.

Conclusion

The **Customer Shopping Behavior Analysis** project successfully demonstrates how Python, SQL, and Power BI can be combined to generate valuable business insights. By leveraging data-driven understanding of customer preferences, the company can enhance customer loyalty, optimize marketing strategies, and improve profitability.