Customer Behaviour Analysis on E-commerce Dataset

Joshua Amade

Dataset Source: E-commerce Customer Behaviour Dataset (https://www.kaggle.com/datasets/uom190346a/e-commerce-customer-behavior-dataset)

Key Questions:

- Total Customers: How many unique customers are in our dataset?
- Gender Distribution: What is the distribution of customers across genders?
- Customer Concentration: Which cities have the highest concentration of customers, and what proportion of our customer base do they represent?
- Average Spending by Gender: How does average spending compare between male and female customers?
- Average Spending by Membership Tier: How does average spending differ across the Gold, Silver, and Bronze membership tiers?
- Top Cities by Gender: Do male and female customers show preferences for different major cities?
- Repeat Purchase Rates: What are the repeat purchase rates for each membership tier, and which tier has the highest retention?
- Spending Preferences: How do male and female customers distribute across 'High Spend' and 'Low Spend' categories?

Notes:

- Dataset: The dataset contains detailed information on customer demographics, spending patterns, satisfaction levels, and product preferences.
- Exploration: The initial exploratory phase will focus on understanding overall customer distribution and identifying key metrics.

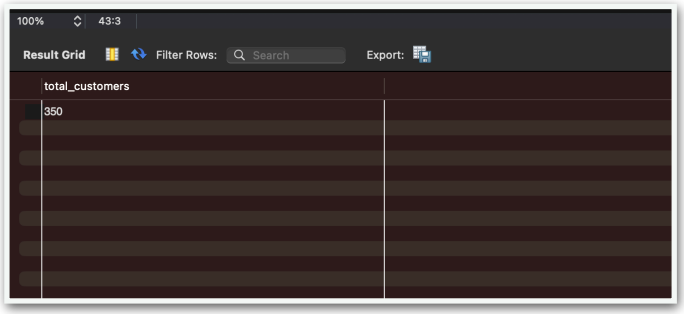
SQL Queries:

```
SELECT COUNT(DISTINCT Customer_ID) AS total_customers
FROM customer_analysis.customer_behaviour;


SELECT Gender, COUNT(*) AS count
FROM customer_analysis.customer_behaviour
GROUP BY Gender;


SELECT City, COUNT(*) AS count
FROM customer_analysis.customer_behaviour
GROUP BY City
ORDER BY count DESC
LIMIT 10;

SELECT Gender, AVG(Total_Spend) AS avg_spend
FROM customer_analysis.customer_behaviour
GROUP BY Gender;

SELECT Membership_Type, AVG(Total_Spend) AS avg_spend
FROM customer_analysis.customer_behaviour
GROUP BY Membership_Type;

SELECT Gender, City, COUNT(*) AS customer_count
FROM customer_analysis.customer_behaviour
GROUP BY Gender, City
ORDER BY customer_count DESC
LIMIT 10;

SELECT Membership_Type,
```

```
    COUNT(DISTINCT Customer_ID) AS total_customers,
    COUNT(DISTINCT CASE WHEN Days_Since_Last_Purchase <= 30 THEN Customer_ID END) AS
repeat_customers,
    (COUNT(DISTINCT CASE WHEN Days_Since_Last_Purchase <= 30 THEN Customer_ID END) /
COUNT(DISTINCT Customer_ID)) * 100 AS repeat_purchase_rate
FROM customer_analysis.customer_behaviour
GROUP BY Membership_Type;

WITH avg_spend_data AS (
  SELECT AVG(Total_Spend) AS avg_total_spend
  FROM customer_analysis.customer_behaviour
)
SELECT Gender,
    CASE
      WHEN Total_Spend >= (SELECT avg_total_spend FROM avg_spend_data) THEN 'High Spend'
      ELSE 'Low Spend'
    END AS spending_category,
    COUNT(*) as item_count
FROM customer_analysis.customer_behaviour
GROUP BY Gender, spending_category
ORDER BY item_count DESC;
```

Results:

Query 1



Query 2

Query 3

Query 4

Gender	avg_spend
Female	703.8285714285722
Male	986.9348571428568

Query 5

Membership_Type	avg_spend
Gold	1311.1444444444455
Silver	748.4324786324781
Bronze	473.38879310344856

Query 6

Gender	City	customer_cou...
Male	Los Angeles	59
Female	New York	58
Female	Chicago	58
Male	San Francisco	58
Female	Houston	58
Male	Miami	57
Female	Miami	1
Male	New York	1

Query 7

Membership_Type	total_customers	repeat_custome...	repeat_purchase_r...
Bronze	116	56	48.2759
Gold	117	111	94.8718
Silver	117	59	50.4274

Query 8

Gender	spending_categ...	item_count
Female	Low Spend	117
Male	Low Spend	116
Male	High Spend	59
Female	High Spend	58

Insights:

- Total Customers: The dataset contains 350 unique customers.
- Gender Distribution: The dataset exhibits a perfectly balanced distribution between female and male customers (50/50 split).
- City Concentration: There is a significant concentration of customers within a few major cities (New York, Los Angeles, San Francisco, Chicago, Houston), representing approximately 33.4% of the total customer base.
- Spending Habits by Gender: Female customers demonstrated a notably lower average spend ($703.83) compared to male customers ($986.93)
- Spending by Membership Tier: Gold tier customers exhibit significantly higher average spending ($1311.14) compared to Silver ($748.43) and Bronze ($473.39) tiers.
- City Preferences by Gender: The most popular cities show differences by gender. Top cities for male customers include San Francisco and Miami, while female customers are concentrated in New York and Chicago.
- Repeat Purchase Rate: All membership tiers demonstrate a repeat purchase rate above 48%, which indicates good customer loyalty in general. The Gold tier exhibits the highest repeat purchase rate (94.87%), suggesting that the benefits of this tier might strongly encourage repeat business, While Silver and Bronze tiers show decent retention, there's still a potential to nudge those repeat purchase rates closer to the Gold tier.
- Spending Preferences: Both female and male customers exhibit a nearly even split between high spending and low spending purchases. This suggests that gender alone might not be the most significant predictor of big-ticket purchases.

Next Steps

Customer Acquisition
- Beyond the Top Cities: How can we expand our customer base and reduce reliance on the current top cities? What marketing strategies would be effective in less concentrated regions?

Gender-Based Spending Gap
- Understanding the Difference: What factors contribute to the lower average spending by female customers? Are there differences in product preferences or price sensitivity?
- Targeted Offerings: How can we design product selections, promotions, or marketing messages tailored to female customers to encourage higher spending?

Membership Tier Optimisation
- Unlocking Gold Tier Success: What specific benefits or incentives within the Gold tier drive its high retention and spending? Can successful elements be adapted for other tiers?
- Boosting Silver & Bronze: How can we make the Silver and Bronze tiers more attractive to encourage upgrades? Should we adjust benefits, pricing, or targeted promotions?

General Optimisation
- High-Spending Drivers: Beyond gender and membership tier, what other customer characteristics correlate with high spending? Can we build a "high-spender" profile for targeted marketing?
- Repeat Purchase Incentives: Are there loyalty programs, personalised recommendations, or other tactics that might further boost repeat purchase rates across all tiers?

Additional Considerations
- Product-Level Data: If possible, obtaining product category data would enable a deeper analysis of the differences in preferences between men and women and the popularity of certain categories across cities.
- Seasonality: Could you extract order dates to examine if purchasing patterns change throughout the year or around specific holidays?