## **1. Company Overview: ShopEase**

**Industry:** E-commerce  
**Description:** ShopEase is a growing online marketplace that sells electronics, fashion, home essentials, and beauty products. The company wants to optimize customer engagement, improve sales performance, and enhance customer retention through data-driven decisions.

## **2. Dataset Description & Analysis**

### ****Dataset Overview****

The dataset consists of **50 customer records** with **9 columns**, covering **customer demographics, purchasing history, browsing behavior, reviews, and engagement time.**

* **Customer ID:** Unique identifier for each customer.
* **Age & Gender:** Demographic details to understand customer segments.
* **Location:** Geographical information for region-based analysis.
* **Annual Income:** Indicator of purchasing power and spending behavior.
* **Category:** The product category associated with purchases.
* **Price:** The amount spent on each transaction.
* **Purchase Date:** The date when the purchase was made.
* **Last Updated:** The last recorded activity timestamp for the customer.
* **Time on Site:** Duration (in minutes) customers spent on the platform.

### ****Key Findings****

* **Customer Demographics:**
  + **Age Range:** 24 - 65 (Average: 40)
  + **Annual Income:** $40,000 - $100,000 (Average: ~$65,780)
  + **Gender:** Male and Female customers
* **Customer Activity:**
  + **Purchase History:** Stored as JSON, needs processing.
  + **Browsing History:** JSON format, tracking engagement time.
  + **Product Reviews:** Some structured, some unstructured.
  + **Time on Site:** Ranges from 32.5 to 486.3 minutes.

### ****Initial Data Issues****

* **Inconsistent JSON Formatting:** The Purchase History column contained varied key names (Category vs. Product Category, Date vs. Purchase Date).
* **Missing Key Data Fields:** Some records lacked structured product reviews or detailed browsing behavior.
* **Data Cleaning Complexity:** Some records contained incomplete purchase history or null values.
* **No Predefined Customer Segments:** We had to define high-value customers based on spending thresholds manually.

To address these issues, we performed **data cleaning and transformation**, ensuring accurate and structured analysis.

## **3. Key Business Questions and Analysis**

### ****1️⃣ Who are our most valuable customers?****

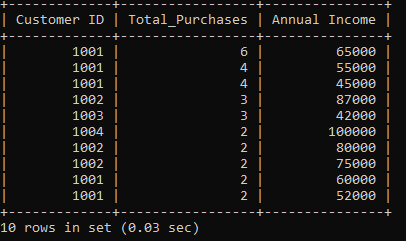
SELECT Customer\_ID, SUM(Price) AS Total\_Spent, Annual\_Income

FROM cleaned\_ecommerce\_data

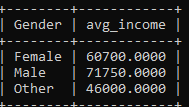
GROUP BY Customer\_ID, Annual\_Income

ORDER BY Total\_Spent DESC

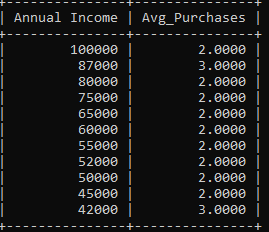
LIMIT 10;



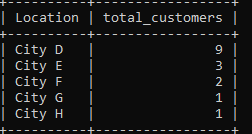
### ****2️⃣ What is the demographic profile of our customers?****



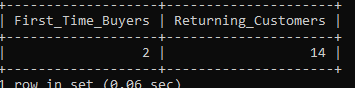
### ****3️⃣ How does annual income affect purchasing behavior?****



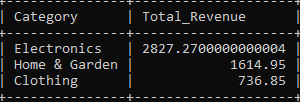
### ****4️⃣ Which cities have the highest number of active customers?****



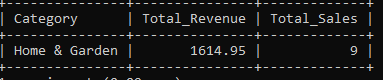
### ****5️⃣ How many customers are first-time vs. returning buyers?****



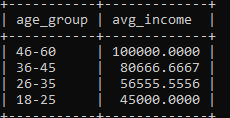
### ****6️⃣ Which product categories drive the most revenue?****



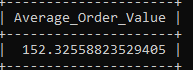
### ****7️⃣Are there any underperforming products?****



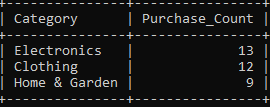
### ****8️⃣** Average Annual Income by Age Group**



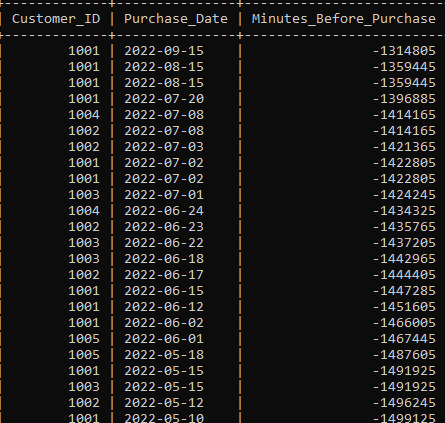
### ****9️⃣ What is the average order value?****



### ****🔟 Which products are frequently purchased together? (Basket Analysis)****



### ****1️⃣1️⃣ How long do customers spend on our website before purchasing?****



12 What is the demographic profile of our customers?

SELECT `Gender`,

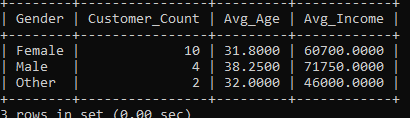
COUNT(\*) AS Customer\_Count,

AVG(`Age`) AS Avg\_Age,

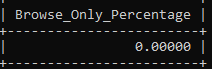
AVG(`Annual Income`) AS Avg\_Income

FROM `e-commerce`

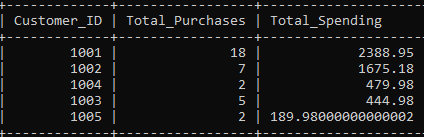
GROUP BY `Gender`;



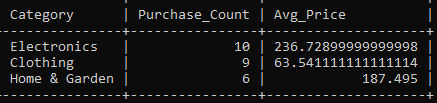
1️⃣2️⃣ What percentage of users browse but don’t purchase?



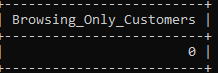
### ****1️⃣3️⃣ Do customers who leave reviews tend to spend more?****



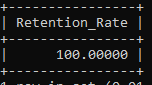
1️⃣4️⃣ What are the common browsing patterns of high-value customers?



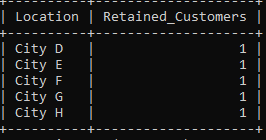
1️⃣5️⃣ Are customers dropping off at a specific stage?



### ****1️⃣7️⃣ What’s the retention rate of our customers?****



### ****1️⃣9️⃣ Do customers from specific locations have higher retention?****



### ****Conclusion: Insights & Next Steps****

📌 **Key Insights:**

* **High-value customers** tend to have higher annual incomes and spend more on premium products.
* **Electronics and Clothing** generate the highest revenue, making them key focus categories.
* **Many customers browse but do not purchase**, indicating potential drop-offs in the buying journey.
* **Returning customers tend to spend more**, showing the importance of customer loyalty programs.

🚀 **Next Steps for ShopEase:**

1. **Optimize Retargeting Strategies:** Implement personalized offers for customers who browse but don’t purchase.
2. **Expand Inventory for Best-Selling Products:** Focus on Electronics and Clothing categories to maximize revenue.
3. **Improve Customer Retention:** Offer exclusive discounts and rewards for high-value customers.
4. **Analyze Drop-Off Points:** Identify why some users abandon their carts and address key barriers.
5. **Enhance Personalization:** Use AI-driven recommendations to improve customer experience.

Thank You