Advanced Data Management, Engineering, and Preprocessing for AI

Assignment **1**
Professor: **Dr. Ruchika Arora**
Submitted By: **Rohit kumar** (Mtech_AI-**25901334**)

Customer Segmentation and Marketing Strategy Improvement using Retail Sales Data

---

Question

You work for a retail company that wants to improve its marketing strategies by targeting specific customer segments.

Objective

To analyze retail transaction data and identify key customer segments, popular product categories, and high-value customers in order to recommend marketing strategies for improving sales and customer engagement.

Dataset Description

The dataset contains retail transactions with the following fields:

- **Transaction ID** – Unique ID for each purchase.

- **Date** – Transaction date.

- **Customer ID** – Unique ID for each customer.

- **Gender** – Male / Female.

- **Age** – Age of customer.

- **Product Category** – Category of purchased product (Beauty, Clothing, Electronics).

- **Quantity** – Number of units purchased.

- **Price per Unit** – Price of one unit.

- **Total Amount** – Total transaction value.

Step 1: Load and Explore the Dataset

The dataset was loaded, initial exploration was performed to understand its structure, missing values were checked to ensure data cleanliness, and relevant fields were prepared for analysis, including grouping by age, gender, and product category to facilitate insights into spending patterns and customer behavior

Table: First 5 Transactions from Dataset

Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount
1	2023-11-24	CUST001	Male	34	Beauty	3	50	150
2	2023-02-27	CUST002	Female	26	Clothing	2	500	1000
3	2023-01-13	CUST003	Male	50	Electronics	1	30	30
4	2023-05-21	CUST004	Male	37	Clothing	1	500	500
5	2023-05-06	CUST005	Male	30	Beauty	2	50	100

Step 2: Customer Segmentation by Age

Age Group Segmentation

To better understand customer behavior, the data was divided into four age groups: **18–25**, **26–40**, **41–60**, and **60+**.

- Customers aged **18–25** spent a total of **₹84,550**.

- Customers aged **26–40** spent a total of **₹144,345**.

- Customers aged **41–60** spent a total of **₹193,880**.

- Customers aged **60+** spent a total of **₹33,225**.

Result: The **41–60 age group** shows the highest overall spending, making them the most valuable segment, followed by the **26–40 group**.

Step 3: Total Spending by Gender

The total spending was calculated separately for male and female customers:

- **Female customers** spent a total of **₹232,840**.

- **Male customers** spent a total of **₹223,160**.

Result: Female customers contribute slightly higher spending compared to male customers, indicating that marketing campaigns targeted towards women may yield better engagement.

# Step 4: Product Category Analysis

- **Most Popular Product Categories (by number of purchases):** Clothing leads with **351 purchases**, followed closely by Electronics (**342 purchases**) and Beauty (**307 purchases**).
- **Revenue by Product Category:** Electronics generates the highest revenue (**₹156,905**), followed by Clothing (**₹155,580**) and Beauty (**₹143,515**).

Result: While **Clothing** is purchased most frequently, **Electronics** brings in the highest revenue.

Step 5: High-Value Customers

High-value customers are those with spending over 1000. Examples include **CUST013 (1500), CUST015 (2000), and CUST016 (1500)**.

Result: These customers should be prioritized for **loyalty programs and personalized offers**.
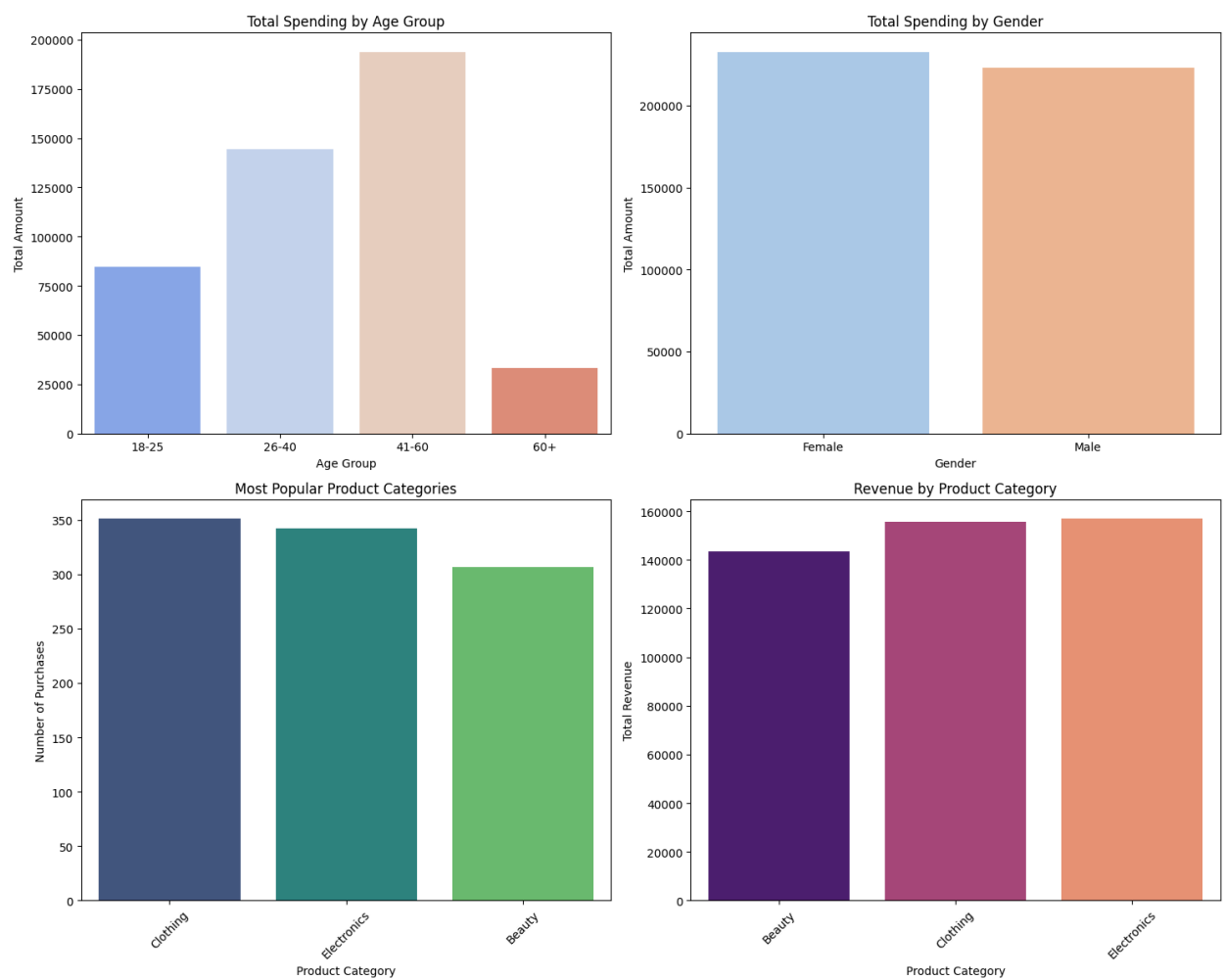
Analytical Skills Demonstrated

1. **Data Exploration and Cleaning**

   ○ Imported and reviewed dataset structure using Pandas.

   ○ Checked for missing values to ensure data quality before further analysis.

2. **Customer Segmentation**

   ○ Created age-based groups to analyze customer spending patterns across demographic segments.

   ○ Applied grouping and aggregation to evaluate spending by age and gender.

3. **Category and Revenue Analysis**

   ○ Identified most frequently purchased product categories using.

   ○ Calculated revenue contribution per category to highlight high-value product lines.

4. **High-Value Customer Identification**

   ○ Used customer-level aggregation to calculate total spending.

   ○ Defined thresholds to identify high-value customers for potential loyalty or VIP programs.

5. **Data Visualization**

   ○ Developed multi-panel charts using Matplotlib and Seaborn to visualize customer segmentation, gender-based spending, and product category trends.

   ○ Focused on clear, business-oriented visual representation of insights.

6. **Strategic Recommendations**

   ○ Translated analytical findings into practical marketing strategies, demonstrating the ability to connect technical analysis with business decisions.

[Code](#)

Customer Segmentation and Marketing Strategy Improvement using Retail Sales Data

Business Impact

1. **Targeted Marketing**

   ○ Age-based and gender-based segmentation supports the design of customized campaigns, improving engagement and conversion rates.

2. **Inventory and Product Strategy**

   ○ Insights into popular versus high-revenue categories guide stock management, product bundling, and promotional strategies.

3. **Customer Retention**

   ○ Identification of high-value customers enables the introduction of loyalty programs, exclusive offers, and personalized engagement strategies.

4. **Revenue Optimization**

   ○ Data-driven decisions on product focus and campaign design contribute to higher sales, improved ROI on marketing spend, and stronger customer lifetime value.

Marketing Recommendations

1. **Age-Based Targeting**:

   ○ 18–25: Social media and influencer promotions.

   ○ 26–40: Email campaigns and bundled offers.

   ○ 41–60: Personalized promotions and loyalty perks.

   ○ 60+: Senior discounts and special packages.

2. **Gender-Specific Campaigns**:

- Females: Fashion, beauty, and home décor.

- Males: Electronics, gadgets, and premium items.

3. **Product Strategy**:

   - Promote popular categories (clothing, accessories) for traffic.

   - Focus on high-revenue categories (electronics, appliances) for profitability.

   - Use cross-selling (e.g., electronics with accessories).

4. **High-Value Customers**:

   - Introduce loyalty programs, VIP memberships, and exclusive offers to retain top spenders.