Objective

The goal of this framework is to estimate customer demographics—specifically age group (<25, 25–40, 40+) and gender (Male/Female)—using only transaction-level data. In the absence of explicit demographic attributes, we rely on behavioral signals such as category preferences, purchase frequency, average spend, and shopping times to develop a rule-based system for demographic classification.

1. Methodology and Behavioral Indicators

We derive demographic insights by analyzing the following behavioral patterns observed in transaction data:

- **Product Categories:** Certain types of products tend to attract specific age groups or genders.
- **Spending Habits:** High-value purchases are typically associated with older or more affluent segments.
- **Shopping Frequency:** Younger users generally make more frequent, lower-value purchases.
- **Time of Purchase:** Late-night activity is more common among younger users.
- **Brand Affinity:** For example, premium electronics often appeal more to men, while beauty and fashion categories show a female preference.

2. Age Group Classification Rules

Customers are segmented into one of three age groups based on transaction characteristics:

- **Under 25 Years**
    - Frequent, lower-value purchases
    - Shopping primarily between 10 PM and 2 AM
    - Interest in gadgets, gaming, budget fashion
- **25–40 Years**

- Likely professionals or young parents
- Moderate to high average order value
- Spending in Kids, Lifestyle, and Premium Electronics categories
- **40+ Years**
  - Fewer purchases but higher spend per order
  - Preference for Healthcare, Home Appliances, Jewelry
  - Mostly daytime shopping activity

3. Gender Prediction Guidelines

Gender is inferred using category-based preferences and purchase behavior:

- **Male**
  - Higher share of spending on electronics, gadgets, and gaming
  - Frequent purchase of premium tech brands (e.g., Apple, Sony, OnePlus)
  - Larger average transaction values
- **Female**
  - Greater activity in Beauty, Fashion, Footwear, and Apparel
  - More consistent mid-sized purchases
  - Increased shopping activity during sales and festive periods

4. Rule-Based Scoring Framework

Each transaction contributes to a cumulative score across gender and age group dimensions. The demographic group with the highest score is assigned to the customer.

Sample Scoring Table

Rule Condition	Inferred Demographic	Score
Over 60% of spend in Beauty/Apparel	Female	+2
More than 50% spent in Kids category	Age 25–40	+2

Avg. order value exceeds ₹80,000	Male, Age 25–40	+2
Frequent purchases between 10 PM and 2 AM	Age <25	+1
Over 70% of spend in Healthcare/Appliances	Age 40+	+2
Dominant purchases of brands like Apple or Sony	Male	+1
More than 10 transactions in a 2-month period	Age <25	+1

5. Example Inferences

- A customer primarily buying Apple/Sony products, spending ₹90,000+, and shopping during festive seasons → **Male, 25–40**
- A frequent buyer of fashion accessories with late-night activity → **Female, <25**
- Infrequent yet high-spend purchases focused on healthcare and appliances → **Male/Female, 40+**

6. Validation Plan

To validate the inference model once real demographic data becomes available:

- **Precision & Recall:** Evaluate accuracy of predicted vs. actual demographics.
- **Confusion Matrix:** Analyze common misclassification patterns for model refinement.
- **Lift vs. Random Baseline:** Compare model performance against random assignment.
- **A/B Testing:** Test marketing effectiveness by targeting inferred segments.

7. Assumptions

- Product preferences are demographically skewed and consistent across regions.
- High-value orders are linked with older or more affluent groups.
- Shopping behavior (e.g., timing, frequency) is indicative of age group.
- Gender-based product choices are relatively stable.

8. Strategic Applications

This demographic inference model offers several business advantages:

- Segment customers effectively, even without explicit demographic data.
- Personalize marketing campaigns (e.g., tech offers for men, fashion for women).
- Design targeted offers like student discounts or family bundles.
- Optimize inventory planning based on predicted customer profiles.

By leveraging behavioral signals from transactions, this rule-based model enables intelligent demographic profiling, which in turn drives better engagement, personalization, and operational decisions.