

**Customer Purchase Analysis Report**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

🟢 **Business Problem**

Walmart aims to better understand customer purchase behavior across various cities, product categories, and demographic groups. The primary goal of this analysis is to extract insights that influence repeat purchases, purchase amounts, and customer satisfaction (ratings). These insights can help Walmart optimize product offerings, promotions, and marketing strategies, and ultimately improve customer retention and profitability.

**🛠️ Data Cleaning and Preparation**

The dataset consisted of 50,000 customer purchase records and included columns like Age, Gender, City, Product Category, Product Name, Purchase Date, Purchase Amount, Payment Method, Discount Applied, Customer Rating, and Repeat Customer status.

The following preprocessing steps were performed:

* Converted Purchase\_Date from string to datetime format for time-based analysis
* Verified data types and converted categorical and numerical features appropriately
* Confirmed there were no null or missing values in the dataset
* Removed outliers from Purchase\_Amount based on visual inspection
* Standardized category labels where inconsistencies were found (if any)

**🔎 Exploratory Data Analysis (EDA)**

To identify trends, customer behavior, and correlations, the following EDA was conducted:

* Analyzed **purchase amounts across product categories** to identify high-revenue items
* Explored **gender distribution** and differences in product preferences
* Visualized **repeat customer patterns by city** to determine high-retention areas
* Assessed **average ratings by product category** to gauge satisfaction levels
* Examined the influence of **discounts and payment methods** on purchasing decisions
* Detected seasonal or date-related trends in purchasing behavior

**Key Findings**

**👥 1. Gender-based Purchase Behavior**

* Males contributed to **~75% of total purchases**.
* Females made fewer but higher-average-value purchases.

**Insight:** Marketing campaigns could be tailored differently for male and female segments.

**🧓 2. Age Group Distribution**

* Most shoppers were in the **26–35** and **36–45** age brackets.
* The 18–25 group was highly active online, but total purchase value was lower.

**Insight:** Focused advertising towards 26–45 age group can yield better ROI

**🧳 3. Occupation vs Purchase Amount**

* Certain job categories like tech professionals and managers had **higher spending habits**.
* Students had the lowest average spending.

**Insight:** Personalized discount programs for high-spending jobs can increase retention.

**🏙 4. City-Wise Trends**

* Category **B cities** led in total purchases, followed by A and C.

**Insight:** Expanding logistics and same-day delivery in B cities can be beneficial.

**🛒 5. Product Category Popularity**

* Top categories: **Electronics**, **Clothing**, **Household Items**.
* Certain niche categories showed seasonal spikes.

**Insight:** Seasonal inventory stocking can be optimized using this data

**📆 6. Time Lived in City vs Loyalty**

**Graph:** Line Graph

* Customers living in a city for 3+ years showed higher repeat purchases.

**Insight:** Long-term residents are more loyal – consider loyalty programs.

**📊 Key Visual Insights**

* 🧔 Males aged 26–45 are the biggest spenders.
* 🏙 Tier B cities are Walmart’s strongest market.
* 📦 Electronics and clothing dominate product categories
* 🧍 Long-term residents show higher loyalty.
* Targeted offers for specific occupations and cities can improve revenue

**✅ Conclusion**

The exploratory data analysis of Walmart’s customer transactions reveals valuable insights into consumer behavior and purchase patterns. The data shows that:

* **Male customers** contribute to the majority of purchases, although **female customers tend to spend more per transaction**.
* The **26–45 age group** is the most active and valuable customer segment, suggesting that targeted marketing in this age range could drive higher returns.
* **City Category B** dominates in sales, indicating strong market potential in mid-tier cities.
* **Product categories like electronics and clothing** are the most frequently purchased, highlighting customer preferences and seasonal demands.
* **Long-term city residents** tend to be more loyal, showing higher repeat purchase behavior.