RetailX Walmart Customer Purchase Behavior Analytics

Thanh Vu, Hoa Ly, Kevin Tran

Department of Data Analytics, Dickinson College

DATA300-02, Spring 2025
Statistical and Machine Learning

INTRODUCTION

Retail businesses rely on customer data to optimize operations, improve customer experience, and increase sales. In this project, we analyze a dataset of transactions from Walmart, which contains 50,000 customer transactions and is of sufficient size to generate an initial understanding of consumer shopping behavior. The dataset records demographic information, product categories, purchase amounts, discounts, and ratings using 12 variables.

OBJECTIVE

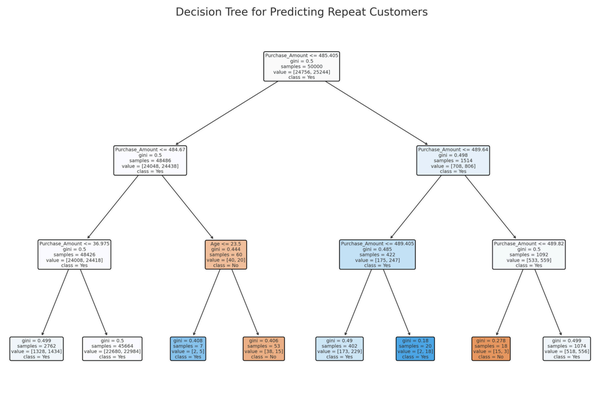
- Customer Segmentation: Grouping customers based on shopping behaviors for targeted marketing and strategic decision-making.
- Behavioral & Trend Analysis: Identifying purchase patterns, peak shopping times, and high-demand products to improve inventory management and staffing efficiency.
- Revenue Forecasting with Machine Learning: Using time series analysis to predict future sales trends and optimize stock levels.

METHODOLOGY

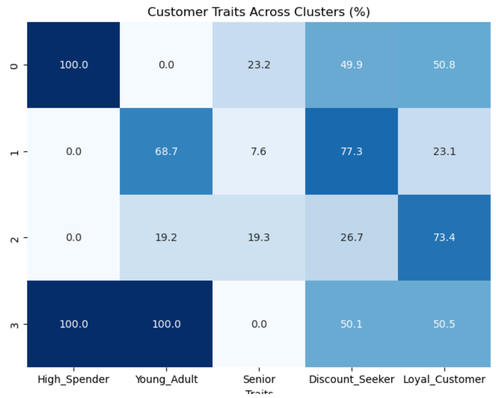
- **Customer Segmentation**: Walmart uses clustering analysis to group customers based on shared traits (e.g., frequent, high-value, or discount-driven shoppers). This helps tailor promotions, ads, and loyalty programs to specific customer types for better retention and sales.
- **Behavioral & Trend Analysis**: By studying shopping patterns—like peak hours, seasonal trends, and best-selling items—Walmart can optimize staffing, inventory management, and pricing strategies to improve operational efficiency and customer satisfaction.
- **Revenue Forecasting with Machine Learning**: Utilizing time-series forecasting techniques, we analyze past sales data to predict future revenue. This supports better inventory planning, pricing adjustments, and financial forecasting aligned with seasonal and market trends.

KEY FINDINGS

- Electronic payment infrastructure is crucial due to dominant credit/debit card use.
- High demand for smartphones and smartwatches necessitates strong inventory and accessory promotion.
- Marketing and product mix should focus on the core 26-45 age demographic.
- Key customer segments include:
  - Discount Seekers (young, price-sensitive, loyal, promo-responsive)
  - High Spenders (affluent, discount-insensitive, premium upsell targets)
  - Loyal Seniors (value consistency/service, benefit from personalized loyalty)
  - New Impulsive Shoppers (young, impulsive, engage with trendy/limited offers)
- Peak shopping times are in the afternoon and evening.
- Revenue fluctuates between days, and stays relatively stable through weeks and months.



Decision Tree Analysis to predict repeat customers



Customer Traits Across Clusters

ANALYSIS

- Customers spending less than $485 tend to have lower repeat rates, while those spending more are more likely to return.
- Receiving a discount improves the chance of returning, especially important for customers with moderate-to-high spend.
- Younger customers (e.g., Age <= 27.5) combined with a decent spend level may return, but this influence is less than Purchase Amount or Discount.

CONCLUSION

- Strengthen customer relationships by offering personalized follow-ups, exclusive offers, or loyalty incentives after a purchase.
- Focus retention strategies on customers with high lifetime value, particularly those who respond positively to discounts. These customers drive a significant portion of revenue and are more likely to remain loyal when engaged effectively.
- Develop a more robust loyalty program that offers meaningful rewards for repeat shoppers. Providing additional benefits such as early access to deals, special discounts, or free shipping can further encourage continued patronage and deepen brand loyalty.

Acknowledgements: Big thanks to Dr. Lulu Wang for supporting us throughout the course and project, and LogicCraftByHimanshi for providing this dataset via Kaggle.