Customer Transaction & Behavior Analysis

1. Introduction

Customer transaction analysis plays a vital role in understanding purchasing behavior, identifying valuable customer segments, and reducing churn. This project analyzes a **publicly available Kaggle customer transaction dataset** to demonstrate how data analytics techniques can be used to extract actionable insights from large-scale transactional data.

The objective of this study is to perform **customer segmentation, purchase pattern analysis, and churn risk evaluation** using industry-standard analytical methods such as **RFM (Recency, Frequency, Monetary) analysis** and data visualization. The findings are presented as a **general business case study** applicable to e-commerce and retail organizations.

---

2. Dataset Overview

The dataset used in this project is sourced from **Kaggle** and contains **250,000 transaction records** representing customer purchases across multiple product categories.

Key attributes include:

- **Customer ID** – Unique customer identifier
- **Purchase Date** – Timestamp of each transaction
- **Product Category** – Category of purchased product
- **Product Price** – Unit price
- **Quantity** – Number of units purchased
- **Total Purchase Amount** – Total transaction value
- **Payment Method** – Mode of payment
- **Returns** – Indicator of returned items
- **Age & Gender** – Customer demographic attributes
- **Churn** – Binary churn indicator

After aggregation, the dataset represents **49,673 unique customers**, enabling customer-level behavioral analysis.

---

3. Data Cleaning & Preparation

To ensure data quality and analytical reliability, the following preprocessing steps were performed:

- Converted **Purchase Date** to datetime format for temporal analysis
- Handled missing values in the *Returns* column through logical imputation
- Removed duplicate records to prevent inflated transaction counts
- Eliminated redundant attributes to reduce noise and improve clarity

The cleaned dataset contains **no missing values** and is structured appropriately for customer segmentation and churn analysis.

---

4. Feature Engineering (RFM Analysis)

RFM (Recency, Frequency, Monetary) analysis was used to summarize customer behavior at an aggregate level.

Definitions:

- **Recency (R):** Days since the most recent purchase
- **Frequency (F):** Total number of transactions
- **Monetary (M):** Total spending by the customer

A snapshot date was defined as one day after the most recent transaction to calculate recency consistently. Transaction-level data was aggregated into a customer-level RFM table.

---

5. Customer Segmentation Results

Customers were scored using **quantile-based RFM scoring** and classified into three segments:

Segment	Description
High Value	Recent, frequent, high-spending customers
Medium Value	Moderately engaged customers
Low Value	Infrequent and low-spending customers

Segment Distribution:

- **Low Value:** 20,146 customers
- **Medium Value:** 16,704 customers
- **High Value:** 12,823 customers

Although High Value customers are fewer in number, they play a critical role in revenue generation.

---

6. Purchase Pattern Analysis

Analysis of purchase behavior revealed:

- **Revenue Concentration:** High Value customers contribute approximately **41% of total revenue**
- **Frequency Trends:** Purchase frequency increases significantly from Low to High Value segments
- **Recency Patterns:** Low Value customers exhibit higher recency, indicating disengagement

These patterns validate the effectiveness of RFM-based segmentation.

---

7. Churn Risk Analysis

Churn rates across segments were found to be relatively similar:

Segment	Churn Rate
High Value	~19.6%
Medium Value	~20.3%
Low Value	~20.0%

Interpretation:

The similarity in churn rates suggests that churn is influenced by factors beyond transaction behavior alone, such as customer experience or external market conditions.

---

8. Business Recommendations (Generalized)

The following recommendations are framed as **general insights applicable to e-commerce and retail platforms**:

1. Focus retention efforts on high-value customers to protect revenue
2. Use recency-based campaigns to re-engage inactive customers
3. Upsell moderately engaged customers through bundles and promotions
4. Investigate non-transactional churn drivers such as service quality
5. Combine behavioral and demographic features for future churn prediction

---

9. Conclusion

This project demonstrates how customer transaction data from a **public Kaggle dataset** can be leveraged to perform meaningful customer segmentation and behavioral analysis. By applying RFM analysis and churn evaluation, the study highlights key customer patterns and provides actionable insights applicable to real-world business scenarios.