Predictive Customer Analysis
WiDS 2025

Rishika Dhiman

January 2026

Contents

1 Introduction

Predictive Customer Analysis is essentially predicting customer behaviour in the future, which includes their frequency of consumption of the product, using the data available about their purchasing patterns. It is actively being used by almost all companies for customer retention purposes. There are many methods for extracting useful information using machine learning such as using RFM (Recency, Frequency, Monetary) analysis and Customer Lifetime Value (CLV) estimation. Using these techniques business can strategise their marketing schemes and increase their profits.
This project focuses on building a minimal end-to-end customer analytics pipeline using transactional retail data. The objective is to segment customers based on their purchasing behaviour, identify customers who are likely to churn, and estimate the long-term value of each customer. The analysis is performed using RFM metrics derived directly from transaction records, ensuring interpretability and simplicity. Churn prediction is modelled using classical machine learning techniques, including Logistic Regression and Random Forest, based solely on RFM features. Customer Lifetime Value is then estimated using a deterministic formulation that incorporates purchase behaviour and predicted churn probability.

2 Recency, Frequency, Monetary Analysis

2.1 What is RFM Analysis

The RFM (Recency, Frequency, Monetary) analysis methodology can assess the behaviour of customers based on their past transaction activities. The 'recency' variable captures the timing of the customer's most recent transaction, the 'frequency' variable measures the number of transaction activities from the customer, whereas the 'monetary value' variable calculates the total or mean amount spent by the customer on the product or service. As a combination methodology that considers all three aspects, the RFM methodology holds an important position as an effective mechanism for measuring the loyalty score, value score, or churn score of the customers. The customers can be scored using the quantile-based scoring mechanism, ensuring that the customers are divided into distinct segments on the basis of their behaviour. The RFM methodology has gained extensive popularity owing to its ease and relevance to business decision-making.

- Recency: How recently a customer has made a purchase. Indicates engagement and potential interest. Customers who have purchased recently are more likely to respond to marketing efforts and promotions.

- Frequency: How often a customer makes a purchase. Measures loyalty and ongoing engagement. Frequent buyers have greater attachment to the business and can be targeted with loyalty programs or special offers.

- Monetary: How much a customer spends. Reflects customer value and profitability. High spenders are valuable for driving revenue and can be rewarded with exclusive perks.

2.2 Usage?

RFM is commonly used to

- Segment customers into groups such as loyal customers, inactive customers, or high-value customers

- Identify customers likely to respond to marketing campaigns

- Prioritize customers for targeted promotions and retention strategies
- Understand purchasing behaviour patterns

2.3 Importance

- It is simple and easy to interpret
- Helps businesses focus on high-value customers
- Enables personalized marketing strategies
- Improves customer retention and campaign effectiveness
- Requires minimal data (transaction history only)

2.4 Math behind RFM

Let:

- t_i = time of most recent purchase by customer i
- T = reference date (e.g., analysis date)
- n_i = number of purchases by customer i
- m_i =total monetary value of purchases by customer i

Recency

$$R_i = T - t$$

Frequency

$$F_i = n_i$$

Monetary

$$M_i = m_i$$

These values are typically ranked or bucketed

$$R_i^{'}, F_i^{'}, M_i^{'} \in \{1, 2, 3, 4, 5\}$$

The final RFM score can be expressed as:

$$RFM_i = w_R R_i^{'} + w_F F_i^{'} + w_M M_i^{'}$$

where w_R, w_F, w_M are weights.

3 Customer Lifetime Value Analysis

3.1 What is CLV Analysis

Customer Lifetime Value or CLV is an important business metric that calculates the overall potential profit a customer can give a business over a lifetime. The Customer Lifetime Value can be calculated using the past behaviour, like the number of purchases and the amount involved in each, on the part of a customer with the presumption that past behaviour can forecast future behaviour. The Customer Lifetime Value can be calculated on the basis that past behaviour can forecast future behaviour, which can help the business decide the future investments on retaining a particular customer or acquiring a new one, as per their lifetime value.

3.2 Usage

CLV is commonly used to

- Identify the most profitable customers
- Decide how much to invest in acquiring or retaining customers
- Allocate marketing budgets efficiently
- Compare the long-term value of different customer segments

3.3 Importance

- Shifts business focus from short-term profits to long-term value
- Helps optimize customer acquisition costs
- Improves decision-making for loyalty programs and retention strategies
- Enables sustainable business growth by prioritizing valuable customers

3.4 Math behind CLV

3.4.1 (a) Simple CLV Model

Let:

- AOV = Average Order Value
- F = Purchase frequency per period
- L = Customer lifespan (in periods)

$$CLV = AOV \times F \times L$$

3.4.2 (b) Discounted CLV Model

Let:

- r = discount rate
- R_t = expected revenue at time t

$$CLV = \sum_{t+1}^{T} \frac{R_t}{(1+r)^t}$$

3.4.3 (c) CLV using Retention Rate

Let:

- m = average profit per period
- p = retention probability

$$CLV = \frac{m \times p}{1 + r - p}$$

4 Churn Prediction

4.1 What is Churn prediction

Customer churn can be understood as the situation where customers no longer conduct business with an organization or business. However, customer churn prediction is helpful since, with the results, companies can act preventatively and prevent revenue losses. The predictive features, which are derived from RFM analysis and include recency and frequency, are good indicators of churn. Behavioural features are also included in addition to enhancing performance. The churn feature takes a binary form based on customer inactivity during a particular time period. There are a variety of machine learning algorithms that can be used for churn prediction. These algorithms include Logistic Regression and Decision Trees. Logistic Regression algorithms are particularly helpful because they are easy to interpret. Tree-based algorithms are helpful because they identify non-linear patterns within the data.

4.2 Usage

CHurn prediction is commonly used to

- Predict which customers are likely to leave
- Identify causes of customer dissatisfaction
- Design proactive retention strategies
- Improve customer experience and service quality

4.3 Importance

- Retaining customers is generally cheaper than acquiring new ones
- Reduces revenue loss due to customer attrition
- Improves customer satisfaction and loyalty
- Enables businesses to take preventive action before customers churn

4.4 Math behind Churn Analysis

4.4.1 (a) Churn Rate

Let:

- C = number of customers lost during a period
- N = total customers at the start of the period

$$Churn\ Rate = \frac{C}{N}$$

4.4.2 (b) Retention Rate

$$Retention\ Rate = 1 - Churn\ Rate$$

4.4.3 (c) Predictive Churn Model (Logistic Regression)

Let X be customer features and $y \in \{0, 1\}$ indicate churn.

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta^T X)}}$$

Customers with $P(y = 1|X)$ above a thershold are predicted to churn.

5 Relationship Between RFM, CLV, and Churn

- RFM captures past customer behaviour
- CLV estimates future customer value
- Churn analysis predicts customer attrition