Advance Database Project Proposal - Group 8

Shashank Gupta, Jay Trivedi, Nikhil Chavan, Charu Joshi

October 2022

1 Problem Statement

In this session, we will try to predict the customer's lifetime value and perform customer segmentation using the CLV estimates. Customer Lifetime Value (CLTV or CLV) can be defined as the estimated net profit the customer will contribute during their future relationship with the company. CLTV provides a picture of the business long-term and its financial viability. A high value of CLTV indicates product-market fit, the brand's goodwill, and expected recurring revenue from the existing customers. E-commerce businesses keep track of their customer lifetime value and optimize it time-to-time to confirm their standings in the market. Estimating CLTV requires RFM analysis, where RFM stands for Recency, Frequency, and Monetary value, each corresponding to some key customer trait.

- **Recency**: Refers to the freshness of customers, be it visits & purchases.

- **Frequency**: Refers to the frequency of customer transactions or visits.

- **Monetary**: Refers to the intention of customer to spend.

Further, the Customer Lifetime Value can help explore customer groups via Customer Segmentation which can be used for building customer-specific marketing campaigns and strategies.

2 Objective

We have to estimate the customer's lifetime value (CLVT) and group similar customers based on CLVT & RFM features to provide this information to the marketing team for the campaign & Cost per acquisition (CPA) optimization.

3 Dataset

The data source for this project is available at the UCI repository and Kaggle. The dataset contains over 1 Million rows and eight attributes. Following is the description of the attributes:

- **InvoiceNo**: Invoice number. Nominal. A 6-digit integral number is uniquely assigned to each transaction. If this code starts with the letter 'c,' it indicates a cancellation.

- **StockCode**: Product (item) code. Nominal. A 5-digit integral number is uniquely assigned to each different product.

- **Description**: Product (item) name. Nominal.

- **Quantity**: The quantities of each product (item) per transaction. Numeric.

- **InvoiceDate**: Invoice date and time. Numeric. The day and time when a transaction was generated.

- **UnitPrice**: Unit price. Numeric. Product price per unit in sterling.

- **CustomerID**: Customer number. Nominal. A 5-digit integral number is uniquely assigned to each customer.

- **Country**: Country name. Nominal. The name of the country where a customer resides.

4 Methodology

The data is already available in a clean format; however, we still have to apply some basic data processing operations. We can perform some data summarizing and exploratory data analysis to explore trends in the data. Moving onward, we will perform feature engineering to calculate features like Recency, Frequency, Monetary value, Retention rate, and Monthly Revenue. We will use these features to model the Customer's Lifetime Value (CLTV). Finally, we will use the predicted CLTV and other features to model customer segmentation; accordingly, we can prepare customer group-specific marketing strategies.

5 Requirements

- Jupyter Notebook (Python 3.7)

- Microsoft Excel