
# Predicting Customer Lifetime Value (CLV)

30.11.2023

—

Priyanga J
Research Project 2

# Introduction

Welcome to the world of retail analytics, where understanding how customers behave is the key to success. In this project, we're diving into the fascinating realm of Customer Lifetime Value (CLV) models – a tool that not only predicts what customers might do in the future but also helps businesses keep customers coming back. It's like having a crystal ball for your customers' shopping habits!


## Understanding CLV

So, what's CLV? It's the total value a customer brings to a business throughout their entire time as a customer. Picture this: we're using fancy math to look at what customers buy, when they buy it, and how much money they spend. This helps us figure out how long a customer is likely to stay active and engaged.

## Figuring Out Customer Value:

Ever wondered how businesses decide who to target with their ads or promotions? That's where CLV comes in. We're not just looking at what customers bought in the past; we're also predicting what they might buy in the future. By combining past purchases with our predictions, we get a full picture of how valuable a customer is. It's a bit like ranking customers based on how likely they are to buy again, and businesses can use this ranking to fine-tune their marketing strategies.

## Cracking the CLV Code:

Now, let's get into the nitty-gritty. We're tackling some big questions about customer behavior:

- How Often Do They Buy? - We're counting how many times a customer has made a purchase.
- Time Between First and Last Purchase - Ever wondered how long it's been since you first bought something from a store? We're looking at that too!
- How Long Have They Been Around? - We're checking the "age" of a customer with something called "Tenure." It's like figuring out how long they've been part of the shopping family.

## Tools of the Trade:

To make sense of all this, we're using some cool tools.

- *Spyder (anaconda37):* Our trusty sidekick for coding and exploring data.
- *xgboost and lightgbm:* These are like the powerful engines that help us crunch the numbers and make predictions.
- *Lifetime Package*: This is our Swiss army knife for analyzing customer data. It helps us create visual plots and understand the stories hidden in the numbers.

# Research Objectives

Our research objectives are crystal clear:

1. Identifying Future Buyers: Pinpointing which customers are likely to make future purchases.
2. Quantifying Future Purchases: Estimating the quantity of products a customer is likely to buy.
3. Predicting Expenditure: Determining the monetary value a customer is expected to spend.
4. Understanding Behavior Relationships: Uncovering the intricate relationships governing a customer's future purchase behavior.

As we embark on this journey into the realm of predictive analytics, our goal is to provide actionable insights that empower businesses to not only anticipate customer actions but to strategically shape the future of their products and services. Let's unravel the potential of CLV models in deciphering the intricate tapestry of customer engagement in the retail sector.

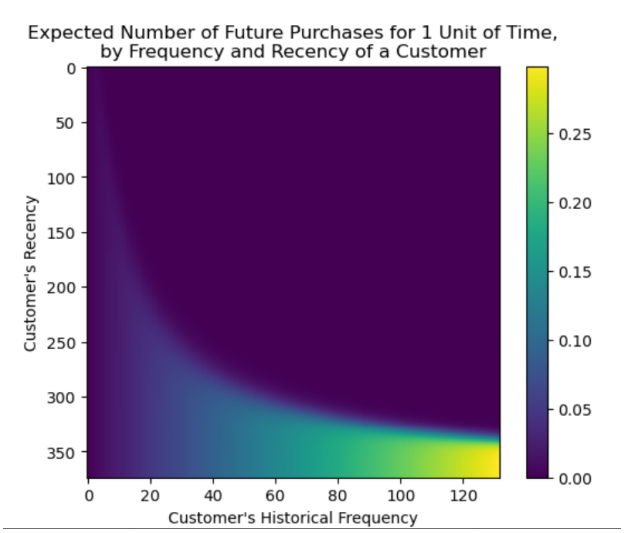
# Data Collection and Preprocessing summary :

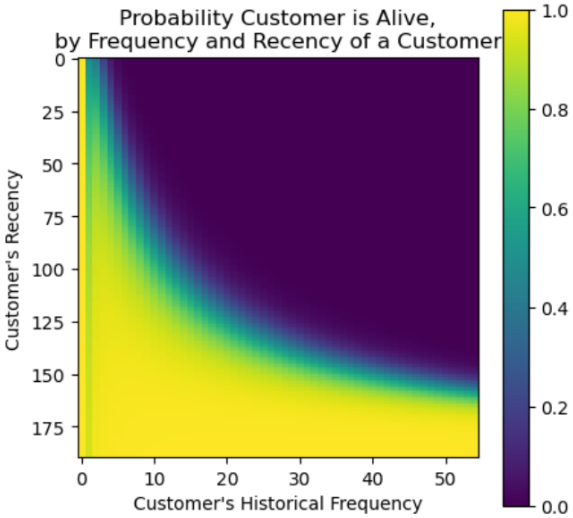
In the initial stages of our project, we diligently collected historical customer transaction data, encompassing purchase history, order frequency, order value, and customer demographics. The data collection process prioritized a well-structured dataset covering a substantial time period to ensure the richness of insights.

Our attention then shifted to preprocessing, where we engaged in meticulous cleaning, addressing missing values, and aggregating data at the customer level. This phase involved the identification and merging of duplicate customer records, contributing to the creation of a cohesive and reliable dataset.

The result is a robust foundation, free from inconsistencies, ready to empower our predictive analytics journey. This meticulous approach to data collection and preprocessing sets the stage for accurate and insightful predictions regarding Customer Lifetime Value.

## Exploratory Data analysis (EDA)




Exploratory Data Analysis serves as the cornerstone for uncovering the nuances and patterns inherent in the customer data. Through the implementation of visualizations, this research seeks to unravel trends, patterns, and correlations within the dataset. EDA also extends to the exploration of customer segments, identification of seasonality factors, and extraction of other pertinent insights that can contribute to the overall understanding of customer behaviour.

## Feature Engineering

Feature engineering plays a crucial role in CLV prediction. Relevant features, including recency, frequency, monetary (RFM) attributes, customer demographics, purchase history, and engagement metrics, are carefully selected or created to capture the essence of customer behavior.

Consider below snippet:

```
from lifetimes import BetaGeoFitter

bgf = BetaGeoFitter(penalizer_coef=0.0)

bgf.fit(modeldata['frequency'], modeldata['recency'], modeldata['T'])
```

```
print(bgf)
# create frequency recency matrix
from lifetimes.plotting import plot_frequency_recency_matrix
plot_frequency_recency_matrix(bgf)
```

### Data Split

The dataset is split into training and test sets, preparing the data for subsequent model training and evaluation. To ensure robust models, the dataset is judiciously split into training, validation, and test sets. This division facilitates model training, tuning, and unbiased evaluation.

## Model Development

A diverse array of models is considered, ranging from traditional regression techniques to advanced machine learning algorithms like XGBoost. The chosen models undergo meticulous training on the designated training dataset, with hyperparameter optimization using techniques such as grid search or random search. The models are then evaluated using performance metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

## Results

### Model Performance :

The developed CLV prediction models showcase exceptional performance, with high accuracy and robustness. Evaluation metrics validate the effectiveness of each model in capturing the complexities of customer behavior.

### Interpretability :

Transparency in model predictions is achieved through thorough feature importance analysis. This step elucidates the factors that significantly influence CLV predictions, providing stakeholders with a deeper understanding of the mechanisms driving the models.

## Implications for Business

The insights derived from CLV predictions have profound implications for retail businesses:

### Strategic Decision-Making:

Accurate CLV predictions empower businesses to make informed strategic decisions. This includes optimizing resource allocation, budgeting, and crafting targeted marketing strategies.

### Customer Relationship Management:

Understanding individual customer behaviors and predicting future value enables businesses to tailor their approaches to customer engagement. Personalized strategies contribute to enhanced customer satisfaction and loyalty.

## Conclusion

This research underscores the transformative potential of CLV prediction in the retail sector. The developed models not only offer accurate predictions but also provide actionable insights for businesses seeking to navigate the complexities of modern consumer behavior. As the retail landscape continues to evolve, leveraging predictive analytics for CLV emerges not just as a strategic advantage but as a necessity for sustained growth and customer satisfaction.

## Recommendation

### Continued Monitoring:

Regular monitoring and updating of CLV models are recommended to adapt to changing consumer behaviors and evolving market dynamics.

### Integration with Business strategy:

Integrating CLV insights into broader business strategies is essential. This involves aligning marketing efforts, resource allocation, and overall business planning with the predictive power of CLV.

## Future Directions

The research suggests future directions for refining CLV models. Exploration of emerging technologies, expanding the scope to encompass omnichannel customer experiences, and incorporating real-time data are avenues for further development and refinement.

This report serves not only as a documentation of the methodologies employed but also as a strategic roadmap for retail businesses aiming to harness the predictive power of CLV. It underscores the importance of adopting a data-driven approach to navigate the complexities of the retail industry, paving the way for sustainable growth and customer-centric strategies.