REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Predicting CLTV using Machine Learning

Name: Anand KN

SRN: R19MBA52

Date: Aug 2022

PGDM/MBA in Business Analytics

Capstone Project Presentation
Year: I

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Agenda

01 Introduction
Back Ground | Current status | Why this study

02 Literature Review
Seminal works | Summary | Research Gap

03 Problem Statement
Business Problem | Analytics Solution

04 Project Objectives
Primary & Secondary Objectives | Expected Outcome

05 Project Methodology
Conceptual Framework | Research Design

06 Business Understanding
Business Context | Monetary Impact

07 Data Understanding
Data Collection | Variables

08 Data Preparation
Pre-processing | Process | Techniques

09 Descriptive Analytics
Univariate | Bivariate | Hypothesis

10 Modeling
Machine Learning | Model Evaluation | Insights

11 Model Deployment
Applications | Demo

12 Suggestions and Conclusions
Insights | Next Step | Future Scope

13 Annexure
References | Publications | Plagiarism Score

Introduction

Customer Value or Customer Lifetime Value is picking up a lot of attention as a metric in the marketing domain of business. CLTV has been used by businesses like Retail, Telco and others as a measure of their success. There is an increased pressure in businesses to make marketing accountable. Businesses are not able to realize the return on investment (ROI) (Marketing Investment) by using traditional metrics like brand awareness, attitudes, shares and stocks. Marketing actions that are taken to improve sales or shares may impact long term profitability of the brand.

Next important thing is financial metrics like stock price and aggregate profit of the firm or a business unit that are measured are useful, but they have limited diagnostic capability. Some recent studies have found that not all customers are equally profitable. So, it may be wise to target some customers or allocate different resources to specific segments of customers. In contrast, CLTV is a disaggregate metric that can be measured to identify customers who are profitable and resources can be allocated accordingly.

Current technology makes it possible to leverage these insights and customize marketing programs for individual customers or customer segments. Nowadays, 69% of firms monitor CLTV, but they do it inefficiently. Instead, 81% of firms are doing a good job in measuring CLTV to increase their sales.

According to a survey conducted by one of the firms, 55% of developing companies believe that it is “Very important” to invest in customer service programs. Study by another company showed that a 5% increase in the retention rate could result in a 25% to 95% increase in profits.

Calculating CLTV for different customers helps in a number of ways, mainly regarding business decision making.

Literature Review

CLTV stands for Customer Lifetime Value and measures all the potential profits a particular customer can bring to the organization. For instance, you have an online shop selling bicycles and all the additional products, and a new customer has just bought one. In the future, they may buy a helmet, new tires, a basket, etc. At some point, they may come for another bike. All these potential purchases and revenues are CLTV.

Lifetime Value is generally defined as the present value of all future profits obtained from a customer over his or her life of relationship with a firm. CLV is similar to the discounted cash flow approach used in finance. However, there are two key differences.

First, CLTV is typically defined and estimated at an individual customer or segment level. This allows us to differentiate between customers who are more profitable than others rather than simply examining average profitability. Second, unlike finance, CLV explicitly incorporates the possibility that a customer may defect to competitors in the future.

By Understanding their customers, a company can make strategies by which they can retain their customers and increase overall profitability. Calculating CLTV can help companies to investigate the parameters that companies generally ignore. At the beginning of a relationship, customers are more valuable due to the future potentials that they offer.

Many studies use different methods, including generalized regression, logistic regression, quantile regression, latent class regression, CART, Markov chain modelling, neural network to create past customer behaviour models.

Problem Statement

Predicting Customer Life Time Value for a retail business

- Customer lifetime value (CLTV) is one of the key stats likely to be tracked as part of a customer experience program.
- CLTV is a measurement of how valuable a customer is to your company with an unlimited time span as opposed to just the first purchase.
- This metric helps you understand a reasonable cost per acquisition. CLTV is the total worth to a business of a customer over the whole period of their relationship.
- It's an important metric as it costs less to keep existing customers than it does to acquire new ones, so increasing the value of your existing customers is a great way to drive growth.

Project Objectives

Here are the key objectives of predicting CLTV:

Analyze the transactional data of the retail apparel business

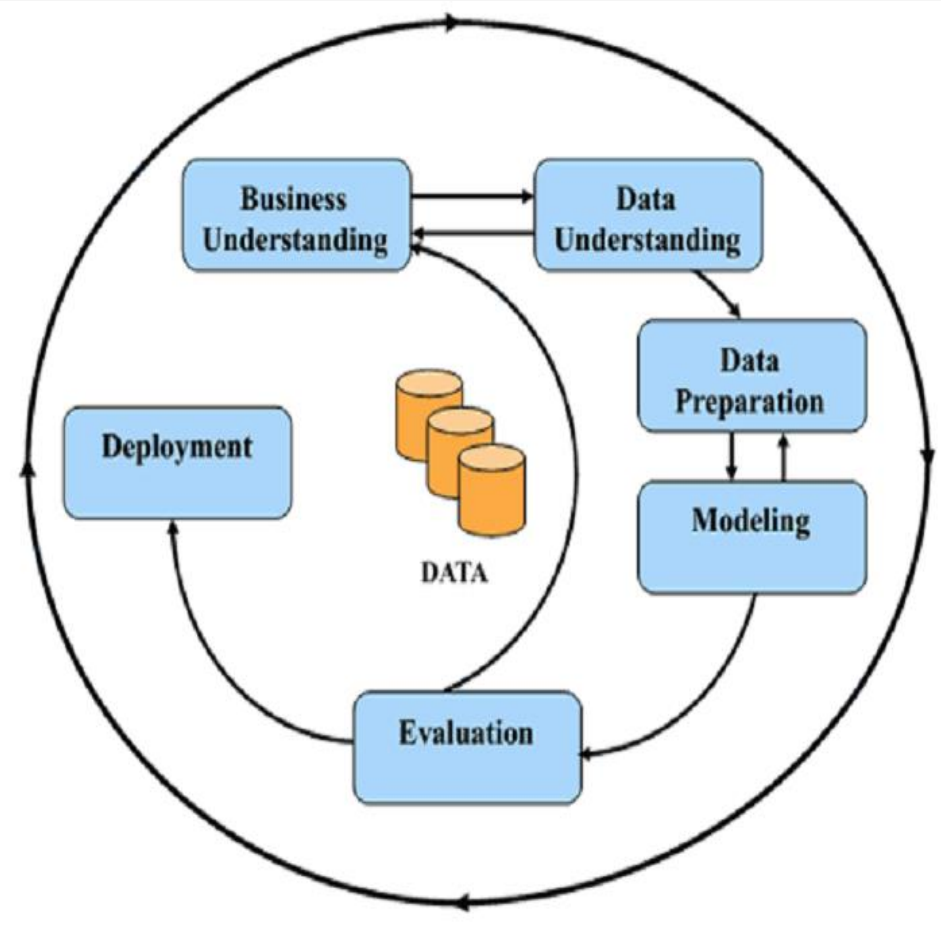
Classify the customers based on the purchases into different groups based on their value

Compute CLTV for the customers on a periodic basis

CLTV is an important metric, and the way we approach it can both define the business and could vary significantly depending on what we're trying to get from the business. CLTV is a measurement of how valuable a customer is to the business over time than just a simple exchange of goods for money

Of course, not all customers are valued equally. Keeping CLTV high can be essential to the success of the business After all, a higher CLTV means that the customers are more loyal. This helps a company forecast profitability, set customer acquisition budgets and determine the growth and improvement

Project Methodology



Project methodology that will be used as part of this project is being discussed in this chapter with details of the approach.

The process that will be followed as a project methodology is CRISP-DM framework that starts with understanding the business and then narrowing down into specific areas of interest like understanding the data, preparing the data for more insights, building the model, evaluating the model and deployment.

CLTV for the Apparel_POS data: Data used for this project consists of all the purchases from 2018 to 2020 for a period of 3 years. The company is an Apparels Retailer that sells apparels across India.

The main target variable that needs to be predicted here by looking at the data is if a customer repeats his purchase or not with the retailer. If the customer repeats his purchases, then the CLTV value would be on the higher side. As part of predicting the CLTV, we are more interested in identifying the repeat customers who give more business.

Using Logistic Regression machine learning algorithm, we derive the Churn Propensity which is used to arrive at the Churn Indicator based on the first purchase date and the recent purchase date. Using this Churn Indicator, we come to a conclusion if a customer has Churned or not.

Business Understanding

Retail business is more challenging than ever. Competition is exponential along with Amazon being a major player in this domain which drives the acquisition costs to shoot up while customer bringing down the profitability. It's a double edged sword.

The days of increasing the size of the businesses by acquiring new customers are gone, it is the value that these customers can bring is what matters.

To increase revenue from the existing client base and to get a deeper understanding of the clients with high value is the only way today for the retailers to grow the business. Keeping the customers repeat their purchases is the key to get more value.

To understand the client's preferences and behaviors, we need to make use of next level of analytics that is advanced in functionality and machine learning techniques.

Predicting the value add by a client is key activity to address these objectives. Client value is the best and candid way to run the business.

Data Understanding

Data used in this project is from retail Apparels business that sells apparels across India. This is set contains all the purchases occurring between 2018 and 2020 for a period of 3 years.

Here is the information on some of the key attributes:

- **Order No**: Order Number. Nominal, an integer number uniquely assigned to each purchase. The code starts with letter 'M'.
- **Quantity**: The number of each product (item) per purchase. Numeric.
- **Order Date**: Order Date and time. Numeric, the day and time when each item was ordered.
- **MRP**: MRP. Numeric, price of each item in Rupees.
- **Customer_ID**: Customer identification number. Nominal, an integer number uniquely assigned to each customer.

As per the formula for calculating CLTV below: (Hardie, 2006)

$$CLTV = \sum_{t=0}^{T} \frac{(p_t - c_t)\, r_t}{(1 + i)^t} - AC$$

Where

p_t = price of the item that a consumer pays at time t,
c_t = direct cost for servicing the customer at time t,
i = discount rate or cost of capital for the firm,
r_t = probability of customer repeat buying or being "alive" at time t,
AC = Acquisition cost, and
T = time horizon for estimating CLTV.

Data Preparation

As the data that we have for predicting CLTV in this project is a historic transactional data from a retail apparel firm, we need to perform initial exploratory data analysis (EDA) to understand the data fields by using python.

We need to see if there is any need for cleanup of the data by looking for duplicate records, null value rows, etc.

Checking for missing values in the data as part of preparing for data modeling.
Below is a snapshot of the data preparation steps to check for missing values based on Customer_ID feature.

	Customer_ID	Order No	Order Date	Quantity	MRP	TotalSales
0	14955	M011000	2018-01-05 02:08:00	1.0	945	945.0
1	2532	M0110001	2019-08-22 15:12:00	1.0	3095	3095.0
2	10850	M0110002	2019-08-22 16:01:00	1.0	1095	1095.0
3	19829	M011001	2018-01-05 05:16:00	1.0	995	995.0
4	10993	M0110013	2019-08-22 22:49:00	1.0	2995	2995.0

	Count	Proportion
Customer_ID	0	0.0
Order No	0	0.0
Order Date	0	0.0
Quantity	0	0.0
MRP	0	0.0
TotalSales	0	0.0

Descriptive Analytics

Descriptive analysis of the data to check on the key metrics has been carried out using python pandas library methods. Descriptive analytics summarizes the data by computing mean, median, mode, standard deviation likewise.

Below are the screenshots of descriptive analytics performed on the data along with further analysis on the data for details like the time range, total number of unique customers, total quantity sold, etc.

	Customer_ID	Quantity	MRP	TotalSales
count	17782.000000	17782.000000	17782.000000	17782.000000
mean	10263.435778	1.038916	1027.847824	1058.721179
std	6055.332535	0.284685	537.298888	568.953100
min	2.000000	1.000000	95.000000	95.000000
25%	4746.000000	1.000000	595.000000	595.000000
50%	10239.500000	1.000000	995.000000	995.000000
75%	15539.000000	1.000000	1295.000000	1295.000000
max	20610.000000	20.000000	3495.000000	6725.000000

```
The Time range of transactions is: 2018-01-03 to 2020-12-02
Total number of unique customers: 8381
Total Quantity Sold: 18474.0
Total Sales for the period: 18826180.0
```

Modeling

As part of this project, Logistic Regression machine learning algorithm has been used to create the model for calculation of CLTV for the retail apparel dataset. logistic regression is used to predict if the customer will churn or not. Churn indicator is an important metric that we are using here to classify customers based on their purchase.

Churn propensity estimates the likelihood of a customer to leave in the next period of time. In our case, churn propensity is based on if the customer has repeated purchasing from the retailer or not. If the customer has purchased only once and has not purchased anything again, then the customer is considered as churned and if the customer has repeat purchases, then the customer is considered as not churned.

For calculation of CLTV using the formula, we need alive propensity, as we have derived the Churn propensity as part of the logistic regression model. Alive propensity can be calculated by using 1-Churn propensity.

Below screenshot shows the total number of transactions considered for prediction of CLTV and it shows the first 5 rows of the CLTV calculations

Total number of transactions happened in the given period: 15908

	Name	Sum of Gross Margin	Sum of Product Discounted %	Prop_churn	prop_alive	t	Acquisition cost	Numerator	1+i	(1+i)t	CLTV
0	Priyanka Khandelwal	-4504.690	0.000000	0.00795	0.99205	0.033333	0	651.528838	1.99205	1.023238	636.732428
1	Protima Tiwary	153.000	0.499371	0.00924	0.99076	0.000000	0	650.681630	1.99076	1.000000	650.681630
2	arjita grover	3025.979	4.297736	0.01411	0.98589	3.666667	0	647.483258	1.98589	12.373779	52.327042
3	Swati Gandhi	-3532.410	0.500000	0.03802	0.96198	0.000000	0	631.780365	1.96198	1.000000	631.780365
4	Sonal Somani	365.170	0.400000	0.04787	0.95213	5.300000	0	625.311378	1.95213	34.649446	18.046793

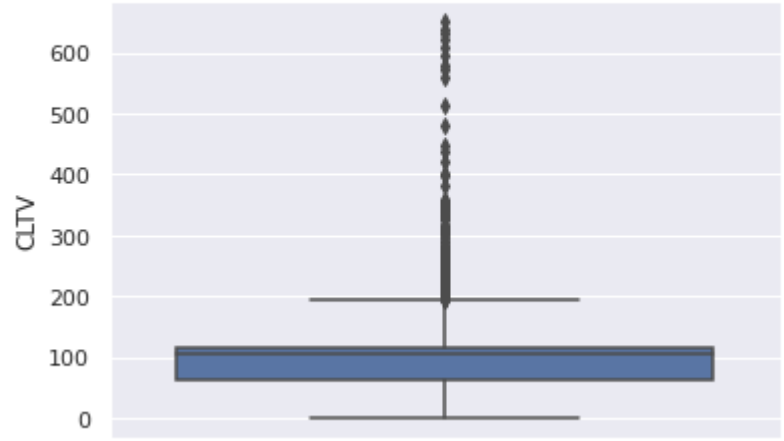
Model Evaluation

CLTV has been calculated or predicted using the Logistic Regression machine learning algorithm. CLTV thus calculated can be used to make marketing decisions to target the high value customers who could increase the profit margins of the firm.

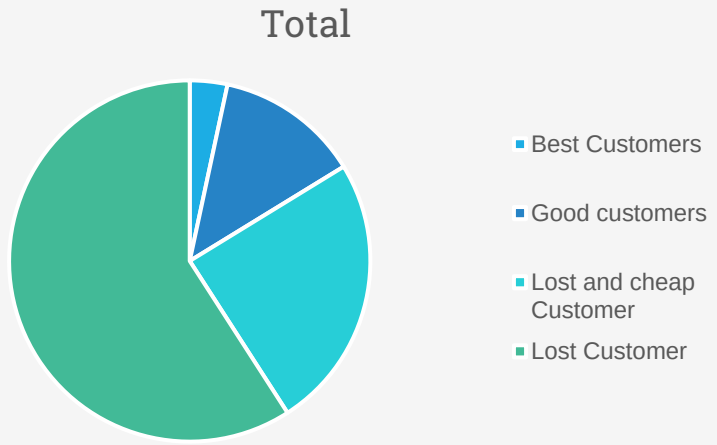
Below are the outcomes of the model evaluations – The table shows the customer classification based on their CLTVs derived from the box plot. The classification is self-explanatory.

```
sns.boxplot(y='CLTV',data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1642c99090>
```



Row Labels	Count of Row Labels
Best Customers	535
Good customers	2050
Lost and cheap Customer	3920
Lost Customer	9403
Grand Total	**15908**



Model Deployment

As part of the first year project work, the modeling has been carried out using the data from csv file with python scripting on the Google Colab.

Modelling of the data has been carried out by making use of Logistic Regression algorithm using IBM SPSS.

As an enhancement and an overall completion part of the second year project work, this work will be carried out through a well-planned deployment.

Results and Insights

The marketing team can now make use of the CLTV values to target high value customers and increase the sales.

Also, it is hard for the firms to target individual customers. We need to segment customers based on demographic data if it was available. Segmentation based on demographics could provide more insights into the customer profile to focus more on the customers.

Conclusion and Future Work

This project has been developed for first year considering the logistic regression machine learning algorithm to model the data for a retail apparel firm that sells apparels in India.

The CLTV calculated or predicted helps the firm to take decision in terms of promotions and other offers that can be extended to their high value customers.

The objective of predicting CLTV for retail firms is to ensure that the firms are in a state of mind to know who their high value customers are and can accordingly work on retaining them to drive higher profit margins.

This project work will be further continued by considering complex ML algorithms to predict and compare CLTV.

CLTV helps to design an effective business plan and also provide a chance to scale the business. However, as already discussed, a lot will depend on marketing strategy to extract profit. In general, marketing automation platform manages the customer lifecycle.

References

Analyzing and Predicting Customer Lifetime Value (CLTV) | Tellius. (n.d.). Retrieved March 23, 2022, from https://www.tellius.com/use-cases/analyzing-and-predicting-customer-lifetime-value-cltv/

Customer Lifetime Value Prediction using Machine Learning | Addepto. (n.d.). Retrieved March 23, 2022, from https://addepto.com/customer-lifetime-value-prediction-machine-learning/

Hardie, B. (2006). Modeling Customer Lifetime Value. *Journal of Service Research*, *Volume 9*(Clv), 17. https://doi.org/10.1177/1094670506293810

Predicting Customer Lifetime Value : A Definitive Guide. (n.d.). Retrieved March 23, 2022, from https://www.analyticsvidhya.com/blog/2020/10/a-definitive-guide-for-predicting-customer-lifetime-value-clv/

Predicting Customer Lifetime Value with AI Platform: training the models | Cloud Architecture Center | Google Cloud. (n.d.). Retrieved March 23, 2022, from https://cloud.google.com/architecture/clv-prediction-with-offline-training-train

"5 Simple Ways to Calculate Customer Lifetime Value | by Marie Sharapa | The Startup | Medium." https://medium.com/swlh/5-simple-ways-to-calculate-customer-lifetime-value-5f49b1a12723 (August 6, 2022).

"Customer Lifetime Value Modelling." https://dr.lib.iastate.edu/entities/publication/61e87bd9-a2ca-4fce-b7e5-5e196a332d48 (August 6, 2022).

"Customer Lifetime Value Prediction Using Machine Learning | Addepto." https://addepto.com/customer-lifetime-value-prediction-machine-learning/ (March 23, 2022).
"Logistic Regression — Detailed Overview | by Saishruthi Swaminathan | Towards Data Science.
" https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc (August 6, 2022).

"Predicting Customer Lifetime Value : A Definitive Guide." https://www.analyticsvidhya.com/blog/2020/10/a-definitive-guide-for-predicting-customer-lifetime-value-clv/ (March 23, 2022).
Sharma, Shreya. 2021. "Customer Lifetime Value Modelling.": 1–31.

Annexure

Additional Information | Plagiarism score

Prediction of Customer Lifetime Value (CLTV) using Machine Learning

ORIGINALITY REPORT

11%	10%	5%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	mafiadoc.com Internet Source	2%
2	dr.lib.iastate.edu Internet Source	2%
3	global.oup.com Internet Source	2%
4	www.analyticsvidhya.com Internet Source	2%
5	Submitted to University of Durham Student Paper	1%
6	Submitted to Universiti Malaysia Perlis Student Paper	<1%
7	towardsdatascience.com Internet Source	<1%
8	www.bjmc.lu.lv Internet Source	<1%
9	Lavneet Singh, Nancy Kaur, Girija Chetty. "Customer Life Time Value Model Framework	<1%

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Annexure

Publications | Conferences

REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Thank You!