REVA UNIVERSITY
Bengaluru, India

A Project Report on

Prediction of Customer Lifetime Value (CLTV) using Machine Learning

Submitted in Partial Fulfilment for Award of Degree of
Master of Business Administration
In Business Analytics

Submitted By
Anand Kumar N
R19MBA52

Under the Guidance of
Mithun DJ
Senior Manager – Data Science,
RACE, REVA University

REVA Academy for Corporate Excellence - RACE
REVA University
Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064
race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Anand Kumar N** hereby declare that I have completed the project work towards the first year of Master of Business Administration in Business Analytics at, REVA University on the topic entitled **Prediction of Customer Lifetime Value (CLTV) using Machine Learning** under the supervision of **Mithun DJ**. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year **2022**

Place: Bengaluru — Name of the Student: Anand Kumar N

Date: — Signature of Student

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to Certify that the Project work entitled **Prediction of Customer Lifetime Value (CLTV) using Machine Learning** carried out by **Anand Kumar N** with **R19MBA52,** is a bonafide student at REVA University, is submitting the first-year project report in fulfilment for the award of **Master of Business Administration** in Business Analytics during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said Degree.

Signature of the Guide	Signature of the Director
Name of the Guide: **Mr. Mithun DJ**	Name of the Director: **Dr. Shinu Abhi**

External Viva

Names of the Examiners

1. Harsh Vardhan, Chief Digital Technology Architect, Capgemini
2. Pradeepta Mishra, Director – AI, L&T InfoTech
3. Rajib Bhattacharya, Director – Data & Analytics, Cargill

Place: Bengaluru

Date: 20 Aug 2022

Acknowledgement

I am highly indebted to **Dr. Shinu Abhi**, Director, and Corporate Training for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

I would like to thank my project guide **Mr. Mithun D. J.** for the valuable guidance provided to understand the concept and execute this project. It is my gratitude towards **Dr. Jay Bharateesh Simha** and all other mentors for the valuable guidance and suggestions in learning various data science aspects and for their support. I am thankful to my classmates for their aspiring guidance, invaluable constructive criticism, and friendly advice during the project work.

"I would like to acknowledge the support provided by the founder and Hon'ble Chancellor, **Dr. P Shayma Raju**, Hon'ble Vice-Chancellor, **Dr. M. Dhanamjaya**, and Registrar, **Dr. N. Ramesh**."

It is my sincere gratitude towards my parents and my family for their kind co-operation and encouragement which helped me in the completion of this project.

Place: Bengaluru

Date: 20 Aug 2022

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **Prediction of Customer Lifetime Value (CLTV) using Machine Learning** was scanned for similarity detection. Process and outcome are given below.

Software Used: Turnitn
Date of Report Generation: 19/08/2022
Similarity Index in %: 11%
Total word count: 5561
Name of the Guide: Mithun DJ

Place: Bengaluru
Date: 20 Aug 2022
Verified by: Dincy Dechamma

Name of the Student: Anand Kumar N
Signature of Student

Signature
Dr. Shinu Abhi,
Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	CLTV	Customer Lifetime Value
2	AC	Acquisition Cost

List of Figures

List of Tables

List of Equations

Abstract

At this current times of customer focussed marketing, where every business is competing to get and be top in the market, businesses are required to consider all the factors that will enable them to be successful in the long run. Providing best customer experience is the most crucial factors for the businesses.

Business world today has changed its focus from a product development and sales to pivoting around their customers more likely than before. Putting at the customer at the centre is the new approach that's getting normal across the industry. The main reason being the numerous choices or options that people get to choose an offering.

With the current trend, retail companies are in a race and are competing against one another to both serve better and also to attract more clients from their immediate competitors, the main need for the businesses to expand their clientele and retaining existing client base is very important.

The exercise of retaining existing clients (by giving discounts, targeted offers, etc.) involves a huge cost and is as good as the process of acquiring new customers. With this, out of all the customers a business attracts towards it, not everyone is same in-terms of value that can be derived as some of them add more value or profit by purchasing more frequently than just a one time.

For the businesses to at-least sustain in competitive markets, the key is to identify such categories of clients and targeting only the clients that add more value. Brand awareness, client attitudes, or even sales and shares that fall under the marketing metrics category are not good enough to show a return on the investment in marketing the offering as these are traditional in nature.

This drives the need for predicting the **Customer Lifetime Value** (CLTV) for a business. In the process of making internal decisions like: from allocating

budget marketing spends, to minimizing potential losses and preventing customer churning, CLTV plays a key role

The value that a client adds through his lifetime with a business represents the total amount of money a customer spends in a given business or product during their life time which means time during which the customer is actively purchasing from that business.

In a big picture, the customer value can be looked up as a measure of the profit associated with a particular customer relationship, which should be the guiding factor on how much the business is interested in retaining that customer or customer segment.

In order to foster the full potential of loyal customers, the customers need to be first identified based on their value to the business and then segmented to group them into targeted and profitable customers, prediction of CLTV is the initial point.

The ability of an organization to identify, create, nurture and sustain a loyal and high valued client experience and relationships is an important aspect of corporate success.

The proposed solution framework here provides a step-by-step process of calculating CLTV for a retail apparel business using different methods of prediction. CLTV is a recent marketing paradigm that helps to pursue long term relationships with profitable customers.

We go through different approaches like Historical and Predictive methods to perform a detailed analysis of the purchase data to predict the CLTV.

***Keywords:* CLTV, CLV, Customer Value, Sentiment Analysis,**

Contents

Chapter 1: Introduction

This chapter provides a brief introduction to the concept of CLTV – Customer Lifetime Value and its importance in increasing the profit margin of businesses. It also introduces the formula that is being used to predict the CLTV for the retail apparel business.

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 success for a business. Business have pressure 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 to improve sales or shares could have an impact on the long term profitability of the business

Financial metrics like aggregate profit of a business or stock price which are measured are useful with diagnostic capability. Recent studies in the area customer value measurement have come up with the observations that there will be no uniformity across all clients being profitable. It may be a wise strategy to reach out to target some clients or allocate resources differently to specific segments of clients.

Resources can be allocated by identifying customers who are profitable based on prediction of CLTV, a metric that is disaggregate in nature.

Technology advancements are enabling the businesses gather infinite amount of client activity data which lets them use this data as preferred rather than intentional. Modelling sophistications have enable marketers to convert the data into valuable insights.

With making use of the current technological enhancements, it enables the businesses to use these insights to introduce custom marketing programs for individual clients or segments of clients.

Nowadays, 69% of firms monitor CLTV, but they do it inefficiently. Instead, 81% of firms are doing an extraordinary job in gauging customer value to expand their selling power.

As per a study conducted by one of the firms, 55% of companies that are developing think that it is “Very important” to invest in customer focused service programs. Observations from another company showed that a getting the retention rate up by 5% could result in a 25% to 95% expansion in profits.

Calculating CLTV for different customers helps in several ways, mainly regarding business decision making. Knowing CLTV, we can determine many things not limited to some of the below key items:

To calculate how much a business needs to spend to acquire a similar customer and have a relationship that is profitable

To predict the kind of products customers with the highest CLTV want?

Identify the highest profitability products

Identify the most profitable clients

Together, these decisions can significantly boost profitability

Calculating or Predicting CLTV is an important step towards increasing the profitability of the business for a firm:

Value of customer through the lifetime with a business is something that can be gauged in terms of monetary is the total value that a client adds to the business from all the purchases they make his entire time of relationship. The lifetime in this context is the entire time the client is purchasing from the business before moving to competitors.

The general formula to calculate **CLTV** is as below: (Hardie, 2006)**.**

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

Where

p_t = price paid by a consumer at time t,

c_t = direct cost of 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.

Equation 1.1 CLTV Formula

The above formula to calculate CLTV looks simple but the complexity around it will unfold as we look at the definition of CLTV. Future sales and values in monetary terms is being predicted by making use of the client purchase data that is historic in nature brings in the complexity to some extent.

As part of this project, we would be calculating the CLTV using Logistic Regression machine learning algorithm from the data that we have for the retail apparel business.

This chapter provided an introduction along with the formula to predict CLTV. The next chapter is all about the literature review of the CLTV concept by reviewing the various research papers that are being referred as part of this exercise.

Chapter 2: Literature Review

The second chapter here talks about the literature review of the CLTV concept by referring the various research papers not just limited to the ones listed in the bibliography.

CLTV means the value that a client adds to the business which gauges all the potential profits a particular client can bring to the organization. For example, let's consider an online shop selling sports goods and all the additional products, and a new client has just purchased a bat. In the future, they may purchase a ball, wicket, gloves, etc. At some point, they may come for another bat. All these potential purchases and revenues are the value that customer would add.

CLTV 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 like the discounted cash flow approach used in finance. However, there are two key differences.

First, the prediction of CLTV is carried out at an individual client or segment level. This lets the businesses differentiate between clients who are of more value than others rather than simply examining an average across profit. Second, unlike finance, CLTV explicitly incorporates the possibility that a customer may defect to competitors in the future (Hardie, 2006).

This value is one of the most important factors and plays a vital role when it comes to maximizing the company's efficiency. We have already mentioned some of the benefits of CLTV. However, here is a more detailed example: when the total cash flow of a given customer is known, it is straight forward to understand how far the business has got with customer retention and maximize ROI (return on investment) (Customer Lifetime Value Prediction using Machine Learning | Addepto n.d.).

Value of customer through the lifetime with a business is something that can be gauged in terms of monetary is the total value that a client adds to the business from all the purchases they make his entire time of relationship. The lifetime in this context is the entire time the client is purchasing from the business before moving to competitors

(Predicting Customer Lifetime Value: A Definitive Guide n.d.).

A company can come up with strategies by which they can retain their clients and the ability to increase the overall profit by understanding the customers. Companies can investigate the parameters that companies generally ignore by predicting CLTV. At the beginning of a relationship, customers are more valuable due to the future potentials that they offer.

There are numerous CLTV models that have been developed. These models are the PCV model (Past Customer Value), RFM model (Recency, Frequency, Monetary), SOW model (Share of Wallet) and future-past customer behaviour model (Sharma 2021).

Future behaviour of customers is considered by the future - past customer behaviour models whereas some analytical models include acquisition cost when calculating lifetime value while some others do not.

Retention rate is used by most of the future-past customer behaviour studies to determine the activation period

Different methods like generalized regression, logistic regression, quantile regression, latent class regression, CART, Markov chain modelling, neural network to create past customer behaviour models, etc. are used by many studies.

Techniques like decision trees, clustering, logistic regression, artificial neural networks, support vector machine, random forests, etc. are used by industries like retail, insurance, banking, telecommunication, financial services taking advantage by significantly using data mining techniques for identifying CLTV and performing analysis based on CLTV.

Other methodologies that are available like the proposed extended RFM analysis method with one additional parameter called Count Item can be used. When the results of these approaches are compared, it's understood that there is no difference to the clustering result with addition of the count item as a new parameter to the RFM method. So, the weighted RFM method is used to calculate CLTV for each segment. Marketing and sales strategies of the company can be explained by the results of calculated CLV for different segments.

Hence Logistic Regression has been used in this project work to predict the customer churn or customer alive propensity using which CLTV is predicted with the formula discussed earlier.

Once the literature review of the CLTV concept is completed, the next important step is to define the problem. The next chapter states the problem that will be addressed by predicting the CLTV.

Chapter 3: Problem Statement

Predicting Customer Lifetime Value for a retail business

Problem statement provides a clear definition of the key area that needs to be focused as part of this project. Prediction of CLTV is required to address problem areas not just limited to the below:

One of the key stats that is likely to be tracked as part of customer experience program is CLTV.

To measure how valuable a customer is to a business with an unlimited span of time as opposed to just the first purchase is based the CLTV prediction.

The total worth of a customer to a business over the whole period of their relationship is based on the CLTV calculated. This helps the business to understand the reasonable cost per acquisition.

The cost to retain and keep the existing customers is comparatively lesser than it would cost to acquire new customers, in order to drive growth it makes a great sense to increase the value of existing customers.

As the problem statements have been defined, the next chapter moves into listing down the key objectives of the project.

Chapter 4: Objectives of the Study

As part of this chapter, the key objectives to be achieved are listed down.

Below are the objectives for predicting CLTV:

- Analyze the transactional data of the retail apparel business
- Classify the customers based on the purchases into different groups based on the CLTV
- Compute CLTV for the customers on a periodic basis

Listing the key objectives is an important activity, with that completed. The next chapter provides details about the project methodology.

Chapter 5: 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.

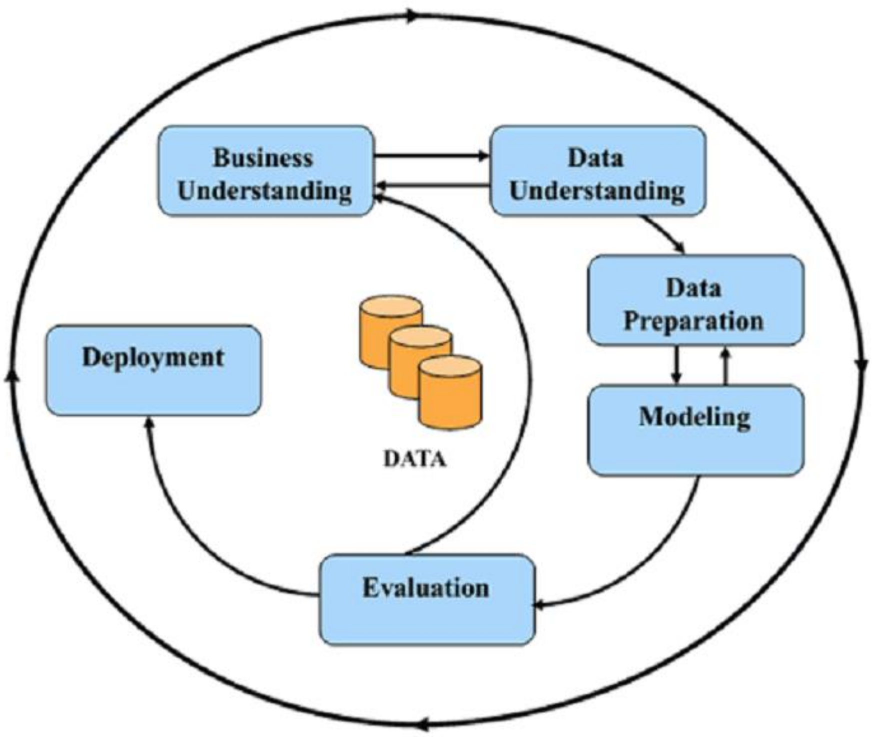

Figure 5.1 CRISP DM Methodology

Looking into various studies, we understand that there are two broad approaches to modeling the CLTV problem(*Predicting Customer Lifetime Value : A Definitive Guide*, n.d.):

- **Historical Approach:**

  Aggregate Model —this method calculates the CLTV based on past purchases taking into consideration the average of revenue per customer It gives us a single value for the customer.

  Segmentation Model —this method groups the customers into different segments based on the transaction date, etc., and calculates the average revenue per segment. This method gives CLTV value for each segment.

- **Predictive Approach:**

  Model based on ML—this method makes use of different regression techniques to fit on past data to predict the CLV.

  Probabilistic Model — estimates the count of purchases to happen in the future with monetary value for each purchase by making use of the data distribution based on probability

As part of the first year project, CLTV is predicted using Logistic Regression Machine Learning Algorithm to identify if a customer has churned or not which is needed in the formula to predict CLTV (*Customer Lifetime Value Prediction Using Machine Learning | Addepto*, n.d.).

Below are the details about the approach that has been used to predict 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.

Attribute Information - Some of the key attributes are listed here to get a glimpse into the dataset:

Order No.
Order Date
Customer ID
Quantity
Unit Price
Status

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.

In-order to identify the customers that are of high value to the business, we need to identify the customers that churn meaning who do not repeat their purchases after their initial or first purchase. 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. Churn is an important flag type variable whose value helps group the customers into repeat or non-repeat customers leading the way to calculate the CLTV for those repeat customers.

CLTV is calculated on a monthly basis using the formula discussed in the Data Understanding section for the customers.

As the project methodology and approach has been covered in this chapter, the next chapter onwards, each of the steps part of the methodology are covered in more detail. To start with, the next chapter elaborates the Business Understanding area.

Chapter 6: Business Understanding

Here the first step of the CRISP DM methodology, Business Understanding has been elaborated along with this relevance to the prediction of CLTV.

As part of this project, we are focusing on retail business. 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.

“In the moment” personalization, using technology like black box which is gaining a lot of momentum to convey the right information at the correct point of time to each client.

Getting to know for sure who the customer is and what they intend to do by matching the customer journeys with insights is the key driving factor of customer engagement that brings in value.

The websites and algorithms part of the recommendation engines for products available today are not in position to deliver this kind of engagement with the customers.

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.

Activation of these insights in an easier way that enables retailers to attract more clients that are of higher value, increase loyalty within the existing client base and step in at the right time when clients deviate from their expected purchase journey is what is needed from a system that would be used.

To achieve the business goals like transforming the buyers who purchase for one time to purchase again and again, decrease the number of clients that are churning out, and getting increased client value can be driven by making use of predictive analytics that combines details of clients, steps and quantification for the marketers. The below pictures represent the key steps:

Figure 6.1 Customer Conversion

Metrics like count of clicks on the item, scores that indicate client engagement , percentage of clients that are churning, number of items added to the cart and so on that are tracked by the marketers today are kind of distractions. The key insights like the base-line impact to the business – number of purchases and profits from those purchases can be captured making use of the metrics captured.

Retailers need to shift the thoughts to thinking about their customers holistically in addition to looking at the channels that are isolated from others and additionally utilize metrics that are channel-specific like the metrics discussed earlier to keep an eye on the performance.

"Let me go and click on that email I saw on my mobile device while I am in the store to ensure the retailer can attribute my purchase" may not be something a customer would say.

Looking at the long term profitable growth by focusing on customer-centric metrics will let the retailers to come up with the best decisions.

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.

By predicting CLTV, below are some of the key business problems that we could try to provide an answer to:

1. Identifying the most profitable customer/s
2. How best offers could be offered by the company to make the most?
3. Drive segmentation of profitable customers
4. Budget allocation for acquiring new customers to the business

Client information that is important in nature can be pulled out, that supports making of decisions that are critical for the business by leveraging the data mining techniques.

Once the Business Understanding has been clearly documented, the next important step in the process is understanding the data.

Chapter 7: Data Understanding

As part of this chapter, the data is looked at very closely from different perspectives to ensure the data is all set and is in good shape for analysis and prediction of CLTV.

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.

Below table provides information about the different columns that are part of the data that is being used for data mining and analysis to calculate the CLTV for the customers that have purchased apparels from this firm.

Column	Data Types
Order No	Object
External Order No	Int64
Order Date	DateTime64[ns]
Order Type	Object
Status	Object
Customer Name	Object
Country	Object
State	Object
City	Object
Email	Object
Color	Object
Quantity	float64
Return Qty	float64
Order Currency	Object
Price	Object
Ship Cost	Float64
Packing Cost	Float64

Discount	Object
Discount Code	Object
Tax	Float64
Invoiced	Object
Base Currency	Object
COGS	Object
Invoiced In Base Currency	Object
Gross Margin	Object
GM Percent	Float64
On Hold Status	Object
Replacement Order	Object
Primary color	Object
MRP int64	Object
product Discounted?	Object
Product Discounted %	Object
Account object	Object
New order No.	Object
Customer_ID	Int64
State_City	Object
Final state	Object
Final City	Object

Table 7.1 Attributes and Data types

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.

Data that is considered for this project has around 37453 transactions that include purchases made by different customers from across the country. This data will be prepared as part of data mining in terms of cleanup, getting rid of duplicate records if any, removal of null value rows, etc. to make it meaningful for the purpose of our analysis.

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.

Equation 7.1 CLTV Formula

Columns/attributes from the above table will be used to come up with new fields as per the formula requirement.

As the understanding of the data which is a key step is completed along with the basic details of the data like the attributes, their data types, etc., then next activity is preparing the data.

Chapter 8: 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.

For our Lifetime value calculation, we need to make use of the existing columns and calculate the parameters as per the formula to get the target variable - CLTV. Below are the key parameters that we need to calculate:

t = this is value derived from the difference of recent order date and first order date
Sum of Product Discounted – This is based on the discount rate
Acquisition Cost = 0 (0 has been assumed as this is an online retailer)
Churn Propensity which is used to calculate Alive Propensity
Alive Propensity is 1 – Churn Propensity

To be able to calculate these values, we need to prepare the data as below:

Here is a snapshot of the top 5 rows of the data after constructing the dataset as needed:

	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

Figure8.1 First 5 rows

Checking for missing values in the data as part of preparing for data modeling.

Below is a snapshot of the data preparation step to check for missing values based on Customer_ID feature.

	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

Figure 8.2 Missing Values Check

As we can see there are no missing values in the data based on Customer_ID field. Data looks to be clean and ready for the next step.

Descriptive statistics on the data is as below.

	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

Figure 8.3 Descriptive Statistics

Here we can see some of the key metrics of the data.

Further analysis on the data for details like the time range, total number of unique customers, total quantity sold, etc. we have the below snapshot.

```
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
```

Figure 8.4 Key Metrics

This shows that data is all set for the next step where we can model the data for calculating CLTV using the Logistics Regression machine learning algorithm.

Data preparation is the crux of the steps in the CRISP DM methodology as the data is thoroughly examined and prepared for analysis and modelling. Once the data is prepped, data will be modelled as part of the next chapter.

Chapter 9: Modeling

Modeling the data to predict CLTV is the agenda of this chapter, the data prepared as part of the previous chapter is now being modeling using machine learning.

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 a supervised learning algorithm used when the target (dependent) variable is categorical. It is used to predict a binary (yes/no) event occurring.

In our case, 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.

Once we classify the customers to two groups as Churned and Not Churned, we predict the CLTV for all those customers that are classified as Not Churned.

Below screenshot shows the total number of transactions considered for prediction of CLTV.

Also, 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

Figure 9.1 Top 5 Rows of CLTV

Below is the box plot of the CLTVs predicted using the formula:

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

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

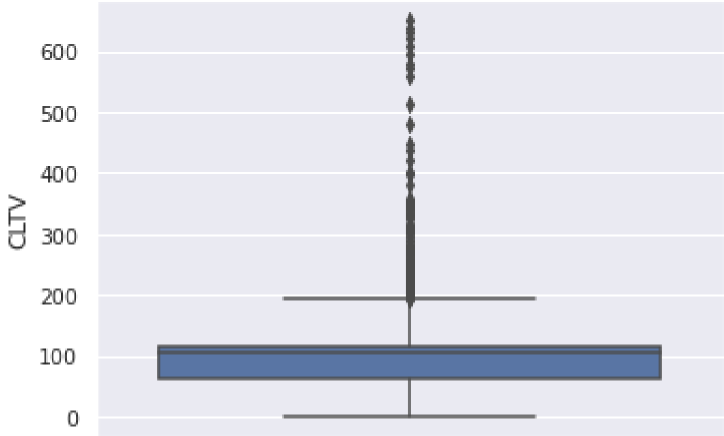

Figure 9.2 Boxplot of CLTV values

As the data is modeled using Logistic Regression machine learning algorithm, the CLTVs are being predicted or calculated. The next chapter runs through the model results to evaluate them.

Chapter 10: Model Evaluation

Model built as part of the previous chapter are being evaluated with the results to check on the predicted CLTVs.

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. The classification is self-explanatory.

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

Figure 10.1 Customer Classification based on CLTV

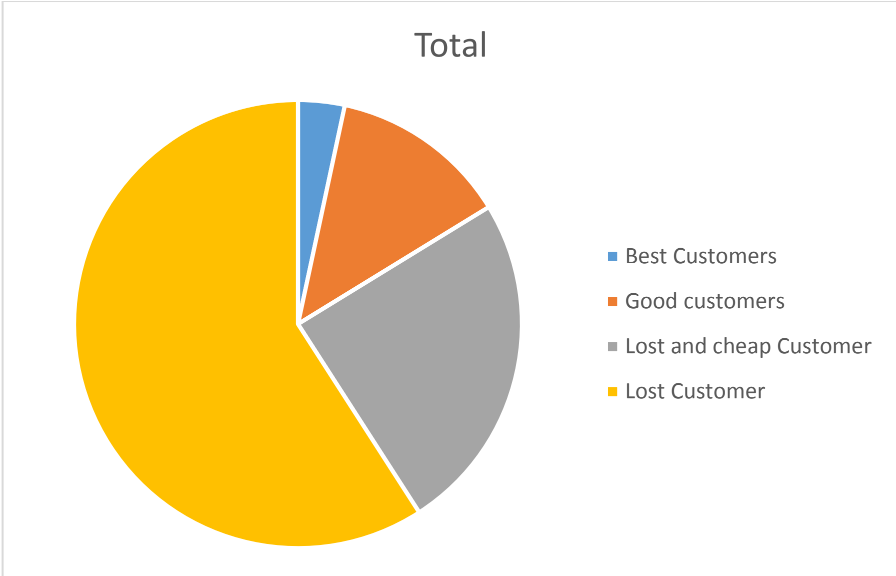

Figure 10.2 Customer Classification Chart

Based on this the retail firm can allocate budget for retaining the high value customers by offering good discounts.

As the evaluation of model is completed, the next chapter talks about the deployment process that needs to be carried out.

Chapter 11: Deployment

This chapter looks like into the deployment options and the details of deploying this solution.

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 Collab. 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.

The next chapter looks into analyzing the results of the model developed.

Chapter 12: Analysis and Results

Here in this chapter, the CLTV predicted is analyzed to derive the results.

The marketing team can now make use of the CLTV values to target high value customers or group of customers to drive more sales and profit.

Also, it is hard for the firms to reach individual customers. This calls for the 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.

In the next chapter, this analysis helps with the recommendations and conclusions.

Chapter 13: Conclusions and Recommendations for future work

This chapter concludes this exercise of predicting the CLTV and provides recommendations based on the CLTV.

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 more efficient ML algorithms to predict CLTV.

Customer value can be used to come up business plan that can be effective in terms of driving the value add from each of the customers along with providing room for scaling the business. The strategy defined by the marketing teams can indicate the profit extracted. In general, customer lifecycle is managed by the automation platform under marketing. Marketing campaigns are orchestrated by the marketing platforms which also automate the movement of leads and the client pipeline. The planning, coordination, execution, management and measurement of the campaigns and automation of various repetitive tasks are carried out by using these software applications.

This chapter concludes the project with providing the suggestions based on the predicted CLTVs in order to achieve the objectives.

Bibliography

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/

“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.

Appendix

Plagiarism Report[1]

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%

[1] Turnitn report to be attached from the University.