Spring 2021

Customer Lifetime Value Modelling

Shreya Sharma

Customer Lifetime Value Modelling

by

Shreya Sharma

A Creative Component submitted to the graduate faculty in fulfillment
of the requirements for the degree of
MASTER OF SCIENCE

Program of Study Committee:
Major Professor: Dr. Anthony M Townsend

Ivy College of Business
Iowa State University
Ames, Iowa
2021

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my Major Professor Dr. Anthony M Townsend, for his insights and support throughout my graduate studies and related research for the Creative Component.

I feel incredibly grateful to my parents for their love, affection, care, and sacrifices they made for me. I would like to extend my thanks to my friends, teachers, and peers for the timely help and encouragement.

ABSTRACT

In this era, when every organization competes to stay on the top in the market, organizations need to ensure that they should consider all the factors that will result in their long-term success. One of the most crucial factors among all is to provide the best customer experience. Customer Lifetime Value is an important factor that helps in understanding customers. It allows organizations to understand the importance level of each customer. By segmenting customers into different groups, analysts can build tailored strategies for customers. With data mining approaches, critical customer knowledge can be extracted, which could further help in critical decision-making. This paper aims to segment customers into groups, calculate customer lifetime value, and determine the best prediction model with maximum accuracy. The evaluation was carried out within customer segmentation, using a database of a company operating in the retail sector. The results indicated that developing prediction models by dividing CLTV into clusters is a better approach with a good accuracy rate and provided many beneficial insights.

1. INTRODUCTION:

CLV is one of the most critical metrics for any company, but they have ignored it. According to a UK study, it was found that **only 34% of the marketers they surveyed were "completely aware of the term and its connotations."** And only 24% of respondents felt their company was monitoring CLV effectively ("How to Optimize Customer Lifetime Value (CLV)—15 Effective Tactics That Every Marketer Needs to Know," 2020). Many start-ups also fail to establish themselves in the market because the cost of acquiring customers is higher than expected. Existing customers spend 67% more than new customers. So, encouraging repeat purchases is a good strategy to have in your marketing mix ("Fine-Tune Your SaaS Business with CLV Metrics", 2018). CLV can help businesses to make cost-effective and time-saving decisions. It will also allow companies to measure the financial impact of various activities like advertisements.

In the last few years, there is a massive transition in business as the whole business revolves around the customers. The approach that companies or organizations have started to use is the customer-centric approach. To sustain and maintain their value and simultaneously grow in the market, companies need to acquire new customers and retain their existing customers. But retaining existing customers is associated with a huge cost. This cost is called Customer Lifetime Value (also called CLV, CLTV, LCV). The customer lifetime value (CLV) is the total value customers bring to a business throughout their lifetime. This cost helps a business decide how much money they should invest in acquiring new customers and retaining their existing customers.

This paper intends to focus on the importance of CLV and build a CLV model to predict the cost, further determining how much money a business should spend on customer retention or how customer segmentation could be done based on value. This paper is organized as follows: Studies related to the customer lifetime value are provided in Section 2. Section 3 presents detailed information about customer lifetime value and methods to calculate it. Section 4 is the methodology used in the paper. These methods include cleaning and preparing the dataset for analysis, calculating customer lifetime values, customer segmentation, and building predictive models for CLTV. Section 5 and 6 of the paper consists of conclusions and references, respectively, from academic and practical points.

2. LITERATURE REVIEW:

As business strategies change every day and become more customer-centric, any company needs to understand their customers' needs and behavior. By understanding customers, a company can make strategies by which they can retain their customers and increase overall profitability. CLV is defined by (Kumar and Ranjan, 2009) "The sum of cumulated cash flows-discounted using the weighted average cost of capital (WACC)-of a customer over his or her entire life with the company."

Calculating CLV (Customer Lifetime Value) 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 (Ryals, 2002).

There have been many CLV models in the related literature that have been built. These models are the PCV Model (past customer value), RFM Model (recency, frequency, monetary), SOW model (share of wallet), and future- past customer behavior model.

As far as the future - past customer behavior models are concerned, although they all consider the future behavior of customers, some analytical models (Berger & Nasr, 1998; Gupta & Lehmann, 2003; Venkatesan & Kumar, 2004) include acquisition cost when calculating lifetime values while some others (Bauer, Hammerschmid & Braehler, 2003) do not (Hiziroglu et al., 2018). Most of the studies on future-past customer behavior models use retention rate to determine the activation period (Hiziroglu et al., 2018). Many studies use different methods, including generalized regression, logistic regression, quantile regression, latent class regression, CART, Markov chain modeling, neural network to create past customer behavior models (Haenlein, Kaplan & Beeser, 2007). The significant usage of data mining techniques provides advantages in the areas of modeling CLV, including performing analysis based on CLV and evaluating the optimal method for identifying customer lifetime value in many industries such as retail, insurance, banking, telecommunication, financial services (Kim et al., 2006; Chen et al., 2009; Khajvand & Tarokh, 2011; Lin et al., 2011; Golmah & Mirhashemi, 2012; Hu et al., 2013). These techniques include decision tree, clustering, logistic regression, artificial neural network, support vector machine, random forests (Hiziroglu et al., 2018).

There are other methodologies as well, where the proposed extended RFM analysis method with one additional parameter—called Count Item—is used. Comparing results of these approaches shows that adding count Item as a new parameter to the RFM method makes no difference to clustering result, so CLV is calculated based on the weighted RFM method for each segment. The results of calculated CLV for different segments can explain marketing and sales strategies by the company.

In this paper, I will be finding how customers are segmented between active and inactive customers. And finally, I will build machine learning models for customer lifetime value.

3. CUSTOMER LIFETIME VALUE IN DETAIL:

Customer Lifetime Value is how much amount a customer is predicted to spend on a particular business in their lifetime. It measures how valuable a customer is for a business. There are two most important factors for any business growth: acquiring new customers and retaining old customers. The cost of building new relationships(customers) is expensive compared to the cost of retaining customers. So, whenever a business focuses on retaining their old customers, they increase Customer Lifetime Value (CLV). That means the higher the CLV, the higher the loyalty of customers towards a particular business.

3.1. How to calculate CLV?

Many factors come into the role while calculating CLV, but majorly few factors affect it the most. These are customer lifespan, retention rate, customer churn rate, and the average profit margins. CLV can be of two types: Historic CLV or predictive CLV.

3.1.1. Historic CLV:

It uses historical transactions or events of an existing customer from a specific period. In simple words, it is the total profit from a customer's past purchases. To calculate the historical CLV, the following step needs to be done:
Historical CLV = (Sum of all transaction) * Average gross margin

3.1.2. Predictive CLV:

It is an algorithmic process that uses transaction history and behavioral patterns to determine a customer's current value and forecast how customer value will evolve with time. As customer's purchases and interactions increase with time, this value becomes more accurate and a better method to calculate CLV. This method considers customer service costs that include the cost of returns, acquisition costs, cost of marketing tools, etc.). There are five steps to calculate predictive CLTV. These are as follows:

Step 1: Calculating Average order value (AOV)
AOV is how much revenue is generated from the average customer. To calculate AOV, choose a time of business that is a good representation of customer purchase behavior. This time can be six months, one year, etc.
Average order value (AOV) = Total revenue/Total number of orders

Step 2: Average purchase frequency (APF)
It calculates how frequently a customer purchases from a particular business. The customer who buys regularly is considered a loyal customer. To calculate APF, determine total sales and unique customers in the time frame used for average order value. Unique customers refer to the customer who has frequently purchased in the selected period.
Average purchase frequency (APF) = Number of purchases/Number of unique customers

Step 3: Customer value (CV)

CV is one of the vital parameters as loyal customers may help advertise the business, which will further acquire new customers at a low cost. Customer Value tells us how much money is generated by the average customer over the selected period.
Customer value (CV) = Average order value * Average purchase frequency

Step 4: Average Customer Lifespan (ACL)
Customer's lifespan is the time between a customer's first purchase and their last purchase before they end their purchases from the business, i.e., end relationship with a business. The average of this timeframe across all customers can give the average customer lifespan.
Average customer lifespan (ACL) = Total of customer lifespans/Total customers

Another method to calculate average customer lifespan is by using Churn Rate. Churn Rate tells how many customers a business is losing.
Churn Rate = (Number of customers at the beginning of a time period-Customers at the end of a time period) / (Number of customers at the beginning of a time period)
So, Average customer lifespan = 1/Churn Rate

Step 5: Customer lifetime value (CLV)
After calculating the first four steps, the final step is to calculate Customer Lifetime Value.
Customer lifetime value (CLV) = Customer value * Average customer lifespan

Further, if we want to gain insights into each customer's profitability, gross margin needs to be included in the formula. Gross margin is revenue earned by a company after incurring all the costs associated with the business.
Gross margin = (Sum of revenue - Total cost of goods sold)/ Total Revenue

So, customer lifetime value when accounting for profit:
Customer lifetime value (CLV) = Customer value * Average customer lifespan * Gross margin

Note: Timeframe should remain the same while calculating each step.

4. METHODOLOGY:

4.1. DATASET:

There are several methods to calculate CLTV. To predict CLTV for both old and new customers, we can use machine learning methods. This paper has built a prediction model for existing customers using supervised and unsupervised machine learning methods.

For the dataset used in this paper, I have chosen an open-source online retail dataset. This dataset is a transactional dataset that contains all transactions occurring during a year for a UK-based and registered non-store online retail. I have downloaded this dataset from UCI Machine Learning Repository. This database contains information about the transactions done by 3921 unique customers. The response variable for this dataset is Customer Lifetime Value (CLV). There are a total 541909 number of instances with eight attributes. The information about the attributes are as follows:

4.1.1. Information About Attributes of Dataset:

Column Name	Datatype	Description
InvoiceNo	Object	Invoice number. A unique six-digit integral number assigned to each transaction. If InvoiceNo starts with the letter 'c,' it indicates a cancellation.
StockCode	Object	Product (item) number. A unique five-digit integral number assigned to each product or item to differentiate products from each other.
Description	Object	Name of product or an item.
Quantity	Integer	The number of quantities of each product (item) per transaction
InvoiceDate	Datetime64	Date and time of each generated transaction.
UnitPrice	Float64	Unit price. Numeric, Product price per unit in sterling.
CustomerID	Float64	It represents customer number. A unique 5-digit number assigned to each customer.
Country	Object	Name of the country where each customer resides.

Figure 1: Dataset Attributes

4.2. Data Loading:

Below is the subset of the data after loading the database used in this paper:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Figure 2: Subset of Dataset

4.3. Data Cleaning and Data-Preprocessing:

Data cleaning is converting the unstructured data to structured data. It is fixing or removing the duplicate values, missing values, incomplete values, and unnecessary columns within a dataset.

4.3.1. Remove Duplicates:

To make analysis more efficient and minimize distraction from the primary target and create a more manageable and more performant dataset, I removed all the duplicate and unnecessary values from the dataset. Initially, our database had 541909 rows. But out of 541909, there were 5268 duplicate rows. After removing duplicates, the final count of rows was 536641, and the number of columns was 8.

4.3.2. Removing Irrelevant and missing Values:

After examining the dataset, I found that there were many missing values, and the dataset contained negative values in the column Quantity. There was total 136538 missing and irrelevant data. That means 25% of data from our database is unstructured. These inconsistencies can cause mislabeled categories or classes. Hence, I removed these values from the database, resulting in 392732 rows and eight columns. Data is free from missing and duplicate value at this stage, but after preliminary data exploration, I found that most of the customerID belongs to the United Kingdom. Below bar graph shows the top ten country's customer.

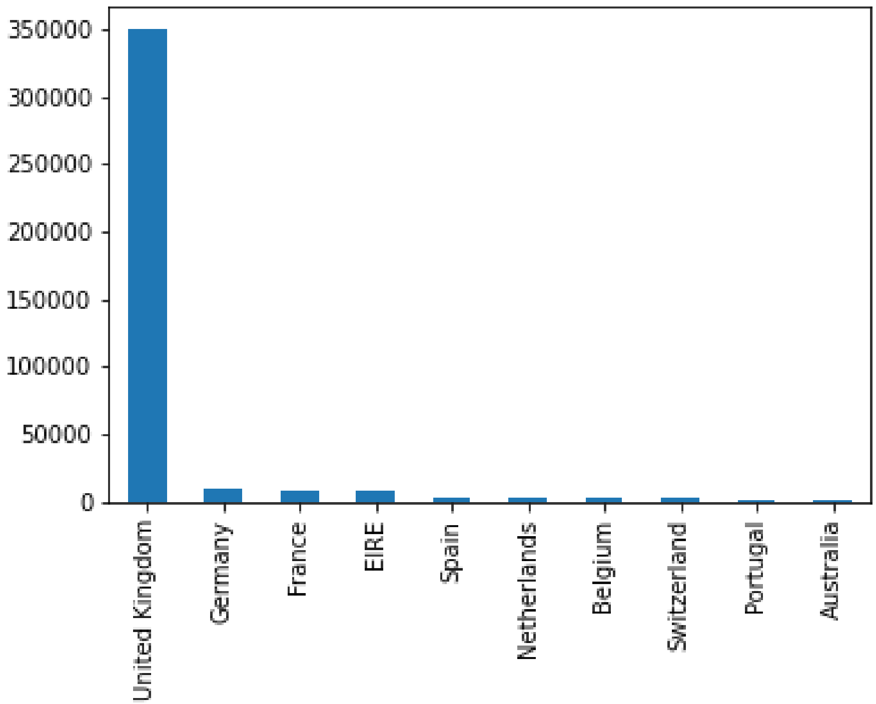

Figure 3: Top Ten Countries Customers

Out of 392732 rows, 349227 rows belonged to customers from the United Kingdom, so filtering out records that contain data only for United Kingdom customers. So, the final count of rows was 349227, and columns were 8.

4.3.3. Converting Datatype:

This dataset contains a column Invoice Date with datatype as datetime64. To make calculations and visualization easy, I have converted the datatype of this column as int. For this, I have extracted only the month and column from the InvoiceDate and renamed the column name as InvoiceYearMonth. The below figure represents the Invoice Date with updated datatype and name.

```
Retail_new['InvoiceYearMonth']

0         201012
1         201012
2         201012
3         201012
4         201012
           ...
541889    201112
541890    201112
541891    201112
541892    201112
541893    201112
Name: InvoiceYearMonth, Length: 349227, dtype: int64
```

Figure 4: InvoiceYearMonth

4.4. Feature Selection:

Feature selection is made to reduce the training time and increase the predictive model's accuracy; it is crucial to select only those features or columns that contribute the most towards the prediction. So, I removed StockCode, Description, and Country from the data frame.

4.5. Customer Segmentation:

Customer segmentation divides customers into segments or groups based on their common attributes to market to each group effectively and appropriately. It also helps in increasing the retention rate, which further helps to increase Customer Lifetime Value. One of the most useful segmentation techniques is calculating RFM, which stands for Recency, Frequency, and Monetary Value. I have segmented the customers into three groups. These three groups are as follows:

High: This group contains customers who were the most active and generated maximum revenue for the company.

Medium: This group contains customers who were less active and generated mediocre revenue for the company.

Low: This group contains the most inactive customers and generated the least or negative revenue.

4.5.1. Calculating RFM:

I have used unsupervised machine learning method - K-Means Clustering to calculate Recency, Frequency and Monetary Value.

4.5.1.1. Recency:

To calculate recency, I calculated each customer's most recent purchase date and then compared the last transaction of the dataset with the individual customer IDs' last transaction dates. Figure 5 represents the recency for five customers in our database.

	CustomerID	LastDateOfPurchase	Recency
0	12346.0	2011-01-18 10:01:00	325
1	12747.0	2011-12-07 14:34:00	1
2	12748.0	2011-12-09 12:20:00	0
3	12749.0	2011-12-06 09:56:00	3
4	12820.0	2011-12-06 15:12:00	2

Figure 5: Recency for five customers

4.5.1.1.1. Recency Cluster:

To calculate the recency score, I divided above calculated recency into clusters. To achieve that, I applied K-means clustering on my dataset for a range of values from 1 to 15. I further used Elbow Method to calculate the optimal number of clusters. After carefully analyzing the below graph, the value for the number of clusters is four.

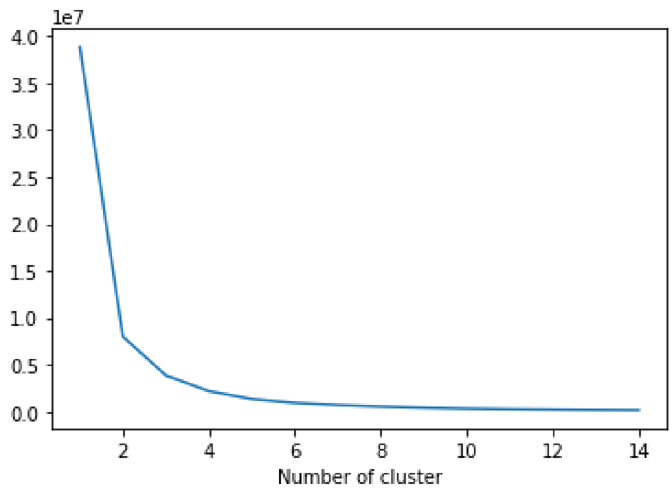

Figure 6: Elbow method to find recency clusters

Figure 7 shows number of recency clusters and information about those in detail. According to the figure, there are four clusters as 0,1,2 and 3. Cluster 0 has maximum mean recency, whereas cluster 3 has minimum mean recency that means cluster 0 contains the least active customers, and cluster 3 contains customers who are recent.

```
Retail_new_m.groupby('RecencyCluster')['Recency'].describe()
```

RecencyCluster	count	mean	std	min	25%	50%	75%	max
0	442.0	307.135747	39.591504	250.0	272.25	302.5	336.0	373.0
1	940.0	82.265957	24.205930	51.0	63.00	76.0	99.0	137.0
2	556.0	191.841727	31.690226	138.0	165.00	189.0	217.0	249.0
3	1983.0	19.063036	14.161087	0.0	7.00	17.0	30.0	50.0

Figure 7: Recency Cluster in Detail

4.5.1.2. Frequency:

Frequency is the total number of purchases or transactions performed by a customer. The below figure shows the frequency for five customers in our database.

	CustomerID	Frequency
0	12346.0	1
1	12747.0	103
2	12748.0	4413
3	12749.0	199
4	12820.0	59

Figure 8: Frequency for five customers

Calculating frequency can further help us to find out how many customers are infrequent or purchased the least. In this dataset, there were 35.68 % of the total customers have made transactions only once.

```
print("Percentage of customers who purchased only once:", one ,"%")
Percentage of customers who purchase the item only once: 35.68 %
```

Figure 9: One-time shopper

4.5.1.2.1. Frequency Cluster:

To calculate the frequency score, I again used K-mean Clustering. The method to find out clusters was like the method used to calculate the recency score.

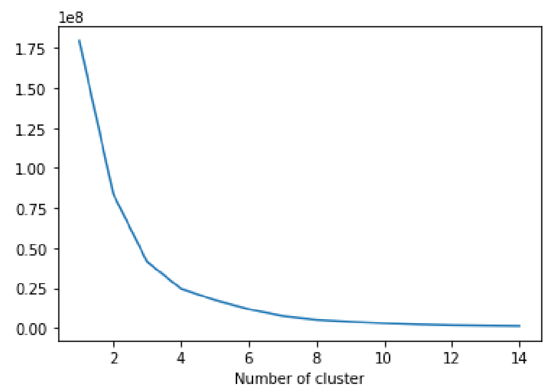

Figure 10: Elbow method to find frequency clusters

After observing the above graph, I can see that four should be the optimal number of clusters. Like the recency cluster, I have again divided frequency into four frequency clusters- 0,1,2 and 3. Cluster 3 has the maximum mean frequency, while cluster 0 has the minimum frequency. That means cluster three contains the most frequent customers, whereas cluster 0 contains the least frequent customers.

FrequencyCluster	count	mean	std	min	25%	50%	75%	max
0	3469.0	48.253675	43.411575	1.0	15.0	33.0	71.0	184.0
1	427.0	320.072600	130.103306	185.0	218.0	277.0	385.0	785.0
2	22.0	1271.090909	489.487284	828.0	951.0	1100.0	1423.5	2677.0
3	3.0	5733.333333	1718.216032	4413.0	4762.0	5111.0	6393.5	7676.0

Figure 11: Frequency Cluster in detail

4.5.1.3. Monetary Value (Total Revenue/Total Sales/Total Purchase):

To calculate the monetary value, I multiplied the Quantity column with the UnitPrice column. I named the resultant as TotalPurchasePerCust. The graph below shows the total purchase for five customers within our data frame.

	CustomerID	TotalPurchasePerCust
0	12346.0	77183.60
1	12747.0	4196.01
2	12748.0	33053.19
3	12749.0	4090.88
4	12820.0	942.34

Figure 12: Total Purchase for five customers

4.5.1.3.1. Total Purchase Cluster:

These clusters divided the database into four clusters 0,1,2, and 4 according to the amount of purchase made by a customer. Once again, I used K-mean clustering to calculate clusters to calculate. The method to find out clusters was similar to the method used to calculate the recency score.

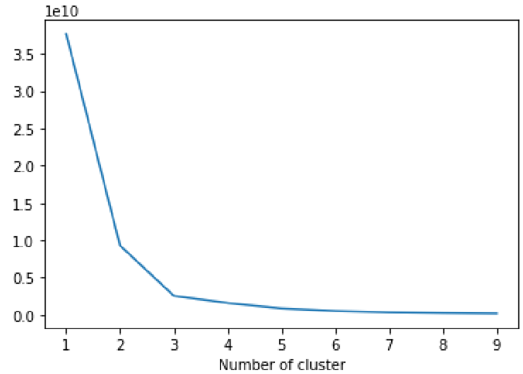

Figure 13: Elbow method to find Total Purchase Clusters

TotalPurchaseCluster	count	mean	std	min	25%	50%	75%	max
0	3715.0	998.859659	1049.577531	0.00	284.505	598.73	1341.2350	5298.48
1	180.0	9669.716389	5134.151793	5343.28	6286.135	7777.09	10854.1025	28882.44
2	23.0	52660.498696	16523.446096	31833.68	38535.175	51527.30	62836.1850	91062.38
3	3.0	207506.863333	46986.066791	168472.50	181431.645	194390.79	227024.0450	259657.30

Figure 14: Total Purchase Clusters in detail

From figure 14, I could infer that cluster 3 has maximum mean purchase and cluster 0 has minimum mean purchase. That means cluster 3 has customers who generated the maximum revenue, whereas cluster 0 has customers who generated the minimum revenue.

4.5.1.4. RFM Score:

To identify the best and worst customers in our dataset, I added recency, frequency, and total purchase. The higher the RFM score, the higher is the CLTV. The customers with the maximum Score were the best, whereas the customers with the lowest customer score were the worst.

Score	Recency	Frequency	TotalPurchasePerCust
1	306.370370	24.148148	381.842025
2	64.214286	45.178571	857.300000
3	56.770711	116.713389	2432.485562
4	264.765957	20.297872	427.257447
5	81.116667	51.831111	982.080869
6	66.862385	88.917431	3243.418716
7	89.250000	101.250000	1141.775000
8	76.210526	40.473684	807.371316

Figure 15: RFM Score

4.5.2. Segmenting the Customers:

The last step in customer segmentation is to divide customers among segments-High, Medium, and Low. To accomplish that, I have used RFM Score. So, the customers with Scores 0 and 1 will be part of the Low segment. Customers with Scores 2,3 and 4 will be part of the Medium segment. Customers whose Score is 5,6,7 and 8 will be part of the High segment. The figure shows customer segmentation for five customers.

	CustomerID	LastDateOfPurchase	Recency	RecencyCluster	Frequency	FrequencyCluster	TotalPurchaseCluster	TotalPurchasePerCust	Score	Segment
0	12346.0	2011-01-18 10:01:00	325	1	1	0	2	77183.60	3	Medium
1	12829.0	2011-01-07 11:13:00	336	1	11	0	0	293.00	1	Low
2	12831.0	2011-03-22 13:02:00	261	1	9	0	2	215.05	3	Medium
3	12834.0	2011-03-02 09:49:00	282	1	18	0	0	312.38	1	Low
4	12845.0	2011-03-17 13:34:00	266	1	27	0	0	354.09	1	Low

Figure 16: Customer Segmentation

4.6. Calculating CLTV:

4.6.1. Average Order Value:

To calculate the average order value, I divided the Total Purchase column by their respective InvoiceNo.

```
# Average Order Value
Retail_new_m['avg_order_value']=Retail_new['TotalPurchasePerCust']/Retail_new['InvoiceNo']
```

Figure 17: Calculation of Average Order Value

4.6.2. Average Purchase Frequency:

Average Purchase Frequency for this database was 89.065.

```
#Average Purchase Frequency
Avg_purchase_frequency=sum(Retail_new_m['Frequency'])/Retail_new_m.shape[0]
```

```
Avg_purchase_frequency
```

```
89.06579954093344
```

Figure 18: Calculation of Average Purchase Frequency

4.6.3. Customer Value:

To calculate the average order value, I multiplied the Average Order Value with the Average Purchase Frequency.

```
#Customer Value
Customer_Value= Retail_new_m['avg_order_value'] * Avg_purchase_frequency
```

```
Customer_Value
```

```
0       0.002541
1       0.003378
2       0.003653
3       0.003378
4       0.003378
```

Figure 19: Calculation of Customer Value

4.6.4. Customer Lifespan:

Calculation of customer lifespan is a stepwise process. First, I calculated the repeat rate, and with the value of the repeat rate, I found out the churn rate. The lower the churn rate, better the company's growth. Here, the churn rate was only 0.18. Lastly, I divided the churn rate by 1 to get the average customer lifespan which was 55.25.

4.6.4.1. Repeat Rate:

```
# Customer Lifespan
repeat_rate=Retail_new_m[Retail_new_m.Frequency > 1].shape[0]/Retail_new_m.shape[0]
```

```
repeat_rate
```

```
0.9818923743942872
```

Figure 20: Calculation of Repeat Rate

4.6.4.2. Churn Rate:

```
#Churn Rate
churn_rate=1-repeat_rate
```

```
churn_rate
```

```
0.018107625605712774
```

Figure 21: Calculation of Churn Rate

4.6.4.3. Average Customer Lifespan:

```
#Average customer lifespan
Avg_customer_lifespan = 1/ churn_rate
```

```
Avg_customer_lifespan
```

```
55.22535211267622
```

Figure 22: Calculation of Average Customer Lifespan

4.6.5. Customer Lifetime Value:

There are two ways to calculate CLTV. One is including the profit rate (Gross Margin), and the other is by excluding the profit rate (Gross Margin). To include Gross Margin in CLTV calculation, we can either choose some random Gross Margin value like .05% or calculate Gross Margin value using mathematical formulas. In this paper, I have calculated the Gross Margin value as mentioned in the figure 24. After looking at both the CLTV's, I can say that for this dataset, gross margin did not impact CLTV.

4.6.5.1. CLV Without considering Profit:

```
# Customer Lifetime Value Without Profit

Retail_CLV_Without_Profit['CLV_W/O_Profit']= Customer_Value * Avg_customer_lifespan
```

```
Retail_CLV_Without_Profit.head()
```

ID	Recency	RecencyCluster	Frequency	FrequencyCluster	TotalPurchasePerCust	TotalPurchaseCluster	Score	Segment	avg_order_value	CLV_W/O_Profit
.0	325	1	1	0	77183.60	1	2	Mid-Value	0.000029	0.140307
.0	21	0	82	0	42055.96	1	1	Low-Value	0.000038	0.186526
.0	181	2	3	0	39916.50	1	3	Mid-Value	0.000041	0.201749
.0	234	1	10	0	44534.30	1	2	Mid-Value	0.000038	0.186526
.0	6	0	130	0	56252.72	1	1	Low-Value	0.000038	0.186526

Figure 23: CLTV without profit

4.6.5.2. CLV with Profit:

```
Gross_Margin = (Retail_new_m['TotalPurchasePerCust'] - Retail_new_m['Frequency'])/Retail_new_m['TotalPurchasePerCust']
```

```
Gross_Margin
```

```
0       0.999987
1       0.998050
2       0.999925
3       0.999775
4       0.997689
          ...
3916    0.893324
3917    0.775470
3918    0.999982
3919    0.998272
3920    0.998340
```

Figure 24: Calculation of Gross Margin and resultant Gross Margin

```
CLV_try= Customer_Value * Avg_customer_lifespan * Gross_Margin
```

Figure 25: Calculation of CLTV with Gross Margin

CustomerID	Recency	RecencyCluster	Frequency	FrequencyCluster	TotalPurchasePerCust	Score	Segment	TotalPurchaseCluster	Gross_Margin	CLV-Profit
12346.0	325	1	1	0	77183.60	3	Medium	2	0.999987	0.140306
12931.0	21	0	82	0	42055.96	2	Low	2	0.998050	0.186163
15098.0	181	2	3	0	39916.50	4	Medium	2	0.999925	0.201734
15749.0	234	1	10	0	44534.30	3	Medium	2	0.999775	0.186484
15769.0	6	0	130	0	56252.72	2	Low	2	0.997689	0.186095

Figure 26: CLTV with Profit

4.7. Prediction Model:

For this paper, I have used two methods to build a prediction model for Customer Lifetime Value. The first method included building a linear regression model by taking data from the last six months and their respective CLTV. In the second method, I built machine learning models by dividing CLTV into different clusters. The following methods are mentioned in detail below:

4.7.1. Linear Regression Model:

4.7.1.1. Feature Selection:

I have extracted the sales in the last six months of a year, i.e., from June to December, and the sum of CLTV in those months.

4.7.1.2. Splitting Dataset:

I have used function train_test_split() to split the dataset into training and test set. The test size is 0.30; training size is 0.70, and random_state is 56.

4.7.1.3. Model Development:

To develop the model, I imported the Linear Regression module and created a Linear Regression object. Then, I fitted my training set model using the model.fit() function and performed prediction on the test set using predict() function.

```
#Linear Regression Model
# import model
from sklearn.linear_model import LinearRegression

# instantiate
linearreg = LinearRegression()

# fit the model to the training data
linearreg.fit(X_train, y_train)

# make predictions on the testing set
y_pred = linearreg.predict(X_test)
```

Figure 27: Linear Regression Model

4.7.1.4. Model Evaluation:

This is a regression model, so I have calculated Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and R-Square for the model. As R-Square's value is 0.906, close to 1, that suggest that this the model fits the dataset better.

```
from sklearn import metrics

# compute the R Square for model
print("R-Square:",metrics.r2_score(y_test, y_pred))
```

```
R-Square: 0.9068202503712575
```

```
# calculate MAE using scikit-learn
print("MAE:",metrics.mean_absolute_error(y_test,y_pred))

#calculate mean squared error
print("MSE",metrics.mean_squared_error(y_test, y_pred))
# compute the RMSE of our predictions
print("RMSE:",np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

```
MAE: 535.430761778359
MSE 1527857.3937410605
RMSE: 1236.0652870059334
```

Figure 28: Evaluation of Linear Regression Model

4.7.2. Machine Learning Models:

Like linear regression model, I have used transactions made in the last six months of the year and calculated the total amount generated in the last six months. I named this calculated column as Revenue_6Mo. To get more insights on the CLTV, I divided CLTV into four clusters by using K-means clustering. These four clusters were 0,1,2, and 3. The figure represents the last six months' data with CLTV clusters and Revenue_6Mo.

CustomerID	Recency	RecencyCluster	Frequency	FrequencyCluster	TotalPurchaseCluster	TotalPurchasePerCust	Segment	Revenue_6Mo	CLTVCluster
17340.0	29	0	407	3	0	12353.29	Medium	6550.48	2
17381.0	8	0	109	3	0	20275.61	Medium	6149.32	2
17396.0	38	0	27	3	0	7330.80	Medium	6933.30	2
17581.0	0	0	439	3	0	11025.24	Medium	6621.00	2
17706.0	3	0	382	3	0	10504.49	Medium	8839.80	2
17735.0	1	0	690	3	3	13110.02	High	7197.41	2

Figure 29: CLTV Clusters and Revenue generated in last six months

CLTVCluster	count	mean	std	min	25%	50%	75%	max
0	2770.0	253.873315	233.372652	0.00	0.00	213.755	409.775	795.12
1	833.0	1337.616002	395.125723	798.74	1000.50	1256.900	1631.770	2265.17
2	221.0	3199.297919	680.773653	2279.49	2607.25	3045.010	3641.820	4838.40
3	57.0	6495.665439	1178.552958	4904.36	5530.67	6354.670	7165.270	9112.68

Figure 30: CLTV Cluster in Detail

From above Figure 30, it can be observed that cluster three has a maximum mean value (6495.66) so, this is the best cluster, whereas cluster 0 has a minimum mean value (253.87), so this is the worst cluster.

4.7.2.1. Feature Engineering:

Some columns are categorical; I used get_dummies() to convert them into numerical values.

Next, I plotted a Correlation matrix after converting all variables to numerical values to know how columns are interrelated.

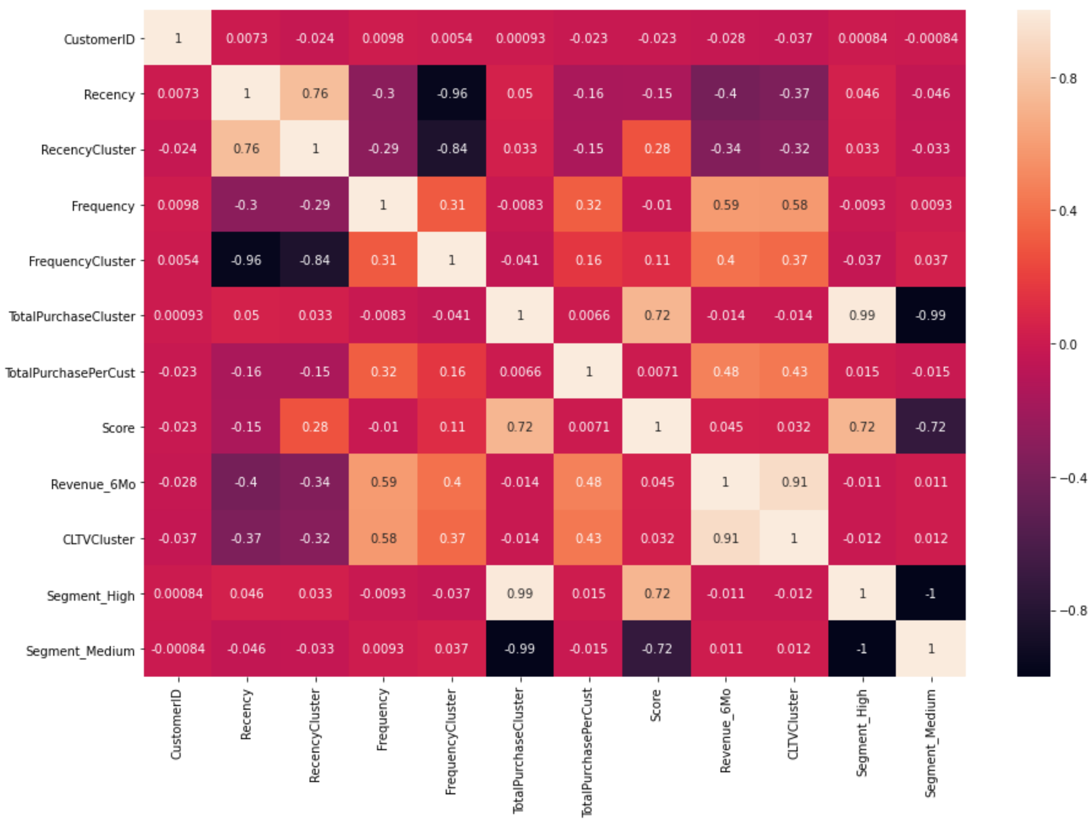

Figure 31: Correlation Matrix

4.7.2.2. Splitting Dataset:

To improve model performance, I split my dataset into training and test. The training set prepares the model, and the test set makes new predictions, from which I can evaluate the performance of the model. The test size is 0.20; training size is 0.80, and random_state is 2020.

4.7.2.3. Machine Learning Model Selection:

4.7.2.3.1. XGBoost:

It is a decision-tree-based ensemble Machine Learning algorithm and works very well when we encounter classification and regression predictive modeling problems. It gives high performance with good accuracy in less amount of time. I loaded the dataset to develop the model, and then I imported the xgboost package and used XGB Classifier to classify the model. The parameters I have used for XGB classifier were max_depth=5, learning_rate=0.1, n_jobs=-1. To fit the model, I used scikit-learn API and the model.fit() function.

4.7.2.3.2. Light Gradient Boosting Model (LGBM):

It is also based on the decision tree algorithm. It generates an efficient model because it uses discrete bins, which fastens the training process. Before building this model, I used MinMaxScaler() to transform the dataset and used function fit() to fit scaled training data. To build this model, I imported lightgbm package. I further imported LGBMClassifier from lightgbm package. The parameters passed in LGBMClassifier were n_estimators = 100, random_state = 2020, and learning_rate = 0.1. For the fourth parameter, num_leaves, I created an array using np.array with a range between 2 to 12.

4.7.2.3.3. Gradient Boosting:

Gradient Boosting trains models in a gradual, additive, and sequential manner while optimizing the differentiable loss functions. I imported GradientBoostingClassifier for this with parameters n_estimators=20, max_features=2, max_depth=2, random_state = 0. For learning_rate, I had chosen two values, 0.1 and 0.25. Then I used model.fit() function to fit my scaled training set.

4.7.2.4. Model Comparison:

4.7.2.4.1. XGBoost:

The figure shows the XGBoost classifier's accuracy on training and testing set.

```
print('Accuracy of XGB classifier on training set: {:.2f}'
       .format(Cltv_XGB_model.score(X_train, y_train)))
print('Accuracy of XGB classifier on test set: {:.2f}'
       .format(Cltv_XGB_model.score(X_test[X_train], y_test)))
```

```
Accuracy of XGB classifier on training set: 0.94
Accuracy of XGB classifier on test set: 0.73
```

Figure 32: Accuracy of XGB classifier

4.7.2.4.2. LGBM:

The figure shows the LGBM classifier's accuracy on training and testing set.

```
print('Accuracy of LGB classifier on training set: {:.2f}'
      .format(LGB_GS.score(X_train_sub, y_train_sub)))
print('Accuracy of LGB classifier on test set: {:.2f}'
      .format(LGB_GS.score(X_validation_sub, y_validation_sub)))
```

```
Accuracy of LGB classifier on training set: 0.70
Accuracy of LGB classifier on test set: 0.73
```

Figure 33: Accuracy of LGBM classifier

4.7.2.4.3. Gradient Boosting:

The figure shows the GB classifier's accuracy on training and testing set with two different learning rates.

```
print("Learning rate: ", learning_rate)
print("Accuracy score (training): {0:.3f}".format(gb.score(X_train_sub, y_train_sub)))
print("Accuracy score (validation): {0:.3f}".format(gb.score(X_validation_sub, y_validation_sub)))
print()
```

```
Learning rate:  0.1
Accuracy score (training): 0.878
Accuracy score (validation): 0.857

Learning rate:  0.25
Accuracy score (training): 0.887
Accuracy score (validation): 0.860
```

Figure 34: Accuracy of GB classifier

5. CONCLUSIONS:

Companies can use CLTV models to determine the characteristics of their customers. Moreover, customer segmentation could be carried out based on these value-based characteristics; organizations can develop appropriate strategies to support their decision-making processes in the customer relationship management context. This paper has successfully segmented customers among the three groups- Low, Medium, and High. Customers in the group High generate maximum revenue for the company; however, customers in the group Low generates minimum revenue. There are 35.68% of customers who purchased the product only once. Customers falling under cluster 3 in recency cluster, frequency cluster, and total purchase cluster could be promising customers, i.e., they can be retainable.

We have also built predictive models for CLTV. The Linear Regression model can be a good option because of the high R-squared value.

After building different machine learning models, we can conclude the following things:

- There is a strong correlation between CLTV and the last six months' revenue.
- CLTV cluster 3 is the best; customers falling under that cluster can retain retention.
- After comparing all the three models' accuracy, I could say that XGBoost and Gradient Boosting with a learning rate of 0.25 works well with our dataset. But the difference between training and testing accuracies in XGBoost is comparatively more than Gradient Boosting. So, I can conclude that Gradient Boosting is the best model because the lesser the difference between training and testing accuracy, the better the model is.

Between the method of model development, the first method where I built linear regression model, and the second method where I built different machine models, method two is better. The reason for this is the first model does not provide many insights on data. But the second method gives significant insights from the business point of view.

Finally, I can conclude that companies should focus on customer retention as most of the revenue was generated by frequent customers.

6. REFRENCES:

"A Guide to the Types of Machine Learning Algorithms." Retrieved From https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html.

Baker, Kristen. n.d. "The Ultimate Guide to Customer Segmentation: How to Organize Your Customers to Grow Better." Retrieved From https://blog.hubspot.com/service/customer-segmentation.

Bauer, H. H., Hammerschmidt, M., & Brähler, M. (n.d.). The Customer Lifetime Value Concept and its Contribution to Corporate Valuation. Yearbook of Marketing and Consumer Research, 47–67.

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. Journal of Interactive Marketing, 12(1), 17–30. https://doi.org/10.1002/(SICI)1520-6653(199824)12:1<17::AID-DIR3>3.0.CO;2-K

Chen, C.-W., Yang, C., & Lin, C.-S. (2009). A Study of Discovering Customer Value for CRM :Integrating Customer Lifetime Value Analysis and Data.

https://doi.org/10.30136/ZXYGLKX.200907.0002

Chugh, A. (2020, December 8). "MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric Is Better?" Analytics Vidhya | Medium. Retrieved From https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e

"Elbow Method — Yellowbrick v1.3.Post1 Documentation." Retrieved From https://www.scikit-yb.org/en/latest/api/cluster/elbow.html.

Feiz, S., Ramezani Ghotbabadi, A., & Khalifah, Z. (2016). Customer Lifetime Value in Organizations. Asian Journal of Research in Social Sciences and Humanities, 6, 53. https://doi.org/10.5958/2249-7315.2016.00103.9

"Fine-Tune Your SaaS Business with CLV Metrics." SaaSOptics. (2018, May 28). Retrieved From https://saasoptics.com/blog/fine-tune-your-saas-business-with-clv-metrics/

Golmah, V., & Mirhashemi, G. (2012). Implementing A Data Mining Solution To Customer Segmentation For Decayable Products—A Case Study For A Textile Firm. *Undefined*. /paper/Implementing-A-Data-Mining-Solution-To-Customer-For-Golmah-Mirhashemi/6f868fff8a644a607f977ecf70cb80fb481de009

Gupta, S., & Lehmann, D. R. (2003). Customers as assets. Journal of Interactive Marketing, 17(1), 9–24. https://doi.org/10.1002/dir.10045

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., & Sriram, S. (2006). Modeling Customer Lifetime Value. Journal of Service Research, 9(2), 139–155. https://doi.org/10.1177/1094670506293810

Haenlein, M., Kaplan, A., & Beeser, A. (2007). A Model to Determine Customer Lifetime Value in a Retail Banking Context. European Management Journal, 25, 221–234. https://doi.org/10.1016/j.emj.2007.01.004

Hiziroglu, A., Sisci, M., Cebeci, H. I., & Seymen, O. F. (2018). An Empirical Assessment of Customer Lifetime Value Models within Data Mining. Baltic Journal of Modern Computing, 6(4). https://doi.org/10.22364/bjmc.2018.6.4.08

"How to Optimize Customer Lifetime Value (CLV)—15 Effective Tactics That Every Marketer Needs to Know." Insider. (2020, May 14). Retrieved From https://useinsider.com/how-to-optimize-customer-lifetime-value-clv-15-effective-tactics-that-every-marketer-needs-to-know/

Hu, Y.-H., Huang, T. C.-K., & Kao, Y.-H. (2013). Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. *Journal of Systems and Software*, *86*(3), 779–788. https://doi.org/10.1016/j.jss.2012.11.016

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, *31*(1), 101–107. https://doi.org/10.1016/j.eswa.2005.09.004

Khajvand, M., & Tarokh, M. J. (2011). Analyzing Customer Segmentation Based on Customer Value Components (Case Study: A Private Bank) (Technical note). *Undefined*. /paper/Analyzing-Customer-Segmentation-Based-on-Customer-A-Khajvand-Tarokh/42694f5f4840c251476ff210e2bf795e9b7fe315

Kumar, V. and Rajan, B. (2009). Profitable Customer Management: Measuring and Maximizing Customer Lifetime Value. Management Accounting Quarterly. 10 (3), 1-18.

Lin, C. C., & Shih, D. H. (2011). Data Mining Techniques to Enhance Customer Lifetime Value. *Advanced Materials Research*, *225–226*, 3–7. https://doi.org/10.4028/www.scientific.net/AMR.225-226.3

"Linear Regression." Retrieved From http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm.

"Predictive Modelling." *Wikipedia*. Retrieved From https://en.wikipedia.org/w/index.php?title=Predictive_modelling&oldid=1009799430

"Python API Reference — Xgboost 1.4.0-SNAPSHOT Documentation." Retrieved From https://xgboost.readthedocs.io/en/latest/python/python_api.html

"RFM (Market Research)." *Wikipedia*. Retrieved From https://en.wikipedia.org/w/index.php?title=RFM_(market_research)&oldid=1000773304

Ryals, L. (2002). Are your customers worth more than money? Journal of Retailing and Consumer Services, 9(5), 241–251. https://doi.org/10.1016/S0969-6989(02)00005-X

"Sklearn.Cluster.KMeans — Scikit-Learn 0.24.1 Documentation." Retrieved From https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html.

Solutions, BEPEC. (2020, April 16). "Predictive Analytics: Artificial Intelligence For Marketing And Growth." *BEPEC*. Retrieved From https://www.bepec.in/post/predictive-analytics

"UCI Machine Learning Repository." Retrieved From https://archive.ics.uci.edu/ml/index.php.

Venkatesan, R., & Kumar, V. (2004). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing Journal of Marketing*, *68*, 106–125. https://doi.org/10.1509/jmkg.68.4.106.42728