ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Bachelor Thesis Economie en Bedrijfseconomie

Title thesis: Integrating engagement metrics into historical revenue analysis for better customer profitability at Gites.com

Name student: Rico Koevoets
Student ID number: 585330
Supervisor: Mateo Merlo
Second assessor: Prof. Ying Gan
Date final version: 12/07/2024

Abstract:
This thesis investigates the relationship between customer engagement metrics and customer lifetime value for Gites.com, a digital holiday home rental platform. A regression analysis is conducted to explore the relationship between historical customer behaviours and characteristics and lifetime value. Our results show that engagement metrics like GuestTenure and RecencyOfLogin significantly influence lifetime value. It is found that recent engagement contributes to higher lifetime value. Additionally, after evaluating both CPA and CLV practises, a customer segment analysis is conducted. Further analysis show that Belgian and Swiss customers should be classified as high-value segments, while French customers are classified as a low-value segment. This study advances the scientific literature by showing the importance of incorporating engagement data in non-contractual, continuous customer relationships when looking at lifetime value.

Table of contents

1. Introduction

Since the launch of the World Wide Web into the public domain in 1993, the business landscape has experienced a significant increase in technological innovation. This digital transformation has provided companies with new perspectives on business development. The widespread adoption of digital platforms has further accelerated this shift. Digital platforms are capable of collecting extensive data on customer characteristics and their purchasing behaviours (Horngren, Datar and Rajan, 2012). Platforms benefit from the ongoing advancements in information technology, including cloud computing, in-memory databases, and big data analytical tools (Bharadwaj, El Sawy, Pavlou, and Venkatraman, 2013). Platforms are able to collect data on a customer's return history and click-through behaviour. The data that is collected can be used by marketers and accountants alike.

Gites.com is a digital holiday home rental platform that specialises in connecting travellers with the property owners of unique holiday homes across various regions of France. In 2023, Gites.com facilitated the booking of over 4,000 holiday homes. Over this same period of time they attracted nearly a million unique visitors to the site. For digital holiday home rental platforms like Gites to reach their objectives efficiently, effective allocation of resources and strategic decision-making are necessary. We believe that understanding the lifetime value of customers is instrumental in determining budgets for customer acquisition campaigns, for example. We will determine Lifetime Value (LTV) by looking at the historic customer revenue, analysing total revenues of individual customers. Lifetime value measures the long-term value of a customer to a company, taking into account not only the initial purchase but also repeated transactions. By accurately understanding lifetime value, companies like Gites.com can enhance their customer profitability analysis (CPA) practices.

CPA is a management accounting tool for identifying the most valuable customer segments and those that may be costing more than they contribute. We determine net profitability by analysing the revenues generated from a customer and the costs associated with acquiring, serving, and retaining that customer. CPA can identify the most and least profitable customers to inform customer segmentation. Companies can base segmentation not only on traditional criteria like demographics but also on profitability (Mulhern, 1999).
Customer Relationship Management (CRM) allows accountants to collect the relevant data to provide managers with reports on. CRM is based on the idea that companies can improve their

profits by identifying their most profitable customers and directing more marketing resources toward them. Firms will attempt to create long term relationships with their most profitable customers by doing so (Payne and Frow, 2005). It is possible for a customer to have a negative net profitability over time, when acquisition, retention and selling costs associated with servicing the customer exceed their revenue. Digital platforms should focus on only retaining their profitable customers (Gupta et al. 2006).

1.1 Motivation of the Research

To my knowledge, there is limited research on integrating CRM and accounting systems to enhance customer analysis. Most research focuses on the best practices for utilising data for a singular module of an ERP system. Furthermore, It is not known whether a regression-based model can accurately explain variance in customer behaviour over an extended period. In my opinion, understanding the correlation between historical customer engagement metrics and lifetime value is crucial for the strategic decision-making of digital platforms like Gites.com. Traditional lifetime value models often overlook these metrics, which can result in incomplete insights into customer value.

Digital platforms are capable of gathering detailed transactional data necessary for calculating lifetime value, often with additional data on personal characteristics. Businesses can use this information to examine past customer behaviors and uncover significant correlations with lifetime value. Traditional lifetime value models typically focus solely on historical transaction data and may neglect the importance of customer engagement activities. Analysing the relationships between engagement metrics, such as the length and frequency of customer interactions, and lifetime value can provide a more comprehensive understanding of the factors that influence customer value.

On the revenue side of the analysis, Gites considers only the rental income from bookings made as a direct result of a marketing campaign. This does not take into account recurring customers and misses out on the value of branding and customer relations which have been key to shareholder value in modern time (Amir, E. and Lev, B. 1996). This use of just historical data is outdated in my opinion, Gites should utilise the advantages that come along with being a platform. The digital nature of the platform allows data to be gathered on a very large scale. In my opinion, Gites.com needs a more clear picture of their customers' value which can be realised by a regression analysis. I propose that this analysis should be focused on providing a basis to segment customers. This segmentation is based on the historic value of the current customer base. In turn, this segmentation can be used in CPA to rationalise targeted marketing efforts, since Gites.com in essence is a marketing company. In the marketing and management accounting literature, CPA and CLV models have typically been studied separately. This research examines the strengths and limitations of both models from a combined perspective. I hope that insights from my analysis will aid in possible market expansion.

1.2 Research questions and hypotheses

The objective of this thesis is to improve understanding on how historical customer behaviours and characteristics correlate with their lifetime value. By examining past data, we aim to show patterns and associations that can inform strategic decision-making at Gites.com. This thesis builds upon Kumar et al. (2010) which discusses the importance of customer engagement in creating enduring value to the platform context. We attempt to establish a relationship between high lifetime value customers and high engagement with the firm.

This is expressed in my first hypothesis:
"Customer engagement metrics, like GuestTenure and RecencyofLogin, significantly correlate with the customer lifetime value for Gites"

My second hypothesis is as follows: "Gites.com can enhance customer segment profitability analyses to identify high value segments by using customer segmentation based on nationality."

My central research question is as follows:
"How can incorporating customer engagement metrics, such as GuestTenure and RecencyOfLogin, into regression analysis improve lifetime value prediction and enhance customer profitability analysis by segmenting for more effective customer acquisition and retention efforts."

1.3 Research Structure

This thesis will be structured as follows: First, we will conduct a literature review on the relevant scientific literature on lifetime value prediction, CRM, CPA and engagement. This is followed by the methodology, which includes a description of the data, the research design, data pre-processing, data stationary, descriptive statistics of variables used in the analysis, and the software used for model interpretation. Subsequently, we present and discuss the results of the models. The report concludes with the main findings of the research and suggestions for future work.

2. Literature review

In this report, we will be looking at significant correlations between customer characteristics and lifetime value. The practices of forecasting Customer Lifetime Value (CLV) will serve to inform us of potential variables which drive revenue over time. Guptaet al. (2006) describe that “Firms generally define CLV as the present value of all future profits obtained from a customer over his or her lifetime relationship with the firm”. CLV can be determined at individual or aggregate level. At the aggregate level, CLV is calculated for an entire customer segment or the entire customer base. While aggregate CLV fails to account for variations in individual lifetime values, dividing distinct groups can assist managers with identifying promising market segments for firm level strategy according to Kumar and George (2007). This method can be used to implement resource allocation across customer segments. Fader and Hardie (2009) stated that historic profitability is not the same as CLV and should not be confused. CLV is a method to forecast the value of new customers who have just made their first purchase and predicting the future value of the existing customer base.

2.1 Individual CLV

Businesses evaluate the projected revenue and expenses made while servicing each customer to calculate Individual CLV.

The most common equation to calculate individual CLV is:

$$\text{CLV}i = \sum_{t=i}^{T} \frac{(p_{it} - c_{it})\, r_{it}}{(1+d)^{t}} - AC_{i},$$

Each customer is represented by an index i. At any given time t , p_{it} represents the price paid by customer i, and c_{it} denotes the direct cost of servicing that customer. The term $(p_{it} - c_{it})$ is referred to as the customer margin at time t. r_{it}, the probability that customer i remains active at time t, is known as the customer retention rate. AC_{i} is the acquisition cost for customer i, and T is the forecast horizon used to estimate CLV.

The aforementioned model differs from how CLV is presented by Berger and Nasr (1998). In their model, the acquisition costs and the direct cost of servicing that customer are left out of the equation. This irregularity in definition can be solved by deriving the true meaning of value from finance and accounting theory as discussed by Pfeifer, Haskins and Conroy (2005). Pfeifer et al. (2005) define lifetime value as follows: “Customer Lifetime Value (CLV) is the present value of the future cash flows attributed to the customer relationship”. Excluding costs to service a single customer is in line with the practices of digital platforms. The automated nature of platforms ensures that there are practically no differences between the costs to service a customer. The dataset in this thesis does not include data on acquisition costs. These cost do vary per person as the acquisition can occur through very different channels, through word of mouth referral or a paid advertisement. Excluding acquisition cost does not take away from our analysis. Reinartz, and Kumar (2000) argue that correlation between lifetime value and customer characteristics can be analysed without taking acquisition cost into account.

The RFM method, which stands for recency, frequency, and monetary value, is the most commonly known approach for forecasting the profitability of a customer according to Fader, Hardie and Lee (2005). These parameters reflect customers' past purchases and are used in the CLV calculation. Solely relying on RFM metric is flawed because it relies completely on past customer behaviour, leading to outdated information for customer selection and resource allocation (Reinartz and Kumar, 2002). Furthermore, solely relying on RFM variables is unsuitable for a dataset where repeated purchases are scarce. RFM variables are essentially imperfect representatives of other customer characteristics. However, including RFM metrics into the regression analysis can improve explanatory power. In the case of Gites.com, repeated purchases are not guaranteed and often do not happen in a short amount of time. People often do not repeat holiday destinations year after year.

When considering CLV, two types of customer-company relationships should be considered: contractual and non-contractual (Reinartz and Kumar, 2000). A contractual relationship involves a legal agreement where the company knows precisely when a customer becomes inactive. In a contractual relationship, the company precisely knows when a customer becomes inactive. Additionally, customers can make both discrete purchases, occurring at specific times, and continuous purchases, occurring at any time. In our dataset, the customer-company relationship is non-contractual and continuous. This attribute of the dataset means that the exact moment of customer churn is unknown, we do not know when a customer will not be using the services of the company again as stated by Castéran, Meyer-Waarden, and Reinartz (2021).

2.2 Engagement metrics

Engagement metrics, such as duration of customer relationship and frequency of interactions, have gained recognition for enhancing predictive power in customer behaviour models (Kitchens, Dobolyi, Li, and Abbasi, 2018). This is indicative of a correlation between engagement and the lifetime value of a customer, as it is a signal of a customer's interest in a continuing relationship with the firm. This signal may be a link between customer engagement and customer satisfaction. Intuitively, a customer who is satisfied with the services and gets enjoyment from using the platform is more likely to engage with the platform. Reichheld and Teal (1996) argues that a customer's profitability increases over time, making tenured customers the most valuable. However, according to Reinartz and Kumar (2000), this pattern does not occur in situations where customers are not bound by long-term contracts. Instead, customer behaviour and profitability in non-contractual settings can be more unpredictable and do not necessarily follow the same upward trend. Traditional models focusing solely on transactional data thus may often miss the broader picture of customer interaction with the brand. Research indicates that the duration of a customer's relationship with the firm represents a crucial component of lifetime value (Kumar et al., 2010). Long-term customers often correspond to higher lifetime value as they are more likely to engage in positive behaviours that enhance their value to the firm, such as repeat purchases and advocacy through word-of-mouth (WOM) referrals as stated by Kumar, V., Petersen, and Leone (2010). Marketeers should be careful with trying to force engagement upon customers, however. Venkatesan and Kumar (2004) found that the effectiveness of marketing spend has diminishing returns on lifetime value. This is reiterated by (Fournier, Dobscha, and Mick, 1997) as they state that excessive marketing interaction may overwhelm a customer, and thereby potentially harm the customer relationship.

2.3 Customer Relationship Management (CRM)

Over the last years, a change in philosophy can be seen in many companies in service-oriented industries. Multiple studies in marketing highlight the shift in focus from focusing on individual transactions to gaining long-term relationships with high value consumers.

Customer Relationship Management (CRM) systems are instrumental in this shift, as they have been shown to improve customer satisfaction (Mithas, Krishnan and Fornell, 2005). The importance of CRM is not only building long term relationships with customers, but also targeting those who will generate the most profit for the firm in the long term (Reichheld & Kenny, 1990). CRM systems can collect detailed data on the various forms of engagement, which is crucial for further research on this topic. Customer loyalty can be established through satisfaction, which is crucial for strong customer retention (Kincaid, 2003). Digital platforms have to fight for these loyal customer as low switching costs are an industry characteristic. CRM vendors have embraced these studies to indicate the value of their products, suggesting that successful CRM implementation helps increase organisational effectiveness and efficiency in managing customer relationships. From a project management perspective, CRM systems implementation may even be linked to improved project completion metrics (Wachnik, 2017).

The long-term financial impact of CRM implementation for firms has been subject to scrutiny, however. Research shows that implementing CRM systems does not improve stock returns or profitability (Hendricks, Singhal and Stratham, 2007). These are discouraging results considering the broad sample of publicly traded firms and the documented effect on long-term stock price performance and profitability instead of immediate post-implementation success metrics like time, budget, and scope. More recent research however indicates that CRM might be on the right track. By a broader set of performance metrics, including improvements in sales, sales efficiency, operating margins, and accounts receivable collectability a study found that CRM system implementations lead to several operational benefits that do not directly translate to immediate stock market performance. (Haislip & Richardson, 2017). This hint at technological maturity may also be helped by a more rigorous matching process, ensuring that the control firms were similar to the treatment firms, providing a clearer comparison of the impact of CRM implementations. Specifically, Haislip and Richardson (2017) examine how CRM systems enhance customer satisfaction and improve communication between the firm and its customers, leading to better management of accounts receivable. The presented results indicate that making sales to a customer who is already happy with your firm supports the inclusion of customer engagement metrics in our lifetime value model.

2.4 Customer Profitability Analysis (CPA)

Customer Profitability is the difference between customer-related revenues and costs (Pfeifer et al., 2005). Customer profitability analysis is essentially the retrospective version of CLV analysis as both attempt to identify the most valuable customers to inform resource allocation decisions. The main insights offered in Customer Profitability Analysis (CPA) is split into two categories: it determines the profitability of each customer individually and examines how profitability is distributed across the entire customer base. This allows businesses to not only identify their most profitable customers but also understand the overall profitability landscape (Van Raaij, E. M. 2005). In customer accounting, CPA can also be conducted for a customer segment (Ward, 1992). Businesses can make use of individual and aggregated lifetime values of their customers in this equation. Customer profitability analysis is standard practice in market-oriented businesses where strategic positioning is key. This report will seek a bottom-up approach to CPA as the individual customer segment will be the subject of our analysis (Helgesen, Sandanger and Sandbekk, 2018). In order for customer segment analysis to be successful the characteristics of the segment must be observable and readily available. When a customer visits Gites.com one of the first pieces of information that can be gathered on this person is their nationality. McManus (2007) constructed a CPA analysis to discover if statistically significant changes to profitability could be explained by customers being located in different geographical location in Australia. Significant differences between profitability were determined when customers from metropolitan and urban areas were compared. These finding could translate to Europe, as European nations differ drastically in their makeup. Including the nationality of our analysis of individual lifetime value could prove valuable in discovering profitable customer segments when the data is aggregated.

Gites.com is limited in implementing CLV forecasting due to considerable variability in customer retention durations in their dataset. Holm, Kumar and Rohde (2012) develop an integrated approach to CPA and CLV where they propose sophisticated CLV models for such a company. However, it is previously established that low repeated purchases and significant amount of time between initial and repeated purchase severely limits forecasting ability. Considering this, we find an integrated approach to the solution. Holm et al. (2012) poses that lifetime value of customers can only be enhanced when the drivers behind CLV like "(retention probabilities, depth and breadth of engagements, and direct marketing investment requirements)" are the focal points of improvement for such a strategy. This reinforces our understanding that solely considering financial measures will not lead to a comprehensive understanding of the drivers of revenue for a firm.

2.5 Performance measures

Comparison of the baseline model and the enhanced is conducted by comparing the values of R-Squared, Mae and RSME. These figures will determine the model which provides the most explanatory power. This model will be chosen for further analysis.

2.5.1 R-Squared

R-squared (R^2) is a measure of goodness-of-fit for linear regression models (Cameron and Windmeijer, 1997). R-squared is included in reports as the coefficient of the determinant, which indicates the accuracy with which the explanatory variables can predict the dependent variable. It is important to note that R-squared should not be used to draw conclusions on causal links between the dependent variable and the independent variables. Usage of R-squared should be reserved for quantifying predictive ability of the model (Moksony and Heged, R. 1990). The R-squared is calculated by dividing the sum of squares by the total sum of squares. The sum of squares due to regression quantifies how effective the regression model is in fitting the data (Miles, 2005). The total sum of squares measures the overall variation in the observed data used in the regression analysis (Miles, 2005).

2.5.2 RMSE and MAE

Hodson (2022) states that “for a sample of n observations y (y_i, i = 1,2,..., n) and n corresponding model predictions $\widehat{y}$, the MAE and RMSE are:”

$$\text{RMSE}= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y})^2_i} \qquad \text{MAE}= \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}\right|^2$$

The root-mean-squared error (RMSE) and mean absolute error (MAE) are also commonly used statistical metrics. Hodson (2022) explains that lower values for RMSE and MAE indicate a better fit of the model, in the case of normal errors. RSME is found to be the better performing metric under normal errors, while MAE is more robust (Hodson, 2022).

3. Methodology

In this section, I will present the research methodology used in this paper. Our study employs a quantitative approach, and the following sections outline this methodology. This will provide an overview of the research design and details how data and models are utilised to analyse the correlations and associations between historical customer data and lifetime value. This approach is taken to seek significant patterns and relationships that can inform strategic decision-making at Gites.com.

3.1 Data description

The transactional data was directly exported from the Gites.com superadmin. Gites.com introduced the booking module in 2016, taking responsibility for the collection and payment of rental sums in exchange for a commission. Before this, hosts could place advertisements of their gite for an annual fee.

The first data set, called *transactions, consists* of transactional data on almost 20000 bookings, placed between 1 January 2016 and 1 June 2024.

The second data set, called *customers* consists, was exported from Campaign Monitor, the email marketing software that Gites.com uses. This data set contains personal information on approximately 16500 customers that placed a booking between 1 January 2016 and 1 June 2024.

For each booking in the *transactions* data set, the following attributes were recorded.

• Booking ID: The unique ID of an individual booking.

• Customer ID: The unique ID of the customer who placed the booking.

• LandcodeGuest: This abbreviation gives the landcode for the nationality of the guest who placed the order.

• Placed on : The date when the booking was completed by the payment of the deposit.

• Arrival date: The date that the guests will arrive at the vacation home.

• Number of nights: The number of nights the vacation homes were reserved.

• Number of guests: The number of persons making up the travel party.

• Rental sum: The rental sum calculated for the travel party and amount of nights, excluding discounts and additional requests.

• Commission basis: Total amount to be paid by the guest excluding booking fee and tourist tax.

• Commission percentage: Percentage of commission according to the host's subscription

• Customer ID verhuurder: The unique ID of the host whose vacation home was booked.
•Advert ID: The unique ID for the relevant advert of the vacation home.

For each booking in the *customers* data set, the following attributes were recorded.
• Customer ID: The same unique ID as mentioned in the *transactions* data set.
• Date Joined: The date on which the customer first used their email address on the site, either by creating an account or placing an order as a guest.
• Last login: The date the customer last logged in or used their guest account.
• Status: The status of the customer in the mailing system.

In order to conduct the regression analysis properly, a few steps were taken to customise the data set to my liking. I sorted the data to each unique customer ID. For each customer, I created the following variables:
•MonetaryValue: The total commission basis of bookings amassed by the customer. This variable accurately represents the basis for the revenue that a customer generates. Note that the actual amount of revenue is given by the actual commission, however this is also influenced by the commission percentage, which is not visible to the guests.
•AvgNumOfNights: The average number of nights were reserved by a customer per booking, rounded to the nearest integer.
•AvgNumOfGuests: The average number of persons making up the travel party per booking, rounded to the nearest integer.
•Recency: Time between the customer's first purchase and their last purchase, in months.
•Frequency: Number of purchases (transactions) a customer made.
•Time: Time between the customer's first purchase and the end of the observation period, in months.
•RecencyOfLogin: Time between the last login of a customer and the end of the observation period, in months
• Guest Tenure: Time between the creation of an account by a customer and the end of the observation period, in months

3.2 Data pre-processing

To prepare the sample for analysis, the customer and transaction datasets were merged using Customer ID as the key identifier. This creates our dataset that integrates both transactional and engagement information for each customer.

In the Gites.com dataset, a lot of customers have not made a repeated purchase. If we assume that a repeated purchase is still possible for these customers, then the value of Recency and Time is equal. This causes Recency and Time to be highly correlated, confirmed by a correlation matrix (Appendix A), potentially causing inconsistencies in the coefficients and p-values due to multicollinearity. This moved me to drop Recency from the analysis.

Before the analysis commences, we have to ensure that the data is clean and suitable for analysis. During checks on the engagement metric data some unusual discrepancies, namely RecencyOfLogin having greater values than GuestTenure. After some manual checks, I concluded that only customers with an “Active” status in the mailing systems should be included in the analysis. I excluded all observations from the *customers* data set that did not have an “Active” status and dropped all rows with missing engagement metrics from the sample. This resulted in 5,682 dropped observations. A few customers have logged into Gites.com between the end of the observation period and the time of the analysis, resulting in negative values for RecencyOfLogin. This appeared in very few observations and with very small negative values so these values were rounded to zero.

I also transformed the dependent variables “Monetary value” to its logarithmic scale. When the dependent variable is log-transformed, the coefficients of the dependent variables can be interpreted as the percentage change in the dependent variable for a one-unit change in the dependent variable. This change allows for a more intuitive interpretation.

In order to transform the categorical variable “LandcodeGuest” into a format suitable for regression analysis by creating dummy variables.

• NL: indicating a Dutch guest. •BE: indicating a Belgian guest, •CH: indicating a Swiss guest, •DE: indicating a German guest, •FR: indicating a French guest, •GB: indicating a guest from Great Britain, •IE: indicating an Irish guest. •MG: indicating a guest from a not previously mentioned country. The countries from MG were merged together, as they made up 0.55% of the sample size altogether. Grouping these observations together created a clearer analysis.

3.3 Research design

We use a train-test split methodology to evaluate the performance of the baseline and enhanced lifetime value models. First, the sample is split into two subsets: one for training the model and the other for testing. First, 80% of the data was allocated to the training set, while the remaining 20% was allocated to the testing set. This split was implemented using a randomization process to ensure unbiased data partitioning. This method of testing for correlation can be seen before in a lot of papers (Tan,Yang, Wu, Chen, and Zhao, 2021). The linear regression is set up using the STATA software.

The baseline regression model includes traditional financial metrics, customer demographics, and control variables such as Frequency, AvgNumOfNights, and AvgNumOfGuests to analyze their effects on MonetaryValue. The dependent variable for our regression analysis will be the log transformed variable “MonetaryValue”. For the traditional financial metrics, our choice of the “RFM” metrics is in line with current literature on lifetime value calculation. The baseline regression model includes several control variables to account for potential confounding factors. These control variables are Frequency (the number of transactions), AvgNumOfNights (the average number of nights booked per transaction), AvgNumOfGuests (the average number of guests per booking) and the dummy variables representing the categorical variable LandcodeGuest, which indicates different geographical regions of the customers. These variables help isolate the effect of RecencyOfLogin and GuestTenure on the dependent variable, MonetaryValue, because the control variables themselves have an expected significant effect on the size of MonetaryValue for each customer. The most common LandcodeGuest “NL” is chosen to serve as the reference category. All interpretations of the dummy variables should be done so in comparison to Dutch guests. The enhanced regression model is an extension of the baseline model, which incorporates engagement metrics. The chosen metrics are GuestTenure and RecencyOfLogin. In addition to the regression analysis, I constructed charts which could provide alternative explanations for the results. In order to identify which nationalities are more likely to return, a figure is constructed to illustrate how customers who made more than one booking (repeat customers) are distributed across different nationalities. Geographical characteristics of a customer could have a big impact on the type of holiday in France is preferred. To verify the assumption that these differences would be reflected in the control variables, I constructed a figure illustrating the average number of nights and average number of guests per stay by nationality.

4. Results

Table 4.1 Descriptive statistics of the regression model variables for the relationship between lifetime value (log-transformed monetary value) and customer characteristics and engagement metrics for Gites.com customers from 2016 to 2024.

Variable	Observations	Mean	Std. dev	Min	Max
Log Monetary value	10,900	7.0979	(0.7354)	1.9459	10.2692
Frequency	10,900	1.2290	(0.6045)	1	7
Time	10,900	28.0036	(17.3447)	.0334	91.5
AvgNumofNights	10,900	8.9712	(3.8719)	1	63
AvgNumOfGuests	10,900	3.7284	(2.3735)	1	21
County Codes	10,900				
NL	7,897	.7245	(0.4468)	0	1
BE	1,803	.1654	(0.3716)	0	1
CH	22	.0020	(0.0449)	0	1
DE	188	.0172	(0.1302)	0	1
FR	747	.2527	(0.2527)	0	1
GB	184	.1288	(0.1288)	0	1
IE	12	.0332	(0.0332)	0	1
MG	47	.0655	(0.0655)	0	1
RecencyOfLogin	10,900	18.8848	(19.4410)	0	111.9
GuestTenure	10,900	27.8033	(13.9087)	0	125.9

Note: This table presents descriptive statistics for the variables used in the analysis. The variables include log-transformed monetary value, frequency of transactions, time, average number of nights, average number of guests, RecencyOfLogin, GuestTenure and country codes. Standard deviations are in parentheses

Table 4.2

Linear regression results for the relationship between lifetime value (log-transformed monetary value) and customer characteristics and engagement metrics for Gites.com customers from 2016 to 2024.

	Baseline model		Enhanced Model	
Variable	Coefficient	Std	Coefficient	Std. err.
Frequency	0.5716***	(0.0080)	0.5328***	(0.0095)
Time	-0.0035***	(0.0003)	-0.0005	(0.0005)
AvgNumberofNights	0.0968***	(0.0012)	0.0958***	(0.0012)
AvgNumberOfGuests	0.1551***	(0.0020)	0.1544***	(0.0020)
BE	0.1042***	(0.01286)	0.1031***	(0.0128)
CH	0.5300***	(0.11183)	0.5437***	(0.1115)
DE	0.0417	(0.0361)	0.0403	(0.0360)
FR	-0.2734***	(0.0192)	-0.2906***	(0.0192)
GB	-0.0480	(0.0352)	-0.0499	(0.0350)
IE	0.04326	(0.1444)	0.0472	(0.1439)
MG	-0.0138	(0.0686)	-0.0112	(0.0684)
RecencyOfLogin			-0.0039***	(0.0005)
GuestTenure			-0.0006**	(0.0003)
Constant	5.0538***	(0.0185)	5.1191***	(0.0203)
R-squared	0.6518		0.6606	
RMSE	0.4325		0.4296	
MAE	0.3382		0.3340	

*Note: This table shows the coefficients, standard errors (in parentheses), and significance levels for the baseline and enhanced regression models. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.*

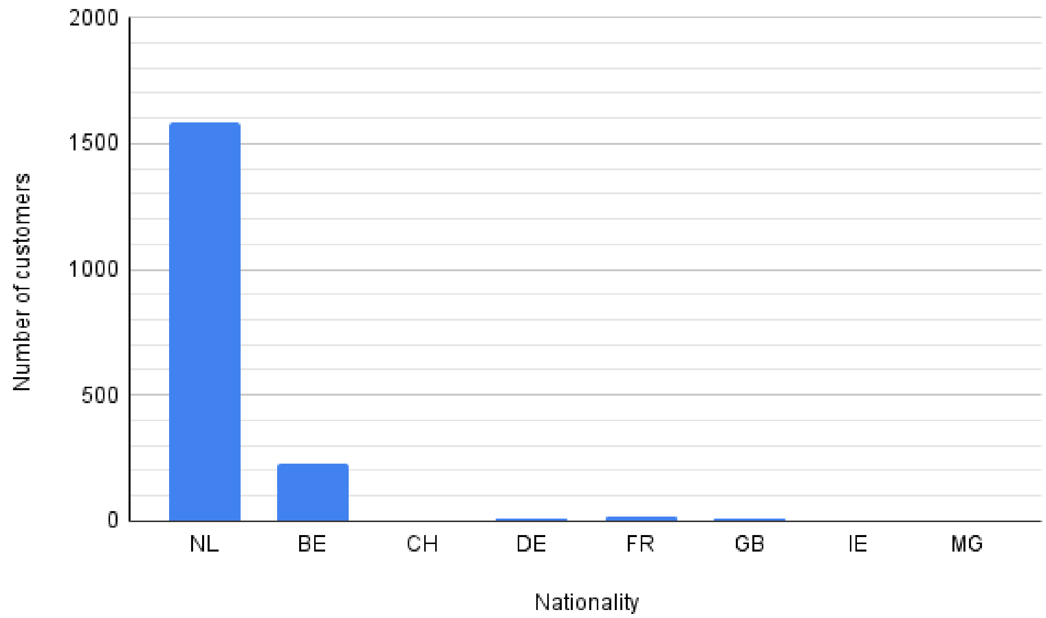

Figure 4.1 Number of repeat customers by nationality for Gites.com 2016-2024.

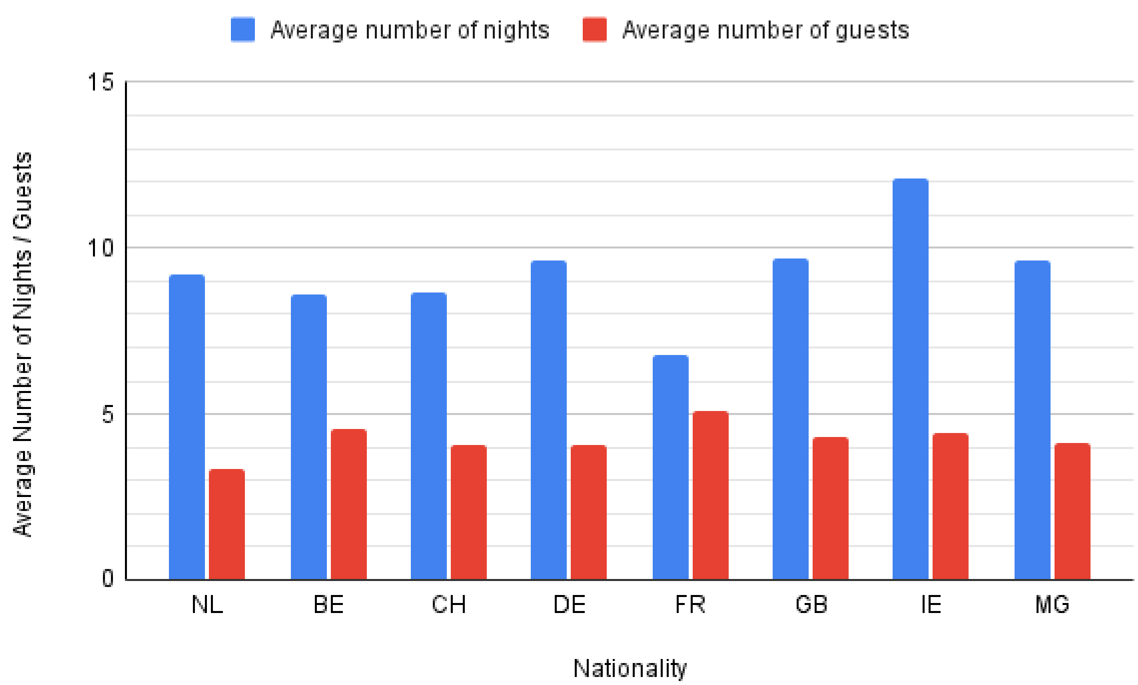

Figure 4.2 Average number of nights and average number of guests per stay by nationality for Gites.com 2016-2024.

4.1 Interpretation of results

In the following section, the results obtained from the regression analysis conducted on the Gites.com dataset will be presented and discussed. First, the baseline model and enhanced model will be compared to determine the most appropriate model to use in further lifetime value analysis and inform on segmentation.

4.1.1 R-squared

Both models have very close R-squared values (0.6518 vs. 0.6606), indicating similar explanatory power. The slight increase in the enhanced model suggests that the added engagement metrics provide minimal additional explanatory power.

4.1.2 RMSE and MAE

Both RMSE and MAE are slightly lower in the enhanced model, indicating a marginally better fit. Even though the improvement is statistically significant, it might not be practically significant given the complexity added by the new variables. The enhanced model explains a greater proportion of variance in the dependent variable. Incorporating engagement metrics allows for a better prediction of lifetime value. All further conclusions will be drawn from the results of the enhanced model.

4.1.3 Coefficient interpretation

We will focus the interpretation of the results from the enhanced regression model on the variables that have been found to significantly affect expected revenue. The constant coefficient in the enhanced regression model represents the value of log-transformed MonetaryValue when all independent variables are zero. In practice, the value of the constant of 5.1191 gives the expected baseline revenue of €167.50. This baseline serves as the reference point to understanding how the experimental variables impact the revenue. The negative coefficient of RecencyOfLogin indicates that the expected revenue from a customer decreases over time if the customer does not engage with the platform. For each additional month since the last login, the log of the monetary value decreases by 0.0039 units. This means that an additional month of inactivity leads to 0.39% decrease in expected revenue. The negative coefficient for GuestTenure indicates that the expected revenue declines as more time passes since the customer first joined. Conversely, if revenue per customer has been increasing over time, new customers are more valuable than older customers. The log of the monetary value decreases by approximately 0.0006 units every month for each customer. The decrease in the revenue for every month that the customer has been registered with the firm is approximately 0.063%. All things considered, the change of expected revenue over time is not very significant. However, from the results on the correlation between engagement metrics and expected revenue,we can still draw the conclusion that customers who have been recently active on the platform should be the main focus of retention and acquisition efforts. Keep in mind that making an account and engaging does not mean that a first booking has been placed.

The positive coefficient for Frequency indicates that the expected revenue increases as the number of transactions increases. For each additional transaction, the log of the monetary value increases by 0.5328 units. This translates to a 70.3% increase in revenue. This can be used to set an upper bound for the budget allocated to customer retention efforts. If Dutch guests have a 20% chance of making a repeated purchase, Gites.com can allocate up to 14.06% of the first order value (70% of the first order value multiplied by the 20% repeat purchase probability) for retention efforts. The expectation of the coefficient of frequency was higher than the observed coefficient. One would think that if a customer was happy with the provided service, the subsequent purchase would be the same if not larger. A possible explanation of would be that if a holiday was so great you want to return to France, it is imaginable that you would want to see a different part of the country. Since holiday budgets do not change drastically on a yearly basis, getting the most sightseeing out of this budget is making 2 smaller bookings at different locations. The positive coefficient for AvgNumofNights indicates that for each additional night a customer books, the log of the monetary value increases by 0.0958 units. This translates to a 10.05% increase in revenue per additional night booked. This confirms the intuition that longer stays result in higher revenues. The positive coefficient for AvgNumOfGuests also confirms the assumption that larger groups generate more revenue. For each additional guest, the log of the monetary value increases by 0.1544 units. This translates to a 16.7% increase in revenue per additional guest. The positive coefficient for BE indicates that Belgian guests contribute more revenue compared to Dutch guests. The log of the monetary value for Belgian guests is approximately 0.1031 units higher. This indicates an 10.86% increase in revenue for Belgian guests compared to Dutch guests. The positive coefficient for CH indicates that Swiss guests significantly contribute more revenue compared to Dutch guests. The log of the monetary value for Swiss guests is approximately

0.5437 units higher. This translates to an 72.2% increase in revenue for Swiss guests compared to Dutch guests. The negative coefficient for FR suggests French guests are a less valuable customer segment for Gites.com. The log of the monetary value being -0.2906 translates to French guests generating approximately 25.2% less expected compared to Dutch guests. This can be attributed to a multitude of possible factors. Since Gites.com specialises in facilitating booking for holiday homes in France, there is competition from platforms who offer the same services that have more brand recognition, like Gites-de-France. Families are probably also less likely to spend big on a holiday in their own country. Data from figure 3.2 supports this theory. On average, French customers spend the least number of nights in their accommodation. Coincidentally, French customers also have the largest travel company. Packing light with larger company points at French customers using Gites.com as a way to facilitate a weekend getaway with family and friends instead of an extended holiday. Another possible theory is that size restrictions in European cars and a larger travel company necessitates a shorter stay.

4.1.4 Practical recommendations for Gites.com

We derived these recommendations from the significant findings of the employed regression analysis. The marketeers at Gites.com should develop strategies to engage recently active customers, such as personalised communication and targeted promotions. Trying to re-activate customers who have not been engaging with the platform recently might still be a successful strategy, as the decline in expected revenue is minimal. The choice of which strategy should be the focal point of the customer retention strategy largely depends on the effectiveness and costs of campaigns to convert engaged potential customers and re-activate less engaged customers.

When only the geographical characteristics of potential customers are known, Belgian and Swiss guests should be prioritised, as they are likely to be included in the high-value customer segment. Figure 3.1 displays that, apart from Dutch guests, only Belgian guests have a notable history of repeated business. Figure 3.1 also displays the power of branding for Gites.com in the Netherlands. Around 20% of Dutch customers place more than 1 booking with Gites.com. Repeated business indicates customers being satisfied with the service. Satisfied customers are more likely to engage in word of mouth referring (Anderson, E. W. 1998).

5. Conclusions

The regression results show that both GuestTenure and RecencyOfLogin have statistically significant impacts on lifetime value. The negative coefficients found in the linear regression results indicate that longer periods of inactivity and extended customer tenure reduce revenue. We can conclude that recent engagement drives higher lifetime value. While the baseline model provides sufficient explanatory power, a slightly higher R-squared value and lower RMSE and MAE indicate that including engagement metrics results in a marginally better fit. Thus, engagement metrics should be considered when predicting lifetime value and customer behaviour. Our assumption that the number of nights and guests per booking drives revenue is confirmed by the significant positive relationship between AvgNumOfNights, AvgNumOfguests and revenue. Customer segmentation based on nationality proved that Belgian and Swiss guests are particularly high-value segments. French customers are found to be significantly lower value customers. Geographical information is part of most easily extracted information when potential customers interact with one of the touchpoints of Gites.com. This serves as a good starting point for improved acquisition efforts.

The results indicate that marketing efforts and resource allocation can be improved by Gites.com by targeting the high-value customer segments that were identified. In practice, targeting recently active customers and re-engaging inactive ones are likely to result in the very similar revenue that is generated per customer. Still, segmenting customers by engagement levels provides a framework for personalised marketing campaigns, proving that advice derived from lifetime value prediction and customer profitability analysis can be improved by taking engagement metrics into consideration.

5.1 Contributions to Society.

My interest in this topic started by thinking about the practical challenges managers face at businesses similar to Gites.com. Business owners want their strategies to be rational and founded in logic. The analysis was conducted on a dataset with limited repeated purchases, which can happen in real world businesses as situations with perfect information is rare. The exercise of dealing with an imperfect dataset and coming up with a scientifically correct solution motivated me. By incorporating engagement metrics (such as RecencyOfLogin and GuestTenure), it is demonstrated that conclusions supporting customer acquisition and retention efforts can be drawn from datasets with minimal repeat purchase behaviour.

Furthermore, we were able to develop segmentation to be used by Gites.com as a part of customer acquisition and retention efforts. We found that using this segmentation and taking into account engagement metrics is able to improve CPA practices at Gites.com

5.2 Contributions to Scientific Literature

This study contributes to the scientific literature on lifetime value by demonstrating that customer engagement data can compensate for the lack of frequent transactions and still provide a meaningful prediction of customer value. Our approach to the research design also mimics an integrated approach to CPA and CLV practices. This approach allows managers to be informed on promising market segments when reliable forecasting methods are not available. Most existing lifetime value models are designed for contractual settings where the end of the customer relationship is clear. This thesis searches for a solution to a lack of a model adapted to a situation where the customer-business relationship is non-contractual and continuous. It is concluded that more lifetime value models need to be developed that consider both transactional and engagement data. The positive impact of more recent engagement on lifetime value found in this thesis rationalises further research into the implementation of CRM systems with big data analytics. My thesis suggests that there is potential in these technologies being able to enhance customer value prediction and management.

5.3 Future research

The main limitation of the current study is that the dataset does not have a lot of observations with repeated purchases. Future research should gather more transactional data over a longer period or from a larger customer base. Conducting a longitudinal study may be the way to go about this. A longitudinal study can provide data into how customer engagement and transaction behaviour evolves over time. This can help in understanding the long-term impact of engagement strategies on lifetime value. Furthermore, this report is unable to make any analysis on the effects of acquisition costs. Future research could test for correlations between levels of acquisition costs and lifetime value. While this thesis' usage of the variables RecencyOfLogin and GuestTenure as engagement metrics may be sufficient to draw conclusions, future research could expand the range of engagement metrics. This can include social media interactions, customer service interactions, reviews and detailed usage patterns. In particular, I would be interested in the correlation between review scores and repeated purchases. Combining extensive transactional data with more detailed engagement metrics, which can be gathered by employing a more extensive CRM system, may allow future research to capture insights lost upon this study. Detailed customer segments based on engagement patterns which additional research can construct can lead to improved customer profitability analysis practices.

6. References

Amir, E., & Lev, B. (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of accounting and economics*, *22*(1-3), 3-30.

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of interactive marketing*, *12*(1), 17-30.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. V. (2013). Digital business strategy: toward a next generation of insights. *MIS quarterly*, 471-482.

Cameron, A. C., & Windmeijer, F. A. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of econometrics, 77(2), 329-342.

Castéran, H., Meyer-Waarden, L., & Reinartz, W. (2021). Modeling customer lifetime value, retention, and churn. In *Handbook of market research* (pp. 1001-1033). Cham: Springer International Publishing.

Fader, P. S., & Hardie, B. G. (2009). Probability models for customer-base analysis. *Journal of interactive marketing*, *23*(1), 61-69.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of marketing research*, *42*(4), 415-430.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N. al. (2006). Modeling customer lifetime value. *Journal of service research*, *9*(2), 139-155.

Haislip, J. Z., & Richardson, V. J. (2017). The effect of Customer Relationship Management systems on firm performance. *International Journal of Accounting Information Systems*, *27*, 16-29.

Helgesen, Ø. (2007). Customer accounting and customer profitability analysis for the order handling industry—A managerial accounting approach. *Industrial marketing management*, *36*(6), 757-769

Helgesen, Ø., Sandanger, H. M., & Sandbekk, J. (2018). Do customer profitability analyses pay? A survey of large Norwegian companies. *International Journal of Managerial and Financial Accounting*, *10*(4), 352-377.

Hendricks, K. B., Singhal, V. R., & Stratman, J. K. (2007). The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations. *Journal of operations management*, *25*(1), 65-82.

Kincaid, J. W. (2003). *Customer relationship management: getting it right!*. Prentice Hall Professional.

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Information Systems*, *35*(2), 540-574.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1-10.

Holm, M., Kumar, V., & Rohde, C. (2012). Measuring customer profitability in complex environments: an interdisciplinary contingency framework. *Journal of the Academy of Marketing Science*, *40*, 387-401.

Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or overvalued customers: Capturing total customer engagement value. *Journal of service research*, *13*(3), 297-310.

Kumar, V., & George, M. (2007). Measuring and maximizing customer equity: a critical analysis. *Journal of the Academy of Marketing Science*, 35, 157-171.

Kumar, V., Petersen, J. A., & Leone, R. P. (2010). Driving profitability by encouraging customer referrals:

Who, when, and how. *Journal of Marketing*, *74*(5), 1-17.
Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of marketing*, *80*(6), 36-68.
McManus, L. (2007). The construction of a segmental customer profitability analysis. *Journal of Applied Management Accounting Research*, *5*(2), 59-74.
Miles, J. (2005). R‐squared, adjusted R‐squared. *Encyclopedia of statistics in behavioral science*.
Mithas, S., Krishnan, M. S., & Fornell, C. (2005). Why do customer relationship management applications
Moksony, F., & Heged, R. (1990). Small is beautiful. The use and interpretation of R2 in social research. *Szociológiai Szemle, Special issue*, 130-138.
affect customer satisfaction?. *Journal of marketing*, *69*(4), 201-209.
Mulhern, F. J. (1999). Customer profitability analysis: Measurement, concentration, and research directions. *Journal of interactive marketing*, *13*(1), 25-40.
Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of marketing*, *69*(4), 167-176.

Pfeifer, P. E., Haskins, M. E., & Conroy, R. M. (2005). Customer lifetime value, customer profitability, and the treatment of acquisition spending. *Journal of managerial issues*, 11-25.
Reichheld, F. F., & Kenny, D. W. (1990). The hidden advantages of customer retention. *Journal of Retail Banking*, *12*(4), 19-24.
Reichheld, F. F., & Teal, T. (1996). The loyalty effect: The hidden force behind growth, profits and lasting. *Harvard Business School Publications, Boston*.
Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of marketing*, *64*(4), 17-35.
Tan, J., Yang, J., Wu, S., Chen, G., & Zhao, J. (2021). A critical look at the current train/test split in machine learning.
Van Raaij, E. M. (2005). The strategic value of customer profitability analysis. *Marketing Intelligence & Planning*, *23*(4), 372-381.
Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of marketing*, *68*(4), 106-125.
Wachnik, B. (2017). An analysis of ERP and CRM system implementations in Poland between 2013 and 2016. *Journal of Economics and Management*, (27), 134-149.
Ward, K. (2012). *Strategic management accounting*. Routledge.

7. Appendices

Appendix A.Correlation matrix of independent variables included in the enhanced linear regression.

Variables	Recency	Time	GuestTenure	RecencyOfLogin
Recency	1			
Time	0.8862	1		
GuestTenure	0.6699	0.7171	1	
RecencyOfLogin	0.7211	0.6705	0.5625	1