MASTER

Data-driven Drivers of Customer Loyalty in a Business-to-business Environment for the Software as a Service Industry

van Belle, E.P.J.

Award date:
2022

Eindhoven, February 2022

Data-driven Drivers of Customer Loyalty in a Business-to-business Environment for the Software as a Service Industry

Master Thesis

E.P.J. van Belle
Student identity number 1388088

In partial fulfillment of the requirements for the degree of

Master of Science

in Innovation Management

Supervisors Eindhoven University of Technology

Dr. N. Raassens	ITEM
Dr. S.E.C. Gelper	ITEM

Company Supervisors

P.L.M. Camps	Sales Operations
M.E.M. Bogman	E-Commerce

TU/e, Eindhoven University of Technology, School of Industrial Engineering
Master Thesis Innovation Management

Due to publication in the Eindhoven University of Technology library, the company where this graduation project was carried out is anonymized. From here on, the fictive name “Digidata” is used.

Keywords: customer loyalty, customer retention, cross-buying, software as a service, business-to-business, drivers, hierarchical logistic regression

Management Summary

Introduction

The software-as-a-service (SaaS) industry is growing at an increasing rate. Due to this growth, competitors have joined the industry, making competition fierce. Technology and information are more accessible than ever before, making it easy for customers to compare prices and functionalities and choose the best offer available. Furthermore, unsatisfied customers can easily end a contract and switch providers due to its subscription-based business model. Since churn levels over 20% are fatal for subscription-based services and winning new customers can cost up to six times more than to keep existing ones, customer loyalty is becoming an increasingly important topic for SaaS companies. Drivers of customer loyalty have been widely researched, but a recent study by Kocaman et al. (2020) shows that drivers can have a different effect in a SaaS environment. Furthermore, information about customer loyalty is usually gathered through surveys, which can be expensive and time-consuming, lead to skewed results, and make it challenging to derive implications. Fortunately, due to the contractual setting with SaaS companies, large amounts of transactional and behavioral data are continuously collected through SaaS platforms, which can be used for analytical purposes. Therefore, this study contributes to SaaS literature by exploring drivers of customer loyalty, about which, information is available purely on data instead of surveys.

Theoretical Background

To identify different drivers that influence customer loyalty, a systematic literature is performed. From this, four drivers are identified. The first driver of customer loyalty is platform value. Platform value in a B2B setting, is the degree of engagement on a service platform by users to fulfill functional social and/or hedonic objectives to ultimately enhance firm performance. When customers increase their number of interactions with the platform, they rely more on the services, knowledge and information that is provided by the platform. Furthermore, when services are important for a customer, there is a higher level of intention to do business with the service provider. It is, therefore, hypothesized that platform value has a positive effect on customer loyalty.

The following three drivers are categorized as hybrid drivers, meaning that information about them can either be gathered through a database or a survey. The first one is relationship value, namely, the personal relationship between people. Business relationships are established between organizations but are managed by individuals. An individual relationship can bring people closer together, and companies often try to create this relationship through account management, where individual account managers have the edge in building one-to-one relationships making customers feel attached to a company. Thus, it is hypothesized that relationship value has a positive effect on customer loyalty.

The second hybrid driver is communication. Communication between a customer and a company can enhance loyalty in multiple ways. It can mitigate conflict, improve service quality, and generate positive customer outcomes, but it can also function as a marketing channel in which products or services can be promoted and highlighted. Past promotions may have a long-term effect on customer purchases and create a positive association about a brand, ultimately enhancing customer loyalty. Therefore, the third hypothesis state that communication has a positive effect on customer loyalty.

The last driver analyzed in this study is social bonds, which can occur in many forms, such as attachment, commitment, trustworthiness, conflict, benevolence, equity, shared values, and common reference points, all of which enhance customer loyalty. On the opposite one finds dissimilarities, such as different working cultures, standards, values, and nationalities which can pose risks due to different backgrounds. Similarities on the interpersonal and inter-organizational level favor business, and parties' collaboration and stimulate goal seeking. Thus, it is hypothesized that social bonds has a positive effect on customer loyalty.

Lastly, the moderator relationship length was added to the theoretical framework since customers tend to stay longer with a company if they are satisfied, and a relationship can evolve from weak to robust over time increasing loyalty.

This study includes two dependent variables representing customer loyalty. The first is customer retention: customers who extend their contract with a company show a sign of commitment in terms of repurchasing a product or service. The second dependent variable is cross-buying: buying multiple products or services from a company indicates that a customer is willing to continue doing business. Furthermore, if a customer buys multiple products from a company, they greater attachment towards it. Both dependent variables are measures of behavioral loyalty.

Methodology

This study was structured according to the six steps of the cross-industry standard process for data mining (CRISP-DM) methodology, which is a life cycle of data mining projects. Then, the variables were defined in the theoretical background, extracted from a database, and implemented in a hierarchical logistic regression model to shed light on their and the moderator's predictive performance.

Results

The results showed a positive significant effect of platform value on both customer retention and cross-buying. Confirming that platform engagement enhances a customer's loyalty. However, a negative effect of the relationship between platform value cross-buying, moderated by relationship length was found. A possible explanation for this could be that, when a relationship ages, customers are more experienced and aware of the negative aspect of being dependable on an external service. Relationship value showed mixed results. A negative effect was found on customer retention, whereas a positive effect was found on cross-buying. The negative effect of relationship value on customer retention can be explained by the job description of an account manager by Digidata. The account manager is also the salesperson who closed the deal between Digidata and the customer, and is not trained for the actual job as an account manager. However, this person is trained to sell more products to the customer. Communication had a positive direct effect on both dependent variables. Nevertheless, when moderated by relationship length, communication only had a significant positive effect on customer retention. Overall, it can be concluded that marketing efforts, such as past promotions, are a way to increase loyalty. The results showed that social bonds do not affect either dependent variables. However, a positive effect was shown on the relationship between social bonds and cross-buying when moderated by relationship length. This can be explained by the similarity-attraction theory, whereby relationships with similar others maintains the balance and support relationships.

Discussion and Conclusion

If SaaS providers want to increase their customer loyalty, they should focus on customers by creating a certain urgency to use the provider's services. Selling multiple services increases platform value. However, companies should not overdo it since, over time, customers can become suspicious by feeling too dependable on a provider which can decreases loyalty. Building a relationship with one's customers enhances loyalty; however, companies must make sure that the person in charge can perform all competencies of an account manager. Communication through marketing efforts increases loyalty. Companies should promote their services to increases customer's lifetime, since it positively affects customer purchases. Independent variable social bonds showed no effect on customer loyalty because the current similarities are of no importance between a customer and a company. However, future research could focus on similarities that are more relevant to the current business culture, such as sustainability.

Preface

Dear Reader,

Before you lie my thesis, that is, my graduation project in completion of the Innovation Management master program at the Eindhoven University of Technology. This project was conducted at Digidata, which is a SaaS company in The Netherlands. This document will be the end of my study, and my time as a student. I am glad I made the decision to apply at the Eindhoven University of Technology, in which I have made new friends and grown personally and professionally. Before I welcome you to read my thesis, I would like to express my gratitude to several people.

First, I would like to thank my university supervisors, Dr. N. Raassens and Dr. S.E.C. Gelper. Néomie Raassens was always an excellent mentor when supporting me thought out the process. The feedback was always clear and she took the time to clarify aspects that were unclear to me. Furthermore, I would like to thank Sarah Gelper as second supervisor for her sharp and welcome feedback.

Second, I would like to thank my company supervisors Pim Camps and Manon Bogman, for the opportunity of writing my thesis at Digidata. Their direct and indirect support was always very helpful during the writing of my thesis.

Lastly, I would like to thank my parents for their continuous support and my friends, who made my time as a student unforgettable.

Thank you all and enjoy reading my thesis.

Etienne Pieter Julien van Belle

Eindhoven, February 2022

Contents

Management Summary ... iii
Preface ... vi
List of Tables ... ix
List of Figures ... x
1 Introduction ... 1
1.1 Research Background ... 1
1.2 Company Background and Context ... 2
1.3 Problem Statement ... 2
1.4 Research Question ... 4
1.5 Academic Relevance ... 5
1.6 Practical Relevance ... 6
1.7 Research Approach ... 6
1.8 Report Structure ... 6
2. Theoretical Background ... 7
2.1 Definition of Customer Loyalty ... 7
2.2 Measures of Customer Loyalty ... 8
2.3 Drivers of Customer Loyalty ... 9
2.3.1 Systematic Literature Review ... 9
2.3.2 Survey Drivers ... 13
2.3.3 Database Drivers ... 14
2.3.4 Hybrid Drivers ... 14
2.3.5 Moderator ... 16
2.4 Control Variables ... 17
2.5 Theoretical framework ... 17
3 Methodology ... 18
3.1 CRISP-DM ... 18
3.2 Hierarchical Logistic Regression ... 20
4. Quantitative Study: Hierarchical Logistic Regression ... 22
4.1 Data Understanding ... 22
4.1.1 Database Overview ... 22
4.1.2 Initial Collection of Data ... 23
4.1.3 Data Quality ... 23
4.2 Data Preparation ... 23
4.2.1 Operationalization ... 24

4.2.2 Data Cleaning ... 25
4.2.3 Data Conversion ... 26
4.2.4 Final Dataset ... 27
4.3 Modeling ... 28
4.3.1 Descriptive Statistics ... 29
4.3.2 Performance Evaluation Metrics ... 29
4.3.3 Results ... 30
4.4 Model Development ... 35
4.5 Findings ... 35
4.5.1 Customer Retention Findings ... 35
4.5.2 Cross-Buying Findings ... 36
4.5.3 Control Variables ... 38
5. Discussion ... 39
5.1 Theoretical Implications and Contributions ... 39
5.2 Managerial Implications ... 41
5.3 Limitations and Future Research ... 43
References ... 45
Appendices ... 53
Appendix A: Data Sources ... 53
Appendix B: Structural Literature Review ... 54
Appendix C: Pearson Correlation Coefficients ... 55
Appendix D: VIF Scores ... 56

List of Tables

Table 2.1: Identified articles from systematic literature review 9
Table 4.1: Database overview 22
Table 4.2: Overview of all variables in the different databases 23
Table 4.3: Overview of variables containing missing values 23
Table 4.4: Operationalization of the variables 24
Table 4 5: All variables used in this study 28
Table B.1: Journal Ranking List 54
Table C.2: Pearson correlation coefficients 55
Table D.1: VIF Scores 56

List of Figures

Figure 2.1: Theoretical framework of customer loyalty drivers 17
Figure 3.1: CRISP-DM cycle (Wirth and Hipp, 2000)………………………………………………………………….18
Figure 4.1: Number of customers per country………………………………………………………………………….27
Figure 4.2: Effect of platform value (A) and communication (B) on customer retention moderated by relationship length……..36
Figure 4.3: Effect of platform value (A) and social bonds (B) on customer retention moderated by relationship length……..37

1 Introduction

1.1 Research Background

The software-as-a-service (SaaS) industry is an increasingly growing market, with a total market value of 120 billion U.S. dollars in 2020 (The Business Research Company, 2021). Furthermore, the number of players within this market is increasing, and an increasing number of companies have infiltrated the SaaS industry and tried to use the market growth to their advantage (The Business Research Company, 2021). SaaS has been defined as “an application or service that is deployed from a centralized data center across a network, providing access and use on a recurring fee basis, where users normally rent the applications/services from a central provider” (Seethamraju, 2015, p.476).

With today’s technology, companies within the SaaS industry can offer their services in a more accessible and faster way than ever before (Figalist et al., 2020). Consequently, customers can easily compare prices and choose the best offer. Therefore, companies should distinguish themselves to retain their customers. According to Chen and Hitt (2002), a competitor is just one click away, whereas Lah and Woord (2016) state that churn levels over 20% are fatal for subscription-based services. Hence, the urgency for SaaS companies to transform their customers into loyal customers is even more important since it can cost up to six times more to win a new customer than to keep an existing one (Rosenberg et al., 1984).

Furthermore, companies want to know whether their customers are loyal; hence, they invest significant resources towards customer loyalty (Russo et al., 2016). However, constantly measuring customer loyalty is often time and cost-intensive or requires effort on the customer side (Figalist et al., 2020). As the construct of customer loyalty has been studied over the past couple of decades, practitioners within the field of marketing have discovered that customers express their devotion in many ways. Specifically, loyalty can be is industry, company, or brand-specific (Backman and Crompton, 1991; Day, 1969; Jacoby and Chestnut, 1978; Mägi, 1999; Narayandas, 1998; Pritchard and Howard, 1997).

The SaaS industry has some unique aspects (Mäkilä et al., 2010). Services are offered through the cloud and relationships are created in an online environment instead of a more natural way (Steinhoff, Arli and Weaven, 2019). Because of the online environment many different types of interactions and information between the customer and the SaaS platform are recorded and stored within a database. However, companies have difficulties analyzing this information to achieve customer loyalty (Ambulkar and Borkar, 2012), and even top firms are challenged when trying to determine the best “recipe” for customer loyalty (Russo et al., 2016).

1.2 Company Background and Context

Digidata is a service provider covering SMS messaging services. Over the years, it has broadened its SMS messaging services to communication through voice (messages/calls), e-mail, and other messaging apps (e.g., WhatsApp for Business, Apple Business Chat, Google RCS, WeChat, Telegram, and Viber) in the business-to-business (B2B) industry. Digidata has also acquired different products in its portfolio, such as innovative payment solutions, identity verification, and ticketing services, and it has integrated them within a customer data platform (CDP) to optimally arrange customer service and marketing communication channels. All these products can be grouped as SaaS. Specifically, Digidata positions itself as a SaaS company instead of a telecom company because it has expended its product portfolio beyond text messaging products. Digidata acts as a postman that delivers the message to the customer, in which it uses the network of network operators, while providing customers with the related technology.

Currently, 95% of Digidata's revenue is generated through the transactions of its core product (i.e., SMS messaging services), entailing that it is dependent on one source of income. A large portion of this revenue is generated when a customer, sends an SMS through Digidata's platform. Digidata then receives a fee for this transaction. The firm wants to focus more on its platform customers who have a subscription and thus generate monthly recurring revenue. Moreover, subscriptions within the SaaS platform have no intermediary from which the product is bought, which results in higher margins.

Digidata operates four systems to store and manage data: a local customer relationship management (CRM) system, which merges the other databases to provide a data warehouse, UNIT4 CODA, where the financial information is stored, Salesforce, where customer-specific data are stored, and a customer data platform, which stores marketing and customer platform activity. Appendix A provides an overview of all available databases within Digidata and their functionality. These systems offer Digidata the opportunity to gain insights into its customer base and manage its marketing and sales purposes.

1.3 Problem Statement

Customer loyalty is a company's most enduring asset since it helps develop a long-term, mutually beneficial relationship with its customers (Pan et al., 2012). Customer recommendations, preferences, and feedback have become key brand loyalty factors (Nyadzayo et al., 2018), especially in a SaaS setting where revenue models have become more dependent on the customer's success (Satyanarayana, 2012). Customers can now switch more easily from one vendor to another since new business models lower switching costs (Ojala, 2012). Furthermore, due to the straightforward access to information through the internet, customers can compare prices and pick their most favorable SaaS provider effortlessly. With increasing competitiveness, customer loyalty has become an even more critical aspect of SaaS organizations (Tvrdíková and Koubek, 2011). For SaaS companies, the percentage share of customer lifetime value is distributed more evenly over longer periods of time due to subscriptions can increase in value and lock-in customers due to usage growth, add-ons, and cross-selling over time (Pineda and Izaret, 2013). Furthermore, the duration of a customer relationship is a relational characteristic that is a critical determinant between customer satisfaction and customer loyalty (Bolton, 1998). Relational characteristics pertain to the relationship between a firm and its customers (Seiders et al., 2005). Balaji (2015) analyzes the core literature of relational characteristics by focusing on the moderating role of relationship length. Even though Bolton (1998) found a stronger link between satisfaction and retention for old customers compared to new customers, Digidata's focus is on acquiring new customers, whereas little is known about the loyalty of its current customers. For businesses to survive, organizations not only need to know how to attract new customers, but also how to retain their current customers (Bloemer and Kasper, 1994).

In sum, for SaaS companies, customer loyalty is a crucial metric (Javed and Cheema, 2017). It is much more expensive and time-consuming to acquire new businesses than to build upon an existing relationship (Singh and Khan, 2012). Furthermore, SaaS functionalities has become tremendously popular in recent years, and many companies, independent of their size started to adopt SaaS functionalities within their business (Goyal, 2014). Companies could therefore focus on customer loyalty, since loyal customers will most likely retain at a company (Coyles and Gokes, 2005).

In the literature, multiple drivers have been identified that represent characteristics of customer loyalty (Yuan et al., 2020; Pan et al., 2012; Lee and Bellman, 2008; Stock, 2005). A majority of them have been identified through customer surveys and without the use of the company's available data. The use of surveys is time-consuming, and results can be skewed (Peterson and Wilson, 1992). Thanks to the current information technology developments, SaaS companies can acquire large amounts of transactional and activity customer data (Achache et al., 2020) to develop a more targeted retaining strategy towards customer loyalty (Buckinx et al., 2007).

Accordingly, the following problem definition was formulated:

SaaS organizations encounter increasing competitiveness. Hence, they have shifted their focus towards retaining customers. To retain customers, they need to track their customer loyalty. However, analyzing customer loyalty can be a costly and time-consuming process. While SaaS companies extensively gather transactional and activity customer data, only a small part is extracted and used for analytical purposes. Therefore, organizations should explore more possibility of determining customer loyalty.

1.4 Research Question

This study sheds light on customer loyalty dimensions in a SaaS context without resorting to the active participation of customers (e.g., through the use of surveys). Therefore, the following research question was formulated:

What drivers that can be obtained without the active participation of customers through surveys determine customer loyalty within a business-to-business context?

Digidata never measured customer loyalty until recently, when it introduced the net promoter score (NPS) as a metric to assess feedback on how loyal its business customers are through the following question: “How likely are you to recommend Digidata to a friend or colleague?” Due to its simplicity and ease of measurement, many other companies within different industries have also adopted the NPS as their corporate loyalty metric (Gupta and Zeithaml, 2006). However, people are skeptical about the NPS because a single question cannot represent a multi-dimensional, high-level construct such as customer loyalty, especially in a B2B environment where a multidimensional approach to predict customer loyalty is to be preferred (Keiningham, et al., 2007; Wiesel et al., 2012). Therefore, a more detailed understanding of the construct of customer loyalty and its measurements is needed. Consequently, the following sub question was formulated:

Sub question 1: How can customer loyalty be measured?

After examining how customer loyalty can be measured, the drivers affecting customer loyalty must be identified. Common drivers identified in the literature are discussed in Chapter 2. Therefore, the following three sub question were formulated based on the division of data acquisition strategies:

Sub question 2: What survey drivers affect customer loyalty?
Sub question 3: What database drivers affect customer loyalty?
Sub question 4: What hybrid drivers affect customer loyalty?

Several studies investigated the moderating role of relationship length (Balaji, 2015). According to Balaji (2015), relationship length is the amount of time a customer and a firm have known each other. The relationship between a customer and an organization evolves with experience, and the length of a relationship has a moderating effect between platform value (He and Zhang, 2022), relationship value (Ulaga, 2003), communication (Dagger et al., 2009), social bonds (Cater and Zabkar, 2009) and customer loyalty. The effect of person-related characteristics such as similarity, empathy, and politeness on loyalty decreases as the length of a relationship increases. The last sub question is based on the moderating effect of relationship length between independent variables and dependent variable customer loyalty (Coulter and Coulter, 2002).

Sub question 5: What is the moderating effect of relationship length on the relationship between loyalty drivers that can be measured through a database and customer loyalty?

All findings from the main research question and sub questions are combined to develop a recommendation on how to measure customer loyalty in the SaaS industry with B2B customers.

1.5 Academic Relevance

Customer loyalty is the most enduring asset of a company (Pan et al., 2012). However, while marketing scholars and academic researchers have studied the concept of customer loyalty in many service industries (Cooil et al., 2007, Gustafsson et al., 2005, Reichheld and Teal, 1996), it remains one of the most challenging issues for firms in the modern era that is characterized by intense competition (Zeithaml et al., 1996). Several studies indicate that loyalty strategies are not equally effective across industries and firms and that different variables have distinct effects on customer loyalty (Haan et al., 2015; Eisenbeiss et al., 2014; Kumar et al., 2013; Rust et al., 2004). Kocaman et al. (2020) build on this argument by showing that drivers might have a different effect in a SaaS setting.

Furthermore, the standard technique to acquire customer loyalty information is field surveys (Hallowell, 1996; Ndubisi, 2007; Chen, 2015). To gain more insights into customer loyalty, it is important to consider context-specific characteristics, particularly in a B2B context where purchasing decisions tend to be rational and the end-users are often other people than the people who buy the product from the company (Patterson et al., 1996). Since multiple people are involved in a purchase decision, the relationship in a B2B setting is rather complex (Lages et al., 2008) and thus difficult to assess.

In sum, customer loyalty is a widely studied concept in the academic literature. However, there is a lack of research whether and how customer loyalty can be measured through databases. Organizations within the SaaS industry possesses a continuous stream of transactional and behavioral customer data through their provided software or platforms, thanks to the contractual setting. A new approach is proposed that examines drivers of customer loyalty using these customer data and has not been studied in the literature before. Also, as mentioned by Kocaman et al. (2020), drivers can have different effects in a SaaS setting. Thus, exploring drivers using data from a database will contribute to the SaaS literature.

1.6 Practical Relevance

An insight of this study is that while organizations spend a significant amount of their budget on marketing activities, many firms neglect to use their customer base as a marketing instrument. In particular, organizations within the SaaS industry gather a wide variety of data through their platforms. However, a large amount of the gathered data is not used for analytical purposes, while data can be used to analyze customer loyalty. Companies can focus on gathered data by the platform to gain more insights into customer loyalty, before using surveys that are expensive and time consuming. Behaviorally loyal customers are also thought to act as information channels who informally link friends, relatives, and other potential customers to the organization (McMullar and Gilmore, 2008).

1.7 Research Approach

The systematic literature review is a method in which a body of literature is aggregated, reviewed, and assessed while utilizing pre-specified and standard techniques to reduce bias (Torgerson et al., 2012). The literature reviewed in this report supports and sheds light on the current studies on the drivers of customer loyalty within the B2B industry. Based on the identified drivers, a theoretical framework was created that was tested during the empirical analysis. Specifically, the snowball sampling method was used to identify other potential drivers, and a hierarchical logistic regression was implemented.

1.8 Report Structure

The remaining chapters of this report are structured as follows. Chapter 2 presents the theoretical background concerning customer loyalty and a theoretical framework, whereas Chapter 3 elaborates on the methodologies used in this thesis. Chapter 4 provides the results, and Chapter 5 discusses them and elaborates the answers to the research questions by interpreting them. The chapter then illustrates this study's theoretical contributions and implications and managerial implications. Finally, it presents the thesis's limitations and suggests directions for future research.

2. Theoretical Background

First, the concept of customer loyalty is discussed to better understand its different aspects. Second, measures of customer loyalty are identified, and a combination of them is proposed to elaborate on the research question "How can customer loyalty be measured?" Third, drivers of customer loyalty mentioned in the literature are analyzed and categorized into survey drivers, database drivers, and hybrid drivers. Lastly, based on the literature, a conceptual model is shown.

2.1 Definition of Customer Loyalty

This study defines B2B customer loyalty consistent with prior literature, as a deeply held commitment to re-patronize a preferred product or service for a consistent time in the future, thereby causing repeat purchase even though marketing efforts and situational influences have potentials to induce switching (Oliver, 1999). Despite the high number of studies published about customer loyalty, according to Pan et al. (2012), there are several factors that limit the comprehensive understanding and prevent generalizations of research findings. First, a consensus is not often found in the accumulated research findings. For example, where some studies acknowledge the link between customer satisfaction and customer loyalty (Jones and Reynolds, 2006; Meuter et al., 2000), others do not find support for this relationship (Khatibi et al., 2002; Stoel et al., 2004). Second, the many inconsistencies are confounded due to the various research contexts of the previous research studies. Because of the diverse study conditions and different results in the literature, the relationship between various correlates and loyalty cannot be determined upfront (Pan et al., 2012). Building on this, Kocaman et al. (2020) showed that drivers may have different effects in the SaaS context compared to other environments. Lastly, within the loyalty literature, no agreement can be made upon conceptualizing and operationalizing the loyalty construct (Gil-Saura et al., 2008). However, according to Ramaseshan et al. (2013), customer loyalty in a B2B context is less likely to be derived from habits or routines and is more focused on behavioral facets such as purchase intentions. Furthermore, SaaS companies possess large amounts of transactional and activity data to analyze these behavioral facets (Achache et al., 2020). According to Tucker (1964), the behavior such as past purchases of a brand or product completely account for loyalty. Jacoby and Chestnut (1978) elaborate on this by pointing out that the focus in behavioral loyalty is on interpreting patterns of repeat purchasing in primarily panel data as a revelation of loyalty. In a more recent article by Rauyruen and Miller (2007), behavioral loyalty is defined as the willingness of average customers to repurchase the service or the product and maintain a relationship with the service provider or supplier. The main concern of behavioral loyalty studies is how repeat purchasing patterns and cross-buying are a manifestation of loyalty (Curtis, 2009). In the loyalty literature, attitudinal loyalty is another type of loyalty that one usually finds in the relationship marketing and business marketing (Rauyruen and Miller, 2007). However, it is still largely untested that attitudinal measures are abundant replacements for, or good supplements to behavioral measures (Pan et al., 2012). Also, attitudinal loyalty focuses on attitudinal concepts such as positive word of mouth, encouraging services to others, and recommending services to others (Rauyruen and Miller, 2007). Without the use of surveys, such as the Net Promoter Score (Reichheld, 2003), this type of loyalty is extremely difficult to measure. Since this study focuses on database drivers, attitudinal loyalty is outside of the scope of this study.

2.2 Measures of Customer Loyalty

While customer loyalty has become an ever more prevalent construct in multiple research fields and the knowledge on customer loyalty has increased tremendously, the measurement of customer loyalty remains heterogeneous (Wallenburg, 2009). Wallenburg (2009) reviewed recent empirical articles and showed that no standard measurements of customer loyalty have emerged. In line with Wallenburg (2009), this research bases customer loyalty on a broader view that incorporates purchases. Moreover, a further distinction for purchases can be made between retention and cross-buying. First, customer retention is prevalent. Customer retention refers to the ability of a company or product to retain its customers over a specific period (Vroman et al., 1996). Customers who extend their contract with a company, show a sign of specific loyal behavior in term of repurchasing and can be easily extracted from a database. Kuehn (1962) and Lipstein (1959) confirm this in their studies by using the probability of customer retention to measure loyalty. Second, cross-buying is defined as the customers' practice of buying additional products and services from its current provider (Ngobo, 2004). Cross-buying is an important loyalty measure since it is an indicator of a customer's willingness to continue doing business with a company. Reinartz et al. (2008) confirm this by stating that cross-buying is a main consequence of customer loyalty. Additionally, Liu and Wu (2007) state that customer loyalty has been traditionally viewed as a behavioral construct that includes cross-buying and customer retention. Thus, the second loyalty measure is cross-buying. Stone and Woods (2000) argue that cross-buying can build loyalty and retain customers. Cross-buying has been associated with higher levels of customer loyalty due to an increase in relationship duration, buying frequency, revenues, or share of wallet (Ramaseshan et al., 2017). The rationale behind this relation is that if a customer buys from different categories offered by the same company, it should experience greater attachment towards that firm. Also, it becomes more dependent on the supplier's products or services (Reinartz et al., 2008).

2.3 Drivers of Customer Loyalty

In this section a systematic literature review is conducted to give an overview of customer loyalty drivers within the B2B environment, and their loyalty measurement instruments. Afterwards, drivers are put into different categories that correspond with how information about the drivers is acquired. Since this research is focused on predicting customer loyalty through the solely use of SaaS databases, hypotheses are proposed for drivers where information can be acquired through Digidata's databases (see Appendix A).

2.3.1 Systematic Literature Review

To identify the drivers of customer loyalty, a systematic literature review was performed. Appendix B shows the systematic research approach in combination with snowball sampling that was applied to identify drivers of customer loyalty from previous studies. As a result, Table 2.1 was generated in which all identified relevant articles are summarized. The table also illustrates the key takeaways and the dependent and independent variables, as well as the mediators and moderators.

Table 2.1: Identified articles from systematic literature review

Reference	Nature of Research	Specifications	Findings	Journal	Data Collection Method
Stock (2005)	• Empirical study among salespeople in different industries to measure the impact of customer satisfaction on price sensitivity • Methodology: confirmatory factor analysis on a dyadic dataset	• DV: customer price sensitivity • IV: customer satisfaction • Moderators: product/service complexity, product/service specificity	• High customer satisfaction has a lower effect on price sensitivity. • The inverse relationship between customer satisfaction and price sensitivity is stronger in case of high as opposed to low product/service complexity. • The inverse relationship between customer satisfaction and price sensitivity is stronger in case of high as opposed to low product/service specificity.	Journal of Business-to-Business Marketing	Surveys
Lee and Bellman (2008)	• Empirical study among business customers for financial services to understand how the drivers of loyalty might be moderated by short- versus long-term relational orientation • Methodology: partial least squares analysis and classical regression	• DV: customer loyalty • IVs: customer satisfaction, perceived value, perceived corporate image, customer expectations, perceived value of hardware (i.e., product quality), perceived value of human ware (i.e., service quality) • Moderator: Buyer's Relational orientation (BRO)	• Perceived value, customer satisfaction, hardware quality, and human ware quality have a positive significant effect on customer loyalty. • Customers with a high BRO were more loyal compared to customers with a low BRO.	Journal of Business-to-Business Marketing	Surveys
Callarisa Fiol (2009)	• Empirical study among an industrial cluster to explain customer loyalty antecedents, where perceived	• DV: customer loyalty • IVs: perceived value, satisfaction • Moderator: number of suppliers	• Perceived value and satisfaction have a positive influence on customer loyalty. • Number of suppliers does not affect the relationship between perceived value and satisfaction on customer loyalty.	Journal of Business-to-Business Marketing	Surveys

	value acquires a multidimensional perspective. • Methodology: structural equation models				
Cater and Zabkar (2009)	• Empirical study among services providers and their clients to study the relationships between the three dimensions of commitment, social bonds, and trust and satisfaction and their effect on customer loyalty. • Methodology: structural equation models	• DV: customer loyalty • IVs: trust, social bonds, satisfaction • Mediators: affective commitment, normative commitment, calculative commitment	• Trust, social bonds, and satisfaction have a positive effect on customer loyalty through affective commitment. • Normative and calculative commitment has no significant effect on customer loyalty.	Industrial Marketing Management	Surveys
Wallenburg (2009)	• Empirical study among logistic service providers (LSPs) to analyze to what extent LSPs may utilize their proactive improvement to create customer loyalty and whether a focus on either cost or performance improvements is preferable • Methodology: structural equation models	• DV: customer loyalty (retention, extension, referral). • IVs: proactive cost improvement, proactive performance improvement	• Proactive cost improvement and proactive performance improvement are both strong drivers of all core dimensions of loyalty (retention, extension, and referrals). • Cost improvement and thus efficiency are the main drivers of customer loyalty when the outsourced services are simple, and the contracting period is relatively short. • When services increase in complexity and the contracting period lengthens, customer loyalty is primarily driven by proactive performance improvement and thus effectiveness, whereas cost improvement plays a subordinate role.	Journal of Supply Chain Management	Surveys
Scheer et al. (2010)	• Empirical study among purchasing managers to examine whether suppliers' capabilities impact customer's dependence on the supplier and thereby generate customer loyalty. • Methodology: structural equation models	• DVs: relational loyalty, insensitivity to competitive offerings, future purchase expansions • IVs: core offering capabilities, communication capabilities, operations capabilities • Mediators: benefit-based dependence, cost-based dependence	• Core offering capabilities have a positive effect on benefit-based dependence. • Operations based capabilities have a positive and significant effect on benefit- and cost-based dependence • Benefit-based dependence has a direct positive and significant effect on customer loyalty. • Cost-based dependence has a positive effect on core offerings. • Relational loyalty has a positive effect on insensitivity to competitive offerings and future purchase expansion. • Through the positive effect on relational loyalty, benefit-based dependence has an indirect significant positive effect on insensitivity to competitive offerings and future purchase expansion. • Benefit- and cost-based dependence mediate the impact of core offering and operations capabilities. •Relationship duration has a positive significant effect on benefit-based dependence.	Journal of the Academy of Marketing Sciences	Interviews and surveys
Eggert et al. (2012)	• Empirical study among the medical instruments industry to analyze the interplay between different	• DVs: distributor loyalty, brand loyalty, channel switching • IVs: brand loyalty, customer	• Customers' brand loyalty has positive and unidirectional spill-over effect on distributors • From the brand manufacturer's perspective, the loyalty spill-over can have positive or negative consequences,	Journal of Supply Chain Management	Surveys

	forms of customer loyalty in distribution channels • Methodology: structural equation models	distributor satisfaction, number of distributors, number of brands	depending on the level of vertical competition among channel members. • While the spill-over increases end customers' loyalty toward the channel, it decreases the brand manufacturer's odds of keeping end customers when it comes to the contest between a brand manufacturer and its distributor.		
Hartmann and Grahl (2011)	• Empirical study among logistics service providers (LSP) to analyze the role of flexibility as a potential source of competitive advantage • Methodology: structural equations models	• DV: customer loyalty (retention, extension, referrals) • IVs: supply chain partner insight, communication • Mediators: LSP flexibility, collaboration	• LSP flexibility is a strong driver of all core dimensions of customer loyalty (i.e., retention, extension, and referrals) and thus a source of competitive advantage for LSPs. • Collaboration positively influences LSP flexibility and the loyalty dimensions. • Knowledge resources have a positive effect on LSP flexibility and collaboration.	Journal of Supply Chain Management	Surveys
Molina and Saura (2012)	• Empirical study among retailers to analyze the influence of both market and relationship conditions on trust, commitment, and customer loyalty • Methodology: structural equations models	•DV: customer loyalty •IVs: relationship value, provider dependence • Mediators: trust, long-term orientation (LTO), commitment	•Relationship value has a positive effect on trust and a negative effect on commitment •Trust has a positive effect on LTO and commitment. • Dependence has a negative effect on LTO and loyalty. •LTO has a positive effect on commitment. •Commitment has a positive effect on loyalty.	Journal of Business-to-Business Marketing	Surveys
Pan et al. (2012)	• Meta-analysis of empirical findings on the predictors of customer loyalty • Methodology: Fisher Z-transformation to obtain average statistics	• DV: customer loyalty • IVs: customer satisfaction, trust, psychological commitment, LP membership, Perceived value, product/service quality, perceived fairness, switching costs, and brand reputation • Moderators: goods versus service, behavioral versus attitudinal loyalty, single versus multi-item measures, business vs. consumer market, regular vs. irregular purchase cycle	• Customer satisfaction, trust, psychological commitment, LP membership, perceived value, product/service quality, perceived fairness, switching costs, and brand reputation all have a positive effect on customer loyalty. • Customer loyalty is stronger with service than with goods. • Customer loyalty is stronger for attitudinal loyalty than for behavioral loyalty • Customer loyalty is stronger with multi-item measures than with single-item measures. • Customer loyalty is stronger in consumer markets than in business markets. • Customer loyalty is stronger with an irregular purchase cycle than with a regular one.	Journal of Retailing and Consumer Services	Other studies
Stock and Zacharias (2013)	• Empirical study among multi-industry B2B companies to analyze a multidimensional conceptualization on innovativeness at the product program level • Methodology: moderated regression analysis	• DV: customer loyalty. • IV: product program newness, product program meaningfulness • Moderator: brand association with innovativeness, customer integration	• A negative effect of product program newness on customer loyalty • A positive effect of product program meaningfulness • A brand's close association with innovativeness reduces the negative effect of product newness • Integrating customers into the value-creating process fosters the loyalty effect of product meaningfulness.	Journal of Product Innovation Management	Surveys
Kofi Amoako et al. (2020)	• Empirical study among business	• DV: intention to continue business	• A positive and significant relationship exists between commitment and	Journal of Business-to-	Surveys

	distributors to analyze the relationship between trust, commitment, relative dependence, customer satisfaction, and perceived value on customer loyalty • Methodology: structural equation modeling technique of partial least squares	• IVs: trust, commitment, relative dependence, customer satisfaction, and perceived value.	intention to continue business, relative dependence and intention to continue business, customer satisfaction and intention to continue business, and distributor perceived value and intention to continue business. • No relationship was found between trust and related intention to continue business uniquely unless they were joined to other variables.	Business Marketing	
Yuan et al. (2020)	• Empirical study among third-party organizations (TPOs) to examine how unobservable product quality and information value can be signaled • Methodology: structural equation models	• DV: customer loyalty • IVs: expertise of TPO endorsement, trustworthiness of TPO endorsement • Mediator: perceived value. • Moderator: parasocial relationship	• A positive association between the expertise and trustworthiness of a TPO endorsement, and if characteristics of TPO endorsement positively affect perceived value. •A positive association between perceived value and B2B customer loyalty • Parasocial relationship moderate the relationship between TPO endorsement and perceived value.	Industrial Marketing Management	Interviews and surveys
Bill et al. (2020)	• Empirical study among sales managers, salespeople, and customers to analyze what drives salesperson social media use in B2B relationships and under which circumstances social media use affects customer loyalty • Methodology: structural equation models	• DV: customer loyalty • IVs: performance expectancy, effort expectancy, social influence, facilitating conditions • Mediator: salespeople's social media use • Moderators: salespeople's characteristics, characteristics of salespeople-customers relationship, characteristics of the customer's buying task, and resources of the salesperson	• Social media's effect on customer loyalty strongly depends on the context. • Salespeople' social media use increases customer loyalty only for high-status customers and customers with small buying centers.	Journal of the Academy of Marketing Sciences	Interviews and surveys
Mangus et al. (2020)	• Empirical study among salespeople in B2B firms to analyze the interplay between business and personal trust on performance on the one hand and customer relationship satisfaction and customer loyalty on the other hand • Methodology: structural equation models	• DVs: customer loyalty, customer satisfaction • IVs: business trust, personal trust • Moderator: market turbulence	• Personal trust has a positive and significant direct effect on customer loyalty. • Business trust has a positive but statistically insignificant effect on customer loyalty • Personal trust and business trust have a positive and significant interactive effect on customer loyalty • Personal trust and market turbulence have a positive and significant interaction effect on customer loyalty • Business trust and market turbulence have a positive and significant interaction effect on customer loyalty • A negative and significant three-way interaction effect among personal trust, business trust, and market turbulence on customer loyalty	Journal of the Academy of Marketing Science	Surveys

Agnihori et al. (2020)	• Empirical study among frontline employees (FLE) working in B2B sales/service roles, their customers, and managers to analyze attributes of FLE social capital and how the use of online social networks (OSNs) enables social capital development and social capital maintenance • Methodology: structural equation models	•DVs: FLE social capital development, FLE social capital maintenance, customer loyalty with the firm, FLE's sales performance •IVs: FLE's use of OSN's, FLE social capital development, FLE social capital maintenance •Moderator: time management skills, perceived innovation climate, customer perceived FLE responsiveness	• OSNs relate to social capital development and maintenance. • Time management skills strengthen links between OSNs and both forms of social capital. • Perceived innovation climate plays a moderating role only in the social capital development process. • Unique pathways connecting social capital development to customer loyalty with the firm and social capital maintenance to FLE sales performance were noted	Decision Sciences	Surveys

DV= Dependent Variable, IV= Independent variable

The systematic literature review led to a division of drivers into survey drivers, database drivers, and hybrid drivers. This study only mentions drivers applicable to the general B2B customer loyalty context. If drivers were industry-specific, they were not included in this study. After analyzing all articles, four survey drivers, one database driver, and three hybrid drivers were identified.

For the completeness of identifying all possible drivers, survey drivers are acknowledged in the next paragraph. However, due to the focus of this study on drivers that can be identified through databases, thus without the active participation of customers, drivers that are solely identified through surveys are not further analyzed in this study.

2.3.2 Survey Drivers

The first driver is customer satisfaction, which has often been considered as an essential driver (Stock, 2005; Lee and Bellman, 2008; Cater and Zabkar, 2009; Eggert et al., 2012; Pan et al., 2012; Kofi Amoako et al., 2020). However, Khatibi et al. (2002) and Stoel et al. (2004) state that other studies fail to provide a strong linkage between satisfaction and loyalty or that the satisfaction-loyalty relationship is indirect and complex (Anderson and Mittal, 2000; Magi, 2003). Nevertheless, in general, studies find a linear and positive effect of satisfaction on loyalty.

Trust is the second driver, whereby a consumer who trusts a product is more likely to have a positive attitude towards that product, is willing to pay a premium price, remains loyal, and spreads positive word-of-mouth (Pan et al., 2012). Furthermore, trust can be a critical driver that focuses on positive motivations to maintain relationships (Scheer et al., 2010).

The third driver is perceived value, whereby the benefits in combination with the costs are used to determine the overall perceived value of a product or service that can ultimately motivate loyalty

(Scheer et al., 2010). When a customer's perceived value meets or even exceeds its expectations, customers view a product as a worthy buy. In contrast, a low perceived value leads to customers' higher willingness to switch to a different brand, resulting in a decline in loyalty (Pan et al., 2012).

The fourth driver is brand reputation. Brand reputation is often considered as a mechanism of assuring the trustworthy behavior of a firm. Thus, customers are more likely to perceive a brand with a good reputation as reliable than a brand with a poor reputation (Pan et al., 2012). Pan et al. (2012) also argue that a product with a good reputation has less perceived risk associated with performance ambiguity and information asymmetry, which leads to a favorable purchase and repurchase intent. Customers' loyalty to a brand signals a motivation to maintain a relationship with that brand (Eggert et al., 2012). However, Lee and Bellmann (2008) posit that the image of an organization has less impact on business markets than on consumer markets.

2.3.3 Database Drivers

The drivers discussed in Section 2.3.2 focus on variables where information is acquired only through surveys. When focusing on database drivers, one particular driver that can show an effect on customer loyalty, which is platform value (He and Zhang, 2022). According to He and Zhang (2022), platform value in a B2B setting can be described as engagement on the service platform by users to fulfill functional, social and/or hedonic objectives to ultimately enhance firm performance. Users obtain more platform value from interactions through multiple channels. As a result, customers rely more on the platform services, knowledge, and information provided on the platform. Consequently, a certain dependability on a product or service will lead to an intention to continue the relationship (Kofi Amoako et al., 2020). Xie et al. (2013) support this statement by reporting that when an organization provides important products or services for the partner, the customer has a higher level of intention to continue business with the organization. Furthermore, the more a customer engages with the platform, the more they use it in their daily activities (Brendt et al., 2012). Ultimately, customers who repeatedly use the platform services, and create higher customer share due to the better customer-platform relationship quality, are therefore more likely to stay with the current service provider (He and Zhang, 2022). Therefore, the proposed hypothesis is:

H1a: Platform value has a positive effect on customer retention.
H1b: Platform value has a positive effect on cross-buying.

2.3.4 Hybrid Drivers

Drivers who can be identified in both ways will be discussed within this section. The first driver is relationship value, which concerns the people within a relationship. All sides develop trust and mutual benefits to create value (Molina and Saure, 2012). The literature emphasizes the role between relationship value and long-term orientation (Ulaga, 2003; Bondar et al., 2007). As it can be inferred by the dimensions identified by Ulaga (2003), the value of personal interaction can exist in the form of a dedicated account manager from the supplier's organization towards the customer. Business relationships are established between organizations; however, they are managed by individuals. It is the people who make the relationship work or fail (Ulaga, 2003). An individual relationship brings people closer and impacts the relationship's performance, which ultimately promotes stronger relationship outcomes (Mangus et al., 2020). Research shows that companies who understand the importance of building one-to-one relationships will have the edge (Bondar et al., 2007). However, dedicating an account manager to a customer can be expensive. Therefore, account managers are often assigned to customers who only generate sufficient revenue for the company to maintain its profitability. The proposed hypotheses are:

H2a: Relationship value has a positive effect on customer retention.
H2b: Relationship value has a positive effect on cross-buying.

The second driver is communication, which can mitigate conflict and operation costs, enhance interfirm cooperation, improve service quality, and generate positive customer outcomes, all of which eventually increase loyalty (Scheer et al., 2010). According to Hartmann and Grahl (2011), communication with the customer represents a crucial resource and should be used to generate and increase flexibility capabilities in the specific customer relationship. Marketing efforts represent a way of communicating with customers to increase their loyalty (Hänninen and Karjaluoto, 2017). Knott et al. (2002) agree with this statement and suggest that past promotions may have long-term effects on customer purchases, increasing the customer's lifetime. According to Sashi (2012), engaging the customers in interaction with their suppliers may increase customer loyalty. Moreover, positive associations about a brand may be strengthened by communication (Hartmann and Grahl, 2011). Additionally, Scheer et al. (2010) argue that communication capabilities may enhance social relationships and impact interpersonal dependence and loyalty. The proposed hypotheses are:

H3a: Communication has a positive effect on customer retention.
H3b: Communication has a positive effect on cross-buying.

The third driver is social bonds and occurs in many different forms such as attachment, commitment, trustworthiness, conflict, benevolence, and equity (Perry et al., 2002). Bardauskaite's (2014) determinant in business services loyalty is similarities, whereby people or organizations may signal that they will work toward mutually essential goals for both parties. In a B2B setting, the relationship is more important than the actual exchange of goods or services, and shared values are critical in facilitating exchange relationships (Friman et al., 2002). According to Palmatier et al. (2005), a common reference point helps strengthen and maintain the relationship and can be one of the most effective relationship-building strategies. Similarities on interpersonal and interorganizational levels make it easy to do business with each other, bring both parties close to collaboration, and stimulate mutual goal-seeking (Bardauskaite, 2014), whereas dissimilarities pose risks. For instance, different nationalities often mean different working cultures, making it difficult to come to an agreement (Hagigi and Sivakumar, 2009). Therefore, this study uses social bonds to indicate similarities between firms. Hence, the proposed hypotheses are:

H4a: Social bonds have a positive effect on customer retention.
H4b: Social bonds have a positive effect on cross-buying.

2.3.5 Moderator

The last hypotheses are based on whether relationship length affects the relationship between customer loyalty drivers and customer loyalty. Wallenburg (2009) analyzes the moderating effect of contract length on customer loyalty. In his study, a long contract duration had a positive effect on drivers that influenced customer retention and customer extension of the contract, and it led to more referrals. Oliver (1999) argues that customer loyalty evolves, through the length of a relationship, from weak to robust forms based on different determinants. As argued by Wang and Wu (2012), customer perceived value has a stronger positive effect on switching costs in longer-term relationships than shorter-term relationships. Verhoef (2003) discovered that relationship length positively affects the relationship between satisfaction and customer retention. Thus, the longer customers stay with a company, the more satisfied they become. As shown by previous research, relationship length can have a moderating effect on the relationship between customer loyalty drivers and customer loyalty. Hence, the length of a relationship was incorporated in the theoretical framework of this study. Consequently, the following hypotheses are proposed:

H5a: The relationship between platform value and customer retention is positively moderated by relationship length.

H5b: The relationship between relationship value and customer retention is positively moderated by relationship length.

H5c: The relationship between communication and customer retention is positively moderated by relationship length.

H5d: The relationship between social bonds and customer retention is positively moderated by relationship length.

H5e: The relationship between platform value and cross-buying is positively moderated by relationship length.

H5f: The relationship between relationship value and cross-buying is positively moderated by relationship length.

H5g: The relationship between communication and cross-buying is positively moderated by relationship length.

H5h: The relationship between social bonds and cross-buying is positively moderated by relationship length.

2.4 Control Variables

Next to the main drivers, two control variables were included in this study's model. If such variables are not controlled for, the statistically estimated impact of the identified drivers on customer loyalty may be biased. Accordingly, this study controlled for firm size and country (Stock, 2005; Stock and Zacharias 2013). Firm size is an often-used control variable in the systematic literature review. Country was chosen as control variables since Digidata has many customers in different countries, and cultural differences between countries may thus play a role.

2.5 Theoretical framework

Not all the reviewed articles are recent, but the B2B industry has remarkably changed in the last 10 years (Dempsey and Kelliher, 2018), which makes the four stated drivers considered in this study relevant to the current state of affairs. In addition to the four drivers, one variable was considered to check for moderation. This section summarizes the theory discussed in a theoretical framework. Figure 1 features the quantitative drivers that were investigated through empirical research.

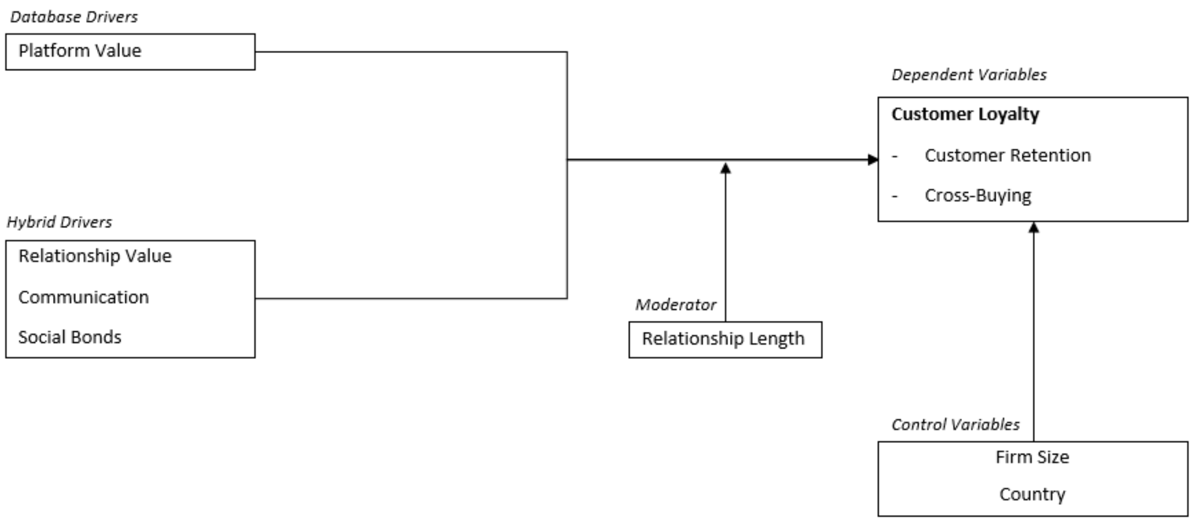


Figure 2.1: Theoretical framework of customer loyalty drivers

3 Methodology

This section describes the cross-industry standard process for data mining (CRISP-DM) that was employed in this study, followed by the quantitative modeling technique known as hierarchical logistic regression.

3.1 CRISP-DM

CRISP-DM was developed by Wirth and Hipp (2000) to provide an overview of the life cycle of a data mining project. This life cycle is broken down into six stages, which are shown in Figure 3.1.

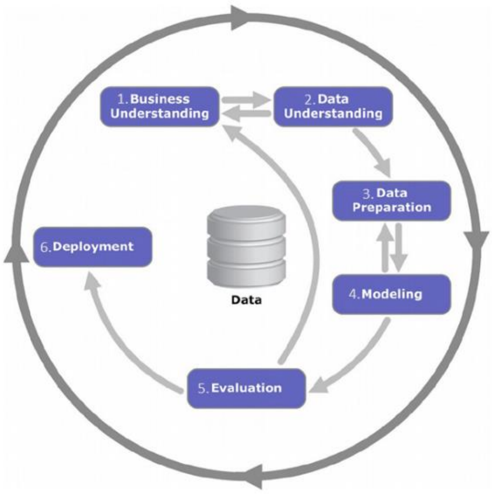


Figure 3.1: CRISP-DM cycle (Wirth and Hipp, 2000)

Even though the figure indicates a sequential procedure in steps, the arrows only display the most important and frequent dependencies between different phases. Within a particular project, the outcome of each phase depends on which phase or task has to be performed next, with the exception of phase six. By moving in an iterative way from step one to five, the overall quality of any data mining research can be improved. Each phase and its implementation in this study is briefly described based on Wirth and Hipp (2000).

1. The first phase is *Business Understanding*. In this phase, a literature review was performed to shed light on the theoretical background concerning the concept of customer loyalty and the drivers of customer loyalty. The focus was on understanding the objectives and requirements of a study. Furthermore, the situation was assessed for risks, requirements, and contingencies. The data mining goals were determined to gather a deeper understanding of the data available and expert knowledge regarding the concept of customer loyalty and the customer base of Digidata. This phase is elaborated in Chapters 1 and 2.

2. The second phase is *Data Understanding*. Here, the datasets are first collected. In the case of Digidata, a CRM database, a financial database, an internal database regarding the customers' behavior, and an external database regarding customer information were used. Subsequently, the data are described to document their descriptive statistics. Next, the data are explored to identify relationships. Lastly, the data quality was verified, and quality issues were documented. Simple statistical analysis could also be performed in this stage, such as the handling of missing values and boxplots. Section 4.1 further explains this phase.
3. *Data Preparation* is the third phase. First, it was determined which datasets would be used, and reasons were documented for inclusion and exclusion. Afterwards, all different datasets at Digidata (CRM database, financial database, internal database, and an external database) had to be merged to perform an analysis. All data had to be prepared so that they would include a type of variable such as ratio, nominal or binary. For example, all dependent variables had to be converted to binary values to perform the hierarchical logistic regression. Then, the data were cleaned to prevent garbage-in, garbage-out. Digidata's datasets might contain missing values and outliers, which had to be handled in this phase. Afterwards, the data were constructed to derive new attributes. Next, a new dataset was created by combining data from multiple sources. Lastly, the data were reformatted to later analyze them using a statistical program. Section 4.2 further explains this phase.
4. The fourth phase is *Modeling*, in which the modeling techniques are selected, and assumptions are made according to the data. Within this study, different hierarchical logistic regression models were created to test the relationship between drivers and customer loyalty. This phase is further elaborated in Section 4.3.
5. Fifth, the *Evaluation* phase includes the assessment of the modeling according to the goals set in the previous phase. When the research goals are not met, previous steps should be repeated, different models can be built, or the goals should be adapted. A confusion matrix can support the evaluation using accuracy, sensitivity, precision, recall, error rate, and F1 score. In this study, some variables were used multiple times; therefore, variance inflation factor (VIF) scores and correlation coefficients were used to check for multicollinearity, as shown in Appendixes C and D. This phase is illustrated in Section 4.5.
6. The last step is the *Deployment* phase. In this phase, the conclusion and discussion are reported. Additionally, recommendations from an organizational and literate perspective regarding this study are provided. This phase is implemented in Chapter 5 in this study.

While going through the first five phases, it is possible and common to go back to a previous phase to better align the research goals or revise the problem statement.

3.2 Hierarchical Logistic Regression

Supervised modeling techniques are generally categorized into two types of boxes (Dreiseitl and Ohno-Machado, 2012). The first type contains techniques such as decision trees, k-nearest neighbors, and logistic regression, which are usually applied to interpret model parameters (Dreiseitl and Ohno-Machado, 2012). Techniques in this box are categorized as white-box models. The opposite of white box models is black box models, which include support vector machines and artificial neural networks that do not allow for the model to be verified externally or the possibility to interpret its parameters. (Dreiseitl and Ohno-Machado, 2012). Hierarchical logistic regressions are also a common technique within the field of marketing. Applications range from estimating customers' churn probability to predicting a customer's tendency to purchase a product or service (Akinci et al., 2007).

White box models—in particular the hierarchical logistic regression model—were the option preferred for this study. Binary dependent variables were the main reason for choosing the hierarchical logistic regression (King, 2008), as in hierarchical logistic regression, variables are used to predict outcomes, which are known as regression coefficients or beta coefficients (B) who define the direction. In hierarchical logistic regression, the magnitude of the direction must be compared with the magnitude of the features used in the model since it cannot be interpreted directly. Furthermore, the odds ratio indicates that one outcome will occur. To model the effect of dependent variables customer retention and cross-buying on the four data-driven drivers, platform value (PV), relationship value (RV), communication (CO), and social bonds (SB), the moderating effect of relationship length (RL) on the relationship between the data-driven drivers and customer loyalty, and the effect of control variables firm size (FS) and country (Belgium (BEL), France (FRA), Germany (GER), The Netherlands (NED), South Africa (SA) and Others (OTH)), two logistic regression formulas were formulated:

$$Logit_{Customer\ Retention} = B_0 + B_1PV + B_2RV + B_3CO + B_4SB + B_5(PV\ x\ RL) + B_6(RV\ x\ RL) + B_7(CO\ x\ RL) + B_8(SB\ x\ RL) + B_9(FS) + B_{10}(BEL) + B_{11}(FRA) + B_{12}(GER) + B_{13}(NED) + B_{14}(SA) + B_{15}(OTH)$$

(Formula 3.1)

$$Logit_{Cross-buying} = B_0 + B_1PV + B_2RV + B_3CO + B_4SB + B_5(PV\ x\ RL) + B_6(RV\ x\ RL) + B_7(CO\ x\ RL) + B_8(SB\ x\ RL) + B_9(FS) + B_{10}(BEL) + B_{11}(FRA) + B_{12}(GER) + B_{13}(NED) + B_{14}(SA) + B_{15}(OTH)$$

(Formula 3.2)

In Formulas 3.1 and 3.2, $Logit_i$ represent the dependent variables customer retention and cross-buying, B_0 represent the intercept term, B_{1-4} are the corresponding beta coefficients of the independent variables, B_{5-8} are the corresponding beta coefficients of the moderating effect on the relationship between the independent and the dependent variables, B_{9-15} are the corresponding beta coefficients for the control variables. Chapter 4 further explains the operationalization of the variables.

4. Quantitative Study: Hierarchical Logistic Regression

This chapter illustrated the quantitative research component of this thesis, which was carried out through a hierarchical logistic regression, and it is structured following the guidelines of the CRISP-DM methodology described in Section 3.1. Specifically, the model's performance is analyzed along with the final results.

4.1 Data Understanding

This section assesses the available data. The first paragraph provides an overview of the different databases and their corresponding variables available at Digidata. In the second paragraph, the quality of the data is evaluated by analyzing the number of missing data and checking for outliers.

4.1.1 Database Overview

Digidata features four databases where its customer-related, firmographic, transactional, and activity-based data are stored. Table 4.1 provides an overview of all databases.

Table 4.1: Database overview

Database:	**Description:**
Customer information database	Customer relationship management (CRM) information on the customer's address, country, contact person, and account manager.
Financial database	Financial information concerning length of contract, contract start, and end date, number of subscriptions, and number of products bought.
Customer activity database	Internal database concerning the last login date, number of logins, and number of email campaigns sent
External database	External database from which customer information, such as primary industry, sector, type of enterprise, and commercial employees can be extracted.

All databases are linked through Microsoft PowerBI—a collection of software services, applications, and connectors that work together to turn unrelated data sources into coherent, visually immersive, and interactive insights. Since all organizations are identified through a unique customer ID, all data were easily combined. All the variables available in the different databases are listed in Table 4.2.

Table 4.2: Overview of all variables in the different databases

Customer Information Database	Financial Database	Customer Activity Database	External Database
Company ID	Company ID	Company ID	Company ID
Company number	Company name	Company number	Country
Account manager	Business lines	Platform login date	Revenue of company
Contract length	Month of	Number of application starts	Major industry
Contract start date	Invoice	Number of subscriptions	Country
Sector	Revenue in EUR	Number of email campaign sent	Line of business
Language	Margin in EUR		Year started
Origin			Legal type
Email			Commercial employees
Currency			DUNS* number
Classification			

*DUNS= Data universal numbering system is a unique, nine-digit series of numerals that identifies a business.

4.1.2 Initial Collection of Data

This section covers the timespan of collection of the dataset and the number of unique data points before cleaning the data. In January 2016, Digidata introduced a new financial system to enhance the quality and number of gathered data. Due to a large number of missing values, customers that were invoiced before January 1st 2016, are excluded from this analysis.

After the initial collection, 2,470 unique customers operating within the SaaS industry between January 2016 and August 2021 were extracted from Digidata's databases.

4.1.3 Data Quality

To analyze the quality of the data, all databases were checked for missing values. Table 4.3 shows the variables that contain missing values.

Table 4.3: Overview of variables containing missing values

Variable	Data Type	% Missing Values	# Missing Values
Company Activity Database			
Contract start date	Numerical	6.2%	152
External Database			
Primary industry	Categorical	16.5%	407
Country	Categorical	16%	394
Legal type	Categorical	18.8%	465
Commercial employees	Numerical	16%	394
DUNS number	Numerical	11%	271

Note: N=2,470

Enders (2003) states that a missing rate of 15% to 20% is frequent in educational and psychological studies, with 97% of those that showed evidence of missing data using the listwise deletion or pairwise deletion method to deal with missing data.

4.2 Data Preparation

Data preparation is the third phase of the CRISP-DM methodology, and it covers all the steps to arrange the data for modeling. Accordingly, the predictors were operationalized and linked to the available data in Digidata' databases and then cleaned to prevent garbage-in, garbage-out. Some variables required a conversion to be able to perform a logistic regression. Lastly, all variables were combined into a single dataset.

4.2.1 Operationalization

This section illustrates the operationalization of the variables in the theoretical framework as discussed in Section 2.5. Table 4.4 shows all variables, their measurements, and the data type. The first column lists variable types, and the second column reports variable names according to the theoretical framework. The third column represents how the variable is measured, while the last column indicates the type of data per variable.

Table 4.4: Operationalization of the variables

Type of Variable	Name	Operationalization	Type of Data
Dependent Variable	Customer Retention	Number of months billed by Digidata.	Ratio
	Cross-buying	The customer purchased 1, 2, 3, or 4 products.	Ratio
Independent Variables	Platform Value	A score based on platform login date, number of application starts, commercial employees and months a customer is subscribed at Digidata.	Ratio
	Relationship Value	The customer has a dedicated account manager with a specific name attached.	Nominal
	Communication	Number of email campaigns sent.	Ratio
	Social Bonds	Number of similarities between commercial sector, primary industry, country and legal structure.	Ratio
Moderator	Relationship Length	Number of months billed by Digidata.	Ratio
Control Variable	Firm Size	Number of employees.	Ratio
	Country	The top 5 countries were analyzed separately. Other countries were merged into one category.	Nominal

Customer retention is the first dependent variable representing customer loyalty and refers to long-lasting relationships maintained between a provider and a customer (Bó et al., 2018). In line with the research by Ozuem et al. (2016), who measured loyalty strength using the length of the relationship, customer retention was measured by the number of months an organization was Digidata's customer. Section 4.3.3 further explains when customer retention occurs. The second dependent variable, cross-buying, was operationalized through the number of products a customer was subscribed to. The rationale behind this operationalization is that if customers buy from different categories offered by the same firm, they will experience greater attachment to that firm and thus greater loyalty (Reinartz et al., 2008).

The first independent variable, platform value, was operationalized through a market segmentation analysis technique based on the customer's recency, frequency, and engagement (RFE; Kholief, 2021). The RFE analysis is a broader version of the recency, frequency, monetary value analysis and reflects platform activity in combination with the length of the customer relationship. Platform login date, number of application starts, number of subscriptions, commercial employees, and contract length were used to calculate an RFE score per customer. Section 4.2.3 provides a more detailed explanation of how this score was calculated.

Relationship value is the second independent variable and was operationalized as a dummy variable indicating whether an account manager was assigned to a customer or not (Ulaga, 2003). The third independent variable is communication and was operationalized through the number of email campaigns sent by Digidata (Hänninen and Karjaluoto, 2017). Email campaigns provide information useful for customers to gain personal benefits through discounts on other services and about new features on current services. The fourth and last independent variable is social bonds, operationalized through similarities, which were measured in the areas of primary industry, sector, country and legal type to assess a party's willingness to work toward goals that were mutually important for both sides (Bardauskaite, 2014).

This study's moderator, namely relationship length, was operationalized through the number of months a customer was subscribed to Digidata (Wallenburg, 2009).

The control variable firm size was operationalized through the number of employees per unique customer (Stock, 2005), whereas country was operationalized through a selection of the customers in the top five most common countries that do business with Digidata. All other countries were grouped together.

4.2.2 Data Cleaning

During the phase of data cleaning, data is corrected, imputed, or removed when erroneous or missing. The first step in cleaning the data is to generate an overview to check them for outliers and nonrepresentative or missing data that must be removed or accounted for. No outliers were found in any of the datasets. The missing values shown in Table 4 were listwise deleted to reduce bias, which resulted in a sample of 2,005 customers. In a listwise deletion a case is dropped from the analysis because it has a missing value in at least one of the specified variables. Then, the data were checked for multicollinearity. The correlation coefficients table in Appendix C shows no significant correlation between the variables. Appendix D shows the Variance Inflation Factor (VIF) values. No values exceed the limit of 10, which indicates no multicollinearity (O'brien, 2007). However, some variables required a conversion from ratio to binary or from textual to numerical, as described in the following section.

4.2.3 Data Conversion

The variables shown in Table 4.4 were converted to perform a hierarchical logistic regression. The two dependent variables were converted into binary values. Specifically, customer retention was converted into two groups. The first group includes customers who are loyal, that is, who had been subscribed to Digidata for longer than 33 months. This group was converted into the binary value of "1." For this study, 33 months was chosen as the tipping point since this is the average lifetime of a customer at Digidata. All other customers were categorized as "not loyal" and converted into the binary value of "0." The second dependent variable is cross-buying Here, customers are also divided into two groups. The first group is customers who are subscribed to multiple products of Digidata. This group is converted into the binary value of "1." Customers who were subscribed to only one product received a target value of "0."

To calculate the RFE score, all three values (last login date, number of application starts, and number of months with Digidata) were categorized. Each variable was split into 10 categories in which a customer was placed using the Excel formula PERCENTRANK.EXC(). The categories were relative to each other: the customers with the lowest 10% were put into category 1, those between 10%–20% were put into group 2, and so. This procedure was completed separately for each variable. Afterwards, all three categories were added up to calculate a final RFE score. As an example, a customer who was in the lowest 10% as login date, between 50%–60% in number of application starts, and between 60%-70% for its time with Digidata received a score of 1 +6 +7 = 14.

The independent variable relationship value converted from account manager names to binary values. Customers with an account manager received the value "1," and customers who didn't have it received the value "0."

The independent variable social bonds was converted from text (primary industry, sector, country and legal type) to binary values. Customers received for each category the value "1" when their industry, sector, country, or legal type was similar to the category Digidata operates in; if not, they received a value of "0." The overall score within the independent variable social bonds was calculated through a count-based measurement (Lavine et al., 2012), with scores ranging between 0 and 4.

Lastly, a dispersion was made for control variable countries, as shown in Figure 4.1. Five countries proportionally had more customers than other countries. Therefore, they were extracted from the data, whereas other countries were grouped into the category "others." All categories were then converted into dummy variables to perform hierarchical logistic regressions.

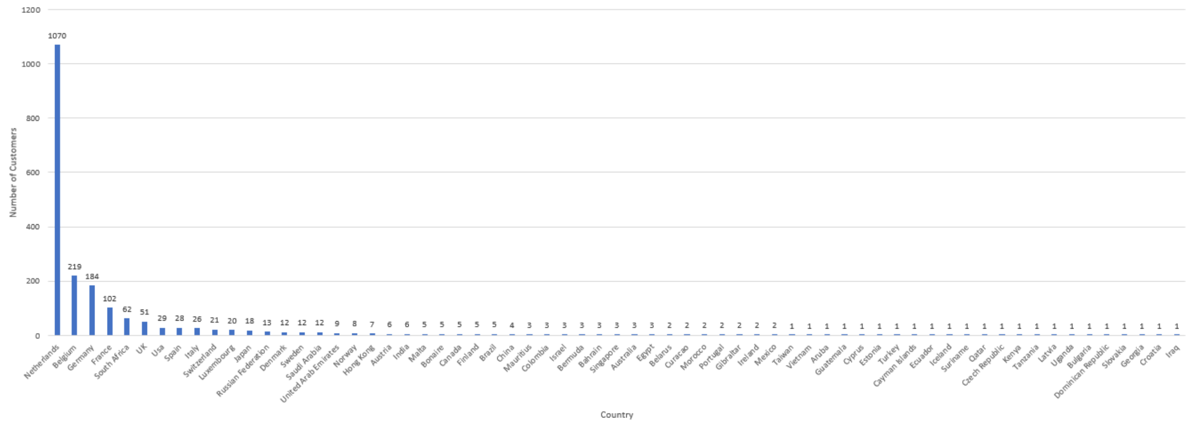


Figure 4.1: Number of customers per country

4.2.4 Final Dataset

The final dataset was assembled by combing data from the literature with those available at Digidata. The time span of the dataset is between 01/01/2016 and 08/01/2021, amounting to five years and eight months. Due to a new billing system, data before 01/01/2016 were either unstructured or lost and could not be used for analysis. After cleaning the initial dataset of 2,470 unique customers, 2,005 unique customers were identified and represented the total sample for this study. In total, seven variables were identified within the theoretical framework. Its corresponding parameters are defined in Table 4.5. Two variables were defined as dependent variables and four variables as independent variables; one variable was defined as moderator, and two variables were defined as control variables. According to van Smeden et al. (2016), when performing logistic regression analysis, 10 events is the minimum criterion for the sample size. Hence, the population of this study is within limits.

Table 4.5: All variables used in this study

Type of Variable	Name	Measurability	Type of Data
Dependent Variables	Cross-buying	The customer purchased one (0) or multiple (1) products.	Binary
	Customer Retention	The organization is a customer of Digidata under (0) or over (1) 33 months.	Binary
Independent Variables	Platform Value	RFE Score based on last login date, number of application starts and number of months with Digidata.	Ratio
	Relationship Value	The customer has a dedicated account manager (1) or not (0).	Binary
	Communication	Number of email campaigns sent.	Ratio
	Social Bonds	Number of similarities between commercial sector, major industry, country and legal structure.	Ratio
Moderator	Relationship Length	Number of months billed by Digidata.	Ratio
Control Variable	Firm Size	Number of employees per customer.	Ratio
	Country	Dummy variable with the four most common countries and a last dummy variable for all other countries.	Dummy

4.3 Modeling

After collecting and preparing the data, the dataset was ready for analysis through the program IBM SPSS Statistics 27. First, the descriptive statistics were elaborated to shed light on the distribution of the dataset. Second, metrics to evaluate the performance of the logistic regression are described. Third, different models were developed by subsequently adding different types of variables. The model with the best explanatory power is the focus of the discussion.

4.3.1 Descriptive Statistics

The descriptive statistics are shown in Table 4.6 to provide a summary of the sample and its measures.

Table 4.6: Descriptive Statistics

Descriptive Statistics					
Categorical Variable	***N***	***%***	**Categorical Variable**	***N***	***%***
Customer retention			Countries		
Yes	788	39.3%	Netherlands	1070	53.4%
No	1217	60.7%	Belgium	219	10.9%
Cross-buying			Germany	184	9.2%
Yes	439	21.9%	France	102	5.1%
No	1566	78.1%	South - Africa	62	3.1%
Customer Retention			Other	368	18.4%
Yes	870	43.4%			
No	1135	56.6%			
Relationship Value					
Yes	1282	63.9%			
No	723	36.1%			
Social Bonds					
0 similarities	82	4.1%			
1 similarity	703	35.4%			
2 similarities	910	45.4%			
3 similarities	310	15.5%			
4 similarities	0	0%			
Continuous Variable	***N***	**Mean**	***SD***	**Min**	**Max**
Platform Value	2005	12.9	4.9	3	24
Communication	2005	27.8	90	0	2386
Relationship Length	2005	33.8	20.3	2	55
Firm Size (Employees)	2005	3256	16763	1	407492

4.3.2 Performance Evaluation Metrics

Four metrics are widely used to evaluate logistic regression models and understand a model's actual performance. These metrics are calculated based on the confusion matrix, which contains numbers of observations that are correctly predicted as positive (True Positive; TP), false predicted as positive (False Positive; FP), correctly predicted as negative (True Negative; TN), or false predicted as negative (False Negative; FN), and can be used to calculate four different metrics that are Accuracy, Precision, Recall, and F1 score. These metrics are defined as follows:

- Accuracy: the ratio of total number of correct predictions to all observations, which generally indicates the performance of predictions in terms of frequency.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

- Precision: the ratio of correctly predicted loyal customer to the total predicted positive observations. It is used to focus on the number of incorrect predicted loyal customer.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: the ratio of correctly predicted loyal customers to all loyal customers, which indicates how many truly loyal customers were predicted correctly.

$$Recall = \frac{TP}{TP + FN}$$

- F1 score: weighted average of Precision and Recall, which also indicates overall prediction accuracy.

$$F1\ score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$$

- Accuracy and F1 score are comparable overall performance indicators. The disadvantage of Accuracy is the dependence on a balanced data classification, whereas F1 score can handle imbalanced data issues, which makes it an important metric that predominates over Accuracy. Precision is another important indicator for loyalty detection. A low score on precision indicates that many customers are predicted as loyal customers, whereas, in reality, they are not loyal to Digidata. A consequence is that Digidata will not target those customers with discounts on services or other promotional advertisements to improve their loyalty. The last indicator is Recall, which values false positives as most significant and is considered the least important among other measures. In this study, loyal customers were predicted as not loyal customers. If these customers were targeted for Digidata's loyalty programs, this would only increase their loyalty even more.

4.3.3 Results

For both dependent variables (customer retention and cross-buying), a separate hierarchical logistic regression was performed. In the first model, only the control variables were added. In the second model, the independent variables were added together with the control variables. In the last model, the moderator relationship length was added. The results of the performance metrics are shown in a separate table under the results of the logistic regression. In all results, the odd ratios (OR) are provided to understand the effect of predictors. Odds ratios greater than 1 indicate that the event is more likely to occur as the predictor increases, whereas odds ratios smaller than 1 indicate that an event is less likely to occur as the predictor increases (Norton et al., 2018).

4.3.3.1 Customer retention results

Table 4.7 shows the results of the logistic regression for the dependent variable customer retention.

Table 4.7: Logistic regression results of dependent variable customer retention

	Model 1a: Control variables			Model 2a: Independent variables + control variables			Model 3a: All variables + moderator		
Factor	OR	S.E.	p	OR	S.E.	p	OR	S.E.	p
Constant	1.200	0.063	0.004	2.294	0.233	0.000	20.089	0.370	0.000
Platform value				0.884	0.023	0.000	1.074	0.044	0.106
Relationship Value				0.311	0.214	0.000	0.380	0.371	0.009
Communication				1.000	0.001	0.773	1.027	0.008	0.000
Social Bonds				1.398	0.149	0.025	1.210	0.247	0.439
Moderator									
Contract Length							1.487	0.040	0.000
Platform value x Relationship Length							1.021	0.004	0.000
Relationship Value x Relationship Length							1.023	0.027	0.401
Communication x Relationship Length							1.005	0.001	0.000
Social Bonds x Relationship Length							0.979	0.017	0.226
Control variables									
Firm Size	1.000	0.000	0.093	1.000	0.000	0.657	1.000	0.000	0.249
Countries									
Belgium	3.264	0.180	0.000	4.485	0.388	0.000	4.773	0.425	0.000
France	0.971	0.209	0.887	1.754	0.476	0.813	1.006	0.505	0.991
Germany	3.318	0.196	0.000	3.463	0.410	0.002	3.706	0.421	0.002
Netherlands	*Reference*			*Reference*			*Reference*		
Others	0.627	0.122	0.000	1.632	0.284	0.085	1.685	0.307	0.089
South Africa	0.119	0.385	0.000	1.754	0.490	0.251	1.823	0.576	0.297

Diagnostics			
N	2005	2005	2005
-2LL	2562.112	754.492	657.811
Cox & Snell *R* squared	0.087	0.629	0.647
Nagelkerke pseudo *R*-squared	0.117	0.844	0.867
Percentage correctly predicted	61.3%	93.0%	93.7%

Note: OR= odds ratio, S.E.= standard error, p= probability

Table 4.8 shows the performance metrics as explained in Section 4.3.2.

Table 4.8: Performance metrics of dependent variable customer retention

Performance metrics	**Model 1a**	**Model 2a**	**Model 3a**
Accuracy	61.30%	93.02%	93.72%
Precision	61.37%	95.60%	95.49%
Recall	85.47%	91.89%	93.30%
F1 Score	71.46%	93.71%	94.38%
AUC	0.681	0.974	0.999

4.3.3.2 Cross-Buying results

Table 4.9 shows the results of the logistic regression for the dependent variable cross-buying.

Table 4.9: Logistic regression results of dependent variable cross-buying

	Model 1b: Control variables			Model 2b: Independent variables + control variables			Model 3b: All variables + moderator		
Factor	OR	S.E.	p	OR	S.E.	p	OR	S.E.	p
Constant	-1.170	0.72	0.000	0.171	0.129	0.000	0.181	0.135	0.000
Platform value				1.111	0.014	0.000	1.119	0.014	0.000
Relationship Value				2.267	0.143	0.000	2.304	0.149	0.000
Communication				1.003	0.001	0.000	1.003	0.001	0.002
Social Bonds				0.924	0.094	0.400	0.900	0.095	0.266
Moderator									
Relationship Length				1.008	0.003	0.020	1.005	0.006	0.416
Platform value x Relationship Length							0.998	0.001	0.010
Relationship Value x Relationship Length							1.006	0.007	0.383
Communication x Relationship Length							1.000	0.000	0.971
Social Bonds x Relationship Length							1.012	0.004	0.004
Control variables									
Firm Size	0.000	0.000	0.394	1.000	0.000	0.273	1.000	0.000	0.342
Countries									
Belgium	-0.429	0.194	0.651	0.576	0.222	0.013	0.585	0.222	0.016
France	-0.773	0.306	0.011	0.451	0.329	0.015	0.428	0.327	0.010
Germany	-0.704	0.228	0.002	0.477	0.246	0.003	0.453	0.246	0.001
Netherlands	*Reference*			*Reference*			*Reference*		
Others	0.038	0.141	0.787	0.867	0.166	0.391	0.799	0.169	0.185

Diagnostics			
N	2005	2005	2005
-2LL	2083.341	1875.846	1859.976
Cox & Snell R squared	0.012	0.109	0.116
Nagelkerke pseudo *R*-squared	0.018	0.168	0.179
Percentage correctly predicted	78.1%	78.7%	78.9%

Note: OR= odds ratio, S.E.= standard error, p= probability

Table 4.10 shows the performance metrics as explained in Section 4.3.2.

Table 4.10: Performance metrics of dependent variable cross-buying

Performance metrics	**Model 1b**	**Model 2b**	**Model 3b**
Accuracy	78.1%	78.65%	78.86%
Precision	0%	55.79%	57.89%
Recall	0%	12.07%	12.53%
F1 Score	0%	19.85%	20.60%
AUC	0.573	0.733	0.740

4.4 Model Development

As with hierarchical logistic regression, the model was developed by subsequently adding different variables as specified in Section 4.3.3. The regression models are compared based on their diagnostics and performance metrics, and the diagnostics of Tables 4.7 and 4.9 and the performance metrics of Tables 4.8 and 4.10 were analyzed. From here, based on the results of both regression models of the dependent variables customer retention and cross-buying, the third model had the best fit and the highest explanatory power. Hence, in both analyses, Model 3 will be the focus of the discussion. Furthermore, to illustrate the moderating effect of relationship length, the moderator variable relationship length was dichotomized using a mean-split plus (long relationship length) and a minus-one (short relationship length) standard deviation (*SD*; Berry et al., 2010). Figures 4–7 illustrate the significant moderating effects.

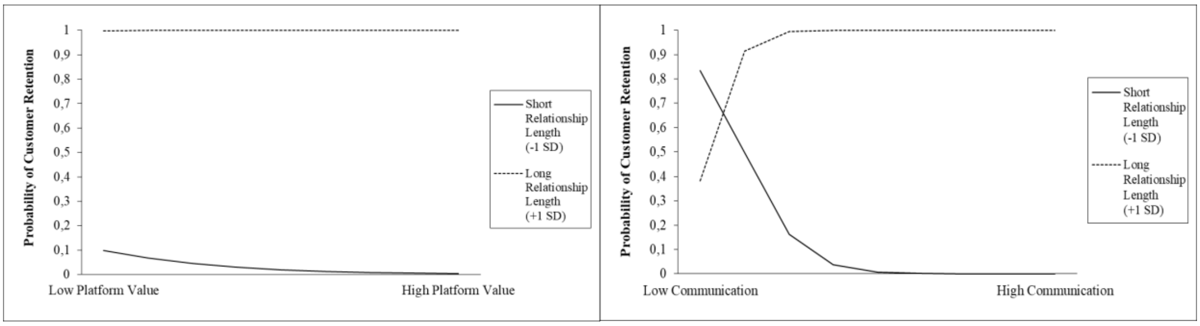
4.5 Findings

Two multiple hierarchical regression models—one for each dimension of customer loyalty, which acted as dependent variable—were configured to test the hypotheses. The (a) independent variables (platform value, relationship value, communication, and social bonds), (b) two control variables (firm size and countries), and (c) one moderator (relationship length) are analyzed in both models.

4.5.1 Customer Retention Findings

In Section 2.3, it was hypothesized that platform value, relationship value, communication, and social bonds have a positive influence on customer retention. Platform value positively influences customer retention and is marginally significant (OR: 1.074, *p*: 0.106), finding weak support for H1a. No support was found for a positive significant relationship between relationship value and customer retention. The relationship was significant but negative (OR: 0.380, *p*: 0.009), leading to rejecting H2a. In line with H3a, a significant positive effect of communication on customer retention (OR: 1.027, *p*: 0.000) was found. No support was found for H4a, which analyzed the effect of social bonds on customer retention (OR: 1.210, *p*: 0.439).

The moderating role of relationship length on the effect between platform value and customer retention is positive and significant (OR: 1.021, *p*: 0.000), leading to accepting H5a. As shown in Figure 4.2, the probability of customer retention is high during all levels of platform value when contract length is long, whereas when contract length is short, it diminishes once platform value increases. A significant positive moderating role of relationship length on the relationship between of communication and customer retention was found (OR: 1.005, *p*: 0.000), providing support for H5c. Figure 4 shows this moderating effect: the probability of customer retention increases for long contract length when communication increases, whereas it decreases for short contract length when communication increases. No significant support was found for the moderating role of relationship length on the effect between relationship value and customer retention (OR: 1.023, *p*: 0.401), nor for the moderating effect between social bonds and customer retention (OR: 1.210, *p*: 0.439), leading to reject H5b and H5d. These findings suggest that the relationship between platform value and customer retention on the one hand and between communication and customer retention on the other hand is stronger for customers who have a long relationship with Digidata.



A B

Figure 4.2: Effect of platform value (A) and communication (B) on customer retention moderated by relationship length

Table 4.11 provides an overview of all hypotheses based on customer retention.

Table 4.11: Summary of results on dependent variable customer retention

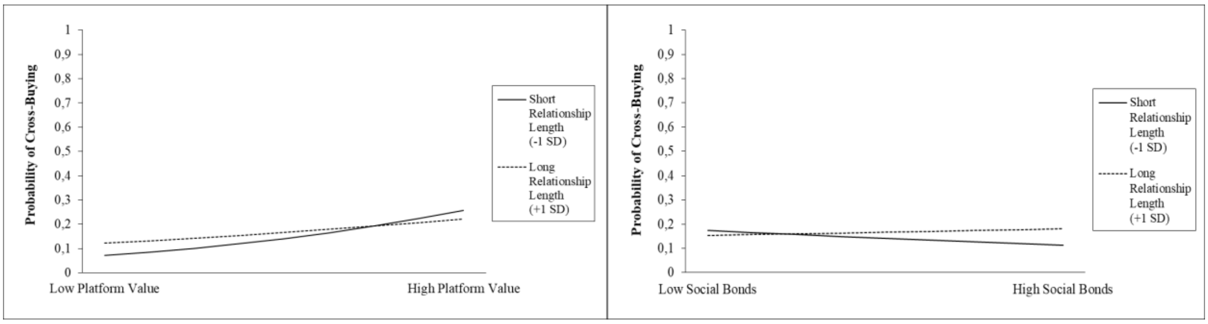
Hypotheses	Support
H1a: Platform Value → Customer Retention	+
H2a: Relationship Value → Customer Retention	--
H3a: Communication → Customer Retention	++
H4a: Social Bonds → Customer Retention	0
H5a: Relationship Length on Platform Value → Customer Retention	++
H5b: Relationship Length on Relationship Value → Customer Retention	0
H5c: Relationship Length on Communication → Customer Retention	++
H5d: Relationship Length on Social Bonds → Customer Retention	0

++: significant positive effect, +: partially significant positive effect, 0: no significant effect, -: partially significant negative effect, --: significant negative effect

4.5.2 Cross-Buying Findings

Section 2.3 also illustrates the hypothesis for the independent variables, as well as the moderator on cross-buying. First, H1b is supported, whereby platform value has a significant positive influence on cross-buying (OR: 1.119, *p*: 0.000). As expected, a significant positive link was found between relationship value and cross-buying (OR: 2.304, *p*: 0.000), confirming H2b. In line with H3b, a positive significant link was also found between communication and cross-buying (OR: 1.003, *p*: 0.002), whereas no support was found for H4b, which analyzed the effect of social bonds on cross-buying (OR: 0.900, *p*: 0.266). The moderating role of relationship length on the effect of platform value on cross-buying is significant (OR: 0.998, *p*: 0.010). However, this significant effect is slightly negative, leading to reject H5e. As shown in Figure 6, the probability of cross-buying increases when platform value increases for both short and long relationship lengths. However, for high values of platform value, the probability that cross-buying occurs is greater for customers with a short relationship length than customers with a long relationship length.

The moderating role of relationship length on the effect of social bonds on cross-buying is positive and significant (OR: 1.012, *p*: 0.004). As shown in Figure 4.3, the probability of cross-buying increases for long relationship length when social bonds increase, whereas it decreases for short relationship length when social bonds increases. However, the overall moderating effect of relationship length on the relationship between social bonds and cross-buying is positive. Therefore, H5h is accepted. These findings suggest that the relationship between high platform value and cross-buying is stronger for customers who have a short relationship length with Digidata, whereas the relationship between high social bonds and cross-buying is stronger for customers who have a long relationship length. The moderating role of relationship length on the effect of relationship value on cross-buying (OR: 1.006, *p*: 0.383), and the moderating role on the effect of communication on cross-buying (OR: 1.000, *p*: 0.971) are both not significant. Thus, H5f and H5g are both rejected.



A B

Figure 4.3: Effect of platform value (A) and social bonds (B) on customer retention moderated by relationship length

Table 4.12 provides an overview of all hypotheses based on dependent variable cross-buying.

Table 4.12: Summary of results on dependent variable cross-buying

Hypotheses	Support
H1b: Platform Value → Cross-buying	++
H2b: Relationship Value → Cross-buying	++
H3b: Communication → Cross-buying	++
H4b: Social Bonds → Cross-buying	0
H5e: Relationship Length on Platform Value → Cross-buying	--
H5f: Relationship Length on Relationship Value → Cross-buying	0
H5g: Relationship Length on Communication → Cross-buying	0
H5h: Relationship Length on Social Bonds → Cross-buying	++

++: significant positive effect, +: partially significant positive effect, 0: no significant effect, -: partially significant negative effect, --: significant negative effect

4.5.3 Control Variables

The effect of the control variable firm size is not significant for customer retention and cross-buying. Some country dummies, however, do have an effect on customer loyalty. In particular, Belgium (OR: 4.773, *p*: 0.000) and Germany (OR: 3.706, *p*: 0.002) has a positive significant effect on customer retention, and Belgium (OR: 0.585, *p*: 0.016), France (OR: 0.428, *p*: 0.010), and Germany (OR: 0.453, *p*: 0.001) have a negative significant effect on cross-buying.

5. Discussion

This chapter discusses the results while basing the structure of its sections on the four drivers, the moderator, and the two dependent variables that represent customer loyalty, followed by the illustration of the managerial implications, the limitations of the study, and the possible directions for future research.

5.1 Theoretical Implications and Contributions

The goal of this study was to identify drivers of customer loyalty that can be operationalized through transactional and behavioral data gathered by the SaaS platform. Using transactional and behavioral customer data, multiple customer loyalty insights were obtained. The database driver that was found positively significant was platform value, which is a driver that can only be identified through behavioral data. The results about platform value are consistent with the theory by He and Zhang (2022), who state that once customers perceive platform value, they proactively establish a long-term relationship with the platform, and their purchase share of a firm. The results are also in line with the theory by Xie et al. (2013), who state that the more important a product or service becomes over time to fulfill a company's activities, the more likely this company will remain with its provider.

Furthermore, the role of relationship length on the relationship between platform value and customer retention is stronger when relationship length increases. This is in line with the studies by Verhoef (2003) and Wangenheim (2003), which show a positive moderating effect of relationship length on customer loyalty. This study confirms those statements by demonstrating a positive effect of platform value on customer retention and cross-buying, about which data is obtained through platform activity. The positive significant effect of this driver on customer loyalty shows that customer loyalty can be measured through behavioral data which is continuously gathered in a SaaS platform.

Next to the database driver, this study also highlights that information about hybrid drivers can be successfully obtained through databases instead of surveys. This study confirms the positive effect of communication on customer retention, as well as a positive effect of relationship value and communication on cross-buying. This is in line with several studies, that show the positive effect of these independent drivers on customer loyalty (Mangus et al., 2020; Hartmann and Grahl, 2011; Knott et al., 2002; Hänninen and Karjaluoto, 2017).

However, contrary to H2a, a negative significant effect was obtained between relationship value and customer retention. An explanation for this result can be found in the research by Urry (2015), who state: "Relationships between variables can be non-linear, with abrupt switches occurring so the same 'cause' can, in specific circumstances, produce different effects." Thus, these effects can be more complex than they first appear, as an individual relationship can be more complex in reality. Complexity theory can help create a more accurate understanding of what generates customer loyalty (Russo et al., 2016). Another explanation can be found in the job description of an assigned account manager at Digidata. With the exception of the top 20 largest customers in terms of revenue at Digidata, an assigned account manager at Digidata is the same person who closed the deal between Digidata and the particular customer. Thus, this person is not trained nor has the actual knowledge to perfectly execute the job as an actual account manager. Hence, even though researchers put much time and effort into gathering information about loyalty drivers through surveys, this study showed that information about those drivers might already be at hand through databases.

The last element that was included in this research paper was the moderating effect of relationship length on the relationship between the independent variables and customer loyalty. This study showed a significant positive effect of the moderating effect of relationship length on the relationship between platform value and customer retention and on communication and customer retention, thus complementing research by Verhoef (2003), Verhoef et al. (2002), Wangenheim (2003), and Wallenburg (2009), who confirm that as relationship ages, the intimacy between the customer and the supplier increases. Positive impressions of the relationship can lead to increased loyalty towards the supplier. A significant positive effect of the moderating effect between on social bonds and cross-buying was found. This effect can be explained through the similarity-attraction theory by Byrne (1971), who states that people prefer to work with their significant other, as it helps to maintain balance and support the relationship between two parties. Furthermore, a specific social identity promotes in-group solidarity and fosters cooperation. Hence, companies favor working together because they enjoy the resulting mutual cooperation. However, not all interaction terms were found significant and positive. Contrary to H5e, a significant negative moderating effect of relationship length on the relationship between platform value and cross-buying was found. An explanation for this result can be found in Ping (1993), who states that the decision to exit a current relationship is more complicated than it looks. Relationships can be linked to high switching costs and lack an attractive alternative. Hence, platform value can have a rather negative connotation. When the relationship ages, customers are more experienced and aware of the negative aspect of being dependable on an external service (Verhoef et al., 2002). Services, knowledge, and information that is gathered from the platform can play a crucial role to perform the company’s daily activities, making companies less inclined to purchase more services, since it reduces their switching possibilities and increases switching costs.

Next to the several implications, this study contributes to the current literature in two main ways: a new approach to measure customer loyalty, and an extension on the SaaS literature of customer loyalty drivers.

Hence, this study not only takes a whole new approach by actively using transactional and behavioral customer data of a SaaS company to predict database and hybrid loyalty drivers; it also shows that customer loyalty drivers can be measured through ways other than surveys, using the length (customer retention) and breadth (cross-buying) of the relationship between the customer and a company (Bolton et al., 2004). The digital revolution, which began in the second half of the 20th century, is still ongoing, with companies increasingly using digital devices to perform their daily activities (Brynjolfsson and McAfee, 2011). This study shows that with the increase of available data (Schermann et al., 2014), information about SaaS drivers and constructs can be readily obtained, and practitioners only have to mine and transform the data to discover what these data can reveal.

Furthermore, many studies highlight that loyalty strategies are not equally effective across industries and firms and that different variables have varying outcomes (Eisenbeiss et al., 2014; Haan et al., 2015; Kumar et al., 2013; Rust et al., 2004). While the SaaS industry is growing at a rapid pace (The Business Research Company, 2021) and customer loyalty is becoming an increasingly important topic (Tvrdíková and Koubek, 2011), very few studies focus on drivers of customer loyalty drivers within this industry (Kaiser and Würthner, 2020). Research by Kocaman et al. (2020) indicates that drivers in the SaaS environment might have different effects than in other industries since the SaaS industry is a radically changing environment. Therefore, findings from other research cannot be adapted to the SaaS industry. Thus, this study contributes to the current literature by acknowledging drivers of customer loyalty that have not been analyzed within the SaaS industry before.

5.2 Managerial Implications

This study was introduced by signaling the fast growth in the SaaS industry, which consequently led to an increase in the number of competitors in its market (The Business Research Company, 2021). To maintain their customer base, companies must focus on customer loyalty, as argued by Rosenberg et al. (1984) and Reichheld (1993), who indicate that retaining loyal customers is more profitable than acquiring new ones. The first insight is that companies, especially within the SaaS industry, often have a large amount of unused transactional and behavioral customer data at hand, which they should use to track customer activities and measure customer loyalty levels. As shown in this study, platform dependency, relationship value, and communication can increase customer loyalty. If applicable, companies can gather these metrics and analyze them to improve their loyalty levels. This study not only presents drivers that SaaS companies can use to increase their customers' loyalty; it also acts as an eyeopener for those companies that do not actively use every single customer interaction, sales transaction, operations report, or other valuable data that can be used to shape business strategy, enhance customer experience, and most importantly, increase their customers' loyalty. Second, through behavioral and activity data, companies can track customer loyalty in a continuous way. Interaction with the customer occurs on a daily basis through the platform, e-mails, and phone calls. Through continuous measurement a better overall prediction about customer loyalty can be made, while surveys provide a snapshot, and customers can fill out surveys differently when they are in a particular mood. Up-to-date data about how customers behave towards a company allow for better overall loyalty prediction. Furthermore, companies can promptly react to disturbances in the data, whereas the ability to track live data is not possible with surveys. Lastly, since data are already at hand through databases, expensive surveys do not need to be designed, delivered, returned, and processed, saving companies resources for other critical activities.

Another important recommendation from this study is to formulate the job description of an account manager more carefully. As shown in the results, relationship value has a significant positive effect on cross-buying and a significant negative effect on customer retention. At Digidata, sales representatives who onboarded the customer become their designated account managers, but they have relatively little experience in the role of account manager. Hence, even though they can maximize the positive effect of selling additional services to the customer, they are not trained to retain them. Digidata can focus on training the current sales representatives to become more experienced in account management. If the current role as a sales representative is already full-time, Digidata can hire account managers who have the corresponding competencies and can fulfill the customer's needs. Furthermore, cross-buying is also an indicator of customer loyalty (Stone and Woods, 2000), which makes the role of the sales representative as important compared to account management for customer loyalty. Therefore, both functions should complement each other to optimally transform customers into loyal customers.

Lastly, managers should look beyond the drivers proposed in this study. Limited data were available at Digidata to perform an analysis. Other drivers can impact customer loyalty as well. Research by Goodman (2019) shows that in almost all business sectors, a customer who complains and is satisfied by the complaint's resolution is 30% more loyal than a noncomplainer and 50% more loyal than a complainer who remains dissatisfied. Digidata could gather different measurements, such as the NPS (Reichheld, 2003), or it can actively measure the number of times a customer contacted the support department for complaints or questions (Goodman, 2019).

5.3 Limitations and Future Research

The first limitation of this study is the scope. Due to the scarcity of available data, the study was limited to four drivers and one moderator. As mentioned in Section 2.3, other drivers that have been extensively researched using surveys and in other industries can be added to increase its explanatory power. Goodman (2019) shows that customer complaints impact customer loyalty. Providing discounts to customers is a predictor that has been proven to be effective towards customer loyalty in other industries (Jahromi et al., 2014). Adding other drivers could enrich the knowledge of loyalty drivers in the SaaS domain. Furthermore, Reichheld (2003) shows that the NPS can account for attitudinal loyalty in terms of referrals. This variable can be added to broaden the customer loyalty construct, as mentioned in Section 1.4. Even though the NPS can be asked on a regular basis for continuous measurement, the NPS is a survey question. Future research may discover a way of measuring attitudinal loyalty through databases.

The second limitation can be found within the customer retention variable. Due to time constraints, the average length a customer is with Digidata was chosen as the tipping point between loyal and nonloyal customers. However, other constructs can represent customer retention as well. As argued by Ascarza et al. (2018), predicting customer retention lies at the center of any attempt to calculate customer lifetime value and customer equity. Thus, future research could analyze the effect of drivers mentioned in this study on customer lifetime value and customer equity rather than customer loyalty.

Thirdly, the effect of countries on customer loyalty showed different effects. Significant positive effects were found on customer retention, whereas significant negative effects were found on cross-buying. While not the focus of this study, future research could focus on the reasons for these findings. Research by Lee et al. (2019) shows a significant moderating effect of cultural differences between independent and dependent variables in the service industry. Thus, future research could focus on the effect of cultural differences on customer loyalty drivers in the SaaS industry.

A fourth limitation is the number of similarities found for social bonds. In this study, similarities are only measured using primary industry, commercial sector, country, and legal type. However, other similarities based on shared values and common reference points might also be of importance. Furthermore, Gelderman et al. (2021) show that sustainability has become a business imperative, and Wallenburg (2009) emphasizes the importance of innovation for customers. Thus, an extension of this study can focus on integrating other types of similarities that focus on aspects that impacts society.

Lastly, a limitation is found in the measurement of the dependent variable customer retention. Due to time constraints, this study could not integrate advanced models to represent customer retention.

Aspinall et al. (2001) provide simple types of measurement to measure customer retention. However, these measurements are based on industries in which single purchases occur instead of subscription-based business models, and it is common practice to estimate churn probabilities instead of customer retention in a contractual setting (Fader and Hardie, 2007). However, the focus of this study was on customer loyalty. Research by Fader and Hardie (2007) explored different advanced models representing customer retention in a contractual setting, which future research could implement to obtain an additional measurement of customer retention.

References

Achache, A., Baaziz, A., and Sari, T. (2020). The Impact of Data Mining and SaaS-Cloud Computing: A Review. *Data Science and Applications*, *3*(1), 23-35.

Akinci, S., Kaynak, E., Atilgan, E., and Aksoy, Ş. (2007). Where does the logistic regression analysis stand in marketing literature? A comparison of the market positioning of prominent marketing journals. *European Journal of Marketing*.

Ambulkar, B., and Borkar, V. (2012, April). Data mining in cloud computing. In *MPGI National Multi Conference* (Vol. 2012).

Anderson, E.W., Mittal, V., (2000). Strengthening the satisfaction profit chain. *Journal of Services Research 3* (2), 107–120.

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., and Schrift, R. (2018). In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, *5*(1), 65-81.

Aspinall, E., Nancarrow, C., and Stone, M. (2001). The meaning and measurement of customer retention. *Journal of Targeting, Measurement and Analysis for Marketing*, *10*(1), 79-87

Backman, S.J. and Crompton, J.L. (1991). "The usefulness of selected variables for predicting activity loyalty*", Leisure Sciences*, Vol. 13, pp. 205-20.

Balaji, M. S. (2015). Investing in customer loyalty: the moderating role of relational characteristics. *Service Business*, *9*(1), 17-40.

Bardauskaite, I. (2014). Loyalty in the business-to-business service context: A literature review and proposed framework. *Journal of Relationship Marketing,* 13(1), 28-69.

Berry, W. D., DeMeritt, J. H., and Esarey, J. (2010). Testing for interaction in binary logit and probit models: is a product term essential?. *American Journal of Political Science*, *54*(1), 248-266.

Bill, F., Feurer, S., and Klarmann, M. (2020). Salesperson social media use in business-to-business relationships: An empirical test of an integrative framework linking antecedents and consequences. *Journal of the Academy of Marketing Science*, 1-19.

Bloemer, J. M. and Kasper, H. D. (1994). 'The impact of satisfaction on brand loyalty: Urging on classifying satisfaction and brand loyalty', *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 7.

Bó, D. G., Milan, G. S., and Toni, D. D. (2018). Proposal and validation of a theoretical model of customer retention determinants in a service environment. RAUSP Management Journal, 53(2), 202–213.

Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing science, 17*(1), 45-65.

Bolton, R. N., Lemon, K. N., and Verhoef, P. C. (2004). The theoretical underpinnings of customer asset management: A framework and propositions for future research. *Journal of the Academy of Marketing Science*, *32*(3), 271-292.

Bondar, L. Wittkopf W. And Lacki T. (2007). Turning The Corner In Automotive Marketing. *A Research Report By Carlson Marketing And Peppers And Rogers Group.*

Brynjolfsson, E., and McAfee, A. (2011). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee.

Byrne, D. (1971). The Attraction Paradigm, New York: Ac.

Buckinx, W., Verstraeten, G., and Van den Poel, D. (2007). Predicting customer loyalty using the internal transactional database. *Expert systems with applications*, *32*(1), 125-134.

The Business Research Company. (2021, January). *Software As A Service (SaaS) Global Market Report 2021: COVID 19 Impact And Recovery To 2030.* https://www.thebusinessresearchcompany.com/report/software-as-a-service-saas-global-market-report

Callarisa Fiol, L. J., Bigne Alcaniz, E., Moliner Tena, M. A., and Garcia, J. S. (2009). Customer loyalty in clusters: perceived value and satisfaction as antecedents. *Journal of Business-to-Business Marketing, 16*(3), 276-316.

Cater, B., and Zabkar, V. (2009). Antecedents and consequences of commitment in marketing research services: The client's perspective. *Industrial Marketing Management, 38*(7), 785-797.

Chen, P-Y., Hitt, L.M., (2002). Measuring switching costs and the determinants of customer retention in internet-enabled business: A study of the online brokerage industry. *Information Systems Research.* 13 (3): pp. 255-274.

Chen, S. C. (2015). Customer value and customer loyalty: Is competition a missing link? *Journal of retailing and consumer services, 22*, 107-116.

Cooil, B., Keiningham, T. L., Aksoy, L., and Hsu, M. (2007). A longitudinal analysis of customer satisfaction and share of wallet: Investigating the moderating effect of customer characteristics. *Journal of marketing,* 71(1), 67-83.

Coulter, K. S., and Coulter, R. A. (2002). Determinants of trust in a service provider: the moderating role of length of relationship. *Journal of services marketing*.

Coyles, S. and T.C. Gokey, (2005). 'Customer Retention Is Not Enough'. *Journal of Consumer Marketing*, Vol. 22, No. 2 (April), pp. 101–05.

Curtis, T. (2009). Customer Satisfaction, Loyalty, and Repurchase: Meta-Analytical Review, and Theoretical and Empirical Evidence of Loyalty and Repurchase Differences.

Dagger, T. S., Danaher, P. J., and Gibbs, B. J. (2009). How often versus how long: The interplay of contact frequency and relationship duration in customer-reported service relationship strength. *Journal of Service Research*, *11*(4), 371-388.

Day, G.S. (1969). "A two-dimensional concept of brand loyalty", *Journal of Advertising Research*, Vol. 9 No. 3, pp. 29-35.

Dempsey, D., and Kelliher, F. (2018). Industry Trends in Cloud Computing. *Pallgrave McMillian, New York*.

Dreiseitl, S., and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, *35*(5-6), 352-359.

Eggert, A., Henseler, J., and Hollmann, S. (2012). Who owns the customer? Disentangling customer loyalty in indirect distribution channels. *Journal of supply chain management, 48*(2), 75-92.

Eisenbeiss, M., Cornelißen, M., Backhaus, K., and Hoyer, W. D. (2014). Nonlinear and asymmetric returns on customer satisfaction: do they vary across situations and consumers? *Journal of the Academy of Marketing Science, 42*(3), 242–263.

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological methods*, *8*(3), 322.

Fader, P. S., and Hardie, B. G. (2007). How to project customer retention. *Journal of Interactive Marketing*, *21*(1), 76-90.

Figalist, I., Dieffenbacher, M., Eigner, I., Bosch, J., Olsson, H. H., and Elsner, C. (2020). Mining Customer Satisfaction on B2B Online Platforms using Service Quality and Web Usage Metrics. In *2020 27th Asia-Pacific Software Engineering Conference (APSEC)* (pp. 435-444). IEEE.

Friman, M., Gärling, T., Millett, B., Mattsson, J., and Johnston, R. (2002). An analysis of international business-to-business relationships based on the Commitment–Trust theory. *Industrial marketing management,* 31(5), 403-409.

Gil-Saura, I., Ruiz Molina, M. E. and Servera Frances, D. (2008). Logistic service quality and technology: a comparison between supplier–retailer and retailer–consumer relationships. *The International Review of Retail, Distribution and Consumer Research 18* (5), 495-510.

Goodman, J. (2019). *Strategic customer service: Managing the customer experience to increase positive word of mouth, build loyalty, and maximize profits*. Amacom.

Goyal, S. (2014). Public vs private vs hybrid vs community-cloud computing: a critical review. *International Journal of Computer Network and Information Security*, *6*(3), 20.

Gupta, S., and Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing science*, *25*(6), 718-739.

Gustafsson, A., Johnson, M. D., and Roos, I. (2005). The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of marketing,* 69(4), 210-218.

Haan, E., de, Verhoef, P. C., and Wiesel, T. (2015). The predictive ability of different customer feedback metrics for retention. *International Journal of Research in Marketing, 32*(2), 195–206.

Hagigi, M., and Sivakumar, K. (2009). Managing diverse risks: An integrative framework. *Journal of International Management*, *15*(3), 286-295.

Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International journal of service industry management.*

Hänninen, N., and Karjaluoto, H. (2017). The effect of marketing communication on business relationship loyalty. *Marketing Intelligence & Planning*.

Hartmann, E. V. I., and De Grahl, A. (2011). The flexibility of logistics service providers and its impact on customer loyalty–an empirical study. In *Success Factors in Logistics Outsourcing* (pp. 7-51). Gabler Verlag.

Jacoby, J. and Chestnut, R.W. (1978). *Brand Loyalty,* Wiley, New York, NY.

Jahromi, A. T., Stakhovych, S., and Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, *43*(7), 1258-1268.

Javed, F., and Cheema, S. (2017). Customer satisfaction and customer perceived value and its impact on customer loyalty: the mediational role of customer relationship management. *The Journal of Internet Banking and Commerce*, 1-14.

Jones, M.A., and Reynolds, K.E., (2006). The role of retailer interest on shopping behavior. *Journal of Retailing 82 (2)*, 115–126.

Kaiser, L. K., and Würthner, A. F. (2020). Keeping SaaS business clients loyal: An exploratory multi-case study on how to design loyalty initiatives.

Keiningham, T. L., Bruce C., Tor W. Andreassen, and Aksoy, L. (2007). "A Longitudinal Examination of Net Promoter and Firm Revenue Growth.," Journal of Marketing, 71 (3), 39–51.

Khatibi, A.A., Ismail, H., Thyagarajan, V. (2002). What drives customer loyalty: an analysis from the telecommunications industry*. Journal of Targeting, Measurement & Analysis for Marketing 11* (1), 34–43.

Kholief, M. (2021). RFMx-An Approach for an Enhanced RFM Analysis of Customers Multiple-Product Transaction Based Data. In *Informatics and Cybernetics in Intelligent Systems: Proceedings of 10th Computer Science On-line Conference 2021* (Vol. 3, p. 456). Springer Nature.

King, J. E. (2008). Binary logistic regression. *Best practices in quantitative methods*, 358-384.

Knott, A., Hayes, A., and Neslin, S. A. (2002). Next-product-to-buy models for cross-selling applications. *Journal of interactive Marketing*, *16*(3), 59-75.

Kocaman, B., Gelper, S. and Langerak, F. (2020). Working paper.

Kofi Amoako, G., Baadu Adams, G., and Doe, J. K. (2020). Intention to continue relationship in B to B setting–case of Guinness Ghana breweries–some preliminary findings. *Journal of Business-to-Business Marketing, 27*(1), 81-94.

Kuehn, A. A. (1962). Consumer Brand Choice-As a Learning Process. *Journal of Advertising Research*, 2, 10-17.

Kumar, V., Dalla Pozza, I., and Ganesh, J. (2013). Revisiting the satisfaction–loyalty relationship: empirical generalizations and directions for future research. *Journal of retailing, 89*(3), 246-262.

Lages, L. F., Lancastre, A., and Lages, C. (2008). The B2B-RELPERF scale and scorecard: Bringing relationship marketing theory into business-to-business practice. *Industrial Marketing Management, 37*(6), 686-697.

Lavine, H. G., Johnston, C. D., and Steenbergen, M. R. (2012). *The ambivalent partisan: How critical loyalty promotes democracy*. Oxford University Press.

Lee, Y. W., and Bellman, S. (2008). An augmented model of customer loyalty for organizational purchasing of financial services. *Journal of Business-to-Business Marketing*, *15*(3), 290-322.

Lee, M., Kang, M., and Kang, J. (2019). Cultural influences on B2B service quality-satisfaction-loyalty. *The Service Industries Journal*, *39*(3-4), 229-249.

Lipstein, B. (1959). The dynamics of brand loyalty and brand switching. In *Proceedings of the fifth annual conference of the advertising research foundation* (pp. 101-108). New York: Advertising Research Foundation.

Liu, T. C., and Wu, L. W. (2007). Customer retention and cross-buying in the banking industry: An integration of service attributes, satisfaction and trust. *Journal of financial services marketing, 12*(2), 132-145.

Mäkilä T., Järvi A., Rönkkö M., Nissilä J. (2010). How to Define Software-as-a-Service – An Empirical Study of Finnish SaaS Providers. In: Tyrväinen P., Jansen S., Cusumano M.A. (eds) Software Business. ICSOB 2010. Lecture Notes in Business Information Processing, vol 51. Springer, Berlin, Heidelberg.

Mägi, A.W. (1999). "Store loyalty? An empirical study of grocery shopping", PhD thesis, The Economic Research Institute, Stockholm, Sweden.

Mangus, S. M., Jones, E., Folse, J. A. G., and Sridhar, S. (2020). The interplay between business and personal trust on relationship performance in conditions of market turbulence. *Journal of the Academy of Marketing Science, 48*, 1138-1155.

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., and Bitner, M. J. (2000). Self-service technologies: understanding customer satisfaction with technology-based service encounters. *Journal of marketing*, *64*(3), 50-64.

Narayandas, D. (1998), "Measuring and managing the benefits of customer retention: an empirical investigation", *Journal of Service Research,* Vol. 1 No. 2, pp. 108-28.

Ndubisi, N. O. (2007). Relationship marketing and customer loyalty. *Marketing intelligence & planning.*

Ngobo, P. V. (2004). Drivers of customers' cross-buying intentions. *European Journal of Marketing.*

Norton, E. C., Dowd, B. E., and Maciejewski, M. L. (2018). Odds ratios—current best practice and use. *Jama*, *320*(1), 84-85.

Nyadzayo, M. W., Matanda, M. J., and Rajaguru, R. (2018). The determinants of franchise brand loyalty in B2B markets: An emerging market perspective. *Journal of Business Research*, *86*, 435-445.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, *41*(5), 673-690.

Ojala, A. (2012). Comparison of different revenue models in SaaS, Proceedings of 5th Computer Games, Multimedia and Allied Technology Conference.

Ojala, A. (2013). 'Software-as-a-Service Revenue Models', IT Professional, *15*(3).

Oliver, R. (1999). *Satisfaction: A Behavioral Perspective on the Consumer* (New York : McGraw-Hill).

Ozuem, W., Thomas, T., and Lancaster, G. (2016). The influence of customer loyalty on small island economies: an empirical and exploratory study. *Journal of Strategic Marketing*, *24*(6), 447-469.

Palmatier, R. W., Dant, R. P., Grewal, D., and Evans, K. (2005). A meta-analysis of the nomological network surrounding relationship marketing. *Marketing Science Institute, Mass, Cambridge.*

Pan, Y., Sheng, S., and Xie, F. T. (2012). Antecedents of customer loyalty: An empirical synthesis and reexamination. *Journal of retailing and consumer services, 19*(1), 150-158.

Patterson, P. G., Johnson, L. W., and Spreng, R. A. (1996). Modeling the determinants of customer satisfaction for business-to-business professional services. *Journal of the academy of marketing science, 25*(1), 4-17.

Peterson, R. A., and Wilson, W. R. (1992). Measuring customer satisfaction: fact and artifact. *Journal of the academy of marketing science, 20*(1), 61-71.

Pineda, J., and Izaret, J.-M. (2013). Profiting from the Cloud: How to Master Software as a Service. Retrieved May, 10, 2020, from https://www.bcg.com/publications/2013/pricing-technology-software-profiting-from-the-cloud.aspx.

Pritchard, M. P., and Howard, D. R. (1997). The loyal traveler: Examining a typology of service patronage. *Journal of travel research*, *35*(4), 2-10.

Ramaseshan, B., Rabbanee, F. K., and Hui, L. T. H. (2013). Effects of customer equity drivers on customer loyalty in B2B context. *Journal of Business & Industrial Marketing*.

Ramaseshan, B., Wirtz, J. and Georgi, D. (2017). "The enhanced loyalty drivers of customers acquired through referral reward programs", *Journal of Service Management,* Vol. 28 No. 4, pp. 687-706.

Rauyruen, P. and Miller, K.E. (2007). "Relationship quality as a predictor of B2B customer loyalty", *Journal of Business Research*, Vol. 60 No. 1, pp. 21-31.

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review*, *81*(12), 46-55.

Reichheld, F. F., and Teal, T. (1996). The loyalty effect: The hidden force behind growth, profits and lasting. *Harvard Business School Publications, Boston.*

Reinartz, W., Thomas, J. S., and Bascoul, G. (2008). Investigating cross-buying and customer loyalty. *Journal of Interactive Marketing*, *22*(1), 5-20.

Rosenberg, L., Czepiel, J.A., (1984). A Marketing Approach to Customer Retention. *Journal of Consumer Marketing.* Vol. 1, pp. 45-51.

Russo, I., Confente, I., Gligor, D. M., and Autry, C. W. (2016). To be or not to be (loyal): Is there a recipe for customer loyalty in the B2B context? *Journal of Business Research*, *69*(2), 888–896.

Rust, R. T., Lemon, K. N., and Zeithaml, V. A. (2004). Return on marketing: using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109–127.

Sashi, C. M. (2012). Customer engagement, buyer-seller relationships, and social media. *Management decision*.

Satyanarayana, S. (2012). *Cloud computing: Saas. Computer Sciences and Telecommunications*, (4), 76–79.

Scheer, L. K., Miao, C. F., and Garrett, J. (2010). The effects of supplier capabilities on industrial customers' loyalty: the role of dependence. *Journal of the Academy of Marketing Science, 38*(1), 90-104.

Schermann, M., Hemsen, H., Buchmüller, C., Bitter, T., Krcmar, H., Markl, V., and Hoeren, T. (2014). Big data. *Business & Information Systems Engineering*, *6*(5), 261-266.

Seethamraju, R. (2005). Adoption of Software as a Service (SaaS) Enterprise Resource Planning (ERP) Systems in Small and Medium Sized Enterprises (SMEs). *Information Systems Frontier 17*, 475–492.

Seiders, K., Voss, G. B., Grewal, D., and Godfrey, A. L. (2005). Do satisfied customers buy more? Examining moderating influences in a retailing context. *Journal of marketing*, *69*(4), 26-43.

Singh, R., and Khan, I. A. (2012). An approach to increase customer retention and loyalty in B2C world. *International journal of scientific and research publications*, *2*(6), 1-5.

Smeden, M. V., Groot, J. D., Moons, K., Collins, G., Altman, D., Eijkemans, M., and Reitsma, J. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*.

Steinhoff, L., Arli, D., Weaven, S., and Kozlenkova, I. V. (2019). Online relationship marketing. *Journal of the Academy of marketing science*, *47*(3), 369-393.

Stock, R. M. (2005). Can customer satisfaction decrease price sensitivity in business-to-business markets?. *Journal of Business-to-Business Marketing, 12*(3), 59-87.

Stock, R. M., and Zacharias, N. A. (2013). Two sides of the same coin: How do different dimensions of product program innovativeness affect customer loyalty?. *Journal of Product Innovation Management, 30*(3), 516-532.

Stoel, L., Wickliffe, V., Lee, K.H. (2004). Attribute beliefs and spending as ante- cedents to shopping value. *Journal of Business Research 57*, 1067–1073.

Stone, M. and Woods, A. (2000). 'Making sense of cross-purchasing: A note', Journal of Financial Services Marketing, Vol. 5, No. 2, pp. 129–134.

Torgerson, C., Hall, J., and Light, K. (2012). Systematic reviews. *Research Methods and Methodologies in Education. London: Sage*, 217-230.

Tucker, W.T. (1964). "The development of brand loyalty", *Journal of Marketing Research,* Vol. 1 No. 3, pp. 32-35.

Tvrdíková, M., and Koubek, O. (2011). The Use of Cloud Computing (SaaS) for Small and Medium Enterprises to Improve the Quality of Their Information Systems. In *IDIMT* (pp. 389-390).

Ulaga W. (2003). Capturing value creation in business relationships: A customer perspective. *Industrial Marketing Management*, 32 (8), pp. 677-693.

Urry, J. (2005). The complexity turn. *Theory, culture & society*, *22*(5), 1-14.

Verhoef, P. C. (2003). "Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development*," Journal of Marketing,* 67 (October), 30-45.

Verhoef, P. C., Franses, P. H., and Hoekstra, J. C. (2002). The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: does age of relationship matter?. *Journal of the academy of marketing science*, *30*(3), 202-216.

Vroman, H. William; Reichheld, Frederick F. (1996). "The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value". *Academy of Management Perspectives.* Harvard Business School Press. 10 (1): 88–90.

Wangenheim, F. (2003). Situational characteristics as moderators of the satisfaction-loyalty link: an investigation in a business-to-business context. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *16*, 145.

Wallenburg, C. M. (2009). Innovation in logistics outsourcing relationships: proactive improvement by logistics service providers as a driver of customer loyalty. *Journal of supply chain management, 45*(2), 75-93.

Wang, C. Y., and Wu, L. W. (2012). Customer loyalty and the role of relationship length. *Managing Service Quality: An International Journal*.

Wiesel, T., Verhoef, P., and De Haan, E. (2012). There is no single best measure of your customers.

Wirth, R., and Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-40).

Xie, E., Peng, M. W., and Zhao, W. (2013). Uncertainties, resources, and supplier selection in an emerging economy. *Asia Pacific Journal of Management,* 30(4), 1219-1242.

Yuan, C. L., Moon, H., Kim, K. H., Wang, S., and Yu, X. (2020). Third-party organization endorsement impacts on perceived value and B2B customer loyalty. *Industrial Marketing Management, 90*, 221-230.

Zeithaml, V. A., Berry, L. L., and Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of marketing, 60*(2), 31-46.

Appendices

Appendix A: Data Sources

Table A.1 : Overview of the different data sources available at Digidata

Database	Type of Information	Description
Customer relationship management (CRM) system	Customer/External	Internal CRM-based platform that provides an accessible overview of all supporting systems (Unit4 CODA, Salesforce, Customer Data Platform) and displays externally loaded data (type of organization, number of employees, industry, etc.).
Unit4 CODA	Financial	Information regarding monthly invoices, contract length, and different products used by customers.
Salesforce	Customer	Collects customer-specific information, such as contact details, demographics, and VAT numbers.
Customer Data Platform	Customer	Marketing-related data source in the form of a platform. This platform collects the behavioral/activity information from customer who create an account on the platform.

Appendix B: Structural Literature Review

During the first phase of this research project, a structural literature review was performed. This review supports and expands previous research in the field of customer loyalty within a B2B environment. The search engine “Scopus” was used to analyze and select the studies. The keywords used to select the appropriate literature are “customer loyalty,” “business-to-business” or “B2B,” and “drivers” or “antecedents” or “variables” or “enablers.” After this search query, 50 journals were found. Abstract were analyzed from this search query to assess the relevance of each article.

Inclusion and exclusion criteria. The first inclusion criterion was journals from the Amsterdam Business School (ABS) Journal List with an A* or A rating. The second inclusion criterion was journals from the Erasmus Research Institute of Management (ERIM) journal list with a STAR or P rating. The third inclusion criterion was journals listed in both the ABS and the ERIM journals. If journals were listed in both lists, they were automatically included. Table 15 shows the journals that were included with their corresponding ratings. After editing the search query, a total of 17 articles were found. The researcher read all articles and assessed all journals. One article— “Salesperson social media use in B2B relationships: An empirical test of an integrative framework linking antecedents and consequences” — was excluded since it was not relevant for this study and focused on the drivers of salespersons in social media use.

Table B.1: Journal Ranking List

Journal	ABS rating	ERIM rating
Industrial Marketing Management	B	S
Journal Of Business To Business Marketing	B	S
Journal of Business Research	B	S
Journal of Supply Chain Management	A	P
Journal Of The Academy Of Marketing Sciences	A*	STAR
Decision Sciences	A	P
Journal Of Product Innovation Management	A	P
Journal Of Retailing And Consumer Services	B	S

Appendix C: Pearson Correlation Coefficients

Table C.1: Pearson correlation coefficients

	Customer retention	Cross-buying	Platform Value	Relationship Value	Communication	Social Bonds	Relationship Length	Firm Size	Belgium	Germany (REF=Netherlands)	France (REF=Netherlands)	Other (REF=Netherlands)	South Africa (REF=Netherlands)
Customer Retention	1	-.004	.045*	-.268**	-.057*	-.078**	.816**	.045*	.165**	.153**	-.008	-.128**	-.158**
Cross-buying	-.004	1	.284**	.179**	.192**	.019	.053*	.009	-.042	-.064**	-.051*	.029	.038
Platform Value		.284**	1	.265**	.314**	.031	.183**	.077**	-.030	-.023	-.069**	.039	.103**
Relationship Value	-.268**	.179**	.265**	1	.116**	-.010	-.205**	.028	-.040	-.060**	.075**	.120**	.134**
Communication	-.057*	.192**	.314**	.116**	1	.001	-.018	.102**	-.051*	-.051*	-.042	.016	.042
Social Bonds	-.078**	.019	.031	-.010	.001	1	-.095**	-.125**	-.243**	-.122**	-.201**	-.206**	-.089**
Relationship Length	.816**	.053*	.183**	-.205**	-.018	-.095**	1	.057*	.165**	.138**	-.004	-.146**	-.145**
Firm Size	.045*	.009	.077**	.028	.102**	-.125**	.057*	1	-.020	.080**	.049*	-.050*	.005
Belgium (REF = Netherlands)	.165**	-.042	-.030	-.040	-.051*	-.243**	.165**	-.020	1	-.111*	-.081**	-.116**	-.063**
Germany (REF = Netherlands)	.153**	-.064**	-.023	-.060**	-.051*	-.122**	.138**	.080**	-.111*	1	-.074**	-.151**	-.057*
France (REF = Netherlands)	-.008	-.051*	-.069**	.075**	-.042	-.201**	-.004	049*	-.081**	-.074**	1	-.110**	-.041
Other (REF = Netherlands)	-.128**	.029	.039	.120**	.016	-.201**	-.146**	-.050*	-.116**	-.151**	-.110**	1	-.085**
South Africa (REF = Netherlands)	-.158**	.038	.103**	.134**	.042	-.089**	-.145**	.005	-.063**	-.057*	-.041	-.085**	1

Appendix D: VIF Scores

Table D.1: VIF Scores

Factor	**VIF Score**
Platform Value	1.323
Relationship Value	1.216
Communication	1.211
Social Bonds	1.419
Moderator	
Contract Length	3.098
Platform Value x Relationship Length	1.200
Relationship Value x Relationship Length	3.093
Communication x Relationship Length	1.174
Social Bonds x Relationship Length	1.046
Control variables	
Firm Size	1.049
Countries	
Belgium	1.305
France	1.186
Germany	1.175
Netherlands	*Reference*
Others	1.346
South Africa	1.155