Customer Lifetime Value Drivers of Independent Retailers within South Africa's Informal Trade Environment: An Empirical Case Study

A Research Report

submitted to

Wits Business School

in partial fulfillment of the requirements of the degree

Master of Management in Strategic Marketing

by

Balekile Catherine Msibi

Supervisor: Neale Penman

March 2021

Abstract

Customer Lifetime Value (CLV) is a "technique for analysing all the potential value purchase values from a single customer over the lifetime of a relationship with a supplier or service provider" (Doyle, 2016). It is a valuable measure that can be used to segment customers according to their profitability in order to implement customer-focused marketing strategies. The study looks at independent retailers that operate within the informal trade environment and are directly serviced by a consumer packaged goods (CPG) company within the snacking category. Through the analysis of these customers, this study aims to provide empirical evidence to support the utilisation of CLV as a customer segmentation tool within a business-to-business (B2B) setting based on CLV drivers. This study also seeks to understand whether short-term macroeconomic indicators such as inflation do indeed have an effect on CLV. Through the use of the Venkatesan-Kumar method for calculating CLV as well as classification trees, the study aims to create a decision rule process whose end result will be a CLV classification system. The results will equip companies with the tools and ability to identify which customer segment a customer belongs to in order to ensure that they use the appropriate customer relationship management strategies to drive customer development and retention.

Keywords*: Customer Equity, Business-to-Business, Customer Lifetime Value, Consumer Packaged Goods, Classification and Regression Trees*

Declaration

I hereby declare that this research report is my own unaided work. It is being submitted for the Degree of Master of Management in Strategic Marketing at the Wits Business School. It has not been submitted for any other degree or examination at any other university.

Dedication

To my Lord God Jesus Christ, Thank you. “I can do all things, through Christ who strengthens me.”

To my daughter, Azaria Natsai Letsokuhle, who throughout the process has been my sole driving force. May God’s hand ever be a guiding light in your life.

To my mother, Nombuyiselo Prudence Sylvia Msibi, for your words of encouragement and constant reminder to finish what I start. I know I may not always be the easiest of your four, but your love never goes unnoticed. Thank you for always being my one true cheerleader.

To my late father, Judas M. Msibi, your approach to life and our upbringing has proved to be one of the most valuable life lessons I have come to appreciate. I have finally done it.

“Inspiration, from whatever the source, arouses feelings within us that rekindle hope, ambition, and determination. It is a momentary whisper of encouragement and reassurance that causes us to become aware of our potential.”

Jim Rohn

Acknowledgements

First and foremost, thanks and appreciation goes out to my supervisor, Mr Neale Penman. Thank you for your guidance and support throughout the process and for pulling me out of the woods when I couldn't see the forest for the trees.

To these great individuals: Unathi, Jerome, Gert and Matias – thank you for allowing me the time and patience to indulge me as my dedicated sounding boards. Your experience, expertise and constant reminders to make it to the finish line were always and will continue to be always welcome.

To Mr David Kaplan, thank you for all your assistance despite the tight timelines.

To my friends and classmates, a special mention to Sibongani Malumo and Harold Mogale; I could have not asked for anyone better to stand in my corner. Your reassurance and pep talks to stay the course were the oil needed to keep the lamp burning bright.

To my adversities, without you I would never have realised the faith, resilience and strength I possessed.

"You may have to fight a battle more than once to win it."

Margaret Thatcher

Contents

List of Figures

List of Tables

List of Abbreviations

ABS	Average Basket Size
B2B	Business-to-Business
B2C	Business-to-Consumer
CART	Classification and Regression Trees
CBBE	Customer-Based Brand Equity
CHAID	Chi-Squared Automatic Interaction Detection
CLV	Customer Lifetime Value
CP	Complexity Parameter
CPG	Consumer Packaged Goods
CPI	Consumer Price Index
CS	Cross Shopping
DC	Distribution Centre
DMS	Distribution Management System
GDP	Gross Domestic Product
HHC	Handheld Terminal
IPR	Inter-Purchase Rate
KPI	Key Performance Indicator
MMSM	Master of Management in Strategic Marketing
PV	Present Value
RFM	Recency, Frequency and Monetary
SDV	Store Door Value
SP	Set Point
VK	Venkatesan-Kumar

1 Introduction

Customer Lifetime Value (CLV) is a "technique for analysing all the potential value purchase values from a single customer over the lifetime of a relationship with a supplier or service provider" (Doyle, 2016). It is a valuable measure that can be used to assess the profitability of a customer to a business. With the increasing popularity of consumer-centrism and the growing need to move towards a more individualised customer relationship management approach, it is becoming more pertinent that companies understand who their customers are, what drives them and what value they bring to the business. CLV is a tool that can help with this as it looks at the different drivers of customer behaviour and can group customers into homogenous segments based on their profitability (Frow & Payne, 2009). These segments can in turn be used to formulate customer-focused marketing strategies to drive customer development and retention through the adjustment of drivers such as basket size, purchase frequency and basket composition.

1.1 Background

Africa and in particular, Sub-Saharan Africa, has promising economic growth prospects which are set to fuel future global economic growth (Wroughton, 2018). This view is confirmed by the World Bank with data indicating that emerging markets are set to grow at a faster pace of +4.6% in 2021 in comparison to advanced economies, which only had a growth projection of +3.9% (World Bank, 2020a). Sub-Saharan Africa, although set to have slightly subdued growth of +3.1% in 2021, will have regional pockets that will experience strong growth such as Cote d'Ivoire and Ghana in the west at +8.7% and +3.4%, as well as Kenya, Uganda and Ethiopia in the east at +5.2%, +3.7% and +3.6%, respectively (World Bank, 2020a). Corporations with global operations looking to expand market territories find economies with such high growth prospects attractive to invest in, thus making Sub-Saharan Africa a region of interest for most multinational corporations. Prior to a corporation setting up operations within a market, it first needs to understand the consumer landscape and socio-economic dynamics specific to that market.

The first notable difference of the African landscape in comparison to developed economies is the income disparity. Although emerging markets have high growth prospects, the GDP per capita remains significantly low at USD1,585 for Sub-Saharan countries, in comparison to the Global average of USD11,429 (World Bank, 2020b). This is further emphasised by the Gini index tracked by the World Bank, as it shows greater income inequality for some Sub-Saharan markets in comparison to high-income countries (World Bank, 2020b), as depicted by Figure 1 below.

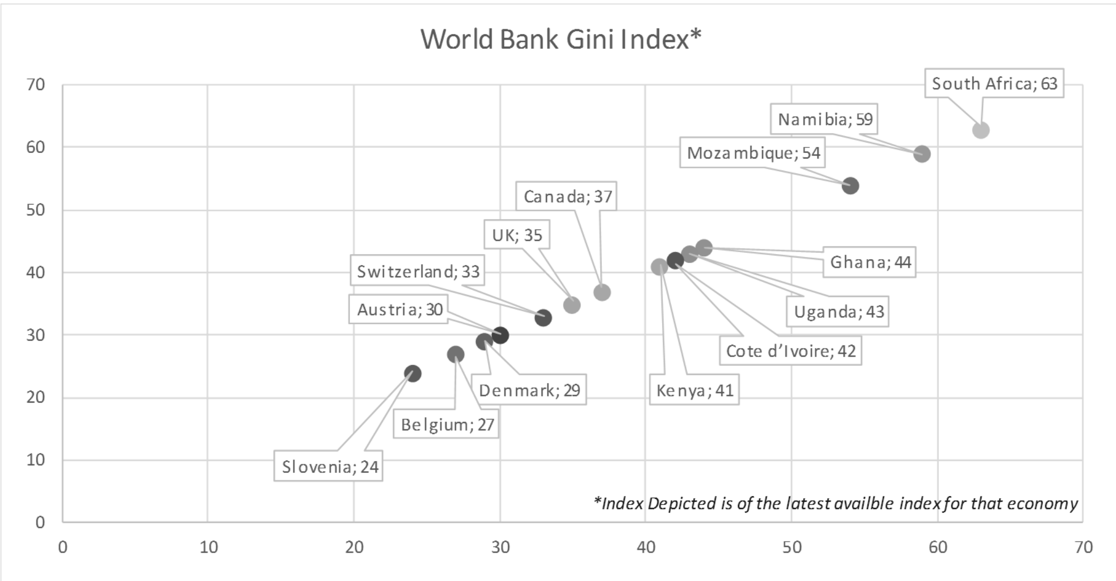

Figure 1: Gini Index by Country (World Bank, 2020b)

Fisher, Johnson, Smeeding, and Thompson (2020) confirmed that "low-wealth households had a higher marginal propensity to consume than high-wealth households", and thus their consumption patterns tend to respond more quickly to changes in income (Fisher et al., 2020). This means that with only "1% to 2% of the population classified as middle class" (Johnson, 2018), the majority of consumers within Sub-Saharan Africa are living in low-wealth households. For these households, as the economy expands and their living standards and disposable income improve, so too will their purchasing power. This is a key consideration for any corporate looking to enter or further invest in a market as it gives an indication of the possible future earnings that can be made, as the level of disposable income directly impacts the ability to engage in economic activity.

This study will focus on the CPG industry, which is an industry characterised by goods that are "highly in-demand, sold quickly and are affordable" (Corproate Finance Institute, 2015). Within this industry, most manufacturers do not retail directly to their end consumer, but rather use a network of retail partners to get their product to the end consumer. This network can differ vastly depending on the socio-economic aspects of the target market, as depicted in Figure 2 below.

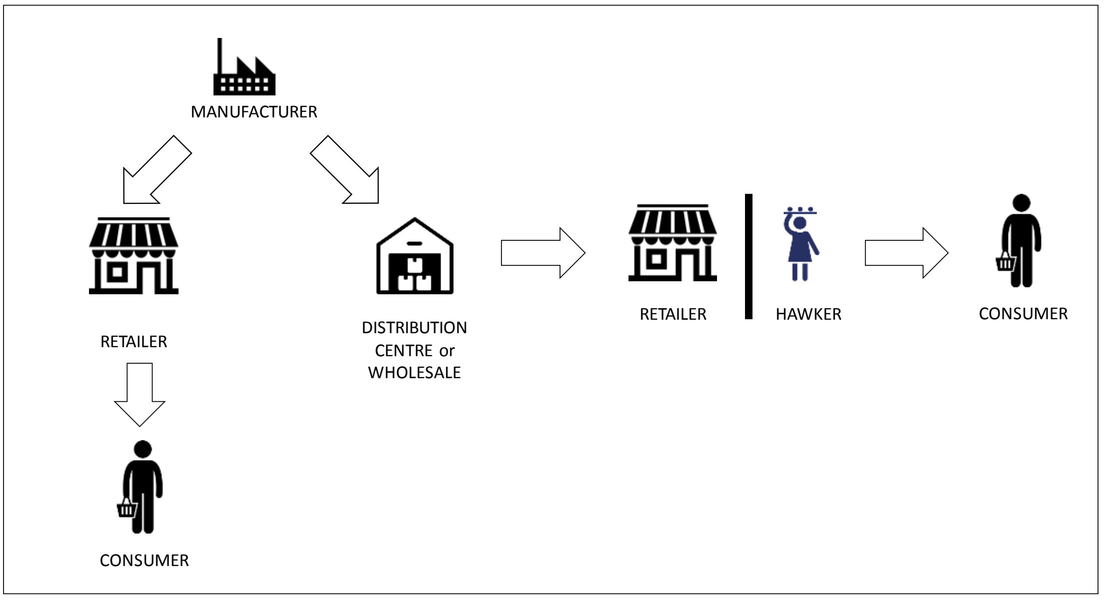

Figure 2: Various Route-to-Market to End Consumer

Based on Figure 2, it can be said that the CPG industry may resemble a hybrid industry with both a business-to-business (B2B) and business-to-consumer (B2C) component. The B2B component refers to the interaction between the manufacturer and the intermediary or retail partner, through its sales function. Whilst the B2C component is the interaction between the manufacturer and the end consumer, through its marketing function. This study will look specifically at the B2B component of the CPG value chain.

A South African CPG manufacturer that manufactures snacking goods will be used for this empirical study. Snacking goods make an ideal case study as they are a category that is shopped on impulse. This implies that it is not a planned purchase and therefore is a category whose performance is sensitive to changes in push and pull marketing initiatives. The dynamic of the snacking category is that it is occupied by many brands and options for consumers to choose from. In such a business environment, growing and maintaining market share becomes a crucial key performance indicator that companies try and achieve by outdoing their competitors through both marketing initiatives and optimised supply chain processes.

Although South Africa is one of the more developed countries within Sub-Saharan Africa, it is still a viable country to use in an empirical study as it also faces similar socio-economic dynamics as other less developed countries on the continent. In 2015, Statistics South Africa (2017) recorded that "30.4 million (55.5%) South Africans were living under the poverty line (R992 per person per month) and 13.8 million (25.2%) under extreme poverty (R441 per person per month)". This is further reiterated by a Gini Index of 63 for South Africa (World Bank, 2020b) and implies that consumers in South Africa also face the same concerns with regard to household expenditure given the income constraints as consumers in less developed countries. A second consideration is that most business in Africa is conducted within the informal sector which can account for 20% of trade sales in more developed countries, such as

South Africa, to 90% of trade sales in less developed countries, such as Nigeria. In order for a business to thrive on the continent, it would need to understand how best to optimise their sales to consumers within the informal sector.

The study will look specifically at the retail partners within the informal trade environment, which will allow the results of the study to be generalisable across markets with similar trade infrastructures. In order for businesses to be able to capture a large share of the informal trade wallet, they need to understand how to optimise the customer equity of the retail touch points to these consumers to ensure the future profitability of the organisation.

1.2 Problem Definition

Customer Lifetime Value is a topical theme in today's marketing circles. CLV has been seen as a way in which marketing executives can become more accountable for marketing initiatives as the approach is centred on "redirecting the marketing focus from mass product one-way communication to a more customer-focused, customised bi-directional communication with personalised offerings" (Rozek & Karlícek, 2014); or simply put, "a more long-term relationship marketing approach" (Rozek & Karlícek, 2014). Gupta and Lehmann (2003) in their study showed that when customers are viewed as assets, CLV can be used as a powerful tool to assist managers in shifting from a product-orientated view to a long-term, relationship-building customer-focused view.

Jain and Singh (2002) in their review identified three main streams of thought in which research on CLV has been undertaken within the marketing field. These areas are: the derivation of CLV; the use of CLV for customer base analysis; and "CLV and its implications for managerially relevant decisions through analytical models" (Jain & Singh, 2002).

Focusing on the second stream of research, previous studies have yielded the following:

i. Hansotia (2004b) posited the use of customer equity scorecards to track "customers' performance over time by looking at cash flows, acquisition costs, operating expenses, add-on selling expenses and retention expenses". These would assist managerial staff in not only being able to assign a financial measure to their performance, but would also assist in streamlining budgets and providing a clearer link between customer equity and shareholder value.

ii. Kumar, Venkatesan, Bohling, and Beckmann (2008) found that using CLV as a customer segmentation tool led to "improved resource allocations for IBM and increased productivity for marketing investments" (Kumar et al., 2008). By using CLV as a customer segmentation tool, their study found that it was able to classify customers into low, medium, high and super-high CLV customers. They then assessed the customer firmographics in each segment in order to identify commonalities in discernible traits that would aid management in quickly identifying which segment a new customer is likely to fall into.

iii. Hansotia (2009) later proposed analytical tools that could be used to enhance customer equity by ensuring that the correct goals are set for each customer segment. These tools included a behavioural customer segmentation tool as well as a product propensity model. The behavioural customer segmentation tool is used to group customers based on their key characteristics. This is done through the use of clustering methods such as K-Means clustering, or tree methods such as chi-square automatic interaction detection (CHAID) or classification and regression trees (CART). Product propensity models are "a system of logit models that estimate the likelihood of a customer buying specific product categories over some future period of time" (Hansotia, 2009). The customer segmentation would assist the marketing and sales functions to understand which strategy to employ when recruiting new customers or driving sales from lapsed or loyal customers. The product propensity model would then assist in identifying the products that are most likely to be taken up by the customers based on their behavioural profile as well as what products to suggest for add-on selling.

iv. Hosseini and Shabani (2015) used a Recency, Frequency and Monetary (RFM) model and amended it by adding a time effect in order to understand the customer's CLV classification over time and then used the results in a K-Means clustering to classify customers. The addition of the time element helped to better segment customers into micro segments, which allowed for more accurate managerial decisions to be implemented by understanding which elements of the RFM model the customer needs an intervention for.

The focus of this study will be the use of classification trees and CLV to provide a customer segmentation framework. There has been limited research on the use of this method and CLV with some studies noted below:

i. Tirenni, Kaiser, and Herrmann (2007) used decision trees in their empirical study to provide "marketing managers with the means to optimise customer equity". Their findings resulted in prediction rules which managerial staff could utilise to classify airplane customers (based on ticket value and seat preference) into the appropriate customer segment in order to ensure that profitability is maximised through appropriate resource allocation.

ii. Haenlein, Kaplan, and Beeser (2007) used CART to create a subset of homogenous groups, from which a first order Markov chain analysis was applied to create a dynamic model that was then used to calculate CLV.

iii. Ekinci, Ulengin, and Uray (2014) used CART to segment customers in a retail bank based on the monetary values of different types of products. These segmentations were then used to understand which promotional strategies would best optimise customer profitability based on the segment the customer belonged to.

iv. Carr, Drennan, and Andrews (2016) used CART to examine how different types of resorts use customer equity strategies. Their study found that chain hotels focused more on acquisition strategies, whilst resort hotels balanced acquisition and retention strategies and independent hotels leaned more towards retention strategies.

This paper aims to add to this body of knowledge through an empirical analysis of the CPG industry by using a CART to classify customers into different CLV bands based on the drivers of CLV.

1.3 Research Question

Can purchase frequency, basket size, basket composition and inflation influence CLV? Can these variables be further used to segment customers into homogenous groups based on their CLV?

1.4 Dissertation Goals

The goal of this dissertation is to create a classification system based on the components of CLV of independent retailers across the South African landscape. The study builds on existing literature, looking at CLV literature through the examination of past purchasing behaviour (basket size, frequency and basket composition/cross-shopping) as well as the use of classification trees to create a decision rule for segmenting customers. This will enable the business to understand the future profitability and customer equity within this traditional trade channel and assist managers in optimising their resource allocation and, ultimately, profits.

1.5 Benefits of Research

In their article on managing customers for value, Kumar, Lemon, and Parasuraman (2006) cited that CLV can be used in creating CLV-based customer segmentation models. In addition, K. N. Lemon and Mark (2006) posited that CLV allows one to better segment customer micro segments by looking at profitability and would be a valuable exercise in a B2B or direct marketing setting. This paper aims to explore this area of research through a B2B empirical study. The benefits of this study will showcase how drivers of CLV can be used early on to classify and segment customers as a function of their purchasing behaviour and macro-economic environment and thus assist the firm in allocating resources appropriately.

1.6 Limitations of the Study

One limitation to the study is that it looks at a single manufacturer in a specific product category, thus its generalisability would be limited to CPG companies with similar traits in terms of firmographics, category dynamics and business model. A secondary limitation is that the study assumes a steady state with respect to CLV and does not consider a transition matrix and the possibility that the B2B customers can move between CLV segments over time.

1.7 Chapter Outline

This dissertation is divided into seven sections:

Chapter One covered the introduction and background to the study.

Chapter Two provides a literature review on customer lifetime value looking at both the aggregate and disaggregate schools of thought with respect to calculating CLV. Based on the literature, the chapter then goes on to identify the hypothesis that will be tested in this study and concludes with a theoretical framework.

Chapter Three gives an account of the research design and methodology. It looks at the source and structure of the data as well as detailing the calculations and analysis to be carried out, concluding with a discussion on validity and reliability for both CLV and CART.

Chapter Four analyses the secondary data and provides the research findings.

Chapter Five discusses the research findings and draws conclusions with respect to the hypotheses tested and their similarities or differences to existing literature.

Chapter Six gives the theoretical, managerial and policy implications as a result of the research findings.

Chapter Seven concludes this study with a section on its contribution to the current literature as well as recommendations for future research.

2 Conceptualisation and Literature Review

Customer Lifetime Value (CLV) and Customer Equity are “powerful and fundamental” tools in driving corporate growth (Taybi & Frankel, 1989). Within the marketing practice, these two concepts are split between two schools of thought: an external, demand-side-orientated view versus an internal, supply-side-orientated view.

2.1 CLV and Customer Equity: External School of Thought

The external school of thought or demand-side approach was introduced by Rust, Zeithaml, and Lemon (2000), who captured the notion of CLV as “the total of the discounted lifetime values of all its customers” and identified value equity, brand equity and retention equity as the drivers of customer equity (Rust et al., 2000). In unpacking the concept, Rust et al. (2000) aimed to move companies away from product-profitability-led strategies to customer-led strategies. This framework aimed to give a company the tools to identify strategic initiatives that would have a long-term impact on customer profitability by understanding “(1) What leads the customer to do business with a company (brand equity); (2) What leads the customer to repurchase (retention/relationship equity); and (3) What influence can the company have on these decisions (value equity)” (Rust et al., 2000). Through this study, Rust et al. (2000) aimed to make companies understand that although they may have a superior product, chasing profitability at the expense of the customer relationship will inevitably lead to the “death spiral”, resulting in a company having to discontinue a product or close down completely (Rust et al., 2000).

2.1.1 Brand Equity

Brand equity is the "customers' perceptions of the brand" (Rust et al., 2000) and is defined as the "difference in outcomes arising from the added value endowed to a product as a result of past marketing activity" (K. L. Keller, 2013). It is "what leads the customer to do business with a company" (Rust et al., 2000) and has been known to influence the performance of a company. In a Business-to-Business (B2B) setting, corporate branding and the corporate brand image are important to a firm's success due to the nature of the relationship that has been built over time (Wiedmann, Hennigs, Schmidt, & Wuestefeld, 2011). This value is more pronounced when it is a heritage brand, as "the brand's roots add to the authenticity and result in a very strong equity for the brand and parent company" (Wiedmann et al., 2011). This is even more apparent when the corporate and brand names are synonymous, for example, Revlon and Revlon, Pepsi and PepsiCo, Colgate and Colgate-Palmolive, Coke and Coca-Cola, and so forth.

Brand equity is driven by brand awareness, customer attitudes as well as customer perceptions towards the brand (Rust et al., 2000). In B2B environments and, in particular, within the consumer packaged goods (CPG) industry, brand equity is synonymous with corporate brand equity that the retail partner ascribes to the supplier and the products they sell. Brand equity can be measured by looking at four dimensions developed by Aaker (1991): awareness, associations, perceived quality and brand loyalty. These dimensions were later improved upon by Keller's customer-based brand equity (CBBE) model, wherein the brand equity can be measured by measuring brand awareness and brand image through a measurement scale (K. L. Keller, 2013). Within a B2B context, brand equity can also be measured by looking at the same dimensions. An empirical study conducted by Biedenbach

(2012) provides a 5-point scale framework for measuring brand equity in a B2B context based on Aaker's four dimensions.

Brand equity is the pinnacle of the customer equity journey, as Stahl, Heitmann, Lehmann and Neslin (2012) have found brand equity, and especially knowledge or awareness, to be "positively related to acquisition and retention".

2.1.2 Retention/Relationship Equity

Relationship equity is defined as "the customer's tendency to stick with the brand above and beyond objective and subjective assessments of the brand" (Rust et al., 2000). Palmatier and Sridhar (2017) better captured the concept of relationship equity as:

> *"the aggregation of relational assets and liabilities, associated with the firm's boundary-spanning employees and social networks linked to the offering or experience, that add to or subtract from the value provided by the firm's offering"*.

Relationship equity is important in a B2B setting as it directly influences customer equity.

Palmatier, Gopalakrishna, and Houston (2006) note that relationship equity can be affected by three types of interventions, namely, structural, financial and social, and these are defined as follows:

> *"Financial efforts include activities such as discounts and free products. Social efforts include activities such as special treatment, entertainment and personalised information. Structural efforts include activities such as customised ordering processes, dedicated personnel and tailored packaging"*.

Structural efforts have been found to "generate short-term positive economic returns" (Palmatier et al., 2006) and, if managed well, to improve customer equity. Servicing plays a crucial part in driving relationship equity as servicing customers directly saves the customer time and money, and thus has a positive impact on the relationship equity (Efanny, Haryanto, Kashif, & Widyanto, 2018; Vogel, Evanschitzky, & Ramaseshan, 2008). Biedenbach, Hultén, and Tarnovskaya (2019) found that human capital “had a major influence on brand associations, perceived quality and brand loyalty” (Biedenbach et al., 2019).

In Palmatier’s (2008) study on relational drivers of customer value, he identifies three relational drivers in this regard, namely, “relationship quality, contact density, and contact authority” (Palmatier, 2008). These drivers fall under what he termed the social efforts of relationship equity. Each of these factors have what is termed “first-order factors”. For relationship quality, these are comprised of trust, commitment, reciprocity norms and exchange efficiency. Contact density refers to the level of interconnectedness that the sales person has with their clients, or more simply put, the breadth of their customer portfolio (Palmatier, 2008). Having a strong network as a salesperson has benefits as it allows one to be better equipped at “building and maintaining relationships” (Palmatier, 2008) due to the positive word-of-mouth publicity generated by their clientele. Contact authority refers to the decision-making abilities of the salesperson. Sales personnel with higher decision-making capabilities tend to perform better and thus drive higher customer value amongst their clients (Palmatier, 2008). In his 2008 study, Palmatier found that these three factors not only have a positive influence on customer equity, but also have synergistic effects, with “the interaction between relationship quality and contact authority having positive effects on customer equity” (Palmatier, 2008).

Other factors influencing relationship equity include pricing and push promotions which influence the customers’ propensity to purchase a product (Efanny et al., 2018). These factors form part of the financial effort of relationship equity and are key in a B2B setting as all businesses aim to maximise their bottom line through cost savings. Thus, the trade discounts offered by manufacturers will influence the retail partners' propensity to want to purchase directly from the manufacturer and not via a third party, which in turn will drive retention and therefore relationship/retention equity.

Chahal and Bala (2017) outlined customer retention strategies that can be employed by companies in order to drive retention equity. These include loyalty schemes, affinity schemes, special treatment benefits, customer feedback surveys, content marketing and courtesy schemes. Through a 5-point Likert Scale, Chahal and Bala (2017) created a scale that can measure customer retention equity based on the six strategies, with loyalty and affinity schemes grouped as one.

2.1.3 Value Equity

This refers to the customer's value perception of the firm (Rust et al., 2000) and is regarded as the "link between perceived quality", which is derived from the consumers' perceptions of the brand, as well as the "monetary sacrifice" (Dodds, 1996) spent for the purchase. In a B2B environment, factors influencing value equity include price, quality and the convenience of conducting business with a particular supplier (Rust et al., 2000). In their work, Palmatier and Sridhar (2017) refer to this as offering equity and posit that offering equity has the potential to drive "superior sales and profits by providing customers with innovative offerings, through products and processes, that will unlock more value for the customer (i.e. reduced price, better performance, better customer experience)". This, in turn, gives the business a competitive advantage through increased sales and recruitment of new customers, ultimately improving the business's customer equity (Palmatier & Sridhar, 2017).

Treacy and Wiersma (1995) posited value equity or the value hypothesis to be comprised of three disciplines, namely: operational excellence, product leadership and customer intimacy. Operational excellence is defined as a combination of "quality, price, and ease of purchase".

Product leadership is defined as innovation that continuously aims to “leapfrog” a company’s existing products. Customer intimacy is defined by the relationships an entity has with their customers and their constant pursuit to achieve service excellence. With respect to value equity, the definition is restricted to the operational excellence definition, as product leadership and customer intimacy would fall under brand equity and relationship equity respectively. With regard to operational excellence, companies can choose to either focus on low-cost leadership, convenience through service delivery, virtual integration through adoption of technology, or all three (Treacy & Wiersma, 1995).

2.2 CLV and Customer Equity: Internal School of Thought

The concept of customer equity within a marketing context was first introduced by Blattberg and Deighton (1996), who in their work defined customer equity as “the sum of the discounted expected customer contributions of current customers over their expected lifetime”. In their initial model, Blattberg and Deighton’s aim was to find the optimal balance between returns from acquisition costs and returns from retention costs that would allow a firm to optimise revenues from its customer base. The model did unlock the ability to mathematically model optimised marketing spend, based on acquisition or retention; however, it failed to “identify the components of marketing spend” and thus led to a revised model in 2001 (Bick, 2009). In his 2001 revised model, Blattberg took into account acquisition equity, retention equity and add-on selling equity. Despite improvements, the model still suffered from limitations, being internally focused and not taking into account customer inputs and competition effects (Bick, 2009). This school of thought was regarded as the “internal school of thought with a supply-side orientation” (Bick, 2009). Kumar discussed various ways in which CLV can be calculated, looking at both aggregate and disaggregate methods (Kumar, 2006).

2.2.1 Aggregate Approach

Aggregate approach models are models that use "segment or firm-level information to calculate the average or expected CLV and multiply that by the number of customers" (Kumar & George, 2007). Four common approaches include the Berger-Nasr approach, the Gupta-Lehmann approach, the Rust-Lemon-Zeithaml approach as well as the Blattberg-Getz-Thomas approach. The Berger-Nasr, Rust-Lemon-Zeithaml as well as the Gupta-Lehmann approaches are all calculated at the firm level whilst the Blattberg-Getz-Thomas approach is calculated at the segment level (Kumar & George, 2007). Each of these approaches is briefly explained below.

i. Berger-Nasr looks at CLV from a simplistic point of view with the assumptions that sales take place once a year, and that the yearly spend, gross margin and retention rate remain constant over time (Berger & Nasr, 1998). The drawback to this approach is that it is difficult to "segregate the components of customer equity" and that estimated acquisition and retention probabilities have "low predictive accuracy" due to the subjectivity of their nature (Kumar & George, 2007). The impact of this is the method's inability to improve specific customer-equity drivers and thus strategies cannot be set in place using this method to optimise customer equity (Kumar & George, 2007).

ii. Gupta-Lehmann, in addition to considering a constant average margin and retention rate, also considered the CLV over an infinite projection period. This additional aspect helped to improve the CLV projections from those of other methods such as Berger-Nasr, as the income streams were not overstated as a consequence of retention rate being converted into an expected lifetime (Gupta & Lehmann, 2003). This method uses publicly available information and is a useful method with applications in quantifying the value of firms when doing mergers and acquisitions (Gupta & Lehmann, 2003). Its drawback, as with that of the Berger-Nasr approach remains its inability to identify specific customer equity drivers which can be improved upon (Kumar & George, 2007).

iii. Blattberg-Getz-Thomas calculates CLV using return on acquisition, return on retention, as well as return on add-on selling. It has the added benefit of looking at the data from a segment level, yet remains an aggregate model as the CLV cannot be estimated at a customer level (Kumar & George, 2007). The main drawback of this method is that amendments to a company's strategy with regard to optimising customer equity may be difficult to implement as changes need to be actioned at a firm level and not at an individual customer level (Kumar & George, 2007).

iv. Rust-Lemon-Zeithaml created a CLV approach that incorporates customer-specific brand-switching matrices, by considering consumers' responses to the likelihood of them purchasing another brand in subsequent time periods (Kumar & George, 2007). This approach improves on the others by taking into account competitive effects through the switching element, yet falls short as it does not account for cross-shopping (Rust, Lemon, & Zeithaml, 2004), which is a high probability especially in CPG environments, with most manufacturers having a house of brands model when it comes to their product portfolio.

Despite improvements across the years, aggregate methods are useful for high-level CLV measurement, yet remain at a disadvantage as they cannot be used to provide information at a customer level and thus impede the ability to put in place customer-specific interventions in order to maximise equity, unlike disaggregate methods (Kumar, 2006).

2.2.2 Disaggregate Approach

The disaggregate approach, on the other hand, utilises customer information at an individual level to calculate the cash flows of each customer and, resultantly, the CLV of the customer based on that customer's behaviour (Kumar & George, 2007). The advantage of the disaggregate approach is that it "identifies customer-specific variables as drivers of CLV such as purchase frequency, contribution margin, marketing costs, etc." (Kumar & George, 2007). Another advantage of using a disaggregate approach is that customer equity can be maximised through the utilisation of customer-specific strategies and the ability to optimise resource allocations through customer segmentation and analysis of customer equity drivers for each segment (Kumar & George, 2007). Two common approaches include the Verhoef-Donkers approach as well as the more widely-known Venkatesan-Kumar (VK) approach. Both of these are briefly explained below.

i. Verhoef-Donkers created a way in which the CLV or "customer potential value" could be calculated by looking at a customer's purchasing behaviour as well as socio-demographic attributes (Verhoef & Donkers, 2001). The method used the insurance industry as a case study and noted what insurance products customers purchased as well as their socio demographics such as age, education and income. They found that the univariate probit method was a decent technique to use to predict insurance ownership, although it did not always predict this well (Verhoef & Donkers, 2001).

ii. The Venkatesan-Kumar (VK) approach uses customer-level data to calculate individual-level CLVs from which customer-level strategies can be implemented to improve customer equity (Kumar & George, 2007). The VK method will be expanded upon in the following section as well as in Chapter Three.

2.3 Preferred CLV Approach for this Study

This study will utilise the VK approach to calculate the CLV due to the availability of individual-level customer data. The VK method is grounded on the notion of "purchase frequency, contribution margin, and marketing costs" (Venkatesan & Kumar, 2004), which are among the driving factors when it comes to relationship equity through customers' engagement and contribution to a company's profitability and customer equity (Ramaseshan, Rabbanee, & Hui, 2013). The VK approach is mathematically defined as follows:

$$CLV_i = \sum_{y=1}^{T_i} \frac{CM_{i.y}}{(1+r)^{\frac{y}{frequency_i}}} - \sum_{l=1}^{n} \frac{\sum_m c_{i,m,l} \times x_{i,m,l}}{(1+r)^{l-1}}$$

The first part of the mathematical equation: $\sum_{y=1}^{T_i} \frac{CM_{i.y}}{(1+r)^{\frac{y}{frequency_i}}}$ refers to the customers' purchasing behaviour, with CM being the contribution margin of customer i in purchase occasion y, frequency being the purchase frequency and r being the discount rate.

2.3.1 Contribution Margin

Contribution margin is a metric based on previous studies focusing on purchase quantity and customer revenue (Venkatesan & Kumar, 2004). Customer purchasing behaviour can be decomposed into three components: "purchase frequency, spend per purchase occasion as well as purchase composition (single or cross category)" (Reinartz & Kumar, 2003). As a customer intensifies their purchase behaviour, so their relationship with a firm deepens, thus fostering customer loyalty and thereby driving customer lifetime value and customer equity. Reinartz and Kumar (2003) showed that consumers who have a high investment with respect to share of wallet tend to have a high customer lifetime duration. This leads to the following hypothesis:

Hypothesis One (H_1): A customer's basket size influences CLV

$$H_0: Customer's\ spend\ does\ not\ affect\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ who\ spend\ more\ with\ the\ firm\ are\ likely\ to\ have\ a\ higher\ CLV$$

Purchase composition, which refers to the extent to which the customer buys across several categories has been known to have an influence on the customer value. Reinartz and

Kumar (2003) found that cross-purchasing had a positive influence on customer lifetime, whilst on the other hand, their study concluded that focused or single category buying had a negative impact on customer lifetime value and, in turn, on customer equity. For the purposes of this study, category will refer to individual brands, given that the manufacturer in this study offers a portfolio of products.

Hypothesis Two (H_2): Basket composition influences CLV

$$H_0: Cross\ shopping\ does\ not\ affect\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ who\ cross\ shop\ tend\ to\ have\ a\ higher\ CLV$$

2.3.2 Frequency

Frequency forms part of the VK approach, based on the assumption that "customers are most likely to reduce their frequency of purchase before terminating a relationship" (Venkatesan & Kumar, 2004) and is better at estimating CLV than the "lost-for-good scenario which has a tendency to underestimate CLV" (Venkatesan & Kumar, 2004).

Purchase frequency is a fluid concept, as it depends on the consumers' objective for the purchase being made and the type of category that they are purchasing. High consumption categories, such as toilet paper, would have a high frequency rate, whilst more durable goods, such as automobiles, will have a low frequency rate as the time interval between one purchase and the next can range anywhere from three to ten years, depending on the consumer. Reinartz and Kumar (2003) propose the notion of time intervals between purchases as a measure of frequency and posit this relation to be U-shaped, thus leading to the following hypothesis:

Hypothesis Three (H_3): Purchase frequency influences CLV

$$H_0: Customer's\ purchase\ frequency\ does\ not\ affect\ a\ customers'\ CLV\ classification$$

$$H_A: Customers\ with\ a\ high\ purchase\ frequency\ are\ more\ likely\ to\ have\ a\ higher\ CLV$$

2.3.3 Marketing Costs

The second part of the equation $\sum_{l=1}^{n} \frac{\sum_m c_{i,m,l} \times x_{i,m,l}}{(1+r)^{l-1}}$ deals with the marketing costs to service a customer, with $c_{i,m,l}$ referring to the unit marketing cost per customer *i* within channel *m* in year *l*, and $x_{i,m,l}$ referring to the number of contacts/interactions with the customer *i* within channel *m* in year *l*. For the purposes of this study, neither the effect of marketing costs on CLV, nor their ability to segment customers into low, medium or high value will be tested.

2.4 Macro-economic Factors

Macro-economic factors, such as inflation, are overlaid as an additional layer which takes into consideration the environment in which the firm is operating and its effect on CLV.

The macro-economic environment influences household consumption, which in turn influences the retailer's customer value as it is dependent on how much the household is willing to spend within that retail outlet. The extent to which a household can consume products is determined by the level of disposable income available to them, which is defined as "the income left to an individual or household after the deduction of personal taxes" (Akekere & Yousuo, 2012). The relation between disposable income and consumption is succinctly stated by Keynes' Absolute Income Hypothesis, which states that an individual will increase their consumption as their income rises, but not to the same extent (Alimi, 2013).

However, Bonsu and Muzindutsi (2017) and Mallick and Mohsin (2016) found that inflation has both a long- and short-term negative effect on household consumption and thus would inversely affect a retailer's customer value for the supplier. What should be noted is that the influence of inflation is dependent on a household's disposable income, as "high-income households tend to spend more money for the same bundle of products than low-income households" (Reinartz & Kumar, 2003). Therefore the following hypothesis is considered for the inflation factor:

Hypothesis Four (H_4): Socio-economic elements influence CLV

$$H_0: Inflation\ affects\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ subjected\ to\ a\ higher\ inflation\ have\ a\ lower\ CLV$$

2.5 CLV as a Customer Segmentation Tool

CLV has been regarded as a useful customer segmentation tool as it enables a company to "improve the effectiveness of marketing expenditures" as well as resource allocation by identifying the specific levers to address per CLV classification. In their work, K. N. Lemon and Mark (2006) assessed the use of CLV as a segmentation tool against the criteria set out by Wedel and Kamakura (1998) and confirmed that CLV does have the potential of meeting this criterion with due considerations. Later studies such as the one by Tukel and Dixit (2013) confirmed that CLV can be a powerful segmentation tool that can be used by manufacturers to apply appropriate customer relationship management (CRM) strategies in order to optimise production and maximise profits. For the purpose of this study, the following can be confirmed against the criteria set out by Wedel and Kamakura (1998):

1. **Identifiability:** The CLV used as a basis for the CART process can identify homogenous groups of customers as it uses the purchasing behaviours of the customers as the deciding factor.
2. **Sustainability:** In a B2B environment and for this study, the segments would be large enough to justify the marketing resources, as this study relates to customers serviced directly by the business and the marketing efforts in question would be specifically linked to those that aim to drive retention/relationship equity and,

therefore, the classification would have the benefit of giving management guidelines on how to optimise resources by segment.

3. **Accessibility:** CLV allows the business to create targeted promotional strategies that drive volume uplift by enabling the business to customise the promotions given to the retail partners in order to drive either increased spend or frequency to improve customer equity. What is to be noted is that, given that the retailer is an intermediary, this needs to be supported by demand-pull activities for the end consumer to ensure a thorough pull through of sales.
4. **Stability:** Stability may not apply to the use of CLV as a segmentation tool. CLV in this study is calculated at an individual level. This implies that the segments would be dynamic over time, based on the marketing interventions that the business takes which may change the customer's purchasing behaviour. Thus, the CLV approach requires the segmentation to be updated at frequent intervals and, given the nature of the business, a suggested time frame may be every two to three years.
5. **Actionability:** The CLV segmentation approach results in "effective targeted marketing efforts" through the optimisation of the servicing strategy that the business employs, which ensures that customers' purchasing behaviours are changed through either driving a bigger basket, increased frequency, or cross-shopping.
6. **Responsiveness:** Given the B2B nature of this study, it would take a short period of time for managers to be able to see the change in customer purchasing behaviour, with the interventions put in place. In addition, due to the identifiability of the customers at an individual level, it would also make it easy

for management to know when to place extra resources on a customer or terminate the relationship to improve overall profitability.

2.6 Summary of Hypothesis and Theoretical Model

The table below summarises what this study assesses and the main thought leadership contributors.

Table 1: Summary of Hypotheses and Key Contributors

Hypothesis	Description	Main Author
H_1	Customer' s spend does not affect a customer's CLV classification	Reinartz & Kumar
H_2	Cross-shopping (Basket Composition) does not affect a customer's CLV classification	Reinartz & Kumar
H_3	Customer's purchase frequency does not affect a customer's CLV classification	Venkatesan & Kumar
H_4	Inflation (socio-economic elements) does not affect a customer's CLV classification	Reinartz & Kumar; Alimi, Bonsu and Muzindutsi

Below is the proposed theoretical model outlining the constructs to be tested.

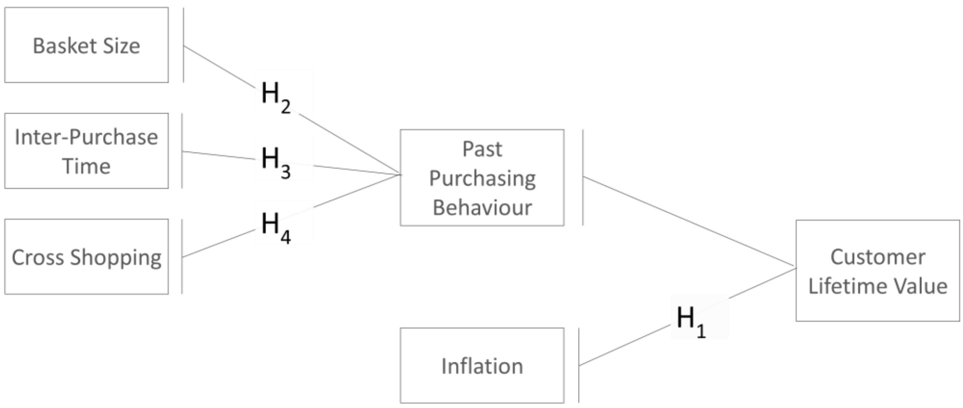

Figure 3: Theoretical Diagram for MMSM Dissertation

3 Research Methodology and Data Collection

3.1 Research Design

Research design is defined as "a logical plan for getting from here to there" and provides the researcher with a "blueprint" for their research (Yin, 2018). For this research, an empirical case study approach was used as it allowed the researcher to "investigate a contemporary phenomenon within the real world context" (Yin, 2018).

The research followed a quantitative, multiple-case study approach. This design considered each independent retailer in the sample as a single-unit case study. This design was ideal and logical for the current research as it removed the complexity associated with sample design given the nature of the independent trade in South Africa.

3.2 Population and Sample

3.2.1 Population

When choosing a representative sample on which to conduct research, the population of interest must first be defined. There are four different population types to take into consideration during the survey design. These are: the population of inference, the target population, the frame population and, lastly, the survey sample. The populations are defined as follows by several scholars (Bracht & Glass, 1968; Schonlau, Fricker, & Elliott, 2002):

Table 2: Types of Populations for Survey Design

	Theoretical Definition	Application in Current Study	Population Size
Population of Inference	The population on which conclusions or inferences are to be drawn	Independent retailers within South Africa	138,392
Target Population	The population of inference excluding non-target items or groups	Independent retailers within South Africa that are serviced directly by the CPG company and are neither bulk retailers nor redistributors, or which service an institution (i.e. school tuckshop/canteen/etc.)	28,830
Frame Population	The portion of the target population that will be enumerated	Independent retailers within South Africa that are serviced directly by the CPG company and are neither bulk retailers nor redistributors, or which service an institution (i.e. school tuckshop/canteen/etc.) and who have purchased from the FMCG company since 2015 for a period of at least 3 years	15,350
Survey Sample	Members of the frame population who were chosen to be surveyed	The full frame population for customers with a positive CLV will be used, with the population split into 2 samples: the 1st for the main study, and a 2nd for validation	15,339

3.2.2 Sample

Given that a non-parametric research method and in particular classification trees will be used. The full frame population was utilised as the survey sample and no further considerations need to be taken into account, such as sample stratification. Using a 60/40 split, 9,073 of the observations were used for the training data set and 6,136 for the validation data set.

3.3 Data Sources

Data was collected from secondary sources for both customer and macroeconomic data.

3.3.1 Customer Data

The customer data used for the CLV calculations was sourced from internal sales data recorded PepsiCo (from here on referred to as the company). The internal sales are a longitudinal data set with transaction-level data for customers over a five-year period.

The customer data is stored in a data warehouse and is collected via a handheld terminal (HHC). This device is connected to the distribution management system (DMS) which transmits data such as stock movement, invoicing, and returns to and from the HHC. The sales representative requests stock from the distribution centre (DC) to service the customer via the HHC. This instruction is then sent to the DMS system, wherein the warehouse manager and the stock controller will have visibility of the request and pick the stock for collection. Once the stock is picked, it then appears as an open transaction on the sales representative's HHC. The sales representative then chooses to either accept or reject the transaction. An accepted transaction officially becomes an order which the sales representative must go and collect from the DC in order to service the customer. This data, which is stored in a central warehouse can also be accessed through the use of the business intelligence tool called QlikSense, a self-service analytics tool.

Two files were extracted from the QlikSense platform and combined to create the final input file for the CLV calculations. The initial data file extraction contained daily sales data by customer and included the following fields:

i. ***Brand:*** the brand of the snacking product that the customer purchased.
ii. ***Customer Name:*** the name of the customer as recorded on the company's database.
iii. ***Customer ID:*** the unique identifier assigned to the customer by the CPG company.
iv. ***DC:*** the distribution centre from which the customer is serviced.

v. ***DTS Region:*** the company's operationally-defined region wherein the distribution centre is located.
vi. ***Route ID:*** the unique identifier of the route which the delivery truck travels from the distribution centre to the customer in order to deliver the order placed.
vii. ***SAP Customer ID:*** the unique identifier for the customer on the SAP system used by the company.
viii. ***Year:*** the year in which the transaction/sale was made.
ix. ***Transaction Date:*** the date on which the transaction/sale happened.
x. ***Invoice Sales Volume:*** the quantity in volumes, measured in kilograms, of the product purchased as per the invoice billed.
xi. ***Stales Volume:*** the quantity in volumes, measured in kilograms of the product that was returned back to the distribution centre by the customer. These products are defined as stales as they are either past their best before date or are close to reaching their best before date.
xii. ***Buyback Volume:*** the quantity in volumes, measured in kilograms, of the product that was returned to the distribution centre by the customer. These products are still viable to be resold by the business to other customers as they are not past their best before date.

In the second file, the data that was extracted was for annualised sales and included the following fields:

i. ***Customer Name:*** the name of the customer as recorded on the company's database.
ii. ***Customer ID:*** the unique identifier assigned to the customer by the CPG company.
iii. ***Year:*** the year in which the transaction/sale was made.
iv. ***SDV:*** Store door value, which is the calculated price per kilogram which each customer paid for that year based on what was invoiced to them throughout the year.

These fields from both extracts were then manipulated to create the variables that are used in the CLV calculation as well as the classification tree. These variables and how they were calculated are explained below:

i. ***IPR:*** the inter-purchase rate was calculated by taking the number of days between each transaction per customer and averaging those over an annual period for each respective year.

ii. ***ABS:*** the average basket size was calculated by summing the total volume for each transaction as follows:

$$Total\ Volume = Invoice\ Sales\ Volume + Stales\ Volume + BuyBack\ Volume$$

The reason for this is because the stales and buyback volumes are recorded as negative figures as they are returns. Once calculated, the total volume per transaction is then averaged by customer for each respective year to obtain the ABS figure.

iii. ***CS:*** Cross-shopping was derived by counting the number of different brands the customer purchased per transaction. This was then averaged for each respective year by customer.

A practical example of these calculations is provided in Appendix I.

3.3.2 Macroeconomic Data

The macroeconomic data on inflation was sourced from national statistics databases housed by Statistics South Africa. This data is available in monthly time intervals, however, for the purposes of this study, the data was rolled up to annualised values, based on the annual average.

3.4 Research Methods

3.4.1 Classification and Regression Trees

Classification and regression trees, better known as CART, is a statistical method founded by Breiman et al. (1984). The technique identifies patterns in datasets and aims to classify them into a series of homogenous groups by using recursive partitioning (Galletta, 2016; Shmueli, Bruce, Yahav, Patel, & Lichtendahl, 2017). To ensure that the classification tree is not over-fit, some of the decision nodes, which "hardly reduce the error rate" (Shmueli et al., 2017) are redesigned as terminal nodes, thus retaining the pattern in the data set (Shmueli et al., 2017). The difference between a classification and a regression tree is that the terminal nodes of a classification tree indicate a binary outcome (High or Low; Yes or No; 0 or 1), whilst those of a regression tree indicate a continuous outcome. Given the nature of the data set, the method will be a classification tree, the outcome of which could then be used to enable the business to classify customers' CLV potential, based on purchasing habits. Decision trees have been used in CLV literature to some extent; for example, in their study, Tirenni et al. (2007) used decision trees to find a manner to classify customers of a European airline into a high-value segment based on spend along with other variables, such as nationality, seat preference, family size and frequency of purchase. The advantages of using decision trees is that, "No distributions need to be pre-specified; no transformations or pre-selection of variables is required; single rules or terminal nodes can provide valuable insight" (Tirenni et al., 2007). In addition, they "create law" (Kashani & Shahmirzaloo, 2017) by giving the user a set of rules from which the classifications were obtained. This allows the user to classify new observations or, in this case, new customers, based on these rules without having to rerun the analysis.

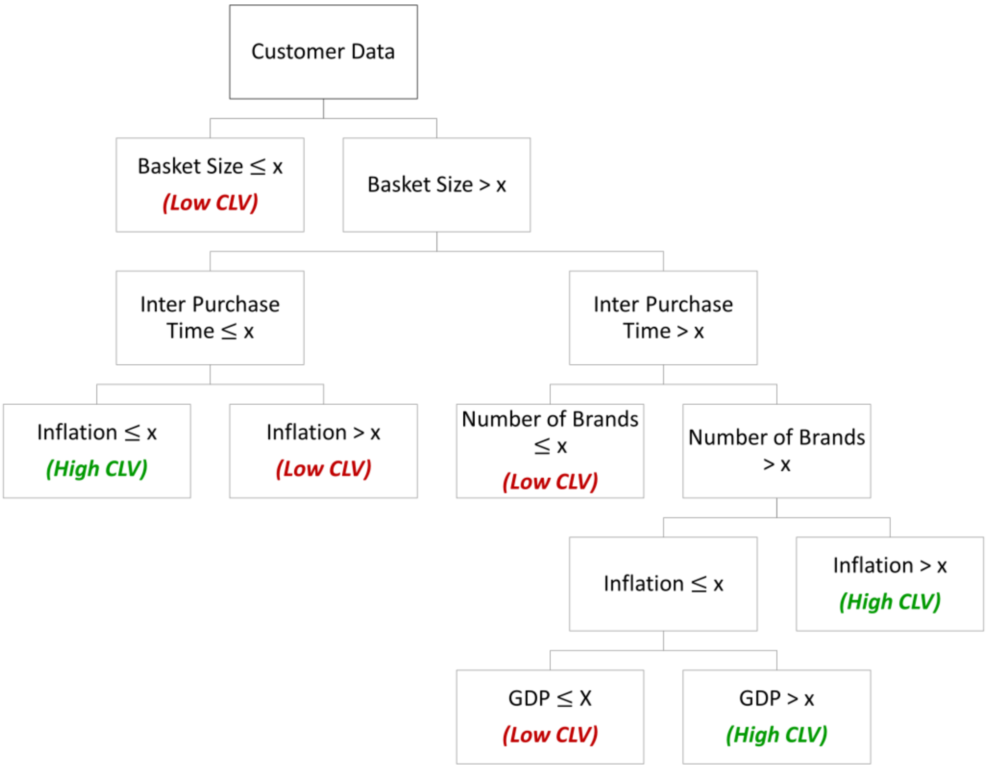

Figure 4: Proposed Classification Tree Model for MMSM Dissertation

One concern with classification trees is that they may have a tendency to over-fit the data. In order to overcome this issue, one can set limiting factors such as "determining the number of splits; determining the number of records in a terminal node; or stipulating the minimum reduction in impurity" (Shmueli et al., 2017). Alternatively, through the use of the validation sample, the tree created from the training dataset can be pruned by reducing the number of decision nodes through the removal of the weakest branches (Shmueli et al., 2017). An added feature of using CART is its ability to select the "right-sized" tree through an in-built pruning algorithm which prunes the large tree by selecting the sequence which optimises the error rate (Thrasher, 1991).

3.4.1.1 Limitations

Classification trees tend to be sensitive to changes in data, given that they are designed off a particular data set, and thus may return different splits based on new information in the data set. In addition, classification trees do not necessarily consider the relationship between predictors; therefore, unless the researcher specifically creates a variable to account for this relation, the tree may be expected "to perform lower than methods such as discriminant analysis" (Shmueli et al., 2017).

3.4.2 Estimating Customer Lifetime Value

In this study, the disaggregate VK approach was used to determine the CLV of the retailers (Kumar & George, 2007). This approach uses longitudinal customer transaction data, which was available in this study and was calculated as follows:

$$CLV_i = \sum_{y=1}^{T_i} \frac{CM_{i.y}}{(1+r)^{\frac{y}{frequency_i}}} - \sum_{l=1}^{n} \frac{\sum_m c_{i.m.l} \times x_{i,m,l}}{(1+r)^{l-1}}$$

i. $x_{i,m,l}$ is the number of contacts of customer i in channel m in year l. Given that this study looked at a single channel:
   - m was not a consideration as it is equal to 1 and therefore did not form part of the summation;
   - In terms of number of contacts, customers are called on once a week; therefore, the number of contacts per customer per annum would be equal to the number of trading weeks per annum. Accordingly, 52 weeks was used as the number of contacts, as the company operates all year round.

ii. $frequency_i$ was calculated on a customer by customer basis using the average inter-purchase times (measured in days) per year as the customer's purchase frequency. The average IPR over the five-year period was used as the frequency.

iii. n is the number of years to forecast – a five-year horizon was used.

iv. $CM_{i,y}$ is the predicted contribution margin of customer i in purchase occasion y.

- For purposes of this study, the sales – which are measured in physical quantity – were regarded as the contribution margin;
- To get the predicted $CM_{i,y}$, the finance calculation of future value: $FV = PV(1+i)^n$ was adapted as follows:
  - PV is the arithmetic mean over the historical years 2015 to 2018. Use of the arithmetic mean is favoured over the geometric mean, given that some customers will have negative annualised values due to returns exceeding sales within that year. The PV value is the nominal value as it was not adjusted for inflation.
  - i is determined by the compounded annual growth rate (CAGR) over the past five years for the customer and applied as the future expected interest rate in order to calculate the FV. The CAGR is a useful measure and replacement for i in the sense that it assists in providing a view of the historical growth pattern based on a smoothed rate of return over a set number of years. The method used to calculate the CAGR was the logarithm method $i = e^{\frac{ln\left(\frac{FV}{PV}\right)}{n}} - 1$ as shared by Robert (1969).

v. $c_{i,m,l}$ is the unit marketing cost for customer i in channel m in year l. For the purpose of this study, only a single channel was considered:

- m is not a consideration as it is equal to 1 and therefore did not form part of the summation;
- The marketing cost is based on annual budgets set for the channel and proportioned, based on the customer's percentage contribution;
- Given that $CM_{i,y}$ is measured in physical quantities, the $c_{i,m,l}$ is converted to a physical unit by using the average price per physical unit for customer i in year l.

vi. r is the discount rate; for purposes of this study, the weighted average cost of capital currently set at 9% was utilised.

vii. T_i is the predicted number of purchases and, given that the $CM_{i,y}$ is based on annualised values, T_i represented years one through five.

Thus, the amended equation for the VK approach based on the considerations above is as follows:

$$CLV_i = \sum_{y=1}^{T_i} \frac{CM_{i.y}}{(1+r)^{\frac{y}{frequency_i}}} - \sum_{l=1}^{n} \frac{\left(c_{i.l} \times x_{i,l}\right) / \left(Price / KG \right)}{(1+r)^{l-1}}$$

3.4.2.1 Limitations

The drawback of using the VK method is that the effect of purchases from competitor brands and/or alternative channels is not accounted for due to the non-availability of this data at a transactional level. In addition, using the future value calculation in point iv above smoothens the data with regard to customer purchasing trends over time.

3.4.3 Macro-Economic Factors

Economic factors were collected from Statistics South Africa’s database. The inflation figures were sourced from the CPI dataset (Statistics South Africa, 2018). The unit of measure was the percentage change versus last year for inflation. The data was assigned to the customer, based on their geographical location.

Given that the CLV used the historical five-year interval from 2015 to 2019 to calculate future values, an arithmetic mean for the inflation figures in order to calculate a single inflation figure applicable to the five-year timeline was employed. The arithmetic mean is in line with the annual average reported by Statistics South Africa (Statistics South Africa, 2019).

3.5 Validity and Reliability

In any empirical research undertaking, it is important to test the quality of the research by conducting validity and reliability tests. In terms of the case study approach, the reliability as well as three validity tests are applicable, namely: construct, internal and external validity (Yin, 2018).

3.5.1 Construct Validity

This refers to the ability of a "measure to assess the construct it is purported to assess through the assessment of the magnitude and direction of the representative sample of the characteristics of the construct as well as the degree to which the measure is not subjected to an error term" (Peter, 1981). In the review by Kane (2001) on construct validity, he refers to the strong programme on construct validity by Cronbach and Meehl (1955), which posits that in order to meet the criteria of construct validity, one has to first "lay out theoretical assumptions and conclusions and then subject these to empirical challenges" (Kane, 2001).

3.5.2 Internal Validity

Internal validity is concerned with "trying to explain how and why event x led to y" (Yin, 2018). The key consideration with internal validity is in understanding the cause and effect sequence of events in order to prevent ambiguous temporal precedence (Clark & Middleton, 2010). This is when a researcher is not clear on which variable goes first, thus resulting in research design failure, which negatively impacts the outcome of the research. Clark and Middleton (2010) posited other factors that can compromise the research design's internal validity, such as sampling bias in treatment effects, attrition/case study drop-outs, and changes in the methods used to measure responses. With respect to the current study, internal validity was ensured by assessing the complexity parameter (CP) of the classification output, and then using it to prune the tree so that the output did not result in an over-fitted tree with nested rules of similar variables.

3.5.3 External Validity

External validity deals with whether or not "a study's findings are generalisable across different settings (Calder, Phillips, & Tybout, 1982; Yin, 2018). When considering external validity, there are two major types of threats that one has to try and mitigate, namely, population validity and ecological validity (Bracht & Glass, 1968). Population validity threat deals with the ability to generalise findings in a study given the sample population characteristics, whilst ecological validity threat deals with the ability to generalise findings in a study given the environment in which the experiments were conducted (Bracht & Glass, 1968).

In this study, population external validity was controlled through having a training data set to initially run the classification tree, as well as a validation data set to pressure-test the results from the classification tree obtained using the training dataset. In terms of ecological external validity, it was much more difficult to mitigate this threat from a macro-economic environmental influence, given the current state of the South African market.

3.5.4 Reliability

This deals with the ability to replicate the same findings in a study if the experiment were to be conducted again (Yin, 2018).

3.6 Validity and Reliability for Predictive Models

To ensure validity with classification trees, the research used a method called cross-validation in which the classification tree is run several times using samples of the data set, each outcome of which is then used to create a set of prediction rules by pooling the results from the various trees. An alternative and simpler method is to use a training data set, which is normally 60% of the entire data set, as well as a validation data set, the remaining 40%, to create a set of trees (Shmueli et al., 2017). The training data is used to create the initial set of rules which are then validated by the validation data set. The set of rules should be similar in most instances with a few changes in leaf order. Once observed, the researcher can then finalise the set of rules for the data set, based on both trees.

4 Data Analysis

4.1 Resources

Resources to carry out the data analysis included Microsoft Excel and a statistical analysis software called R Studio.

4.2 Inflation

In order to ensure that the inflation with the strongest correlation to the volume movements over time was used in the study, three variations of inflation were considered, namely: Headline CPI, CPI for Food and Non-alcoholic Beverages, as well as CPI for Food. Having considered all three, it was found that Headline CPI (CPI) had a marginally stronger correlation to volume movements in the data set at 12.18% in comparison to CPI for Food and Non-Alcoholic Beverages (CPIFNAB) at 10.46% and CPI for Food (CPIF) at 9.79%. Based on these findings, Headline CPI was retained in the study. Given the high correlation between Headline CPI and CPI for Food and Non-Alcoholic Beverages, it stood to reason that Headline CPI was the better measure to use, as the operating margins in B2B environments are affected by a plethora of factors such as petrol, electricity and wages, which in turn impacts the extent to which a retail outlet can adjust its margins in order to maintain demand, thus affecting CLV.

Table 3: Table of Correlations for Macro Variables

	Volumes	**CPI**	**CPIFNAB**	**CPIF**
Volumes	1.00000000	0.12187501	0.10461945	0.09794341
CPI		1.00000000	0.90749450	0.89595910
CPIFNAB			1.00000000	0.99852156
CPIF				1.00000000

Considering that customers are located in different geographical regions; the five-year arithmetic mean of the provincial level CPI were used for the study. These are noted below:

Western Cape	Eastern Cape	Northern Cape	Free State	KwaZulu-Natal	North-West	Gauteng	Mpumalanga	Limpopo
5.5	5.1	4.4	5.1	4.8	4.5	5.0	4.6	5.0

4.3 CLV

For CLVs, only the observations with positive CLVs were retained for the study. This implied that in total 15,339 observations were retained for the study. In order to be able to run a CART, the CLVs first need to be classified into low and high. To do this, a box plot of the 15,339 retail outlets was generated and, based on the quartiles, the split between low and high CLV was determined. The median was used to determine the split between low and high. This is because the median or the 50th percentile is "a measure of central location and gives the observation that falls into the middle of a data set" (G. Keller, 2005). The initial box plot, Figure 4 below, revealed extreme outliers, as well as little differentiation between low and high. Extreme outliers were then removed and the CLV classifications reset from two to three. Having assessed the quartiles, using deciles to split the CLV into three classifications proved more apt in creating more homogenous groups. Those classified as low had values up to the 50th percentile; those classified as medium were between the 50th percentile and the 90th percentile, and those classified as high were above the 90th percentile. This led to the box plot distribution as shown in Figure 6 below.

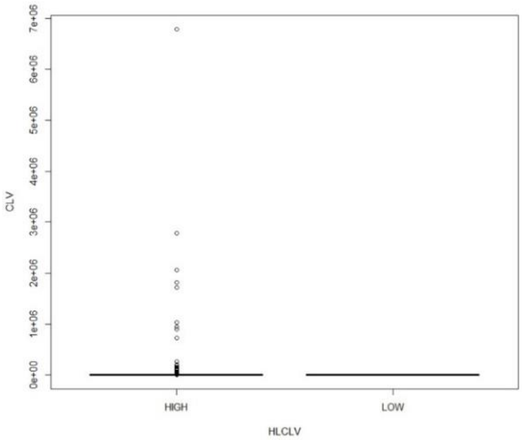

Figure 5: Preliminary Box Plot

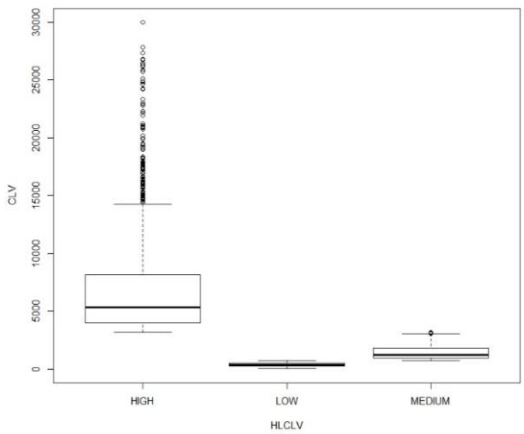

Figure 6: Box Plot with 3 CLV Classifications

Given that classification trees are a non-parametric statistical method and require no need for outliers to be removed, in this next section, the outliers were removed to obtain a clearer understanding as to whether there were any clear differences in traits amongst the variables used to classify the data set. Once outliers were removed, the observations were reclassified based on the new decile figures (as shown in Figure 7 below). The following boxplots were obtained for the variables included in the model: inter-purchase rate (IPR), average basket size (ABS), cross-shopping (CS) and inflation.

```
CLV5$CLV
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
   15031        0    15031        1     1228     1304    130.8    195.8    372.5    717.7   1420.9   2828.6   4379.5

lowest :    12.40351    12.72181    13.78918    21.77601    23.16538, highest: 10017.15033 10023.97149 10026.93929 10027.87302 10041.60162
```

Figure 7: Percentile Values for CLV Data Set

The removal of the outliers was merely to give a clearer view of the boxplot distributions. For the actual study, all 15,339 observations were retained.

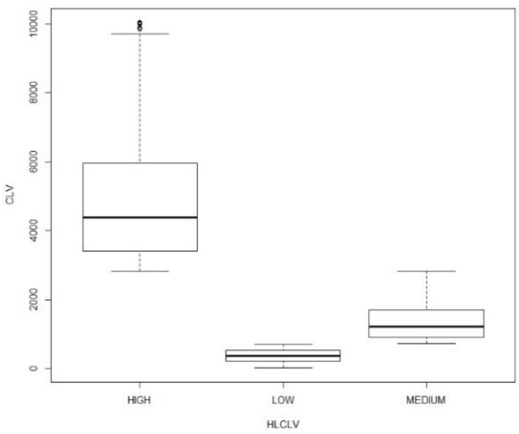

Figure 8: CLV Box Plot by CLV Classification

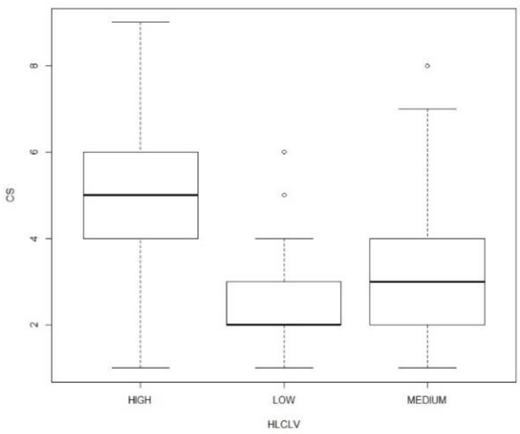

Figure 9: CS Box Plot by CLV Classification

The CLV box plot (Figure 8) reveals a clear distinction between high, medium and low CLV customers. Similarly, for cross-shopping (Figure 9), despite all customer types ranging from 1, the high CLV customers had a higher median, followed by the medium CLV customers and, lastly, by the low customers. This shows that high CLV customers had a tendency to purchase more brands in the portfolio than medium and low CLV customers.

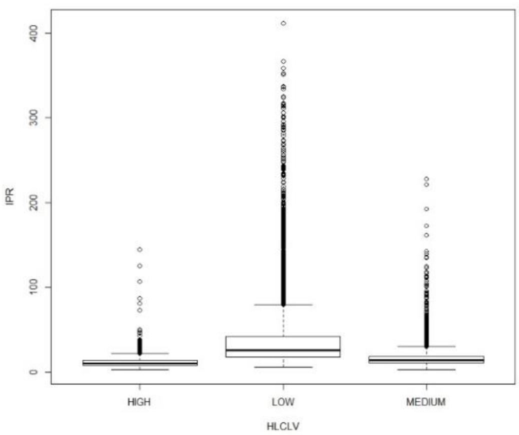

Figure 10: IPR by CLV Classification

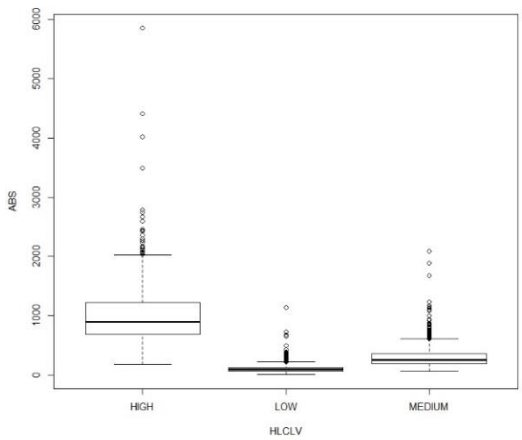

Figure 11: ABS by CLV Classification

Looking at the IPR in Figure 10, high and medium CLV customers tend to have a lower IPR indicating that they tend to purchase more frequently, whilst low CLV customers have a higher IPR indicating that they purchase less frequently. Similarly, with the ABS in Figure 11, high CLV customers on average purchased more per occasion than medium and low CLV customers, with low CLV customers having the smallest basket on average.

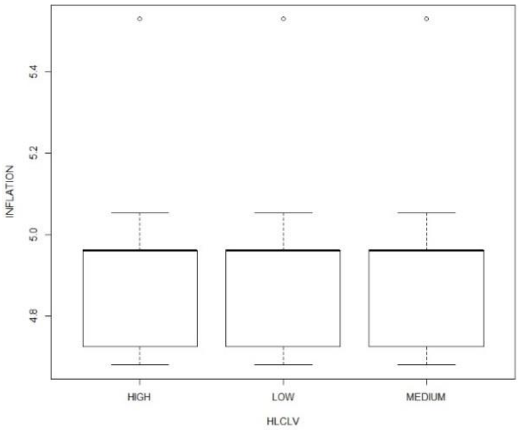

Figure 12: Inflation by CLV Classification

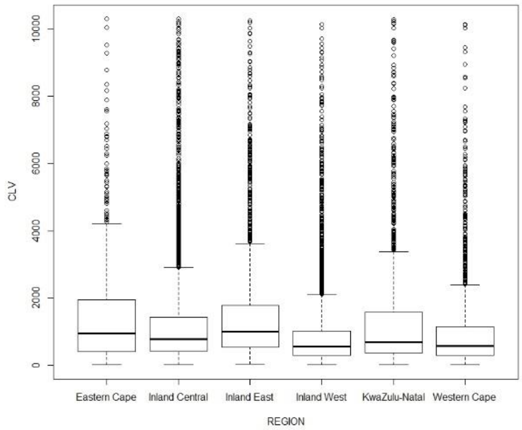

Figure 13: CLV by Regional Classification

Finally, looking at the inflation level (Figure 12) as well as provincial split (Figure 13), there are no clear distinctions amongst the groups. This gives an indication that these two variables may not necessarily be considered as distinguishing factors when generating the classification tree.

4.4 Classification Tree Analysis

To start the analysis, the original data set including outliers was used, but split into the training and validation sets based on the 60/40 principle. The high, medium and low split for the full data set was as indicated in Figure 14, with values greater than 3,226 denoted as high, those lower than 736 as low, and the remainder as medium.

```
> describe(SPLITS$CLV)
SPLITS$CLV
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
   15339        0    15339        1     2856     4496    132.9    197.4    378.4    735.6   1504.0   3225.9   5557.0

lowest : 1.240351e+01 1.272181e+01 1.378918e+01 2.177601e+01 2.316538e+01
highest: 1.715335e+06 1.813888e+06 2.059162e+06 2.787668e+06 6.783386e+06

Value            0   50000  100000  150000  200000  250000  700000  900000  950000 1050000 1700000 1800000 2050000
Frequency    15282      36       7       3       1       1       1       1       1       1       1       1       1
Proportion   0.996   0.002   0.000   0.000   0.000   0.000   0.000   0.000   0.000   0.000   0.000   0.000   0.000

Value      2800000 6800000
Frequency        1       1
Proportion   0.000   0.000

For the frequency table, variable is rounded to the nearest 50000
```

Figure 14: Descriptive Stats of Full Data Set to Determine CLV Intervals

The classification tree from the training data set was first created with all the variables set out in the study, namely: ABS, IPR, inflation, CS and regional splits. In the first run of the classification tree, only the ABS was a significant determinant in classifying customers with a complexity parameter (CP) cut off of 0.01 and a relative error of 0.1933, indicating that the tree's predictive accuracy is about 81% as the $rel\ error = 1 - R^2$. Despite the good accuracy, this tree did not take into account shopping behaviours of the customers, which is the main purpose of this study.

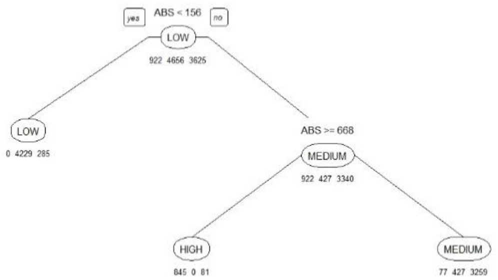

Figure 15: Initial Training Classification Tree

```
> printcp(default2.ct)

Classification tree:
rpart(formula = HLCLV ~ ., data = train2.df, method = "class")

Variables actually used in tree construction:
[1] ABS

Root node error: 4547/9203 = 0.49408

n= 9203

        CP nsplit rel error  xerror      xstd
1 0.64064      0   1.00000 1.00000 0.0105482
2 0.16802      1   0.35936 0.36332 0.0080969
3 0.01000      2   0.19133 0.19573 0.0062357
```

Figure 16: Initial Training Tree CP Output

A second run of the classification tree using the training data set was executed with a relaxed complexity parameter of 0.001. This resulted in a deeper classification tree as indicated in Figure 17. Inspection of the tree reveals that the first set of rules still holds with ABS being the primary driver, however, for customers sitting in the mid-range $156 < ABS < 196$ the tree then starts to take the other purchasing behaviours such as IPR and CS into consideration.

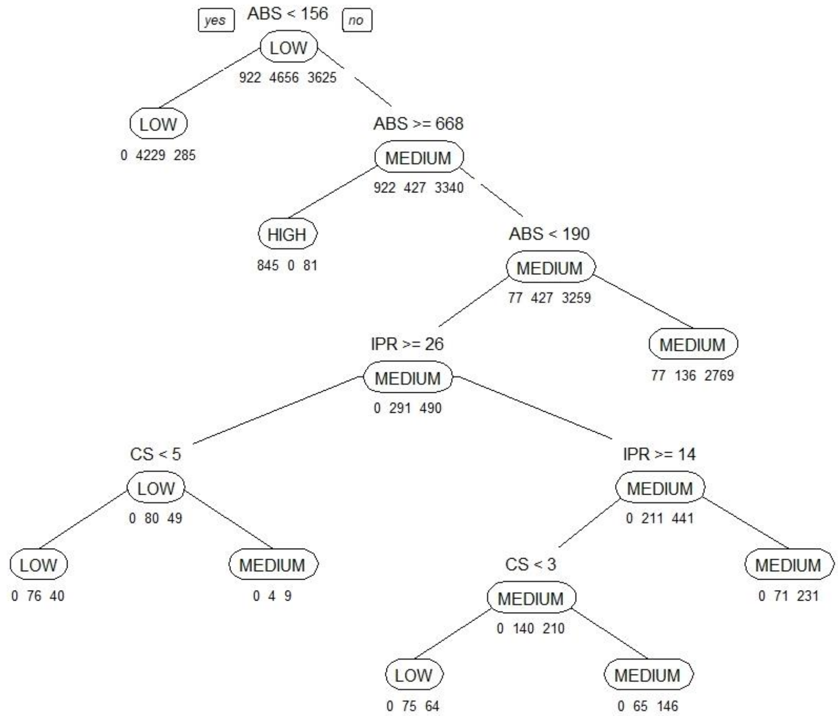

Figure 17: Training Set Classification Tree with Relaxed CP to 0.001

Examining the deeper tree in Figure 17, it is observed that the tree has a mean square error of 18.1% implying an R^2 of 81.9%. However, when examining the lower branches of the tree, the number of observations in each cell becomes too small.

```
> printcp(default4.ct)

Classification tree:
rpart(formula = HLCLV ~ ABS + IPR + CS + INFLATION, data = train2.df,
    method = "class", cp = 0.001)

Variables actually used in tree construction:
[1] ABS CS  IPR

Root node error: 4547/9203 = 0.49408

n= 9203

         CP nsplit rel error  xerror      xstd
1 0.6406422      0   1.00000 1.00000 0.0105482
2 0.1680229      1   0.35936 0.36244 0.0080892
3 0.0034088      2   0.19133 0.19485 0.0062232
4 0.0012096      4   0.18452 0.19068 0.0061631
5 0.0010996      6   0.18210 0.19046 0.0061599
6 0.0010000      7   0.18100 0.18980 0.0061503
```

Figure 18: Pruning Table for Deeper Classification Tree from Training Data Set

This is further substantiated by the pruning table in Figure 18, wherein split 6 and 7 only improve the relative error marginally compared to split 4. Thus, this gives grounds to prune the tree back using a CP of 0.012096. Once pruned, the result is a smaller classification tree as shown in Figure 19. Based on the pruned training data set tree shown in Figure 19, the following classification rules were defined and used to assess the tree's predictive ability:

i. If $ABS < 156$ then customer is classified into the LOW CLV group.
ii. If $ABS \geq 668$ then customer is classified into HIGH CLV group.
iii. If $156 \leq ABS < 668$ AND $IPR \geq 26$ then customer is classified into the LOW CLV group.
iv. If $156 \leq ABS < 668$ AND $IPR < 26$ then customer is classified into the MEDIUM CLV group.
v. If $190 \leq ABS < 668$ then customer is classified into the MEDIUM CLV group.

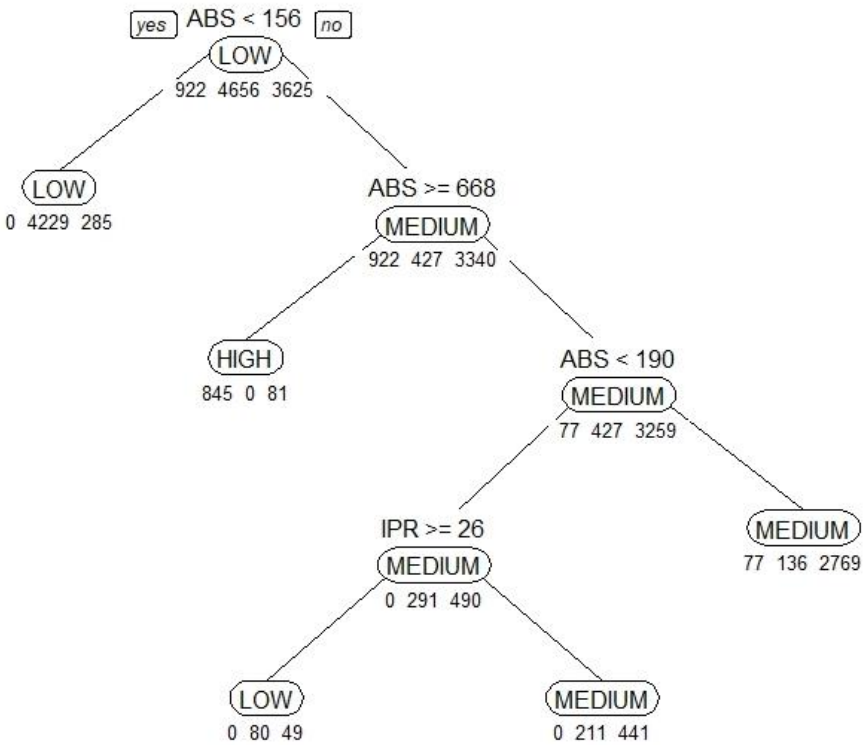

Figure 19: Final Pruned Training Data Set Tree with CP of 0.012096

Based on the above rules, a manually constructed classification matrix reveals that the classification tree has a high probability of classifying the customers into the correct *a priori* groups.

Table 4: CART Prediction Accuracy from Training Data Set

		A Priori CLV Classifications		
		LOW	**MEDIUM**	**HIGH**
CLV Classification according to Training Tree	**LOW**	89%	11%	-
	MEDIUM	9%	89%	2%
	HIGH	-	9%	91%

This, along with the low mean square error confirms that the classification tree has given a good set of rules which can be used to assess a customer's CLV potential, based on purchasing behaviour.

In order to validate the rules from the training data, a similar exercise was executed on the validation data set. An initial tree was constructed with a CP value of 0.001, which resulted in the classification tree indicated in Figure 20 below. From the initial inspection of the tree, it is observed that the tree has 11 nodes, some of which have too few cells. Thus, the pruning table must be examined in order to ascertain how far back to prune the tree.

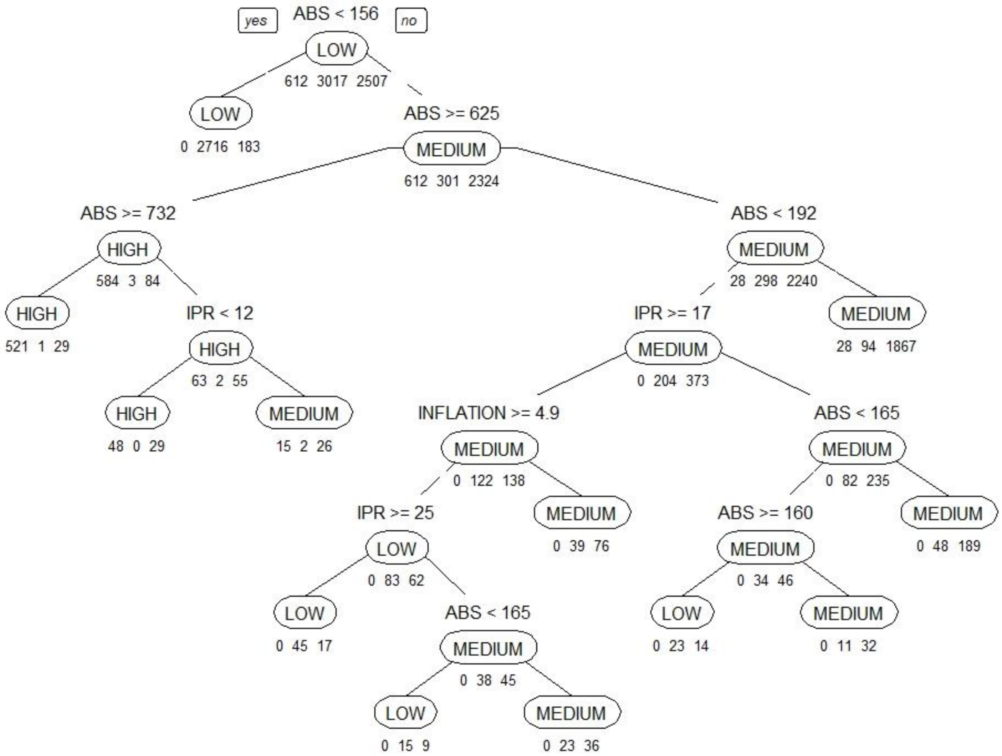

Figure 20: Classification Tree from Validation Data Set

Upon initial inspection, a major difference between the training tree and the validation tree can be observed, and that is that the validation tree includes inflation as one of its decision nodes. Pruning the tree back to a CP of 0019237, based on the pruning table in Figure 21, results in a much smaller tree that can still explain 81.8% of the variation in the data set.

```
> printcp(validate.ct)

Classification tree:
rpart(formula = HLCLV ~ ABS + IPR + CS + INFLATION, data = valid.df,
    method = "class", cp = 0.001)

Variables actually used in tree construction:
[1] ABS       INFLATION IPR

Root node error: 3119/6136 = 0.50831

n= 6136

         CP nsplit rel error  xerror      xstd
1 0.6486053      0   1.00000 1.00000 0.0125556
2 0.1603078      1   0.35139 0.35460 0.0096539
3 0.0022443      2   0.19109 0.20071 0.0076016
4 0.0019237      6   0.18211 0.19910 0.0075746
5 0.0017634      7   0.18019 0.20006 0.0075908
6 0.0014428      9   0.17666 0.19622 0.0075257
7 0.0010000     11   0.17377 0.19333 0.0074762
```

Figure 21: Pruning Table for Deeper Classification Tree from Validation Data Set

The smaller validation tree set still retains inflation as one of its decision nodes but has cut the tree back to 6 splits instead of 11, as shown by Figure 22.

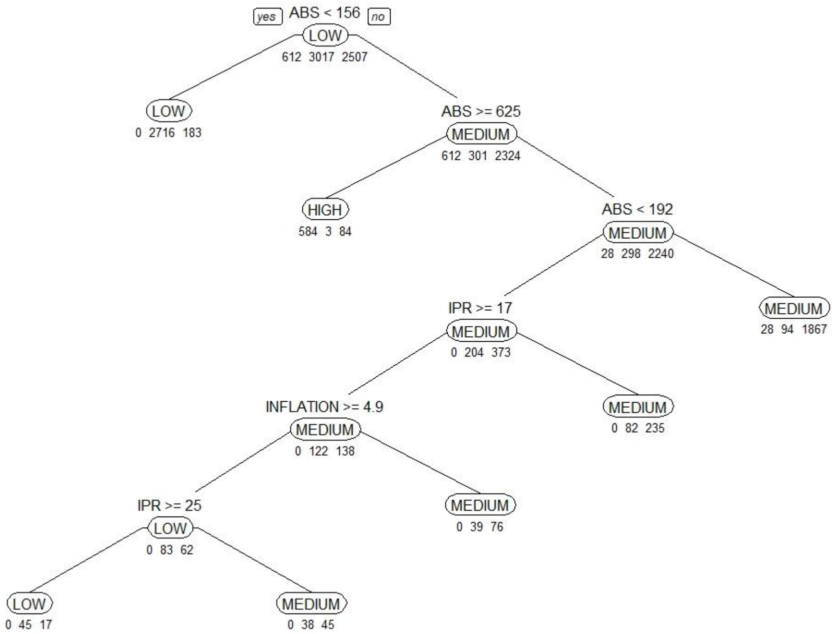

Figure 22: Initial Pruned Validation Data Set Tree with CP of 0.0019237

From the above tree, the classification rules were then defined in order to assess the prediction accuracy of the validation tree. The rules were defined as:

i. If $ABS < 156$ then customer is classified into the LOW CLV group.

ii. If $ABS \geq 625$ then customer is classified into HIGH CLV group.

iii. If $ABS < 192$ AND $IPR < 17$ then customer is classified into the MEDIUM CLV group.

iv. If $ABS < 192$ AND $IPR \geq 17$ AND $INFLATION < 4.9$ then customer is classified into the MEDIUM CLV group.

v. If $192 < ABS < 625$ then customer is classified into the MEDIUM CLV group.

vi. If $ABS < 192$ AND $IPR \geq 17$ AND $INFLATION \geq 4.9$ then customer is classified into the LOW CLV group.

vii. If $ABS < 192$ AND $IPR \geq 25$ AND $INFLATION \geq 4.9$ then customer is classified into the LOW CLV group.

viii. If $ABS < 192$ AND $IPR < 25$ AND $INFLATION \geq 4.9$ then customer is classified into the MEDIUM CLV group.

Given the small size of the cells generated from rule vii. and viii., these rules were not taken into account when classifying the observations; thus, only rules i. through vi. were used to generate the prediction matrix below.

Table 5: CART Prediction Accuracy from Validation Data Set

		A Priori CLV Classifications		
		LOW	**MEDIUM**	**HIGH**
CLV Classification according to Training Tree	**LOW**	92%	8%	-
	MEDIUM	9%	90%	1%
	HIGH	-	13%	87%

The validation tree's prediction accuracy is on par with that of the training data set. The one question that remains is to confirm whether removing inflation from the classification process will impact the prediction accuracy of the tree or not. To do so, another tree was created from the validation set with the inflation variable removed.

```
> printcp(validate.ct)

Classification tree:
rpart(formula = HLCLV ~ ABS + IPR + CS, data = valid.df, method = "class",
    cp = 0.001)

Variables actually used in tree construction:
[1] ABS IPR

Root node error: 3119/6136 = 0.50831

n= 6136

         CP nsplit rel error  xerror      xstd
1 0.6486053      0   1.00000 1.00000 0.0125556
2 0.1603078      1   0.35139 0.35749 0.0096844
3 0.0018168      2   0.19109 0.20135 0.0076123
4 0.0017634      5   0.18564 0.20071 0.0076016
5 0.0016031      7   0.18211 0.20006 0.0075908
6 0.0014428      9   0.17890 0.20071 0.0076016
7 0.0010000     11   0.17602 0.20006 0.0075908
```

Figure 23: Pruning Table for Validation Classification Tree without Inflation

Based on the pruning table output, this tree was then pruned back to a CP value of 0.0017634 with a cross-validation error of 20.071%. The pruned tree had five node splits and is depicted in Figure 24 below.

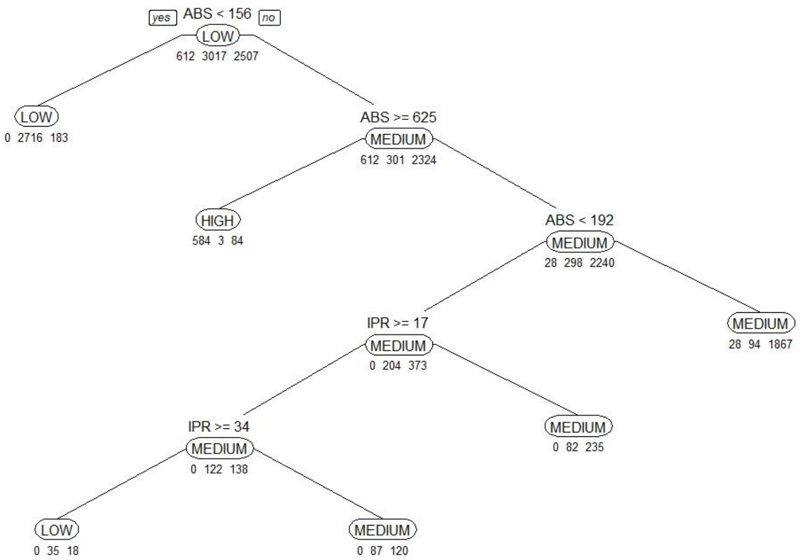

Figure 24: Pruned Validation Classification Tree with No Inflation Variable

The above tree led to the following classification rules:

i. If $ABS < 156$ then customer is classified into the LOW CLV group.
ii. If $ABS \geq 625$ then customer is classified into HIGH CLV group.
iii. If $ABS < 192$ AND $IPR < 17$ then customer is classified into the MEDIUM CLV group.
iv. If $192 < ABS < 625$ then customer is classified into the MEDIUM CLV group.
v. If $ABS < 192$ AND $IPR \geq 34$ then customer is classified into the LOW CLV group.
vi. If $ABS < 192$ AND $IPR < 34$ then customer is classified into the MEDIUM CLV group.

Given the small size of the cells generated from rule v. and vi., these rules were not taken into account when classifying the observations; thus, only rules i. through iv. were used to generate the prediction matrix below.

Table 6: CART Prediction Accuracy from Validation Data without Inflation

		A Priori CLV Classifications		
		LOW	**MEDIUM**	**HIGH**
CLV Classification according to Training Tree	**LOW**	90%	10%	-
	MEDIUM	8%	91%	1%
	HIGH	-	13%	87%

From Table 6, it is confirmed that removing the inflation variable has not resulted in a decrease in prediction accuracy, but rather marginally improved the prediction of medium CLV customers. Thus, based on the above prediction accuracy table, findings from the training classification tree, as well as the initial box plots, the decision to retain the validation classification tree without the inflation variable was made.

5 Discussion

In Chapter Two of this paper, the hypothesis in question and theoretical model of this study were provided. The theoretical model hypothesised that basket size, inter-purchase time, cross-shopping (CS) and inflation had an effect on customer lifetime value (CLV) and that these could be used to create a CLV classification system as indicated by the theoretical model below.

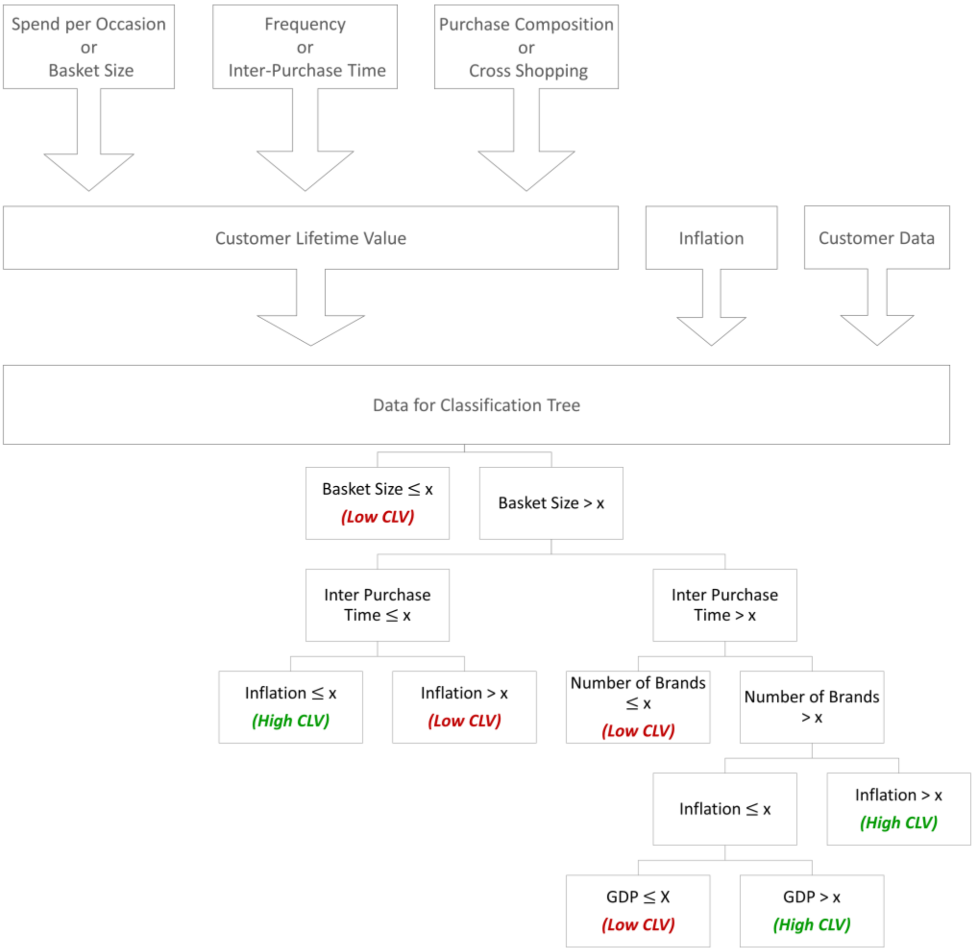

Figure 25: Theoretical Model for MMSM Dissertation

Based on the findings in Chapter Four, the following assertions can be made conclusively:

Table 7: Summary of Acceptance or Rejection of Hypothesis

Acceptance or Rejection of Hypothesis

	Accept	Reject
$\boldsymbol{H_1}$	✓	
$\boldsymbol{H_2}$		✓
$\boldsymbol{H_3}$	✓	
$\boldsymbol{H_4}$		✓

5.1 Hypothesis One (H_1): A customer's basket size influences CLV

$$H_0: Customer's\ spend\ does\ not\ affect\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ who\ spend\ more\ with\ the\ firm\ are\ likely\ to\ have\ a\ higher\ CLV$$

Average basket size (ABS) was the first variable to be included in the classification tree as a decision node, with very distinct partitioning between customers who were to be classified as having a low CLV and those classified as having a high CLV. This implied that there was sufficient evidence to reject the null hypothesis in favour of the alternate hypothesis H_A. In addition, without specifying the complexity parameter (CP), the R software ran a classification and regression tree (CART) on the training data set that automatically optimised the tree after the ABS variable had been included.

This reaffirms the work done by Reinartz and Kumar (2003), which proved purchase amount or basket size to be positively correlated to profitable lifetime duration or CLV. This assertion was also made by Zhang, Dixit, and Friedmann (2010) who found that customer revenue, comprised of basket spend and frequency, resulted in increased CLV. Thus, the results from the study are aligned to previous findings, confirming the rejection of H_0 in acceptance of H_A, namely that the customers with a larger basket size tend to have a higher CLV.

5.2 Hypothesis Two (H_2): Basket composition influences CLV

$$H_0: Cross\ shopping\ does\ not\ affect\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ who\ cross\ shop\ tend\ to\ have\ a\ higher\ CLV$$

Upon examining the tree results, there is insufficient evidence to reject the null hypothesis, and thus conclude that CS does not affect a customer's CLV classification. This was evidenced by the fact that the CS variable, even with deeper trees was not considered a significant factor to include in the classification tree for both the training and validation data sets. This observation is therefore in direct contradiction to Reinartz and Kumar (2003). To ensure that this was indeed the case, a tree was rerun (see Appendix I) excluding the ABS variable although CS was included, and the overall prediction accuracy of the tree was compromised. Thus, it can be said with certainty that CS is not a significant factor in determining whether a customer will have a low, medium or high CLV. Additional research has given light to the fact that although CS or add-on selling is important, it will only enhance customer equity if a company is able to do so at efficiency rates higher than those used to initially acquire customers (Hansotia, 2004a). Leone et al. (2006) further noted that the extent of cross/add-on selling "is influenced by the branding strategy a firm uses, be it house of brands, or branded house" (Leone et al., 2006), with firms that employ a house of brands strategy likely to have greater success. In the case of this study, despite having a house of brands strategy, the manufacturer is only active in a single product category and thus cannot leverage CS to the extent that it will improve CLV. This further substantiates the inability to reject H_0.

5.3 Hypothesis Three (H_3): Purchase frequency influences CLV

$$H_0: Customer's\ purchase\ frequency\ does\ not\ affect\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ with\ a\ high\ purchase\ frequency\ are\ more\ likely\ to\ have\ a\ higher\ CLV$$

The tree produced by the CART process included purchase frequency as one of the decision nodes determining a customer's classification. This implied that given the purchase behaviour exhibited by the customers in the data set, there was sufficient evidence to reject the null hypothesis in favour of the alternate hypothesis H_A. Through the study, it was found that customers who made purchases more frequently tended to be classified into a higher CLV grouping than those who made purchases less frequently. This assertion has also been found to hold in previous literature. In their study, Rust et al. (2004) took into account the notion of inter-purchase intervals or equivalent purchase frequencies when improving the CLV calculation and noted that customers with higher inter-purchase intervals or lower frequency were more likely to be subjected to higher discounting and therefore a lower CLV, in comparison to customers with a lower inter-purchase interval or higher frequency. Subsequently, Furinto, Pawitra, and Balqiah (2009) also found that loyal customers who tended to have a higher purchase frequency had a higher CLV. Thus, the results from the study are aligned to previous findings confirming the rejection of H_0 in acceptance of $H_{A,}$ namely that a higher purchase frequency does indeed lead to a higher CLV.

5.4 Hypothesis Four (H_4): Socio-economic elements influence CLV

$$H_0: Inflation\ does\ not\ affect\ a\ customer's\ CLV\ classification$$

$$H_A: Customers\ subjected\ to\ a\ higher\ inflation\ have\ a\ lower\ CLV$$

Based on the study results, inflation was not a factor in determining how a customer is classified. This was due to the fact that it was not included in the classification tree derived from the training data set. Although it was included in the classification tree derived from the validation data set, its inclusion did not significantly improve the prediction accuracy of the classification tree. This implies that there is insufficient evidence to reject the null hypothesis and thus conclude that inflation is not a significant factor in determining whether customers will have a low, medium or high CLV. Therefore, H_0 still holds.

Accordingly, this does not improve the current body of knowledge. However, a study by Kumar and Pansari (2016) did look at economic factors utilising GDP per capita as a proxy for disposable income and found that for multinationals, GDP per capita does indeed have an impact on purchase frequency and contribution margin and in turn CLV (Kumar & Pansari, 2016).

5.5 Summary of Findings

The focus of this study was to look at the CLV of business-to-business (B2B) customers and understand how it can be used for customer segmentation and with the aid of CART to provide decision rules for classifying new customers. The analysis revealed that basket size and inter-purchase times were significant factors in determining whether a customer would be classified as low, medium or high CLV. These findings are in line with the study conducted by Fader, Hardie, and Lee (2005). Fader et al. (2005) found that “monetary value per transaction appears stable over time and independent of recency and frequency” (Fader et al., 2005). Fader’s finding links to this study in that firstly, ABS represents monetary value as the initial variable to define whether a customer is low or high value. Secondly, given that this is based off the Venkatesan-Kumar (VK) approach for CLV calculation, the ABS per annum was found to be relatively stable per customer over time similar to Fader’s assertion. Fader et al. (2005) also found that recency (defined as: time of most recent purchase) was found to be a more “powerful discriminator of CLV than frequency and monetary value” (Fader et al., 2005). The similarity in these findings is that although the ABS is the initial differentiator, inter-purchase rate is the second most important, in that it is able to provide a way to fine-tune the classification of the customers. The difference between these two studies would then be the order of importance of the variables. For instance, RFM is so named in order of importance of the variable to influence CLV, whereas in this study, the ordering would be MR(F), with F being brackets as it did not explicitly form part of this study, but was taken into account in the VK method of calculating CLV.

When compared to the work done by Kumar and George, this work reiterates their findings in that ABS, or what they term as customer-spending level has a positive influence on (relationship to) CLV (Kumar & George, 2007). With respect to the IPR, this study found a shorter IPR to result in a higher CLV, thus assuming a linear relationship. This, however is not directly in line with Kumar and George, as they found IPR or average inter-purchase time to have a U-shaped relationship to CLV (Kumar & George, 2007), implying that after a certain point, the increase in number of purchases or decrease in time between purchases does not have an incremental effect on CLV. This study, however departs from Kumar's work with respect to CS as CS was not found to have a significant impact on CLV or to aid in determining a customer's CLV classification.

Therefore, based on the study's findings it can be concluded that the ABS and IPR are significant influencers of CLV and these variables can be used to create decision rules to classify customers into homogenous CLV groups. The study has added to the current body of work by reiterating previous findings, thus confirming the validity of these assertions (Kumar & George, 2007; Reinartz & Kumar, 2003; Rust et al., 2004). Furthermore, by adding to the current body of work, it provides greater generalisability of the CLV findings within a consumer packaged goods setting.

5.6 Problem Definition and Answer

The research question posed in section 1.3 was stated as follows: *Can purchase frequency, basket size, basket composition and inflation influence CLV? Can these variables be further used to segment customers on the basis of their CLV?*

Through the findings in this study, it has been confirmed that basket size and frequency influence CLV and can be further used to segment customers. This is in line with prior research (Kumar et al., 2008; Reinartz & Kumar, 2003; Tirenni et al., 2007; Venkatesan & Kumar, 2004; Zhang et al., 2010). With respect to basket composition and inflation, this study did not find these variables to be significant in segmenting customers based on their purchasing behaviour. These findings are in contradiction to current literature (Hansotia, 2004a; Leone et al., 2006; Reinartz & Kumar, 2003) and may be as a result of the nature of the dataset being limited to one supplier and one macroeconomic landscape.

6 Research Implications

6.1 Methodological Implications

This study used the classification tree method to derive a set of rules by which customers can be classified into predefined CLV segments. The use of classification trees is not new and has been used in prior research. Examples of this include studies by Tirenni et al. (2007); Manlio Del, Campanella, and Dezi (2016); and Bezabeh (2017).

Tirenni et al. (2007) used Classification trees to find a manner to classify customers of a European airline into a high-value segment based on spend.

Manlio Del et al. (2016) used CART in his study in order to determine whether the internet of things has an effect on a bank's profitability. The CART method was used to classify customers into Low vs. High Return on Equity based on predetermined variables and was found it be an apt method as it: Identified the determinants of the dependent variable; Discarded redundant data; Searched for interactions between predictive variables; generated classification rules; searched for non-linear or non-monotonic relationship (S. C. P. Lemon, Roy, Clark, Friedmann, & Rakowski, 2003; Manlio Del et al., 2016)

Bezabeh (2017) used a decision tree method to create a classification model for a customer relationship management system for the Ethiopian Revenue and Customs Authority. In his study of Bezabeh (2017) compared decision trees with neural networks and found the decision trees to be better at creating homogenous clusters based on customer characteristics.

This study has been similar in this respect using CART as a classification tool. It has contributed to the current body of work in that it provides an empirical example as to the value of using CART as a segmentation and classification method that can generate classification rules with high accuracy.

6.2 Theoretical Implications

In keeping with the current body of literature; this study found basket size and purchase frequency to be factors that influence CLV.

This study did not find basket composition to influence CLV. This is in contrast to the current body of literature and may have been as a result of the data available in the study. What it alludes to is that basket composition will not influence CLV if it is a single category. This instance may only be applicable to a subset of industries such as FMCG wherein the manufacturer only plays in one category.

In this study, Socio-economic elements, specifically inflation was not found to influence CLV. This implies that in a single market economy; socio-economic elements have limited ability in influencing CLV in a B2B setting. Thus further studies similar to Kumar and Pansari (2016) need to be conducted in a B2B setting looking at various socio economic indicators such as inflation; GDP and exchange rate impacts on CLV.

6.3 Managerial Implications

Findings in this study have shown that there is a definite direct relation between customer spend and purchase frequency to CLV with spend having a stronger impact on a customer's CLV than purchase frequency. This assertion is useful for companies in any industry as it affects their bottom line. Gross revenue is determined as price multiplied by quantity $(r = p \times q)$ which in CLV terms would refer to customer spend. For Luxury goods; price would be high with low quantities sold; whilst for commodities; price would be low with high quantities sold.

For managers, this then becomes a balancing act on one of the key 4Ps which is price. Optimising the price based on the target market and the product type will optimise the revenue and thus customer spend for a firm. The purchase frequency, can then be considered as a multiplying factor for the revenue from a customer $rt = t(p \times q)$ where t is the number of purchases made in a given time period. The only way in which management can influence this factor is through value which the customers' places on doing business directly with the manufacturer especially in a commodities environment. This then makes the servicing of customers key in any business in order to drive repeat business.

Thus in order to for management grow the customer equity; they need to ensure that 1) the price of the goods sold is optimised within both the competitive set with respect to competing goods as well as alternative sourcing options in a B2B setting and 2) that the value equity derived from doing business with the manufacturer drives repeat purchasing and fosters customer loyalty.

6.4 Policy Implications

The Competition Act is a policy that "aims to promote and maintain competition in the Republic of South Africa" (The Competition Act, 1998). This act governs the activities of most industries and is applicable to the CPG industry, wherein one of the key success metrics is to gain share of market. In section 2 of the act, it details economic activities that are prohibited in South Africa as these may have a detrimental impact on competitiveness and thus may hamper the development of the economy.

In putting in place customer-centric initiatives to drive growth through customer development and/or retention, companies should be cognisant of the legislative limitations as governed by the Competition Act as it now stands. Thus, when influencing CLV through financial or structural initiatives such as discounts and service agreements, they should ensure that these activities do not have the effect of contravening the prohibited restrictive vertical practices regulation, which is defined as follows: "An agreement between parties in a vertical relationship is prohibited if it has the effect of substantially preventing or lessening competition in a market, unless a party to the agreement can prove that any technological, efficiency or other pro-competitive gain resulting from that agreement outweighs that effect".

6.5 Recommendations

The traditional trade market is a distinct market in that it is a market driven by volumes and is hinged on the relationships that sales personnel have with the shop owners. It is for this reason that understanding the customers' (retail partners') purchasing behaviours and being able to segment them into homogenous groups according to their customer lifetime value (CLV) may assist manufacturers to service this part of the trade more efficiently. This study has learnings which can be used to make business function more efficiently. Through the knowledge and understanding of the customer's basket size and purchase frequency, the business can optimise its operations through the sales team to either drive "customer development", by increasing the basket size, or "customer retention", by increasing purchase frequency (Hansotia, 2004a).

6.5.1 Classification Rules

The study enables us to create classification rules based on purchase behaviour in order to classify customers into High; Medium; and Low CLV. A final set of rules was derived based on the rules generated by the training and validation datasets in section 4.4. wherein limits differ; the median of the values determined by the training data set and the values determined by the validation dataset were used. This led to the following classification rules:

i. If $ABS < 156$ then customer is classified into the LOW CLV group
ii. If $ABS \geq 645$ then customer is classified into HIGH CLV group.
iii. If $156 \leq ABS < 645$ AND $IPR \geq 30$ then customer is classified into the LOW CLV group

iv. If $156 \leq ABS < 645$ AND $IPR < 30$ then customer is classified into the MEDIUM CLV group

6.5.2 New Customers

The rules in section 6.5.1 can assist management in classifying new customers into CLV segments. By observing a customer's purchasing behaviour within the first three months of trade, management would be able to tell what CLV segment they belong to and thus adjust their engagement and investment level based on this.

6.5.3 Low CLV Customers

With low CLV customers, also known as "strangers" according to Kumar (2018), management needs to first segment them into a further two groups: those whose CLV result in a break-even financial position for the manufacturer, and those whose CLV result in a profit position for the manufacturer. Those in the break-even segment are going to erode the manufacturer's financial reserves and will not be an efficient use of the manufacturer's resources. Thus, management should consider no longer directly servicing this group, but instead having them serviced by other trade partners such as wholesalers (Kumar, 2006).

For those who generate profits for the manufacturer but have low CLV, the management needs to put in place initiatives that will influence the purchasing behaviour of these customers such that they increase their average basket size (ABS). These customers in a consumer packaged goods (CPG) environment are strategic, as being stocked in these outlets implies that a brand has penetration into the traditional trade market and assists with maintaining brand awareness and consumption amongst consumers.

A way in which management can increase the basket size is to consider bundle pricing the products. Bundle pricing, which is defined as the "pricing strategy in which the price of a set of products is lower than the total of the individual prices of the components" (Law, 2016), can be used to drive a larger basket by coupling a brand that is a high-volume driver with one that still needs to gain traction in the market. By bundle pricing the brands, this will influence low CLV customers to purchase a larger basket size given the pricing structure and thereby improve the customer equity.

In addition, the business can optimise its operations when servicing its low CLV customers by ensuring that they only call upon these customers as per their defined inter-purchase rate and not on a weekly basis. This will free up human resources for the sales team to place more focus on the more profitable customer segments.

6.5.4 Medium CLV customers

Given the findings in the study, medium CLV customers have a decent-sized basket, however, they shop less frequently than high CLV customers. Kumar (2018) identified medium CLV customers as “butterflies or barnacles”, depending on whether they had low repeat rates and high profitability or high repeat rates and low profitability. With these type of customers, management should aim to retain these customers whilst at the same time influencing them to purchase more frequently (in the case of butterflies), or drive a bigger basket (in the case of barnacles).

In order to drive this change in behaviour, management can consider either one of two strategies. The first is hybrid-bundle pricing, which is when the price of the goods and the cost to service them are bundled together and provided at a discounted price to the customer. The manufacturer can achieve this by offering a discount on the product price and, in addition to that, having trading terms that entice the customer to purchase directly from the manufacturer instead of going to a trade partner such as a wholesaler. An example of such would be offering a product at a discounted rate, providing point of sale material at no additional cost to the customer, and increasing the number of interactions with sales personnel per week (e.g. two deliveries per week instead of one).

An alternative strategy in order to drive a bigger basket would be to upsell these customers to purchase a wider range of the manufacturer's portfolio, either by getting them to stock different variants of the same brand or to instead stock a range of brands.

6.5.5 High CLV Customers

High CLV customers have a large basket size and shop relatively frequently; Kumar (2018) referred to these customers as "true friends". These are top-tier customers which a CPG company should invest in for the long term. Given that these customers already purchase frequently and with a sizeable basket, what is key is to develop the relationship equity in order to drive retention, as loyalty will drive repeat purchases which in turn drives CLV and profitability.

Areas which can improve the relationship equity of high CLV customers include: pre-sales support, post-sales support, order management and fulfilment and ongoing relationship support (Piscopo, 2013). An example of each which may improve customer relations and thus result in trust, commitment and loyalty include:

i. **Pre-Sales Support:**
    a. Provision of samples prior to new product launches.
    b. Provision of merchandising and point of sale material to facilitate sales.
ii. **Order Management and Fulfilment:**
    a. Consideration in creating an app for replenishing stock in order to drive a faster turnaround time as customers are currently serviced once a week.
        i. This could either take the form of an app, a USSD-type platform or a traditional call centre.
iii. **Post-Sales Support:**

a. Courtesy call from a call centre after order fulfilment to check if all was to the client's expectations.

iv. Ongoing Relationship Support:

a. Regular face-to-face meetings with senior sales managers.

6.5.6 Decision Tree

Based on the recommendations given above; the guidelines unearthed in this study and based on the data analysis can be summarised as follows:

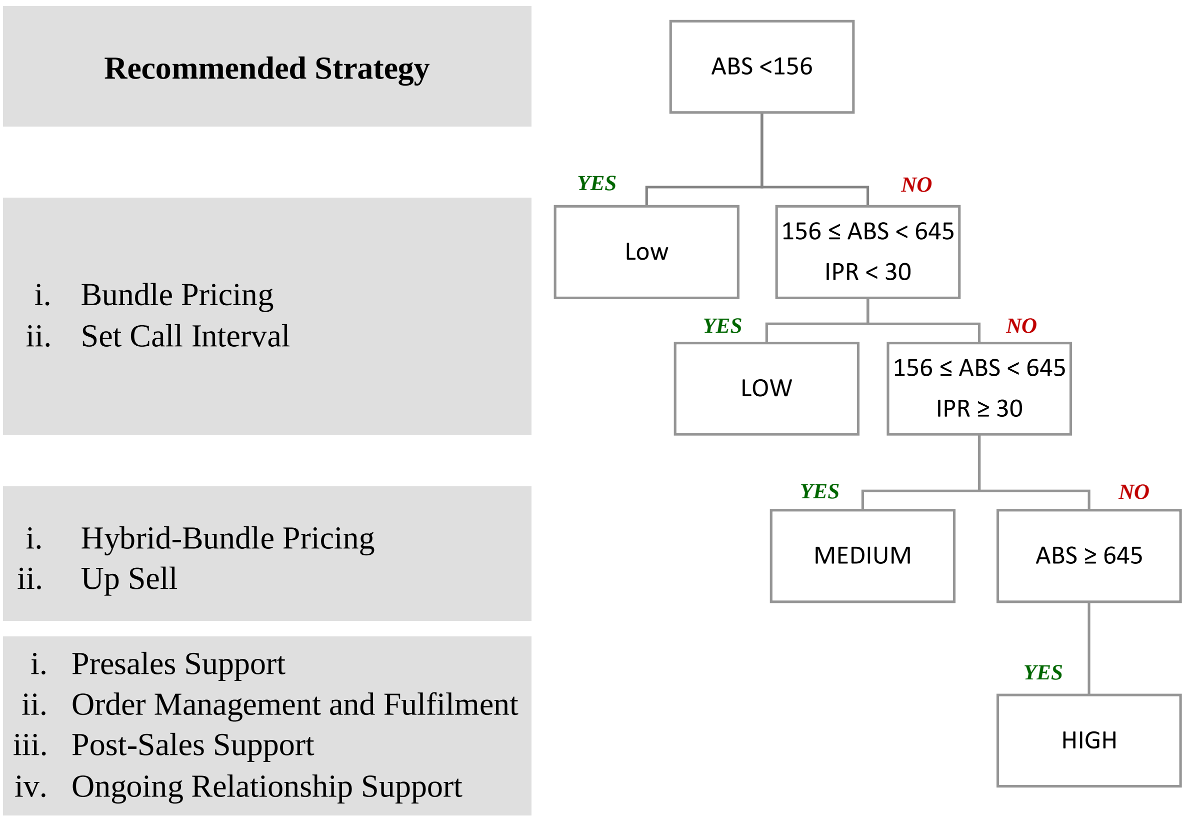

Figure 26: Final Decision Tree and Recommended Strategies

with respect to generalisability; the same recommendations hold in other studies taking into consideration the set points for each respective variable. This is diagrammatically depicted in Figure 27 Below with SP_A_{ABS} referring to set point A for variable ABS; SP_B_{ABS} referring to set point B for variable ABS and SP_A_{IPR} referring to set point A for variable IPR.

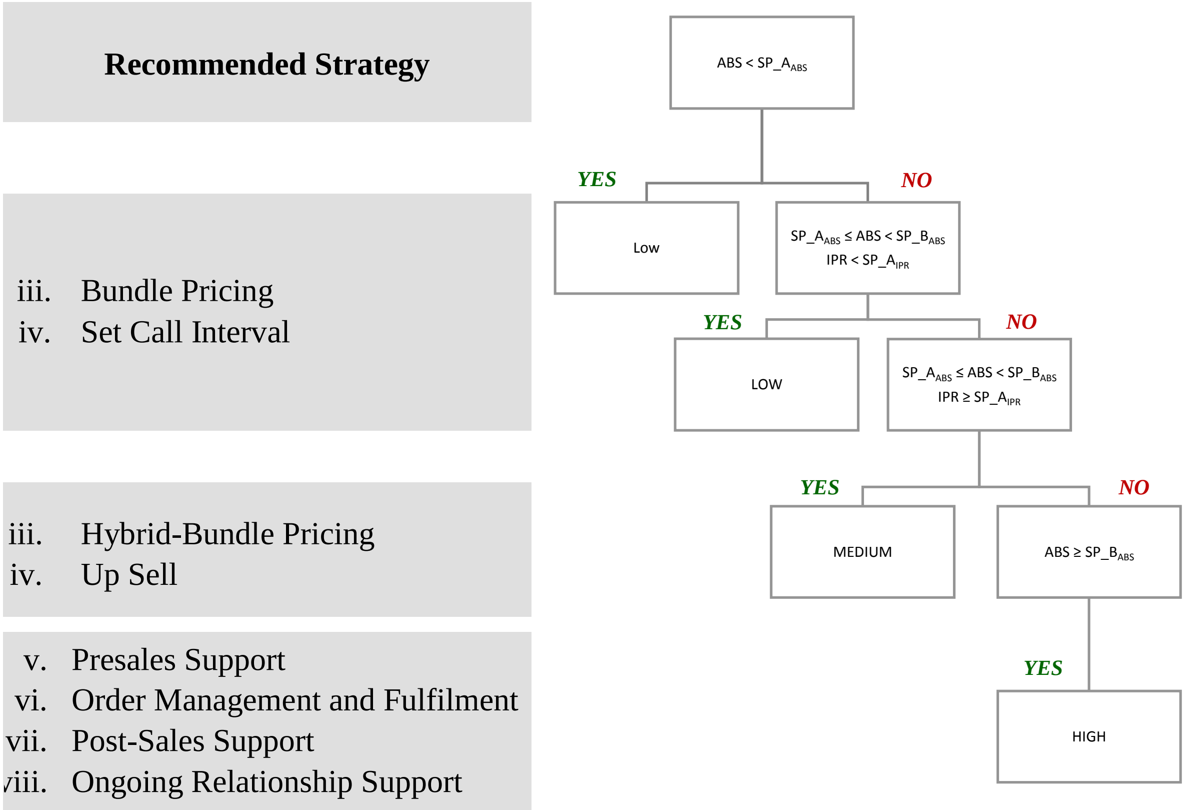

Figure 27: Generalised Decision Tree and Recommended Strategies

6.5.7 Performance Tracking

A final step would be for management to track the initiatives that they have put in place to maximise customer equity through customer development and retention to see if this is indeed achieved. This can be done by setting up scorecards to track purchase frequency, basket size, as well as marketing and/or acquisition costs by segment (Hansotia, 2009). By creating these scorecards, management will be able to track their investment in both sales and marketing initiatives in order to understand what the level of return on investment is. By understanding this, management will work more towards a customer-specific marketing strategy, adjusting the investment and support as is deemed necessary dependent on the customer's level of engagement and CLV potential.

6.6 Conclusion

Not all hypothesis in the study were supported. The hypothesis on basket size (H_1) impacting CLV was supported indicating a direct relationship between a customers' spend and their worth to a company's profitability.

The hypothesis on Cross Shopping (H_2) was not supported and this may be attributable to the study's limitation of only looking at one product category. This would thus require future research or an amended research design to assess whether this is indeed the case for B2B environments.

The hypothesis on purchase frequency (H_3) was supported also proving that a direct relationship exists between how often a customer shops and their CLV. This finding can help customer strategy and affirms business models focused on volumes such as low cist retail stores like USave and PEP stores wherein the value of the customer basket is small, yet the frequency is high thus driving volume sales and profitability.

The inflation hypothesis (H_4) was found to not have a strong enough relationship and thus did not attribute to CLV. This is in contradiction to findings from earlier studies and may indicate that additional customer elements, such as total business turnover, may need to be included in future studies in order to conclusively affirm or deny this hypothesis.

In light of the above findings, the study is generalisable to other industries based purely on measured purchasing behaviours such as spend and frequency. Further investigation is required on the additional layers of purchasing behaviour such as cross shopping and the impact of socio—economic dynamics. Proposed improvements to the study are noted in the upcoming section.

7 Contributions to Literature and Future Considerations

7.1 Contributions to Literature

This study has contributed to the current body of literature by providing empirical evidence confirming the drivers of customer lifetime value (CLV) as well as the benefits of using CLV as a segmentation tool. It enhances the literature by providing a practical example based on an emerging market with application in a consumer packaged goods (CPG) business-to-business environment. It also explored the use of decision trees as an analytical method.

7.2 Future Considerations

Future considerations based on the outcomes of this study can be grouped into two buckets; the first being considerations based on the study's findings and the second being considerations based on marketing principles.

7.2.1 Considerations based on the study's findings

In this study, cross-shopping did not prove to be a significant enough factor to include in the current model. This may be due to the fact that in the current study, the manufacturer only caters for a single food category. Future studies may perhaps revisit this topic by looking at a CPG manufacturer that manufacturers multiple categories. By doing so, future studies may reiterate the findings by Reinartz and Kumar (2003) and thus the generalisability of this assertion within the context of a CPG environment.

Secondly, with respect to inflation and macroeconomics, expanding a similar study to a global setting with the inclusion of macroeconomics such as inflation in addition to GDP per capita may expand on the work done by Kumar and Pansari (2016). This would result in enriching the current body of knowledge by forging further links between macroeconomic indicators and customer equity across different markets. This may also have the added benefit of providing a means in which macroeconomic indicators may be used as a proxy to determine the customer equity potential of a market when no customer-specific data is available, thus broadening the application of CLV.

Another element that can be considered as an additional layer to the socio-economic dynamic would be the customer's annual turnover in monetary value. This would be similar to considering an end consumer's disposable income and thus their marginal propensity to consume which would have an impact on their willingness to do business and ultimately CLV standing.

7.2.2 Future considerations based on Marketing Principles

The 7Ps of marketing are defined as: Product, Price, Place, Promotion, Processes, People and Physical Evidence. In both B2C and B2B environments all of these principles need to be considered in order to ensure the overall success of any business venture. With respect to this study; certain elements of the marketing principles can be used to include additional layers into this piece of research in order to confirm possible relationships between variables that may exist.

Price: The price consideration here would be with respect to competition with other distribution channels. Further studies can consider a retail switching matrix wherein they look at how the customer is switching between purchasing directly from the manufacturer and purchasing from an intermediary.

Processes: This could look at the end to end process from placing an order to delivery or requesting a return. What futures studies can consider is to assess whether different types of service platforms; i.e. Hard Copy Order Book in Triplicate; Call Centre; Online Ordering

System/App; have an influence on the customers' perception of value equity and thus resultantly customer equity through their inclination to continue doing business with the

People and Physical Evidence: Although these are separate principles in a B2B environment similar to the one in this study; the customer in most instances has a direct relationship with the ales representative, this would imply that that the sales representative's business acumen and how they present themselves would impact the customers perception of the business and willingness to want to do ongoing business. Future research can look at this relationship in order to understand whether High, Medium or Low CLV customers shared similar perception with respect to the sales representative that services them.

8 References

Aaker, D. A. (1991). *Managing brand equity: capitalizing on the value of a brand name*. New York: The Free Press.

Akekere, J., & Yousuo, P. O. J. (2012). Empirical Analysis of Change in Income on Private Consumption Expenditure in Nigeria from 1981 to 2010. *International Journal of Academic Research in Business and Social Sciences, 2*(12), 188-197.

Alimi, R. S. (2013). *Keynes' Absolute Income Hypothesis and Kuznets Paradox*. Working Paper. Federal Reserve Bank of St Louis. St. Louis. Retrieved from *https://search.proquest.com/docview/1699266132?accountid=15083*

http://ideas.repec.org/p/pra/mprapa/49310.html

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

Bezabeh, B. B. (2017). The Application of Data Mining Techniques to support Customer Relationship Management: The Case of Ethiopian Revenue and Customs Authority. *International Journal of Advanced Studies in Computers, Science and Engineering, 6*(6), 35-41.

Bick, G. N. (2009). Increasing shareholder value through building customer and brand equity. *Journal of Marketing Management, 25*(1-2), 117-141.

Biedenbach, G. (2012). Brand equity in the business-to-business context: Examining the structural composition. *Journal of Brand Management, 19*(8), 688-701. doi:*http://dx.doi.org/10.1057/bm.2011.60*

Biedenbach, G., Hultén, P., & Tarnovskaya, V. (2019). B2B brand equity: investigating the effects of human capital and relational trust. *Journal of Business & Industrial Marketing, 34*(1), 1-11.

Blattberg, R. C., & Deighton, J. (1996). Manage Marketing by the Customer Equity Test. *Harvard Business Review, 74*(4), 136-144.

Bonsu, C. O., & Muzindutsi, P.-F. (2017). Macroeconomic Determinants of Household Consumption Expenditure in Ghana: A Multivariate Cointegration Approach. *International Journal of Economics and Financial Issues, 7*(4).

Bracht, G. H., & Glass, G. V. (1968). The External Validity of Experiments. *American Educational Research Journal, 5*(4), 437-474. doi:10.2307/1161993

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*: Wadsworth Advanced Books and Software, Belmont, CA.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The Concept of External Validity. *Journal of Consumer Research, 9*(3), 240-244.

Carr, A., Drennan, J., & Andrews, L. (2016). Measuring customer equity and managing the customer asset: an explanatory study in the Australian hotels. *Journal of Marketing Management, 32*(17/18), 1670-1700.

Chahal, H., & Bala, R. (2017). Role of Customer Retention Equity in Creating and Developing Brand Value. *Journal of Relationship Marketing, 16*(2), 119-142.

Clark, M. H., & Middleton, S. C. (2010). Internal Validity. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education (Third Edition)* (pp. 90-96). Oxford: Elsevier.

Corproate Finance Institute. (2015). Fast-Moving Consumer Goods (FMCG). Retrieved from *https://corporatefinanceinstitute.com/resources/knowledge/other/fast-moving-consumer-goods-fmcg/*

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. doi:10.1037/h0040957

Dodds, W. B. (1996). Perceived Value: A Dimension of the Consumer Risk Construct. *American Journal of Business, 11*(1), 15-22. doi:10.1108/19355181199600001

Doyle, C. (2016). *A Dictionary of Marketing* (4 ed.): Oxford University Press.

Efanny, W., Haryanto, J., Kashif, M., & Widyanto, H. A. (2018). The relationship between marketing mix and retailer-perceived brand equity. *IMP Journal, 12*(1), 192-208. doi:10.1108/IMP-12-2016-0019

Ekinci, Y., Ulengin, F., & Uray, N. (2014). Using customer lifetime value to plan optimal promotions. *The Service Industries Journal, 34*(2), 103-122.

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research, 42*(4), 415-430.

Fisher, J. D., Johnson, D. S., Smeeding, T. M., & Thompson, J. P. (2020). Estimating the marginal propensity to consume using the distributions of income, consumption, and wealth. *Journal of Macroeconomics, 65,* 103218. doi:*https://doi.org/10.1016/j.jmacro.2020.103218*

Frow, P., & Payne, A. (2009). Customer relationship management: a strategic perspective. *Journal of Business Market Management, 3*(1), 7-27. doi:10.1007/s12087-008-0035-8

Furinto, A., Pawitra, T., & Balqiah, T. E. (2009). Designing competitive loyalty programs: How types of program affect customer equity. *Journal of Targeting, Measurement and Analysis for Marketing, 17*(4), 307-319. doi:*http://dx.doi.org/10.1057/jt.2009.20*

Galletta, S. (2016). On the Determinants of Happiness: A Classification and Regression Tree (CART) Approach. *Applied Economics Letters, 23*(1-3), 121-125. doi:*http://www.tandfonline.com/loi/rael20*

Gupta, S., & Lehmann, D. R. (2003). Customers as Assets. *Journal of Interactive Marketing, 17*(1), 16.

Haenlein, M., Kaplan, A. M., & Beeser, A. J. (2007). A Model to Determine Customer Lifetime Value in a Retail Banking Context. *European Management Journal, 25*(3), 221-234. doi:*https://doi.org/10.1016/j.emj.2007.01.004*

Hansotia, B. (2004a). Company activities for managing customer equity. *Journal of Database Marketing & Customer Strategy Management, 11*(4), 319-332.

Hansotia, B. (2004b). Customer metrics and organisational alignment for maximising customer equity. *Journal of Database Marketing & Customer Strategy Management, 12*(1), 9-20.

Hansotia, B. (2009). Marketing by objectives: Using segmentation based on purchase timing to enhance customer equity. *Journal of Direct, Data and Digital Marketing Practice, 10*(4), 336-355. doi:*http://dx.doi.org/10.1057/dddmp.2009.1*

Hosseini, M., & Shabani, M. (2015). New approach to customer segmentation based on changes in customer value. *Journal of Marketing Analytics*, *3*(3), 110-121. doi:*http://dx.doi.org/10.1057/jma.2015.10*

Jain, D., & Singh, S. S. (2002). Customer Lifetime Value Research in Marketing: A Review and Future Directions. *Journal of Interactive Marketing, 16*(2), 13. doi:IO.IOO2/dir.lOO32

Johnson, S. (2018). African consumer companies see earnings go up in smoke. *FT.com*.

Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement, 38*(4), 319-342.

Kashani, F. H., & Shahmirzaloo, Z. (2017). Developing Marketing Strategies using Customer Relationship Management and Data Mining (Case Study: Perperook Chain Restaurants). *Разработка стратегий маркетинга с использованием управления взаимоотношениями с клиентами и сбора данных (пример исследования: сеть ресторанов Perperook).*(3), 135-148. doi:10.21272/mmi.2017.3-13

Keller, G. (2005). *Statistics for Management and Economics* (C. Levesque Ed. Seventh ed.). USA: Thomson Higher Education.

Keller, K. L. (2013). *Strategic Brand Management: Building, Measuring and Managing Brand Equity* (S. Wall Ed. 4th ed.). England: Pearson Education Limited.

Kumar, V. (2006). CLV: The Databased Approach. *Journal of Relationship Marketing, 5*(2/3), 7.

Kumar, V. (2018). A Theory of Customer Valuation: Concepts, Metrics, Strategy, and Implementation. *Journal of Marketing, 82*, 1-19. doi:10.1509/jm.17.0208

Kumar, V., & George, M. (2007). Measuring and Maximizing Customer Equity: A Critical Analysis. *Journal of the Academy of Marketing Science, 35*(2), 15. doi:DOI 10.1007/s11747-007-0028-2

Kumar, V., Lemon, K. N., & Parasuraman, A. (2006). Managing Customers for Value: An Overview and Research Agenda. *Journal of Service Research*, *9*(2), 87-94.

Kumar, V., & Pansari, A. (2016). National Culture, Economy, and Customer Lifetime Value: Assessing the Relative Impact of the Drivers of Customer Lifetime Value for a Global Retailer. *Journal of International Marketing, 24*(1), 1-21.

Kumar, V., Venkatesan, R., Bohling, T., & Beckmann, D. (2008). The Power of CLV: Managing Customer Lifetime Value at IBM. *Marketing Science, 27*(4), 585-599,749,751,753.

Law, J. (2016). *A Dictionary of Business and Management* (6th ed.): Oxford University Press.

Lemon, K. N., & Mark, T. (2006). Customer Lifetime Value as the Basis of Customer Segmentation: Issues and Challenges. *Journal of Relationship Marketing, 5*(2/3), 55-69. doi:10.1300/J366v05n02_04

Lemon, S. C. P., Roy, J. P., Clark, M. A. P., Friedmann, P. D. M. D. M. P. H., & Rakowski, W. P. (2003). Classification and regression tree analysis in public health:

Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine, 26*(3), 172-181. doi:*http://dx.doi.org/10.1207/S15324796ABM2603_02*

Leone, R. P., Rao, V. R., Keller, K. L., Luo, A. M., McAlister, L., & Srivastava, R. (2006). Linking Brand Equity to Customer Equity. *Journal of Service Research : JSR, 9*(2), 125-138.

Mallick, S. K., & Mohsin, M. (2016). Macroeconomic Effects of Inflationary Shocks with Durable and Non-Durable Consumption. *Open Economies Review, 27*(5), 895-921. doi:*http://dx.doi.org/10.1007/s11079-016-9405-0*

Manlio Del, G., Campanella, F., & Dezi, L. (2016). The Bank of Things. *Business Process Management Journal, 22*(2), 324-340. doi:*http://dx.doi.org/10.1108/BPMJ-10-2015-0139*

Nielsen South Africa. (2016). SA FMCG 2016-17 Update.

Palmatier, R. W. (2008). Interfirm Relational Drivers of Customer Value. *Journal of Marketing, 72*(4), 76-89.

Palmatier, R. W., Gopalakrishna, S., & Houston, M. B. (2006). Returns on Business-to-Business Relationship Marketing Investments: Strategies for Leveraging Profits. *Marketing Science, 25*(5), 477-493,548-549.

Palmatier, R. W., & Sridhar, S. (2017). *Marketing Strategy based on First Principles and Data Analytics*. London: Palgrave.

Peter, J. P. (1981). Construct Validity: A Review of Basic Issues and Marketing Practices. *Journal of Marketing Research, 18*(2), 133-145. doi:10.2307/3150948

Piscopo, M. G. (2013). *Effect of Relationship Quality and Cost to Serve on Customer Value in Business Market.* (text), Retrieved from *http://0-search.ebscohost.com.innopac.wits.ac.za/login.aspx?direct=true&db=ddu&AN=5AC5F50E4DBB50AA&site=ehost-live&scope=site* Available from EBSCOhost ddu database.

Ramaseshan, B., Rabbanee, F. K., & Hui, L. T. H. (2013). Effects of customer equity drivers on customer loyalty in B2B context. *The Journal of Business & Industrial Marketing, 28*(4), 11. doi:*http://0-dx.doi.org.innopac.wits.ac.za/10.1108/08858621311313929*

Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, *67*(1), 77-99.

Robert, D. M. (1969). Mathematics of Growth: I. *Financial Analysts Journal, 25*(1), 11-14.

Rozek, J., & Karlícek, M. (2014). Customer Lifetime Value as the 21st Century Marketing Strategy Approach. *Central European Business Review*, *3*(2), 28-35.

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing, 68*(1), 109-127. doi:10.1509/jmkg.68.1.109.24030

Rust, R. T., Zeithaml, V., & Lemon, K. N. (2000). *Driving Customer Equity: How customer lifetime value is reshaping corporate strategy*. New York: The Free Press.

Schonlau, M., Fricker, R. D., & Elliott, M. N. (2002). Background on the Survey Process. In *Conducting Research Surveys via E-mail and the Web* (pp. 5-18): RAND Corporation.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl, K. C. (2017). *Data Mining for Business Analytics : Concepts, Techniques, and Applications in R*. Hoboken, New Jersey: Wiley.

Stahl, F., Heitmann, M., Lehmann, D. R., & Neslin, S. A. (2012). The Impact of Brand Equity on Customer Acquisition, Retention, and Profit Margin. *Journal of Marketing, 76*(4), 44-63.

Statistics South Africa. (2017). *Poverty Trends in South Africa: An examination of absolute poverty between 2006 and 2015* (Report No. 03-10-06). Retrieved from South Africa: *https://www.statssa.gov.za/publications/Report-03-10-06/Report-03-10-062015.pdf*

Statistics South Africa. (2018). Time Series Data Sets (National Statistics). Retrieved 02 March 2019, from Statistics South Africa *http://www.statssa.gov.za/?page_id=1847*

Statistics South Africa. (2019). *Consumer Price Index December 2019*. Pretoria Retrieved from *http://www.statssa.gov.za/publications/P0141/P0141December2019.pdf*.

Taybi, P., & Frankel, J. (1989). Calculating Lifetime Value. *Catalog Age, 6*(4), 101.

The Competition Act, 89 of 1998 C.F.R. (1998).

Thrasher, R. P. (1991). CART: A Recent Advance in Tree-Structured List Segmentation Methodology. *Journal of Direct Marketing, 5*(1), 35-47. doi:10.1002/dir.4000050107

Tirenni, G., Kaiser, C., & Herrmann, A. (2007). Applying decision trees for value-based customer relations management: Predicting airline customers' future values. *Journal of Database Marketing & Customer Strategy Management, 14*(2), 130-142. doi:10.1057/palgrave.dbm.3250044

Treacy, M., & Wiersma, F. (1995). *The Discipline of Market Leaders*. London: HarperCollins Publishers.

Tukel, O. I., & Dixit, A. (2013). Application of Customer Lifetime Value Model in Make-to-Order Manufacturing. *The Journal of Business & Industrial Marketing, 28*(6), 468-474. doi:*http://dx.doi.org/10.1108/JBIM-04-2013-0099*

Venkatesan, R., & Kumar, V. (2004). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing, 68*(4), 106-125.

Verhoef, P. C., & Donkers, B. (2001). Predicting customer potential value an application in the insurance industry. *Decision Support Systems, 32*(2), 189-199. doi:*https://doi.org/10.1016/S0167-9236(01)00110-5*

Vogel, V., Evanschitzky, H., & Ramaseshan, B. (2008). Customer Equity Drivers and Future Sales. *Journal of Marketing, 72*(6), 98-108.

Wedel, M., & Kamakura, W. (1998). *Market Segmentation: Conceptual and Methodological Foundations* (Vol. 22). Boston, Dordrecht, London: Kluwer Academic Publishers.

Wiedmann, K.-p., Hennigs, N., Schmidt, S., & Wuestefeld, T. (2011). The Importance of Brand Heritage as a Key Performance Driver in Marketing Management. *Journal of Brand Management, 19*(3), 182-194. doi:*http://dx.doi.org/10.1057/bm.2011.36*

World Bank. (2020a). Global Economic Prospects. Retrieved 15 Oct 2020 *https://0-databank-worldbank-org.innopac.wits.ac.za/source/global-economic-prospects#advancedDownloadOptions*

World Bank. (2020b). World Development Indicators. Retrieved 15 Oct 2020 *https://0-databank-worldbank-org.innopac.wits.ac.za/source/world-development-indicators#*

Wroughton, L. (2018). Emerging markets set to drive 2018 global growth - World Bank. Retrieved from *https://www.cnbc.com/2018/01/09/reuters-america-emerging-markets-set-to-drive-2018-global-growth-world-bank.html*

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (C. Neve Ed. 6th ed.). London: Sage Publications.

Zhang, J. Q., Dixit, A., & Friedmann, R. (2010). Customer Loyalty and Lifetime Value: An Empirical Investigation of Consumer Packaged Goods. *Journal of Marketing Theory and Practice, 18*(2), 127-139.

9 Appendix

9.1 Appendix I: Example of Data Transformation and Variable Creation

The following steps were executed on the extracted data within Excel in order to get the data set ready to be analysed within the R software. Computed Fields/Variables will be shown with headings highlighted in yellow.

Table 8: Extract of Data Extract from QlikSense Tool

Brand	Customer Name	Customer ID	DC	DTS Region	Route ID	SAP Customer ID	Year	Transaction Date	Invoice Sales Volume	Stales Volume	Buyback Volume
DORITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	0,540	0,000	0,000
GHOST POPS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	0,180	0,000	0,000
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	1,614	0,000	0,000
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	0,864	0,000	0,000
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/17	1,944	0,000	0,000
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/17	0,864	0,000	0,000
DORITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	1,830	0,000	0,000
FRITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	1,110	0,000	0,000
GHOST POPS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	0,150	0,000	0,000
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	1,125	0,000	0,000
NIKNAKS ALL	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	0,330	0,000	0,000
PEANUTS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	0,300	0,000	0,000
PEANUTS AND RAISINS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	2,100	0,000	0,000
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	0,932	0,000	0,000
DORITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	0,870	0,000	0,000
GHOST POPS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	0,180	0,000	0,000
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	0,432	0,000	0,000
NIKNAKS ALL	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	0,600	0,000	0,000
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	0,716	0,000	0,000

i. First take the initial data extraction from the QlikSense database and calculate total volumes as follows: $Total\ Volume = Invoice\ Sales\ Volume + Stales\ Volume + Buyback\ Volumes$. This will then give you the total volume per brand per transaction that will be used to calculate the average basket size.

ii. Create a new column called Transaction Date2 by copying the Transaction Column into this one and change the date format to number. This will then be used to calculate the IPR.

Table 9: Data Extract with Modified Columns

Brand	Customer Name	Customer ID	DC	DTS Region	Route ID	SAP Customer ID	Year	Transaction Date	**Transaction Date2**	Invoice Sales Volume	Stales Volume	Buyback Volume	**TOTAL VOLUME**
DORITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	42341	0,540	0,000	0,000	0,54
GHOST POPS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	42341	0,180	0,000	0,000	0,18
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	42341	1,614	0,000	0,000	1,61
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/03	42341	0,864	0,000	0,000	0,86
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/17	42355	1,944	0,000	0,000	1,94
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/17	42355	0,864	0,000	0,000	0,86
DORITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	1,830	0,000	0,000	1,83
FRITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	1,110	0,000	0,000	1,11
GHOST POPS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	0,150	0,000	0,000	0,15
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	1,125	0,000	0,000	1,13
NIKNAKS ALL	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	0,330	0,000	0,000	0,33
PEANUTS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	0,300	0,000	0,000	0,30
PEANUTS AND RAISINS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	2,100	0,000	0,000	2,10
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/24	42362	0,932	0,000	0,000	0,93
DORITOS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	42369	0,870	0,000	0,000	0,87
GHOST POPS	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	42369	0,180	0,000	0,000	0,18
LAY'S PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	42369	0,432	0,000	0,000	0,43
NIKNAKS ALL	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	42369	0,600	0,000	0,000	0,60
SIMBA PC	JULANE SUPERMARKET	000829	PORT ELIZABETH	Eastern Cape	PEL056	0200405116	2015	2015/12/31	42369	0,716	0,000	0,000	0,72

iii. Create a pivot table with Customer ID in the rows; Year in the Column and Total Volume as the values. Copy the data and paste as values into a new sheet

iv. Calculate the arithmetic mean. The resultant table will look as follows and this will be the data used for the PV values in the CLV calculation

Table 10: Table with PV Values for CLV Calculation

Sum of TOTAL VOLUME	Year					
Customer ID	2015	2016	2017	2018	2019	***PV***
000829	16,681	103,33				60,0055
000907	12,04					12,04
000911	788,321	808,155	616,536	160,182	0	474,6388
001580	3,888	-14,76				-5,436
009498	60,602	65,797	42,433	19,824	0	37,7312
009504	331,429					331,429
009510	-2,136					-2,136
009512	1746,83	1495,254	1352,891		8,858	1150,958
009516	9,936					9,936
009520	821,853	792,344	967,031	486,438	0	613,5332
009815	39,7					39,7
009843	10,841					10,841
009860	31,016					31,016
009873	128,752	126,072	100,8	44,208	0	79,9664
009902	28,08					28,08
010169	748,009	501,369	192,988	703,798	103,072	449,8472
010225	58,642	5,221				31,9315
010228	990,49	884,459	886,261	929,936	905,515	919,3322
010258	61,831	40,054	60,805	56,07	82,765	60,305
010662	130,195	127,518	248,672	768,352	203,668	295,681
010687	239,668	182,527	12,6			144,9317
010688	411,314	398,668	355,461	206,456	262,992	326,9782
010705	3312,604	2004,502	1928,954	2001,642	576,756	1964,892
010706	3107,681	2978,368	3204,67	2668	209,73	2433,69
010710	401,336	875,7	451,7	491,381	375,014	519,0262

v. In the pivot table created in point iv; add the transaction date in the rows after the customer ID and remove the year. This will give the total volumes per transaction per customer.

vi. Add brand into the values and select Count of method. This will give the number of brands per transaction.

vii. Copy this table into a new sheet.

viii. Using the table in point vii to calculate the IPR by taking the difference between the transaction dates for each customer.

ix. Add subtotals using the subtotal option on the Data tab and select average at each change in customer ID for the Sum of Volumes; Count of Brand and IPR

x. The result will be a table of customers with their:

- Average volumes per transaction –the average basket size (ABS)
- Average number of brands per transaction –the cross-shopping (CS)
- Average number of days between orders – the inter-purchase rate (IPR). This will also be used as the frequency proxy in the CLV calculation.

Table 11: Table with Calculated Variables for Analysis

Customer ID	Transaction Date2	Sum of TOTAL VOLUME	Count of Brand	IPR
000829 Average		4,000	5	12
000907 Average		3,010	2	65
000911 Average		15,717	6	11
001580 Average		-5,436	4	274
009498 Average		5,240	4	44
009504 Average		13,257	6	12
009510 Average		-0,712	2	63
009512 Average		27,081	8	10
009516 Average		9,936	3	#DIV/0!
009520 Average		20,316	6	11
009815 Average		19,850	3	239
009843 Average		10,841	5	#DIV/0!
009860 Average		15,508	3	56
009873 Average		13,328	4	53
009902 Average		28,080	3	#DIV/0!
010169 Average		12,358	3	9
010225 Average		12,773	5	21
010228 Average		15,688	7	6
010258 Average		2,792	4	17
010662 Average		11,116	4	14
010687 Average		7,128	2	13
010688 Average		8,934	3	10
010705 Average		36,658	6	7
010706 Average		38,266	7	6
010710 Average		21,099	3	15

9.2 Appendix II: B2B Brand Equity Measurement Framework

Constructs	*Scale items*	*Standardized loadings one-dimensional model**	*Standardized loadings multidimensional model**
Brand awareness	*AW1.* I can quickly recall the logo of XXX (*1-strongly disagree … 5-strongly agree*)	0.17	—
Brand associations	*AS1.* XXX show empathy (*1-strongly disagree … 5-strongly agree*)	0.59	0.62
	AS2. XXX are flexible (*1-strongly disagree … 5-strongly agree*)	0.69	0.71
	AS3. XXX are reliable (*1-strongly disagree … 5-strongly agree*)	0.65	0.67
	AS4. XXX are pragmatic (*1-strongly disagree … 5-strongly agree*)	0.68	0.73
Perceived quality	*PQ1.* How would you evaluate overall quality of XXX services (*1-very low … 5-very high*)	0.70	0.75
	PQ2. How consistent is quality of XXX services (*1-very inconsistent … 5-very consistent*)	0.59	0.64
	PQ3. How would you evaluate quality of XXX services compared with quality of services provided by their competitors (*1-much lower … 5-much better*)	0.53	0.55
Brand loyalty	*LO1.* XXX would be the first choice if my company would need auditing services (*1-strongly disagree … 5-strongly agree*)	0.60	0.69
	LO2. I would recommend XXX services to others (*1-strongly disagree … 5-strongly agree*)	0.71	0.83

*$P < 0.01$.

Figure 28: B2B Brand Equity Measurement Framework (Breidenbach, 2012)

9.3 Appendix III: Classification Tree excluding ABS

A classification Tree built off the Training data set excluding the absolute basket size variable (ABS) is shown below. Although this tree used more of the available variables such as Cross-Shopping (CS); Inter-Purchase Rate (IPR) as well as regions, it resulted in poor prediction accuracy as noted by the prediction table below.

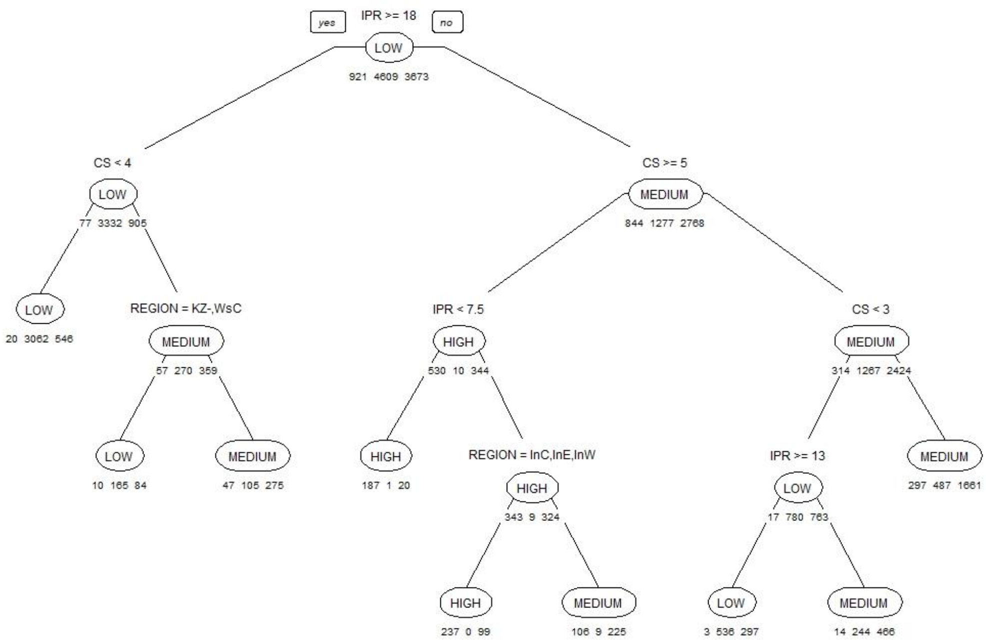

Figure 29: Training Data CART excluding ABS

Table 12: Prediction Accuracy Table for Training CART excluding ABS

		A Priori CLV Classifications		
		LOW	**MEDIUM**	**HIGH**
CLV Classification according to Training Tree	**LOW**	77%	21%	2%
	MEDIUM	33%	59%	8%
	HIGH	1%	37%	61%