**Gordon Institute
of Business Science**
University of Pretoria

The effectiveness of a loyalty programme across income segments

Mahendran Thamannna

14455413

A research project submitted to the Gordon Institute of Business Science, University of Pretoria, in partial fulfilment of the requirements for the degree of Master of Business Administration.

9 November 2015

Abstract

Increased focus on customer relationship management in retail industries has resulted in a growing trend of the use of loyalty programmes to influence customer behaviour. Currently loyalty programmes often fail to return the expected benefit that businesses wish to achieve given the significant costs associated with these programmes. Businesses could seek to increase the effectiveness of a loyalty programme by segmenting its customers and tailoring the loyalty programme design to each segment. Current research offers methods of segmentation based on clients behaviour while on the programme but does not cover any means of upfront segmentation.

The aim of this research is to test if level of income can be used as an effective segmentation tool that allows business to segment loyalty programmes customers upfront in an effort to offer them a more effectively designed loyalty programme.
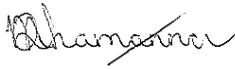
Secondary data was collected from a South African retail bank, comprising of two years transactional information for credit card customers. Analysis and statistical tests, in the form of regressions, was performed on the data to determine whether a significant relationship exists between income and effectiveness of a loyalty programme. The results are then used to establish if income provides an effective means of customer segmentation for loyalty programmes.

Declaration

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.

_____

Mahendran Thamanna

9 November 2015

List of Figures

List of Tables

iii

Contents

1 Chapter 1: Introduction to the Research Problem

1.1 Introduction

In several retail industries the reduction in barriers to entry due to the advancements in technology has led to intense competition and as a result more customer choice (Coussement, 2014). As a result of this customer relationship management has steadily grown in terms of business and academic focus (Stading & Johnson, 2012). Businesses have begun to understand that the acquisition of new customers can be far more expensive than retaining existing ones, and that long, enduring relationships with customers can be highly profitable (Dagger & O'Brien, 2010). The growing trend in retail is to move from a product focused to a customer focused strategy and companies are turning to customer relationship management in an effort to optimise customer lifetime value (Coussement, 2014).

This shift toward customer centricity has led to the introduction of loyalty programmes (Breugelmans, Bijmolt, Zhang, Basso, Kopalle, Minnema, Mijnlieff & Wünderlich, 2014). Dorotic, Bijmolt, & Verhoef (2012) stated that the main purpose of a loyalty programme is to reward customers for loyal behaviour.

The key objective of most loyalty of most loyalty programmes is to increase purchase behaviour and customer retention (Dorotic, Bijmolt, & Verhoef, 2012). Successful loyalty programmes are able to encourage increases in purchase frequency and value, as well motivate the customer to remain with the company (Dorotic, Bijmolt, & Verhoef, 2012). The costs and effort which companies invest into loyalty programmes are often quite substantial (Coussement, 2014). Despite this many loyalty programmes fail to achieve the expected returns and many are forced to terminate (Melnyk & Bijmolt, 2015).

It is thus clear that a deeper understanding of loyalty programme will be beneficial in assisting companies with increasing the effectiveness of their loyalty programmes and thereby achieving its intended objectives.

1.2 Background of the Research Problem

The main variable, within the control of the company that is used to influence the effectiveness of the loyalty programme is its design. Dorotic, Bijmolt, & Verhoef (2012) stated that a company should tailor the design of its loyalty programme to suit its target market in order to maximise its

return on investment. The better the fit of the design of a loyalty programme to its customer, the more effective that loyalty programme will be (Daryanto, Ruyter, Wetzels, & Patterson, 2010).

The design of a loyalty programme is directly affected by the customer's behaviour (Stading & Johnson, 2012). A customer's behavioural traits affect how the customer receives the specific design of the loyalty programme and thereby its effectiveness in encouraging the desired behaviour. Therefore to maximise effectiveness the design of the loyalty programme should align to the customers behaviour on the programme (Dorotic, Bijmolt, & Verhoef, 2012).

Currently literature on loyalty programmes only focuses on customers' behaviour within the loyalty programme (Dagger & O'Brien, 2010), (Kang, Alejandro, & Groza, 2015), (Stading & Johnson, 2012). Research on consumer behaviour and consumer phycology also attempts to link a customer's actions to behavioural traits, and suggests methods of segmentation based on this (Göritz, Büttner, & Anja, 2015), (Hansen, Samuelsen, & Sallis, 2013), (Horváth & Birgelen, 2015). Current literature suggests that segments derived in this manner will result in varying levels of effectiveness of a loyalty programme design and varying designs by segment may increase the loyalty programmes effectiveness.

The gap in current literature is in that there is no research done any methods of segmenting clients upfront, on enrolment to the loyalty programme. An upfront segmentation method would allow companies to predict the customers behaviour once on the programme and thereby offer an appropriately designed loyalty programme.

1.3   Research Objectives and Motivation

Income is a common segmentation method and is used extensively in marketing for this purposes (Jobber & Shipley, 2012). For this reason income was selected as the possible method of segmentation to be tested.

The objective of this research is to test if level of income offers an appropriate means of segmenting customers of loyalty programmes. The aim is to prove that there is a statistically significant relationship between income level and the effectiveness of a loyalty programme and by doing so, offer a new means of customer segmentation for loyalty programmes.

From and academic perspective this research adds to the current body of research on loyalty programmes by expanding on the behavioural segmentation approaches of (Dagger & O'Brien,

2010), (Kang, Alejandro, & Groza, 2015), (Stading & Johnson, 2012). This research follows on from current literature and uses it to suggest other means of segmentation. The current literature focuses on customer behaviour in a loyalty programme and attempts to explain this behaviour using various aspects of behavioural traits (Göritz, Büttner, & Anja, 2015), (Hansen, Samuelsen, & Sallis, 2013), (Horváth & Birgelen, 2015). This research attempts to further this by suggesting a predetermined segmentation approach that may align to the behavioural traits. If successful the research offers a method of predicting the behavioural traits and ultimately the way a customer will receive a particular design of a loyalty programme and answers the call of Melnyk & Bijmolt (2015) who stated, "a call in the literature to broaden research on LP to account for effects of LP designs and cross-customer effects remains unaddressed to date". This will also open the door to new possibilities for customer segmentation approaches under the subject of loyalty programmes.

From a business perspective this will give companies a method of segmenting loyalty programme customers upfront at enrolment or even before, from a targeting perspective. The method is objective and uses information that is easy to obtain thus making it feasible for businesses to use. Companies should tailor the design of their loyalty programme based on customer's behaviour and then use this upfront segmentation method to determine which design is most relevant to a particular customer to ensure maximum effectiveness (Dorotic, Bijmolt, & Verhoef, 2012).

Coussement (2014) stated that the costs associated with loyalty programmes are generally significant. This method of segmenting customers and then appropriately tailoring the loyalty programme design will allow for companies to reduce costs or increase the return on those costs. Melnyk & Bijmolt (2015) stated that a lack of understanding of loyalty programmes and its customers lead to the termination of many and Hoseong & Choi (2013) stated that it is the reason for many companies not realising the expected benefits. This research offers a means to deepen that understanding and translate it into increased effectiveness and ultimately profits.

The hypothesis is that a relationship exists between income level and effectiveness of a loyalty programme will be tested using comprehensive data from one loyalty programme. Effectiveness will be analysed in the components of purchase behaviour and retention. The component of purchase behaviour will be further broken down into purchase frequency and purchase amount. The effectiveness of the loyalty programme will be tested across income segments for each of these components to prove if it indeed a relationship does exist.

3

2 Chapter 2: Literature Review

2.1 Introduction

With the shift toward a customer centric mind-set and the growing popularity of customer relationship marketing, the use of loyalty programmes have increased significantly in recent times (Dorotic, Bijmolt, & Verhoef, 2012). With most industries facing an environment that is continually increasing in competitiveness, it makes sense for businesses to attempt to retain their existing customers (Dorotic, Bijmolt, & Verhoef, 2012). Dagger & O'Brien (2010) stated that in most instances the costs of acquiring new customers outweighs the costs of retaining current customers and longer term relationships offer greater probability than short term ones. It is argued though that simply retaining customers is not the sole purpose of loyalty programmes and in fact the other key objective is to influence purchase behaviour (Zhang & Breugelmans, 2012). Most of the research to date focuses on these two aspects when discussing the effectiveness of loyalty programmes, the aspects of retention and purchase behaviour.

Current literature on loyalty programmes indicates that they are generally effective at encouraging loyal behaviour or customer loyalty (Dorotic, Bijmolt, & Verhoef, 2012). Zhang & Breugelmans (2012) proved that loyalty programmes are effective at both retention and influencing purchase behaviour. As a result of this proven effectiveness we see a proliferation of loyalty programmes in the market, with various structures and designs, however with similar objectives (Hutchinson, Donnell, Gilmore, & Reid, 2015). The effectiveness of loyalty programmes may be clear from its rapid adoption and prior research, however despite this many businesses do not reap the expected financial benefits from their loyalty programmes and low performance of many loyalty programmes result in their termination (Hoseong & Choi, 2013). Based on this is it is evident that a better understanding of loyalty programmes is required to allow for business to increase their effectiveness.

The majority of research to date focuses on the design of loyalty programmes and how it influences the overall effectiveness of the programme (Dorotic, Bijmolt, & Verhoef, 2012). There is also a considerable amount of literature describing the behaviour of customers within a loyalty programme (Stading & Johnson, 2012). Current research suggests that behavioural and psychological traits of consumers influence the effectiveness of the loyalty programme. Thus the

4

two key aspects that influence the effectiveness of loyalty programme, emerging from current literature, are programme design and consumer behaviour.
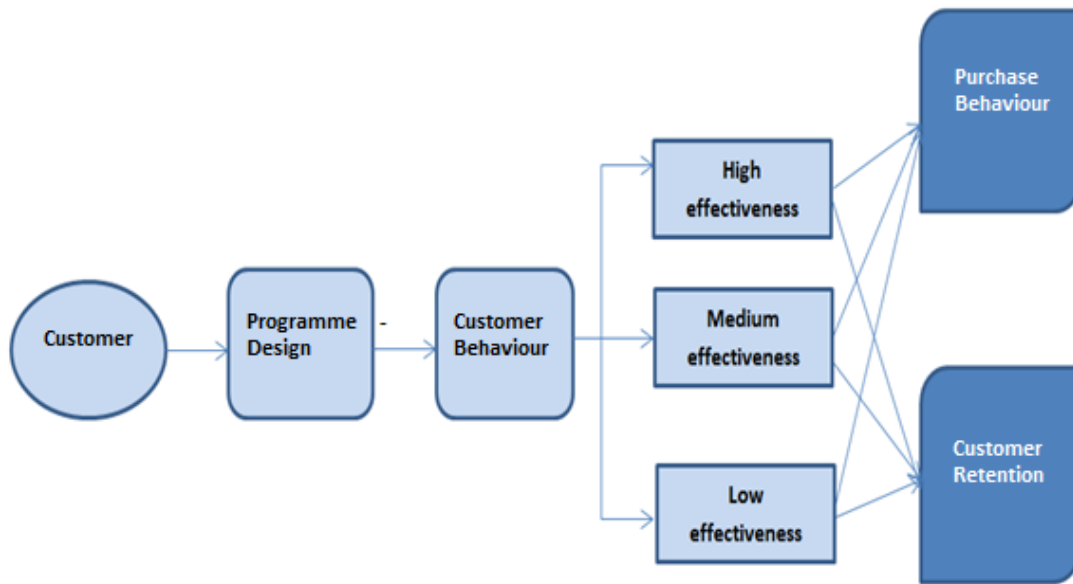
Figure 1 : Current model for loyalty programme effectiveness

In order to achieve a greater understanding of loyalty programme performance, an expanded theoretical model is required (Steinhoff & Palmatier, 2014). Where current research is lacking is to prove a relationship between an objective variable and the effectiveness of a loyalty programme. This can then be used to segment customers upfront and allow companies to tailor a programme's design to the relevant segment, thus improving its effectiveness (Dorotic, Bijmolt, & Verhoef, 2012).

2.2 Customer segmentation

Melnyk & Bijmolt (2015) stated that the call in literature to broaden research on loyalty programmes to account for effects of loyalty programme designs and cross-customer effects remained unaddressed to date. Marketing literature places great focus on customer segmentation but there is a gap when it comes to segmentation and loyalty programmes. Fuchs

5

& Diamantopoulos (2010) proved how differentiation in brand positioning and marketing strategies is more effective if based on a segmented market. It was proved that if segmented correctly the optimal strategy for each segment could vary significantly. Jobber & Shipley (2012) furthered this concept of differentiated strategies based on segmentation, where they proved that different segments would respond significantly differently to the same pricing structure. A theme in marketing research is that a "one size fits all" strategy is rarely successful and marketing efforts need to be tailored to the customer (Dorotic, Bijmolt, & Verhoef, 2012).

Despite this the majority of loyalty programmes incorporate a "one size fits all" strategy. Variations in design between programmes exist, however variations within a specific loyalty programme are not common (Zhang & Breugelmans, 2012). Stading & Johnson (2012) proved that there are different segments within a loyalty programme and the behaviour within these segments varies. Therefore if companies could find a method of segmenting customers with different behaviours upfront, they could vary the design of their loyalty programme to suit each segment. However current literature does not offer a means to identify these segments upfront but rather focuses on the differences in behaviour after enrolment to the programme.

The most common variable used to segment customers is income. Income segmentation is used by retailers to develop customer value propositions that are tailored for each segment. Jobber & Shipley (2012) showed that income, and in particular the ability to pay, is a classic differentiator for customer segmentation. Given its wide use, proven ability to segment a customer base and other reasons expanded in this literature review, income was determined as the best method to use for an upfront segmentation. An added advantage of using income as a metric for segmentation is that it is a clear and objective measure, thus enabling feasible execution in a business context. A statistically significant relationship between income level and effectiveness of the loyalty programme would indicate the need for differentiation in programme design across different income segments. The model would thus be adapted as depicted in figure 1.
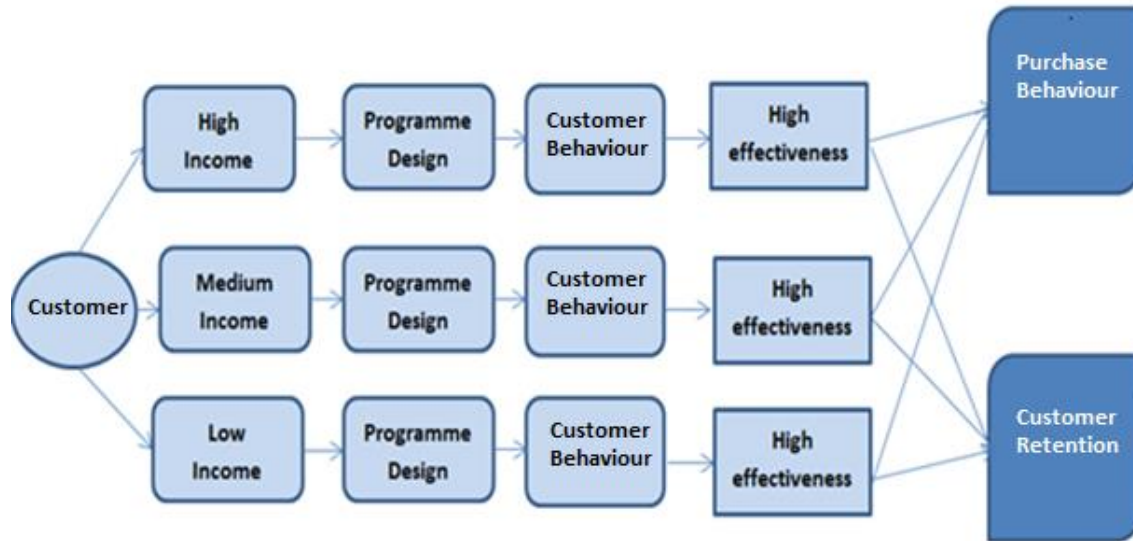
Figure 2 : Desired model for loyalty programme effectiveness

The literature review will now expand on themes and findings in literature that suggest that there could be a difference in effectiveness of loyalty programmes across income segments.

2.3   Financial versus Social benefits

Loyalty programme benefits are often looked at in a one dimensional manner, with only the practical value being focused on. However customers derive a wider range of benefits from loyalty programmes (Henderson, Beck, & Palmatier, 2011).

Mimouni-Chaabane & Volle (2010) stated that there are three collectively exhaustive categories of benefits that consumers derive from loyalty programmes. The first category is Utilitarian benefits, which are your practical benefits such as money, time saved or discounts, the second is Hedonic, which is typically your benefits of entertainment or exploration and finally Symbolic that deals with the social benefits such as status. However most of the current literature breaks the benefit down into two groups. Melancon, Noble, & Noble (2010) refer to the two groups as social and economic benefits whereas, Göritz, Büttner & Anja (2015) refer to the groups as monetary and nonmonetary and the consumers that fit within the groups as task and experiential consumers. Table 1 summarises the characteristics of both groups which will be referred to as social and financial rewards.

7

Table 1: Categories of benefit from a loyalty programme

Perceived benefits of loyalty programmes		
Dimensions of Benefits	**Sub Dimensions of Benefits**	**Definition**
Financial	Monetary and Time savings	To spend less and save money. To reduce choice, and save time and effort.
Social	Recognition and Preferential treatment	To have a special status, to feel distinguished and treated better. To belong to a group that shares the same values.

Mimouni-Chaabane & Volle (2010) proved that within this broad grouping of benefits lie five dimensions of perceived value a consumer seeks to attain from a loyalty programme. Variations of these dimensions and the extent of value desired in each dimension will result in consumers wanting either more social or more financial benefits from a programme. Mimouni-Chaabane & Volle (2010) also revealed in their research that consumers across different groups will vary in these dimensions, which suggest that a programme that is consistent in design across these groups will have varying degrees of effectiveness.

Göritz, Büttner, & Anja (2015) looked specifically at the two groups of consumers and their preference to the aforementioned groups of benefits. These were referred to as task and experiential consumers. Göritz, Büttner, & Anja (2015) described task consumers as consumers that preferred the financial benefit; they were typically described as being utilitarian in their nature and hence chose the more economically sensible benefit. Experiential consumers responded better to the social elements. (Horváth & Birgelen, 2015) had similar findings in their research, the only difference being that they labelled the customers as compulsive and non-compulsive buyers and spoke about emotional and value benefits.

When these findings are overlaid onto income segmentation, lower income consumers would typically fall under task consumers seeking financial benefits, which is a consequence of a lower disposable income. The inverse would apply to higher income consumers.

The typical design of loyalty programme is to reward customers for exhibiting loyal behaviour (Dorotic, Bijmolt, & Verhoef, 2012). The reward is usually in the form of discounts or points that can be used as currency. By this definition loyalty programme offer financial rewards as the key benefit in influencing purchase behaviour. However from a retention perspective the inverse may apply as loyalty programmes offer the social benefit of belonging to a group (Mimouni-Chaabane & Volle, 2010). Therefore is can be established that loyalty programmes typically use financial benefits to influence purchase behaviour but social elements to influence customer retention.

This suggests that loyalty programmes in their current design should vary in effectiveness with income. They may be more effective in the lower income segment at increasing spend behaviour but more effective in higher income segments at enhancing customer retention.

2.4   Reciprocity and Gratitude

Palmatier, Jarvis, Bechkoff, & Kardes (2009) stated that gratitude, the emotional appreciation for benefits received, accompanied by a desire to reciprocate is an important construct for understanding relationship management effectiveness. They proved that when consumers received benefits from a business that they perceived as additional benefit, they felt the need to reciprocate the gesture. This reciprocation of benefit would typically manifest itself in the consumers' loyal behaviour toward the business thus resulting in improved seller performance and purchasing behaviour (Palmatier, Jarvis, Bechkoff, & Kardes, 2009). This is the general intent and purpose of loyalty programmes (Dorotic, Bijmolt, & Verhoef, 2012).

However the simple provision of a benefit is not enough to invoke a feeling of gratitude among customers, it is dependent on how it is perceived by the customer. Palmatier, Jarvis, Bechkoff, & Kardes (2009) stated that in order for a customer to feel gratitude, two very important criteria must be met. The first is that customers must perceive that benefit as being at the expense of the seller and not earned by the customer; and the second is that the customer must believe that the benefit was provided intentionally. Palmatier, Jarvis, Bechkoff, & Kardes (2009) stated that if the customer perceives the benefit as being provided at the discretion of the seller, with a benevolent motive, or with some risk to the seller, he or she should feel more grateful and be more likely to reciprocate.

Based on the design of loyalty programme, the intent of the business to provide a benefit is clearly evident. It can thus be deduced that the key aspect in determining the level of effectiveness of a loyalty programme in generating feelings of gratitude is the customer's perception of whether or not the benefit is earned. Palmatier, Jarvis, Bechkoff, & Kardes (2009) proved that though the provision of benefit may result in reciprocation, it's the level of gratitude invoked that ultimately determined the level at which in is shown.

Jiang, Hoegg, & Dahl (2013) actually had an opposing view; they stated that customers reacted better to earned benefits versus unearned benefits. However their research was based on customers receiving preferential treatment in front of other customers for no effort at all. Jiang, Hoegg, & Dahl (2013) research is therefore not relevant to loyalty programme as the benefit is usually not overtly presented, customers within the programme are generally given the same benefits and at least some effort is required. It is therefore more appropriate to use the stance of Palmatier, Jarvis, Bechkoff, & Kardes (2009).

Reczek, Haws, & Summers (2014) stated that customers reacted differently to a business' promotions based on the level of effort they believed to have invested. The level of effort refers to the historic purchase behaviour of the consumer. Consumers with a perceived high investment, in the form of high purchase behaviour, would feel a sense of entitlement (Reczek, Haws, & Summers, 2014). Even where benefits were randomly distributed, high purchasing customers felt that they were entitled to the benefit. The inverse was true for customers with lower purchase behaviour who were less prone to believing that a benefit was entitled to them. Reczek, Haws, & Summers (2014) stated that the highly invested customer's perception of entitlement would work in the business' favour for promotions with random winners, however the design of most loyalty programme provide a benefit that is certain.

Wagner, Hennig-Thurau, & Rudolph (2009) illustrated this point even further when they proved that high purchasing customers felt entitled to benefits and preferential treatment even if they stopped exhibiting their high purchasing behaviour. They proved that reducing their benefits or perceived status when their purchases reduced actually had a negative effect that was greater than the positive effect of gaining the benefit in the first instance. It can be concluded that with high purchasing customers, the benefit may be seen as a hygiene factor.

When overlaid over the income segmentation, high income customers are typically high purchasers and low income are low purchasers. This would suggest that lower income

customers will feel a greater sense of gratitude to the benefits received from a loyalty programme.

Given this the loyalty programme effectiveness of invoking gratitude and generating customer reciprocity that results in increasing the current levels of purchase behaviour should be higher on lower income customers.

2.5 Programme versus Company Loyalty

Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus (2010) stated that in companies that have loyalty programmes, there are two types of loyalty, programme loyalty and company loyalty. Kang, Alejandro, & Groza (2015) confirmed these two types of loyalty in their research. Stading & Johnson (2012) extended it to four types of loyalty; however the two additional types are basically groups of no loyalty. The distinction between programme and company loyalty as well as the effects of each are consistent across current literature.

Programme loyalty is derived from the positive attitude or response by the customer in relation to the benefits received from and directly associated to the programme (Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus, 2010). The benefits associated with a programme are generally economic in nature such as loyalty points or discounts. Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus (2010) stated that consumers who are loyal to the programme continue to purchase in order to reap the tangible benefit that is received by being a member of the programme. This ties in with the definition of "spurious loyalty", that is high purchase behaviour with no or little affiliation to the company itself (Stading & Johnson, 2012). Kang, Alejandro, & Groza (2015) stated that programme loyalty is a far more fickle relationship than company loyalty and though effective at influencing purchase behaviour, programme loyalty will not be as effective at retention.

Company loyalty is derived from the perceived value the customer derives from their relationship with the company (Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus, 2010). Company loyalty is derived from emotion and is linked to social identity theory. Kang, Alejandro, & Groza (2015) stated that company loyalty occurs when a customer sees their relationship with the company as being part of their social identity. This ultimately leads to the customer exhibiting behaviours that signal their intention to maintain a long enduring relationship with the business (Kang, Alejandro, & Groza, 2015). Company loyalty can be linked

to brand salience which is the prominence of a particular brand or company in the mind of the consumer (Vieceli, Deakin, & Shaw, 2010). This correlation suggests that company loyalty will have a positive impact on customer choice when joining the organisation in additional to the customers decision to stay with the organisation. However company loyalty and brand salience may not be as effective in influencing purchase behaviour (Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus, 2010).

Lower income consumers are in nature utilitarian due to the fact that they have a limited amount of disposable income, which would predispose them to programme loyalty. Mende, Bolton, & Bitner (2013) stated that a customer's attachment style is a critical component in developing company loyalty. Attachment style refers to the customers need to have a close relationship with the company which is a derivative of trust in the business (Mende, Bolton, & Bitner, 2013). Trust in businesses being an issue with lower income consumers further alludes to the fact that lower income consumers would be more prone to programme loyalty as appose to company loyalty.

Higher income consumers in contrast have greater trust in businesses and by virtue of the higher purchase behaviour and interaction with the organisation; they would typically prefer a closer relationship. Higher income consumers are more emotional than utilitarian (Göritz, Büttner, & Anja, 2015), and therefore would be more prone to company loyalty than programme loyalty.

Ideally a company would want the design of its loyalty programme to create both company loyalty and programme loyalty. Company loyalty would get consumers to choose the company and ultimately stay with them while the elements of programme loyalty will encourage purchase behaviour resulting in optimal seller performance (Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus, 2010). This could only be possible if the loyalty programme design was tailored to suit each group. Since loyalty programmes are generic in their design adopting a one size fits all approach, the research suggest the loyalty programme will be more effective at influencing purchase behaviour in lower income groups, who are prone to programme loyalty and more effective at retention in higher income groups that are more prone to company loyalty. This is further evidence to the possibility of a significant relationship between income and the effectiveness of a loyalty programme.

12

2.6   Timing of Rewards

Loyalty programme enable customers accrue or earn points based their behaviour, most commonly their purchase behaviour. These points can then be used as currency to purchase items of goods or services offered by the programme (Dorotic, Bijmolt, & Verhoef, 2012). However how and when customers use these points can vary (Stourm, Bradlow, & Fader, 2015). Behavioural traits of customer segments influence when and how they redeem and it is this timing of rewards that influences the effectiveness of the loyalty programme.

Lower income consumers typically redeem often and for low value items whereas higher income customers tend to stockpile points and redeem infrequently for high value items (Stourm, Bradlow, & Fader, 2015).

Drèze & NuNes (2011) proved that the attainment of frequent recurring goals was much more effective at motivating the desired behaviour than infrequent attainment. They state that achieving targets and attaining recurring goals enhanced perceptions of self-efficacy and instilled self-belief which motivated consumers to pursue the recurring goals to an even greater extent. Drèze & NuNes (2011) found that the effects were most apparent where the goal could be attained frequently yet still remained challenging. Roehm & Roehm (2010) had similar findings when testing customers' response to long term versus short term incentives. They found that customers were more receptive to incentives with a short term horizon. This research suggests those loyalty programmes are more effective at influencing purchase behaviour in lower income consumers than higher income consumers based on their preferred redemption behaviour.

Huang & Zhang (2011) researched at which points of goal pursuit a customer is most motivated. What they found is that a customer is highly motivated at two distinct points in the pursuit of a reward. The first is at the start of the pursuit and the second is when the customer is nearing their goal (Huang & Zhang, 2011). This again suggests a greater effectiveness in lower income segments as the attainment of their goals are frequent so they are very often close to the goal or starting a new pursuit. In contrast higher income customers collect points over a long period of time before redeeming and therefore spend the majority of the pursuit somewhere in the middle.

13

Another aspect that furthers this suggestion is that customers react negatively to diminishing returns when receiving incentives or rewards (Hiasley & Loewenstein, 2011). If high income customers earn points at a lower rate at any time during their lengthy pursuits, the impact on their motivation will be negative.

Sayman & Hoch (2014) tested the differences in behaviour and the effectiveness of a loyalty programme on customers with varying balances. They stated that light and moderate purchasers increased their purchase behaviour but high purchasers didn't, which further suggests the greater effectiveness of loyalty programme on lower income consumers. However according to Sayman & Hoch (2014) customers with higher balances were more entrenched in the programme. Their high balances increased their switching costs and they would rather pay a premium than leave the company. This suggests that though loyalty programme may be more effective in lower income segments at increasing purchase behaviour, they could be more successful at retaining customers in the higher income segment.

The manner in which customers redeem points can be significantly influenced by the design of the programme (Stourm, Bradlow, & Fader, 2015). Variations in design could assist in optimising the effectiveness, however most loyalty programme are static in design across segments. What is clearly evident though is the suggested relationship between income level and the effectiveness of the programme.

2.7   Desire for the Loyalty Programme

Raciti, Ward, & Dagger (2013) stated that a customer's desire to join the programme at the start of the relationship directly influences the level of motivation of the customer while on the programme. They found that customers with a stronger desire to join had a greater motivation to increase purchase behaviour during membership. Raciti, Ward, & Dagger (2013) stated that the main measure used in determining a customer's level of desire was whether or not they had to opt in to the programme and how significant the perceived costs of joining were.

From a banking perspective membership to loyalty programmes are often imbedded into the value proposition of the products for higher income segments. This is not the case for the lower income segment products which suggests a greater proportion of customers in the lower income segments have to opt in and thus display greater desire. Most loyalty programme have a joining fee and some a monthly membership as well (Dorotic, Bijmolt, & Verhoef, 2012). Though the

14

fees are generally consistent across income bands, the perception of the costs will vary for obvious reasons. Based on this it is suggested that lower income consumers that join a membership programme may have a stronger desire to do so and thus will be more motivated (Raciti, Ward, & Dagger, 2013). It can thus be argued that level of income is correlated to the level of desire and thus ultimately the level of effectiveness of loyalty programme.

2.8   Relevance of the Benefit Received

The literature review has already covered the split between financial and social benefits, which already indicates that a static design may be more relevant, in terms of benefits, to one segment than another. However further differentiation can exist within these groups based on the type of benefits offered. Daryanto, Ruyter, Wetzels, & Patterson (2010) stated that the key to an effective loyalty programme is to align the rewards with the customer's needs, wants and objectives in order to enhance the perceived value of the programme. They prove that rewards that are relevant to the customer result in a greater appreciation for the programme. Daryanto, Ruyter, Wetzels, & Patterson (2010) introduce the concept of "regulatory fit" which states that the rewards offered by a loyalty programme must not only align to customer needs and wants but also to the behaviour the company is attempting to drive.

This need for personalisation and fit in rewards offered is echoed in the study by Wierich & Zielke (2014), who looked at the design elements of retailer coupons and how these elements affected purchase behaviour. It was found that personalisation was the greatest driver of loyalty and that the benefit should be personalised to get the best response. The customers cognitive processing of the reward received results in them perceiving a greater benefit when the reward is personalised (Wierich & Zielke, 2014).

Coelho & Henseler (2012) further illustrate the need for customisation as they proved that customisation is a key component in creating customer loyalty. "The key marketing objective is to meet the customer's specific needs, wants and aspirations" (Paswan, Blankson, & Guzman, 2011, p. 312).

Recent research has shown considerable heterogeneity among customer preferences with regards to the rewards offered by loyalty programmes (Daryanto, Ruyter, Wetzels, & Patterson, 2010). Thus it has become more complex for loyalty programme to align its rewards to customers' needs and wants and maintain a regulatory fit. This suggests that the types of

15

rewards offered, both financial and social should vary across different customers segments. Once again in a programme with a design that remains unchanged across income segments the perceived value of the rewards offered will vary and so too will the loyalty programme effectiveness. Which segments derive greater value will be dependent on the rewards offered by the programme, however what can be deduced if the rewards that are preferred vary with income then a relationship between income and effectiveness will exist.

2.9 Familiarity with the programme

One may instantly relate familiarity with the programme with enhanced customer loyalty, if not to the company then to the programme itself. Research suggests that this is indeed not the case; however what level of familiarity does indicate is how a programme should be designed and marketed (McMullan & Gilmore, 2008).

Dagger & O'Brien (2010) examined the differences in relationships between what they deemed experienced and novice consumers. In their study they calculated a customer's level of experience based on the frequency of encounters with the company. Customers with high frequency of encounters or high purchase behaviour would be deemed experienced customers based on their level of interaction with the company (Dagger & O'Brien, 2010). This would be applicable to higher income customers who are typically high purchasers. They labelled low frequency purchasers as novice consumers and would be applicable to lower income customers. Dagger & O'Brien (2010) stated that there were significant differences between novice and experienced customers with the key finding being that the impact of relationship benefits on perceptions of satisfaction, trust and commitment was more profound on experienced customers. This relates to the aspects of social benefits and since loyalty programme are designed more around financial benefits, they may be less effective on experienced or higher income consumers.

Hansen, Samuelsen, & Sallis (2013) expanded on this by introducing a concept called "need for cognition" in the mind of the consumer. Need for cognition is described as the manner in which customers draw on past experiences and satisfaction as appose to external stimuli when deciding on future purchase behaviour (Hansen, Samuelsen, & Sallis, 2013). They stated that the level of the need for cognition would determine what benefits appeal to the customer. More experienced customers had a higher need for cognition and less experienced a lower need for

16

cognition. Hansen, Samuelsen, & Sallis (2013) demonstrated that customers with a low need for cognition prefer financial benefits and customers with need a high for cognition preferred social benefits. This again highlights a disparity between high and low income customers.

McMullan & Gilmore (2008) suggested that customers could be segmented by their level of loyalty, which could again be measured by frequency of purchase. Their findings indicated there should be differentiated approaches to relationship management based on the level of loyalty. According to Noble, Esmark, & Noble (2014), the perceived complexity of the programme should be considered when developing the marketing approach for the programme. Dagger & O'Brien (2010) suggested that frequency of purchase dictates level of experience which ultimately results in the degree of complexity with which the programme is perceived by the customer. Both studies thus suggest that a programme that is static in design and promotion will have differing levels of effectiveness between different levels of experience. With the correlation between income and experience, it can be inferred that there should be a relationship between income level and effectiveness of the programme.

2.10 Churners versus non-churners

Current literature recognises the impact of customer attrition on companies and the strong business need to limit or reduce customer attrition (Dagger & O'Brien, 2010). Customer retention is crucial to business success. A five per cent increase in customer loyalty can improve firm profits anywhere between twenty-five and eighty-five per cent (David & Dagger, 2012). Given this companies incur significant costs on efforts to reduce customer attrition such as introducing loyalty programme.

Coussement (2014) stated that companies could save a portion of these costs if they were able to segment customers based on the probability of them attriting. Coussement (2014) referred to them as churners, customers who were more likely to attrite, and non-churners, customers who were less likely to attrite. He stated that companies should reduce focus on non-churners thereby reducing costs. The core finding was that the benefit or effectiveness of relationship management efforts on non-churners were significantly lower (Coussement, 2014).

David & Dagger (2012) furthered this argument in their research which examined other aspects that may affect the impact of customer relationship management on loyalty. They stated that switching costs had a direct impact on the effectiveness of relationship management tools such

17

as loyalty programme. The level of switching costs is directly related to Coussement (2014) concept of churners and non-churners as the level of perceived switching costs influences the customer's probability of attriting. In fact, David & Dagger (2012) infer that switching costs are the strongest determinant of a customer's probability of attriting.

Across many retail industries switching costs can vary significantly among income bands, depending on the industry itself. As a result the effectiveness of a loyalty programme with regards to retention may vary across income segments.

2.11 Competitiveness within a segment

Loyalty programmes have grown in popularity in recent times and as a result there is a proliferation of them in the market today (Dorotic, Bijmolt, & Verhoef, 2012). This proliferation of loyalty programmes reflects a changing market environment that is increasingly characterized by intense competition, more demanding and knowledgeable consumers, and a development toward relationship marketing and customer relationship management in marketing thinking and practice (Liu & Yang, 2009). "A consequence of this enthusiasm toward loyalty programmes is the increasing competition among rival programmes, especially in sectors such as airlines, financial services, retail, hotels, and gaming. In the credit card industry, for example, half of the general purpose credit cards offer a reward programme." (Liu & Yang, 2009, p. 93)

A consequence of this market saturation is that it has a negative impact on the effectiveness of loyalty programmes, especially in industries where the goods or services are homogenous, such as banking (Liu & Yang, 2009). However the level of market saturation can differ significantly between income segments. Most advertising and marketing around loyalty programme seem to target higher income customers and therefore creates the perception of greater market saturation in this segment.

This variance in market saturation between high and low income customers could result in varying degrees of effectiveness. Liu & Yang (2009) stated that differentiation in programme offerings would mitigate the effects of market saturation which suggest variations in design across market segments.

2.12 Reward Distances and Step-Size Ambiguity

The proliferation of retail loyalty programmes has been one of the marked features of retail marketing, encompassing different types of schemes including point cards, coalition cards and store cards (Hutchinson, Donnell, Gilmore, & Reid, 2015). The most popular and arguably the most effective types of loyalty programmes is a points based loyalty programme (Zhang & Breugelmans, 2012). With points based loyalty programme companies can vary their design in terms of value of the points and how these points are earned. These two variables were referred to as rewards distance and step-size (Bagchi & Li, 2010).

In a points based loyalty programme a customer earns points at a certain rate, perhaps one point for every ten rand spent, this is referred to as step-size. Then the points are used as currency to redeem for actual rewards, for example a hundred points for a kit bag, this is referred to a reward distance. Bagchi & Li (2010) stated that how a programme is designed with regards to these two components can influence customers' motivation toward attaining a desired reward. Optimal designs would create the perception of earning more value than customers actually received, however according to Bagchi & Li (2010) poor designs will have the inverse effects.

Bagchi & Li (2010) stated that companies would generally choose one of two designs, either a small rewards distance and small step-size or a large rewards distance and large step-size. The first would give the customer the illusion that the reward was close but their slow step-size meant it took longer to get there. This was more effective for customers with a high rate of earning and preferred high value rewards. According to Bagchi & Li (2010) the latter design gave the illusion of earning faster and making progress but the reward was actually further away. This design was more effective with customers that earned at a low rate and preferred low value rewards.

Based on this the effectiveness of the design of the elements of reward distance and step-size is influenced by the customers rate of earning and the value of reward the customer is aspiring toward. (Stourm, Bradlow, & Fader, 2015) stated that the there is a difference between both the earn rates and redemption values of high and low income customers. Therefore depending on the design of a loyalty programme and the structure of reward distance and step-size, its effectiveness may vary between high and low income segments indicating the existence of a relationship.

19

2.13 Programme Advertising

Like any product, the way a loyalty programme is received by customers will be influenced by how it is advertised. In most cases the advertising is consistent and the marketing message unchanged between segments. This consistent advertising may result in varying levels of effectiveness across income segments if the advertising is received differently across income segments.

Jin, Liu, & Knight (2014) stated that there are two types of advertising, rational and emotional. Their finding was that certain behavioural attributes of the customer determined which of these types of advertising was more effective. According to Jin, Liu, & Knight (2014), customer that prefer financial benefits respond better to rational advertising and customers that prefer social benefits respond better to emotional advertising. The correlation between financial and social benefits and level of income has been covered already, thus it can be inferred that consistent advertising in loyalty programme, as is current practice, will result in variances in how it is received across income segments.

2.14 Summary and formulation of propositions

The literature review revealed that customers react differently to the design of a loyalty programme based on their specific behavioural traits. Current research provides methods of segmenting of customers based on the actual behaviour of the customer once they are enrolled into the programme. Business would benefit from the establishment of an upfront method of customer segmentation as this could be used to vary the design of their loyalty programme based on the behavioural traits of the segment. This way business would be able to optimise the effectiveness of their loyalty programme and in some instances even reduce costs.

The segmentation approach that was proposed is to use income level as a means to segment customers at the enrolment stage. This method was chosen for the following reasons:

- Income is an objective measure.
- Income is easily obtained.
- Income is a commonly used in other customer segmentation approaches.
- This will can be executed easily by business

20

The gap is the current literature is that no upfront segmentation has been tested to ascertain if a statically significant relationship exists between an objective variable and the effectiveness of the loyalty programme. Income was introduced as a possible objective variable that could be used for this objective. The following themes were covered from current literature that suggests the possibility of a relationship between income level and the effectiveness of a loyalty programme:

- Financial versus Social benefits – two broad categories of benefit that vary in appeal to different income segments.

- Reciprocity and Gratitude – feelings of gratitude in customers cause them to reciprocate with desired behaviour; however the level of gratitude varies with income segments.

- Programme versus Company Loyalty – two different types of loyalty that result in different behaviours which could vary between income segments.

- Timing of Rewards – when and how a reward is redeemed varies across income segments and has an impact on perceived value received from the loyalty programme.

- Desire for the Loyalty Programme – the level of desire at enrolment influences the effectiveness of the loyalty programme. Desire may be different across income segments.

- Relevance of the Benefit Received – different income segments may perceive the benefits received with varying levels of relevance.

- Familiarity with the programme – may vary with income and influences effectiveness

- Churners versus non-churners – retention objectives of a loyalty programme may be more effective for churners than non-churners.

- Competitiveness within a segment – effectiveness of a loyalty programme is influenced by market saturation. Competitiveness may vary across income segments.

- Reward Distances and Step-Size Ambiguity – two important aspects in the design of a loyalty programme and are received differently across income segments.

- Programme Advertising – two broad categories of advertising with each being received differently by high and low income consumers.

Based on these themes it is proposed that level of income offers a means for segmentation and that a relationship should exist between income level and the effectiveness of a loyalty programme.

21

3 Chapter 3: Research Hypothesis

3.1 Introduction

Following up from the review of the relevant literature and the arguments that were constructed, this chapter serves to define the research questions and detail the relevant hypotheses for investigation.

3.2 Research question one

Research question one deals with the effectiveness of the loyalty programme on influencing purchase behaviour in relation to income levels. The effectiveness will be tested between income segments to ascertain if a statistically significant relationship exists between the two variables. Research question one is split into two parts, namely:

- Purchase frequency
- Purchase amount

Both of these components will be tested separately.

3.3 Research question two

Research question two deals with the effectiveness of the loyalty programme in influencing customer retention in relation to income levels. The level of effectiveness will be tested between income segments to ascertain if a statistically significant relationship exists between the two variables.

3.4 Research Objective

This study intends to validate the applicability of the use of level of income as a method of segmentation for loyalty programme customers by focussing on the desired objectives of increased purchase behaviour and customer retention.

The research questions have been split into three hypotheses, which if proved will validate the use of income level as a method of segmentation.

3.4.1 Hypothesis 1A:

Income is a significant predictor of the effectiveness, related to average purchase value, of a loyalty programme.

H_0: $\beta_1 = 0$;   Income is not a significant predictor of effectiveness.

H_1: $\beta_1 \neq 0$;   Income is a significant predictor of effectiveness.

3.4.2 Hypothesis 1B

Income is a significant predictor of the effectiveness, related to average number of purchases, of a loyalty programme.

H_0: $\beta_1 = 0$;   Income is not a significant predictor of effectiveness.

H_1: $\beta_1 \neq 0$;   Income is a significant predictor of effectiveness.

3.4.3 Hypothesis 2

Hypothesis 2: Income is a significant predictor of the effectiveness, related to the retention rate, of a loyalty programme.

23

H_0: $\beta_1 = 0$;   Income is not a significant predictor of effectiveness.

H_1: $\beta_1 \neq 0$;   Income is a significant predictor of effectiveness.

4 Chapter 4: Research Methodology

4.1 Introduction

The previous chapter served to summarize the purpose of the research, detailing the research question and laying out the hypotheses. The aim of this chapter is to outline the research methodology of the study. It forms the basis of the approach used to answer the research questions, beginning with the research design and outlining the methodology structure. This is followed by population and sampling, variables and its constructs, the unit of analysis, data gathering process, the data analysis approach, principles of reliability and validity and concludes with the research limitations.

4.2 Research Design

The research method adopted for this study is descriptive, as it aims to identify whether or not there exists a relationship between the effectiveness of a loyalty programme and income. Given the nature and structure of the research question, a quantitative statistical method of analysis was employed (Saunders & Lewis, 2012). This approach is appropriate since the data extracted consists of numbers, representing either measurements or counts (Triola & Franklin, 1994). According to Jupp (2006) analysing data using quantitative methods involves a process of collecting the data in a numerical format and then performing statistical tests on it. These tests include descriptive statics, regression analysis and correlation analysis. Inferential statistics is then used to draw conclusions about the factors being examined. In this study the data extracted is in a numeric format and the results from the tests completed were used to measure the effectiveness of a loyalty programme, in terms of the purchase behaviour and retention rate, at different levels of income.

The basic flow of the research design is outlines in the figure below:
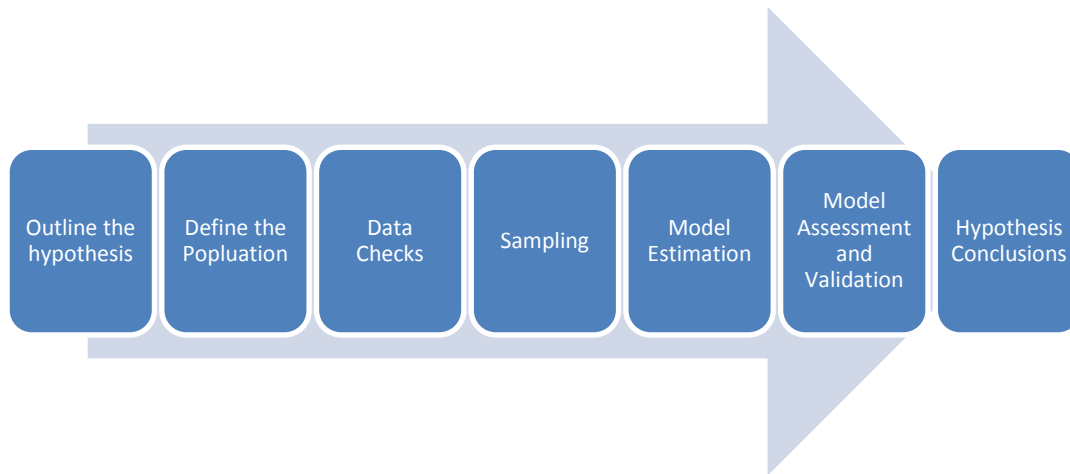
Figure 3 : Research Methodology Process Flow

The data utilized in the study was extracted from a large retail bank in South Africa. According to Saunders & Lewis (2012) this is a form of secondary data collection as the data was already collected by other people for another purpose. The required variables and fields where extracted from this source file and used for statistical analysis.

Prior to testing the hypothesis at different income levels, an overall test of effectiveness was performed. This was done using a two sample t-test for equal means with a pooled variance. According to Seo & Srivastava (2000) this test can be employed to determine if two means (of a population) are equal. In this study, the average purchase value is used to test the hypothesis of whether or not there is a difference in the effectiveness between the two groups (loyalty verses non-loyalty). A second form of analysis determining overall significance was then employed using a binomial test for proportions. A binomial test can be employed on large samples to perform hypothesis tests for the equality of two binomial proportions (Massey, 1957). In this study, the average closure rate *(1-retention rate)* is used to test the theory.

The first hypothesis examining if there is a relationship between effectiveness and income, in terms of purchase behaviour, is tested using linear regression analysis. According to Carlson & Thorne (1997), regression analysis is a widely used tool to measure the relationship between a dependent variable and independent or explanatory variables. Regression analysis was performed using a Statistical Analysis Software (SAS) procedure for this study. When using regression analysis, the independent variable is used to measure the variation in the dependent variable (Carlson & Thorne, 1997). When testing this hypothesis, effectiveness enters the model as a dependent variable and is measured in two ways; the average number of purchases or the

26

average purchase value (spend). The independent variable is income and is the account holder's level of annual income. The variables are regressed against each other and the results analysed to determine whether firstly; the model was significant and secondly whether or not is possible to reject the null hypothesis.

The second research question examines the relationship between a different measure of effectiveness, namely the rate of customer retention, and income. The higher the retention rate the more effective is the loyalty programme in that income segment. This hypothesis was examined using a logistic regression. According to Hosmer & Lemeshow (2000) the goal of logistic regression is the same as any other statistical model building technique; to find the most parsimonious and best fitting relationship between an outcome variable and predictor or independent variable(s). However the distinct feature of a logistic regression is that the outcome, or the predicted variable, is binary or dichotomous (Hosmer & Lemeshow, 2000). For the purposes of testing the hypothesis, a logistic regression, where the independent variable is income and the dependent variable is closure rate (1-retention rate) was employed. The model was then tested for significance and the results assessed.

4.3   Population and Sampling

The chosen statistical study is in the form of analytics. Carlson & Thorne (1997) iterate that an analytic study can be used to measure a group of items. The results obtained are then used to infer attributes outside of the group studied. According to Triola & Franklin (1994), a population is the set of data that contains all elements and a sample is subset drawn from this population. The sample of accounts in this study was used to model the relationship between the effectiveness of a loyalty programme and income. This measurement is then used to infer the type, significance and strength of the relationship in question.

The population data was drawn from a large retail bank in South Africa, after permission to access the customer's transactional data was obtained. The population consists of all credit card accounts and their relevant information observed between August 2013 and July 2015. According to Saunders & Lewis (2012), relevance and suitability of the data is an important criterion, as this provides the information required to build the relevant models and test the hypothesis. For the purpose of this study, the most recent set of complete records where utilized as this controls for the data needing to be relevant and suitable when assessing results and drawing conclusions.

27

Saunders & Lewis (2012) define the sampling frame as a complete list of all observations in the population. When it is possible to obtain this comprehensive list, the sample should be selected using a sampling technique called probability sampling. In this study, since the comprehensive list is obtained and available, probability sampling was employed. The technique of Simple Random Sampling (SRS) which is a form of probability sampling (Saunders & Lewis, 2012) was used to draw a subset of the population. The method of SRS ensures that each observation in the population has an equal chance of being selected and no one observation influences the selection of other observations (Carlson & Thorne, 1997). This technique creates ideal samples and is appropriate data sources for most statistical tests and procedures.

A sample size of 50000 accounts was selected during the development stages and the remaining accounts were used for validating the models. This represented a 70:30 % split between the development and the validation sets. This is an appropriate number, since it is a high proportion of the population but is also at a fair level of computational intensity (Saunders & Lewis, 2012). The conclusions resulting from the analysis on the sample should apply to the entire population and the estimates should on average be very close (Carlson & Thorne, 1997). However one needs to be cognizant of the potential bias and random error, discussed in Section 4.9.

4.4   Variable and Constructs

According to Saunders & Lewis (2012) when data is numerical or can be measured, it is referred to as quantitative data. The authors assert that this data can be spilt into two main types; categorical and numerical data. The data extracted for the purpose of this study is numerical.

Discrete data is a form of numerical data in which the values are measured as quantities in discrete units. This results in a discrete variable being able to take on a finite number of values (Saunders & Lewis, 2012). In contrast, the authors explain, continuous data is measured in quantities that may take on any value depending on the level of accuracy in which the variable is measured. For the purpose of this study, there are both continuous and discrete variables. This is outlined in the figure below and explained thereafter.

28
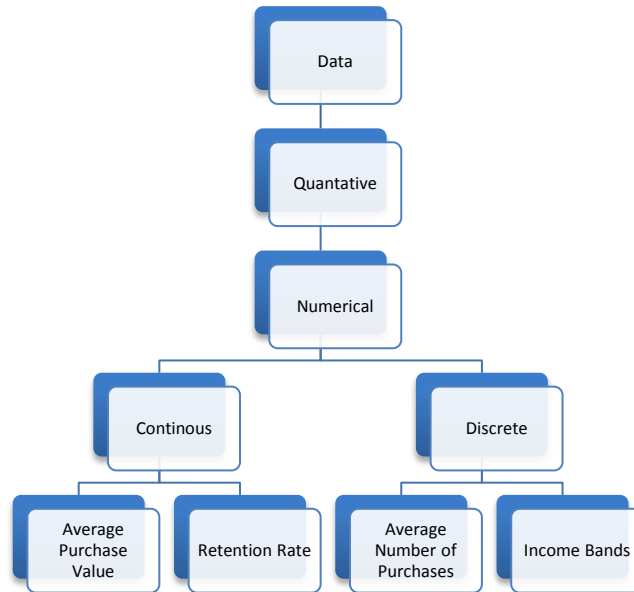
Figure 4: Quantitative Data

The variables used to build the model where extracted from the source data and then transformed as appropriate for modelling. A list of the variable pertinent to the analysis is outlined below:

4.4.1 Construct: Income Bands

Income is as a continuous variable and is measured as the gross annual income each credit card holder receives in rand value. For the purposes of this analysis the population is segmented into thirty income bands, based on the individual's income level relative to the population income level. According to Triola & Franklin (1994), the raw income field is a form of continuous quantitative data as a result of infinitely many possible values. However once the variable is grouped into the bands, this is transformed to a discrete variable. The income band enters the model as the independent variable in the three regressions.

4.4.2 Construct: Effectiveness

Effectiveness is the dependent variable measured in terms of three metrics. The construction of the effectiveness covariate (variable in the model), was created using three different measures.

29

The main role of a loyalty programmes is to entice customers to use the product and to incentivize retention. Utilization and retention are therefore important factors when measuring effectiveness and are used as proxies during the research (Dorotic, Bijmolt, & Verhoef, 2012).

4.4.2.1 Value of purchases

The first measure of effectiveness is the value of purchases. This value is estimated based on the average amount spent on the credit card facility for a month. The scale used to measure the average purchase value is a quantitative ratio scale. Ratio scales indicate both rank and distance from an arbitrary zero with the ratio of the two measures having a meaning (Carlson & Thorne, 1997) .Using this measure, the higher the purchase value of the accounts on the programme relative to the accounts that are not, the more effective is the programme.

The measure of effectiveness is given by the formula below:

$$Effectiveness = \frac{Average\ Spend\ Value(Loyalty = Yes)}{Average\ Spend\ Value\ (Loyalty = No)} - 1$$

This was used to test hypothesis 1a, where effectiveness estimated from the equation above was then regressed against the income band.

4.4.2.2 Volume of purchases

The second measure of effectiveness relates to volume of purchases. The volume is estimated based on the average number of purchases on the credit card facility for a month. Using this measure, the higher the number of purchases on accounts in the programme relative to the accounts that are not, the more effective is the programme. The scale used to measure the average spend value is a quantitative ratio scale. Ratio scales indicate both rank and distance from an arbitrary zero with the ratio of the two measures having a meaning (Carlson & Thorne, 1997).

The effectiveness estimated from the equation above was then regressed against the different bands.

$$Effectiveness = \frac{Average\ number\ of\ purchases(Loyalty = Yes)}{Average\ number\ of\ purchases(Loyalty = No)} - 1$$

This was used to test hypothesis 1b, where effectiveness estimated from the equation above was then regressed against the income band.

4.4.3   Construct: Retention

Retention is measured as a compliment of the average closure rate on accounts per month. Using this measure, if the average closure rate of accounts on the programme is relatively less than the average rate of closure on accounts that are not in the programme, the loyalty programme is most effective in that income segment. The scale used to measure the average spend value is a quantitative ratio scale. Ratio scales indicate both rank and distance from an arbitrary zero with the ratio of the two measures having a meaning (Carlson & Thorne, 1997).

The retention rate is calculated as

$$rr = (1 - closure\ rate)$$

It is estimated based on the average number of closed card accounts within the twenty four month observation period.

$$Effectiveness = \frac{Number\ of\ Closed\ Accounts(Loyalty = Yes)}{Number\ of\ Closed\ Accounts(Loyalty = No)} - 1$$

This was used to test hypothesis 2, where effectiveness estimated from the equation above was then regressed against the income band.

.

31

4.5 Unit of analysis

The unit of analysis is the part of the research project that defines its most basic element which the researcher is able to generalize ((Lewis-Beck & Liao, 2013)**.** The unit of analysis selected as appropriate for this study was all credit card account holders at a large retail bank in South Africa. This population was assessed in terms of purchase behaviour and retention rates over a twenty four month outcome period, given that some accounts were on a loyalty and others not. After all the assumptions of the regression are accounted for, the results from the analysis were used to measure the relationship between the effectiveness of the loyalty programme at different income levels.

4.6 Data gathering process

The data collection process began after the researcher had obtained the necessary approval from the Head of Consumer Card at the large retail bank.

The study used secondary data, which is defined by Saunders & Lewis (2012) as the data used for a research project that was originally created for some other purposes. The bank currently collects and uses this data for internal modelling and other corporate purposes. This is a form of "re-collecting data", as it involves sourcing the data that already exists and providing a context and background to conduct new research (Page & Meyer, 2000). The main use of this data is to look for patterns and themes, rather than for anything specific (Page & Meyer, 2000).

Due to the nature and structure of the research question, this was therefore deemed appropriate, since the data required for analysis was readily available. The information was gathered on a nominal scale and then segmented according to the definitions of each construct. The population which according to Page & Meyer (2000) is the entire set of records as per the defined rules is measured as all active credit card accounts. In order to understand the effectiveness of the programme at different income segments, information about the following variables for the twenty four months was also extracted.

- Observation Month
- Monthly Spend Value
- Volume/Number of Purchases

32

- Account Identification

- Current Account Status

- Initial Account Status

- Current Credit Limit

- Initial Credit Limit

- Account Closed Date

- Age

- Annual Gross Income

- Loyalty

- Account Age

These variables are used to create heterogeneous segments, before performing statistical analysis.

Relevance and suitability of the data is an important criterion, as this provides the information required to build the relevant models and test the hypothesis (Saunders & Lewis, 2012) . For the purpose of this study, the most recent set of records where utilized to account for this.

4.7   Analysis approach

Statistical Analytics Software (SAS) Enterprise Guide and Microsoft Excel were the tools used for all computational and visual analysis. The analysis approach is summarized in the table below.

4.7.1   Assumptions

The following assumptions were made when completing the analysis:

- The Income variable remains constant over the twenty four month observation period
- An account does not move loyalty states i.e. if the account was on the programme at the commencement of the observation period, it remains in that state.

33

4.7.2 Descriptive statistics

Descriptive statistics are employed mathematical techniques that can be used to transform, summarize and process numerical data into information (Carlson & Thorne, 1997). According to Schwandt (2007) the main elements of central tendency are the; mean, median, mode and range. This is used to firstly understand the composition of the data before any statistical tests are employed. Measures of the central tendencies are used to estimate the "typical" value or observation in a dataset (Carlson & Thorne, 1997). These measures give a good understanding of how the data is constructed and assists with verifying the hypothesis. The variability measured by the standard deviation is an important concept and it is required to verify the reliability of the results obtained (Carlson & Thorne, 1997). The effectiveness of the loyalty programme measured in terms of purchase behaviour and retention is assessed using measures of central tendency and variation.

4.7.3 Outlier Analysis

An outlier analysis was performed, the aim of which was to identify extreme values in the data that could interfere with statistical results. Outliers are values that are very different from other values within a data set and if these extreme values are not removed from the initial data set or remedied in some other way, then statistical results obtained may be misrepresented (Weiers, 2010).

A box-and-whiskers plot was employed as a univariate method for detecting outliers. This represents the standardized (z) individual measurement items and is useful for estimating the symmetry of a distribution and to detect any inconsistent values and outliers (Rodriguez, 2007).

4.7.4 Two sample t-test for equal means

According to Dorotic, Bijmolt, & Verhoef (2012) a loyalty programme is effective in most retail products. This means it has a profound effect on the amount that card holders purchase for, the number of purchases a card holder performs, and the ability of bank to retain the card holder. To measure the overall effectiveness of loyalty programmes, a two sample t-test of equality is employed to test the difference between purchase behaviour of accounts on a loyalty program

verses those that are not. Saunders & Lewis (2012) explain the employed t-tests ability to analyse differences when a numerical variable is split into two groups.

The sample is split into two populations; accounts with a loyalty programme and those not on the loyalty programme. A two sample t-test for equality of means is employed. This test can be employed to perform a large sample hypothesis test for the equality of two means (Seo & Srivastava, 2000).

The effectiveness is measured by the difference in the average value of purchases on accounts on the programme versus those that are not. If there is no difference in the average purchase value, then the effectiveness is the same, keeping all other factors constant.

For this t-test, the hypothesis is laid out below:

$$Ho: Average\ spend(Loyalty = Yes) = Average\ Spend\ (Loyalty = No)$$

$$Ha: Average\ spend(Loyalty = Yes) <> Average\ Spend(Loyalty = No)$$

To test whether this difference is significant a t-test is then performed.

$$t = \frac{\overline{X_1} - \overline{X_2}}{s_{x_1 x_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_{x_1 x_2} = \sqrt{\frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}}$$

,

Since this is a test for equality a two-tailed test is employed (Massey, 1957) .The result of this test is to reject the null if the t-statistic calculated is outside the critical region at a 0.05 significance level or fail to reject the null if the t value is within the critical region. An important assumption of this t-test is that data needs to be normally distributed.

4.7.4.1 Q-Q Plot (Normality of Residuals Test)

The normal Quantile-Quantile plot is a plot of the theoretical quantiles from the normal distribution against the sample quantiles from the data under consideration. If the theoretical quantiles have a linear relationship with the sample quantiles, then the sample aligns with the

normal distribution. Furthermore, if the plot aligns with a 45^0 line through the origin, then the sample has the same mean and variance as the assumed normal distribution (Filliben, 1975)

.

4.7.5 Linear Regression

The technique of linear regression according to Oakshott (2006) is the attempt to estimate a linear equation that defines the relationship between a dependent and an independent variable. The results of a linear regression are twofold; the predicted values for the dependent variable as a function of the independent variable and the marginal change in the dependent from a one unit change in the independent variable (Carlson & Thorne, 1997).

Linear models take the following form

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

, where the slope coefficient β_1 is the change in Y for a one unit change in X. β_0 is the value of Y when X_1 is zero. ε is the error term (Carlson & Thorne, 1997).

The dependent variable in the current study is effectiveness while the independent variable is income. This is then used to understand and interpret the relationship between the two variables to reach conclusive results. According to Chatterjee & Hadi (2012) a regression model containing one predicator variable is called a simple regression equation, whereas if a model has more than one predictor variable it is called a multiple regression equation. Similarly Chatterjee & Hadi (2012) explained if the model has only one response variable it is a univariate regression analysis, whereas if there is more than one response variable it is called a multivariate regression analysis. Since there is only one predictor and one response variable in this model, it is explained to be a simple univariate regression analysis.

4.7.5.1 Caveats of Linear Regression

In order to use a Linear Regression the following assumptions need to be met (Segrin, 2010):

- The relationship between Y and X is linear in the parameters $\beta_0 \ and \ \beta_1$.

A visual inspection of the linearity can be assessed using a scatter plot (Oakshott, 2006). The relationship between two variables that are numeric is also best shown using the scatter plot or graph (Saunders & Lewis, 2012).

According to Chatterjee & Hadi (2012), a hypothesized model can be refuted by performing a regression analysis on the data collected. It is important to note that although the form of the model needs to be linear, the functions can be linear or nonlinear (Chatterjee & Hadi, 2012). This is explained by the two different equations below:

$$\text{Linear Function: } Y = \beta_0 + \beta_1 + \varepsilon$$

$$\text{Non-Linear Function: } Y = \beta_0 + e^{\beta_1 X_1} + \varepsilon$$

The term linear does not describe the relationship between Y and X_1, it confirms that the parameters $\beta_0 \text{ and } \beta_1$ enter the equation linearly (Chatterjee & Hadi, 2012). Hence the following model is linear:

$$Y = \beta_0 + \beta_1 ln X_1 + \varepsilon$$

This according to Chatterjee & Hadi (2012) is a transformation, which allows the variable to be re-expressed to form a linear relationship. Thereafter normal regression tests apply.

- The expected value of the error term is zero
- There is no homoscedasticity present. This, according to Thorne means that the variance of the error term needs to be constant
- Statistical independence of the errors
- There is no autocorrelation present
- The error term is normally distributed

4.7.5.2 Standard Error of Estimate

When defining the regression equation, it is referred to as the line that fits the data points the 'best" (Triola & Franklin, 1994). This property, according to the authors is called 'least squares'. This means that the regression line that fits the data best has the lowest sum of squares of the vertical deviation of the sample points from the regression line (Triola & Franklin, 1994)

This sum is referred to as "SSError" and is useful in judging how well the actual regression line fits the data. The total variation in the dependent variable is sum of the explained variation (from the regression line) and the unexplained variation (SSError) (Triola & Franklin, 1994).

4.7.5.3 Coefficient of Determination

The coefficient of determination, sometimes called the R-Squared can be used as metric to describe how well the regression equation fits the actual data (Oakshott, 2006). The coefficient of determination is a measure of how much of the variation in a dataset is explained by a regression model. It is stated as (Cameron, Windmeijer, Gramajo, Cane, & Khosla, 1997):

$$R^2 = 1 - \frac{\sum_i (f_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}. = 1 - \frac{Unexplained\ Variation}{Total\ Variation}$$

A high value of R^2 indicates a well-fitting regression line, whereas a low value of R^2 near 0 indicates a regression line that does not fit the data well (Triola & Franklin, 1994).

4.7.5.4   Correlation Analysis

The correlation coefficient is used to measure the direction and strength of a linear relationship between two variables ((Triola & Franklin, 1994). Since this is a simple linear regression, the correlation coefficient (r) is the square root of the coeffiencient of determination

$$r = \sqrt{R^2}$$

4.7.6   Hypothesis Testing

The hypothesis tests were conducted in the following manner:

4.7.6.1   Formulating the null and alternative hypothesis

The hypothesis is defined as per Chapter 3

4.7.6.2   Choosing a level of significance

The chosen level of significance was 0.05.

4.7.6.3   Computing the relevant test statistic

Hypotheses testing for the slope coefficient is an important decision making tool. A Wald test is usually employed when testing for significant slope coefficients (Carlson & Thorne, 1997).

38

Wald Test (Significance of Parameters Test)

For a given parameter θ in a regression model, for a covariate under consideration, the Wald statistic compares the maximum likelihood estimate of the parameter value $\hat{\theta}$ to a proposed value θ_0. When testing the significance of a parameter in a model, $\theta_0 = 0$. The test statistic, which is assumed to have a chi-squared distribution under the null hypothesis, is as follows (Harrell, 2001).

$$W = \frac{(\hat{\theta} - \theta_0)^2}{Var(\hat{\theta})}$$

According to Triola & Franklin (1994) the value computed from the formula is compared to the critical value which is obtained from statistical table depending on the distribution employed.

If the computed t-statistic is less that the critical t-value, the null hypothesis may be rejected. Since we are testing for equality, a two tailed test needs to be used (Carlson & Thorne, 1997).

4.7.6.4   Has the null hypothesis being accepted or rejected, and drawing conclusions from the results.

4.7.7   Binomial Test for Proportions

Prior to testing the hypothesis of whether a relationship between effectiveness, in terms of retention rates, and income levels exists, it is imperative to test the effectiveness, using retention rates, of a loyalty programme overall. The sample is split into two populations; accounts with a loyalty programme and those not on the loyalty programme. A binomial test for proportions is then used to assess whether a significant difference exists between the two populations. According to Massey (1957) this test can be employed to perform a large sample hypothesis test for the equality of two binomial proportions.

The hypothesis for this two sample test is shown below using the retention rate as a measure of effectiveness.

$$H_0: Average\ Retention\ Rate\ (Loyalty\ =\ Yes)\ =\ Average\ Retention\ Rate\ (Loyalty\ =\ Yes)$$

$$H_a: Average\ Retention\ Rate\ (Loyalty\ =\ Yes)\ <>\ Average\ Retention\ Rate\ (Loyalty\ =\ Yes)$$

To test whether this difference is significant a t-test is then performed.

$$z = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

, where $\hat{p} = \frac{n_1\widehat{p_1} + n_2\widehat{p_2}}{n_1+n_2}$

$$= \frac{X_1 + X_2}{n_1 + n_2}$$

According to Carlson & Thorne (1997) since this is a test for equality a two-tailed test is employed. The result of this test is to reject the null if the z score is outside the critical region at a 0.05 significance level.

4.7.8   Logistic Regression

Hypothesis 2 will be tested using a Logistic Regression. Hosmer & Lemeshow (2000) assert that regression analysis has become an integral part of data analytics when concerned with describing the relationship between a dependent and an explanatory variable. The goal of Logistic Regression is the same as any other statistical model building technique; to find the most parsimonious and best fitting relationship between an outcome variable and predictor or independent variable (Hosmer & Lemeshow, 2000).

The difference between logistic and linear regression is that the in the logistic model dependent variable is binary or dichotomous, and the choice of parametric model and assumptions are different (Hosmer & Lemeshow, 2000). The authors assert that after these are taken into account, the methods that are employed in a logistic regression follow the general principles employed in linear regression.

Logistic regression is a modelling technique that forms part of the class of generalised linear regression. According to Hosmer & Lemeshow (2000) it can be applied to modelling situations where the target variable is binary, to produce a statistical model for the probability that the variable takes on one of the two possible values. When building the model to test the hypotheses, the target variable was the probability of an account closing within the twenty four months.

According to Hosmer & Lemeshow (2000) logistic regression assumes that a linear relationship exists between logit transformation of the probability $P_x(E)$ of and the independent variables.

$$\ln\left[\frac{P_x(E)}{1 - P_x(E)}\right] = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon_i$$

Where ε_i is the error term, β_j are the parameter estimates of the p independent variables X_i.

According to Hosmer & Lemeshow, 2000, the parameters of a logistic model are fitted by maximum likelihood methods. The estimates do not have a closed form solution. The Logistic procedure offers the Newton-Raphson method and the Fisher-scoring method to estimate parameter. The Fisher-scoring method was used (Hosmer & Lemeshow, 2000)

4.7.8.1   Hosmer-Lemeshow

The Hosmer-Lemeshow (HL) test assesses the goodness of fit for logistic regression models. It is a Chi-Squared test that measures how well the predicted event-rates align with the observed rates (Hosmer & Lemeshow, 2000)

To conduct the test, the population is split into m decile groups, so that O_i, E_i, n_i and π_i denote the observed events, expected events, number of observations and predicted event rate for decile group i. The test statistic, which asymptotically follows a χ^2_{m-1} distribution, is as follows (Hosmer & Lemeshow, 2000)

$$H = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{n_i \pi_i (1 - \pi_i)}$$

The Logistic procedure outputs the HL statistic when the Lackfit option is selected.

4.7.8.2   Adjusted Generalised Coefficient of Determination
The generalised coefficient of determination is a measure of how well data points fit a statistical model. It is an extension of the coefficient of determination that is common to the assessment of the fitness of a linear model. It is defined as follows (Cox & Snell, 1989)

$$R^2 = \frac{1 - \left(\frac{L(0)}{L(\hat{\theta})}\right)^{\frac{2}{n}}}{1 - \left(L(0)\right)^{\frac{2}{n}}}$$

41

4.7.8.3 Wald Statistic

For a given parameter θ, for a covariate under consideration, the Wald statistic compares the maximum likelihood estimate of the parameter value $\hat{\theta}$ to a proposed value θ_0. When testing the significance of a parameter in a model, $\theta_0 = 0$. The test statistic, which is assumed to have a chi-squared distribution under the null hypothesis, is as follows (Harrell, 2001):

$$W = \frac{\left(\hat{\theta} - \theta_0\right)^2}{Var\left(\hat{\theta}\right)}$$

4.7.8.4 Parameter Significance

If $\hat{\beta}_k$ is the parameter estimate associated with the covariate X_k and σ_k is the standard deviation associated with the covariate, then we can adjust (standardise) the parameter estimate as follows.

$$\hat{\alpha}_k = \frac{\hat{\beta}_k}{\sigma_k}$$

$\hat{\alpha}_k$ is now insensitive to large variance in covariates i.e. it is not influenced by the size of difference in X_k within the population. The contribution of variable X_k to the model prediction can be measured from $\hat{\alpha}_k$s as follows.

$$contrib_k = \frac{\hat{\alpha}_k}{\sum_k \hat{\alpha}_k}$$

4.8 Validity

The following checks were done to assess the validity and completeness of the data used when testing the hypothesis:

- Line graphs of the number of accounts, average spend and average number of purchases were plotted to ensure that there were no sudden drops in records.

- The data was checked for missing records and duplications. The proportion of accounts that had no account or customer information was checked. There were reported accounts where the income field is blank. These accounts were excluded.

- The data set was split into two parts, a development and a validation set. This validation data is used to assess whether or not the model fits the data that it was not build on well enough.

4.9   Limitations of the Research

4.9.1   Data

The data utilized in the study was extracted from a large retail bank in South Africa. According to Saunders & Lewis (2012) this is a form of secondary data collection as the data was already collected by other people for another purpose. The required variables and fields where extracted from this source file and used for statistical analysis. One of the main drawbacks of using secondary data is the fact that the data is not collected by the researcher. It is therefore a challenge to understand the manner in which the data was collected and how this might influence the results (Saunders & Lewis, 2012). A one on one discussion with the relevant personnel who manages the data sources assisted with understanding the data collection process to draw meaningful conclusion from the tests carried out.

Heterogeneous groups were created on the data extracted. This controlled for the variables within the data. However, variables outside of the scope of this research may have an endogenous effect on effectiveness. These variables include the account holders age, age, age of account, number of other product holdings and risk rating.

A further data limitation results from using a specified two year observation period, due to lack of availability of the data required. Therefore factors that are autonomous to the study, including macroeconomic, industry and market trends, are not taken into account when drawing conclusions. However, the two years chosen is a fair representation of a business cycle and should inherently consider the factors outlined.

4.9.2 Sampling

Triola & Franklin (1994) believe that determining the appropriate sample size is important as large data sets are computational intensive but too small data sets may result in inaccurate predictions. A challenge when extracting samples arise from potential biases which may be created if the sample is not representative of the population. It is essential to monitor whether the sample data reflects the population as accurately as possible.

4.9.3 Business Strategy

Another limitation is the fact that the researcher is unaware of the internal business and marketing strategies of the retail bank. This strategy may have a direct influence the effectiveness of the loyalty programmes. Specific marketing strategies aimed at incentivizing purchase behaviour and spend are outside the scope of the analysis. This may lead to bias results.

Further limitations arise from the fact that the analysis was performed using data from a single company and on a single loyalty program. Caution must be applied, as the results and recommendations may not necessarily be generalized as indicative of all retail industries or banks, but could definitely be used for companies and industries which are similar in nature. Attributes to consider when inferring the results, include the size of the organization and the industry in which it operates in.

5 Chapter 5: Results

5.1 Introduction

Chapter four outlined the research methodology employed during the study. The aim of this chapter is to outline the results obtained from carrying out the analysis stipulated. Descriptive statistical analysis was performed on the data collected, detailing the average number of purchases, the average purchase value and the closure rate over time, on the two populations. This formed part of the completeness and validity checks to ensure the data is of good quality. This included removal of missing data, outlier analysis and sampling. Thereafter the results of the difference in overall effectiveness of a loyalty programme, relating to purchase behaviour is assessed. The results from the linear regression, testing Hypothesis 1a and 1b are then

presented and conclusions are drawn about the relationship between effectiveness, in terms of purchase behaviour, and spend.

Subsequently, the results of the difference in the overall effectiveness of the loyalty programme, relating to retention were examined. Finally, the chapter concludes with the logistic regression results, testing Hypothesis 2 on the relationship between retention and effectiveness.

5.1.1 Research Question One: Effect of Programme on Spending Behaviour

Research question one deals with the effectiveness of the loyalty programme on influencing purchase behaviour in relation to income levels. The effectiveness will be tested between income segments to ascertain if a statistically significant relationship exists between the two variables. Research question one is split into two parts, namely:

- Purchase frequency
- Purchase amount

These components are tested separately.

Hypothesis 1a: Income is a significant predictor of the effectiveness, related to average purchase value, of a loyalty programme. If this hypothesis holds, we expect to see income as a strong predictor of the effectiveness of the loyalty programme on changing purchase behaviour (spend amount). Therefore, a test for this hypothesis can be formulated in terms of a Wald test for parameter significance in linear regression:

$$H_0: \quad \beta_1 \ = \ 0 \quad ; \quad \text{Income is not a significant predictor of effectiveness.}$$

$$H_1: \quad \beta_1 \ \neq \ 0 \quad ; \quad \text{Income is a significant predictor of effectiveness.}$$

Hypothesis 1b: Income is a significant predictor of the effectiveness, related to average number of purchases, of a loyalty programme. If this hypothesis holds, we expect to see income as a strong predictor of the effectiveness of the loyalty programme on changing purchase behaviour (spend frequency). Therefore, a test this hypothesis can be formulated in terms of a Wald test for parameter significance in linear regression:

45

H_0: $\beta_1 = 0$;   Income is not a significant predictor of effectiveness.

H_1: $\beta_1 \neq 0$;   Income is a significant predictor of effectiveness.

5.1.2   Research Question Two: Effect of Programme on Customer Retention

Research question two deals with the effectiveness of the loyalty programme in influencing customer retention in relation to income levels. The level of effectiveness will be tested between income segments to ascertain if a statistically significant relationship exists between the two variables.

Hypothesis 2: Income is a significant predictor of the effectiveness, related to the retention rate, of a loyalty programme. If this hypothesis holds, we expect to see different levels of improvements in attrition rate by across different income groups. Therefore, a test this hypothesis can be formulated in terms of a Wald test for parameter significance in logistic regression:

H_0: $\beta_1 = 0$;   Income is not a significant predictor of effectiveness.

H_1: $\beta_1 \neq 0$;   Income is a significant predictor of effectiveness.

5.2   Descriptive Statistics

5.2.1   Completeness and Validity of Data

The following checks were done to assess the validity and completeness of the data used when testing the hypothesis:

- Line graphs of the number of accounts, average spend and average number of purchases were plotted to ensure that there were no sudden drops in records.
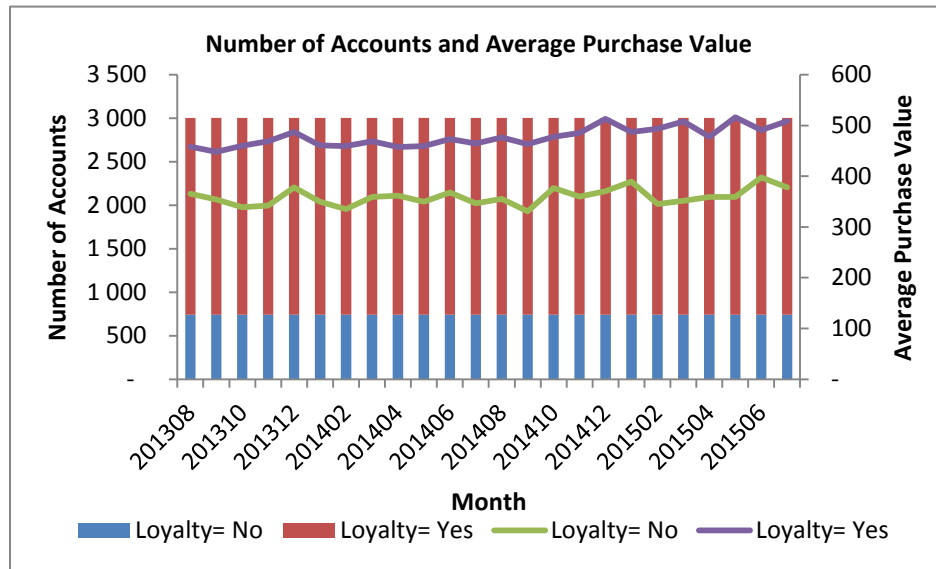
Figure 5 : Number of Accounts and Average Purchase Value over Time
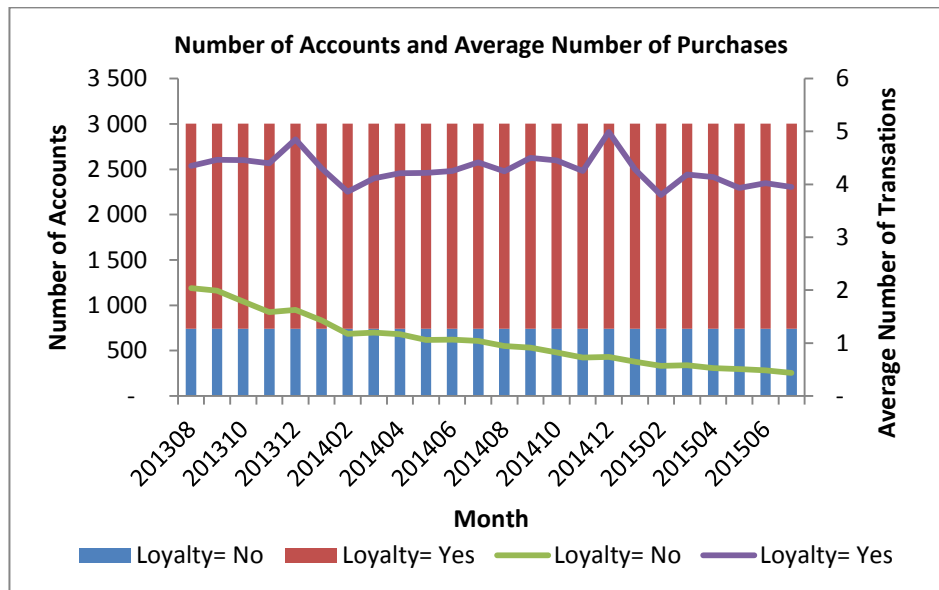
Figure 6 : Number of Accounts and Average Number of Purchases over Time

- The data was checked for missing records and duplications. The proportion of accounts that had no account or customer information was checked. There were reported accounts where the income field is blank. These accounts were excluded.

47

5.2.2 Outlier Analysis

An outlier analysis was performed and all accounts with income greater than the 99th percentile were considered outliers and removed from the development data set.
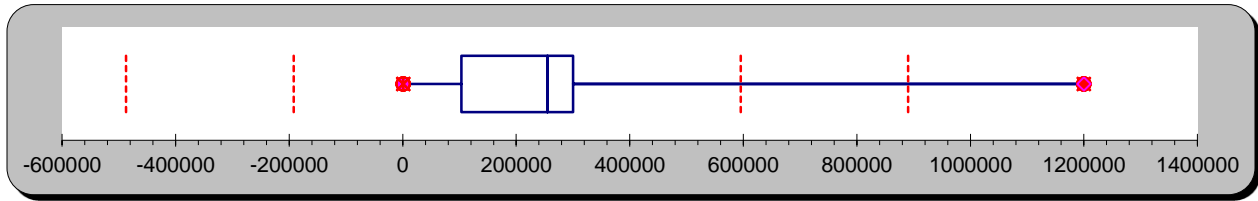
Figure 7 : Box and Whisker Plot

It is evident from the box and whisker figure above that all accounts where the income is greater than R 1,200,000 should be excluded.

5.2.3 Data

The table below outlines the manner in which the sample and validation data set is created. The initial data set consists of 94,751 accounts. The data set is created on a set of active accounts which are then observed for 24 months. The 24 month period during which these accounts are tracked is August 2013 to July 2015.

Table 2 : Data sets used for Modelling

	Loyalty = Yes	Loyalty = No	Total Accounts	Per cent of Base Set
Initial Data Set	72466	22285	94,751	
Percentage of Data Set	76%	24%		
Removal of Accounts with Missing Data	54,273	17,786	72,059	
Percentage of Data Set	76%	25%		
Removal of Outliers	53699	17690	71,389	
Percentage of Data Set	75%	25%		
Development Sample	37621	12379	50,000	70%
Percentage of Data Set	25%	75%		
Validation Sample	16,078	5,311	21,389	30%

48

Percentage of Data Set	25%	75%

The second section represents the number of accounts, which are free from missing data. There were many accounts for which the income variable is missing. Income is a key covariate when building the model and testing the hypothesis. Since it is not straight-forward to build models on accounts where this data is missing, such accounts have been removed from the base set.

The third row represents the remaining accounts after outliers are removed. These include all accounts where the income was more the R 1,200,000.

Using simple random sampling, the sample set of 50,000 accounts was selected. This is represented by 75% with loyalty programme and 25% without a loyalty programme. This is mostly consistent across all sets.

The remaining accounts are then part of the Validation Sample, and will be used to test the models' accuracy. The development set represents about 70% of the population, while the validation set represents the remaining 30% of the population.

5.2.4 Sampling

It is important to ensure that the sample is a good representation of the population being studied. Tests were conducted on the sample to ensure it is representative of the population. The figures below illustrate two variables of interest, average purchase value and average number of purchases on the development and the population sample sets for each month. It is clear from this figure that the sample is a good representation, as it strongly tracks the population.
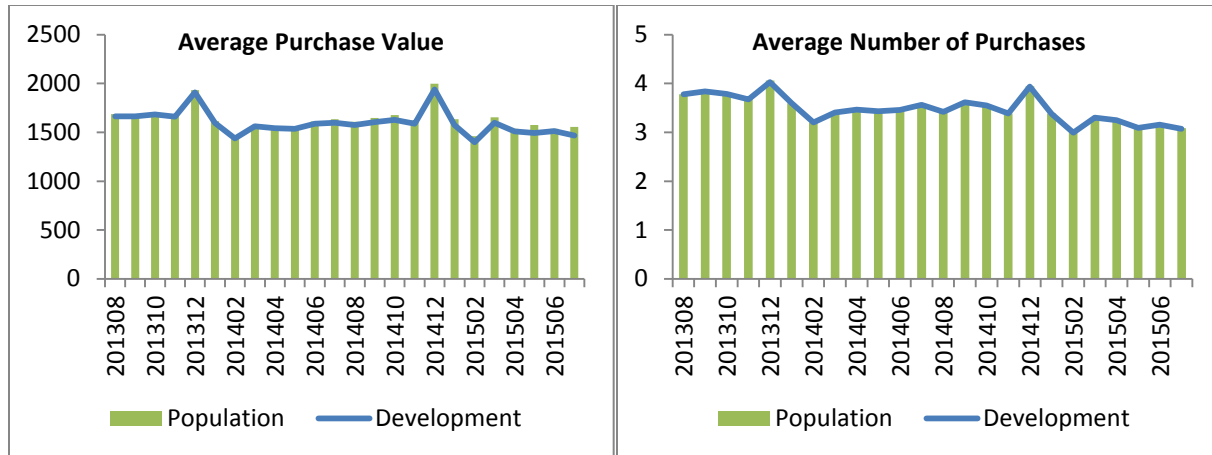
Figure 8 : Sampling Precision

5.3 Effect of Loyalty programme on Value of Spend

Before testing whether or not there is a difference in effectiveness of a loyalty programme between the different income segments, we test whether the loyalty programme is effective in general. For this purpose, a two sample t-test for equality of population means is used.

The t-test relies on the assumption that data is normally distributed. This assumption is tested using a Quantile-Quantile plot. For this, the natural logarithm of the monthly spend per account was taken.
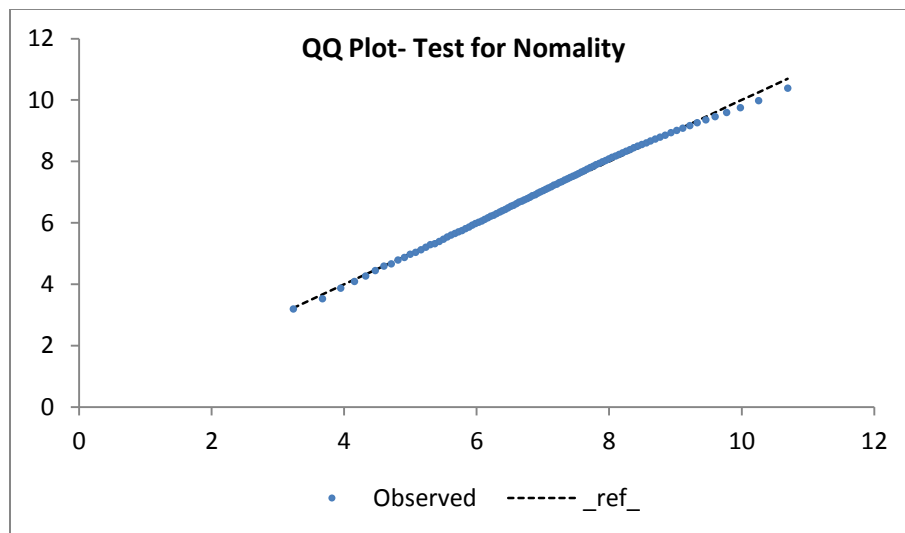
Figure 9 : Quantile-Quantile Plot

50

Visual inspection of the Quantile-Quantile plot suggests that logged spend amount is normal (i.e. spend is log-normally distributed). Therefore, the t-test was performed on the logged spend amount.

The effectiveness is measured by the difference in the average value of purchases on accounts that are on the loyalty programme versus those that are not. If the difference in sample means is not statistically significant, we would conclude that the loyalty programme does not have an effect on spending behaviour.

For this t-test, the hypothesis is laid out below:

$$H_0: Average\ spend(Loyalty = Yes) = Average\ Spend\ (Loyalty = No)$$

$$H_1: Average\ spend(Loyalty = Yes) <> Average\ Spend(Loyalty = No)$$

Average Purchase Value	
Loyalty= No	6.43
Loyalty= Yes	7.13

It is clear from the table, that the average purchase value of accounts on the loyalty programme exceeds that of those that are not. To test whether this difference is statistically significant to reject the null, test statistic (t-statistic) is calculated

This resulted in a t-statistic of - 902.71. Thus, we reject the null hypothesis for any t outside of the region -1.96 < t_c < 1.96. Since - 902.71 is less than -1.96, the null is rejected, asserting the position that there is a significant difference in the average purchase value between accounts with a loyalty programme and those without.

To test the relationship of the effectiveness of a loyalty programme and income, a linear regression is employed.

5.3.1   Income Bands

51

For a linear regression to be accurate and in order to perform tests for significance, a sample size of at least 30 is required. Income is therefore segmented into thirty bands.

5.3.2   Effectiveness

The effectiveness variable is created using the following formula:

$$Effectiveness = \frac{Average\ Spend\ Value(Loyalty = Yes)}{Average\ Spend\ Value\ (Loyalty = No)} - 1$$

This hypothesis is tested using simple linear regression, where the dependent variable is effectiveness and the independent variable is the income band.
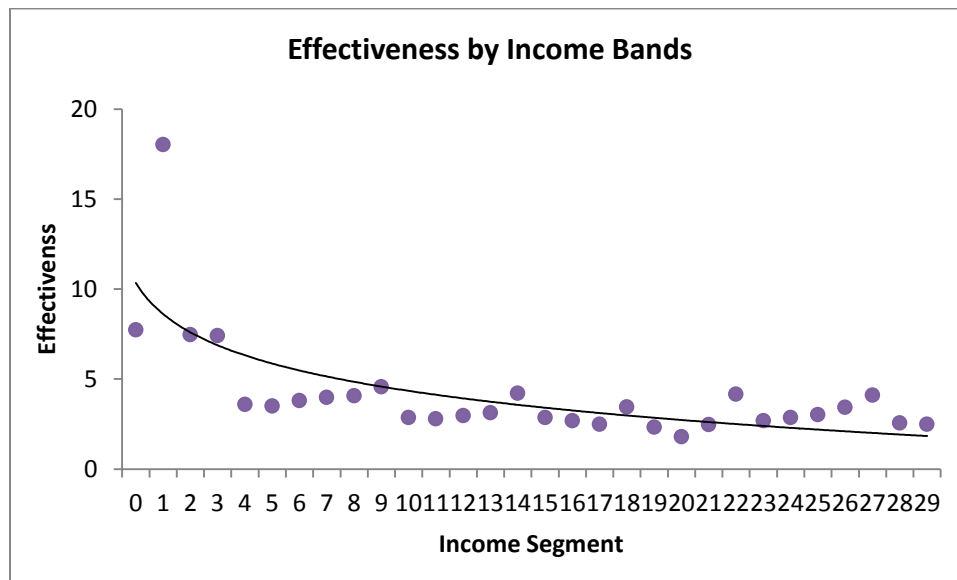
From the scatter plot above, it is evident that a negative relationship exists between income and effectiveness.

5.3.3   Regression Results

Effectiveness is regressed against the different income bands and results from this regression are represented below.

52

Table 3 : Regression Results: Hypothesis 1a

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	131.32542	131.32542	27.73	<.0001
Error	28	132.61796	4.73636		
Corrected Total	29	263.94338			

Root MSE	2.17632	**R-Square**	0.4976		
Dependent Mean	4.11709	**Adj R-Sq**	0.4796		
Coeff Var	52.86059				

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	10.34425	1.24757	8.29	<.0001
income	1	-2.50227	0.47521	-5.27	<.0001

The regression equation of best fit is represented below:

$$y = -2.642 ln(x) + 10.517$$

The table above shows that the model is deemed to be statistically significant, owing to the fact that the F-statistic is at a significance level of 0.0001. Since this is a simple linear regression and there is only one independent variable, this results in overall statically significance model as well.

This regression resulted in a coefficient of determination (R^2) equal to 0.4976. This indicates that income explains about 50% of the variation in effectiveness. The regression coefficient has a negative sign; which indicates that there is a negative relationship between the effectiveness and income levels i.e. the loyalty programme leads to a greater increase in average spend in lower income groups than in higher income groups

5.3.4   Correlation Analysis

Again, since this a simple linear regression, the correlation coefficient (r) is the square root of the coefficient of determination

$$r = \sqrt{R^2}$$

, this results in a correlation coefficient that has a value of 0.70. This asserts that not only is there a negative relationship, it is also relatively strong.

5.3.5 Wald Test for Significance

The output from table, indicates that the independent variable, in this case income, is significant recording a t-statistic of -4.56 and a p-value of <.0001. The null hypothesis can therefore be rejected, asserting that there is a difference in the effectiveness of loyalty programme across income segments.

5.4 Effect of Loyalty programme on Value of Spend

To test the effectiveness of the loyalty programme by income, using purchase behaviour, a linear regression is employed.

5.4.1 Income Bands

For a linear regression to be accurate and in order to perform tests for significance, a sample size of at least 30 observations is required. Income is therefore segmented into thirty bands.

5.4.2 Effectiveness

Effectiveness is created using the following formula:

$$Effectiveness = \frac{Average\ Number\ of\ Transactions\ (Loyalty = Yes)}{Average\ Number\ of\ Transactions\ (Loyalty = No)} - 1$$

This hypothesis is tested using simple linear regression, where the dependent variable is effectiveness and the independent variable is the log of income. (This is the result from the normality test done above).
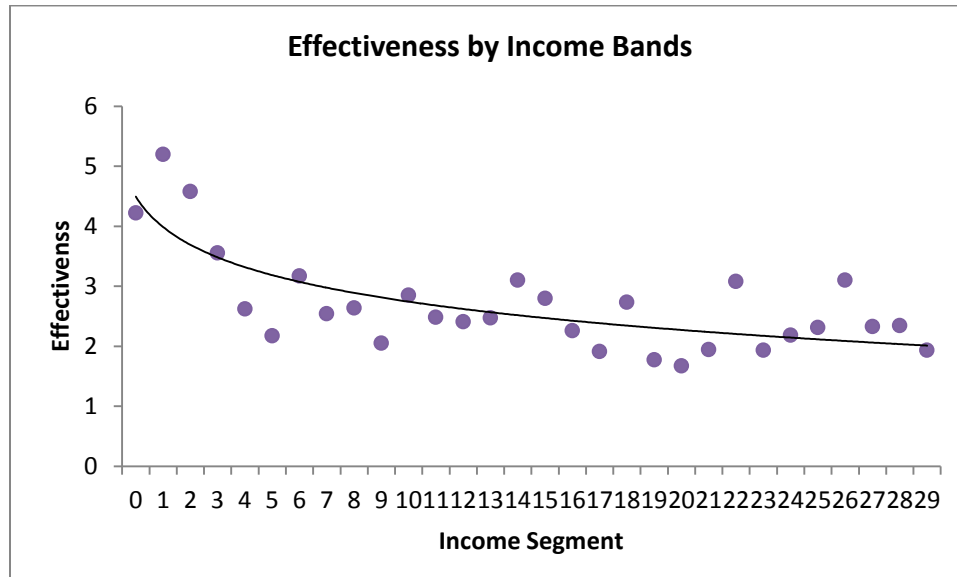
Figure 11 : Scatter Plot – Effectiveness: Average Number of Purchases

From the scatter plot above it is evident that there is a negative relationship between income and effectiveness.

5.4.3 Regression Results

Effectiveness is regressed against the different income bands and results from this regression are represented below:

Table 4 : Regression Results: Hypothesis 1b

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	11.1659	11.1659	37.07	<.0001
Error	28	8.43438	0.30123		
Corrected Total	29	19.60028			

55

Root MSE	0.54884	**R-Square**	0.5697		
Dependent Mean	2.6763	**Adj R-Sq**	0.5543		
Coeff Var	20.50748				

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	4.49208	0.31462	14.28	<.0001
income	1	-0.72964	0.11984	-6.09	<.0001

The regression equation of best fit is represented below:

$$y = -0.73 ln(x) + 4.4921$$

The table above shows that the model is deemed to be statistically significant, owing to the fact that the F-statistic is at a significance level of 0.0001. Since this is a simple linear regression and there is only one independent variable, this results in overall statically significance as well.

This regression resulted in a coefficient of determination (R^2) equal to 0.5697. This indicates that income explain 57% of the variation in effectiveness. The regression coefficient has a negative sign; which indicates that there is a negative relationship between the effective and income levels.

5.4.4 Correlation Analysis

The correlation coefficient in this data set has a value of 0.75. This asserts the fact that there a strong negative relationship between effectiveness and income.

5.4.5 Wald Test for Significance

The output from table, indicates that the independent variable, in this case income, is significant, recording a t-statistic of -5.89 and a p-value of <.0001. The null hypothesis can therefore be rejected, asserting that there is a difference e in the effectiveness of loyalty programme across income segments.

5.5 Effect of Loyalty programme on Customer Retention

5.5.1 Overall Effectiveness Test

Before testing whether or not there is a difference in effectiveness of a loyalty programme between the different incomes segments, we test the overall effectiveness of the programme on retention in general. For this purposes of this, binomial test for proportions equality of population proportions is employed.

The effectiveness is measured by the average retention rate of accounts on the loyalty programme versus those that are not. If the difference is not statistically significant, we would conclude that the programme does not have an effect on customer retention.

For this t-test, the hypothesis is laid out below:

$$H_0: Retention\ Rate(Loyalty = Yes) = Retention\ Rate\ (Loyalty = No)$$

$$H_1: A\ Retention\ Rate(Loyalty = Yes) <> Retention\ Rate(Loyalty = No)$$

Table 5 : Retention Rates

Average Retention Rate	
Loyalty = No	66%
Loyalty = Yes	99%

It is clear from the table, that the retention rate of accounts on a loyalty programme exceeds that of those that are not. To test whether this difference is significant a t-test is then performed.

This resulted in a z-statistic of 655. Thus, we reject the null hypothesis for any t outside of the region -1.96 < Z_c < 1.96. Since 655 is greater than 1.96, the null is rejected; asserting the fact that there is a significant difference in the retention rate between accounts with a loyalty programme and those without.

We proceed to test the effectiveness of the loyalty programme across different income segments.

5.5.2   Effectiveness by Income Segment

A logistic regression is used to test the hypothesis relating to the difference in retention rate across income band. The dependent variable in the regression is dichotomous, it represents a closed indicator. An indicator of 1 means the account has closed, and 0 means it remains open. The retention rate (rr) is then given by the formula:

$$rr = 1 - closure\ rate$$

Before employing the logistic regression a visual inspection of the average closure, can be seen in the figure below. The volatility is a result of using many income segments.
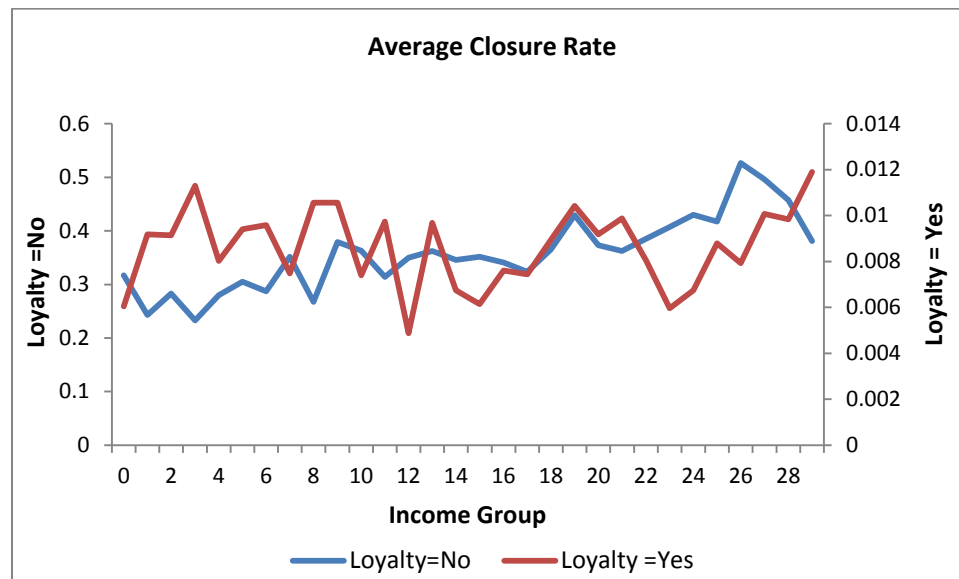
When using Logistic regression, grouping is useful. The thirty income segments are reduced to five groups and then regressed. This is the result of also ensuring that rank ordering takes place.

58

Table 6 : Income Bands

Group	Minimum Income	Maximum Income
0	12	34,200
1	34,800	98,844
2	98,916	156,000
3	156,012	235,368
4	235,428	383,844
5	384,000	1,200,000

5.5.3   Regression Results

The regression result is shown in the figure below:

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
income_group	5	17503.7	<.0001
income_group*LOYALTY	5	41929.05	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-2.1249	0.0101	44527.25	<.0001
income_group	0		1	1.2871	0.0378	1160.19	<.0001
income_group	2		1	1.4705	0.0159	8586.829	<.0001
income_group	3		1	1.4545	0.0164	7880.699	<.0001
income_group	4		1	1.6973	0.0174	9536.195	<.0001
income_group	5		1	1.9126	0.0203	8833.149	<.0001
income_group*LOYALTY	0	Y	1	-3.8257	0.0987	1502.905	<.0001
income_group*LOYALTY	2	Y	1	-4.1854	0.0428	9570.936	<.0001
income_group*LOYALTY	3	Y	1	-4.4163	0.0476	8617.626	<.0001
income_group*LOYALTY	4	Y	1	-4.2353	0.0389	11839.48	<.0001

income_group*LOYALTY	5	Y	1	-4.4109	0.0433	10398.11	<.0001

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	82.8	**Somers' D**	0.743
Percent Discordant	8.5	**Gamma**	0.813
Percent Tied	8.7	**Tau-a**	0.126
Pairs	2.23E+10	**c**	0.871

5.5.4 Wald Test for Significance

Further inspection of the results indicate that all variables are significant as the p-values are <.0001. It is therefore possible to reject the null hypotheses and confirm that the effectiveness, in terms of retention.

The formula for calculating the average closure rate is represented below:

$$f(x) = \begin{cases} \dfrac{1}{1 + e^{-(\alpha + \beta_k)}} & for\ Income\ Group\ k, Loyalty = N \\ \dfrac{1}{1 + e^{-(\alpha + \beta_k + c_k)}} & for\ Income\ Group\ k, Loyalty = Y \end{cases}$$

The parameter c_k visible in the equation above is the difference in the closure rate between accounts on a loyalty programme versus those that are not. Since this value is negative (the coefficient of, income_group*loyalty), this indicated that account on a loyalty programme have a lower closure rates.

The table below demonstrates the overall effectiveness in terms of retention.

	Closure		Retention		Effectiveness
Group	Loyalty=No	Loyalty=Yes	Loyalty=No	Loyalty=Yes	
1	31.7%	0.6%	68.3%	99.4%	45.5%
2	26.9%	0.9%	73.1%	99.1%	35.5%
3	33.8%	0.8%	66.2%	99.2%	49.8%

4	34.9%	0.8%	65.1%	99.2%	52.4%
5	39.7%	0.8%	60.3%	99.2%	64.4%

Table 8: Applying the Logistic Model
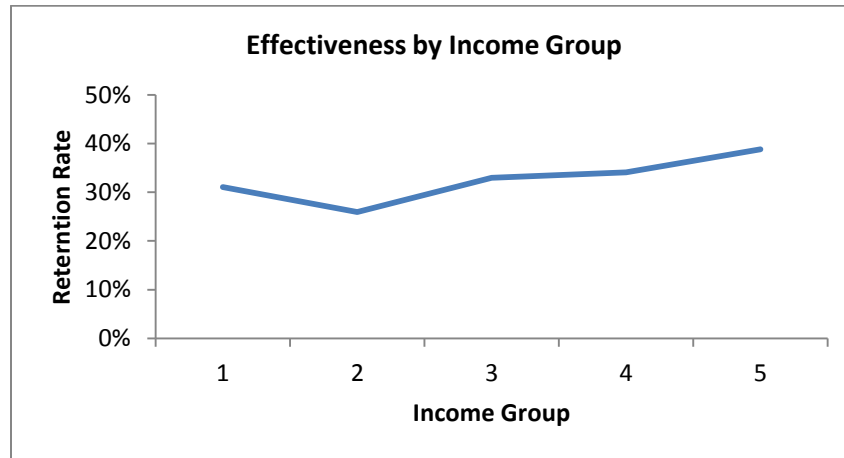
Figure 13 : Retention Effectiveness by Income Group

It is evident from the table above and the figure above, the effectiveness of a loyalty programme is positively correlated to income in terms of retention, i.e. high income customers that are not on the loyalty programme have a much higher attrition rate.

5.5.5 Model Validation

The validation data set was used to assess the accuracy of the models. The Hosmer –Lemshow graphs are seen in the figure below for the development and validation set.

61
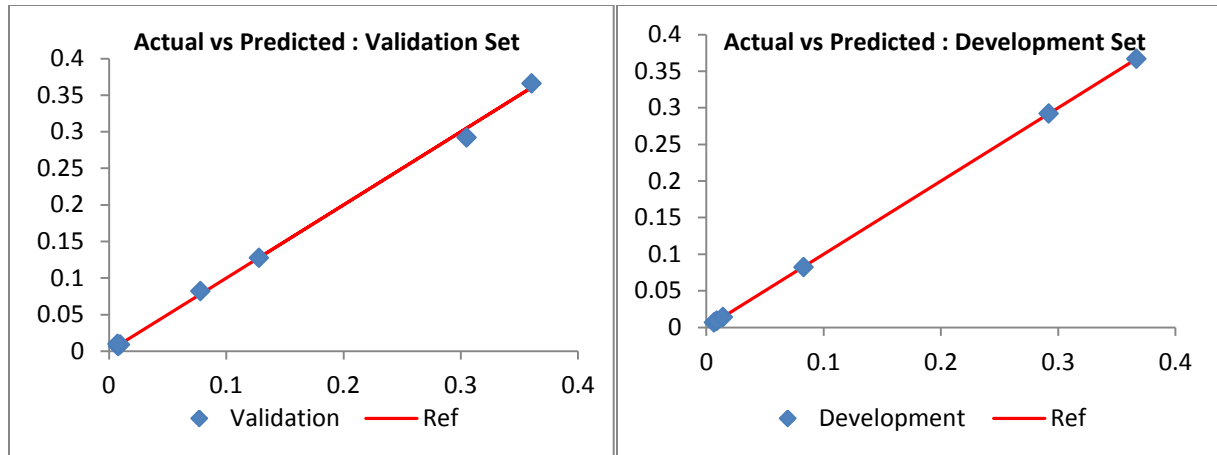
Figure 14 : Hosmer –Lemshow Accuracy Plots

The accuracy plots confirm the models are a good fit.

© University of Pretoria

6 Chapter 6: Discussion of results

6.1 Introduction

This chapter of the research is a culmination of the previous chapters. It serves to explain the results and explanations thereof outlined in chapter five and tested based on the framework and methodology laid out in chapter four. The results and explanations of chapter five are aimed at answering the hypotheses presented in chapter three, which were developed from the major themes in literature and the research problem as described in chapters one and two.

In this chapter the results of chapter five will be analysed and interpreted and related back to major themes and research problems covered in chapters one and two. The confirmation of the hypotheses will be discussed as well as provide any additional insights observed that is relevant to the research problem. The analysis process and discussion is aimed at addressing and answering all three hypotheses in an effort to address the main research objective.

6.2 Discussion of Hypothesis 1A:

Income is a significant predictor of the effectiveness, related to average number of purchases, of a loyalty programme.

Hypothesis 1A is aimed at answering in part the research question of whether there is a relationship between income level and the effectiveness of loyalty programme on influencing purchase behaviour. This hypothesis looks specifically at the value component of purchase behaviour.

6.2.1 Measuring effectiveness

In order to ascertain if a relationship exists, the initial step was to measure the effectiveness of the programme at different income levels on an overall basis. This was calculated by observing the difference in purchase value between loyalty and non-loyalty customers. The results obtained from the t-test performed validate the theory that accounts on a loyalty program

perform better than account not on a loyalty programme in relation to value of purchase. This validates many of the claims of current literature (Zhang & Breugelmans, 2012), (Dorotic, Bijmolt, & Verhoef, 2012), (Henderson, Beck, & Palmatier, 2011) and (Stading & Johnson, 2012).

6.2.2 Validity of the model

Income level was split into 30 different bands, which .ensured accuracy of the regression and validity of the tests. The distribution of income across the bands, as depicted in table 7, was done in an equitable and proportionate manner which again validates the integrity of results obtained.

The F-statistic for the model, as stated in table 3, is at a significance level of 0.0001. This indicates that the results are statistically significant and that the relationship between the independent variable (income) and the dependent variable (value of spend) is not a spurious relationship.

The coefficient of determination for the regression is equal to 0.4976, which means that 49.76% of the variation in the dependent variable is explained by the independent variable. For a simple linear regression, the explanation of the variation by the independent variable is significant.

The correlation coefficient is an indication of the strength of the linear relationship between the variables. The correlation coefficient is presented as 0.70 and indicates a strong relationship between the variables.

These factors combined prove the validity of the results presented in chapter five.

6.2.3 Results

The regression equation of best fit is represented below:

$$y = -2.143 \ln(x) + 9.2311$$

The coefficient of x is negative which indicates that there is a negative relationship between the independent and dependent variables. This means that in this particular loyalty programme, as income increases the level of effectiveness of the programme on increasing value of purchases decreases. This illustrated in figure 10. The regression equation states that moving up one income band results in a decrease in effectiveness of 2.143.

64

A t-test was carried out to test for significance outlined in table 3. At a significance level of 0.05, the p-value is less than 0.001. Given this result the null hypothesis can thus be rejected.

6.3   Discussion of Hypothesis 1B:

Income is a significant predictor of the effectiveness, related to the retention rate, of a loyalty programme.

Hypothesis 1B is aimed at answering the second component of the research question of whether a relationship exists between income level and the effectiveness of loyalty programme on influencing purchase behaviour. This hypothesis looked specifically at the volume component of purchase behaviour.

6.3.1   Measuring effectiveness

As with hypothesis 1A, this hypothesis was tested using a linear regression. Income was thus regressed against effectiveness to determine if a significant relationship exists

6.3.2   Validity of Model

Income was segmented into the thirty bands as per the previous analysis in order to maintain integrity of the results.

The F-statistic for the model, as stated in table 4, is at a significance level of 0.0001. This indicates that the results are statistically significant and that the relationship between the independent variable (income) and the dependent variable (value of spend) is not a spurious relationship.

The coefficient of determination for the regression is equal to 0.5697, which means that 56.97% of the variation in the dependent variable is explained by the independent variable. This is slightly more than recorded in hypothesis 1A and once again the explanation of the variation by the independent variable is significant.

The correlation coefficient is an indication of the strength of the linear relationship between the variables. The correlation coefficient is presented as 0.75 which indicates a strong relationship between the variables.

These factors combined prove the validity of the results presented in chapter five.

6.3.3 Results

The regression equation of best fit is represented below:

$$y = -0.703 ln(x) + 4.5114$$

The coefficient of x is negative which indicates that there is a negative relationship between the independent and dependent variables. This means that, similar to hypothesis 1A, in this particular loyalty programme, as income increases the level of effectiveness of the programme in increasing value of purchases decreases. This illustrated in figure 11. The regression equation states that an increase in one income band results in a decrease in effectiveness of 0.703.

A T-test was carried out to test for significance outlined in table 4. A significance level of 0.05 results in a p-value of less than 0.001Given this result the null hypothesis can thus be rejected.

6.4 Discussion of Research Question 1:

Is there a relationship between income level and the effectiveness of the loyalty programme in influencing purchase behaviour?

The objective of research question is to determine if income would be a suitable method of segmentation for loyalty programme customers. By answering the research question one can establish whether or not the level of effectiveness does vary with income level in a single loyalty programme design. Research question one looks at the effectiveness from the aspect of influencing purchase behaviour.

6.4.1 Existence of a relationship

The F-statistic is both hypothesis 1A and 1B stated that the relationship was significant, which proves that there indeed is a relationship between income level and effectiveness in a specific loyalty programme with no variation in design. This validates the themes built in chapter two on financial versus social benefits from the literature of (Mimouni-Chaabane & Volle, 2010), (Horváth & Birgelen, 2015) and (Göritz, Büttner, & Anja, 2015). They suggested that

66

effectiveness would vary between income levels due to the variations in preference of financial versus social benefits. Therefore if the type of benefits remained constant the effectiveness would vary.

Another theme that came through in the literature was that timing of rewards can affect the effectiveness of a loyalty programme (Drèze & NuNes, 2011), (Huang & Zhang, 2011) and (Sayman & Hoch, 2014). It was built up that timing of rewards can be experienced differently by different income levels and this is indicative in the results. Bagchi & Li (2010) stated that with their concept of reward distance and step-size ambiguity, which again is validated by the existence of a significant relationship.

6.4.2　A negative relationship

Both tests prove that a negative relationship exists between the independent and dependent variables, which means that in this particular loyalty programme an increase in income reduces the effectiveness of influencing purchase behaviour.

Dorotic, Bijmolt, & Verhoef (2012) stated those loyalty programmes are typically designed with financial benefits. These benefits typically appeal to the utilitarian customer (Mimouni-Chaabane & Volle, 2010), (Göritz, Büttner, & Anja, 2015) and (Horváth & Birgelen, 2015). It was inferred that the lower income customer would fall into the utilitarian behavioural group and it is proved by the negative relationship between the variables.

Feelings of gratitude invoke reciprocity and in loyalty programme should influence customers purchasing behaviour (Palmatier, Jarvis, Bechkoff, & Kardes, 2009). It was inferred that the higher the income the lower the levels of gratitude and ultimately the lower the influence on purchase behaviour (Reczek, Haws, & Summers, 2014). This inference is gain validating by the existence of a negative relationship between the variables.

6.4.3　Summary
By rejecting the null for the hypothesis 1A and 1B, research question one is answered in the affirmative. A relationship does exist between income level and the effectiveness of the loyalty programme in influencing purchase behaviour.

This means that income can be used as a predictor of behavioural traits covered in the literature. These behavioural traits affect how customers perceive the elements of a loyalty programme and ultimately the effectiveness of the programme in influencing spend behaviour.

We have thus proved by answering the research question that income can be used as a method of upfront customer segmentation in an effort to predict the effectiveness of a loyalty programme in influencing purchase behaviour.

In this particular study the relationship between the variables in both hypotheses is negative, however that is not indicative of what the relationship should be. This study merely describes the relationship between income and effectiveness in influencing purchase behaviour on one particular loyalty programme. The design of the loyalty programme should dictate whether the relationship is positive or negative. However regardless of the design a relationship does exist which proves the effectiveness of income as a method of segmentation for loyalty customers.

6.5 Discussion of Hypothesis 2:

Income is a significant predictor of the effectiveness, related to the retention rate, of a loyalty programme.

Hypothesis 2 is aimed at answering the second research question of whether or not a relationship exists between income level and the effectiveness of loyalty programme on influencing customer retention.

6.5.1 Measuring effectiveness

As with the first two hypotheses, in order to test the relationship between income and effectiveness one first has to measure the effectiveness of the programme in influencing customer retention. This was again calculated by observing the difference in customer retention between loyalty and non-loyalty customers. The approach here was different from the previous two hypotheses in that an overall closure rate was used to calculate effectiveness. Visual inspection of the data reveals that the rate of retention is higher on loyalty customers versus non-loyalty customers. This is shown in table 5.

A binomial test for proportions was used to test whether the difference in effectiveness is significant. The result of the test indicates that the relationship is significant and it can therefore be concluded that the retention rate is significantly higher on loyalty customer than on non-loyalty customers. As in the case of hypothesis 1A and 1B, the claims of current literature are validated (Dagger & O'Brien, 2010), (Dorotic, Bijmolt, & Verhoef, 2012), (McMullan & Gilmore, 2008).

68

6.5.2   Logistic Regression

For hypothesis two the approach had to be varied from a linear regression to a logistic regression. The reason for this is that the dependent variable is a closed variable, as in an account can only be either open or closed.

The use of thirty income bands causes the volatility in the results. To reduce this and smooth results, the thirty bands were grouped into fewer bands. This is depicted in table 6.

The formula derived from the logistic regression is then applied to estimate the average closure rate. From this the average retention rate is estimated and plotted as an upward sloping line in figure 12.

6.5.3   Results

The results of the regression are plotted in figure 12.As is depicted, the results vary from those of hypothesis 1A and 1B in that the relationship between the dependent and independent variable is recorded as being a positive one. This means that as income increases the effectiveness of the programme with regards to retention increases.

The p-values for all variables are less than 0.001, as indicated in table 8. This proves that the relationship between the variables is significant. As a result, the null hypothesis can thus be rejected.

Model Accuracy

To test the fit of the model an accuracy plot was used as depicted in figure 13. These are the Hosmer-Lemshow plots that depict the predicted results using the regression against the actual results, thereby testing the fit or accuracy of the model. The development and the validation set illustrates that model is a good fit. Refer to figure 13.

Discussion of Research Question 2:

Is there a relationship between income level and the effectiveness of the loyalty programme on influencing purchase behaviour?

The objective of research question is to determine if income would be a suitable method of segmentation for loyalty programme customers. By answering the research question one can establish whether or not the level of effectiveness does vary with income level in a single loyalty programme design. Research question two looks at the effectiveness from the aspect of influencing customer retention.

6.5.4 Existence of a relationship

The p-values for the testing in hypothesis two indicate the existence of a significant relationship between the income level and the effectiveness of the loyalty programmes in relation to customer retention.

This confirms and validates some of the themes introduced in chapter. Coussement, (2014) concept of churners versus non-churners suggested differences in effectiveness between those two. Switching costs were used to identify these two groups of customers (David & Dagger, 2012). It was suggested that switching costs vary with at different levels of income and thus it can de deduced that a relationship exists between income and effectiveness of the loyalty programme. This is confirmed by the research.

Liu & Yang (2009) stated that market saturation has a significant impact on the effectiveness of a loyalty programme. It was suggested that market saturation can vary across income segments and result in varying levels of effectives of a loyalty programme. This is again confirmed by the results obtained.

6.5.5 A positive relationship

Figure 12 clearly indicates a positive relationship between the independent variable (income) and dependent variable (effectiveness). This means that as income level increases so does the effectiveness of the programme with regards to customer retention.

The issue of company loyalty versus programme loyalty was discussed in chapter two. Company loyalty is described as loyalty to the company and is associated more with social benefits (Ramaseshan, Woisetschläger, Richelsen, Blut, & Backhaus, 2010). Company loyalty is more effective when attempting to retain customers and it was suggested that higher income consumers fall into the category. This infers a positive relationship between income and effectiveness and is confirmed by the results.

Dagger & O'Brien (2010) stated that experience played a role in how customers received a loyalty programme. They stated that more experienced customers valued social benefits more

70

and ultimately were enticed by benefits that increased retention. This was overlaid against income criteria and the suggestion was that higher income customers were more experienced. The positive relationship indicated by the results validates the suggestion.

6.5.6   Summary

By rejecting the null for the hypothesis 2, research question two is answered in the affirmative. A relationship does exist between income level and the effectiveness of the loyalty programme in influencing customer retention.

This means that, as with research question one, income can be used as a predictor of behavioural traits covered in the literature. However with research question two we are concerned with the behavioural traits affect how customers perceive the elements of a loyalty programme that influences the effectiveness of the programme with regards to customer retention. By answering the research question and confirming the relationship it has been proven that income can be used as a method of upfront customer segmentation. As with research question one this in an effort to predict the effectiveness of a loyalty programme in influencing customer retention.

Unlike in research question one, here the relationship between the variables is positive. Again this is not an indication of what the relationship should or will always be. It is merely a description of the relationship between income and effectiveness in influencing customer retention on one particular loyalty programme. The aim of the study is not focused on describing the relationship but rather proving that a relationship does exist which proves that income can be used as a method of segmentation for loyalty customers.

6.6   Conclusion

The null hypothesis for all three hypotheses was rejected and thus confirming that a significant relationship exists between income level and the effectiveness of a loyalty programme.

The results and findings were consistent with the themes built upon in the review of the literature. The literature offered different behavioural traits that explained the way a customer received and reacted to the design of a loyalty programme. By proving that a relationship exists between income and effectiveness, it can be deuced that income can be used to predict the aforementioned behavioural traits. This will therefore confirm the validity and effectiveness of using income as a means of customer segmentation in loyalty programme.

71

In addition to proving the existence of a relationship between income and effectiveness, the nature of the relationship was uncovered for a specific programme. The findings, though limited to that of the programme in the study, were once again consistent with the themes raised in current literature. Literature suggested that under the typical design of a loyalty programme, the lower income consumer would react better from a purchase behaviour perspective, while high income consumers would react better from a retention perspective. The higher effect on purchase behaviour and lower impact on retention for low income customers can be labelled as spurious loyalty (Stading & Johnson, 2012), whereas the inverse effects for high income customers can be labelled as latent loyalty. These results are indicative of the design of the programme and as suggested by literature, this is the typical design of most loyalty programme.

In closing this chapter the results of the research in relation to the literature, research questions and proposed hypothesis. Further to this any additional finding were noted and discussed and possible explanations for their occurrence were discussed.

7 Chapter 7: Conclusion

7.1   Introduction

The previous chapter served to discuss the results of the research and relating back to the literature and research problem. The following chapter serves to recap the research objectives, outline the principle findings and discuss the implications to business. It then proceeds to highlight the limitations of the research and provide suggestions for future research.

7.2   Recap of objectives

In an effort to increase customer lifetime value by retaining customers and increasing purchase behaviour, retail industries are exhibiting greater focus on customer relationship management. A direct result of this shift in focus is the growing popularity in loyalty programmes that aim to increase both customer retention and purchase behaviour.

The effectiveness of a loyalty programme at influencing these two variables varies based on the customer's behavioural traits. Current research focuses on segmenting customers based on their behavioural traits once enrolled to a programme but does not provide any method for upfront segmentation of customers. Businesses can later the design of the loyalty programme to increase its effectiveness based on a customer's behaviour. However without an objective upfront method of segmentation, business cannot identify the appropriate design for a customer.

The aim of this study was to investigate if income could be used as a method of segmentation. For income to be a suitable method of segmentation, a relationship must exist between income and the effectiveness of a loyalty programme. The objective was therefore to test if this relationship does indeed exist.

7.3   Summary of principle findings

The findings of current literature as to the effectiveness of loyalty programmes was confirmed (Zhang & Breugelmans, 2012), (Dorotic, Bijmolt, & Verhoef, 2012), (Drèze & NuNes, 2011). It was proven that loyalty programmes are effective at increasing purchase behaviour and retention.

It was proven that a statistically significant relationship exists between income and the effectiveness of a loyalty programme in influencing both retention and purchase behaviour. Income was proven to be a significant predictor of effectiveness in influencing these two

73

variables. This would imply that income could act as a predictor of customer's behaviour in a loyalty programme and thus is proven to be a suitable and effective method for customer segmentation.

It was also proven that in a conventionally designed loyalty programme, such as the one used in this study, the relationship between income and effectiveness in relation to purchase behaviour is negative. Therefore the higher the income, the lower the effectiveness of the programme in increasing purchase behaviour.

The relationship between income and effectiveness in relation to retention was shown to be negative. This means that the higher the income, the greater the effectiveness of the programme in increasing customer retention.

7.4   Implications for business

This study has succeeded in providing business with a means to segment loyalty programme customers upfront on enrolment. The ability to segment customers using an objective measure such as income will allow businesses to predict the customer's behaviour. Dorotic, Bijmolt, & Verhoef (2012) stated that design of a loyalty programme is critical to its success. The ability to segment allows the business to adjust the design of the loyalty programme per segment based on the predicted behaviour.

Coussement (2014) highlighted the significant costs associated with loyalty programmes and Hoseong & Choi (2013) state that many loyalty programmes fail to achieve the expected return. Tailoring the deisgn per segment will result in the removal unnecessary benefits and increase effectiveness by offering relevant benefits per segment. This will result in both a decrease in loyalty programme costs and an increase in the return on investment.

The research has also revealed additional implications for business in terms of the current typical design of loyalty programmes. Aspects of the design outlined in the literature review can be adjusted for low income customers to increase the programme's effectiveness at increasing retention. The design should also be adjusted for high income customers to increase the programme's effectiveness on increasing purchase behaviour.

74

7.5   Limitations of the research

The research was carried out on only one company and therefore only one loyalty programme and. The results are therefore limited to the design and marketing elements of the programme in question. As a result the findings may be a consequence of the particular business strategy of the company and may not necessarily be indicative of all other loyalty programmes. However using one company in the research data allows for variables such as design to remain constant.

The company and programme in question is in the retail banking industry. Though most loyalty programmes in the retail industry are designed similarly, the results may not be the same across all retail industries.

The research data is limited to a two year period only due to the availability of data. The data may as a result be influenced by industry, market and economic trends for the particular period. However a two year period would be sufficient to cover general industry cycles and is therefore deemed sufficient.

Finally other variables outside of the scope of this research may have an effect of effectiveness such as age, age of account, number of other product holdings and risk rating. However these variables were not highlighted in current literature and due to the robust methodology employed it is not deemed a major flaw. However isolating some of these factors an accounting for them could increase the validity of the results.

7.6   Suggestions for future research

Based on the findings and limitations of this research, some suggestions for future research were compiled.

It would be beneficial to replicate the study with loyalty programmes from different companies and different industries than that of this study. It may add value to include more than just one loyalty programme within the data set. This could improve the validity of the results and test the consistency across different loyalty programme designs.

Another avenue for future research would be test whether or not a relationship exists between specific design variables and the effectiveness of a loyalty programme within different income segments. This research offered income as a means of segmentation and the proposed research would seek a method of establishing the appropriate design for each segment. The

75

proposed research would thus further this research and expand the understanding of loyalty programmes.

Finally it is suggested that future research explore other possible variables that could be used for upfront segmentation other than level of income. A new segmentation method could result in different segments and subsequently result in different design variations to maximise effectiveness. Additional options of design variations will offer greater flexibility to business. Alternatively other segmentation variables could be overlaid onto income segmentation to result in a more granular segmentation.

7.7 Conclusion

Income was proven to be a significant predictor of the effectiveness of a loyalty programme. This implies that income can be used as an effective of segmenting customers on enrolment of a loyalty programme and will allow businesses to tailor the design of the programme to suit the income segment.

These findings will only businesses to reduce costs and increase the returns on their loyalty programmes. Academically it has added to the current body of knowledge and understanding of loyalty programmes as well as opened up avenues for further research.

8 Bibliography

Bagchi, R., & Li, X. (2010). Illusionary Progress in Loyalty Programs: Magnitudes, Reward Distances, and Step-Size Ambiguity. . *JOURNAL OF CONSUMER RESEARCH*, 888-899.

Breugelmans, E., Bijmolt, T. H., Zhang, J., Basso, L. J., Dorotic, M., Kopalle, P., et al. (2014). Advancing research on loyalty programs: a future research agenda. *Marketing Letters*, 127-139.

Cameron, C., Windmeijer, F., Gramajo, H., Cane, D., & Khosla, C. (1997). An R-Squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models. *Journal of Econometrics*, 329-342.

Carlson, W., & Thorne, B. (1997). *Applied Statistical Methods.* United States of America: Prentice-Hall Inc.

Chatterjee, S., & Hadi, A. (2012). *Regression Analysis by Example.* New Jersey: Wiley& Sons Inc.

Coussement, K. (2014). Improving customer retention management through cost-sensitive learning. *European Journal of Marketing*, 477-495.

Cox, D. R., & Snell, E. J. (1989). *The Analysis of Binary Data.* London.: Chapman & Hall.

Dagger, T. S., & O'Brien, T. K. (2010). Does experience matter? Differences in relationship benefits, satisfaction, trust, commitment and loyalty for novice and experienced service users. *European Journal of Marketing*, 1528 - 1552.

Daryanto, A., Ruyter, K. d., Wetzels, M., & Patterson, P. G. (2010). Service firms and customer loyalty programs: a regulatory fit perspective of reward preferences in a health club setting. *Journal of the Academy of Marketing Science*, 604–616.

David, M. E., & Dagger, T. S. (2012). Uncovering the real effect of switching costs on the satisfaction-loyalty association: The critical role of involvement and relationship benefits. *European Journal of Marketing*, 447-468.

Dorotic, M., Bijmolt, T. H., & Verhoef, P. C. (2012). Loyalty Programmes: Current Knowledge and Research Directions. *International Journal of Management Reviews*, 217–237.

Drèze, X., & NuNes, J. C. (2011). Recurring Goals and Learning: The impact of successful reward attainment on purchase Behavior. *Journal of Marketing Research*, 268 –281.

Filliben, J. J. (1975). The Probability Plot Correlation Coefficient Test for Normality. *Technometrics*, 111-117.

Göritz, Büttner, O. B., & Anja, A. F. (2015). How shopping orientation influences the effectiveness of monetary and nonmonetary promotions. *European Journal of Marketing*, 170-189.

Hansen, H., Samuelsen, B. M., & Sallis, J. E. (2013). The moderating effects of need for cognition on drivers of customer loyalty. *European Journal of Marketing*, 1157-1176.

Harrell, F. E. (2001). *Regresion Modelling Strategies.* New York: Springer-Verlag.

Henderson, C. M., Beck, J. T., & Palmatier, R. W. (2011). Review of the theoretical underpinnings of loyalty programs. *JOURNAL OF CONSUMER RESEARCH*, 256-276.

Hiasley, E., & Loewenstein, G. (2011). It's Not What You Get but When You Get It: The Effect of Gift Sequence on Deposit Balances and Customer Sentiment in a Commercial Bank. *Journal of Marketing Research*, 103-115.

Horváth, C., & Birgelen, M. v. (2015). The role of brands in the behavior and purchase decisions of compulsive versus noncompulsive buyers. *European Journal of Marketing*, 2-21.

Hoseong, Y. Y., & Choi, J. B. (2013). Segregation vs aggregation in the loyalty program: the role of perceived uncertainty. *European Journal of Marketing*, 1238-1255.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression.* Hoboken, NJ: John Wiley & Sons inc.

Huang, S.-C., & Zhang, Y. (2011). Motivational Consequences of Perceived Velocity in Consumer Goal Pursuit. *Journal of Marketing Research*, 1045–1056.

Hutchinson, K., Donnell, L. V., Gilmore, A., & Reid, A. (2015). Loyalty card adoption in SME retailers: the impact upon marketing management. *European Journal of Marketing*, 467-490.

Jobber, D., & Shipley, D. (2012). Marketing-orientated pricing: Understanding and applying factors that discriminate between successful high and low price strategies. *European Journal of Marketing*, 1647-1670.

Jupp, V. (2006). Quantative Research. *The sage dictionary of social research methods*.

Kang, J., Alejandro, T. B., & Groza, M. D. (2015). Customer–company identification and the effectiveness of loyalty programs. *Journal of Business Research*, 464-471.

Lewis-Beck, M. S., & Liao, A. B. (2013). Unit of Analysis. *Encyclopedia of Social Science Research Methods*, SAGE Publications, Inc.

Liu, Y., & Yang, R. (2009). Competing Loyalty Programs: Impact of Market Saturation, Market Share and Category Expandability. *Journal of Marketing*, 93–108.

Massey, F. (1957). *Introduction to Statistical Analysis.* New York: McGraw Hill Book Company, Inc.

McMullan, R., & Gilmore, A. (2008). Customer loyalty: an empirical study. *European Journal of Marketing*, 1084-1094.

Melnyk, V., & Bijmolt, T. (2015). The effects of introducing and terminating loyalty programs. *European Journal of Marketing*, 398-419.

Mende, M., Bolton, R. N., & Bitner, M. J. (2013). Decoding Customer-Firm Relationships: How Attachment Styles Help Explain Customers' Preferences for Closeness, Repurchase Intentions, and Changes in Relationship Breadth. *European Journal of Marketing*, 125-142.

Mimouni-Chaabane, A., & Volle, P. (2010). Perceived benefits of loyalty programs: Scale development and implications for relational strategies. *Journal of Business Research*, 32-37.

Oakshott, L. (2006). *Essential Quantative Management.* Hampshire: Palgrave McMillan.

Page, C., & Meyer, D. (2000). *Applied Research Design for Buisness and Management.* Sydney: McGraw-Hill Book Company.

Palmatier, R. W., Jarvis, C. B., Bechkoff, J. R., & Kardes, F. R. (2009). The Role of Customer Gratitude in Relationship Marketing. *Journal of Marketing*, 1-18.

Paswan, A. K., Blankson, C., & Guzman, F. (2011). Relationalism in marketing channels and marketing strategy. *European Journal of Marketing*, 311-333.

Raciti, M. M., Ward, T., & Dagger, T. S. (2013). The effect of relationship desire on consumer-to-business relationships. *European Journal of Marketing*, 615-634.

Ramaseshan, H. E., Woisetschläger, D. M., Richelsen, V., Blut, M., & Backhaus, C. (2010). Consequences of customer loyalty to the loyalty program and to the company. *Journal of the Academy of Marketing Science*, 625–638.

Reczek, R. W., Haws, K. L., & Summers, C. A. (2014). Lucky Loyalty: The Effect of Consumer Effort on Predictions of Randomly Determined Marketing Outcomes. *JOURNAL OF CONSUMER RESEARCH*, 1065-1077.

Rodriguez, W. (2007). *Encyclopedia of measuremeant and statistics.* California: Sage Publications.

Saunders, M., & Lewis, P. (2012). *Doing Research in Business & Management.* London: Pearson Education Limited.

Sayman, S., & Hoch, S. J. (2014). Dynamics of price premiums in loyalty programs. *European Journal of Marketing*, 617-640.

Schwandt, T. (2007). Descriptive Statistics. *The SAGE dictionary of qualitative inquiry*, 67.

Segrin, C. (2010). Mutiple Regression. In C. Segrin, *Mutiple Regression* (pp. 845-850). California: SAGE Publications.

Seo, T., & Srivastava, M. S. (2000). Testing Equality of Means and Simultaneous Confidence Intervals in Repeated Measures with Missing Data. *Biometrical Journal*.

79

Stading, G. L., & Johnson, M. (2012). An Examination of the Relationship Between a Firm's Offerings and Different Customer Loyalty Segments. *Journal of Business-to-Business Marketing*, 367–391.

Steinhoff, L., & Palmatier, R. W. (2014). Understanding loyalty program effectiveness: managing target and bystander effects. *Journal of the Academy of Marketing Science*.

Stourm, V., Bradlow, E. T., & Fader, P. S. (2015). Stockpiling Points in Linear Loyalty Programs. *Journal of Marketing Research*, 253-267.

Triola, M. F., & Franklin, L. A. (1994). *Business Statistics.* Boston: Addison-Wesley Publishing Company Inc.

Vieceli, J., Deakin, & Shaw, R. N. (2010). Brand salience for fast-moving consumer goods: An empirically based model. *Journal of Marketing Management*, 1218-1238.

Weiers, R. M. (2010). *Introductory to Business Statistics.* Cengage Learning.

Wierich, R., & Zielke, S. (2014). How retailer coupons increase attitudinal loyalty – the impact of three coupon design elements. *European Journal of Marketing*, 699-721.

Zhang, J., & Breugelmans, E. (2012). The impact of an item-Based loyalty Program on Consumer Purchase Behavior. *Journal of Marketing Research*, 50-65.

Appendices

Appendix 1: Letter of consent from company

⊗ NEDBANK

135 Rivonia Road
Sandown
2196
23 July 2015

Gordon Institute of Business Science
University of Pretoria
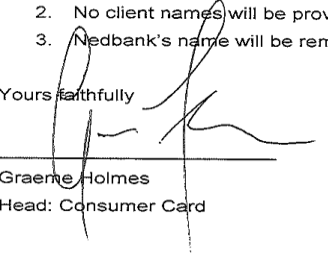26 Melville Road
Johannesburg

Dear Sir/Madam

MBA Research: Mahen Thamanna

This letter serves to confirm that Nedbank Consumer Card has granted Mahen Thamanna permission to conduct his MBA research amongst existing Nedbank Consumer Card clients.

The student will be permitted to conduct research using Nedbank's Consumer Credit Card account data. The permission for use of Nedbank's data will be under the following conditions;

1. The data will be presented in an aggregated format.
2. No client names will be provided in the data.
3. Nedbank's name will be removed from the data to maintain anonymity.

Yours faithfully

Graeme Holmes
Head: Consumer Card

NEDBANK

NEDBANK CARD

135 Rivonia Road Sandown Sandton 2196 PO Box 1777 Johannesburg 2000 South Africa
Tel 0860 555 111 Fax 011 710 4710 Telex 4 22079SA Email name@nedbank.co.za www.nedbank.co.za

Nedbank Limited Reg No 1951/000009/06, VAT Reg No 4320116074, 135 Rivonia Road, Sandown, Sandton, 2196, South Africa.
Directors: Dr RJ Khoza (Chairman) MWT Brown (Chief Executive) TA Boardman TCP Chikane GW Dempster (Chief Operating Officer) MA Enus-Brey ID Gladman*
DI Hope** WE Lucas-Bull PM Makwana NP Mnxasana RK Morathi (Chief Financial Officer) JK Netshitenzhe JVF Roberts* GT Serobe MI Wyman* (*British) (**New Zealand)
Company Secretary: TSB Jali 01.07.2012

We subscribe to the Code of Banking Practice of The Banking Association South Africa and, for unresolved disputes, support resolution through the Ombudsman for Banking Services.
We are an authorised financial services provider. We are a registered credit provider in terms of the National Credit Act (NCR Reg No NCRCP16).

A Member of the 🔷 OLD MUTUAL Group

81

Appendix 2: Ethical clearance

Gordon Institute of Business Science
University of Pretoria

Dear Mahendran Thamanna

Protocol Number: **Temp2015-01461**

Title: **The Effectiveness of Loyalty Programs in the South African Retail Banking Industry**

Please be advised that your application for Ethical Clearance has been APPROVED.

You are therefore allowed to continue collecting your data.

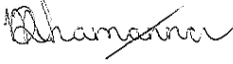We wish you everything of the best for the rest of the project.

Kind Regards,
Adele Bekker

© University of Pretoria

Appendix 3: Copyright declaration form

Gordon Institute of Business Science
University of Pretoria

COPYRIGHT DECLARATION FORM

Student details			
Surname:	**Thamanna**	**Initials:**	**M**
Student number:	**14455413**		
Email:	**mahendrant@nedbank.co.za**		
Cell :	**0846827919**	**Landline:**	**0112622359**
Course details			
Degree:	**MBA**	**Year completed:**	**2015**
Department:	**GIBS**		
Supervisor:	**Pride Morokane**		
Supervisor email:	**Pride Morokane <pridemmm@gmail.com>**		
Confidentiality / Embargo			
Do you need to have your report embargoed? If so, write a letter of motivation to substantiate (please attach letter to this form). Without a letter this will not be granted.			
Yes		**No**	**x**
If yes, please indicate period requested			
Two years		****Permanent**	

83

<table>
<tr><td colspan="4">**Please attach a copy of the letter of permission from the Vice-Principal: Research and Postgraduate Studies, if indicated, permanent. Without a letter this will not be granted.**</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">A copy of your research report will be uploaded to UPetd/UPSpace

Can the Information Centre add your email address to the UPetd/UPSpace web site?</td></tr>
<tr><td>**Yes**</td><td>**x**</td><td>**No**</td><td></td></tr>
<tr><td colspan="4">**If no, please motivate (ignore if report is to be embargoed)**</td></tr>
<tr><td colspan="4">Copyright declaration</td></tr>
<tr><td colspan="4">I hereby certify that, where appropriate, I have obtained and attached hereto a written permission statement from the owner(s) of each third-party copyrighted matter to be included in my research report ("the work"), allowing distribution as specified below. I certify that the version of the work I submitted is the same as that, which was approved by my examiners and that all the changes to the document, as requested by the examiners, have been included.

I understand that all rights with regard to intellectual property in the work vest in the University who has the right to reproduce, distribute and/or publish the work in any manner it may deem fit.

I agree that, a hardcopy of the abovementioned work be placed in the Gordon Institute of Business Science Information Centre and worldwide electronic access be given to the softcopy on UPetd and UPSpace.</td></tr>
<tr><td colspan="2">Signature:</td><td colspan="2">Date: 9/11/2015</td></tr>
</table>

Appendix 4: SAS Code

```
/*Initial Data Set*/

data basedata_rem;
```

84

```
        set basedata;

run;


proc sql;

     create table desc_stats as select

          loyalty, count(distinct(ac_num)) as num_accounts from
basedata_rem

     group by loyalty;

quit;

/*Remove Missing Income*/

data basedata_rem_a (where=(ANUL_GRS_INCM gt 0 ));

     set basedata;

run;


proc sql;

     create table desc_stats1 as select

          loyalty, count(distinct(ac_num)) as num_accounts from
basedata_rem_a

     group by loyalty;

quit;

/*outlier analysis*/


proc univariate data=basedata_rem_a ;

     var ANUL_GRS_INCM;

     output out=outliers pctlpts= 1 25 75 99 mean= mean pctlpre=p;

run;

/*Remove Outliers*/
```

```
data basedata_rem_b (where=(ANUL_GRS_INCM le 1200000 ));

     set basedata_rem_a;

run;


proc sql;

     create table desc_stats1 as select

          loyalty, count(distinct(ac_num)) as num_accounts from
basedata_rem_b

     group by loyalty;

quit;

/*Create Sample*/

proc surveyselect data=basedata_rem_b(where=(tran_cycle=201308))
method=srs n=50000  out=sample;

run;


proc sql;

     create table desc_stats2 as select

          loyalty, count(distinct(ac_num)) as num_accounts from
sample

     group by loyalty;

quit;


Proc sql;

     create table base_sample as

          select a. *,(b. ac_num ne .) as validation, (a.
ANUL_GRS_INCM gt 1200000) as outlier

               from basedata_rem_a as a

                    left join sample as b
```

86

```
                    on   a. ac_num eq b. ac_num;

quit;



/**/

/*proc sql;*/

/*create table descrip as select tran_cycle,loyalty,
sum(ANUL_GRS_INCM) as income,  sum(num_tran) as num_tran,
sum(total_spnd) as total_spend, count(*) as counter*/

/*from base_sample*/

/*group by tran_cycle,loyalty;*/

/*run;*/




/*t-test*/



data qqplot(where=(total_spnd gt 0 and outlier = 0));

set base_sample;

spnd= log(total_spnd);

run ;



proc sql;

create table norm as select loyalty, mean(spnd) as average_spend,
std(spnd) as std_spend, count(*) as counter

from qqplot

group by loyalty;

run;



/*qq plot*/
```

```
data  qqplot(where=(total_spnd gt 0 and outlier = 0));

set base_sample;

where tran_cycl = 201308;

run;


/*Linear Regression*/;

data spend;

    set base_sample;

    income_group = ANUL_GRS_INCM;

run;


proc rank data=spend out=rankedspend groups=30;

    var income_group;

run;


proc sql;

    create table linear_spend as

        select income_group

            ,sum(total_spnd * (loyalty = 'Y')) / sum((loyalty =
'Y')) as yes_spend

            ,sum(total_spnd * (loyalty = 'N')) / sum((loyalty =
'N')) as no_spend

            ,mean(ANUL_GRS_INCM) as income

            ,(calculated yes_spend/calculated no_spend - 1) as
ratio

        from rankedspend

            group by income_group;

quit;
```

```
/*using a regression*/

data a;

set linear_spend;

inc= log(1+income_group);

run;

proc reg data=a;

model ratio=inc;

run;


data spend;

    set base_sample (where=(validation = 0 and outlier=0));

    income_group = ANUL_GRS_INCM;
/*    income_group = age;*/

run;


proc rank data=spend out=rankedtran groups=30;

    var income_group;

run;


proc sql;

    create table linear_tram as

        select income_group

            ,sum(num_tran * (loyalty = 'Y')) / sum((loyalty =
'Y')) as yes_num_tran

            ,sum(num_tran * (loyalty = 'N')) / sum((loyalty =
'N')) as no_num_tran
```

```
              ,mean(ANUL_GRS_INCM) as income

          ,(calculated  yes_num_tran/calculated no_num_tran - 1) as
ratio

          from rankedtran

              group by income_group;

quit;

/*using a regression*/

data a;

set linear_tram;

inc= log(1+income_group);

run;

proc reg data=a;

model ratio=inc;

run;


data basedata_reten;

     set base_sample;


     if close_dte > 0 then

          close_ind=1;

     else close_ind=0;

     income_group=ANUL_GRS_INCM;

run;


proc rank data=basedata_reten out=rankedspend groups=30;

     var income_group;

run;
```

90

```
proc sql;

    create table summary1 as

        select loyalty

            ,count(*) as trials

            ,sum(close_ind) as events

        from rankedspend

            group by loyalty;

quit;

/**/

/*proc sql;*/

/*    create table summary2 as*/

/*        select income_group*/

/*            ,mean(ANUL_GRS_INCM) as ANUL_GRS_INCM*/

/*            ,mean(close_ind) as closed_rate*/

/*            ,loyalty*/

/*        from rankedspend*/

/*            group by income_group,loyalty;*/

/*quit;*/


data rankedspend (where=(validation = 0 and outlier=0));

set rankedspend;

if income_group in (1,2,3,4,5,6) then income_group = 1;

if income_group in (7,8,9,10,11,12) then income_group = 2;

if income_group in (13,14,15,16,17,18) then income_group = 3;

if income_group in (19,20,21,22,23,24) then income_group = 4;

if income_group in (25,26,27,28,29,30) then income_group = 5;
```

91

```
run;


proc sql;

create table a as select income_group,  max(ANUL_GRS_INCM) as
max,min(ANUL_GRS_INCM) as min from rankedspend

group by income_group;

run;




ods output lackfitpartition = logistic_hosmer;

proc logistic data=rankedspend (where=(validation = 0 and outlier=0));

     class income_group (ref='5') loyalty(ref='N') /param=ref;

     model close_ind(event='1') =  income_group income_group *
loyalty;

     score data=rankedspend out=scored;

run;


proc sort data=scored; by validation;run;


data hl;

     set scored;

     decile = p_1;

run;


proc rank data=hl out=hl groups=10;

     by validation;
```

92

```
        var decile;

run;


/*proc sql;*/

/*    create table summary1 as*/

/*         select income_group*/

/*              ,sum(close_ind * (loyalty = 'Y'))/sum((loyalty = 'Y'))
as Yes*/

/*              ,sum(close_ind * (loyalty = 'N'))/sum((loyalty = 'N'))
as No*/

/*         from scored*/

/*              group by income_group;*/

/*quit;*/


/*Accuracy;*/


        /*expected vs actual*/
proc sql;
    create table hosmer_lemeshow as
        select mean(p_1) as predicted
              ,mean(close_ind) as actual
              ,validation
              ,decile

        from hl
              group by validation,decile;
quit;
```