The effects of frequency of purchase and tenure on firm revenue

Saalem Sadeque

BBA, Master of Applied Finance, MSc in Marketing Research and Analysis

This thesis is presented for the degree of Doctor of Philosophy at The University of Western Australia

UWA Business School
Marketing Discipline

2016
Acknowledgements

Firstly, my deepest gratitude goes to my supervisor, Winthrop Professor Dick Mizerski. His thoughtful guidance and constant inspiration were the keys for me to stay on course with this thesis.

My sincerest thanks go to Dr. Wade Jarvis, Winthrop Professor Julie Lee, and Winthrop Professor Geoff Soutar. Their constant encouragement to “get it done” always stiffened my resolve to complete the thesis.

I would also like to acknowledge Dr. Alvin Lee, who was always ready to provide me with invaluable suggestions and encouragements. I would also like to take this opportunity to thank Dr. Sally Gainsbury and Professor Alex Blaszczynski for their support.

I would also like to acknowledge the contributions of the professional staff members at the Business School. Thank you Helen, Sandra, Maryann, Anne-Maree, Robyn, Mei, Isabela, Dani and Adam. Their supports at various times during the course of this epic adventure were priceless. Thanks to all my fellow PhD students who have provided constant support and showed enthusiasm in my work.

My parents, Professor Dr. Md. Sadequzzaman and Nasima Sadeque have always dreamt of this day. I hope that I have been able to fulfil at least a little part of that dream. Thanks to my dear sisters, Nayeema and Sabikeena, who have supported and encouraged me in various ways during this journey.

My daughters’, Nashwa and Nashmia, constant love always provided me with encouragements to complete this task. Lastly, my deepest thanks to my dear wife, Naina, who has constantly and patiently dealt with all the nuisances that comes up when a spouse pursues PhD. The completion of thesis would not have been possible without her understanding and care.
Abstract

A longer tenure for buyers of firm’s products has been argued to bring a number of benefits for the firm; such as generating high profits, lower operating costs, being less price sensitive, etc. This theory-based assumption has led companies to focus on developing and managing customer retention programmes. However, some other studies have cast doubts by reporting that in many marketing contexts, the benefits from longer tenure customers do not actually materialise. For example, some research findings show that short tenure customers can be as profitable as long tenure customers. It appears that buyers’ frequency of purchase may be one important factor that can explain these discrepancies.

Prior studies have not explicitly investigated buyers’ frequency of purchase and tenure in firm revenue. This study investigates these two variables in the context of a very frequently purchased product like gambling. Gambling is now recognized as the largest revenue generating subcategory in the entertainment industry with revenues greater than music, theme parks and movies combined. The study used a dataset obtained from a large gambling service provider in Australia. The dataset contained real time, actual behavioural information collected over a period of over nine years for over 11,000 bettors.
The major contribution of the study is the finding that frequency of purchase has a greater role in firm revenue than tenure with the company. The study also found that longest and shortest tenure customers contribute about the same to firm revenue. The customer groups that exhibited high frequency of purchase contributed the greatest to firm revenue. Interestingly, these high of frequency purchase groups did not account for the longest or the shortest tenure with the company. Prior studies did not report these differences in customer groups based on frequency of purchase and tenure with the firm. One of the major implications from this research is that the objective of loyalty programmes may need to be rethought for of high frequency of purchase products. Buyers with high frequency of purchase in these product categories may be strongly influenced by habit that may leave them less affected by loyalty programmes. This also implies that it may not be necessary to offer loyalty programmes to high frequency buyers in these categories.
Table of Contents

Acknowledgements .......................................................................................................... iii
Abstract ............................................................................................................................ iv
List of Tables ................................................................................................................... ix
List of Figures ................................................................................................................... x

Chapter 1: Introduction ................................................................................................. 1
  Frequency of buying and tenure.................................................................................... 3
  Tenure vs. retention..................................................................................................... 4
  Subscription vs. non-subscription customers............................................................ 5
  Trade-offs between tenure and frequency of purchase ............................................. 5
  Research context ...................................................................................................... 6
  Research Scope ........................................................................................................ 7
  Outline of the thesis ................................................................................................. 8

Chapter 2: Literature Review and Hypotheses Development .................................... 9
  Customer Loyalty..................................................................................................... 10
    Attitudinal loyalty ................................................................................................. 10
    Behavioural loyalty.............................................................................................. 11
    Behavioural loyalty and frequency of purchase.................................................. 15
  Tenure ..................................................................................................................... 17
  Factors affecting tenure with the service provider................................................. 19
  Is long tenure always desirable? ............................................................................. 23
  High frequency buyers........................................................................................... 25
    High frequency user/buyer group based on volume or how often purchased.... 28
    High expenditure group. .................................................................................... 31
  The Gambling industry ............................................................................................ 34
  High frequency of buying in gambling .................................................................. 37
  High spending in gambling .................................................................................... 40
  Frequency of betting and tenure in gambling ....................................................... 42
  Is tenure more important than frequency of buying? ........................................... 46
List of Tables

Table 1: Forms of loyalty (adapted from Dick and Basu, 1994) ...............15
Table 2: Descriptive statistics of the wagering data (n=11,068) .............57
Table 3: Descriptive statistics of low frequency bettors vs. high frequency bettors groups .................................................................61
Table 4: Correlations among gambling behaviour variables for low frequency bettors (presented above the diagonal) and high frequency bettors (presented below the diagonal) .........................................................66
Table 5: Composition of cluster solution ...........................................70
Table 6: Composition of cluster solutions on the resorted data ...............73
Table 7: Composition of cluster solutions for Set A (50% of the sample)...74
Table 8: Composition of cluster solutions for Set B (50% of the sample)...74
Table 9: Classification results for both analysis and holdout samples ........79
Table 10: Summary of discriminant analysis (analysis sample; n=5,534) ...81
Table 11: Summary of discriminant analysis (holdout sample; n=5,534) ....81
Table 12: Results of multiple regression analysis ...............................85
List of Figures

Figure 1: Benefits of long tenure (customer tenure) (Reichheld & Sasser, 1990) .......................................................... 19

Figure 2: Histogram representation of number of bets (the dashed line represents the median value of number of bets = 114) ........................................... 57

Figure 3: Histogram representation of net loss in dollars............................ 57

Figure 4: Histogram representation of tenure............................................ 58

Figure 5: Population pyramid graph for low and high frequency bettors based on median value split of number of bets made (114)......................... 61

Figure 6: Bar chart showing number of bets and total bet value (i.e. gambling revenue)................................................................. 75

Figure 7: Bar chart showing tenure and total bet value (i.e. gambling revenue)................................................................................... 76

Figure 8: Scatterplot for analysis sample (n=5,534)................................. 82

Figure 9: Scatterplot for holdout sample (n=5,534)................................. 83
Chapter 1: Introduction

The importance of customer retention has been recognized since the 1910s (Tadajewski & Saren, 2009). Orange and McBain in 1911 suggested, in the context of mail-order companies, that customers “... in due course, at a small cost, be persuaded to buy other articles, or ... in all probability repeat his order” (cited in Tadajewski & Saren, 2009, p. 197). Similarly, Thompson (1939, cited in Tadajewski & Saren, 2009, p. 201) stated “… it would seem that a store can assure itself of a very substantial growth if it can reduce the number of customers lost”.

However, a more intense research focus on customer retention began in the 1990s (Agariya & Singh, 2011; Sherrell & Bejou, 2007). Customer retention marketing is now viewed as one of the most important marketing strategies for companies because of its impact on the firm’s financial well-being (Ahmad & Buttle, 2001). The key assumption in customer retention marketing is that it is more beneficial to keep existing customers for a longer period of time than to attract new customers.

It has been claimed that attracting new customers can be up to five times more expensive when compared to the costs of keeping existing customers (Weinstein, 2002). An underlying claim of the customer retention strategy is that long tenure customers transact more often
with the firm, and generate greater revenue for the firm, compared to the short tenure customers (Carroll, 1992; McDougall, 2001).

Retaining customers has been argued to lead to a number of benefits for the companies (Reichheld, 1996; Reichheld & Sasser, 1990). Retained customers are claimed to generate higher profits, result in lower operating costs, are more willing to pay premium prices and can account for free advertising resulting from referrals (Buchanan and Gillies, 1990; Reichheld and Kenny, 1990; Reichheld and Sasser, 1990). Other benefits include more opportunity to learn about customers and the potential to build relationships with customers which make switching brands difficult (Kamakura et al., 2003). This viewpoint has led companies to implement customer relationship management (CRM) programmes that are designed to retain customers for longer period of time (Ahmad & Buttle, 2002; Aksoy, Keiningham, & Bejou, 2008; Bolton, Kannan, & Bramlett, 2000; Kale, 2003; 2005; Dowling & Uncles, 1997; Uncles, Dowling, & Hammond, 2003).

However, the idea that the long tenure customers are always desirable may not be true in all marketing contexts (Smith & Lakhani, 2008). Bhatnagar, Maryott, and Bejou (2008) criticised the evidences provided for the support of the benefits of longer tenure as mostly anecdotal. They argued that formal CRM programmes focusing on customer longevity have failed to improve firms’ profitability. Ward and Dagger (2007) found lengthening tenure did not automatically
make the relationship between the company and customers stronger. East, Hammond and Gendall (2006) report that an increase in customers’ tenure (i.e. retention) with their service providers did not necessarily lead to higher spending when investigated across a variety of industries (e.g. credit card, hairdressers, restaurants, etc.), and across four different countries. Reinartz and Kumar (2000) find that some short duration customers may be as profitable as long duration customers in the mail-order catalogue marketing industry. Reinartz and Kumar also report that some long duration customers may not be profitable because they are not high frequency buyers. However, previous studies fall short of explicitly investigating the association between frequency of buying and customers’ tenure. There remains a general lack of understanding of how these two important variables of customer behaviour are linked to each other, especially in the context of high frequency purchase product categories.

**Frequency of buying and tenure**

It has been recognized that all buyers do not buy with the same frequency (Aksoy et al., 2008; Ward & Dagger, 2007; Weinstein, 2002; Vakratsas & Bass, 2002). One of the often mentioned rules in marketing is the “80-20” rule (Goldsmith, 2000; Weinstein, 2002). According to this rule, 20 percent of the buyers account for 80 percent of the sales. Aksoy et al. (2008) states that up to 80% of a firm’s customers do not generate an acceptable rate of return. Regardless of
whether the “80-20” rule precisely holds across different market conditions, it suggests that a relatively small group of customers (approximately 20%) account for a high proportion of sales. This group of buyers has been variously termed as the “heavy-half”, “heavy user” or “high frequency buyer” (Ehrenberg, 1959; Goldsmith, 2000; Twedt, 1964).

**Tenure vs. retention**

Duration has been defined as how long a relationship has existed between exchange partners (Palmatier, et al., 2006). The exchange partners include the customers and the supplier of the product or service. In the current research, the term “tenure” is used to represent how long a customer has stayed with the brand and/or the service provider. Tenure is calculated as the difference between first day of transaction and the last day of transaction. In this study, tenure is considered different from customer retention. The latter term refers to strategies that the companies use to actively attempt to keep customers for a longer period of time. Whereas, tenure is a measure of how long the customers have been transacting with the company. In other words, retention is a marketers’ goal and tenure is the customers’ time with the company. See Appendix D-1, for a discussion of differences between duration of stay and tenure.
Subscription vs. non-subscription customers

Customers will have different tenures with the firm depending on the relationships between these two exchange partners. In a subscription based relationship, customers may have fixed tenure because of contractual agreements. In a non-subscription relationship, customers are free to leave and buy from other companies which may lead to different tenures. In both cases, customers may have long or short tenures with the firm. The only difference being that in a non-subscription context, the customers can decide when to leave.

Trade-offs between tenure and frequency of purchase

Firms have typically focused on increasing tenure with the assumption that this will lead to high frequency of purchase. However, given the criticisms that this does not necessarily eventuate, the possible trade-offs of frequency of purchase and tenure has not been explicitly investigated previously. The trade-offs concern whether the firm actively works toward keeping the customer for longer period of time (lengthening tenure) vs. strategies to increase frequency of purchase with the firm. For example, do high frequency buyers also have longer tenure with the company? What contributes to generating higher revenues – high frequency of buying or long tenure with the company? Are there different frequency buying and tenure groups that contribute
disproportionate amounts of revenue? The aim of this research is to investigate these questions.

Research context

The current study was conducted in the context of consumer purchase behaviour in the gambling industry. The gambling industry is a largely mature industry where the frequency of purchase tends to be relatively high, with very high participation rate in many populations (Mizerski et al., 2013). The gambling industry is now recognized as the world’s largest revenue generator of all forms of entertainment (e.g. movies, music, etc.) combined (McGowan, 2008, cited in Pinto & Mansfield, 2011; Mizerski et al., 2013). Gross revenue from gambling is projected to reach $512.78 billion by 2015 worldwide (H2 Gambling, 2011). The online gambling earning was estimated to be $30 billion in 2012 (Garlitos, 2013).

According to the productivity commission (2010), Australian consumers spent $19 billion on various gambling products in 2008-2009. By comparison, consumers spent $23 billion on footwear and clothing and $12 billion on alcoholic beverages during the same period. While frequency of gambling is affected by the form of the gambling product, there are few consumer goods that can generate as high a frequency of purchase (Mizerski et al., 2013). In addition, one study found that gamblers self-reported to have been gambling for an
average of over 6 years (Grant and Kim, 2001). This implies that some of the gamblers are likely to have long tenure with gambling service providers. These facts make the gambling industry an ideal industry to investigate the relationships of consumers’ high frequency of purchase and their tenure. A particular interest is to identify which of these two factors (i.e. frequency of betting or tenure) have a stronger association with revenue.

Research Scope

It is important to note that this research does not focus on problem gambling. Previous research suggests that only about 2% of the gambling population appear to suffer from these problems (Productivity Commission, 1999; 2010). The majority of the gambling population do not suffer from problem gambling issues and consider gambling as a form of entertainment. For example, Hashimoto (2010) finds that a vast majority of the casino players are aware that casino offers games that have house advantage, but are still willing to participate in these games. Another study by Narayanan and Manchanda (2012) reported that only about 8% of casino gamblers showed evidence of addiction. Thus, the gambling industry is similar to a service industry that offers a particular form of entertainment, involving risk-taking behaviour.
Outline of the thesis

The next chapter reviews the concepts of frequency of purchase and tenure discussed in the prior literature. This chapter also develops the hypothesis tested in the study. Chapter 3 discusses the data used in the study along with the measures used to test the research hypothesis. Chapter 4 provides the results of the hypothesis tests, while chapter 5 provides a summary and the conclusion of the research findings.
Chapter 2: Literature Review and Hypotheses Development

The concepts of tenure and frequency of purchase have been mostly researched in the context of customer relationship management (CRM). There is no single definition for what CRM actually is (see Baran, Galka & Strunk, 2008, for a list of definitions). However, CRM can be seen as a program instituted by companies to increase brand value by recognizing the best customers, retaining them, encouraging increased usage of the company’s product or service (Baran, Galka & Strunk, 2008).

Kutner and Cripps (1997, cited in Baran, Galka & Strunk, 2008) suggested that CRM is based on four tenets. First, customers should be managed as important assets. Second, not all customers are equally desirable. Third, customers vary in their needs, preferences and buying behaviour. Fourth, by better understanding their customers, companies can tailor their offerings to maximize overall value. Ultimate aim for the CRM programmes is to engender loyalty among groups of customers by keeping them for a longer period of time (Aksoy et al., 2008). This notion rests on the idea that keeping customers for longer period of time will encourage repeat purchases and consequently greater profits.
Customer Loyalty

The concept of loyalty and retention are very close. A group of loyal customers are assumed to be retained customers (Reichheld & Sasser, 1990). It is assumed that the longer a customer stays with the company the more likely they are to buy from that company. East et al. (2005) argued that repeat patronage and retention are essentially the same behaviours which are only separated by temporal differences. Specifically, higher repeat purchases lead to greater retention. There is no one accepted definition for loyalty that scholars have agreed on (Baran et al., 2008), or on how loyalty should be measured (East et al., 2005; Rundle-Thiele & Bennett, 2001; Rundle-Thiele & Mackay, 2001). However, researchers agree there are two types of loyalty - attitudinal loyalty and behavioural loyalty (East et al., 2005; Widing et al., 2003).

Attitudinal loyalty

Attitudinal loyalty refers to a deeply held commitment toward the brand or product/service that lead to repurchase behaviour. Proponents of this form of loyalty believe that loyalty cannot be measured by simply observing what is being continually purchased. It is also necessary to understand what the buyer feels toward that brand or product/service (Baran et al., 2008). In other words, the interest is in finding out if the buyer has greater ‘liking’ and ‘preference’ for the
brand or product/service (Baran et al., 2008; Widing et al., 2003). Rundle-Thiele and Mackay (2001) note that altitudinal loyalty is usually measured using self-report surveys. Self-report surveys of behaviours have been criticized for lacking reliability due to problems of recall bias (Baumeister et al., 2007; Nisbett & Wilson, 1977; Woodside & Wilson, 2002) and socially desirable responses (Steenkamp, de Jong & Baumgartner, 2010). This thesis uses real-time gambling data which are assumed to be free of these deficiencies of self-report surveys. This is described more fully in Chapter 3.

**Behavioural loyalty**

Behavioural loyalty refers to a buyer’s consistent repurchase of a brand or product/service (Widing et al., 2003). For example, a complete and consistent repurchase of the same brand would show perfect behavioural loyalty although this seldom happens in reality (Widing et al., 2003). Reichheld, Markey and Hopton (2000) suggest that companies should focus on repurchase loyalty by identifying the customers who frequently purchase from a company. This suggests that focusing on the frequency of purchase is crucial to understanding consumers’ behavioural loyalty. Behavioural loyalty can be measured in three ways (Widing et al., 2003):

- **Proportion of purchase**: it is also called Share of Category Requirement (SCR). Here loyalty is measured as a percentage –
the number of times the most frequently purchased brand is purchased divided by the total purchases. Thus, if a brand has high SCR for a certain customer, then it can be concluded that that customer shows high brand loyalty for that brand (Dawes, 2013; Dawes et al., 2015). For example, Dawes (2013) argued that SCR can measure the level of loyalty for private label brands.

- **Sequence of purchase**: here the consistency with which consumer switches between brands is considered. This is different from SCR as SCR measures proportion of category requirement whereas sequence of purchase measures how often one switches brands. If consumers switch often among available brands then consumers are said to be showing low behavioural loyalty. For example, Brown (1953) defined four different levels of loyalty based on the frequency of purchase (cited in East et al., 2008). An ‘undivided loyalty’ is defined when a consumer purchases the same brand six times in a row. However, if the consumer made six purchases but switched between two brands during these purchases then the consumer is said to have ‘divided loyalty’. If the consumer made three purchases of the same brand in a row and then switched to a different brand, then he is said to have ‘unstable loyalty’. Lastly, if the consumer bought brands in any other different patterns, then he is said to have ‘no loyalty’.
However, these cut-off points are arbitrarily defined and are yet to be empirically supported.

- **Probability of purchase:** here proportion and sequence measures are combined to compute the probability of future purchase of a specific brand (Widing et al., 2003). Thus, the probability of future purchase tends to be positively related to their frequency of purchase. This measure deals with future probability of purchases rather than quantify past behaviour (Iwasaki & Havitz, 1998).

Iwasaki and Havitz (1998) considered three additional measurement for behavioural loyalty. These are duration, frequency of purchases and intensity. Duration is the long-term length of time spent as a participant, patron or user. Frequency is the number of purchases, uses or participation over a period of time (e.g., week, day, month, or year, etc.). Intensity is the number of the hours over a period an individual devotes to using, participating in or purchasing a product or service.

However, it should be noted that the concept of intensity can be encompassed within the concept of frequency. For example, Barwise, Ehrenberg and Goodhardt (1982) showed that the number of hours that people watch television can be very steady over a number of days. People who watch television on one day have a tendency to view television in the next day. Thus, the number of times a person watches
television (frequency) will be closely associated with the number of hours spent watching television (intensity) each day. In the online gambling industry context, Jolley, Mizerski & Olaru (2006) used number of sessions played (intensity) and number of bets made (frequency) to create habit-strength variable and showed that this new variable can predict gamblers’ retention. Thus, Iwasaki and Havitz’s three additional measures of loyalty actually represent only two measures of behavioural loyalty – frequency purchases and tenure. Note that Jolley at al. (2006) defined session as an event whose beginning was when a player logged in and ending was when that player logged out. They also defined bet as the number of times the player placed a bet by hitting the spin button.

Dick and Basu (1994) developed a four-segment typology of loyalty by combining relative attitude and repeat purchasing behaviour (see Table 1). According to this typology, ‘true loyalty’ toward a brand occurs when consumers score high on both relative attitude and repeat purchase. ‘Spurious loyalty’ occurs when relative attitude is low but repeat purchase is high. ‘Latent loyalty’ occurs when relative attitude is high but repeat purchase is low. And, lastly, ‘no loyalty’ exists when both the relative attitude and repeat purchase is low. Dick and Basu’s model, although useful, has been criticised for lacking the insight of the dynamic process underlying the development of loyalty among customers (Iwasaki & Havitz, 1998). Furthermore, it fails to define the
level of the repeat purchase necessary to generate loyalty of any kind. An empirical study by East et al. (2005) showed that a combination of attitude and repeat patronage failed to predict loyalty behaviour. East et al. went on to question the useful of such typology for real world applications.

Table 1: Forms of loyalty (adapted from Dick and Basu, 1994)

<table>
<thead>
<tr>
<th>Repeat purchase</th>
<th>Relative attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>High</td>
<td>True loyalty</td>
</tr>
<tr>
<td>Low</td>
<td>Latent loyalty</td>
</tr>
<tr>
<td>Low</td>
<td>Spurious loyalty</td>
</tr>
<tr>
<td></td>
<td>No loyalty</td>
</tr>
</tbody>
</table>

**Behavioural loyalty and frequency of purchase**

Pure behavioural loyalty does not consider attitudinal factors. A consumer may continue to buy from the same seller even if he does not like or prefer that seller but may buy the same brand simply out of habit or convenience, without thinking too much about it (Liu-Thompkins & Tam, 2013; Widing et al., 2003). It is important to note here that several scholars have argued that behavioural loyalty is related to purchase habit. Lindquist and Sirgy (2009) mention that ‘consumer loyalty, because it takes the form of repeat purchasing, is directly linked to the formation of purchase habits’ (p. 130). East et al. (2008) also emphasize that ‘...most patterns of purchase, including
Reichheld et al. (2000) asserts that companies should focus on continued purchases by the buyer. This implies that companies should focus their attention to understanding the role of frequency of purchases in order to understand behavioural loyalty. Previous researches have shown that past frequency of purchases can be a good predictor of future purchases (Lam & Mizerski, 2009; Uncles et al., 1995).

Opperman (2000) also suggests that loyalty researchers should consider focusing only on the behavioural measures since measuring attitudes over a long time period is difficult and impractical. O’Malley (1998) asserts that behavioural measures ‘provide a more realistic picture of how well the brand is doing vis-à-vis competitors, and the data generated facilitate calculation of customer life-time value, enhance prediction of purchase probabilities, and assist in developing cost-effective promotions’ (p. 49).

It is also possible for attitude to change in a short period of time but not easy to change habit-based behaviour that has been learned over a long period of time. Both O’Malley (1998) and Opperman (2000) suggest behavioural measures, such as frequency of visits, amount spent and purchase sequences, should be used to understand repeat purchase behaviour of customers. The main argument, here, is that
individuals’ purchase behaviour can be understood by analysing their purchase patterns.

However, not all customers may be desirable for companies because of differing buying behaviour (Kutner and Cripps, 1997; cited in Baran, Galka & Strunk, 2008). Prior research have found that customers vary in terms of frequency of purchase (Goldsmith, 2000; Pleshko & Al-Houti, 2011) and also in terms of their tenure with the service providers (Reinartz & Kumar, 2000). The rest of the literature review discusses the frequency of purchase and tenure in detail. It also develops the hypothesis to be tested in the research.

**Tenure**

As noted previously in chapter 1, the importance of tenure as a major factor in relationship marketing was recognized as early as the 1910s (Tadajewski & Saren, 2009). In relationship marketing, customers are considered assets for the firm who contribute to the value of the firm (Gupta, Lehmann & Stuart, 2004). The emphasis in relationship marketing is on lengthening the tenure of the buyers rather than acquiring new customers (Evans, 2002; Weinstein, 2002).

Lengthening the buyers’ tenure is viewed as an easier and cheaper way to increase company’s value compared to customer acquisition (Ahmad & Buttle, 2001; Bolton & Tarasi, 2006). For example, it has been reported that the cost of acquiring a customer can run up to five times
more than the cost of retaining a customer (Baran et al., 2008; Pfeifer, 2005; Weinstein, 2002).

The principal argument for lengthening the tenure of buyers is that the longer the customer stays with the company, the more often he or she is assumed to continue buying the company’s products and services (Reichheld & Kenny, 1990; Reichheld & Sasser, 1990). This assumption has led several researchers to argue for developing long-term relationships with customers as a means to stay profitable (Ahmad & Buttle, 2001; Evans, 2002; Morgan & Hunt, 1994; Sheth & Parvatiyar, 1995). Long-term customers are reported to provide a number of benefits (see Figure 1) that include higher profits, lower operating costs, ability to charge premium prices and free word of mouth advertising from referrals from existing buyers to new buyers (Buchanan & Gillies, 1990; Reichheld & Kenny, 1990; Reichheld & Sasser, 1990).

For example, Buchanan and Gillies (1990) argue that a high retention rate not only implies high loyalty but also higher profits for the firm. Dawkins and Reichheld (1990) report that a five percent increase in customer retention can increase a company’s profits by 25 percent to 85 percent. Similarly, Gupta et al. (2004, p.7) find that “firm’s financial value can rise by almost five times when 1% improvement in customer retention is made when compared to 1% change in discount rate”. In addition, a decrease in customer retention rates can indicate
potential problems in service quality and customer satisfaction (Dawkins & Reichheld, 1990). This emphasis on retaining customers on a long-term basis has led firms to institute customer loyalty programs (Dowling & Uncles, 1997). For example, casinos have increasingly adopted customer retention programmes as part of their overall marketing strategy (Kale, 2005; Loveman, 2003).

**Figure 1**: Benefits of long tenure (Reichheld & Sasser, 1990)

Factors affecting tenure with the service provider

In order to understand long tenure (i.e. longer retained) customers, past research has investigated the factors that affect customer’s tenure (i.e. customer churn). To avoid losing valuable customers, companies with appropriate data attempt to identify and monitor their customers’ purchase behaviour to predict ‘customer churn’ (Jahromi et al., 2010). Customer churn refers to the likelihood that customers are likely to stop transacting with the service provider. Low customer churn implies
that the majority of the customers have a relatively long tenure with the service provider (i.e. they continue to stay with the service provider).

In a contractual setting, up- or downgrades of memberships can indicate the likelihood of churn. For example, Bhattacharya (1998) reported that when museum members opted to downgrade their membership, it indicated a higher likelihood of their discontinuing membership with the institution in the future. In a non-contractual setting, companies generally rely on variations in purchase data as an indication for discontinuation of a relationship with the service providers (Chen & Hitt, 2002; Lopez et al., 2006; Zorn et al., 2010). For example, a decreasing trend in purchase volume (e.g. purchasing fewer number of items) by the buyers can indicate potential for future customer churn (Jahromi et al., 2010).

Other studies, using self-report surveys, have reported non-behavioural variables that can influence customer churn. For example, Bolton (1998) found that customers’ reported high satisfaction levels is related to longer duration of stay. Verhoef (2003) reported that self-reports of affective commitment, payment equity (i.e. whether the customers perceived that the price charged was fair or not) and loyalty programs increased customers’ tenure with the service provider. On the other hand, reported inconvenience and inadequate responses to the customer service failures increase the probability of customers’ leaving
the firm, leading to shorter tenure with the service providers (Keavney, 1995).

A limited number of studies have used actual behavioural data to investigate customers’ tenure. This has probably occurred because of the lack of availability of real-time and actual data. These studies have mostly investigated customers’ tenure with service usage (e.g. fewer banking transactions, number of contacts made to customer service units, telephone service usage per hour, etc.). The number of exchanges between the customers and service provider represent the usage frequency (Lopez et al., 2006).

Wieringa and Verhoef (2007) analysed customers’ reported behaviour with survey data for an energy company in the Netherlands. They reported that high energy usage rate was related to lower switching intentions implying that the high frequency users were less likely to stop transacting with the energy company (i.e. consequently have longer tenure). Lopez et al. (2006) also found that higher service usage of fixed telephony services lengthened tenure with the service provider. In addition, higher expenditures can lead to longer tenure with the service provider. For example, using actual usage data for direct-to-home satellite TV services, Jamal and Bucklin (2006) reported that customers with higher expenditures tended to stay longer with the service provider.
On the other hand, Neto et al. (2011) reported that, in the banking industry, customers with very few transactions (i.e. low frequency users) are most likely to have shorter tenure. Specifically, customers making fewer deposits in their checking accounts were more likely to stop transacting with the bank. Chen and Hitt (2002) reported that higher volume of website usage was negatively correlated with switching and attrition in the online brokerage industry. These studies suggest that increased usage lead to longer tenure.

However, contradictory results were reported by Zorn, Jarvis and Bellman (2010). They examined an online movie rental website and found that an increase in number of visits and payment per month was associated with high churn rate. In other words, higher frequency of usage (i.e. number of visits) and higher expenditures were associated with shorter tenure by the website visitors. One explanation could be that the high frequency visitors are more knowledgeable about offers from alternative movie rental sites and subsequently switched to these alternative sites.

East et al. (2008) have suggested that high frequency users of one brand tend to be high frequency users of other available brands as well. Another possible explanation could be that the negative relationship between number of visits to the website and expenditures with tenure is moderated by product category. It is possible that products that are
available both online and offline may show this relationship when compared to the products that may only be available online or offline.

**Is long tenure always desirable?**

The benefits of long tenure have been challenged in recent research (East, Hammond & Gendall, 2006; Reinartz & Kumar, 2000). The main argument for the benefits of long duration customers assumes that these customers not only stay longer with the firm but also buy frequently resulting in higher revenues (Reichheld & Kenny, 1990). However, several studies have shown that this relationship between tenure and revenue is weak. For example, East et al. (2006) report that an increase in tenure length did not necessarily lead to higher spending for a number of industries (e.g. hairdresser, supermarket, car servicing, coffee shop, dry cleaning, fashion store, etc.) when examined across four different countries (i.e. UK, USA, Mexico and Mauritius).

However, they did not investigate frequency of buying in their study. Dowling and Uncles (1997) note that not all long duration customers may be financially beneficial for the firm, especially if this group of customers expect to receive rewards (or discounts) in return for their long stay. They also suggest that a majority of the buyers of a brand tend to be buyers of other brands as well (polygamous loyalty), and may not generate enough revenues to cover the costs of serving them. East, Lomax and Narain (2001) found that recently acquired customers
recommended more than the long-term customers in the supermarket and hairdressing industries. Dawes (2009) reported that long tenure customers in insurance industry may not necessarily pay higher prices. However, they do tend to be less sensitive to price increases.

Other research suggest that the short tenure customers may be as profitable as the long tenure customers. For example, Reinartz and Kumar (2000) find a short tenure consumer segment (based on a median split) of a catalogue marketing company can generate similar amounts of life-time revenue, profit and are willing to pay higher prices compared to the long tenure customer segment. They report that these short tenure customers are high frequency purchasers and therefore generate higher profit in a shorter period of time.

In another study, Payne and Frow (1997) report that the highest profitability comes from the customer segment with the lowest retention rate in the utilities sector in the United Kingdom. Although they do not report the usage figures, it can be assumed that this segment accounts for the highest usage (hence their highest profitability in a short period of time).

Coyles and Gokey (2005) suggested that managing the buying patterns of existing customers may be more important than focusing on their retention or increasing their tenure. For example, they report that banks lost 3% of their total balance when 5% of their customer
defected but lost 24% of total balance when existing customers reduced their frequency of transactions by 35%.

Defection implies that customers switch brands so there is no purchase whereas reduction in purchase means reducing frequency of purchases (Hammond & East, 2003). This difference can have different consequences on revenues that firm earn. These findings suggest that it may be more important to focus on increasing frequency of transactions (i.e. frequency of usage) of existing customers rather than increasing these customers’ tenure. One of the aim of this thesis is to investigate the relative importance between frequency of purchase and tenure in a high frequency purchased product category.

Previous research has found that there are different groups of customers who exhibit different frequency of buying (Goldsmith, 2000). Specifically, some customers are frequent buyers of products and brands, whereas others tend to be infrequent or low frequency buyers. Businesses usually attempt to target the high frequency buyers because of the potential of generating higher revenues from this group of buyers.

**High frequency buyers**

Repeat buyers have always been the desired target market for marketing managers (Petrick, 2004). There is continual and considerable interest in identifying and understanding consumers who
are frequent buyers (volume) or big spenders (Goldsmith, 2000; Shani et al., 2010). Collectively they have been called the “heavy users”. The high frequency user group is particularly attractive for businesses because of the higher revenue potentials.

There is no formal definition for the heavy user or high frequency buyer group. Pleshko and Al-Houti (2011) suggest that buyers may be classified as heavy users if they buy and/or spend more than the other customers. The following definition for the high frequency group is used in this dissertation – ‘the high frequency group is responsible for a disproportionate share of product usage or expenditures for its group size’. For example, if 10% of the consumers are responsible for 70% of the product usage, then these 10% of consumers could be classified as the ‘heavy users’ or the ‘high frequency group’.

However, there is no single, universally accepted way to define this group. In the past, researchers have arbitrarily defined this group. For example, the high frequency group has been identified based on volume of purchase (Ehrenberg, 1959; Twedt, 1964) or how often customers purchase (Goldsmith & d’Hauteville, 1998; Goldsmith & Litvin, 1999; Woodside, Cook & Mindak, 1987). Other academics have researched high and low spending groups by segmenting customers based on the amount spent on purchases (Mok & Iverson, 2000; Shani et al., 2010; Spotts & Mahoney, 1991).
Ehrenberg (1959) was the first to mention the high and the low frequency buyers for fast moving consumer packaged goods such as bread, cereals, coffee, soaps, detergents, etc. Twedt (1964) reported that for many consumer packaged products a small percentage of households account for a large share of purchases. Using a consumer panel of Chicago Tribune newspaper subscribers, he reported that 39% of the households purchased 90% of the cola soft drinks and that 48% of the households bought 87% of ready-to-eat cereals. He termed this group the ‘heavy half’.

Over time other scholars have used the term ‘heavy user’ to identify this group (Goldsmith, 2000; Litvin, 2000). Twedt’s (1964) results were compared by Cook and Mindak (1984) using a different sample almost two decades later. They found that, for the same product categories studied by Twedt, the findings related to the ‘heavy-half’ group were strikingly similar. It is interesting to note Twedt (1964) failed to find any demographic differences that could explain the different purchasing behaviour. Cook and Mindak (1984), however, did not investigate the demographic differences between the two groups.

The rest of the literature review is presented as follows. The reported findings related to the high and low frequency groups are presented followed by discussion of the reported studies of high and low frequency groups from the gambling sector – the industry context used in this study.
High frequency user/buyer group based on volume or how often purchased

Studies investigating high frequency buyers based on purchase volume or how often customer purchases, seek to identify individuals who account for a large amount of the purchase or usage frequency of products and services. A majority of this research focus on the demographic characteristics in an attempt to distinguish high frequency buyers from the low frequency buyers and/or non-users of products and services.

However, these research findings are not unanimous on whether demographic variables can distinguish the heavy users from the light users. For example, Goldsmith, Flynn and Bonn (1994) did not find demographic differences, in the United States, for heavy users (15% of the sample) of travel agencies when compared to the light and non-users.

However, in a later study, Goldsmith and Litvin (1999) found age and income distinguished the heavy users of travel agencies from the light and non-users. This later study was conducted in Singapore and thus the results can perhaps be related to the socio-cultural differences. This is supported by a follow up study by Litvin (2000), who find that heavy travellers in Singapore are older, married, have higher average income and education levels. Woodside, Cook and Mindak (1987)
report similar findings for the heavy traveller segment in the United States.

In the case of wine consumption, gender and age have been found to be weakly related to the heavy users of wine (i.e. drinking and buying) (Goldsmith & d’Hauteville, 1998). There have been doubts about the usefulness of demographic variables as market segmentation variables (Yankelovich, 1964). For example, Hammond et al. (1996) analysed 20 different product categories in four different countries and reported that the competitive brands could not be distinguished based on different socio-demographic segments. In other words, all brands in the market attracted similar socio-demographic segments. In another study involving the casino market, Park et al. (2002) reported that behavioural characteristics of the casino patrons could not be differentiated in terms of demographic and socio-economic variables such as age, gender, income, marital status, residence, and education.

As mentioned earlier, some studies have identified demographic differences while other studies have not found significant differences between high and low frequency groups. However, among the studies that have reported the differences, there is no agreement on the list of demographic variables, if any, that can accurately help in describing high and low frequency buyers. This implies that there are no common demographic characteristics that may be influencing the frequency of buying of the customers. Customers are more likely to be influenced by
need of the product which may have no direct association with the demographic characteristics.

In addition, the above studies have reported the presence of psychographic differences between heavy and light users. For example, high frequency buyers (e.g. users, travellers) were reported to be more innovative, more knowledgeable about the product or service, more involved with the purchase decision and tended to be opinion leaders (Goldsmith & d’Hauteville, 1998; Goldsmith, Flynn and Bonn, 1994; Goldsmith & Litvin, 1999; Litvin, 2000). However, the list of psychographic variables researched varied across the different studies, making it difficult to identify the common psychographic variables.

The usefulness of psychographic segmentation, has also been questioned. For example, Yankelovich and Meer (2006) state that, ‘Psychographics may capture some truth about real people’s lifestyles, attitudes, self-image, and aspirations, but it is very weak at predicting what any of these people is likely to purchase in any given product category. It thus happens to be very poor at giving corporate decision makers any idea of how to keep the customers they have or gain new ones’ (p. 124). Dibb and Simkin (2008) argue that although variables such as benefits, lifestyle, etc. may be more meaningful, they are unobservable and there are difficulties in collecting such information over a long period of time.
The above discussion suggests that using demographic and psychographic variables may not be reliable for understanding behaviours of high and low frequency groups as they differ across studies. As Goldsmith (2000) states, demographic and psychographic descriptions of the high and low frequency groups do not allow businesses to develop marketing strategies beyond developing media strategies.

In addition, Legoherel (1998) argued segments based on intention to purchase tend to be unreliable because of the discrepancy between reported intention to purchase and reported actual purchase behaviour. For instance, evidence from the tourism research suggest that there are differences between what the tourists state that they intend to spend and the amount they actually spend (Legoherel, 1998). Instead of measuring intentions, as O’Malley (1998) and Opperman (2000) suggest, behavioural measures such as frequency of visits, amount spent, etc. should be used to understand purchase behaviour of customers. In summary, these studies emphasize on the use of behavioural measures over the demographic and psychographic measures.

**High expenditure group.** Previous studies have also investigated high and low frequency buyer groups based on the amounts the customers spent in buying products and services. In effect, these studies have focused on identifying how the high spenders are different from the
low spenders. For example, Smallwood and Weiner (1987) investigated how high spending fashion catalogue shoppers (defined as purchasing more than $100 during the past 12 months) differed from low spending fashion catalogue shoppers (defined as purchasing less than $100 during the past 12 months). Their analysis revealed that high spending fashion catalogue shoppers were from higher income group and performed the role of opinion leaders.

It should be noted that Smallwood and Weiner’s conceptualization of high spending and low spending fashion catalogue shoppers were based on arbitrary cut-off points of the expenditures. They also did not report on how often (i.e. frequency) the shoppers purchased in a given year or how long they have been buying from the catalogues (i.e. tenure). For example, high spending fashion catalogue shoppers may be buying from the catalogues for a long period of time which may partially explain their high spending behaviour.

In another study, Tsiotsou (2005) found ski resort visitors’ experience, satisfaction and income differentiated between high frequency visitors (visiting ski resort on a weekly basis) from the low frequency visitors (visiting ski resort on a monthly basis). The study reported that contrary to expectations, low frequency visitors had higher experience and satisfaction compared to the high frequency visitors. In other words, high visit frequency did not necessarily lead to higher experience or satisfaction. The result could be explained by the fact
that as low frequency visitors visit less often, each visit may be more memorable leading to self-reporting of higher satisfaction. It should be noted that Tsiotsou (2005) study did not investigate the expenditures made on these ski trips by the visitors or how long they have been visiting the ski resort (i.e. tenure).

As expected, the heavy spender group accounts for a disproportionate amount of expenditures compared to the light spenders (Goldsmith, 2000). For example, Spotts and Mahoney (1991) found that 78% of the total camping trip expenditures were accounted for by the heavy spenders group (i.e. the top one-third of the sample based on total expenditures). Similar results were found in the tourism industry research (Shani et al., 2010; Jang et al., 2001; Mok & Iverson, 2000; Craggs & Schofield, 2009). As perhaps expected, these high spending group have also been found to stay longer (i.e. have higher tenure with the service providers) (Jang et al. 2001; Shani et al., 2010). However, as previously noted the assumption that lengthening the tenure with the service provider will lead to higher frequency of buying and consequently higher spending have been challenged (East et al., 2006; Reinartz & Kumar, 2000).

The focus of the current study is on the relationship between frequency of buying and tenure with the service providers. Do high frequency buyers also have high tenure with their goods or service providers? What contributes to generating higher revenues – high frequency of
buying or long tenure with the service provider? Are there different frequency buying groups who contribute different amounts of revenue? These questions are investigated in the context of a highly frequent purchased product category like gambling.

The Gambling industry

Gambling is now recognized as the largest entertainment service industry in the world with revenue greater than both music sales and movies combined (Mizerski et. al., 2013; Pinto & Mansfield, 2011). Global gambling industry revenue was estimated to be $440 billion for 2013 according to a report released by the *The Economist* (Bergman, 2014). The same report found that the Australians spent the most of any nationality on gambling. In Australia, consumers spent around AUD19 billion, or an average of 3.1 percent of household consumer expenditures on gambling products in 2008-09 (Productivity Commission, 2010). A popular form of gambling in Australia – the Electronic Gambling Machines (EGMs), contributed AUD11.9 billion to an AUD19 billion industry in 2010 (Productivity Commission, 2010).

Another major form of gambling, the online or Internet gambling has also experienced a large increase in participation and revenue (Coussement & de Bock, 2013). Worldwide Internet gambling revenue was estimated to be US$29.9 billion in 2010, of which sports betting
accounts for US$12.06 billion (H2GC, 2011). The Internet gambling sector is seen as a lucrative one with large investment banks such as Goldman Sachs, Merrill Lynch, Morgan Stanley, and Fidelity owning hundreds of millions of shares of online casinos and sports betting companies (Richtel, 2005). Wagering, another form of popular gambling product, comprised 15 percent of the total gambling expenditures in 2008-09 in Australia (Productivity commission, 2010). By one estimate, Australia leads the world in terms of per person gambling where online sports betting and gambling has been growing over 30% annually (Yahoo finance, 2014).

Han et al. (2012) found that retail sales of food and beverages were more significantly related to lottery sales than any other product categories such as apparel, furniture, building materials and automobiles. Compared to other consumer product categories such as milk and bread, the product category of gambling is characterised by the highest average frequency of purchases (Mizerski, 2013). For example, typical hamburger products are bought 3.33 times and tuna is bought 2.64 times over a four-week period.

Similarly, coffee and detergents are bought 2.7 and 2.8 times a year respectively (East et al., 2008). In contrast, Mizerski et al. (2004) found that the self-report of lottery purchases in the U.S. state of Florida averaged 14 games over two weeks (or average of 364 times per year). Lam and Mizerski (2009) analysed productivity commission
(1999) data on gambling and found that the average purchase of gambling products in Australia was 55.9 times per year. High frequency of gambling play is also evident in online poker play. Fiedler (2013) reported that frequent online poker players played an average of 92 sessions in a six-month period. These figures suggest that gambling is a very high frequency purchased product.

Bettors not only buy gambling products in high volume but they also play for longer time period. For example, a self-reported survey result of gamblers found that the mean length of time gambling can be up to 6.3 years (Grant & Kim, 2001). Even though the study was not conducted for any specific gambling service provider, it does suggest that bettors are likely to have longer tenure with their gambling service providers.

The gambling industry is also considered to be a mature industry (Productivity commission, 2010) where the gambling participation is very high. As mentioned previously, Australians spent around $19 billion on gambling products in 2008-09 (Productivity commission, 2010). The relatively long tenure of gamblers coupled with highly frequent purchase and consumer spending on gambling products, makes the gambling an ideal industry to investigate the association between frequency of purchase and tenure with the gambling service provider.
High frequency of buying in gambling

A limited number of researchers have investigated the high frequency user concept in the gambling context. Note that the term “frequency of betting” is used when referring to the purchase frequency in gambling. Griffiths and Bingham (2002) reported that high frequency bingo players (i.e. self-reported playing bingo three or more times a week) were more likely to play fruit machines (or EGMs) and less likely to play scratch cards compared to the light bingo players. However, apart from stating this finding, this particular study did not offer much insight into the behaviours of high and low frequency bingo players.

McConkey and Warren (1987) used discriminant analysis to profile the high frequency lottery purchasers (defined as buying 12 times or more a year), the low frequency lottery purchasers (defined as buying 1-11 times a year) and the non-purchasers of lottery tickets. Based on the analysis of survey data, the researchers reported that high frequency lottery buyers were urban based, had higher household income and were in higher age group compared to the low frequency lottery purchasers. On the other hand, the low frequency lottery purchasers were less-traditionalist, generally more optimistic about the future and were younger than the high frequency lottery purchasers. This study did not investigate the tenure or its association with frequency of betting or betting expenditures.
Mizerski et al. (2004) investigated different types of state lottery games in the United States. The data represented lottery purchase from nine-wave quarterly phone surveys. The researchers created a light lottery playing group (purchased 1 to 5 games in last 14 days) and a heavy lottery playing group (purchased 6+ games in last 14 days). They report that the observed purchase frequency of lottery products fit the Negative Binomial Distribution (NBD) for both the light and heavy buyers. Importantly, the study found that the heavy lottery buyers (ranged from 59.79% to 48.79% of the total lottery buyers) account for 93% to 83% of the ‘Lotto’ games, 90% to 75% of the ‘Cash-3’ games and 73% to 83% of the ‘Instant’ games purchased over a 9-waves of surveys.

An important question to investigate is whether high frequency bettors (i.e. high frequency of gambling product buyers) can also be betting on more days compared to the low frequency bettors (i.e. low frequency of gambling product buyers). Previous research suggests high frequency buyers tend to be regular buyers and not random buyers of goods (Vakratsas & Bass, 2002). Regular buyers are buyers who tend to buy brands at regular intervals whereas random buyers buy brands at irregular intervals.

By analysing the scanner data of household purchases, Vakratsas and Bass (2002) reported that ‘average regular buyers purchase more frequently than random buyers’ (p.124) and that ‘...random buyers
appear to buy consistently less frequently than regular buyers...’ (p. 127). This implies that the high frequency buyers not only buy more often but also on more days compared to the low frequency buyers. Therefore, it can be expected that the high frequency bettor not only places greater number of bets, but also tend to bet on more days. Number of bets per day is the number of times a bettor places bets with the betting service provider. This is considered as the frequency of betting in the current study.

**H1: Gamblers’ frequency of betting is positively related to the number of days they bet.**

It is important to note that the number of betting days may be different to tenure even though both are measured in days. It is not necessary for bettors with a long tenure to also have a high number of betting days. For example, a bettor may bet only two times in a year (e.g. on the first and last days of the year) which means the number of betting days of 2 and tenure of 1 year. Another bettor may bet every day in a year with number of betting days of 365 and tenure of 1 year.

Therefore, how often a bettor bets is not the same as tenure. It should also be noted that a high number of betting days may not necessarily suggest high frequency of betting. For example, a bettor may place 100 bets in 100 betting days (in a year) but another bettor may place 300 bets in 100 betting days (in a year). In this example, the first bettor is a
low frequency bettor and the second bettor a high frequency bettor even though both have the same number of betting days and tenure.

**High spending in gambling**

Boldero, Bell and Moore (2010) conducted a latent class cluster analysis on patterns of 12 different varieties of gambling games (e.g. cards, scratchies, lottery, bingo, etc.) and identified six different latent cluster groups from the sample. One of the latent clusters was identified as the ‘heavy gambler’ group who had the highest probability of playing greater variety of games on offer. The size of this group was also the smallest of the six groups identified. Boldero et al. reported that this ‘heavy group’ not only gambled most frequently but also spent the most. But this study did not report on tenure for any of the gambling groups, for example, did the heavy gambling group also bet for longer period of time.

Moufakkir et al. (2004) analysed the high spending group on non-gambling gaming products and services by dividing their sample into three groups based on the amount of non-gambling expenditures made by the visitors to casinos. It is important to note that Moufakkir et al. did not investigate expenditures on gambling products but on *non-gambling* products such as food, beverage and rooms offered to the casino patrons. They created three groups of spenders – the ‘light spenders’ (defined as the lower third of the spending variable), the
‘medium spenders’ (defined as the middle third of the spending variable) and the ‘heavy spenders’ (defined as the upper third of the spending variable). Among these three groups, the heavy spender group of bettors were reported to be younger, more likely to be employed and affluent compared to the other spending groups.

Moufakkir et al. (2004) found that the heavy spender group (i.e. the upper third of the spending group) of non-gambling casino products accounted for 90.5% of the total non-gambling expenditures of the entire sample. This group reported that they spent the most on food, beverages and other amenities offered by the casinos. The research found that the heavy spenders on the non-gaming products and services do not necessarily gamble heavily on casino products and that this group showed no loyalty toward any particular casino.

In fact, the high spending group indicated that visiting casinos were not their main priority. This is perhaps not very surprising since the research focused on non-gambling related expenditures. However, the researchers also found that this high spending group on the non-gaming products and services did not necessarily report spending more hours in the casinos than the low and medium spending groups. The main limitation of this research is that it investigated the individuals who spend on non-gambling products and services associated with casinos rather than on the gambling products. Clearly, the reasons for spending
on gambling products would be different to reasons for spending on non-gambling products.

Hansen and Rossow (2008) found that very frequent gamblers (those who gambled more than 206 times a year) accounted for high slot machine expenditures. This study was conducted on adolescents group and used self-report surveys to determine the frequency of gambling. LaBrie et al. (2007) reported that the self-reported frequency of betting (number of bets made) on an Internet gambling website was positively related to the reported total bet amount. In other words, prior research found that the frequency of gambling was positively related to the amount of gambling expenditures. Accordingly, the following hypothesis is proposed:

\[ H2: \text{Gamblers' frequency of betting is positively related to their total betting expenditures.} \]

**Frequency of betting and tenure in gambling**

The association between frequency of betting and tenure has largely remained under-investigated in gambling literature. Fiedler (2012) reported that gamblers' tenure (defined as the difference between last observed bet date and first observed bet date) was negatively correlated to frequency of playing for an online poker gambling site. This implies that high frequency gamblers are more likely to have
shorter tenure. This is similar to the results reported by Zorn et al. (2010).

Fiedler (2012) further investigated the top 10%, 5% and 1% (in terms of playing frequency) of the online poker bettor groups. He reported that gambling behaviours of these groups were significantly different from the top 50% (median) group of players. For example, the top 1% group of players accounted for higher playing volume (defined as the amount of money a player has paid to the online poker operator in the 6-month observation period), played at greater number of poker tables and played for longer period of time compared to the top 50%, 10% and 5% groups.

Fiedler also reported the tenure for these gambling groups but failed to expand on their significance. For example, he found that the top 50% group has tenure of 27 days, whereas the top 1% has tenure of 182 days. However, he did not link this reported tenure with the gambling behavioural variables for the different gambling groups. It is unclear if the gambler’s longer tenure is related to higher values on the gambling behavioural variables.

Is longer tenure associated with higher number of bets and betting expenditures? In Fiedler’s (2012) study, the gambling groups (i.e. top 50%, 10%, 5% and 1%) were arbitrarily defined. In addition, the top 50% or 10% online poker playing groups also included the top 1%
playing group. In other words, the defined groups were not independent from each other. This makes it challenging to identify how each group is uniquely different from each other in terms of frequency of betting, tenure and gambling expenditures.

Instead of arbitrarily defining the groups, some researchers have used cluster analysis to identify bettor segments. Park et al. (2002) segmented casino gamblers by their involvement with the gambling games. “Involvement” was considered as a state of motivation, arousal or interests that is related to the consumption of products (such as gambling products). They identified four distinct casino gambler clusters in their research. Respondents from one of these four gambling clusters was reported to have stayed at the gambling venue for longer period of time (i.e. have longer tenure) and to hold significantly more gaming club memberships (suggesting to be high frequency gamblers).

Kim and Petrick (2004) also used reported involvement with gambling activities to group the horse racing bettors and identified four distinct clusters. Respondents in one of the identified clusters reported higher frequency of monthly visits (i.e. high frequency of usage) to the racecourse and greater number of years they visited the racecourse (i.e. implying longer duration). The respondents in this cluster (i.e. long duration) also reported spending higher amounts of money on betting but also experiencing higher amounts of losses when compared to the other identified clusters.
Lee et al. (2006) conducted another study investigating the motivation to engage in casino gambling using the self-report survey format. They reported identifying four distinct casino bettor groups who differed by the number of visits made to the casinos and the betting amounts. For example, one of the bettor groups reported the highest number of visits to the casinos and higher total betting amounts compared to the other bettor groups.

Lee et al. (2006) also found another group of bettors who reported fewer numbers of visits to the casinos and lower total betting amounts. The above findings suggest that the frequent gamblers have longer tenure with the gambling service provider. In addition, bettors with high total betting expenditures are likely to have longer tenure with the gambling service provider (Kim & Petrick, 2004). Accordingly, the following hypotheses are proposed1:

\[ H3: \text{Gamblers' frequency of betting is positively related to their tenure.} \]

\[ H4: \text{Gamblers' total betting expenditures is positively related to their tenure.} \]

\[ \text{It is important to note that causality is neither tested here, nor promised.} \]
Is tenure more important than frequency of buying?

Dagger et al. (2009) have investigated whether frequency of contact or duration of relationship (defined as how long a respondent has been a customer of the firms) affected the customer-reported relationship strength (CRRS) for service firms. The CRRS was defined as the strength of the relationship that was perceived and reported by the customers surveyed. The study found that longer tenure tends to be associated with stronger customer relationship strength (CRRS) than frequency of contact with the service provider. This finding led Dagger et al. (2009) suggest that managers should invest more on developing duration-based strategies than frequency-based strategies in order to achieve high CRRS.

However, their study appears to have some limitations. For example, the study did not report on the financial implications of having long-term high CRRS. It is doubtful that high duration with the service provider can always lead to higher profits if frequency of contact or purchase is decreasing (e.g. decreasing frequency of usage).

For example, Reinartz and Kumar (2000) reported that longer tenure does not necessarily lead to higher life-time revenue. They identified a group of shorter duration customers who generated higher life-time revenues because of high frequency of purchases. Coyles and Gokey (2005) also found that banks suffered higher financial losses when
existing customers reduced their frequency of transactions. The review suggests that frequency of purchase may be more important than long duration in different marketing contexts.

Dagger et al. (2009) also argued that the customer-reported relationship strength (CRRS) is a behaviourally based concept. However, due to its self-report nature, it is probable that the measurement of the construct was subject to a number of biases like socially desirable responses (Steenkamp, de Jong & Baumgartner, 2010) and recall bias problems (Baumeister et al., 2007; Woodside & Wilson, 2002). Moreover, a single-item statement was used to measure the constructs of CRRS, frequency and tenure with the service provider. There are debates on the use of single item self-report survey question which potentially affect the validity and reliability of the underlying measured constructs (Diamantopoulos et al., 2012; Bergkvist & Rossiter, 2007). Given these issues, it might not be correct to suggest that managers should always focus on duration-based strategies than frequency-based strategies.

Another area of contention is the frequency of buying of the industries investigated by Dagger et al. (2009). Their findings were based on industries such as travel agents, hairdressers, family doctors, cinemas, pest control, photo printing, general banking and fast-food outlets. In the Dagger et al.’s research, the self-reported frequency of contacts for most of the services (e.g. travel agencies, hairdresser, cinema, pest
control, family doctor, photo processing and airline) under investigation were under 50 times a year. Only banking and fast-foods had self-reported frequency of contacts of greater than 50 times a year. Gambling, on the other hand, presents a high frequency purchased product (Mizerski et al., 2013). For example, a survey of Florida State Lottery showed that bettors bought various lottery products 7 to 14 times on average over a two-week period (Lam, 2008). Similarly, Australian Productivity Commission (1999) reported that bettors purchased wagering products an average of 17 times over 12-month period. Therefore, for a high frequency purchased product such as gambling, it is possible that frequency of betting will be more important in generating revenue (note that gambling expenditure is equivalent to gambling revenue) for the service provider than tenure. Accordingly, the following hypothesis is proposed:

H5: Gamblers’ frequency of betting will be associated with higher revenues for the gambling service provider than gamblers’ tenure.

\[\text{The aim was to compare the relative importance of tenure vs. frequency of betting. In other words, even if there are different measures of revenue, the relative standing of the two types of responses (frequency vs. tenure) is the relevant issue.}\]
Chapter 3: Methodology

The different research methods used in gambling research have been self-report surveys where the respondents answer a series of questions related to their gambling behaviour (Griffiths & Whitty, 2010; Shaffer et al., 2010; Wood & Griffiths, 2007), conducting experiments where the respondents gamble in simulated gambling scenarios (Cloutier et al., 2006; May et al., 2005), or in actual gambling situations (Jolley et al., 2006; 2013). Among these research methods self-report surveys are the most popular choice among gambling researchers (Shaffer et al., 2010).

However, self-report surveys of behaviours have been criticized for lacking reliability due to problems of recall bias (Baumeister et al., 2007; Nisbett & Wilson, 1977; Woodside & Wilson, 2002) and socially desirable responses (Steenkamp, de Jong & Baumgartner, 2010). Bias occurs when respondents’ responses are affected by an inaccurate recalls of events and decisions.

Socially desirable responses are biased responses in which respondents tend to provide responses that they think will be socially acceptable. The analysis of biased responses can lead researchers to inaccurate findings. The conclusions drawn from self-report surveys of behaviour can be inconsistent with that of actual behaviour (Baumeister et al., 2007). For example, Hu et al. (2000) and Lee et al. (2000) investigated
the frequency and duration of phone service usage using self-report surveys and found that the reported frequency and duration of usage did not match with the frequency and duration of usage recorded on the company database. They report that the high frequency and longer duration users underestimated their usage whereas the low frequency and shorter duration users overestimated their usage figures.

Conducting experiments in simulated gambling scenarios or actual gambling situations also have their drawbacks. In a simulated gambling situation, it is possible to maintain control the possible effects of rival explanations of a causal finding. This ensures that the findings meet internal validity (Bryman & Bell, 2003). Internal validity refers idea that manipulated independent variables caused changes to the dependent variable (Ghauri & Grønhaug, 2010).

However, this does not ensure external validity which refers to the idea that the findings obtained can be generalized, for example, to other situations or other populations (Ghauri & Grønhaug, 2010). Experiments in actual gambling situations are likely to meet the external validity, however, there will be doubts about the internal validity of the findings as there are typically no control groups to minimize or eliminate the effects of rival explanations for any changes in the independent variable.
The current study uses behavioural gambling data obtained from a gambling service provider. The use of behavioural gambling data overcomes the disadvantages of collecting data through self-report surveys that may collect inaccurate data, the high costs of recruiting a representative sample and administering the survey, non-response bias, etc. (Gainsbury, 2011). Some researchers have criticized the use of actual gambling data because of lack of informed consent (Griffiths & Whitty, 2010). Getting respondent consent is an important ethical issue when the risk of identifying the respondents is high. However, in the case of unidentified customer accounts, obtaining consent may not be necessary or possible, especially for a data set that contains records of thousands of individuals.

Behavioural gambling data that is collected over a period of time offers the opportunity to investigate longitudinal betting behaviour unlike the cross-sectional data collected using self-report surveys. It also meets external validity which allows for generalizability but not internal validity. However, the real time, actual data provide better understanding of consumer behaviour over demographic or psychographic measures (O’Malley, 1998; Opperman, 2000). Shaffer et al. (2010) also argued that cross-sectional self-report surveys of behaviour are not adequate for establishing cause and effect relationships. Analysing actual gambling data allows the researchers
with objective, detailed information about betting behaviours which are not affected by recall biases and socially desirable responses.

As discussed previously, the claims by Reichheld and his colleagues (1990) and Dagger et al. (2009), that long-term tenure is desirable, were based on findings for product categories that were not relatively high frequency of purchase. The present study will test these claims for a very high frequency of purchase product of gambling.

**Data**

Data containing gambling behavioural information was obtained from a gambling service provider operating in Australia. This gambling service provider is regarded as one of the largest in Australia offering pari-mutuel and fixed-odds betting opportunities through telephone, Internet, on-course, and retail outlets. The definitions for pari-mutuel and fixed-odds betting are provided in Appendix A.

The company provides wagering opportunities for thoroughbred, harness and greyhounds races, and sporting events (including overseas events). In Australia, the principal forms of wagering are racing (i.e. horses and dogs) and sports betting (i.e. Australian Rules Football, soccer, rugby, etc.). Of these, race wagering accounts for 2.6 billion Australian dollars or 14% of the total wagering expenditure. Thoroughbred horse racing is now the third most attended spectator sport in Australia with about A$14.3 billion wagered in 2009-2010.
This represents 60% increase in money wagered compared to 1999/00 figures (Australian Racing Fact Book, 2011). The data set contained information regarding race wagering. The different types of betting opportunities offered by the company are detailed in Appendix B.

A total of 22,819 individuals registered and set up an account with the company from July 2001 to November 2010. A total of 11,380 individuals (approximately 50%) made wagering bets through the gambling service provider. This proportion of registered clients that did not place a bet has been observed in other studies (see Jolley et al., 2006; personal communications with other gambling service providers by Winthrop Professor Dick Mizerski). However, there were 312 cases with missing data on one of the variables which were removed from subsequent analysis. Therefore, the resultant dataset of 11,068 bettors form the 10-year census used in the study. Each individual bettor is identifiable via a unique account number and no bettors have more than one account. No personal identifiable information was available and ethics approval for the research was obtained from the University ethics committee.

The data set contained information on gender (78% males and 22% females) and age of the bettors. These were self-reported by the bettors. However, these numbers were not used in subsequent analysis because they appeared unreliable. For example, 243 individuals reported their birth date as 01/01/1900 which makes them over 100
years old. There were doubts about the reported gender as well. Past research found that some online video game players change their reported gender during game play (Wood & Griffiths, 2007). It is probable that some bettors in the current data set may not be truthful about their gender and age. This finding supports the potential inaccuracy of using self-survey reports. However, the behavioural gambling data collected by the company is in real time (i.e. not self-reports) and cannot be falsified. These data were collected by the company’s database programmes. The current thesis did not consider using survival analysis (See appendix D-2 for the reasons behind this decision).

Measures

The data had a period that ranged from July 1, 2001 to November 24, 2010. The data included the dates of the first bet and the last bet made by each bettor. Consistent with previous literatures (Dagger et al., 2009; Fiedler, 2012; LaBrie et al., 2007; Palmatier et al., 2006), the difference between the date of first bet and the last bet was used to measure the tenure with the gambling service provider.

The number of betting days reports the number of days that the bettor placed any bets over his tenure with the gambling service provider. For example, a bettor with number of betting days of 5 and tenure of 360 days, mean that this particular bettor placed bets on 5 days out of 360
days. The data set also contained information on the number of bets made over each bettor’s tenure. Gamblers’ frequency of betting was measured by the number of bets placed by the bettors over their tenure with the gambling service provider.

The dataset also contained information on the total dollar bets made by each bettor over their tenure. For the purposes of this research, the total dollar bets made by the bettors are considered equivalent to revenues earned by the gambling service provider. Another piece of information available in the dataset was the total dividend earned by the bettors. Total dividend is the dollar earnings for the bettors from their betting activities over their tenure. It is important to note that high frequency of betting may not necessarily lead to higher total bet amount. An infrequent bettor may bet large amounts when he bets.

For the current study, new measures were created in line with LaBrie et al. (2007). These include average dollar per bet (calculated by dividing total bet in dollar by number of bets), average number of bets per day (calculated by dividing number of bets by number of betting days), net loss in dollar (calculated as the difference between total bets in dollar and total dividend paid) and net loss percentage (calculated as net loss in dollar divided by the total bets in dollar). The calculation of net loss percentage allows for easy comparison between the two frequency groups.
Sample characteristics

As can be seen from Table 2, all the gambling behavioural variables are positively skewed except for the net loss in dollars and tenure (which are negatively skewed). Skewness refers to the tendency of the distribution to deviate from the mean by a larger distance in one direction than in the other (Malhotra, 2010; Tabachnick & Fidell, 2007). This implies that the variables are not normally distributed. Other gambling studies using actual betting behavioural data have also reported the presence of similar skewness in their data set (Fiedler, 2012, Jolley et al., 2006; LaBrie et al., 2007).

As an example from the current study, the distribution for number of bets is shown in Figure 2. The skewness in the behaviour shows that a smaller number of the bettors account for a high proportion of bets, betting days and amount bet. Figure 3 shows the histogram distribution of net loss in dollars. It shows that a large number of bettors made losses from their gambling activities. A small negative skewness (-.45) for tenure suggests the number of bettors who stay for short- and long period of time are very similar. Figure 4 shows tenure has a U-shaped distribution. The histogram charts of other variables are provided in Appendix C.
Table 2: Descriptive statistics of the wagering data (n=11,068)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Median</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bets</td>
<td>717.64</td>
<td>2,206.48</td>
<td>114</td>
<td>9.52</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>45.61</td>
<td>63.36</td>
<td>20</td>
<td>2.34</td>
</tr>
<tr>
<td>Tenure (days)</td>
<td>2,117.66</td>
<td>1,304.38</td>
<td>2,665</td>
<td>-0.45</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>7,970.06</td>
<td>60,444.75</td>
<td>966</td>
<td>32.63</td>
</tr>
<tr>
<td>Total dividend ($)</td>
<td>6,314.66</td>
<td>51,655.51</td>
<td>670.37</td>
<td>33.86</td>
</tr>
<tr>
<td>Net loss ($)</td>
<td>-1,723.41</td>
<td>19,165.84</td>
<td>-248.43</td>
<td>-70.37</td>
</tr>
<tr>
<td>Net loss (%)</td>
<td>-34.07</td>
<td>55.84</td>
<td>-29.94</td>
<td>24.97</td>
</tr>
<tr>
<td>Average bets per day</td>
<td>10.18</td>
<td>13.62</td>
<td>6.33</td>
<td>7.58</td>
</tr>
<tr>
<td>Average bet value ($)</td>
<td>16.44</td>
<td>53.9</td>
<td>7.6</td>
<td>20.60</td>
</tr>
</tbody>
</table>

**Figure 2:** Histogram representation of number of bets (the dashed line represents the median value of number of bets = 114)

**Figure 3:** Histogram representation of net loss in dollars
Figure 4: Histogram representation of tenure and stay duration
Table 2 also shows that the bettors (n=11,068) in the dataset placed 717.64 bets on average (mean) and the average (mean) number of betting day was 45.61 between July 1, 2001 and November 24, 2010. As noted earlier, the distributions of all the variables are skewed. Median values have been suggested as the best measure for reporting highly skewed distributions (Campbell, Machin & Walters, 2007; Tabachnick & Fidell, 2007).

The median value of the number of bets placed shows that half of the betting sample placed more than 114 bets over their tenure with the gambling service provider. The median value of 20 for the number of betting days suggests that at least 50% of the bettors placed bets for less than 20 days. The median value of 2,665 days (7.3 years) suggest that at least 50% of the bettors stay with the gambling service provider for that time period. The bettors generally made net loss from their betting activities over their tenure with an average net loss of $1,723 or an average net loss of 34%. At least 50% of the bettors made a loss to $248.43 and an average loss of about 30%.

**High and low frequency betting groups**

H1 to H4 were examined by comparing the hypothesized relationships between two betting frequency (i.e. high and low) groups. These two bettor groups were created by splitting the number of bets variable at its median value (114) for the 11,068 individuals who made at least
one bet during their tenure with the gambling service provider (see Figure 2).

A median split has been considered appropriate when a continuous variable shows extreme distribution such that a large number of data points gather at one end of the distribution (MacCallum et al., 2002; Streiner, 2002). Median value split has been widely used in academic literature (e.g. Bearden, Rose & Teel, 1994; Gauri, Sudhir & Talukdar, 2008; Greene & Greene 2008; Hong & Sternthal, 2010; Kilbourne & LaForge, 2010; Lam & Mizerski, 2009; Mizerski et al., 2004; Reinartz & Kumar, 2000; Schmittlein, Cooper & Morrison, 1993).

As an additional measure of the appropriateness for splitting the dataset to form the two frequency groups, the ratio of larger group standard deviation to the smaller group standard deviation was calculated (Wright & Royston, 1999). Harris and Boyd (1990) suggest that if this ratio equals or exceeds 1.5, then splitting the data set can be useful. The ratio of standard deviations of number of bets variable for the high frequency bettors and the low frequency bettor exceeded this cut-off point. Therefore, splitting the dataset was deemed appropriate.

Figure 5 provides a population pyramid representation of the distributions of the two bettor groups. Population pyramid is useful in simultaneously comparing the frequency distributions of multiple groups visually (Aldrich & Rodriguez, 2013). The population pyramid
shows that the majority of the low frequency bettors made few bets. On the other hand, some high frequency bettors made large number of bets (evident from greater number of bars being shown on the population pyramid graph in Figure 5).

**Figure 5:** Population pyramid graph for low and high frequency bettors based on median value split of number of bets made (114)

![Population pyramid graph](image)

**Table 3:** Descriptive statistics of low frequency bettors vs. high frequency bettors groups

<table>
<thead>
<tr>
<th></th>
<th>Low frequency bettors</th>
<th>High frequency bettors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median</td>
</tr>
<tr>
<td>Number of bets</td>
<td>36.77 (32.09)</td>
<td>27</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>9.28 (10.17)</td>
<td>6</td>
</tr>
<tr>
<td>Tenure (days)</td>
<td>1,915.74 (1,327.20)</td>
<td>2,235</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>619.85 (2,145.82)</td>
<td>206</td>
</tr>
<tr>
<td>Total dividend ($)</td>
<td>370.46 (1,118.07)</td>
<td>99.51</td>
</tr>
<tr>
<td>Net loss ($)</td>
<td>-190.61 (825.58)</td>
<td>-76.32</td>
</tr>
<tr>
<td>Net loss (%)</td>
<td>-43.75 (75.62)</td>
<td>-48.24</td>
</tr>
<tr>
<td>Average bets per day</td>
<td>5.47 (5.98)</td>
<td>3.64</td>
</tr>
<tr>
<td>Average bet value ($)</td>
<td>18.81 (63.97)</td>
<td>8.46</td>
</tr>
</tbody>
</table>
The descriptive statistics for the two frequency groups (i.e. high and low) are reported in Table 3. Because the distributions of the behavioural gambling variables are not normal, Mann-Whitney U tests were conducted to see whether these two frequency groups came from the same basic population. The Mann-Whitney U test is preferable to the t-test when assumptions of normality cannot be met (Allen & Bennett, 2010). The Mann-Whitney U test compares the median values between the two groups (Pallant, 2005). The results of the Mann-Whitney U test showed that the median values of all the variables were significantly different (p<.001) for the two frequency betting groups. This suggests that the high and low frequency bettor groups can be considered having distinct distributions.

As can be seen from Table 3, the high frequency bettors placed higher number of bets compared to the low frequency bettors. High frequency bettors also appeared to bet more regularly as evidenced by higher number of betting days compared to the low frequency bettors. However, low frequency bettors had higher average dollar per bet and suffered a higher percentage of net loss.

**Relationships among gambling behavioural variables**

In order to test H1 to H4, Kendall’s tau-b (represented as ‘\( \tau_b \)’) was used to examine the interrelationships among the gambling behavioural variables. Kendall’s \( \tau_b \) is a recommended alternative when the
assumptions of normality cannot be met (Allen & Bennett, 2010). It is generally preferred over Spearman’s Rho as it tends to provide a better estimate of the true population correlation and is not artificially inflated by multiple tied ranks (Allen & Bennett, 2010; Malhotra, 2010). Kendall’s $\tau_b$ can also maintain adequate protection against Type I errors even under severe conditions (Arndt et al., 1999). One of the key features of Kendall’s $\tau_b$ is that it can be interpreted in the same way as Spearman’s Rho.

The effect size was calculated as a measure to understand the strength of the obtained relationships. Effect size is a measure of the strength of the relationship between variables (Durlak, 2009). A widely used measure of effect size is Cohen’s $d$ (LeCroy & Krysik, 2007). According to Cohen (1988), $d$ values of 0.20 represents small effect size, whereas values of 0.50 and 0.80 represent medium and large effect sizes respectively. Kendall’s $\tau_b$ values were converted to Cohen’s $d$ by following the procedure recommended by Walker (2003).

As can be seen from Table 4, the frequency of betting is positively related to the number of betting days for both the bettor groups. The relationship is slightly higher for the low frequency bettor group ($\tau_b = .60$, Cohen’s $d = 2.75$) than for the high frequency bettor group ($\tau_b = .56$, Cohen’s $d = 2.42$). Both the relationships show large effect sizes.
Overall, the findings support H1, that gamblers’ frequency of betting is positively related to their number of days that they bet. The results suggest that the more frequently gamblers bet, the more days they bet.

Table 4 also shows that the frequency of betting is positively related to total amount bet (expenditures made by the bettors). The amount bet by the bettors is the betting revenue for the gambling service provider. This effect was found for both the high frequency bettors ($\tau_b = .52$, Cohen’s $d = 2.13$) and low frequency bettors ($\tau_b = .59$, Cohen’s $d = 2.66$). The effect sizes for these relationships suggest that the relationships are strong. These findings support H2, that frequency of betting will be positively related to total betting expenditures.

The frequency of betting is positively related to tenure for both the high frequency ($\tau_b = .15$) and low frequency bettor ($\tau_b = .09$) groups. These findings provide support for H3, that frequency of betting is positively related to duration. However, the relationship is weak. The Cohen’s $d$ suggest small to medium effect sizes (Cohen’s $d = .48$ for high frequency bettors and Cohen’s $d = .28$ for low frequency bettors).

From Table 4, it is also evident that the tenure is related to the number of days that bettors bet for both the high ($\tau_b = .24$) and low ($\tau_b = .15$) frequency bettors. However, this relationship is also weak (Cohen’s $d = .79$ for high frequency bettors and Cohen’s $d = .48$ for low frequency bettors).
The total bet value (expenditure) is positively related to tenure for high frequency ($\tau_b = .13$) and for low frequency bettors ($\tau_b = .08$). These findings provide support for H4, that expenditures made by the bettors (revenue for the gambling service provider) is positively related to the bettors’ tenure with the gambling service provider. However, both the relationships show small to medium effect size (Cohen’s $d$ is 0.38 for high frequency bettor group and 0.25 for low frequency bettor group) suggesting that the relationship is weak.

Overall, the Kendall’s $\tau_b$ results suggest that frequency of betting is positively related to the number of days bettors bet and the amount they bet. These relationships are also very strong. Tenure with the gambling service provider is also positively related to both frequency of betting and the amount they bet. However, these relationships are weak.
**Table 4**: Correlations among gambling behaviour variables for low frequency bettors (presented above the diagonal) and high frequency bettors (presented below the diagonal)

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. of betting days</th>
<th>No. of bets</th>
<th>Total bet value ($)</th>
<th>Total Dividend ($)</th>
<th>Duration (days)</th>
<th>Average number of bets</th>
<th>Average dollar per bet ($)</th>
<th>Net Loss ($)</th>
<th>Net Loss (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of betting days</td>
<td>-</td>
<td>.60</td>
<td>.54</td>
<td>.48</td>
<td>.15</td>
<td>-.09</td>
<td>.08</td>
<td>-.32</td>
<td>.25</td>
</tr>
<tr>
<td>No. of bets</td>
<td>.56</td>
<td>-</td>
<td>.59</td>
<td>.55</td>
<td>.09</td>
<td>.34</td>
<td>-.03</td>
<td>-.36</td>
<td>.28</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>.43</td>
<td>.52</td>
<td>-</td>
<td>.71</td>
<td>.08</td>
<td>.17</td>
<td>.38</td>
<td>-.54</td>
<td>.27</td>
</tr>
<tr>
<td>Total Dividend ($)</td>
<td>.43</td>
<td>.52</td>
<td>.88</td>
<td>-</td>
<td>.08</td>
<td>.17</td>
<td>.29</td>
<td>-.23</td>
<td>.56</td>
</tr>
<tr>
<td>Duration (days)</td>
<td>.24</td>
<td>.15</td>
<td>.12</td>
<td>.13</td>
<td>-</td>
<td>-.06</td>
<td>.02*</td>
<td>-.03</td>
<td>.06</td>
</tr>
<tr>
<td>Average number of bets</td>
<td>-.01*</td>
<td>.43</td>
<td>.25</td>
<td>.26</td>
<td>-.05</td>
<td>-</td>
<td>-.17</td>
<td>-.12</td>
<td>.10</td>
</tr>
<tr>
<td>Average dollar per bet ($)</td>
<td>-.04</td>
<td>-.12</td>
<td>.36</td>
<td>.33</td>
<td>-.01*</td>
<td>-</td>
<td>-.28</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Net Loss ($)</td>
<td>-.35</td>
<td>-.43</td>
<td>-.65</td>
<td>-.53</td>
<td>-.09</td>
<td>-.23</td>
<td>-.26</td>
<td>-</td>
<td>.19</td>
</tr>
<tr>
<td>Net Loss (%)</td>
<td>.09</td>
<td>.09</td>
<td>.12</td>
<td>.24</td>
<td>.04</td>
<td>.06</td>
<td>.04</td>
<td>.23</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: Non-parametric Kendall’s tau-b.
- all correlations are significant at 0.01 level unless marked with * and ns.
- *Correlation is significant at 0.05 level
- ns Correlation is non-significant
Identifying different bettor groups

The previous section reported the relationships among gambling behavioural variables for the high and low frequency groups. However, it is possible that there are naturally occurring bettor groups who differ in terms of their frequency of betting, tenure and gambling expenditures. These naturally occurring bettor groups may be different from the bettor groups arbitrarily created based on the median split of number of bets made. Prior studies have used arbitrary cut-off points to create the top 1%, 5% or 10% of the gambling sample (Fiedler, 2012; LaBrie et al., 2007).

A major problem with this approach is that it ignores the naturally occurring independent groups that may exist within the dataset. Reporting the results for these top 1%, 5% or 10% gambling groups muddles the understanding of the gambling behaviour because the top 10% includes the top 5% or 1% group. In addition, focusing on the top 1% clearly ignores the understanding of the behaviour of the other 99% of the gambling sample. Thus, in the current study a two-step cluster analysis was performed on the dataset of bettors (n=11,068) to identify if there are bettor groups that differ by the gambling behaviours, particularly in terms of frequency of betting, tenure and gambling expenditures.
Two step clustering procedure

Cluster analysis has been a widely used method for empirical classification of objects (e.g. respondents, products, or other entities) based on a set of data (e.g. user characteristics) (Hair et al., 2010). Hair et al. (2010) is considered as authoritative reference for cluster analysis in marketing research. Cluster analysis is an interdependence technique which defines groups in a data set without identifying dependent or independent variables (Jurowski & Reich, 2000). Even though it is generally viewed as an exploratory technique, it can be used in a confirmatory role to examine prior assumptions or stated hypothesis (Hair et al., 2010). This technique was deemed appropriate to identify if there are bettor groups who differ in terms of tenure, frequency of betting and gambling revenues.

A two-step clustering of the cases (n=11,068 bettors) with log-likelihood criterion was used to identify the bettor groups within the data set. This method of clustering has been suggested as appropriate for handling large datasets and is thought to behave reasonably well when assumptions of variable independence and normal distribution are not met (Norusis, 2011).

This was an important consideration because the value ranges of the gambling behavioural variables are skewed (see Table 1). Note that the net loss in dollar, net loss in percentage, average bets per day and
average bet value in dollars variables were not used in cluster analysis due to potential collinearity problem (Sambandam, 2003). The decision was taken because the new variables are a combination of the original variables and thus omitting them avoids the problem of singularity (Tabachnick & Fidell, 2007).

The two-step cluster procedure starts with some of the data being used to create a series of small clusters known as the “preclusters”. The remaining data are then included into one of these preclusters, or used to form a new precluster based on a specified distance measure such as the log-likelihood ratio. When the log-likelihood function between two clusters decreases, they are combined to form a single cluster (Zheng, Wang & Hu, 2007). Next, all the preclusters are grouped together using a hierarchical clustering algorithm to determine the final number of clusters.

The optimal level of clusters can be formed based on either the Schwarz Bayesian inference criterion (BIC) or the Akaike information criterion (AIC). The current study uses BIC because it is considered better because it avoids some arbitrariness in traditional clustering techniques (Okazaki, 2006).

**Cluster analysis results.** The two-step clustering of the original wagering dataset (n=11,068) generated a four-cluster solution. Table 5 shows the composition of these clusters with the average values for the variables.
under consideration. Cluster 1 (labelled “long tenure”) has the longest average tenure (3,126.34 days) and cluster 2 (labelled “short tenure”) has the shortest average tenure (688.27 days) among all the clusters. Cluster 3 (labelled “high frequency”) represents a group of bettors who are very active as shown by the high values on all the behavioural variables. Finally, cluster 4 (labelled “very high frequency”) represents a small group of bettors who exhibit extreme frequent betting behaviour. The bettors in this group account for the highest number of bets (14,901.26) and the highest total betting expenditure ($331,257) (i.e. revenue for the gambling service provider). They also earned the highest dividend from their betting activities ($272,691.37).

**Table 5: Composition of cluster solution**

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of the total sample</td>
<td>48.4%</td>
<td>37.6%</td>
<td>13.1%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Tenure (days)</td>
<td>3,126.34</td>
<td>688.27</td>
<td>2,735.45</td>
<td>2,232.03</td>
</tr>
<tr>
<td>Number of bets made</td>
<td>240.92</td>
<td>219.76</td>
<td>3,097.54</td>
<td>14,901.26</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>28.26</td>
<td>21.64</td>
<td>174.04</td>
<td>231.96</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>2,719.39</td>
<td>2,351.15</td>
<td>22,378.5</td>
<td>331,257</td>
</tr>
<tr>
<td>Total dividend ($)</td>
<td>1,996.97</td>
<td>1,732.14</td>
<td>17,298.6</td>
<td>272,691.37</td>
</tr>
<tr>
<td>Cluster label</td>
<td>Long duration</td>
<td>Short duration</td>
<td>High frequency</td>
<td>Very high frequency</td>
</tr>
</tbody>
</table>
One-way analysis of variance (ANOVA) tests showed that all the gambling behavioural variables were significantly different from each other for all the four groups (p<0.001). Further analysis using post-hoc (Scheffe test) examining the mean differences for the variables found that the number of betting days and the tenure are significantly different for all the four groups (p<0.001). In other words, the four clusters represent distinct bettor groups who have different tenure with the gambling service provider. Prior research has not previously identified these differences within gamblers.

Post-hoc analysis also found that there are no significant differences between clusters 1 and 2 in terms of number of bets made, total dollar amount bet, and total dividend earned. This suggests that the clusters with the longest and shortest tenure do not differ from each other in terms of these gambling behavioural variables. It implies that regardless of how long these two groups of bettors stay with the gambling service provider, they appear to bet similar dollar amounts and earn similar returns from their betting activities.

Even though the clusters 1 and 2 did not significantly differ from each other except for their tenures, these two clusters differed significantly from clusters 3 and 4 on all the other gambling behavioural variables (p<0.001). Clusters 3 and 4 significantly differed from each other for all the five gambling behavioural variables (see Table 5).
Cluster analysis also revealed that there is a small group of gamblers (cluster 4) who show extreme gambling behaviour. For example, the total bet value for this group is $331,257 which is 14.8 times greater than the bettors in cluster 3. This finding is similar to reported findings of Perfetto & Woodside (2009). It is interesting to note that cluster 4 has a shorter tenure (2,232.03 days) compared to cluster 3 (2,735.45 days) even though cluster 4 has a higher frequency of betting. This suggests that longer tenure is not necessarily associated with higher frequency of betting.

**Reliability of the cluster solution.** The reliability of the resulting cluster solution was checked using two methods. First, it has been suggested that cluster analysis solutions can be affected by the order of the cases (Norusis, 2011; Hair et al., 2010). Therefore, the dataset was randomly sorted to generate a different order and a two-step cluster analysis was re-performed on this re-sorted dataset. The new cluster solution was compared to the original solution to see if similar results were obtained. A matched cross-classification of cluster membership suggests high stability of the cluster solutions obtained (Hair et al., 2010). Table 6 reports the results of this second cluster solution.
**Table 6:** Composition of cluster solutions on the resorted data

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of the total sample</td>
<td>53.3%</td>
<td>32.6%</td>
<td>12.7%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Tenure (days)</td>
<td>2,978.28</td>
<td>482.44</td>
<td>2,933.65</td>
<td>2,420.89</td>
</tr>
<tr>
<td>Number of bets made</td>
<td>218</td>
<td>269.66</td>
<td>2,712.04</td>
<td>13,291.4</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>26.13</td>
<td>24.16</td>
<td>169.52</td>
<td>231.77</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>2,529.54</td>
<td>2,804.87</td>
<td>19,829.15</td>
<td>233,037.86</td>
</tr>
<tr>
<td>Total dividend ($)</td>
<td>1,842.32</td>
<td>2,076.62</td>
<td>15,409.08</td>
<td>190,393.6</td>
</tr>
</tbody>
</table>

A comparison of Tables 7 and 8 suggest that less than 6% of the cases were placed in different clusters. These suggest a stable cluster solution because less than 10% of the cases were assigned to a different cluster (Hair et al., 2010).

A second way to examine the reliability of the solution is to split the entire dataset into two halves and perform a two-step cluster analysis on each half. If the results of these two cluster analysis are similar, it would again suggest a stable cluster solution. Thus, the dataset was randomly split into two equal parts (labelled Set A and Set B). The results of the two-step cluster analysis on these two datasets are provided in Tables 7 and 8. A comparison of Tables shows that less than 10% of the cases were reassigned to different clusters. Therefore, the cluster solutions can be considered stable and reliable (Hair et al., 2010).
Table 7: Composition of cluster solutions for Set A (50% of the sample)

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of the total sample</td>
<td>50.7%</td>
<td>33.6%</td>
<td>14.4%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Tenure (days)</td>
<td>3,021.41</td>
<td>524.39</td>
<td>2,722.93</td>
<td>2,371.41</td>
</tr>
<tr>
<td>Number of bets made</td>
<td>212.47</td>
<td>205.25</td>
<td>2,660.09</td>
<td>12,743.11</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>25.31</td>
<td>20.53</td>
<td>161.73</td>
<td>217.20</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>2,512.57</td>
<td>2,413.53</td>
<td>19,444.38</td>
<td>259,867.12</td>
</tr>
<tr>
<td>Total dividend ($)</td>
<td>1,868.68</td>
<td>1,678.54</td>
<td>15,156.09</td>
<td>208,117.15</td>
</tr>
</tbody>
</table>

Table 8: Composition of cluster solutions for Set B (50% of the sample)

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of the total sample</td>
<td>48.3%</td>
<td>39%</td>
<td>11.4%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Tenure (days)</td>
<td>3,148.75</td>
<td>738.49</td>
<td>2,891.12</td>
<td>2,478.36</td>
</tr>
<tr>
<td>Number of bets made</td>
<td>244.06</td>
<td>267.18</td>
<td>2,908.72</td>
<td>15,320.19</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>28.85</td>
<td>25.06</td>
<td>178.48</td>
<td>261.82</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>2,619.21</td>
<td>2,671.24</td>
<td>23,591.22</td>
<td>230,696.58</td>
</tr>
<tr>
<td>Total dividend ($)</td>
<td>1,911.49</td>
<td>1,994.64</td>
<td>18,200.36</td>
<td>194,328.01</td>
</tr>
</tbody>
</table>

A summary of the cluster analysis. Cluster 1, which has the longest average tenure with the gambling service provider, is not associated with higher gambling revenues. Cluster 4, which accounts for the highest average frequency of betting (14,901 number of bets) among all the clusters, accounted for the highest gambling revenues.
($331,257) for the gambling service provider. Cluster 3, which accounts for the second highest average frequency of betting (3,097 number of bets) also accounts for second highest average gambling revenues ($22,378.5).

An examination of the four clusters suggests gambling revenues follow the frequency of betting. Figure 7 shows a greater value of frequency of betting is associated with a greater value of gambling expenditures. However, similar association cannot be observed between tenure and gambling expenditures (Figure 8). The cluster analysis results suggest that gamblers’ frequency of betting is associated with greater revenue for the gambling service provider (i.e. greater betting expenditures) compared to tenure with the gambling service provider. The results provide support for H5, that gamblers’ frequencies of bets are associated with gambling revenues compared to their tenure.

**Figure 6:** Bar chart showing number of bets and total bet value (i.e. gambling revenue)
In order to get an understanding of the variables that discriminate among the four clusters obtained from previous analysis, a multiple discriminant analysis (MDA) was performed. MDA was used because the results can provide information on the strength of the relationship between each group, and the variables under consideration (Chen, 2003; Tsiotsou, 2005). MDA can also be used to determine the reliability of the discriminant solutions by correctly classifying or grouping new observations (Malhotra, 2010).
Multiple discriminant analysis (MDA) procedure

The MDA was conducted by considering the cluster memberships as the dependent variable, and the behavioural gambling variables as the independent variables. Because the distributions of the gambling behavioural variables for the two bettor groups are not normal, they were normalized using natural log transformation. The variables used in the MDA were number of betting days, number of bets, tenure and total bet value. Total dividend earned and net loss variables were not included in the MDA because these represent the returns from engaging in gambling activities and not controllable by the bettors. Step-wise MDA was performed because there is no prior theory to determine which of the four gambling behavioural variables is useful in discriminating the clusters.

In a step-wise MDA estimation method, independent variables are entered into the discriminant function each at a time based on their discriminating power (Hair et al., 2010). Mahalanobis $D^2$ was employed to determine overall significance because this has been advocated as the most appropriate for stepwise MDA (Hair et al., 2010; Tabachnick and Fidell, 2007). Mahalanobis $D^2$ is a procedure that is used to maximize a generalized measure of the distance between the two closest groups (Malhotra, 2010). This procedure is particularly preferable over the Euclidean distance measure when the predictor variables are correlated (Feinberg, 2010).
To test the reliability of the MDA solutions, the dataset (11,068 cases) was randomly split into two equal halves (5,534 cases in each half), with one half labelled as the ‘analysis sample’ and other half labelled as the ‘holdout sample’. The analysis sample was used to conduct the MDA, whereas the holdout sample was used to check the reliability of the solution obtained. This procedure is known as the split-sample validation or cross-validation (Hair et al., 2010).

**The MDA results.** The MDA produced three significant discriminant functions from the analysis sample (n=5,534). However, only two functions explained 98.3% of the total variance in the data. Function 1 explained 60.9% of the total variance (Wilks’ $\lambda = 0.32; \chi^2 = 6273.13; df = 12, p<0.001$) and function 2 explained 37.4% of the total variance (Wilks’ $\lambda = 0.62; \chi^2 = 2636.43; df = 6, p<0.001$). The F-test of the equality of group means supports the results obtained (p<0.001).

The canonical correlation for the first function was 0.69 and for the second function was 0.60. The functions could correctly classify 85% and 85% of the cases in both the analysis and the holdout samples respectively. Recall that the holdout sample was created to assess the reliability of the resultant discriminant function. Table 9 shows the hit ratios for the four cluster groups for both the samples.
Table 9: Classification results for both analysis and holdout samples

<table>
<thead>
<tr>
<th>Actual group (%)</th>
<th>Predicted group membership (%)</th>
<th>Size of clusters (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1 (long tenure)</td>
<td>Cluster 2 (short tenure)</td>
</tr>
<tr>
<td>Analysis sample</td>
<td>Cluster 1 (long tenure)</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Cluster 2 (short tenure)</td>
<td>29.8</td>
</tr>
<tr>
<td></td>
<td>Cluster 3 (high frequency)</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>Cluster 4 (very high frequency)</td>
<td>4.3</td>
</tr>
<tr>
<td>Holdout sample</td>
<td>Cluster 1 (long tenure)</td>
<td>98.5</td>
</tr>
<tr>
<td></td>
<td>Cluster 2 (short tenure)</td>
<td>28.2</td>
</tr>
<tr>
<td></td>
<td>Cluster 3 (high frequency)</td>
<td>13.2</td>
</tr>
<tr>
<td></td>
<td>Cluster 4 (very high frequency)</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Note: 84.7% and 85.3% of analysis and holdout sample correctly classified respectively
Hit ratios are the percentage of correctly classified cases. The hit ratios were compared against the proportional chance criterion (50%) and the maximum chance criterion (50%). Hair et al. (2010) recommends that for practical significance, the classification accuracy must exceed these selected comparison standard by a margin of at least 25 percent (i.e. the calculated hit ratios should be greater than 62.5%). The discriminant analysis results from the current study show that the hit ratios for both the analysis (84.7%) and the holdout samples (85.3%) exceeded this limit (62.5%).

Tables 10 and 11 report the mean values of log transformed variables along with the discriminant loadings. Discriminant loadings are considered more useful for interpretive purposes compared to discriminant coefficients (Hair et al., 2010; Tabachnick & Fidell, 2007). Discriminant loadings measure the relative importance of the predictor variables, with greater value suggesting more importance (Malhotra, 2010). The signs (i.e. positive/negative) associated with the discriminant loadings only indicate the direction of the effects (increase/decrease). Tabachnick and Fidell (2007) suggest that loadings over 0.33 can be considered eligible for inclusion in the model.
**Table 10:** Summary of discriminant analysis (analysis sample; n=5,534)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cluster 1 (long tenure)</th>
<th>Cluster 2 (short tenure)</th>
<th>Cluster 3 (high frequency)</th>
<th>Cluster 4 (very high frequency)</th>
<th>Function 1</th>
<th>Function 2</th>
<th>Function 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenure (days)</td>
<td>8.04</td>
<td>5.73</td>
<td>7.79</td>
<td>7.25</td>
<td>.99</td>
<td>-.12</td>
<td>.05</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>2.68</td>
<td>2.35</td>
<td>5.09</td>
<td>5.32</td>
<td>.36</td>
<td>.87</td>
<td>-.25</td>
</tr>
<tr>
<td>Number of bets</td>
<td>4.37</td>
<td>4.17</td>
<td>7.72</td>
<td>9.31</td>
<td>.30</td>
<td>.94</td>
<td>.15</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>11.12</td>
<td>10.85</td>
<td>14.13</td>
<td>16.45</td>
<td>.27</td>
<td>.79</td>
<td>.40</td>
</tr>
</tbody>
</table>

**Table 11:** Summary of discriminant analysis (holdout sample; n=5,534)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cluster 1 (long tenure)</th>
<th>Cluster 2 (short tenure)</th>
<th>Cluster 3 (high frequency)</th>
<th>Cluster 4 (very high frequency)</th>
<th>Function 1</th>
<th>Function 2</th>
<th>Function 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenure (days)</td>
<td>8.04</td>
<td>5.75</td>
<td>7.76</td>
<td>7.4</td>
<td>.99</td>
<td>-.05</td>
<td>-.06</td>
</tr>
<tr>
<td>Number of betting days</td>
<td>2.69</td>
<td>2.39</td>
<td>5.08</td>
<td>5.35</td>
<td>.30</td>
<td>.90</td>
<td>-.16</td>
</tr>
<tr>
<td>Number of bets</td>
<td>4.38</td>
<td>4.22</td>
<td>7.71</td>
<td>9.04</td>
<td>.24</td>
<td>.95</td>
<td>.13</td>
</tr>
<tr>
<td>Total bet value ($)</td>
<td>11.11</td>
<td>10.86</td>
<td>14.05</td>
<td>16.60</td>
<td>.23</td>
<td>.80</td>
<td>.49</td>
</tr>
</tbody>
</table>
The first discriminant function, representing tenure, is the primary source of difference between the short tenure group and the other groups. The second function; representing primarily gambling activities, such as the number of bets made, the number of days the bettors bet and the total amount that they bet differentiates all four groups. The third function, even though it loads high on total bet value, is not useful because of its low explanation of variance. Figures 8 and 9 provide scatter-plot of functions 1 and 2.

**Figure 8:** Scatterplot for analysis sample (n=5,534)
A Summary of the MDA results. The results of MDA show that there are distinct bettor groups who are different in terms of their tenure and frequency of bets (i.e. number of bets made). The long duration bettor (cluster 1) group does not have a higher number of bets or bet on more betting days compared to the high (cluster 3) and the very high frequency bettor groups (cluster 4). The short duration bettor group (cluster 2) have very similar betting activities as that of the long duration group (cluster 1). It is likely that these groups of bettors are looking for quick returns from their activities and may represent a non-
engaged group who do not wish to stay with the gambling service provider for any extended period of time. Both the high (cluster 3) and very high (cluster 4) frequency group show similar trends in gambling activity where these groups not only have greater number of bets but also account for higher total dollar bet value (see Tables 10 & 11).

Is tenure more important than frequency for generating revenue?

Recall that H5 suggested that for a high frequency purchased product such as gambling, frequency of betting will be associated with higher revenue than with tenure. Cluster analysis results previously provided support for H5 by showing that high frequency of betting was associated with high gambling revenue (expenditures for gamblers). In order to test more direct relationships among tenure, number of bets and total bet value, a linear regression analysis was conducted on the original dataset of 11,068 bettors. The aim of this analysis was to investigate how the two variables – tenure or frequency of bets, are related to total bet value. The total bet value represents the gambling revenue for the gambling service provider. In the regression analysis, total bet value was entered as the dependent variable. Tenure and frequency of bets are entered as the independent variables. These variables were normalized using log transformation because of their skewness (see Table 1). An examination of the collinearity statistics showed that the tolerance was greater than the cut-off point of 0.10.
This indicates that the multicollinearity was not present among the variables.

**Multiple regression analysis results.** The regression analysis output is summarized in Table 12. The regression analysis shows that average number of bets made is positively related and statistically significant to the average total dollar bet. However, the relationship between average tenure and the average total dollar bet is not statistically significant.

**Table 12:** Results of multiple regression analysis

```
<table>
<thead>
<tr>
<th>Model summary</th>
<th>Analysis of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.87</td>
</tr>
<tr>
<td>R²</td>
<td>0.79</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.78</td>
</tr>
<tr>
<td>Std. error of estimate</td>
<td>0.998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression coefficients B</th>
<th>Standardized coefficients β</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>7.06</td>
<td>163.51</td>
<td>.000*</td>
</tr>
<tr>
<td>Log duration</td>
<td>.004</td>
<td>.003</td>
<td>.668</td>
</tr>
<tr>
<td>Log number of bets</td>
<td>.915</td>
<td>.885</td>
<td>196.88</td>
</tr>
</tbody>
</table>
```

Note: *p<0.01
Dependent variable: Log total bet value.
n = 11,068
Given the regression results, the following equation can be derived. The equation shows that the total bet value is associated only with number of bets made by bettors. The tenure has no association with total bet value.

\[ \ln(\text{total bet value}) = 7.06 + 0.915 \times \ln(\text{number of bets}) \]

where, \( \ln \) represents the natural logarithm.

**A summary of the multiple regression analysis.** Recall that the total bet value represents revenue for the gambling service provider. The obtained results are contrary to the idea that higher revenues can be generated from longer tenure customers (Ahmad & Buttle, 2001; Buchanan & Gillies, 1990; Dagger et al., 2009; Evans, 2002; Morgan & Hunt, 1994; Reichheld & Kenny, 1990; Reichheld & Sasser, 1990; Sheth & Parvatiyar, 1995). The obtained results appear to support the findings of East, Hammond & Gendall (2006) and Reinartz & Kumar (2000) that tenure does not necessarily lead to higher revenues. The regression result provides a strong support for H5, that gamblers’ frequency of betting will be associated with higher revenues for the gambling service provider than gamblers’ tenure. See Appendix D-3 for a discussion on light customers.

One argument can be whether the bettors that existed in the dataset from the beginning would lead to a different conclusion given that these bettors would have stayed the longest compared to bettors who
enter the dataset in recent period. A regression analysis of bettors existing in the dataset from 2001 to 2003 was conducted to see if the conclusion could have been different. However, the regression results for the bettors from 2001 to 2003 provided the same conclusion as that conducted for the bettors from 2001 to 2010. In other words, frequency of betting was found to be associated with higher revenues compared to tenure of these bettors. See the Appendix D-4 for the table with regression results for bettors from 2001-2003. The implications of these results are further discussed in the next chapter.
Chapter 5: Summary and Conclusion

Summary

A major finding of this research is that, for this gambling service provider, tenure is not significantly related to gambling revenues but frequency of betting is positively and significantly related to gambling revenues. Put in another way, higher frequency of betting is more important than longer tenure for the gambling service provider.

In order to understand the effects of the frequency of betting, a median split of the number of bets variable (i.e. frequency of betting) was used because of the skewed distributions. This split created two bettor groups – the high frequency (top 50%) and the low frequency (bottom 50%) bettor groups. For both bettor groups, frequency of betting was found to be positively related to the number of days that bettors bet and gambling revenues for the gambling service provider. The tenure with the gambling service provider was also found to be positively related to the number of days that bettors bet and gambling revenues for the gambling service provider. However, this latter relationship was much weaker compared to the relationship between frequency of betting and gambling revenues.

Another interesting finding from the study was that the longest and the shortest tenure groups did not differ in terms of their gambling behaviours. These two groups accounted for similar number of bets,
betting days and gambling revenues for the gambling service provider. Moreover, not only did these two tenure groups accounted for similar gambling revenues, but they also accounted for lower amount of gambling revenues when compared to the other two groups with high frequency of betting.

The study also found the presence of an extreme frequency of betting group (representing 1% of the total betting population of this provider) that accounted for disproportionately high gambling revenue. This group made 12% of all the bets placed by the entire wagering population but accounted for 31% of the entire revenue for the gambling service provider. In contrast, the longest tenure group accounted for 13% of all the bets placed but accounted for only 14% of the entire revenue for the gambling service provider.

**Limitations**

There are several limitations in this research. The study was conducted with a single gambling service provider offering race wagering, and may not be generalizable across other gambling products or other product categories. However, it should be noted that other researchers have reported the importance of frequency of betting for other gambling products such as lottery (Mizerski, et al., 2004) and online casinos (Fielder, 2012a, b; Jolley et al., 2006). These studies, however, did not investigate the tenure with the gambling service providers.
Future research should investigate the tenure variable for other gambling products to see if the findings in the current study hold for other gambling contexts.

The current study investigated gambling data obtained over nine year period. It is possible that various economic factors, that have not been measured, may have affected the gambling behaviours of the wagerers. Previous studies have reported that household income can affect the gambling expenditures (Layton & Worthington, 1999; Worthington et al., 2007). Recently, Lee (2009) found that gross domestic product (GDP) was positively related to gambling industry revenue when investigated across 42 different countries. However, these issues could not be explored in the current research because of lack of the firm’s information on the backgrounds of the race wagerers.

Lastly, this research specifically focuses on the gambling behaviours of regular gamblers for a wagering firm. Thus, the findings from this research may not be generalized or applied to problem or pathological gambling behaviours. It is possible that high frequency of betting is associated with high incidences of gambling addictions (Narayanan & Manchanda, 2012) but this could not be concluded in the current research due to lack of data on problem or pathological gambling behaviours.
Conclusions

Longer tenure of customers is argued to be beneficial for companies (Ahmad & Buttle, 2001; Reichheld & Kenny, 1990; Reichheld & Sasser, 1990). These studies did not explicitly measure the effects of frequency of purchase. For example, Dagger et al. (2009) proposed that managers should invest more on developing duration-based strategies than frequency-based strategies in order to achieve stronger customer relationships on a long term basis. In their study, Dagger et al. measured the customer-based relationship strength by measuring how often the customers come in contact with the company. In that study, the self-reported frequency of contacts for most services (e.g. travel agencies, hairdresser, cinema, pest control, family doctor, photo processing and airline) under investigation were under 50 times a year.

Only banking and fast-foods had self-reported frequency of contacts of greater than 50 times a year. In comparison, frequency of gambling product purchases is much higher than these product categories. For example, lottery purchases in Florida averaged 364 times a year (Mizerski et al., 2004). The Productivity Commission (1999) reported that the average purchase frequency of gambling in Australia was 55.9 times per year. In the current study, the number of bets placed by the high frequency and low frequency bettor groups averaged 1,398 times and 36 per year respectively. In addition, the Dagger et al. (2009) study was based on self-reports which have been criticised for
suffering from recall bias. Therefore, the proposed recommendation of Dagger et al. (2009) of investing in duration-based strategies may not be appropriate for a very high frequency purchased product.

In addition, analysis by other researchers on the same industries investigated by Dagger et al. (2009) such as supermarkets, catalogue marketing, car servicing, hairdressers, fashion stores, retail banking, etc. have reported that the relationship between tenure and revenue to be weak (East et al., 2006; Reinartz & Kumar, 2000; Smith & Lakhani, 2008).

The current study finds for high frequency purchase of gambling, the frequency of betting and not how long a customer has been with the company, is associated with generating higher revenue for the gambling service provider. Due to the high frequency of purchase of gambling compared to other product categories investigated in prior studies, the current research findings may represent a boundary condition for the benefits of having longer duration buyers. In other words, the benefits espoused by scholars supporting of longer tenure may not be obtainable for products categories that are bought with high frequency.

**Future research direction**

In the past decade there has been growing interest in the area of customer life-time value (CLV). The aim of CLV is to identify
individual customer who are going to be profitable for the company to serve (Jain & Singh, 2002; Keiningham, Aksoy & Bejou, 2006). CLV is defined as the present value of all cash inflows less expenditures over the lifetime of stay with the company for a customer (Bhatnagar et al., 2008). The current research findings show that customer groups with high frequency of purchase is the most desirable group to target. Adopting the CLV approach can help differentiate the profitable customers from the unprofitable ones within these high frequency buyer groups. Future research can investigate this line of inquiry.

Implications for loyalty programmes

The finding that long tenure may not be important for highly frequent purchased products holds important implications for loyalty programs. Meyer-Waarden and Christophe (2006) reported that loyalty programmes in retail stores did not necessarily lead to high frequency of purchases from these stores. They also report that loyalty card holders of the retail stores were already buying heavily from these stores for some time before subscribing to the loyalty cards, indicating that these customers already had long tenure with these retail stores. These suggest that offering loyalty programmes to high frequency buyers may not be necessary. However, this may not necessarily imply that loyalty programmes have no usefulness. A well-designed loyalty programme may encourage the low frequency buyers to purchase more from the company, thereby increasing the chance that they become
high frequency purchasers in the long-run. For example, if the loyalty programme is accepted by more than one retail store, then it will become easier for low frequency buyers to accumulate points and lead to fewer switching (Meyer-Waarden & Christophe, 2006).

It has also been suggested that high frequency of buying in a mature industry that exhibit Negative Binomial Distribution (NBD) in purchases generally indicate some level of habitual behaviour (East et al., 2013; Mizerski et al., 2013; Rundle-Thiele & Bennett, 2001). Habitual behaviour does not require detailed cognitive processes such as commitment for decision making (East et al., 2013). Loyalty programmes are designed to affect cognitive decision making of the buying process (Dowling & Uncles, 1997). For example, some loyalty programmes may reward the customers for staying longer or spending more with the company. In reality, habit could explain continual purchase of a company’s products which may be mistaken for loyalty (Uncles et al., 2003). For a high frequency of purchase product category investigated in the current research, loyalty programmes may be of limited use due to habitual buying.

The findings also provide some support for the propositions made by Rundle-Thiele and Bennett (2001) that there is need to develop different loyalty measures for consumable, durable and service marketing industries. Specifically, they suggest that different loyalty measures should be developed depending on market characteristics
(e.g. brand switching, purchase frequency, habit, share of category, etc.). The findings from this study show that differences in purchase frequencies can have different implications for consumer loyalty.

**Implications for gambling service providers**

Previous research in gambling has not investigated frequency of betting and tenure with gambling service providers. The finding that high frequency bettors account for greater gambling expenditure supports Hansen and Rossow (2008) who reported similar findings for slot machine gamblers. The current findings challenge the reported findings of Fiedler (2012) by showing that high frequency gamblers do not necessarily have shorter tenure. Specifically, the lowest frequency of betting group accounted for the shortest tenure and that the highest frequency of betting group did not account for the longest tenure. This could be because of different gambling product and medium (i.e. online poker) investigated by Fiedler compared to wagering investigated in the current study.

Another finding that contributes to the gambling literature is the empirical support for the Extremely Frequent Behaviour (EFB) theory (Perfetto and Woodside, 2009). EFB theory proposes that in some industries a very small group of the customers (for example 2% of the overall customers) of a company can account for more than 25% of the company’s product sales or service usage. For example, Perfetto and
Woodside identified that a very small portion (approximately 1.2%) of over 20,000 respondents accounted for 27.4% of all types of casino plays. Other studies have reported similar extreme frequent purchase behaviour without classifying them as EFB. For example, Woodside, Cook and Mindak (1987) found a group (e.g. 4% of the Americans accounted for 70% of the total airline trips) that exhibit EFB. The current study also identifies the existence of such a group in race wagering where approximately 1% of the race wagerers accounted for 31% of the total betting revenue.
References


Appendixes

Appendix A

Pari-mutuel betting: all bets are placed in a pool and payoff odds are calculated by sharing the pool among all winning bets. When the winning outcome occurs, the company deducts commission out of the pool and then distributes the money remaining in the pool to the bettors who had bet on the winning outcome. This payment is the dividend. The dividend is uncertain until the winning event has actually occurred. This form of betting is described by the productivity commission (2010) as “tote-odds”.

Fixed-odds betting: a form of wagering against odds offered by the bookmakers. The bettor knows the exact odds he/she will receive when they place a bet.
Appendix B

Different betting types offered by the gambling service provider

Win:
A Win bet is simply picking the first place getter in a race.

Place:
A Place bet pays if the selected horse comes in first, second or third.

Eachway:
An Eachway bet is a combined Win and Place bet on one transaction. If the selection wins the bettor is paid both the Win and Place dividends. If the selection comes in second or third, then the bettor is paid the Place dividend.

Quinella:
A Quinella is a bet on the first and second placegetters in any one race. The chosen two runners can come first and second in any order.

Exacta:
An Exacta is a bet on the 1st and 2nd placegetters in any one race. The chosen two runners must come 1st and 2nd in the nominated order of the bet.
First4:

A Single First 4 requires the bettors to correctly select the First 4 runners in order in any one race.

Running Double:

To win a Running Double the bettor must select the winner of any two consecutive races at the one race meeting.

Daily Double:

The gambling service provider chooses groups of two races. The bettors can bet on these as Doubles. To win a Double, the bettors must select the winner of the 1st and 2nd ‘legs’ of the nominated races. There are two types of doubles - Daily Doubles and Consolation Doubles, both cover different races. To win a Daily Double the bettors pick the winner in the two nominated races.

Consolation Double:

To win the Consolation Double, the customer must pick the winner in Leg 1 and first or second in Leg 2 (gives the bettors two chances to win).

Quadrella:

The gambling service provider chooses four races. To win a Quadrella the bettors need to pick the winners of all four legs.
**Triwin:**

A Triwin bet is a fixed odds bet. The gambling service provider chooses four races. To win a Triwin the bettor needs to pick a placegetter in the first three legs and the winner of the fourth leg.

**All-up:**

An All-up bet requires the selection of one or more runners in a race with the winning dividend being reinvested in a subsequent race or races. A minimum of 2 races with a maximum of 4 races is required for an All-up bet type. All selections must be successful for the bettor to collect a dividend. Each leg can be selected for either a win or a place.
Appendix C

Number of betting days histogram

Total bet value in dollar histogram
Appendix D

D-1: Duration of stay vs. tenure

A confusion may arise between the terms “duration of stay” and “tenure”. The difference is in fact a semantic one. “Duration of stay” term has been used on tourism and leisure industry (see Ortega, et al., 2014; Peter & Anandkumar, 2015; Quintal et al., 2014). The term has also been used in retail settings (Zhang & Breugelmans, 2012). For example, Peter & Anandkumar (2015) mentioned that their respondents’ duration of stay ranged from 2-3 days, 4-7 days, etc. On the other hand, Dawes (2009) used the term “tenure” to describe how long a customer remains a customer. For example, he described the tenure of his respondents ranging from 1 to 15 years. Thus, it can be concluded that duration of stay and tenure are essentially the same thing. This thesis uses tenure instead of duration of stay to describe the length of time a bettor remains with the gambling service provider.

D-2: Arguments for not considering censored data

One potential alternative was to consider the data as censored and use survival analysis. However, the aim of the current research does not require this analysis. “The primary goal of survival analysis is to describe the proportion of cases surviving at various times, within a single group or separately for different groups” (Tabachnick & Fidell, 2007, p. 507). In the current context, applying survival analysis would answer questions such as what proportion of bettors will stop transacting at a given point of time. However, the aim of this type of analysis does not fit in with the main objective of the thesis which is to investigate the relative effects of frequency of purchase and duration of stay/tenure on firm revenue.
D-3: **Light customers**

One argument against this notion could be that for an individual customer duration may lead to higher revenue. A light customer over 3 years is worth more than light customers over 1 year. However, this may not be necessarily true. In the present context, there were three variables to consider: number of bets (i.e. frequency of betting), number of days bet and total dollar amount bet. If a 3 year tenure bettor has fewer betting days compared to the 1 year tenure customer, they will account for less revenue assuming the same bet size.

D-4: **Results of multiple regression analysis for bettors from 2001-2003**

<table>
<thead>
<tr>
<th>Model summary</th>
<th>Analysis of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.90</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.80</td>
</tr>
<tr>
<td>Std. error of estimate</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of square</td>
<td>22021.12</td>
</tr>
<tr>
<td>df</td>
<td>2</td>
</tr>
<tr>
<td>Mean square</td>
<td>11010.56</td>
</tr>
<tr>
<td>F ratio</td>
<td>12014.42</td>
</tr>
<tr>
<td>Sig.</td>
<td>.000*</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables in equation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression coefficients</td>
<td>B</td>
</tr>
<tr>
<td>Standardized coefficients</td>
<td>$\beta$</td>
</tr>
<tr>
<td>t-value</td>
<td>Sig.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.50</td>
</tr>
<tr>
<td>Log duration</td>
<td>.002</td>
</tr>
<tr>
<td>Log number of bets</td>
<td>.916</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Note: *p&lt;0.01</td>
<td></td>
</tr>
</tbody>
</table>

Dependent variable: Log total bet value.