Customer Retention: Reducing Online Casino Player Churn Through the Application of Predictive Modeling

Eunju Suh
Matt Alhaery

Abstract

With the potential expansion of legalized online gaming in the United States as well as in the global market, customer retention is critical to the continued growth and success of an online casino. While customer churn prediction can be an essential part of customer retention efforts, it has received very little attention in the gaming literature. Using historical online gaming data, this study examines whether player churn (attrition) can be predicted through an application of a decision tree data mining algorithm called Exhaustive CHAID (E-CHAID). The results of this empirical study suggest that the predictive model based on the E-CHAID method can be a valuable tool for identifying potential churners and understanding their churn behavior. Additionally, this study shows how the classification rules and propensity scores extracted from a decision tree churn model can be used to identify players at risk of churn. The patron play and visitation parameters that are closely associated with churn are also discussed. This study contributes to the gaming literatu1re by focusing on online players' churn prediction through a data-driven approach. Finally, it discusses proactive approaches for churn prevention.

Key Words: Churn, Decision Tree, Data Mining, iGaming, Online Gaming, Customer Relationship Management, Marketing

Eunju Suh
Associate Professor
Chaplin School of
Hospitality & Tourism
Management
Florida International
University
suhe@fiu.edu

Matt Alhaery
Analytics Solution Architect
IBM
Global Business Services
malhaer@us.ibm.com

Introduction

In the U.S., online gaming, or iGaming, is currently permitted for the residents of Nevada, New Jersey and Delaware, although online poker is the only legalized form of online gaming in Nevada. While New Jersey produced 91.2% ($122.9 million) of the online gaming revenues in the U.S for the first full calendar year of 2014, gaming revenues in New Jersey were much lower than the initial forecast of $1.2 billion cited in the New Jersey FY 2014 budget (RubinBrown, 2015). Despite the lower-than-expected performance of online gaming in 2014, the revenue growth in the overall U.S. gaming industry for the year of 2014 was credited mainly to online gaming and limited stakes gaming (RubinBrown, 2015). In 2014, the U.S. gaming industry experienced an increase of 0.5 percent in gaming revenues over 2013 with a record high $68.7 billion in gaming revenues (RubinBrown, 2015). In April 2016, New Jersey set a new record for gaming revenues, generating close to $17 million. This totaled approximately $342 million in gaming revenues since the launch of online gaming in December 2013 (Levine, 2016). With respect to Delaware, its online gaming appears to be in good shape financially in spite of the fact that the state has relatively small population of approximately 900,000 people (Glatzer, 2016). According to the revenue statistics reported by Delaware Lottery, online gaming brought in over $1.8 million in total net revenue from January to July in 2016, surpassing the previous net revenues of $1.4 million in FY 2014 and $1.8 million in 2015 (Delaware Lottery, 2016).

With the recent sluggishness in the global economy, it has been challenging for some land-based casino operators to sustain continued growth and expansion, especially in a saturated and competitive market. In fact, the gaming revenues of brick-and-mortar commercial casinos declined 0.6 % in 2014, producing $65.6 billion (RubinBrown, 2015). Furthermore, the closure of several casinos in Atlantic City is often attributed to the legalization of gambling in neighboring states (Magyar, 2014). Given the challenges that the commercial gaming market faces, some industry pundits and published reports suggest that the growth and expansion of online gaming can complement the efforts of the traditional land-based casinos in expanding their customer bases and boosting overall gaming revenues (e.g., Border, 2014; H2, 2014; Philander & Fiedler, 2012; Versace, 2014). In fact, Atlantic City had a substantial increase in gaming profits for the third quarter of 2015, up 61% over the same quarter last year (Amsel, 2015). A large part of the growth was attributed to the iGaming profits from the brick-and-mortar casinos offering online gaming (Amsel, 2015). Furthermore, the TropicanaCasino.com online casino at the Tropicana Casino & Resort in Atlantic City was attributed to revitalizing lapsed casino customers as 40% of online players in its online customer database were those who had not visited the property for a year or longer (Levine, 2016).

In many ways, iGaming for existing casinos is a startup operation. In fact, for several casinos, online gaming is set up as a separate operating entity. As in the early stages of any start-up company, acquiring new customers is of prime importance, and online casinos are not an exception. However, retaining profitable customers and developing loyal and long-term customers can ultimately drive the continued growth and success of a business (Kotler & Keller, 2015; Reichheld & Sasser, 1990). While varying by industry, a 5 percent reduction in the customer defection rate can increase profits from 25% to 85% (Reichheld & Sasser, 1990). In the wireless telecommunications industry, Mozer et al. (2000) claimed that a 1% reduction in the monthly churn rate from 2% to 1% would generate an increase of $54 million or more in annual earnings and an increase of approximately $150 million in shareholder value for a carrier with 1.5 million subscribers. Similarly, Van den Poel and Lariviere (2004) demonstrated that a 1% improvement in the retention rate for a financial services company could result in substantial increase in profit.

With respect to customer retention efforts in the gaming industry, time-based definitions for customer inactivity or churn can often be observed as the primary trigger. While this cut-off time can vary by property, the following examples show that a 90-day period of inactivity is used in some casinos catering mainly to local clients. In the Maryland Live casino located in Maryland, U.S., a player is defined as inactive if there is no active play for 90 consecutive days according to its rewards program rules (Maryland Live! Casino, 2015). Once the player is defined as inactive, reward points and comp-dollars expire off the player's account (Maryland Live! Casino, 2015). Another example is the Boarding Pass rewards program of Station Casinos which has several properties in Nevada, U.S. Its rewards program has multiple tiers, and players are required to earn a certain amount of points within 90 days to qualify for tiered status (Gaming Market Advisors, 2006). Otherwise, tiered status can be lost (Gaming Market Advisors, 2006). At Caesars Entertainment (formerly Harrah's), which has numerous properties across the U.S.A, all reward credits in the player's account expire if there is no account activity for a period of six months (Gaming Market Advisors, 2006). Retention efforts are often initiated by a casino after a certain time period as shown in the following quote by Joe Witterschein, Vice President, Marketing Services, The Innovation Group: "Here's general gaming for example, capture minimum of 25 percent of top tier carded players that have not visited in 90 days" (33rd annual symposium on racing & gaming, 2006, p.4). One common approach for the effort to gain the patrons back is a "retention" direct mail offer sent to inactive players (Lal, 2004).

While the above-mentioned time-based approach may be simple to understand and execute, the approach to define churn status based on an arbitrary time frame can fail to account for the varying differences in individual subjects' visitation cycles. Each customer may have a different visitation pattern and a time interval between visits, and some players may have a longer visitation cycle than others. For example, if a 90-day cut-off point is used to determine a player's inactivity, a player with a longer visitation cycle (i.e., a 180-day cycle) appears to have churned while the player is in fact still active. In other words, the player is not really inactive, but rather has a long visitation cycle, and his/her repeat visit is imminent. If reactivation offers were sent to the player based on the 90-day inactivity criterion, it can result in a costly "incented" visit as the player could have returned to the casino without a retention offer. In fact, it is pointed out that casinos often fail to differentiate customers who visit casinos regularly (e.g., once a month) from those who rarely visit (e.g., once every two years) by sending promotional offerings to the infrequent visitors as often as they would send to the frequent visitors (Hsieh, 2009). Consequently, this can result in the inefficient use of marketing dollars (Hsieh, 2009). With respect to the casino patrons with high frequency, as frequently as daily, their inactivity for 30 days may not be easily detected based on the 90-day cut-off point for inactivity. On the contrary, they can be declared still active even after 60 days of no play according to the 90-day cut-off criterion. However, 60 days could provide enough time to casino customers with daily visitation pattern to establish their status with competitors. Consequently, casino mangers may miss the opportunity to prevent customer churn by waiting 90 days until customers indeed churned. For these reasons, the static time-based determination of inactivity, while simple, may not be able to accommodate individual differences in visitation cycle. This can be more true especially in the online casino context where customers can visit a casino without physically being there and easily switch from one online casino to another. The opposite can be true when a patron has a very erratic pattern, where the 90 days of inactively is within his/her normal level of activity. These individuals can be seen as churners and offered a reactivation mailer, when they are still active.

Purpose of the Study

While customer churn prediction is an important area in customer retention and relationship management (Buckinx & Van den Poel, 2005; Van den Poel & Lariviere, 2004), little is known about casino customers' churn behavior in the gaming literature. This study focuses on customer retention in the online gaming industry by predicting player churn propensity (the likelihood to leave or stop playing at an online casino) at the individual player level. More specifically, this study examines whether a data mining algorithm can be an effective method to predict customer churn based on online players' historical and tracked gaming data. Furthermore, it identifies the important churn predictors and predicts online players' churn behavior by incorporating an individual player's visitation and play pattern in the churn prediction model. Given the lack of research on customer churn in the gaming industry, this study would provide a better understanding of online players' churn behavior and contribute to the gaming literature on attrition behavior. It will also lay the foundation for future research on the analysis of online gaming behavior in the growing online gaming market. The methodological approach advanced herein introduces a data-drive method to predict which customers are likely to churn based on individual player's gaming and demographic data. The application of this approach can help casino managers identify potential churners more precisely at the earliest possible point and eventually develop more targeted retention programs geared towards the customers at high risk of churn. This in turn will not only help them proactively prevent customer attrition but also optimize their marketing campaign and spend based on insights gained by analyzing customer behavioral data. Furthermore, the targeted retention strategy will help casino managers lower their direct marketing costs and save substantial amount of marketing dollars.

Literature Review

Research on Customer Churn

Churn prediction has been an important topic for customer retention and relationship management and discussed in various fields including retail, finance, telecommunication and insurance (Buckinx & Van den Poel, 2005; Van den Poel & Lariviere, 2004). Researchers in marketing and data mining fields have examined various models to predict customer churn (Buckinx & Van den Poel, 2005; Coussement & De Bock, 2013; Kim, 2012). One such example is Lai and Zeng (2014) which examined customer churn behavior in the Chinese digital libraries. In their study, churners were defined as those who did not download records in the library's data warehouses longer than a year. Using Survival Analysis, the authors found that the likelihood of churn was very high during the first three months after the customer registered for library services. Additionally, the authors segmented library customers into three different clusters, "key customers," "common customers" and "loyal customers" based on behavioral characteristics such as the number and frequency of downloads, the number of days between the first and the last downloads (time interval) and the ratio of the number of downloads to the time interval. Of the three clusters, "common customers," which was characterized by a low demand for articles and a short time interval, exhibited a higher likelihood of churn than the other two clusters. In the retail industry, Buckinx and Van den Poel (2005) examined grocery shoppers' churn using data mining algorithms such as Neural Networks and Random Forests. They built a model to predict customer churn based on individual shoppers' transactional and demographic data provided by a global retailer offering fast-moving consumer goods. The results of their study showed that potential churners could be successfully predicted through an application of data mining algorithms.

In the mobile telecommunications service industry, Ahn, Han and Lee (2006) examined customer churn behavior and the factors influencing it. The authors argued that previous studies have focused on a limited set of factors such as customer dissatisfaction and relied mainly on surveys to understand customers' perceived experience with service providers. To fill this research gap, they examined customer churn behavior based on a large data set of customer transactions and billing information provided by a leading provider of mobile telecommunications service in South Korea. Using logistic regression analysis, they built a model to predict customer churn and explored the mediating effects of a customer's status (e.g., active use, non-use and suspended) on the relationship between the churn determinants and customer churn. Customer status was examined in the model as it could be an early indicator of churn. Along with the customer status variable, a number of other variables associated with switching costs, service usage and customer dissatisfaction were examined as the potential predictors of customer churn. Churn in their study was defined as the termination of service subscription during a designated time period. In their study, the authors suggested that active intervention for the customer group which had status changes is crucial for churn prevention.

In the gaming literature, there has been limited attention given to the prediction of customer churn. Coussement and De Bock (2013) applied data mining algorithms to the gaming data of online poker players in order to predict customers at risk of churn. They used online gaming data for 17 months to predict the player's visit to the online gaming website during the next four-month time period after the 17-month time period. The results of their study suggested that ensemble models which combined the prediction results of several algorithms outperformed the models utilizing a single algorithm. Using the play data drawn from the database of a major hotel-casino in the U.S., Suh and Alhaery (2015) built a model to predict the likelihood of a patron returning to the subject casino. More specifically, the authors applied a decision tree data mining algorithm, C5, to the gaming data to predict any given customer's active or inactive gaming status. The results of their study showed that the C5 model was able to predict customer status with high accuracy based on the historical play data. The authors suggested that casino managers pay special attention particularly to the players who were predicted to return to the casino by the model and yet have not returned to the property. Given that irregular deviation from the historical visitation cycle and play amounts can be an early sign of defection, it is recommended that casino mangers develop retention strategies towards the players who are likely to defect (Suh and Alhaery, 2015).

Modeling Churn Behavior Using a Decision Tree Data Mining Algorithm

In the prediction of customer churn behavior, researchers have employed various data mining algorithms (Lai & Zeng, 2014). In the gaming literature, decision tree algorithms were employed in several studies. Braverman et al. (2013) applied Chi-squared Automatic Interaction Detector (CHAID) to individual players' transaction data at an online betting website to identify high-risk players. Coussement and De Bock (2013) and Suh and Alhaery (2015) also utilized decision tree approach, Classification and Regression Trees (CART) and Exhaustive CHAID (E-CHAID) respectively, to predict player attrition behavior. Decision tree analysis methods are best known for mining large data sets. A decision tree divides any given population into subgroups based on the strongest predictors that provide the greatest degree of separation of one group from another in relation to the target variable. The generated decision tree models are relatively easy to understand and act upon (Koh, 2004; McCarty & Hastak, 2007, Sung et al., 1999). CHAID is a decision tree classification algorithm based on chi-square statistics (IBM Knowledge Center, 2012). It handles large data sets and generates a graphical decision tree and classification rules to categorize subjects into their predicted classes (SPSS Inc., an IBM Company., 2010). Furthermore, it assigns a propensity score based on the statistically-derived confidence level for each prediction that is made (IBM Knowledge Center, 2014). Exhaustive CHAID (E-CHAID) which is an extension of CHAID, does a more thorough search than CHAID in that it can merge similar categories within a variable until only two categories remain. (IBM Knowledge Center, 2012; SPSS Inc., an IBM Company, 2010).

The above-mentioned features of CHAID and E-CHAID can be important in analyzing the transactional gaming data of individual customers. Iaci and Singh (2012) pointed out the extremely sparse nature of player tracking data found in many casinos' customer databases (e.g., a large number of zero entries in the customer play data). Hence, applying traditional statistical methods such as cluster analysis and principal component analysis to the sparse- and high-dimensional data may not produce informative and interpretable dimension reduction (Iaci and Singh, 2012). Additionally, gaming data is rather skewed, with a large concentration of players having very little play. In comparison to the traditional statistical analysis, decision trees are known for its robustness to missing data and assumptions regarding the distribution of input variables (Kim, 2009; Neville, 1999). Another trait of decision trees is that it is less sensitive to outliers or extreme values because of binning which transforms a scale numerical variable into categories of discrete values (Ray, 2016). This binning of variables reduces any noise or nonlinearity found in the variable distribution, thereby improving the overall accuracy of the predictive models (Mueller & Massaron, 2015). In decision trees, supervised binning is performed in which the intervals of a scale numerical variable are chosen to optimize the relationship of the variable with a nominal, target variable ("What is optimal," 2010). Among decision tree algorithms, the CHAID decision tree algorithm is known to be robust to the presence of missing values because it treats missing values as a categorical value (SPSS Inc., an IBM Company., 2010).

Defining Churn

While there is no common definition of churn behavior, termination of a contract and/or cancellation of an account are often used to define churn in a contractual setting (Lai & Zeng, 2014). In a non-contractual setting, customer activity status over a certain time period (e.g., account activity and frequency of purchase) and/or time interval between transactions or between the latest activity and the analysis point have been used to define churn (Ahn, Han & Lee, 2006; Buckinx & Van den Poel, 2005; Coussement and De Bock, 2013; Lai & Zeng, 2014). For example, Lai and Zeng (2014) used an arbitrary time point to define a churner in their study. In their study, churners were defined as those who did not download records in the library's data warehouse for more than a year. The average length of the relationship between customers and libraries was slightly less than two years. Similarly, churners in previous gaming research were often defined based on the subject's inactivity for a certain time period (e.g., Coussement & De Bock, 2013; Suh & Alhaery, 2015). For example, in a study done by Coussement and De Bock (2013), churn status was determined based on whether or not a player had a gaming record during the four-month time period after the 17-month data collection period.

Variables Affecting Customer Churn

In predicting customer churn, researchers generally relied on the data relevant to customer demographics and past behavior (Coussement & De Bock, 2013). For example, in Buckinx and Van den Poel's (2005) study on customer churn in the retail sector found that among the input variables for churn prediction, the Recency, Frequency and Monetary Value (RFM) variables representing customers' historical shopping behavior such as shopping frequency, amount spent and inter-purchase time exhibited strong discriminatory power in separating churners from non-churners. On the other hand, variables representing the amount of spending on store brands, loyalty points earned, and the number of categories purchased had low importance in their study. Finally, the authors mentioned the importance of a binary variable which indicated the presence of a shopper's demographic information in predicting customer churn. The authors noted that customers who were not willing to share their personal data with the company might have a higher propensity to churn (Buckinx & Van den Poel, 2005). In mobile telecoms industry, Seo et al. (2008) used the individual call details such as call time and drop call rates as well as data on service plan, handset, length of relationship and demographics to predict customer churn. Their research revealed that a long-term relationship and high-quality wireless connectivity were positively associated with customer retention. Furthermore, the authors stated strong influences of drop-call ratio and handset sophistication on customer retention behavior.

In the gaming industry, Coussement and De Bock (2013) relied on the online poker players' demographic and gaming data for the 17-month time period to predict a given player's churn status in the subsequent 4-month time period. The input variables derived from the data consisted of continuous, binary and categorical variables representing the recency, frequency and monetary value of online play. Other variables included inter-purchase time, length of relationship, availability of promotional playing funds, and demographic characteristics such as language, region of origin and gender. The results of their analysis revealed that variables representing recency ranked higher than the variables representing frequency and monetary value. Demographic variables failed to rank within the top 20 churn predictors. Similarly, Suh and Alhaery (2015) explored the factors associated with the player's likelihood of return to a casino based on the casino customers' demographic and past gaming data. The variables examined in their study included average bet amount, play time, casino trip frequency, and gaming wins and losses. Furthermore, the authors of the study explored additional variables such as the average number of days between trips and the standard deviation of play time between trips, which were derived from the gaming behavior variables. The goal of these derived variables was to capture any deviations from a patron's regular casino visit cycle and play pattern (Suh & Alhaery, 2015). The results of their study revealed that the derived variable representing the standard deviation of play time between trips was the most important predictor for the player's likelihood of return to a casino. Additionally, variables representing play time and casino trip frequency were important predictors of the player's return to a casino.

Research Gap and Proposed Churn Model

While there have been numerous studies on churn prediction across different industries, little research on customer churn has been published in the gaming literature. While Coussement and De Bock (2013) focused on the churn behavior of the online poker players, they pointed out that sports betting was likely to be the main interest of the poker players in their study, given that the data were from *bwin*, a sports betting operator. Alternatively, Suh and Alhaery (2015) applied a data mining algorithm to customers' actual gaming data. However, the focus of their study was about predicting the return of players to a land-based, traditional brick-and-mortar casino. Given the lack of empirical research on online players' churn behavior, the focus of this study was on predicting online players' churn behavior through the application of a data mining algorithm and identifying the predictors of online players' churn. For the prediction of online player churn, E-CHAID, a decision tree algorithm, was applied to the tracked gaming data following the previous gaming research which also used decision tree algorithms in analyzing attrition behavior (e.g., Braverman et al., 2013; Coussement & De Bock, 2013; Suh & Alhaery, 2015). A conceptual model to predict online players' churn was proposed in Figure 1.

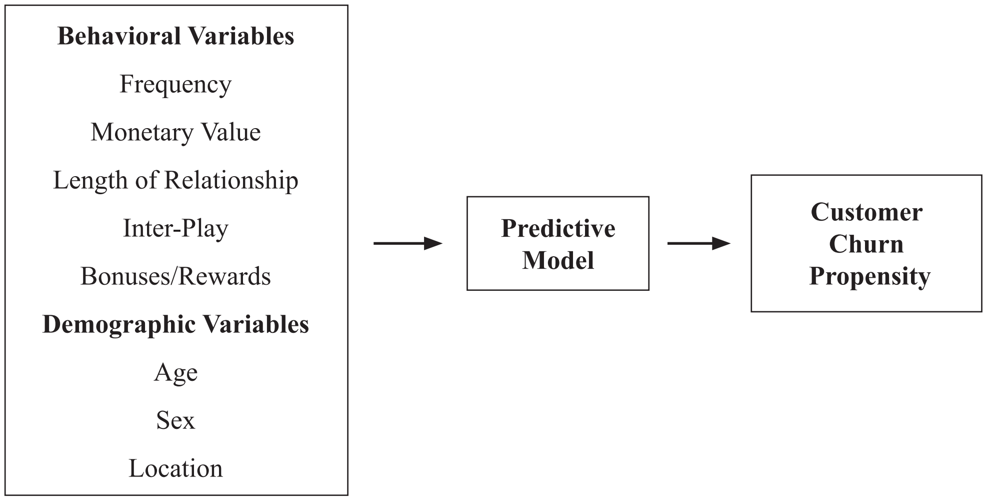

Figure 1. Conceptual Model for Churn Prediction

The conceptual churn model, variables and analytical approach adopted herein were derived from the previously reviewed research on customer churn and predictive modeling of customer behavior (e.g., Braverman et al., 2013; Coussement & De Bock, 2013; Suh & Alhaery, 2015). The input variables advanced in the proposed model were mainly based on the findings of Coussement and De Bock (2013) and Suh and Alhaery (2015) given that they were gaming-related. These studies have demonstrated the importance of the RFM-based measures in predicting customer churn behavior. Hence, the variables proposed in Figure 1 represent online players' gaming behavior such as the frequency and amount of play as well as their demographic characteristics such as gender and age. Additionally, ratio variables were created based on the findings of Coussement and De Bock (2013) in order to capture the player's level of gaming activity (e.g., frequency and deposit amount) relative to the length of relationship. The ratio variables were marked with "r" in their variable names (see Appendix 1). Furthermore, variables representing variations in gaming behavior over time (e.g., standard deviation of daily theoretical win and the same for the number of daily slot spins) were incorporated in the proposed model in order to capture possible deviations from a regular play pattern. According to the findings of Suh and Alhaery (2015), adverse deviations from the casino patron's historical visitation and play pattern could be an early sign of defection. In fact, variables representing such deviations (e.g., standard deviation of play time between trips) were important predictors of the casino customers' retention behavior in their study. With respect to the secondary customer data, researchers derived behavioral and demographic variables to be used in modeling specific behavior such as customer churn. In previous gaming research, researchers relied on real-life data such as the aggregated daily gaming volumes of slot machines and players' past gaming data to model gaming behavior (e.g., Lucas & Dunn, 2005; Lucas, Dunn & Singh, 2005; Lucas et al., 2004; Suh, 2012; Suh & Alhaery, 2015; Tanford & Lucas, 2011). These studies have demonstrated high model performance. Similarly, the goal of the proposed model was to predict customer churn among online players based on a set of the input variables which represented online players' demographics and gaming behavior.

Methodology

Data

The data for analysis was provided by an online casino operated by a land-based casino in the United States. A convenient sample drawn from the subject casino's database included the individual gaming data from March to September in 2015. Given that the time window for the churn behavior analysis and the definition of churn vary by study, the present study adopted the sample selection approach similar to that of previous research (Buckinx & Van den Poel, 2005; Coussement & De Bock, 2013). In both studies, customers who met a certain threshold were used for analysis. For example, Coussement and De Bock (2013) focused on the online poker players who had at least four gaming sessions on three different days during a 17-month time window. Buckinx and Van den Poel (2005) focused only on the loyal customers whose purchase frequency was greater than average and the ratio of the standard deviation of the inter-purchase time to the mean inter-purchase time was lower than average. Based on the review of the sample selection approaches in Coussement and De Bock (2013) and Buckinx and Van den Poel (2005) and the discussion with the managers at the subject online casino, players with a minimum of three play days during the seven-month time period were selected for analysis in the present study. This criterion was applied to exclude first-time and/or one-time players and yet to provide sufficient historical gaming activity of both churners and repeat customers. As a result, the sample included the gaming data of 2,639 players.

Variables

Input Variables. In the previously reviewed studies on customer churn (e.g., Braverman et al., 2013; Coussement & De Bock, 2013; Suh & Alhaery, 2015), researchers used the data for a certain time period to create a set of input variables which in turn were used to predict customers' churn behavior in the subsequent time period. Similarly, play data for the seven-month time period were used in the current research to derive the input variables for modeling churn behavior. The data set provided by the subject online casino contained individual customers' play data and demographic information. Using this data, additional measures that were found to be linked to customer churn in previous research were derived (e.g., Braverman et al., 2013; Coussement & De Bock, 2013; Suh & Alhaery, 2015). Gaming behaviors were expressed in the forms of Recency, Frequency and Monetary Value (RFM) as well as the minimum, maximum and standard deviation for some of the RFM variables. RFM variables mainly represent play frequency, wager amount, win/loss amount, and the number of days since the initial deposit/first play. Several other variables were created to express the monetary amount of incentives (e.g., deposit bonus) and rewards (e.g., loyalty points). Demographic variables were created to represent a player's age and gender. Additionally, a geographic variable was created to represent the distance between the land-based casino offering an online casino and a player's residence. A total of 60 input variables were created to predict the target variable (See Appendix 1 for the descriptions of the input variables).

Target Variables. In defining a customer's churn status (churner vs. non-churner), the following three behavioral measures were used in order to reflect an individual player's visitation pattern in the target variable: number of days since a player's last play (Recency), average days between play (Inter Playday Avg) and the standard deviation of days between play (Inter Playday SD). The player was coded as "1," meaning churner, if the player satisfied the following condition: Recency > (Inter Playday Avg + 2 x Inter Playday SD). Otherwise, the player was coded as "0", or non-churner. In other words, if the number of days since the player's last play was greater than his/her average days between play (inter-play days) plus two standard deviations of the average inter-play days, the player was deemed a churner. Two standard deviations of the mean were chosen because they cover 95 % of the variance. This churner definition resulted in 855 churners (32.4%) out of the 2,639 players in the sample. Regarding the inter-play days, the average number of days between play for the sample was 8 days with a standard deviation of 10 days. By using the individualized inter-play time instead of an arbitrary cut-off point, it is expected that the predictive model can produce more accurate predictions and churn propensity scores for individual players with different play patterns. Propensity scores indicate the likelihood of a specific outcome in reference to the target variable (IBM Knowledge Center, 2014). Given the definition of the target variable herein, propensity scores in this study reflect an adverse deviation from a given customer's normal play activity. Appendix 1 shows the definitions of the target and input variables used for modeling churn behavior.

Analysis

E-CHAID analysis was performed using the IBM SPSS Modeler data mining workbench (version 17.0). As mentioned earlier, approximately 32.4% of the sample belonged to the churner group while the rest of the sample, 67.6% belonged to the non-churner group. Due to an imbalance in sample sizes, a balanced sampling method, which oversamples, or boosts the smaller group, the churner group in this case, was performed for model estimation. As a result, the sample sizes in both churner and non-churner groups became approximately equal: 49.9% non-churner vs. 50.1% churner. Balanced sampling was employed because a predictive model tends to assign cases into a group where the majority of cases belong to in an attempt to improve prediction accuracy (SPSS Inc., 2009). With the balanced sample, the predictive model is expected to better detect the cases in a smaller group. In the current research, this group represents a churner group with players at high risk of churn. While the balanced sample was used for model estimation, the imbalanced, original data set was used to evaluate the model performance. The input variables were fed into the E-CHAID decision tree algorithm to predict customer churn behavior. The E-CHAID model produced a classification tree which consisted of a series of sub-groups that were different from one another with respect to the target variable.

Results

The following table shows the descriptive statistics of the input and target variables used for model estimation. All variables except the age, sex and target variables were continuous variables. The frequency and percentage of each category in these variables are displayed in Appendix 2. For all others, means and standard deviations are shown.

Churn Model and Predictors

The E-CHAID decision tree algorithm employed to construct a churn prediction model produced a model with the tree depth of 5 levels, 16 predictors, and 47 terminal nodes. Within these nodes, a prediction was made with respect to any customer's churn status. Of the 60 input variables fed into the decision tree data mining algorithm, 16 variables were found to be the significant predictors of churn. The variables associated with the length of relationship such as the number of days between first and last play dates and the same between initial deposit and last play dates were significant predictors of customer churn. The RFM-related variables such as the frequency of play and the amount of deposit were also significant. Finally, incentives such as play and deposit bonuses were significant. Table 1 presents the rankings of the statistically significant churn predictors in order of importance.

Table 1. Churn Predictors

Rank	Variables
1	Lor First Last Play
2	Bonus Amt Extra
3	Freq Slot Days
4	Lifetime Spin Slot Max
5	r Days W Deposit
6	Lor InitialDeposit LastPlay
7	Lifetime Dollar Dep. Amt
8	Bonus Amt Discretion
9	Daily Gaming Volume Avg
10	Lifetime Points Table
11	r Lifetime Bonus
12	Playbonus Amt
13	Sex
14	r Lifetime Spin
15	Lifetime Net Dep. Amt
16	Daily Table Volume Avg

Classification Rules and Propensity Scores for Churn Prediction

The churn model based on the E-CHAID decision tree algorithm generates the classification rules for churn prediction. By following these rules, a casino practitioner can understand how the model classified subjects into a churner group or a non-churner group, as well as the factors that are associated with a higher likelihood of churn. A sample excerpt of the classification rules generated by the churn model is as follows:

Lor First Last Play <= 5 [Mode: 1] (385)
    Lifetime Spin Slot Max <= 840 [Mode: 1] (242)
        Playbonus Amt <= 192 [Mode: 1] –> 1 (203; 0.985)

The rules show that if the number of days between first and last play days is equal or less than 5 days; the maximum number of slot spins is equal or less than 840; and the amount of play bonus is equal or less than 192, then the model classifies the player into a churner group coded as "1". Additionally, 200 of the 203 players who satisfied the above-mentioned conditions were correctly classified by the model as churners (98.5% accuracy). In addition to the classification rules, the churn model generates a predicted membership (e.g., churner or non-churner) for each subject along with the propensity scores which indicate the subject's likelihood of churn. For example, a 70% propensity score indicates that a given customer has a 70% likelihood of churn according to the churn prediction model.

Classification Accuracy

The churn model based on the E-CHAID decision tree had 80% accuracy by classifying 2,119 of 2,639 cases accurately. Table 2 shows the model's classification accuracies by category. For the churner prediction, the model accurately classified 670 (78%) of the 855 actual churners. For the non-churner prediction, the model accurately classified 1,449 (81%) of the 1,784 actual non-churners.

Table 2. Classification Accuracy

		Predicted		
		Non-churner	Churner	Accuracy
Actual	Non-churner	1,449	335	81%
	Churner	185	670	78%

Model Evaluation

To measure the model's sensitivity in discriminating between churners and non-churners, the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) was examined. The AUC value ranges from 0.5 and 1.0, and a higher value represents a model's better discriminability of a pair of classes (Fawcett, 2004). The AUC value of the churn model herein was 0.88.

Discussion

The results of the current research suggest that the E-CHAID decision tree can be an effective method for identifying potential churners as well as the important predictors of player churn – at least at the subject online casino. With respect to the predictors of player churn, the present findings are broadly consistent with those of Coussement and De Bock (2013) and Buckinx and Van den Poel (2005) which found RFM-related variables as the most important churn predictors. For example, variables representing the frequency of play as expressed in the number of betting sessions in Couseemen and Debock's study was an important predictor of churn. Similarly, the number of play days and the number of slot/table game spins in the present study were found to be significant. Furthermore, the length of relationship as expressed in the number of days since first betting in Couseemen and Debock's study and the number of days between first and last play dates in the present study were important. With respect to monetary value, the variables representing the wager amount/gaming volume were significant in both studies. Additionally, the amount of deposit, the number of days between initial deposit and last play dates, and the number of deposit days relative to the length of relationship, which were unique to the present study, were found to be the important predictors of churn. Among demographic variables, a variable indicating gender was moderately significant in the present study while it was not in Coussement and De Bock's study. With respect to the variables associated with promotions and rewards, the amount of bonus was an important predictor of churn in the present study. Similarly, the average monetary value of coupons per shopping trip was moderately important for churn prediction in Buckinx and Van den Poel's study.

Managerial Implications

Along with the predicted churner/non-churner membership, the churn model based on the E-CHAID decision tree algorithm produced the churn propensity scores for each player. By sorting the customers based on their churn propensity scores (from highest to lowest), casino managers can identify potential churners with high risk of churn, and develop a targeted retention program for them. Among those who were predicted to be churners according to the model, some players could be currently non-churners. While they could be the misclassified cases, they do have a higher propensity of churn and they will most likely churn in the near future. Because their gaming behaviors were similar to those who had churned in the past, the churn model assigned them to a churner group. While they are still active, casino managers would be wise to pay attention to them considering their high predicted propensity to churn. A special retention program can be designed to prevent these seemingly-active customers from churning, especially the ones with high gaming values.

Offering more generous promotional incentives than usual and/or sending customized messages to these players, especially the ones with high gaming value and high risk of churn, may increase the likelihood of their response to the retention offer. This can possibly motivate their return to the casino and prevent departures from their usual visitation cycles and play patterns. However, higher incentive amounts do not necessarily guarantee higher profits and positive cash flows as play incentives can replace some of the player's out-of-pocket expenses for gaming (Lucas et al., 2005; Suh 2012).

Furthermore, some customers may eventually return to the casino without any retention offers. In that case, the retention offers can over-incentivize players' trips to the casino. These patrons may simply be lapsed players and would have come back anyways. Instead, they could receive a generous retention offer from a casino. This represents wasteful marketing spending which could have been allocated to other customers who are truly at risk of attrition. Hence, it is recommended that casino managers monitor any changes in the respondents' gaming behavior and compare them with those of non-respondents (Suh & Alhaery, 2015). Furthermore, the contribution of retention offers to player retention rate and profit should be evaluated.

With respect to those patrons who were predicted to be non-churners by the model, some of them belonged to a churner group according to the churn definition used in the current study. Despite their defined membership (actual), they were predicted to be non-churners by the E-CHAID classification model because their behavior was similar to that of non-churners. Given that they were predicted to be non-churners, it is possible that repeat visits among some of these actual churners are imminent. It may be that they have not churned but rather have lapsed and their visitation cycles are erratic. Because of their irregular behavior, they were defined as churners according to this study's churn criterion, but the churn prediction model assigns them to a non-churner group based on the analysis of the player's multifaceted gaming behavior. Hence, casino managers may want to review these players' gaming behavior and values along with their churn propensity scores in order to determine the proper course of action. Considering their possible irregular visitation cycles, the ideal time to contact them should be carefully determined. Waiting too long before initiating retention activity could result in the loss of customers. Along with the ideal time to contact a patron, the level of a reactivation offer should be carefully determined not to over-incent the player. Alternatively, a promotional offer that can shorten the time interval between a customer's two successive play sessions can be designed if the customer's vitiation cycle is longer than average.

Overall, the application of the predictive churn model advanced herein can help casino managers identify potential churners before they churn. Once the potential churners are identified, a more effective retention program and targeted marketing messages can be created and deployed to intervene and prevent customer attrition. If a customer appears to have churned, casino managers can prioritize their reactivation efforts according to the customer's churn propensity scores. Those customers with low propensity scores could be relatively easy to convert into active players in comparison to those with high churn propensity scores. It may not be worth spending the time and marketing dollars to convert some of the players with high churn propensity scores and low gaming values to the casino.

Using churn propensity scores, casino marketers can select target customers for their retention efforts more wisely. Casino managers can focus on the players that show the early signs of churning. By precisely marketing to prospects that are most likely to churn or convert them into active players, casino managers can increase response rates and gaming revenues and eventually achieve more efficient use of marketing dollars and greater return on investment. Deployment of the propensity scores can also identify potential churners who have had a recent visit but whose play has diminished in their recent trips. This is because the propensity scores are based on the multifaceted gaming behavior of an individual, and they are a reflection of the level of deviation from the individual's normal gaming activity. On the contrary, the time-based cut-off point applies the same, single criterion (e.g., 3-month inactivity) to every player regardless of the individual differences in visitation cycle and play pattern to determine the player's churn status. Hence, this approach may not be able to detect the above-mentioned seemingly-active and yet potential churners with diminishing play.

Limitations & Suggestions for Future Research

The results of this study are based on the customer play data from a single online casino. Hence, the study should be replicated for other data sets collected at different online casinos over different time periods to assess whether the results can be generalized. Furthermore, more research to define churners in the gaming industry and determine an appropriate time window for the construction of a churn prediction model is recommended. With respect to the input variables examined herein, additional behavioral and demographic variables as well as other data mining algorithms should be explored to improve the churn model's predictive performance. Furthermore, the sample used for analysis can be divided into different groups based on player value and demographics. For each group, a separate churn model can be developed, and a comparative analysis can be performed to explore the differences among groups. Finally, a study to determine the ideal time to refresh or modify an existing churn model is recommended. A churn model's predictive power is likely to decline as consumer behavior or economic conditions change. Given that the data came from an online casino which does not have the same geographic restriction as a land-based casino, online players may shift their gaming behavior faster than offline players. However, frequently updating and/or rebuilding a model can be costly and time consuming. Hence, finding the ideal time to update the model as well as creating a model that is robust over time can be important for casino managers in order to sustain the desired model accuracy while achieving cost and time efficiency. Finally, future research on predicting the potential gaming values of online players can help online casino managers maximize player value and proactively approach the potentially high-value customers with high churn risk.

References

Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications policy*, *30*(10), 552-568.

Amsel, P. (2015). Atlantic City casino profits jump three-fifths in Q3 but city could be broke by April. Retrieved December 19, 2015 from http://calvinayre.com/2015/11/24/casino/atlantic-city-casino-profits-jump-three-fifths/

Border, C. (2014). Can online gambling help brick-and-mortar casinos return to profitability? Retrieved December 19, 2015 from http://www.marketingresults.net/blog/post.php?title=Can+Online+Gambling+Help+Brick-and-Mortar+Casinos+Return+to+Profitability?

Braverman, J., LaPlante, D. A., Nelson, S. E., & Shaffer, H. J. (2013). Using cross-game behavioral markers for early identification of high-risk internet gamblers. *Psychology of Addictive Behaviors, 27*(3), 868-877. doi:10.1037/a0032818

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252–268.

Coussement, K., & De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research, 66*(9), 1629-1636. doi:10.1016/j.jbusres.2012.12.008

Delaware Lottery. (2016). DELAWARE IGAMING Net Proceeds. Retrieved from http://www.delottery.com/games/igaming/finalstats.asp

Fawcett, T. (2004). ROC graphs: notes and practical considerations for data mining researchers. Retrieved from http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf

Gaming Market Advisors. (2006). An analysis of Harrah's Total Rewards Players Rewards Program. Retrieved from http://globalmarketadvisors.com/wp-content/uploads/2015/05/Total-Reward-Program-White-Paper.pdf

Glatzer, J. (June 16, 2016). Online gaming doing well in Delaware, *Delaware Sharelines*. Retrieved from http://www.pokernews.com/news/2016/06/online-gaming-booming-in-delaware-25054.htm

H2. (2014). New Jersey interactive gaming market still can reach its potential, according to the first of a series of new executive market report packages by H2 Gambling Capital. Retrieved December 19, 2015 from http://h2gc.com/news/general/new-jersey-interactive-gaming-market-still-can-reach-its-potential-according-to-the-first-of-a-series-of-new-executive-market-report-packages-by-h2-gambling-capital

Hsieh, C. (2009). Casino CRM: Issues and some implementation, *Communications of the IIMA*, *9*(1), Retrieved from http://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1093&context=ciima

Iaci, R., & Singh, A. K. (2012). Clustering high dimensional sparse casino player tracking datasets. UNLV *Gaming Research & Review Journal, 16*(1), 21-43.

IBM Knowledge Center. (2012). CHAID node. Retrieved from http://www-01.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/chaidnode_general.htm

IBM Knowledge Center. (2014). Decision tree/rule set model nugget settings. Retrieved from https://www.ibm.com/support/knowledgecenter/SS3RA7_16.0.0/com.ibm.spss.modeler.help/clementine/rulemodel_output_settingstab.htm

Kim, Y. (2009). Boosting and measuring the performance of ensembles for a successful database marketing. *Expert Systems with Applications, 36*(2), 2161-2176. doi:10.1016/j.eswa.2007.12.036

Kim, Y. & Moon, S. (2012). Measuring the success of retention management models built on churn probability, retention probability, and expected yearly revenues, *Expert Systems with Applications*, *39*, 11718–11727.

Koh, H. C. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal, 19*(3), 462-472.

Kotler, P., & Keller, K. L. (2015). *Marketing management*. Upper Saddle River, N.J: Pearson *Prentice Hall*.

Lai, Y., & Zeng, J. (2014). Analysis of customer churn behavior in digital libraries, *Program: electronic library and information systems*, *48*(4), 370-382.

Lal, R. (2004). Harrah's Entertainment Inc. *Harvard Business School case*, 9-502-011.

Levine, A. (2016). U.S. iGaming reaches milestone third year. Retrieved June 4, 2016 from *https://www.linkedin.com/pulse/us-igaming-reaches-milestone-third-year-alana-levine*

Lucas, A. F., & Dunn, W. T. (2005). Estimating the effects of micro-location variables and game characteristics on slot machine volume: A performance-potential model. *Journal of Hospitality and Tourism Research, 29*(2), 170-193.

Lucas, A. F., Dunn, W., & Singh, A. K. (2005). Estimating the short-term effect of free-play offers in a Las Vegas hotel casino, *Journal of travel & tourism marketing. 18*(2), 53-68.

Lucas, A. R, Dunn, W. T., Roehl, W. S., & Wolcott, G. M. (2004). Evaluating slot machine performance: A performance-potential model. *International Journal of Hospitality Management, 23*(2), 103-121.

Maryland Live! Casino. (2015). Live! rewards® » Frequently asked questions. Retrieved from https://www.marylandlivecasino.com/live-rewards/frequently-asked-questions/

Magyar, M. (2014). Explainer: drop in casino tax revenue reflects industry's decline in new jersey. Retrieved from http://www.njspotlight.com/stories/14/12/01/explainer-drop-in-casino-tax-revenue-reflects-industry-s-decline-in-nj/

McCarty, J. A. & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, *60*(6), 656-662. doi: 10.1016/j.jbusres.2006.06.015

Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Churn reduction in the wireless industry. In S. A. Solla, T. K. Leen, & K.-R. Mueller (Eds.), *Advances in Neural Information Processing Systems, 12*, 935-941. Cambridge, MA: MIT Press. Retrieved from http://papers.nips.cc/paper/1749-churn-reduction-in-the-wireless-industry.pdf

Mueller, J. P., & Massaron, L. (2015). *Python for data science for dummies*. Hoboken, NJ: *John Wiley & Sons, Inc.*

Neville, P. G. (1999). Decision trees for predictive modeling. Retrieved from https://datamining.bus.utk.edu/Documents/Decision-Trees-for-Predictive-Modeling-(Neville).pdf

Philander, K. & Fiedler, I. (2012). Online poker in north America: empirical evidence on its complementary effect on the offline gambling market, *Gaming Law Review and Economics, 16*(7-8): 415-423. doi:10.1089/glre.2012.1675.

Ray, S. (2016, January 10). A comprehensive guide to data exploration. Retrieved from http://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/

Reichheld, F. F. & Sasser, W. E. (1990). Zero defections: quality comes to services. Retrieved from https://hbr.org/1990/09/zero-defections-quality-comes-to-services#

RubinBrown. (2015). 2015 Gaming Stats. Retrieved December 18, 2015 from http://www.rubinbrown.com/2015-Gaming-Stats.pdf

Seo, D., Ranganathan, C. & Babad, Y. (2008). Two-level model of customer retention in the US mobile telecommunications service market, *Telecommunications Policy*, *32*, 182–196.
SPSS Inc., an IBM Company. (2010). *Introduction to PASW® Modeler and data mining*. Chicago, IL: SPSS Inc.
SPSS Inc. (2009). *Predictive modeling with PASW® modeler*. Chicago, IL: SPSS Inc.
Suh, E. (2012). Estimating the impact of free-play coupon value on players' slot gaming volumes. *Cornell Hospitality Quarterly. 53*(2), 134-143.
Suh, E., & Alhaery, M. (2015). Predicting customer retention for a hotel-casino property. *Direct Marketing Analytics Journal*, 74-78.
Sung, T.K., Chang, N., & Lee, G. (1999). Dynamics of modeling in data mining: interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, *16*(1), 63- 85.
Tanford, S. & Lucas. A. F. (2011). The indirect impact of casual dining and low-end gaming in destination and local casinos. *International Journal of Hospitality Management, 30* (3), 486-94.
Van den Poel, D., & Lariviere, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*, 196–217.
Versace, C. (2014). The time is here for online gaming. Retrieved December 19, 2015 from http://www.forbes.com/sites/chrisversace/2014/02/26/the-time-is-here-for-online-gaming/
What is optimal binning and how is it done? (2010, October 20). Retrieved from https://www-01.ibm.com/support/docview.wss?uid=swg21477421
33rd annual symposium on racing & gaming. (December 7, 2006). Marketing 101 – develop a marketing plan that works! Retrieved from http://ua-rtip.org/sites/ua-rtip.org/files/marketing_101.pdf

Appendix 1. Study Variables

Category	Name	Definition
Recency	Recency	Number of days since last play*
Frequency	Freq Play Days	Number of gaming days
	Freq Slot Days	Number of slot gaming days
	Freq Table Days	Number of table gaming days
	Freq Trips	Number of trips in which consecutive play days were counted as a single trip regardless of length
	Trip Length Avg	Average trip length in days
	Daily Playtime Avg	Average amount of play time per play day
	Lifetime Spin Slot	Number of slot spins since signup
	Lifetime Spin Slot Max	Maximum number of slot spins since signup
	Lifetime Spin Slot Min	Minimum number of slot spins since signup
	Daily Spin Slot Avg	Average number of slot spins per play day
	Daily Spin Slot SD	Standard deviation of daily slot spins
	Lifetime Spin Table	Number of table game spins since signup
	Lifetime Spin Table Max	Maximum number of table game spins since signup
	Lifetime Spin Table Min	Minimum number of table game spins since signup
	Daily Spin Table Avg	Average table game spins per play day
	Daily Spin Table SD	Standard deviation of daily table game spins
	r Lifetime Spin	Number of lifetime spins relative to the length of relationship (Lor First Last Play)
	r Freq Trip	Number of trips relative to the length of relationship (Lor First Last Play)
	Days Deposit Wo Bonus	Number of deposit days without deposit bonuses
	Days W Deposit	Number of days deposit was made
	r Days Deposit Wo Bonus	Number of deposit days without deposit bonuses relative to the length of relationship (Lor First Last Play)
	r Days W Deposit	Number of deposit days relative to the length of relationship (Lor First Last Play)
Monetary Value	Lifetime Dollar Dep. Amt	Deposit amount since signup
	Daily Deposit Avg	Average deposit amount per play day
	Lifetime Net Dep. Amt	Net deposit amount after various bonus deductions since signup
	Lifetime Win Loss	Amount of gaming win or loss since signup
	Lifetime Win Loss Slot	Amount of slot gaming win or loss since signup
	Lifetime Win Loss Table	Amount of table gaming win or loss since signup
	Daily Win Loss Avg	Average amount of win or loss per play day
	Daily Win Loss Avg Slot	Average daily win or loss for slot gaming
	Daily Win Loss Avg Table	Average daily win or loss for table gaming
	Daily Gaming Volume Avg	Average gaming volume per play day
	Daily Slot Volume Avg	Average slot gaming volume per play day

	Daily Table Volume Avg	Average table gaming volume per play day
	Daily T-Win Avg	Average daily theoretical win (T-Win), an indicative of player value to the casino
	Daily T-Win SD	Standard deviation of average daily T-Win
	r Lifetime Win Loss	Lifetime gaming win or loss relative to the length of relationship (Lor First Last Play)
	r Lifetime Dollar Dep. Amt	Lifetime deposit amount relative to the length of relationship (Lor First Last Play)
Length of Relationship	Lor First Last Play	Number of days between the first and last play dates
	Lor InitialDeposit LastPlay	Number of days between the initial deposit date and the last play date
Inter-Play	Inter Playday Avg	Average number of days between play*
	Inter Playday SD	Standard deviation of inter-play days*
	Inter First Deposit First Play	Number of days between the first deposit date and the first play date
Bonuses/Rewards	Playbonus Amt	Amount of bonus award (e.g., Free play) which required play
	Playbonus Amt Deposit	Amount of bonus award which required deposit
	Bonus Amt Extra	Amount of bonus award which required no deposit
	Bonus Amt Discretion	Amount of bonus award based on casino management discretion
	Lifetime Points	Amount of rewards points earned since signup
	Lifetime Points Slot	Amount of slot points earned since signup
	Lifetime Points Table	Amount of table game points earned since signup
	Lifetime Points Redm	Amount of rewards points redeemed since signup
	Daily Points Avg	Average points earned per play day
	Daily Points Redm Avg	Average points redeemed per play day
	Lifetime Cash Redm	Amount of cash rewards via point redemption since signup
	Daily Cash Redem Avg	Average amount of cash rewards via point redemption per play day
	r Lifetime Points	Lifetime points earned relative to the length of relationship
	r Lifetime Points Redm	Lifetime points redeemed relative to the length of relationship
	r Lifetime Cash Redm	Lifetime cash rewards relative to the length of relationship
	r Lifetime Bonus	Lifetime bonus amount relative to the length of relationship
Demographics	Age	Age
	Sex	Sex
	Location	Distance in miles between the customer's place of residence and the location of the subject casino offering online gaming
Target	Churn Status (target var.)	Whether a customer is a churner or not based on the pre-defined criteria ((Recency > (Inter Playday Avg + 2* Inter Playday SD))

Note. * Used for deriving a target variable

Appendix 2. Descriptive Statistics (N = 2,639)

Category	Name	M	SD
Recency	Recency	27.635	40.986
Frequency	Freq Play Days	15.579	20.005
	Freq Slot Days	14.484	19.673
	Freq Table Days	3.253	8.457
	Freq Trips	7.334	6.648
	Trip Length Avg	1.792	1.797
	Daily Playtime Avg	209.638	185.029
	Lifetime Spin Slot	7504.277	14941.366
	Lifetime Spin Slot Max	1309.98	1390.874
	Lifetime Spin Slot Min	43.966	99.528
	Daily Spin Slot Avg	419.286	425.725
	Daily Spin Slot SD	404.814	409.693
	Lifetime Spin Table	821.461	4970.135
	Lifetime Spin Table Max	183.677	597.074
	Lifetime Spin Table Min	7.169	73.56
	Daily Spin Table Avg	55.05	220.49
	Daily Spin Table SD	58.25	193.69
	r Lifetime Spin	183.41	381.35
	r Freq Trip	0.15	0.13
	Days Deposit Wo Bonus	8.25	14.76
	Days W Deposit	10.46	16.82
	r Days Deposit Wo Bonus	0.15	0.21
	r Days W Deposit	0.20	0.24
Monetary Value	Lifetime Dollar Dep. Amt	1636.67	6744.53
	Daily Deposit Avg	157.03	256.52
	Lifetime Net Dep. Amt	1171.55	6506.50
	Lifetime Win Loss	903.69	5829.91
	Lifetime Win Loss Slot	756.73	4443.62
	Lifetime Win Loss Table	146.96	3648.99
	Daily Win Loss Avg	59.95	247.24
	Daily Win Loss Avg Slot	50.81	165.17
	Daily Win Loss Avg Table	9.14	183.80
	Daily Gaming Volume Avg	1336.60	3480.56
	Daily Slot Volume Avg	825.27	1839.25
	Daily Table Volume Avg	511.33	2985.34
	Daily T-Win Avg	41.89	100.80
	Daily T-Win SD	56.58	193.65
	r Lifetime Win Loss	22.44	113.42
	r Lifetime Dollar Dep. Amt	36.38	104.96
Length of Relationship	Lor First Last Play	83.73	66.39
	Lor Initial Deposit Last Play	74.26	64.33
Inter-Play	Inter Playday Avg	8.31	10.07

	Inter Playday SD	56.58	193.65
	Inter First Deposit First Play	9.48	27.18
Bonuses/Rewards	Playbonus Amt	232.56	406.96
	Playbonus Amt Deposit	149.79	303.18
	Bonus Amt Extra	37.28	194.63
	Bonus Amt Discretion	45.49	120.76
	Lifetime Points	8557.36	39608.43
	Lifetime Points Slot	7441.08	34156.10
	Lifetime Points Table	1116.28	16790.35
	Lifetime Points Redm	7610.18	39236.10
	Daily Points Avg	418.92	1008.00
	Daily Points Redm Avg	312.73	961.75
	Lifetime Cash Redm	114.50	820.75
	Daily Cash Redem Avg	4.44	31.84
	r Lifetime Points	151.21	461.41
	r Lifetime Points Redm	110.76	413.65
	r Lifetime Cash Redm	1.74	10.60
	r Lifetime Bonus	16.95	59.37
Demographics	Age	45.65	12.98
	Sex		F: 1443 (55%) M: 1196 (45%)
	Location	81.35	116.20
Target	Churn Status (target var.)		Churner: 1784 (67.6%) Non-churner: 855 (32.4%)