



Consumer Profile Segmentation in Online Lottery Gambling Utilizing Behavioral Tracking Data from the Portuguese National Lottery

Bernardo T. Chagas¹ · J. F. S. Gomes¹ · Mark D. Griffiths²

Accepted: 3 September 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

ABSTRACT

The present study is the first to examine account-based tracking data of Portuguese online lottery players comprising the gambling activity of all active players over a one-year period (N=154,585). The main research goal was the identification of groups or segments of players by their engagement levels (high, neutral, low) and to assess preferences in product category with the use of CHAID (Chi-Square Automatic Interaction Detection) segmentation models, based on expenditure and sociodemographic variables. Findings showed that (i) age was found to be the most influential differentiating variable in player segmentation and had a positive correlation with expenditures and wagers, (ii) gender was the second most influential variable (males represented 78.7% of players), (iii) education the third most influential variable and had a negative correlation with expenditure, and (iv) region was the least relevant variable. The models generated several players segments that engaged in different games. Older males (54-64 years; ≥ 65 years) were the most engaged overall. Younger males (18-34 years) were the least engaged but showed preferences for lotto as did females (35-49 years). Lower educated males and older males (49 years+) with a high school education were the most engaged in instant lottery games. These findings show that Portuguese lottery players can be grouped into several segments with distinct demographic characteristics and corresponding engagement levels. These findings help support more effective marketing segmentation and will help in the targeting of responsible gambling approaches.

Keywords: lottery gambling; internet gambling; online gambling; behavioral tracking; segmentation models; big data.

INTRODUCTION

Up until 2010, self-report surveys dominated the existing lottery literature (LaPlante, Gray, Bosworth, & Shaffer, 2010) but despite increasing use of tracking data to examine gambling behavior, to date, only a few studies have included lottery tracking data in their analysis of gambling behavior. Of the studies that include real lottery playing activity most do not examine lottery exclusively but also include other forms of gambling within the same operator's datasets (e.g., Auer & Griffiths, 2013, 2014, 2017, 2018; Challet-Bouju, et al., 2020; Yuan, 2015). Consequently, this does not allow for an exclusive determination of online lottery players characteristics. The approach taken in previous studies did not focus on segmenting lottery players using real playing data. The present study aimed to identify online lottery players segments using real playing data from the Portuguese national lottery with nationally representative data.

Segmentation is a strategy commonly used in marketing to recognize differences between consumers and their needs or preferences and grouping them according to a set of common characteristics (Dickson & Ginter, 1987; Kotler & Armstrong, 2018). When recognizing the existence of consumers with different preferences, marketers often apply "differentiated" marketing strategies for different market segments (Kotler & Armstrong, 2018). In lottery gambling this can be very useful because lottery operators can better segment their customers and devise the best marketing and responsible gaming strategies to reach different groups of players, with different objectives. The present study uses a behavioral segmentation approach which is a marketing strategy that is considered the best for building market segments (Kotler & Armstrong, 2018), as it can be used to divide buyers into segments based on their usage rate, or responses to a product. It is of particular importance for both researchers and gambling operators because it can differentiate players by engagement levels (e.g., low, medium and high engagement).

Many of the previous studies that used real playing tracking data did not use nationally representative data which does not allow to extrapolate their results to the population nor compare their results with the general population, which was possible in the present study. Additionally, the present study is the first to be conducted on Portuguese lottery players with the use of real player account data. The objectives of the present study stemmed from previous studies' theoretical assumptions and gaps, in which sociodemographic variables were used to study gambling behavior (Gray, Jonsson, LaPlante, & Shaffer, 2015; Kaizeler & Faustino, 2010, 2011, 2012, 2014). Previous studies' limitations in addressing players' common socio-demographic and playing characteristics allowed the present study to go further and understand how online lottery players behavioral activity enables the identification of specific segments with similar sociodemographic characteristics, within the same group but distinguishable from other segments. To identify these groups of players, specific hypotheses were formulated to define online lottery players' segments and meet the present study's main aim. Additionally, the study was able to identify and rank sociodemographic variables by their potential to form specific identifiable segments that can address both marketing and responsible gaming purposes.

LITERATURE REVIEW AND THEORETICAL BACKGROUND

There is a considerable amount of literature on lottery gambling behavior but the body of published research with nationally representative data is considerably narrower. Additionally, there are only a few studies that have used actual playing data and very few studies have included nationally representative

datasets, such as used in the present study. Popular approaches in the study of lottery gambling include socio-demographic and economic analysis of gambling behavior alongside player profiling, but no studies have taken a player segmentation approach by using real playing data, such as the one employed in the present study, which allows a more precise picture of how players group according to sociodemographic variables and game playing preferences. Studies on countries with higher levels of education sell fewer lottery products (Ariyabuddhiphongs, 2011; Kaizeler & Faustino, 2010) and lottery sales are higher in countries in which the percentage of males is higher than that of females for that country's whole population (Clotfelter and Cook, 1989; Kaizeler & Faustino, 2010). A 1% increase in the number of males in the gender ratio of the overall population produces an increase of 13.4% of a country's per-capita lottery sales, which corresponds to about \$270 US annually (Kaizeler & Faustino, 2010), although this gender ratio has not been tested or observed for online lottery gamblers specifically, namely with the use of with real playing data.

Older studies on gambling report that age is negatively related to gambling behavior, so as age increases gambling participation decreases, (Mok & Hraba, 1991). Contradictory findings were reported by (Clotfelter & Cook, 1989) and later by (Kaizeler & Faustino, 2012) as they reported that the pattern of lottery participation by age was an inverted U, with the broad middle range (aged 25-64 years) playing more than the young (18-24 years) and the old (65 years and above) (Ariyabuddhiphongs, 2011) reported that the relationship between age and lottery participation was no longer in an inverted U-shape and reported that all ages play lottery and although the 61+ years age group has the lowest rate of participation in lottery gambling, their mean individual lottery expenditures are the highest. A 2011 study asserted that the shape of this pattern could be somewhat different as Barnes, Welte, Tidwell, & Hoffman, 2011) report that the frequency of gambling on the lottery increased sharply from mid-adolescence to age 18 years (which is the legal age to purchase lottery tickets in most US states) and continued to increase into the thirties when it leveled off and remained high through the sixties and then decreased among those 70 years and older. Despite these findings, Afifi, LaPlante, Taillieu, Dowd, & Shaffer (2014) claimed the role of gender and age in the relationship between gambling and engagement had not been established. The present study analyzed these relationships further to clarify the role of these and other sociodemographic variables in lottery gambling playing and engagement.

Lottery gambling in Portugal

Studies on lottery playing in Portugal are scarce and most research has not been published in peer-reviewed journals (Hubert & Griffiths, 2018). Also, there are no previous studies that focus specifically on online lottery players, especially with the use of real playing data. Additionally, there is only one study on lottery gambling in Portugal that can be considered nationally representative (N=3,850; ages 18–70 years) (Lopes, 2009, 2010). However, the study was based on self-report data, and it has not been published in a peer-reviewed journal and does not analyze specifically online and offline lottery players. In this study, the prevalence rate of lottery gambling was found to be 51.3% over a one-year period in 2007 (Lopes, 2009, 2010). Calado and Griffiths (2016) noted that the prevalence of gambling and problem gambling in Portugal appeared to be similar to other European countries.

In their worldwide studies, Kaizeler, Faustino and Marques (2014) also analyzed lottery sales in Portugal, considering socio-demographic variables for the characterization of the players. They found that for each €1 increase in the income of the habitants of a particular district, there was an increase of 4.4% in the same district's aggregate lottery sales. They also highlighted that richer Portuguese districts spent more on lottery gambling than poorer ones (Kaizeler et al., 2014). Kaizeler et al.'s (2014) five-year analysis of lottery sales in Portugal (2004 to 2008) showed that lottery sales reach their maximum when annual per capita income was €13,208.15, with a corresponding annual lottery sales value per capita of €291.68, and declined thereafter.

Portuguese players' education level has also been found to have a negative correlation with lottery spending. Each 1% increase in a district's secondary school graduation rate leads to an approximate decrease of €162 in lottery sales (Kaizeler et al., 2014). Using a non-representative sample, Brochado, Santos, Brochado and Esperança (2018), also examined the relationship between lottery playing frequency and education among Portuguese lottery games players. They only established a relationship for scratch-cards and did not find a specific relationship between passive/class lottery playing frequency and education. More specifically, Portuguese lower educated lottery players were found to have higher scratch-card gambling frequency than more educated Portuguese lottery players (Brochado et al., 2018). Brochado et al.'s (2018) study focused on motivations to play on *EuroMillions* and passive/class lottery players versus playing scratch-cards. They found that *EuroMillions* players tended to be males with lower incomes who are driven by motivations to buy a car, buy a home, and pay off debt (i.e., *financial motivations*). Among passive/class lottery players, high-frequency gamblers were more likely to be elderly males with lower incomes and who were motivated by increase savings and helping their families (i.e., *safety motivations*). High-frequency scratch-card players presented a different profile and were more likely to be younger females, with lower income and education, but motivated by self-esteem reasons, whereas males who played scratch-cards were more motivated by *financial* or *safety motivations*.

Brochado et al.'s (2018) study did not consider relevant variables such as expenditure, which is one of the most relevant in determining high involvement gambling. Also, they did not consider online gamblers. Additionally, their findings only apply to offline lottery players and their results cannot be extrapolated to all players, nor to the population due the convenience sample used. Only one Portuguese study has examined the effect of age on lottery expenditure (Kaizeler, Faustino, & Marques, 2014). Kaizeler et al. (2014) were only able to establish statistical significance between age and lottery gambling on people aged between 15 and 24 years, where a 1% increase in the population of a particular region led to an annual decrease of approximately €3,700 in the same region. They were not able to establish a statistically significant relationship between age and lottery gambling, for people 25 years and older (a factor which is addressed in the present study).

There is also only one Portuguese study that has compared online and offline gambling habits of Portuguese players but it was not nationally representative and did not focus exclusively on lottery playing (Hubert & Griffiths, 2018). Compared to offline players, online players were found to gamble more days per week but spend less money, consume less alcohol, drugs, and tobacco, and have less suicidal ideation, depression, anxiety, and stress (Hubert & Griffiths, 2018). Most players gambled both online and offline, although they had a preferred channel to play (Hubert & Griffiths, 2018).

To date, studies on lottery gambling among Portuguese players, have all used non-representative samples, with the exception Lopes' (2009, 2010) study which used data collected in 2008 but did not discriminate between online and offline players, and did not use real playing data. Additionally, after 2010, the Portuguese legislation concerning scratch-cards was changed to allow the possibility of increasing the net prize payouts from 48.75% to 50%-70% which is not reflected in Lopes' (2009, 2010) studies.

Gambling studies using behavioral tracking player data

Gambling studies that use account-based behavioral tracking and real playing data to assess gambling behavior were first published in 2007 (LaBrie, LaPlante, Nelson, Schumann, & Shaffer, 2007). Chagas and Gomes (2017) reported that initial studies on lottery tracking data relied on the same database (*Bwin*) and later ones were limited to a few databases and focused on very narrow time frames such as account opening or closing. More recently, additional gambling operators (e.g., *ComeOn Group*, *Íslensk Getspá*, *Kindred*, *Leo Vegas*, *Norsk Tipping*, *Svenska Spel* and *win2day*) have made their databases available for researchers to study gambling behavior (Auer & Griffiths, 2013, 2016, 2017, 2018, 2020; Auer, Hopfgartner, & Griffiths, 2018, 2019a, 2019b, 2020; Fiedler, 2011, 2013; Gainsbury, Sadeque, Mizerski, & Blaszczynski, 2012; Gray, Jonsson, LaPlante, & Shaffer, 2015; Ukhov, Bjurgert, Auer & Griffiths, 2021). These studies cover different forms of gambling and only a few include lottery gambling (e.g., Auer & Griffiths, 2013, 2014, 2016; Auer, Hopfgartner, & Griffiths, 2018; Gray, Jonsson, LaPlante, & Shaffer, 2015) and none of these studies used a player/consumer segmentation method, such as the one we used in the present study, for the study of lottery gambling or even studied the population of a southern European country.

From the studies that include online lottery gambling, some focused on the evaluation of responsible gambling practices. For instance, Auer and Griffiths (2013) found that voluntary limits setting had the highest significant effect on the monetary spending of the most intense players including lottery players. Another study (Auer & Griffiths, 2016) found that personalized behavioral feedback can enable players to gamble more responsibly. Also, gamblers receiving personalized feedback in relation to limit-setting showed significant reductions in the amount of money they gambled (Auer, Hopfgartner & Griffiths, 2018).

OBJECTIVES OF THE PRESENT STUDY

The present study is the first to examine actual online lottery gambling data from Portuguese players. The present study used account-based data made available by the Portuguese national lottery. The dataset used is novel, has never been analyzed before, and was provided exclusively for the purpose of the present research. The dataset is representative of all Portuguese active online lottery players. The study had several research objectives including the identification of lottery players' segments and their engagement levels by analyzing actual online gambling data. The main objective of the study was the identification of different player segments for Portuguese online lottery players and understand the differences between them, including by their engagement levels (high, neutral, low). Another objective was to determine the best approach for the identification of the players' segmentation and behavior.

The study used variables comprising sociodemographic and playing records which were originally in the database and added new variables by using information that was not in the database and was added to or combined with existing variables such as county, district, and NUTS (Nomenclature of Territorial Units

for Statistics), a hierarchical system for dividing the economic territory of the EU (Eurostat, 2019) that was used to better classify and group players by their location/place of residence. The study also assessed differences in gambling from several lottery product categories (i.e., scratch-cards, lotto, Toto, and passive/class lotteries) to understand if there are any distinct sub-groups (segments) of players by gambling engagement in the different forms of lottery games. The cohort was also compared with the general Portuguese population aged 18 years and older to assess for differences in sociodemographic variables distribution between them.

The objectives of the present study led to the following sub-questions: (i) Are there groups of lottery players with different gambling profiles? (ii) Do different types of lottery players engage in different product types and if so, are there significant variations in gambling patterns across the several classes of games? (iii) Are there gender differences across the playing of different lottery products? (iv) Is age a relevant factor in online lottery playing? (v) Does education have a negative correlation with online lottery gambling? (vi) Do players who live in richer regions play lottery games more often and spend more money on lottery games than players who live in poorer areas? (vii) Is there one socio-demographic variable that is more influential in determining online lottery gambling segments than others?

METHODS

Participants

The dataset comprised 218,987 active individual players which accounts for 34.3% of all registered players (Santa Casa da Misericórdia de Lisboa, 2014, 2015). Of the active players, those who did not complete all the data in their player registration form were excluded from further sample analysis. Consequently, the final cohort that underwent data analysis comprised 154,585 valid active players and 14,685,575 data points.

Data

The present study utilized a cross-sectional dataset of a full year aggregate lottery playing activity, from June 30, 2013 to May 31, 2014. The data were anonymized to ensure player identity protection. As data were not normally distributed, missing values were not imputed. The dataset's original variables included date of player registration, age, gender, zip/postal code, education, occupation, amount of money spent (total amount of money spent and amount of money spent per game), wagers¹ (total number of wagers made and number of wagers made per game), number of lottery draws that occurred during the analyzed period (total and per game), number of lottery draws in which each player placed wagers (total and per game), and total number of weeks in which each player placed wagers.

Measures

To identify the players' profiles and to help in segmentation, new variables were created. New variables aggregated products by game design and structural characteristics, features, and play action. Previous research has consistently shown that structural characteristics have an important role in the development

¹We use the term 'wagers' as used by other authors on lottery gambling (Forrest, Gulley, & Simmons, 2004; Gulley & Scott, 1993).

and maintenance of lottery gambling (Griffiths & Wood, 2001; Parke & Griffiths, 2007). The games on the portfolio comprise: *EuroMillions*, the only multi-national game, and *Totoloto* which are lotto games where the first has a 5/50 + 2/12 number picking mechanics and the second is a typical 6/49 lotto game; Toto/sports lottery (*Totobola*) has a 1x2 (first team win, draw, second team win) game mechanic, based on football game outcomes; passive/class lotteries which are lottery games with pre-printed numbers and fixed prize structures, that are also pre-determined, and are not dependent on the number of players and money wagered for the determination of the size of the jackpot (such as in lotto and pari-mutuel games); and scratch-cards which are games based on a card with a section or sections which may be scratched away to reveal a symbol indicating whether a prize has been won in a competition (Ariyabuddhiphongs, 2011; North American Association of State and Provincial Lotteries, 2021). This resulted in five product categories: lotto games (*EuroMillions*; *Totoloto*); lotto games with the inclusion of an add-on game (*EuroMillions*; *Totoloto* and *Joker*; because *Joker* could only be played in association with lotto, it was included in the ‘all lotto’ category analysis); Toto/sports lottery (*Totobola*), passive/class lotteries² (*Lotaria Clássica* and *Lotaria Popular*); and scratch-cards (*Lotaria Instantânea*).

Other new variables included “amount lost” (total amount spent vs. total amount won) and amount spent per wager (mean average). Amount lost variables were created for the total individual gambling activity and were also created for each lottery game individually and for each of the lottery product categories created. These new variables were used to assess playing engagement. Some of the existing variables were converted into new variables by using exogenous information such as the ones created by converting postal codes into NUTS II and III regions.

Data analysis

Considering the large sample size and the known population parameters such as mean, standard deviation, and variance, bilateral Z-tests were used to assess for the final cohort’s representativeness. The cohort proved to be representative of the Portuguese residents who played the national lottery over remote channels. Z-tests were run for total amount spent ($P[Z \leq z]$ bilateral 0.99; confidence interval 0.0007) and for each product group category (lotto [$P[Z \leq z]$ bilateral 0.9918; confidence interval 0.0082); lotto plus add-on game ($P[Z \leq z]$ bilateral 0.9952; confidence interval 0.0048) and scratch-cards ($P[Z \leq z]$ bilateral 0.9606; confidence interval 0.0394). Lotto plus add-on game and scratch-cards represented the bulk of total revenue with 97% of total expenditure (lotto plus add-on game 80.82%; scratch-cards 16.13%). Toto/sports lottery ($P[Z \leq z]$ bilateral 0.2456; confidence interval 0.7544) and passive lotteries ($P[Z \leq z]$ bilateral 0.8937; confidence interval 0.1063) were found to have poor representativeness as Toto represented 0.9% of total revenue and passive lotteries 2.1% of total revenue and cannot be considered representative of all the players that play these games online. Because of this, separate segmentation models for the latter two game categories were not run.

The analysis in this study comprised three main steps. The first step compared the profiles of the individuals in the sample to the general population with estimate information from Statistics Portugal, the Portuguese official agency for economic and demographic data and Pordata. The second step focused on a univariate and bivariate statistical analysis, to characterize the sample. This analysis described the main

characteristics of the players and how they compared to the findings of other studies. The third step comprised a multivariate Chi-square Automatic Interaction Detector (CHAID) analysis (Kass, 1980) which was used to identify different profiles and segments of players. CHAID analysis builds a model, represented as a tree to determine the best merges between variables to explain the outcome in the given dependent variable (Breiman, Friedman, & Stone, 1984; Kass, 1980). The dependent variable in this study was total amount of money spent. An explanatory CHAID model was built to identify player segments and identify the most relevant among them (target market). Separate CHAID analysis was also run to assess amount spent by game type/product category, for the two product categories with the highest expenditures (lotto and scratch-cards), and by amount lost. IBM SPSS 22 was used for the CHAID decision trees.

RESULTS

Characteristics of the online lottery player sample vs. Portuguese population

The study's dataset contained all active Portuguese online lottery players, and the final sample was representative of all active online lottery players. Note that previous studies did not disclose the activity status of players but based their analysis solely on active players, as inactive players do not have any playing records during the specific periods under analysis. Therefore, the first analysis was performed to understand how players' sociodemographic characteristics and profile compared to the general Portuguese adult population (over 18 years of age). The total dataset's active online lottery players accounted for 7.5% of the general adult population and the final sample of this study for 2.6% of the same population. The findings indicated that there was a large discrepancy between the gender distribution of the adult population and online lottery players. Results showed that the differences were statistically significant as there was no overlap and the proportional value was outside the calculated confidence interval (confidence level 96%; $p < 0.05$). Male online lottery players comprised 78.7% of total players ($n=121,601$) whereas in Portugal there are 46.7% adult males (approximately 3.99 million males) (Pordata, 2019b). Female online lottery players comprised only 21.3% of the cohort ($n=32,984$) but are the majority of the adult population (53.3%; approximately 4.56 million females) (Pordata, 2019b).

Another variable used in the comparison was age. Online lottery players and the adult population were split into four sensibly distributed age groups (Table 1). The age group distribution was designed to assess potential skews in lottery players' representativeness when compared to the adult population. Results showed that online lottery players had a different distribution regarding age groups when compared to the Portuguese adult population ($p < 0.05$; see Table 1). Overall, online lottery players' age group representation tended to be younger than the Portuguese adult population given that three-quarters of online lottery players were below 49 years of age (74.4%). Although these are the most represented online lottery players in terms of gender and age, it does not mean that they are the players that had the highest gambling engagement and expenditure.

Table 1*Age groups of the general Portuguese adult population versus active online lottery players*

Age groups	General adult population (≥ 18 years)		Active online lottery players	
18-34	2,005,264	23.5%	44,505	28.8%
35-49	2,342,230	27.4%	70,465	45.6%
50-64	2,097,546	24.5%	31,959	20.7%
≥ 65	2,105,167	24.6%	7,656	5.0%
Total	8,550,207	100.0%	154,585	100.0%

The data were also analyzed for geographic distribution utilizing NUTS III. In general, the geographic location and distribution of players was similar to the geographic concentration of the general Portuguese adult population but there were differences in two main regions. For online lottery players, the Lisbon Metropolitan Area was the most represented while among the general population it is second, after the North region (Table 2).

Table 2*Geographic distribution of the general Portuguese adult population versus active online lottery players.*

Geographic distribution - NUTS II	General adult population (≥ 18 years)		Active online lottery players	
Lisbon Metropolitan Area	2,277,826	26.6%	53,319	34.5%
North	2,992,782	35.0%	46,819	30.3%
Centre	1,898,503	22.2%	31,167	20.2%
Alentejo	615,121	7.2%	8,847	5.7%
Algarve	360,417	4.2%	6,311	4.1%
Mainland	8,144,649	95.3%	146,463	94.7%
Madeira	209,484	2.5%	4,799	3.1%
Azores	195,074	2.3%	3,323	2.1%
Portugal	8,549,207	100.0%	154,585	100.0%

Considerable differences were also found in relation to education (Table 3). Online lottery players had on average a much higher educational level than the general population ($p < 0.05$). Most players had a higher education (51.7%) compared to 20.6% of the general Portuguese adult population and there were no players without any formal education.

Table 3*Education distribution of the general Portuguese adult population versus active online lottery players*

Education	General adult population (≥ 18 years)		Active online lottery players	
Without formal education	167,350	2.6%	0	0.0%
Primary education (4th grade)	1,081,560	16.7%	2,920	1.9%
Primary education (6th grade)	891,880	13.7%	9,745	6.3%
Secondary education	3,015,590	46.4%	61,925	40.1%
Higher education	1,337,700	20.6%	79,995	51.7%
Total	6,494,080	100.0%	154,585	100.0%

Gambling participation

In more specific analysis, the final sample was found to be representative of all active online lottery players on several variables ($p<0.01$), including total expenditure, total amounts lost, and total number and amount of prize money won (Table 4). It was also found that the 25% players with the highest average annual expenditures (Q3=€324), represented 79.7% of total expenditures and showed higher total expenditures than the lower two quartiles combined (Q1 expenditure=€29.40, Median=109.00; IQR=294.6). The most popular games among online lottery players were lotto (including add-on games; 98.05%, n=151,758) and scratch-cards (34.66%, n=53,580). The least played games were passive lotteries (10.34%, n=15,996) and Toto/sports lottery (9.02%, n=13,959).

Table 4

Total online lottery gambling participation (expenditure, prizes and losses in Euros)

Total Sample	n	Mean	Mode	Median	SD of total expenditure	VAR	SD mean expenditure	Q1	Q3
Expenditure	154,585	337.04	10.00	109.00	1,103.08	1,216,786.95	2.81	29.40	324.00
N° prizes won	154,585	23.12	0.00	6.00	143.78	20672.20	0.37	1.00	19.00
Amount won	154,585	116.01	0.00	28.28	642.04	412209.78	1.63	4.32	93.32
Amount lost	154585	-221.13	-10.00	-74.99	724.37	524711.52	1.84	-223.67	-20.00
Amount per wager	154585	3.56	2.00	2.52	3.64	13.24	0.01	1.90	4.21

Sample analysis by gender, age group, education and region

Male online lottery players were found to spend more money on average than female online lottery players (Table 5; $p<0.01$). This trend was also confirmed when assessing for product categories' mean differences, including on lotto plus add-on game ([males: mean=€311.53; median=€109.60; SD=€826.09]; [females mean=€152.11; median=€48.90; SD=€591.90]; [$p<0.01$]); The only game category where no statistically significant differences in mean expenditures were found between males and females was instant lottery games ([males: mean=€157.61; median=€20.00; SD=€1,108.97]; [females mean=€154.36; median=€17.50; SD=€1,073.00]; [$p<0.05$]). During the one-year period analyzed, males placed more and larger wagers. Males' average amount per wager was also higher than that of females ([males: mean=€3.74; median=€2.69; SD=€3.83]; [females: mean=€2.89; median=€2.09; SD=€2.70]; $p<0.01$). Males also placed more wagers than females during the one-year period ([males: mean=117.03; median: 44; SD=546.24]; [females: mean=93.18; median: 26.00; SD=476.40], $p<0.01$).

In assessing age and online lottery gambling, each age group was analyzed individually and compared with all other age groups. Statistically significant differences were found in expenditures (Table 5) and wagers placed across all age groups with a steady increase from the youngest to the oldest age groups ($p<0.01$). Players in the 18-34-year age group placed on average 59.39 wagers per year (median=20.00; SD=143.64), with an average amount per wager of €3.06 (median=€2.18; SD=€3.00). Players in the 35-49-year age group placed on average 103.86 wagers per year (median=38.00; SD=687.14), with an average amount per wager of €3.62 (median=€2.59; SD=€3.56). Players in the 50-64-year age group placed on

average 178.35 wagers per year (median=81.00; SD=470.27) with an average amount per wager of €3.95 (median=€2.87; SD=€4,17). The oldest players (65 years and older) had an annual average of 214.51 wagers (median=99.00; SD=533.43), with an average amount per wager of €4.35 (median=€3.31; SD=€4.81).

The findings also indicated that annual lottery expenditures decreased as education levels increased (Table 5) and differences in expenditures between all the age groups were found to be statistically significant ($p<0.01$). The cohort comprised almost entirely individuals with a high school or higher education (141,920 players: 91.8%).

Table 5

Lottery expenditure by age group, education, and region

	n	Mean	Mode	Median	SD of total expenditure	VAR	SD mean expenditure	Q1	Q3
Age groups									
18-34-years	44,505	151.89	10.00	50.00	402.74	162,197.70	1.91	16.00	154.00
35-49-years	70,465	300.15	10.00	109.00	1,119.11	1,252,401.05	4.22	31.00	304.00
50-64-year	31,959	580.41	10.00	237.00	1,472.23	2,167,450.47	8.24	74.00	595.50
≥ 65 years	7,656	736.95	4.00	326.30	1,559.70	2,432,660.02	17.83	118.80	761.58
Gender									
Male	121,601	371.10	10.00	127.00	1,143.05	1,306,565.56	3.28	34.00	364.40
Female	32,984	211.45	4.00	62.00	930.47	865,782.32	5.12	18.00	200.00
Education									
Fourth grade	2,920	532.90	10.00	193.95	1,348.36	1,818,065.60	24.95	55.88	558.13
Sixth grade	9,745	438.72	10.00	140.00	2,229.05	4,968,680.58	22.58	37.00	411.50
High school	61,925	341.99	10.00	112.00	1,032.95	1,066,981.20	4.15	30.00	338.90
Higher education	79,995	313.67	10.00	101.80	922.29	850,617.93	3.26	28.00	298.55
Region									
Alentejo	8,847	343.51	10.00	116.50	827.83	685,302.80	8.80	30.00	346.95
Algarve	6,311	347.40	10.00	112.50	1,090.45	1,189,073.47	13.73	30.00	330.25
Center	31,167	314.47	10.00	104.00	934.13	872,603.24	5.29	28.00	312.00
Lisbon	53,319	362.61	10.00	116.00	1,079.22	1,164,720.56	4.67	30.00	343.00
North	46,819	317.83	10.00	104.00	1,035.27	1,071,779.44	4.78	28.00	311.00
Madeira	4,799	389.92	10.00	114.60	2,623.19	6,881,102.97	37.87	31.50	325.55
Azores	3,323	295.79	10.00	95.40	766.59	587,659.48	13.30	24.00	276.95

Mean expenditures difference for the regions were more diffused. Analysis showed that the expenditure mean differences were not statistically significant in the five region combinations ([Center vs. North; $p<0.05$]; [Center vs. Azores; $p<0.01$]; [Lisbon vs. Madeira; $p<0.05$]; [North vs. Madeira; $p<0.05$]; [North vs. Azores; $p<0.05$]). For the other five region combinations, mean expenditures differences were found to be statistically significant ([Center vs. Lisbon; $p<0.01$]; [Center vs. Madeira; $p<0.05$]; [Lisbon vs. North; $p<0.01$]; [Lisbon vs. Azores; $p<0.01$]; [Madeira vs. Azores; $p<0.01$]).

Chi-Square Automatic Interaction Detection (CHAID) Segmentation Model

The present study utilized a decision tree technique, based on a chi-square test algorithm to identify the most relevant interacting variables to build a classification model of online lottery players. CHAID analysis was chosen because it does not require the data to be normally distributed and nor does it require the variables to be standardized which was a characteristic of the online lottery gambling activity in this dataset. CHAID was used to identify the relationship between variables and helped understand how variables merged and explained the outcomes on a particular dependent variable. The development of the models considered the total monetary expenditure as a dependent variable to understand which independent variables accounted for the development of single or several groups or segments of online lottery players. The total individual expenditure CHAID decision tree generated 22 nodes at four levels (Figure 1). Each node was considered a different player segment. From these 22 nodes, the model generated 14 final nodes or customer segments (Figure 1). The final nodes corresponded to individual player segments. The model detailed a hierarchical structure of the variables in order, starting with the one (age) that most reduced the variance in the division of the segments (i.e., the variable that maximizes the residual variance). Consequently, in the first level, age was found to be the most influential variable, as it was the variable that made each segment as homogeneous as possible, and had a positive correlation with amount of money spent and number of wagers. Gender was found to be the second most influential variable. Education was the third most influential and had a negative correlation with lottery expenditure. Place of residence was the variable with the least influence in the model.

Figure 1

Online lottery players' total expenditure segmentation CHAID decision tree

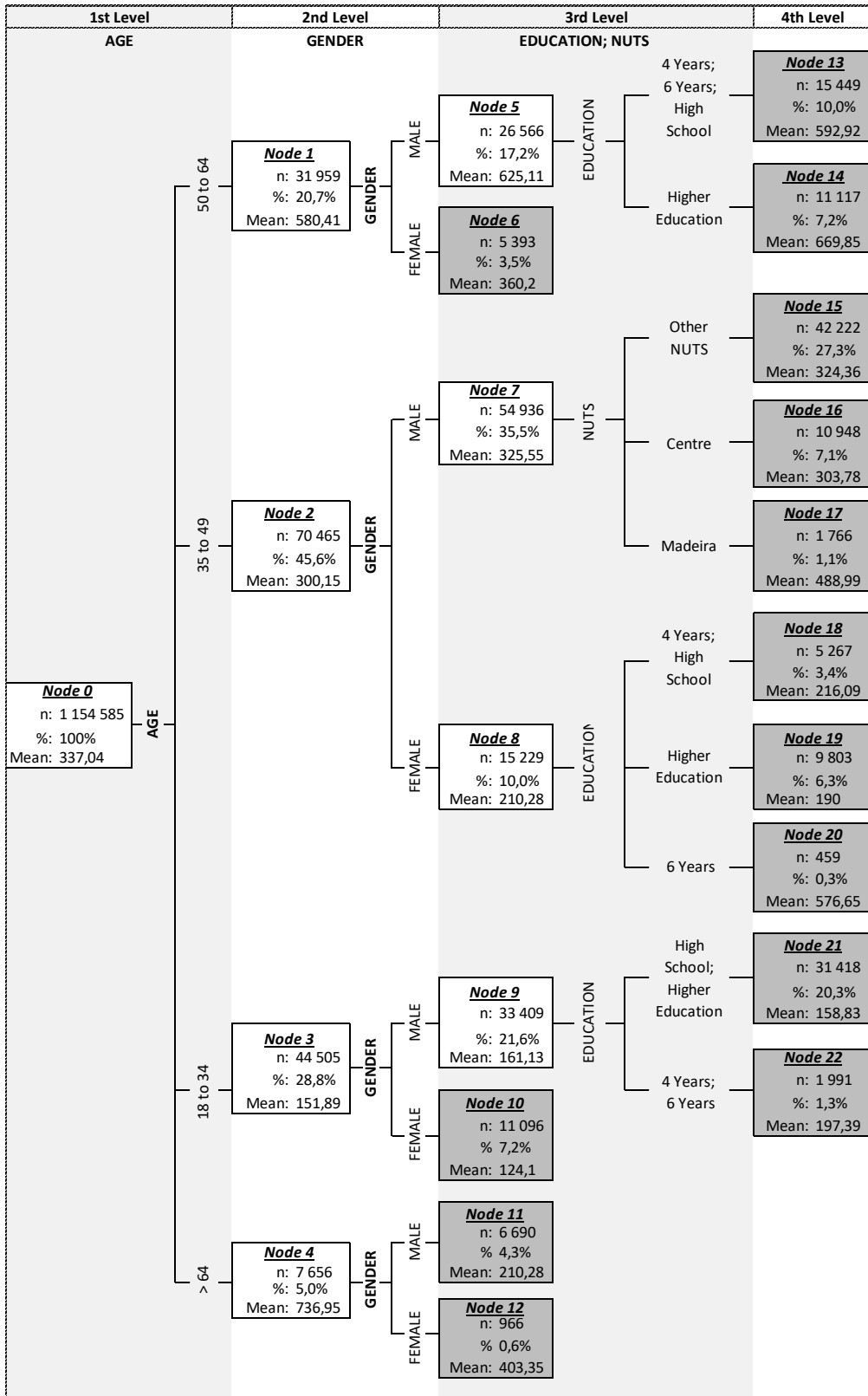


Table 6*High, neutral, and low engagement player segments*

Nodes/Segments		Characterization Variables				Players			Expenditures	
Segment	1	2			3	4	5	6=4x5	7=6/total 6	8=7-4
Engagement	Node n°	Age Group	Gender	Education / Location	N	Segment Weight	Mean	Mean	Player Segment Expenditure %	Player Segment Node Relevance
High	13	50-64	Male	4th grade; 6th grade; High School	15 449,00	10,0%	592,92	9 159 936,30	17,6%	7,6%
	14	50-64	Male	Higher Education	11 117,00	7,2%	669,85	7 446 715,20	14,3%	7,1%
	11	≥ 65	Male	-	6 690,00	4,3%	785,12	5 252 464,40	10,1%	5,8%
	17	35-49	Male	Madeira	1 766,00	1,1%	488,99	863 556,90	1,7%	0,6%
Neutral	6	50-64	Female	-	5 393,00	3,5%	360,20	1 942 559,60	3,7%	0,2%
	20	35-49	Female	6th grade	459,00	0,3%	576,65	264 681,30	0,5%	0,2%
	12	≥ 65	Female	-	966,00	0,6%	403,35	389 632,40	0,7%	0,1%
	22	18-34	Male	4th grade; 6th grade	1 991,00	1,3%	197,39	393 008,30	0,8%	-0,5%
	16	35-49	Male	Centre	10 948,00	7,1%	303,78	3 325 781,20	6,4%	-0,7%
Low	15	35-49	Male	Lisbon; North; Alentejo; Azores	42 222,00	27,4%	324,36	13 695 149,50	26,3%	-1,1%
	18	35-49	Female	4th grade; High School	5 267,00	3,4%	216,09	1 138 152,80	2,2%	-1,2%
	19	35-49	Female	Higher Education	9 803,00	6,3%	190,00	1 862 597,40	3,6%	-2,7%
	10	18-34	Female	-	11 096,00	7,2%	124,10	1 376 962,10	2,6%	-4,6%
	21	18-34	Male	High School; Higher Education	31 418,00	20,3%	158,83	4 990 036,90	9,6%	-10,7%
TOTAL					154 585	100,0%	337,04	52 101 234,30	100,0%	

The CHAID model provided a hierarchical structure that regards the importance of the segments in terms of player engagement. Player segments were divided by node relevance into three lottery engagement categories: high engagement, neutral engagement, and low engagement. The most relevant nodes are those with the highest positive difference between segment weight and player segment expenditure percentage which represents a higher engagement per player. Segments with high engagement are those in which mean expenditures were >1% higher than overall dataset mean expenditure. Segments where parameters fall within 1% more or less than overall dataset mean expenditure were considered neutral engagement segments. Segments in which the mean expenditure was more than 1% lower than overall dataset mean expenditure were considered low engagement (Table 6). The data analysis demonstrated the general level of segment engagement based on individual player mean expenditures. The most expressive nodes for profile characterization were 13, 14, and 11. Node 13 comprised male players aged 50-64 years with one of three educational levels: 4th grade, 6th grade, or high school. The second most relevant node (14) comprised males from the same age group but with a higher education. The third most relevant node (11) comprised male players aged above 65 years or above. These findings tell us that although age is positively correlated with lottery gambling expenditures, the most engaged are not necessarily the elderly (65+) but the 50-64 years age group. The least engaged players were found to be from the youngest age group (18-34 years) with high school or higher education. For the two most engaged groups and for the least engaged group, place of residence was not found to be a relevant variable. To confirm the results of the total expenditures CHAID model, another CHAID model was built with the total amount of money lost variable, which confirmed the results of the first model. The same method was used to identify the most relevant segments for each of the two most participated product categories (lotto and instant lottery games account for more than 90% of total expenditures). Lotto's two most relevant segments comprised younger males (18-34 years) with either a very low or a very high education (4th grade or higher education) or with a mid-level education (6th grade or high school) (segments node relevance of 6.20% and 5.58%).

Another relevant segment in this product category were females aged 35-49 years (segment node relevance of 4.56%). The least engaged players were males aged 50-64 years, independent of their education (segments node relevance of -7.86% and -8.40%). For instant lottery games, the most engaged segments comprised lower educated males (4th and 6th grade) from the following regions: Lisbon; North; Center; Alentejo; Azores, and Algarve (segment node relevance 4.98%); males older than 49 years, with a high school education (segment node relevance 3.96%) and higher educated players older than 49 years (segment node relevance 3.56%). The least engaged instant lottery players were those with higher educated younger males (18-34 years) (segment node relevance -6.69%) and higher educated players aged 35-49 years from the following regions: Lisboa; Madeira; Azores) (segment node relevance -5.23%).

DISCUSSION

In the present study, sociodemographic variables in the dataset, complemented by new variables, were used to build online lottery player segmentation models. Consequently, several different player segments were identified. Consequently, this approach provided a better understanding of which (and how) sociodemographic characteristics of players may be used to create groups of players that present the same

lottery gambling preferences and engagement levels. Moreover, results of the present study are generalizable to the Portuguese population. Generalizability to other populations should be cautious as other countries will have different economic characteristics (GDP, GINI Index, etc.) that may impact the generalizability of the results of the present study, but the same methods used in this study could be applied to other populations in order to establish comparability.

The use of CHAID models is a novel approach in the field of gambling studies which enables a better understanding of how players group together or differentiate from one another to create specific player segments, which are important to understand the relation between player sociodemographic characteristics and its related player activity and engagement. The findings of the present study provide evidence of a hierarchization of the variables (age being the most important in the hierarchy) because it is the one that most reduces the variance in the segmentation process (i.e., it is the variable that maximizes the residual variation). Age was the biggest differentiating factor and the variable that most maximized the residual variation (i.e., made each segment as homogeneous as possible). Hierarchization of the CHAID nodes concerns the definition in terms of variable segment determination relevance. The second most influential variable was gender, education, and place of residence. Results showed several specific segments, which differed in player engagement that was assessed by players' expenditure and number of wagers by product category, and in total. These findings enable both researchers and practitioners to better understand how to better address different groups of players according to these sociodemographic variables. The results have implications for the developments of specific marketing practices or advertising campaigns that can be at the same time more effective while promoting better targeted responsible gambling practices.

Sociodemographic variables were also used to compare the distribution of the Portuguese adult population to online lottery players in the dataset. The findings indicated a large discrepancy between the gender distribution in the Portuguese adult population and Portuguese online lottery players. These findings are of significant relevance for the identification of specific player segments and profiles, including to help build a profile of potential problematic players, which is useful both for practitioner and researcher alike in the identification and prevention of problem gambling and developing better and more efficient and responsible marketing practices. Male online lottery players comprised 78.7% of total players whereas in Portugal there are 46.7% adult male citizens, which shows a large skew of the online lottery player population regarding the general adult population. Female online lottery players comprised 21.3% in the present study but are the majority in the general Portuguese adult population (53.3%). The present study demonstrates that the online lottery player profile distribution differs significantly from the adult population and that males play online lottery games more than females, advancing knowledge that previous studies were not able to establish among the Portuguese population (Brochado et al., 2018; Hubert, 2014; Kaizeler et al., 2014; Lopes, 2009, 2010). Some of the findings from our study are line with Gray et al. (2015), such as males being more involved in gambling overall, whereas other findings in the present study contradicted some of the findings of their study. Gray et al. (2015) found a greater percentage of females played traditional lottery games whereas males were more likely to engage in soccer betting. The present study had different findings because males were much more represented and were more engaged with online lottery gambling overall, which sheds new insight concerning online lottery gambling.

The mean age of online lottery players in Portugal tended to be younger than the national adult population. Millennials, who are considered one of the most important demographic consumer groups (Eastman, Iyer, Liao-Troth, Williams, & Griffin, 2014; McCasland, 2005; Moreno, Lafuente, Ávila-Carreón, & Moreno, 2017; Ordun, 2015; Smith, 2011, 2012), were not very engaged in online lottery playing in the present study. This may imply that young adults have little interest in online lottery gambling, which is an interesting finding given the many studies on underage gambling, including lotteries (Ariyabuddhiphongs, 2011; Derevensky & Gupta, 2001; Felsher, Derevensky, & Gupta, 2004a, 2004b; Wood & Griffiths, 1998, 2004) a few references needed). The present study found a positive association between age and lottery gambling expenditure (i.e., the older an online lottery player was, the higher the expenditure). These findings do not concur with previous studies findings reporting an inverted U shape distribution regarding age and lottery gambling expenditures (Barnes, Welte, Tidwell, & Hoffman, 2011; Clotfelter & Cook, 1989; Kaizeler & Faustino, 2012) and supports Ariyabuddhiphongs' (2011) findings that this inverted U-shaped relationship is no longer present and advances knowledge by confirming that this also applies specifically to online lottery players and not just to offline lottery players. Findings from the present study challenge other previous findings, including age being negatively correlated with gambling behavior (Mok & Hrabá, Age and gambling behavior: A declining and shifting pattern of participation, 1991). Previous studies examining Portuguese online lottery players did not address or failed to establish a relationship between age and lottery playing (Brochado et al, 2018.; Hubert, 2014; Kaizeler et al., 2014; Lopes, 2009, 2010), unlike the present study.

The same trend was observed for the number of wagers placed. This confirms recent findings on the positive association between age and gambling engagement (Auer, Hopfgartner, & Griffiths, 2018). The present study also found that age-related lottery expenditures differed between product categories. Older males (≥ 54 years) tended to be the most engaged players overall. Younger male players (18-34 years old) and females aged 35-49 years tended to be more engaged in lotto games whereas the most relevant socio-demographic variables for instant lottery games were being male over 49 years old, without a higher education, from one of the following regions: Lisbon; North; Center; Alentejo; Azores or Algarve. It was found that the 25% of players with the highest average expenditures, accounted for about 80% of the total amount wagered. Although these results are in line with other studies (e.g., Garibaldi et al., 2015; Tom et al., 2014), further research on these players may be of interest, to assess their specific profiles and gambling habits.

The player profile characterization in the study also showed that the most represented age group was 35-49 years although the largest difference found regarding the player profile and the Portuguese adult population was among those aged 65 years and older (5% in the present sample compared to 24.6% in the general population). Clotfelter and Cook (1990) reported that individuals between 25 and 64 years have a greater propensity to play on lottery products. Although the present study found older players to have higher gambling engagement, older age groups comprised few players. This contrasts with other online gambling activities, especially sports betting and in-play betting, where younger players tend to play more games and be more involved (Gray et al., 2015).

In the present study, players aged 18 to 34 years old represented 28.8% of players, although most of the players in this category were between 25 and 34 years (21.3%). These younger players had the lowest

engagement of all online lottery players. Although the two youngest age groups were the most represented (18-34 years, 35-49 years) – most likely due to the type of sales channel (internet/mobile) – they are not the most active or most engaged. The oldest group (≥ 65 years; 5%) and the youngest subgroup (16 to 24 years; 7.5%) were the least represented. There may be different reasons for this such as disposable income, game design, and play action but further research is needed to confirm such speculations. For the younger age group, some of these games may simply not be attractive and older players may prefer more traditional (offline) venues to engage in lottery playing.

Generally, gambling activities tend to increase with higher educational levels, but previous studies found that lottery gambling tends to show the opposite (Brown & Kaldenberg, 1992; Clotfelter & Cook, 1989; Clotfelter et al., 1999; Forrest & Gulley, 2009; Rogers & Webley, 2001). The present study confirmed this negative association among Portuguese online lottery players because they tended to decrease lottery gambling expenditure and engagement as their education increased. This also confirms the findings from previous studies examining offline lottery players concerning the relationship between education and lottery expenditures (Albers & Hubl, 1997; Clotfelter & Cook, 1989; Griffiths & Wood, 2001), including for Portuguese offline lottery players (Brochado et al., 2018; Kaizeler & Faustino, 2008). The only exception found in the present study was among older scratch-card players where no negative association with expenditure was observed relating to an increase in education level. Although an increase in level of education was associated with a decrease in expenditure, higher educated players were found to be the most represented (52%) in the dataset (50% of males and 61% of female players had a higher education). Playing lottery games online appears to appeal to ‘tech savvy’ educated players.

When examining place of residence, the present study also found differences between the Portuguese adult population’s geographic distribution and online lottery players’ place of residence distribution. This was observed in five of the seven areas, although it was more evident in the two most populated areas. The major difference was found for the Lisbon area, where 34.5% of lottery players reside which is 7.9% higher than the adult population distribution. The second largest difference was found in the north region where the distribution of players was 4.7% higher among online lottery players. Explanation for this may be due to the higher per capita earnings of individuals living in the Lisbon area (Pordata, 2019a, 2020).

As in any research, the present study has some limitations. Typical limitations found in behavioral tracking studies are also applicable to the present study. Data used were from only one website and players might be gambling on various websites and/or gambling in offline land-based venues (Adami, Benini, Boschetti & Canini, 2013; Auer & Griffiths, 2014; Dragicevic, Tsogas, Kudic, 2011, Dragicevic, Percy, Kudic & Parke, 2013; Fiedler, 2011; Griffiths, 2012; LaBrie et al., 2007; Shaffer et al., 2010; Xuan & Shaffer, 2009). Consequently, the present study can only be considered representative of Portuguese active online lottery players and not of offline lottery players or of all other online gambling activities. Moreover, in some cases, users might share their accounts and access passwords with other individuals, which means that an account might be used by more than one person (Auer & Griffiths, 2014; Fiedler, 2011; Griffiths, 2012; Shaffer et al., 2010). Another drawback of using tracking data to study consumer behavior is that actual gambling data does not provide answers on why individuals behave as they do (Griffiths, 2012). Self-report methods are best for collecting such data. This is a common disadvantage of using secondary

data because it is limited in the ability to provide causal relationship explanations. Considering that the data analyzed in the present study was cross-sectional, longitudinal analysis was not possible. Furthermore, a time variable was missing from the present database and there was no direct or indirect way to measure the time spent gambling (although most lottery games tend to be discontinuous forms of gambling, so the time element is not necessarily important).

CONCLUSIONS AND FURTHER RESEARCH

This study demonstrates that the socio-demographic profile distribution of online lottery players differs significantly from the Portuguese adult population and that males were found to be much more represented than females. The mean age of online lottery players in Portugal tends to be lower than in the adult population. The present study found that millennials are not very engaged in online lottery playing. One of the questions that might be addressed in further research is the reason for this. Findings from this study shed new insight into age and lottery gambling and present new empirical evidence on age being negatively associated with online lottery gambling behavior in Portugal. The present study also found considerable differences regarding education. Online lottery players tend to have an average level of education that is much higher than the adult population, although as the present study confirmed, as education level increased, overall lottery gambling expenditures tended to decrease.

The CHAID segmentation models determined a four-level hierarchy of the demographic variables ranked by relevance and engagement. The most influential variable was found to be age, followed by gender, education, and place of residence. Males were found to play online lottery games more and be more engaged in gambling overall, with higher number of wagers and higher expenditures than females. It was found that as age increased, so did lottery expenditures and the number of wagers. Another finding showed that age-related lottery expenditures differed between product categories. It was also found that the 25% of players with the highest average expenditures, accounted for about 80% of the total amount wagered.

In further research, studies focusing on drawing a wider profile of gamblers would be beneficial and add to the existing knowledge in this field namely by including the possibility of identifying the profile of the players across multiple gambling activities, including from different providers. This would help in understanding differences in players and behaviors across multiple platforms. This has obvious challenges, namely ensuring player anonymity and possible conflicts of interest as well as business protection across the several operators. Such studies can be performed with tracking data only or combined with self-report data. Studies combining playing tracking data with self-reported gambling data enabling a wider knowledge of online lottery players which could help further explain motivations that can provide insight into the way players may be segmented.

Since longitudinal data were not available for the present study, further studies on Portuguese lottery players, utilizing longitudinal tracking data, would also help significantly improve behavioral analysis. The Portuguese national lottery added a new odds sports betting game to its portfolio in 2016 (i.e., *Placard*). This new game accounted for 18.9% of overall sales in 2019 (Santa Casa da Misericórdia de Lisboa, 2020). Therefore, it would be useful to assess if these sales represent a shift in lottery players from pre-2016 existing games to this new game or whether these are all entirely new players. It would also be useful to

assess if the current COVID-19 pandemic has driven offline players to start playing online as most points of sale were closed during the lockdowns.

References

- Adami, N., Benini, S., Boschetti, A., & Canini, L. (2013). Markers of unsustainable gambling for early detection of at-risk online gamblers. *International Gambling Studies*, *13*(2), 188–204.
- Albers, N., & Hubl, L. (1997). Gambling market and individual patterns of gambling in Germany. *Journal of Gambling Studies*, *13*, 125–144.
- Afifi, T.O., LaPlante, D.A., Taillieu, T.L. & Shaffer, H. J. (2014). Gambling involvement: considering frequency of play and the moderating effects of gender and age. *International Journal of Mental Health and Addiction*, *12*, 283–294.
- Ariyabuddhipongs, V. (2011). Lottery gambling: A review. *Journal of Gambling Studies*, *27*(1), 15–33.
- Auer, M., & Griffiths, M. D. (2013). Voluntary limit setting and player choice in most intense online gamblers: An empirical study of gambling behaviour. *Journal of Gambling Studies*, *29*, 647–660.
- Auer, M., & Griffiths, M. D. (2014). An empirical investigation of theoretical loss and gambling intensity. *Journal of Gambling Studies*, *30*, 879–887.
- Auer, M., & Griffiths, M. D. (2016). Personalized behavioral feedback for online gamblers: A real world empirical study. *Frontiers in Psychology*, *7*(1875), 1–13.
- Auer, M., & Griffiths, M. D. (2017). Self-reported losses versus actual losses in online gambling: An empirical study. *Journal of Gambling Studies*, *33*(3), 795–806.
- Auer, M., & Griffiths, M. D. (2018). Cognitive dissonance, personalized feedback, and online gambling behavior: An exploratory study using objective tracking data and subjective self-report. *International Journal of Mental Health and Addiction*, *16*, 631–641.
- Auer, M., & Griffiths, M. D. (2020). The use of personalized messages on wagering behavior of Swedish online gamblers: An empirical study. *Computers in Human Behavior*, *110*, 106402.
- Auer, M., Hopfgartner, N., & Griffiths, M. D. (2018). The effect of loss-limit reminders on gambling behavior: A real-world study of Norwegian gamblers. *Journal Behavioural Addiction*, *7*(4), 1056–1067.
- Auer, M., Hopfgartner, N., & Griffiths, M. D. (2019a). An empirical study of the effect of voluntary limit-setting on gamblers' loyalty using behavioural tracking data. *International Journal of Mental Health and Addiction, Advance Online Publication*, 1–12. doi:<https://doi.org/10.1007/s11469-019-00084-3>
- Auer, M., Hopfgartner, N., & Griffiths, M. D. (2019b). The effects of a mandatory play break on subsequent gambling among Norwegian video lottery terminal players. *Journal of Behavioral Addictions*, *8*(3), 522–529.
- Auer, M., Hopfgartner, N., & Griffiths, M. D. (2020). The effects of voluntary deposit limit-setting on long-term online gambling expenditure. *Cyberpsychology, Behavior, And Social Networking*, *23*(2), 113–118.
- Baker, R., Forrest, D., & Pérez, L. (2016). The compatriot win effect on national sales of a multicountry lottery. *Applied Statistics*, *65*(4), 603–618.
- Barnes, G. M., Welte, J. W., Tidwell, M.-C. O., & Hoffman, J. H. (2011). Gambling on the lottery: Sociodemographic correlates across the lifespan. *Journal of Gambling Studies*, *27*(4), 575–586.
- Baumöhl, E., & Výrostováb, E. (2017). Do people gamble more in good times? Evidence from 27 European countries. *Applied Economics Letters*, *24*(18), 1311–1314.

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, Florida: Routledge.
- Brochado, A., Santos, M., Brochado, F. O., & Esperança, J. (2018). Gambling behavior: Instant versus traditional lotteries. *Journal of Business Research*, 88, 560 - 567.
- Brown, D. J., & Kaldenberg, D. O. (1992). Socio-economic status and playing the lotteries. *Sociology & Social Research*, 76, 161–167.
- Byrnes, J. P., Miller, D. C., & William, S. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 25(3), 367-383.
- Calado, F., & Griffiths, M. D. (2016). Problem gambling worldwide: An update and systematic review of empirical research (2000–2015). *Journal of Behavioral Addictions*, 5(4), 592–613.
- Chagas, B. T., & Gomes, J. F. (2017). Internet gambling: A critical review of behavioural tracking research. *Journal of Gambling Issues*, 36, 1-27.
- Challet-Bouju, B., Hardouin, J., Thiabaud, E., Saillard, A., Donnio, Y., Grall-Bronnec, M., & Perrot, B. (2020). Early gambling behaviors of newly registered online lottery gamblers: Longitudinal analysis using gambling tracking data. *Journal of Medical Internet Research*, 22(8), e17675
- Charnessa, G., & Gneezyb, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1), 50-58.
- Clotfelter, C. T., & Cook, P. J. (1989). *The demand for lottery products: Working Paper No. 2928*. Cambridge, MA: National Bureau of Economic Research.
- Clotfelter, C. T., & Cook, P. J. (1990). On the economics of state lotteries. *American Economic Review*, 83(3), 634-643.
- Clotfelter, C. T., Cook, P. J., Edell, J. A., & Moore, M. (1999). *State lotteries at the turn of the century: Report to the National Gambling Impact Study Commission*. Duke University. Retrieved October 10, 2020, from <http://govinfo.library.unt.edu/ngisc/reports/lotfinal.pdf>
- Derevensky, J. L. & Gupta, R. (2001). *Lottery Ticket Purchases by Adolescents: A Qualitative and Quantitative Examination*. Report to the Ministry of Health and Long-Term Care - Alberta Gambling Research Institute. <https://prism.ucalgary.ca/handle/1880/41343>
- Dickson, P., & Ginter, J. (1987). Market segmentation, product differentiation, and marketing strategy. *Journal of Marketing*, 51(2), 1-10.
- Dragicevic, S., Tsogas, S. B., & Kudic, A. (2011). Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. *International Gambling Studies*, 11, 377–391.
- Dragicevic, S., Percy, C., Kudic, A. & Parke, J. (2015). A descriptive analysis of demographic and behavioral data from internet gamblers and those who self-exclude from online gambling Platforms. *Journal of Gambling Studies*, 31, 105–132.
- Eastman, J. K., Iyer, R., Liao-Troth, S., Williams, D. F., & Griffin, M. (2014). The role of involvement on millennials' mobile technology behaviors: The moderating impact of status consumption, innovation, and opinion leadership. *Journal of Marketing Theory and Practice*, 22(4), 455-470.
- Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, 5(3), 159–163.
- Eurostat (2019, October 21). *Eurostat NUTS background*. Retrieved from Eurostat: <https://ec.europa.eu/eurostat/web/nuts/background>
- Felsher, J.R., Derevensky, J.L. & Gupta, R. (2004a). Lottery participation by youth with gambling problems: are lottery tickets a gateway to other gambling venues? *International Gambling Studies*, 4(2), 109-125.

- Felsher, J.R., Derevensky, J.L. & Gupta, R. Lottery Playing (2004b). Amongst Youth: Implications for Prevention and Social Policy. *Journal of Gambling Studies*, 20, 127–153.
- Fiedler, I. (2011). The gambling habits of online poker players. *Journal of Gambling Business and Economics*, 6, 1–24.
- Fiedler, I. (2013). Gamblers' habits: Empirical evidence on the behavior of regulars newcomers and dropouts. *Journal of Gambling Studies*, 29, 289–309.
- Forrest, D., & Gulley, O. D. (2009). Participation and level of play in the UK National Lottery and correlation with spending on other modes of gambling. *International Gambling Studies*, 9, 65–178.
- Gainsbury, S., Sadeque, S., Mizerski, R., & Blaszczynski, A. (2012). Wagering in Australia: A retrospective behavioural analysis of betting patterns based on player account data. *Journal of Gambling Business and Economics*, 6(2), 50-68.
- Garibaldi, S., Frisoli, K., Ke, L., & Lim, M. (2015). Lottery spending: A non-parametric analysis. *PLoS One*, 2: e0115730, 10. doi:https://doi.org/10.1371/journal.pone.0115730
- Gray, H. M., Jonsson, G. K., LaPlante, D. A., & Shaffer, H. J. (2015). Expanding the study of internet gambling behavior: Trends within the Icelandic lottery and sportsbetting platform. *Journal of Gambling Studies*, 31(2), 483-499.
- Griffiths, M. D. (2012). Internet gambling, player protection and social responsibility. In R. Williams, R. Wood, & J. Parke (Eds.), *Routledge handbook of internet gambling* (pp. 227–249). London, United Kingdom: Routledge.
- Griffiths, M. D., & Wood, R. (2001). The psychology of lottery gambling. *International Gambling Studies*, 1, 27–45.
- Gulley, O. D., & Scott, J. F. (1993). The demand for wagering on state operated lotto games. *National Tax Journal*, 46(1), 13-22.
- Hubert, P. (2014, September 9-12). *Online e offline pathological gambler: characterizing and comparison* [Conference presentation]. 10th European Conference on Gambling Studies and Policy Issues, Helsinki, Finland.
- Hubert, P., & Griffiths, M. D. (2018). A comparison of online versus offline gambling harm in portuguese pathological gamblers: An empirical study. *International Journal of Mental Health and Addiction*, 16(3), 1219–1237.
- Kaizeler, M. J., & Faustino, H. C. (2008). *Lottery sales and per-capita GDP: An inverted U relationship*. Lisbon, Portugal: Technical University of Lisbon, School of Economics and Management. Working Paper No. WP 41/2008/DE/SOCIUS.
- Kaizeler, M. J., & Faustino, H. C. (2010). Demand for lottery products: An international study. *Revista de Economia e Administração*, 9(1), 1-16.
- Kaizeler, M. J., & Faustino, H. C. (2011). Is the lottery product an inferior good in higher income countries? *International Journal of Economics and Finance*, 3(4), 54-59.
- Kaizeler, M. J., & Faustino, H. C. (2012). Lottery sales and per-capita GDP: An inverted U relationship. *British Journal of Economics, Management & Trade*, 2(3), 225-238.
- Kaizeler, M. J., Faustino, H. C., & Marques, R. (2014). The determinants of lottery sales in Portugal. *Journal of Gambling Studies*, 30, 729-736.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127.
- Kotler & Armstrong (2018). *Principles of marketing, global edition, 17th Edition*. London: Pearson
- LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of internet sports gambling behavior. *Journal of Gambling Studies*, 23(3), 347–362.

- LaPlante, D. A., Gray, H. M., Bosworth, L., & Shaffer, H. J. (2010). Thirty years of lottery public health research: Methodological strategies and trends. *Journal Gambling Studies*, 26, 301–329.
- Lopes, H. (2009, October 19). *Epidemiologia da dependência do jogo a dinheiro em Portugal* [Conference presentation]. Congresso de Alto Nível promovido pela Santa Casa da Misericórdia de Lisboa. Lisbon, Portugal.
- Lopes, H. (2010, September 14-17). *Update of gambling dependence research in Portugal* [Conference presentation]. 8th European Conference on Gambling Studies and Policy Issues. Vienna, Austria.
- Lutter, M., Tisch, D., & Beckert, J. (2018). Social explanations of lottery play: New evidence based on national survey data. *Journal of Gambling Studies*, 34, 1185–1203.
- McCasland, M. (2005). Mobile marketing to millennials. *Young Consumers: Insight and Ideas for Responsible Marketers*, 6(3), 8-13.
- Mok, W. P., & Hraba, J. (1991). Age and gambling behavior: A declining and shifting pattern of participation. *Journal of Gambling Studies*, 7(4), 313-335.
- Moreno, F. M., Lafuente, J. G., Ávila Carreón, F., & Moreno, S. M. (2017). The characterization of the millennials and their buying behavior. *International Journal of Marketing Studies*, 9(5), 135-144.
- North American Association of State and Provincial Lotteries. (2021, September 1). *Glossary of Lottery Terms*. <https://www.naspl.org/glossaryoflotteryterms>.
- Ordun, G. (2015). Millennial (Gen Y) consumer behavior, their shopping preferences and perceptual maps associated with brand loyalty. *Canadian Social Science*, 11(4), 40-55.
- Parke, J. & Griffiths, M.D. (2007). The role of structural characteristics in gambling. In G. Smith, D. Hodgins & R. Williams (Eds.), *Research and Measurement Issues in Gambling Studies* (pp.211-243). New York: Elsevier.
- Perez, L., & Humphreys, B. (2013). The ‘who and why’ of lottery: Empirical highlights from the seminal economic literature. *Journal of Economic Surveys*, 27, 915-940.
- Pordata (2019a). Poder de compra per capita 2013-2014. *Pordata*. Retrieved August 26, 2021, from <https://www.pordata.pt/Municipios/Poder+de+compra+per+capita-118>
- Pordata (2019b). População residente: total e por grandes grupos etários 2013-2014. *Pordata*. Retrieved August 26, 2021, from <https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente+total+e+por+grandes+grupos+et%C3%A1rios-390>
- Pordata. (2020). PIB per capita (base=2016). *Pordata*. Retrieved August 26, 2021, from [https://www.pordata.pt/Municipios/PIB+per+capita+\(base+2016\)-896](https://www.pordata.pt/Municipios/PIB+per+capita+(base+2016)-896)
- Rogers, P. (1998). The cognitive psychology of lottery gambling: A theoretical review. *Journal of Gambling Studies*, 14, 111–134.
- Rogers, P., & Webley, P. (2001). "It could be us!": Cognitive and social psychological factors in UK National Lottery play. *Applied Psychology: An International Review*, 50, 181–199.
- Santa Casa da Misericórdia de Lisboa (2014). *Relatório e Contas do Departamento de Jogos da Santa Casa da Misericórdia de Lisboa*. Departamento de Jogos da Santa Casa da Misericórdia de Lisboa. Lisboa: Santa Casa da Misericórdia de Lisboa. Retrieved August 26, 2021, from http://www.scml.pt/pt-PT/santa_casa/relatorio_e_contas/
- Santa Casa da Misericórdia de Lisboa (2015). *Relatório e Contas do Departamento de Jogos da Santa Casa da Misericórdia de Lisboa*. Lisboa: Santa Casa da Misericórdia de Lisboa. Retrieved August 26, 2021, from http://www.scml.pt/pt-PT/santa_casa/relatorio_e_contas/
- Santa Casa da Misericórdia de Lisboa (2020). *Relatório e Contas do Departamento de Jogos da Santa Casa da Misericórdia de Lisboa*. Lisboa: Santa Casa da Misericórdia de Lisboa. Retrieved August 26, 2021, from http://www.scml.pt/pt-PT/santa_casa/relatorio_e_contas/

- Shaffer, H. J., Peller, A. J., LaPlante, D. A., Nelson, S. E., & LaBrie, R. A. (2010). Toward a paradigm shift in internet gambling research: From opinion and self-report to actual behavior. *Addiction Research and Theory, 18*, 270–283.
- Smith, K. T. (2011). Digital marketing strategies that millennials find appealing, motivating, or just annoying. *Journal of Strategic Marketing, 19*(6), 489-499.
- Smith, K. T. (2012). Longitudinal study of digital marketing strategies targeting millennials. *Journal of Consumer Marketing, 29*(2), 86-92.
- Tom, M. A., & Shaffer, H. J. (2016). Devil in the details: A critical review of "theoretical loss". *Journal of Gambling Studies, 32*, 865–875.
- Tom, M. A., LaPlante, D. A., & Shaffer, H. J. (2014). Does pareto rule internet gambling? Problems among the "vital few" & "trivial many". *Journal of Gambling Business & Economics, 8*(1), 73-100.
- Ukhov, I., Bjurgert, J., Auer, M., & Griffiths, M. D. (2021). Online problem gambling: A comparison of casino players and sports bettors via predictive modeling using behavioral tracking data. *Journal of Gambling Studies, 37*, 877-897.
- Wood, R.T.A. & Griffiths, M. D. (1998). The acquisition, development and maintenance of lottery and scratchcard gambling in adolescence. *Journal of Adolescence, 21*(3), 265-273.
- Wood, R.T.A. & Griffiths, M. D. (2004). Adolescent lottery and scratchcard players: do their attitudes influence their gambling behaviour? *Journal of Adolescence, 27*(4), 467-475.
- Xuan, Z., & Shaffer, H. J. (2009). How do gamblers end gambling: Longitudinal analysis of internet gambling behaviors prior to account closure due to gambling related problems. *Journal of Gambling Studies, 25*, 239–252.
- Yuan, J. (2015). Examining the gambling behaviors of Chinese online lottery gamblers: Are they rational? *Journal of Gambling Studies, 31*, 573–584.