

# *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*



## **BRIEF OVERVIEW OF PRESENTATION**

- I. Objectives of the Present Study
- II. Gambling studies using behavioral tracking player data
- III. Lottery gambling in Portugal
- IV. Methods
  - a) *Participants*
  - b) *Data*
  - c) *Measures*
  - d) *Data analysis*
- V. Results
- VI. Discussion
- VII. Conclusions And Further Research



## **I. MAIN OBJECTIVES/ MAIN RESEARCH GOAL:**

- The **identification of groups or segments of players** by their **engagement levels** (high, neutral, low)
- **Assess preferences in product category** with the use of **segmentation models**, based on expenditure and sociodemographic variables.
- The present study **aimed to identify remote lottery players segments using real playing data from the Portuguese national lottery** with nationally representative data.



## Segmentation:

- **Segmentation is a strategy 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.**



## Research Questions

**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 remote lottery playing?
- (v) Does **education have a negative correlation** with remote 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 remote lottery gambling segments than others?



## **II. GAMBLING STUDIES USING BEHAVIORAL TRACKING PLAYER DATA**

- **Few studies using behavioral tracking player data include lottery gambling** (e.g., Auer & Griffiths, 2013, 2014, 2016; Auer, Hopfgartner, & Griffiths, 2018; Gray, Jonsson, LaPlante, & Shaffer, 2015)
- The approach taken in **previous studies did not focus on segmenting lottery players** using real playing data.
- **Many of the previous studies** that used real playing tracking data **did not use nationally representative data.**



### **III. 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).
- **No previous studies focus specifically on remote lottery players**, especially with the use of real playing data.
- **Only one study on lottery gambling in Portugal** (based on self-report data) **can be considered nationally representative** (N=3,850; ages 18–70 years) (Lopes, 2009, 2010), but does not analyze remote lottery players nor does it use real playing data.



## IV. 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.





## IV. METHODS

### *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 study's **dataset contained all active Portuguese remote lottery players**, and the **final sample was representative of all remote lottery players** ( $p < 0.01$ ).
- The dataset's sample used in this study accounts for 2.6% of the Portuguese adult population.



## IV. METHODS

### ***Measures:***

- 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),
  - wagers (total number of wagers made and number of wagers made per game),
  - number of lottery draws (total and per game),
  - number of lottery draws in which each player placed wagers (total and per game), and
  - number of weeks in which each player placed wagers.
- 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.



## IV. METHODS

**Measures: The games on the portfolio were aggregated into five product categories**

- Lotto games (*EuroMillions; Totoloto*);
- Lotto games with the inclusion of an add-on game (*EuroMillions; Totoloto and Joker*);
- Toto/sports lottery (*Totobola: 1X2*),
- Passive/class lotteries (*Lotaria Clássica and Lotaria Popular*);
- Scratch-cards (*Lotaria Instantânea*).



## IV. METHODS

### Measures: Other new variables included

- **“Amount lost”** (total amount spent vs. total amount won)
- **“Amount spent per wager”** (mean average).
- New variables were used to assess playing engagement. They were created for the total individual gambling activity, for each lottery game individually and for each of the lottery product categories created.
- Some of the existing variables were converted into new variables by using exogenous information
  - Examples variables created by converting postal codes into local territorial organization regions.



## IV. METHODS

*Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

### *Data analysis*

- **The analysis in this study comprised three main steps.**
  1. The first step **compared the profiles of the individuals in the sample to the general population**
  2. 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.
  3. The third step comprised a **multivariate Chi-square Automatic Interaction Detector (CHAID) analysis** 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
    - The dependent variable in this study was total amount of money spent
    - The CHAID model was built to identify player segments and identify the most relevant among them (target market).



## V. RESULTS

### *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

#### **Comparing the profiles of the individuals in the sample to the general population**

- The findings indicated that there was a large **discrepancy between the gender distribution of the adult population and remote lottery players.**
- Results showed that the differences were statistically significant (confidence level 96%;  $p < 0.05$ ).
  - **Male remote lottery players** comprised **78.7%** of total players ( $n=121,601$ ) and in Portugal there are **46.7% adult males** (approximately 3.99 million males) (Pordata, 2019b).
  - **Female remote 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).



## V. RESULTS

### *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

Considerable differences were also found regarding education and age

- **Most players had a higher education (51.7%) compared to 20.6% of the Portuguese adult population (there were no players without any formal education).**
- Results showed that **remote lottery players had a different distribution regarding age groups** when compared to the Portuguese adult population ( $p < 0.05$ ).
- The mean age of remote lottery players in Portugal tended to be younger than the national adult population.
- **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).**
- Overall, **remote 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%).**



## V. RESULTS

# *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

### **Considerable differences were also found regarding geographic location**

- The data were also analyzed the geographic location and distribution of players which was similar to the geographic concentration of the general Portuguese adult population but there were differences in two main regions.
- For remote lottery players, the Lisbon Metropolitan Area was the most represented while among the general population it is second, after the North region.
- 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).





## Player Characterization

### Analysis by Gender

- **Male remote lottery players were found to spend more money** on average than female online lottery players and placed more wagers
- The only game category where **no statistically significant differences** in mean expenditures were found between males and females was **instant lottery games**



## Player Characterization

### Sample analysis by age group

- The present study found a positive association between age and lottery gambling expenditure. **Expenditures and wagers steady increase from the youngest to the oldest age groups ( $p < 0.01$ ).**
- Results confirm recent findings on the positive association between age and gambling engagement (Auer, Hopfgartner, & Griffiths, 2018).
- The **18-34-year** age group placed on average 59.39 wagers, with an average amount per wager of €3.06.
- The **35-49-year** age group placed on average 103.86 wagers, with an average amount per wager of €3.62.
- Players in the **50-64-year** age group placed on average 178.35 wagers with an average amount per wager of €3.95.
- The oldest players (**65 years and older**) had an annual average of 214.51 wagers with an average amount per wager of €4.35.



## Player Characterization

### Sample analysis by age group

- In the present study, players aged **18 to 34 years old represented 28.8% of players.**
  - 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 ( $\geq 65$  years; 5%) and the youngest subgroup (16 to 24 years; 7.5%) were the least represented.**
  - 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.
  - Although older males ( $\geq 54$  years) tended to be the most engaged players overall.



## V. RESULTS

### Player Characterization

#### Sample analysis by education

- Findings indicated that annual **lottery expenditures decreased as education levels increased** 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%).



## V. RESULTS

# *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

## Player Characterization

### Sample analysis by location/ region

- 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;  $p < 0.01$ ]).



## V. RESULTS

# Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery

Lottery expenditure by age group, education, and region									
	n	Mean	Mode	Median	SD of total expenditure	VAR	SD mean expenditure	Q1	Q3
<b>Age groups</b>									
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
<b>Gender</b>									
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
<b>Education</b>									
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
<b>Region</b>									
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



## **V. RESULTS**

### **Chi-Square Automatic Interaction Detection (CHAID) Segmentation Model**

- **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 remote 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 (total monetary expenditure ).**
- The total individual expenditure CHAID decision tree generated 22 nodes at four levels (14 final nodes )
- Each node was considered a different player segment.



# High, neutral, and low engagement player segments

## V. RESULTS

Nodes/Segments		Characterization Variables			Players			Expenditures		
Segment	1	2			3	4	5	6=3x5	7=6/total 6	8=7-4
Engagement	Node nº	Age Group	Gender	Education / Location	N	Segment Weight %	Mean	Segment Expenditure	Segment	Segment
									Expenditure %	Relevance
High	13	50-64	Male	4th grade; 6th grade; High School	15,449	10.00%	592.92	9,159,936.30	17.58%	<b>7.58%</b>
	14	50-64	Male	Higher Education	11,117	7.20%	669.85	7,446,715.20	14.29%	<b>7.09%</b>
	11	≥ 65	Male	-	6,690	4.30%	785.12	5,252,464.40	10.08%	<b>5.78%</b>
Neutral	17	35-49	Male	Madeira	1,766	1.10%	488.99	863,556.90	1.66%	0.56%
	6	50-64	Female	-	5,393	3.50%	360.2	1,942,559.60	3.73%	0.23%
	20	35-49	Female	6th grade	459	0.30%	576.65	264,681.30	0.51%	0.21%
	12	≥ 65	Female	-	966	0.60%	403.35	389,632.40	0.75%	0.15%
	22	18-34	Male	4th grade; 6th grade	1,991	1.30%	197.39	393,008.30	0.75%	-0.55%
16	35-49	Male	Centre	10,948	7.10%	303.78	3,325,781.20	6.38%	-0.72%	
Low	15	35-49	Male	Lisbon; North; Alentejo; Azores	42,222	27.40%	324.36	13,695,149.50	26.29%	-1.11%
	18	35-49	Female	4th grade; High School	5,267	3.40%	216.09	1,138,152.80	2.18%	-1.22%
	19	35-49	Female	Higher Education	9,803	6.30%	190	1,862,597.40	3.57%	-2.73%
	10	18-34	Female	-	11,096	7.20%	124.1	1,376,962.10	2.64%	-4.56%
	21	18-34	Male	High School; Higher Education	31,418	20.30%	158.83	4,990,036.90	9.58%	-10.72%
<b>TOTAL</b>					154,585	100.00%	337.04	52,101,234.30	100.00%	





## V. RESULTS

### *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

- Segments with high engagement are those in which the difference between the percent contribution of each segment to total expenditure and the percentage of the sample within each segment (segment weight) is >1%.
- Segments where parameters fall within 1% more or less were considered neutral engagement segments.
- Segments in which the difference was greater than -1% were considered low engagement.

**The findings indicated that in the**

- (i) high engagement level, 21.5% of players were responsible for 42.0% of overall expenditure,**
- (ii) neutral engagement level, 13.9% of players were responsible for 13.8% of overall expenditure,**
- (iii) low engagement segment, 64.6% of players were responsible for 44.3% of overall expenditure.**



## V. RESULTS

### *Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery*

The CHAID segmentation models determined a four-level hierarchy of the demographic variables ranked by relevance and engagement.

1. Age (most relevant)
2. Gender,
3. Education,
4. Place of residence.



## Conclusions

- The **CHAID segmentation models** determined a **four-level hierarchy of the demographic variables ranked by relevance and engagement (1. Age, 2. Gender, 3. Education, 4. Place of residence)**.
- 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**.



## Conclusions

- Findings from this study shed **new insight into age and lottery gambling** and present new empirical evidence against age being negatively correlated with **remote lottery gambling behavior in Portugal**.
- **Remote 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**.
- **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.



## **Limitations.**

- Data used were from only one website and players might be gambling on various websites and/or gambling in land-based venues.
- The present study can only be considered representative of Portuguese remote/internet lottery players and not of land based lottery players or of all other online gambling activities.
- Users might share their accounts and access passwords with other individuals.
- 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.



## Further research

- Identifying the profile of players across multiple gambling activities, from different providers
- Studies combining playing tracking data with self-reported gambling data
- 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*). It would be useful to assess if there is 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 COVID-19 pandemic has driven offline players to start playing online as most points of sale were closed during the lockdowns.

# Consumer profile segmentation in lottery gambling utilizing behavioral tracking data from the Portuguese national lottery

*Thank you!*

*Questions?*



*Chagas, B.T.,  
Gomes, J. F. S. &  
Griffiths, M. D.*

Published in the Journal of Gambling Studies  
doi: 10.1007/s10899-021-10072-9.

**Bernardo Chagas**  
bernardo.chagas@scml.pt  
jchagas@iseg.ulisboa.pt