LUISS

Dipartimento
di Impresa e Management

Cattedra Quantitative Methods for Management

Micro-transactions and loot boxes: predicting consumer preferences via machine learning

RELATORE
Chiar.mo Marco Pirra

CORRELATORE
Chiar.mo Paolo Porchia

CANDITATO
Lorenzo De Gasperis
714171

Anno Accademico 2019/2020

Prima di procedere con la trattazione, vorrei dedicare qualche riga a coloro che mi sono stati vicini in questo percorso di crescita personale e professionale.

Un grazie al mio relatore il prof. Pirra per la sua disponibilità e tempestività nel rispondere ad ogni mia richiesta e per i suoi suggerimenti di perfezionamento dell'elaborato.

Un ringraziamento va anche ai miei amici che riuscivano sempre ad allentare la tensione universitaria con una bella risata e quattro chiacchiere in compagnia.

Infine, un ringraziamento speciale va alla mia famiglia senza la quale non avrei ma potuto raggiungere questo importante traguardo.

In particolare, voglio ringraziare mia madre che è sempre stata al mio fianco anche e soprattutto nei momenti di sconforto. Mamma, è solo grazie al tuo sostegno e al tuo incoraggiamento se oggi sono riuscito a raggiungere questo traguardo.

Ora però la sfida si fa più dura ma dopo tutto ciò che ho affrontato, grazie anche al supporto di tutti voi, ora sono sicuro di poter superare qualunque ostacolo la vita mi metterà dinnanzi.

Grazie di nuovo a tutti, senza di voi non ce l'avrei mai fatta.

Table of Contents

1. Introduction

1.1. Research Study background

Quite often, when serious conversation about video games takes place and an open discussion begins, what emerges is just a blurred scenario in which the single components that constitute the vast world of the industry play their role. Video games are, in fact, perceived as partially (or sometimes completely) distant and somewhat detached from reality and are thus easily dismissed from conversation. It is undeniable that computer and mobile gaming has not only become very widespread in recent years, but that competition among developers and producers is increasingly intense. In this extremely fierce scenario in which competition has reached its peak, game developers needed to find alternative ways to generate "extra" return in order to create a sustainable stream of cash and revenues.

In recent years, the industry of video games is no longer merely limited to console or pc gaming as it was in the past, but the entire system has been radically shaken by the development of additional features and content such as mobile gaming and e-sport. This has led to additional revenues for the entire industry that are no longer coming solely from the sale of consoles, video games and technology but also from e-sport tickets and the complete redefinition of an entire business model of the gaming companies: the introduction of in-game content to be purchased via "micro-transactions". We witnessed for the first time the creation and development of this new form of payment that eventually propelled the video game industry and guaranteed astronomical revenues in the years that followed.

During the last two/three years, the video game industry has started to develop an articulated and complex system to generate revenues. A huge portion of the sector's profits, in fact, comes from the design, creation and implementation of this new system of monetization, namely the introduction of micro-transactions in online video games through which, at the cost of a nominal amount of real-world currency (hence the prefix "micro-"), the user purchases certain virtual items to be used in the game. The mechanism of design, development, and implementation of micro-transactions has meanwhile become so relevant that many, nowadays, have begun referring to an entirely new "Business Model". These micro-transactions have been studied to optimize the results of the purchase and the experience both of customers that are more inclined to spend as well as those which are not.[1] As previously explained, optimally selecting the price and tailoring the purchase in order to achieve a revenue-maximizing strategy for video game companies, as well as analysing customers' reactions and propensity to spend according to certain selling mechanisms, has become more and more aligned to what is generally referred to as a "Business Model".

These transactions enable players to access additional game content (generally referred to as "premiums") which include, but are not limited to, virtual items, textures and skins, in-game currency, levels or power-ups to name just a few. These well-known micro-transaction mechanisms are extremely common in mobile gaming and constitute the highest percentage of revenues for most "free-to-play" (also referred to as "f2p", that is to say games that do not require any paying commitment).

In most recent games that include micro-transactions, it is clear to players that developers have included a progression system in which the player's skills and patience is not adequately rewarded. Game progression based on the ability of the player is instead replaced by the purchase of in-game currency in order to accelerate otherwise lengthy processes or to overcome the most challenging obstacles. In these games players have the initial illusion that the equilibrium of the game is well-balanced (e.g. the amount of in-game experience is proportional to the effort of the player in the game, the amount of rewards achieved are reasonable and other additional features that characterize the game are correctly related to the time spent) only to discover, as the game progresses, that such equilibrium is actually based on the speed up processes via in-game currency as mentioned earlier.

Among the methods used to customize and tailor micro-transactions to users' expectations is the presentation of the products and features. One popular way to propose the products is via a so-called loot box (also known as "loot prize/crate"), that is to say a virtual item which can be redeemed in order to receive a selection of additional virtual items or "loot", ranging from simple customization articles for the player's character to armour and weapons. The rise in popularity of micro-transactions and loot boxes in particular also brought about concerns as to whether or not they could simply be considered another form of gambling. To enforce this argument, many claim that the topic is especially relevant if we consider that, in many cases, the virtual items exchanged in the game can be "cashed out" for real-world money.

[1] "Scientific Revenue Introduction Video" available at https://vimeo.com/154271693. Ever wonder why dynamic pricing in video games is so important? This short video explains how Scientific Revenue will help you work less and make more money.

1.2. Reason for the research study

In the history of video games, every so often, the "attention" towards the industry suddenly awakens, and every time this occurs, it is due to the increasing pressure that one of the industries with the highest revenue generating systems exerts on other industries. The last time the entire community rose in response to a scandal in the video game industry can be traced back to the summer of 2019: in that particular instance it became relevant that the already established but forgotten implementation of micro-transactions in the business required the attention of special laws and regulation. For this reason, in many countries all over the world, especially in Asia, games with loot crates systems have become subject to strict regulations. In China, South Korea and Singapore for example, it has become compulsory to represent the allocation probabilities of each virtual item that can be acquired via the purchase of loot crates (which is exactly what happens in the case of gambling). This can be easily understood as certain case studies have proven that when a gaming company wilfully misrepresents the probabilities of the content of their loot boxes, their revenues increase significantly, hence the need for additional and more stringent regulations to discipline the activities of gaming companies.[2] The most publicized of these rulings is the one that dates back to April 2018 when The Netherlands and Belgium determined that some loot box systems in the market violated the local laws on gambling and needed clarification.[3]

1.3. Scope of the research study

The aim of this research study is to provide a general understanding of the phenomenon of loot boxes and of the micro-transaction monetization system. It intends to provide an extensive analysis of the background regarding the topic and to develop a quantitative study that presents an in-depth statistical data analysis in order to better comprehend the potentially related variables among those analysed. The information provided will then be necessary to understand the extent of the analysis developed as well as the rationale underlying the answer to the thesis question: to what extent and with which degree of accuracy can the algorithms developed by the software houses predict the spending habits of video games players?

With this in mind, we will begin by providing an overall description and characterization of the topic in order to allow the reader to acquire the necessary background information and

[2] Fingas 2018 "South Korea fines game studios over deceptive loot box odds" (see references)
[3] Gaming Authority. Study into loot boxes. A treasure or a burden? 10 April 2018 (see references)

fully appreciate the explanation regarding the methods used by software houses to publish and package micro-transactions and loot boxes.

1.4. Methodology

This paper is based on the responses of a questionnaire submitted to video gamers and shared online from the 1st of June until the 10th of September 2020. It sampled 243 video gamers and 134 loot box purchasers using the following criterion:

- the survey was uploaded on 22 Facebook pages and 14 subreddits, most of which were related to either general gaming or to video gaming in particular;
- the questionnaire was submitted in English (with the possibility to be translated into different languages depending on the browser used in the filing) in order to maximize the accessible audience;
- in order to be eligible to take the survey, it was necessary that respondents had played video games in the 12 months preceding their participation to the survey or had purchased loot boxes (or similar in-app purchases) in the past;
- prior to initiating the survey an agree/disagree policy was submitted to reassure participants of the non-disclosure of their personal data and to gather their consensus to use the information provided albeit for academic purposes only.

After the gathering phase, the data collected was re-elaborated to make it easier to manage and "translated" for the envisaged purpose. The converted data was then passed through an encoder which associated and replaced each item and answer with a number.

A series of investigations was performed in order to statistically infer that the data collected was significant. To be more specific, we ran Chi-Square tests of independence, portrayed a Variance-Covariance matrix and elaborated a model that was able to predict the value of a variable after a machine training phase. This last test can be considered the major achievement of the paper: through a simple linear multi regression fit, we were able to estimate the outcomes of certain variables an acceptable number of times.

In this research study we will examine the following:

- the general characteristics of the video game market;
- recent studies on loot boxes and gambling-related behaviours which will be reviewed and commented;

- different in-game purchasing options and other monetisation strategies which will be explained in detail;
- trends and/or relations between the variables analysed; and
- recommendations and outputs of our research study.

2. Video game market – profit and competition

2.1. Sources of revenues and trends

The industry of video games has developed remarkably from the once niche business that involved only a few, and has grown to become a monstrous giant generating billions of dollars in revenues every year. These revenues basically derive from two major sources which can be divided into **hardware**, which includes consoles, processors, screens, controllers and other additional accessories, and **software** which comprises the actual game and the additional content that can be purchased directly from the in-game platform. Software is by far the largest segment of the broader video games industry. The rise of online-only software, such as the astonishingly successful Fortnite, has stolen a considerable share of the traditional boxed and downloadable games even though sales of these latter forms of entertainment have nonetheless continued to grow steadily.

The current leader in the industry is Sony Computer Entertainment (or simply "Sony"), a hardware and software development company with headquarters in Japan. In its 2020 consolidated financial statements for the fiscal year ending March 31, 2020, the company reported 8,259,885 million JPY (equal to approximately 78 billion USD) of sales. Other major players of the market include:

- Tencent Holdings Limited (with headquarters in Shenzhen, China) which has the complete control of Riot Games and Grinding Gear Games (the developer companies of League of Legends and Path of Exile respectively), and whose portfolio includes 84.3% of Supercell, developer company of Clash of Clans, and 40% of Epic Games, developer of Fortnite;
- Microsoft Corporation (with headquarters in Redmond, Washington, in the United States) also specialized in consumer electronics, personal computers, and related services and considered one of the "Big Five" technology companies together with Amazon, Apple, Google, and Facebook;
- and Nintendo Company Ltd. (with headquarters in Kyoto, Japan) which is one of the most successful console developers and manufacturers in the industry, such as the Game Boy, the Super Nintendo Entertainment System, the Wii, and the Nintendo Switch. Nintendo also deserves an honorable mention for having developed some of

the most iconic and influential franchises in the video game industry of all times, such as Donkey Kong, Metroid, Mario, Kirby, The Legend of Zelda, Pokémon and many others.

Sony's PlayStation 4 is the best-selling console among those of the current generation and, in 2020, sales of this console reached more than 108 million units. However, the best-selling gaming console of all times remains PlayStation 2, which was released in the year 2000 and which, to date, has sold over 155 million units. Notwithstanding the predominant position of the Sony company, Nintendo's Wii Sports, a video game released for the Wii console in 2006, is the best-selling console game in the world with more than 82.6 million units shipped worldwide and to this day, the life simulation video game series, The Sims 3, is ranked as the bestselling PC game of all times, with 7.96 million units sold worldwide.[4] Total revenues in the US market alone amounted to 19.8 billion USD in 2018, with a comprehensive annual growth rate of 3.5% from 2014 to 2018. Over the same period the European and Asia-Pacific markets instead had an average comprehensive annual growth rate of 3.1% and 6.7% respectively, and, in 2018[5], reached the respective values of 12.1 billion USD and 13.5 billion USD. The US video game software market remains the world's largest and continues to experience a slow but steady growth. Sales of software for Sony's PlayStation 4 for instance earned the largest market share in 2018 and represented the market's major growth driver.

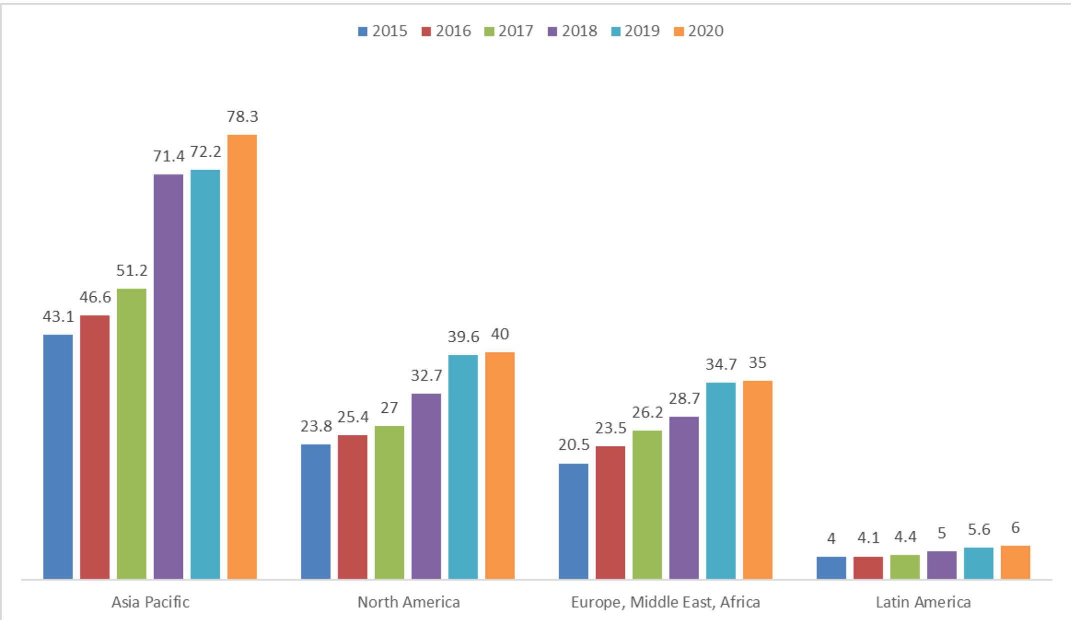


Figure 1 - Games market revenue worldwide from 2015 to 2020, by region (in billion U.S. dollars)

[4] Data gathered by statista.com
[5] See MarketLine references.

Figure 1 above shows the market revenue of the gaming industry from 2015 to 2020. The graph clearly evidences the steady expansion of the gaming industry across the period in question in different geographical locations. With specific regard to online "pure play", that is to say the share of the market that is solely characterized by online purchases of gaming titles and of additional content via micro-transactions, the relative revenues accounted for the largest portion of the US software industry production in 2018. This channel alone generated more than 5.2 billion USD, equal to approximately 26% of the entire market value.

The industry of video games however does not consist only of software. If we consider additional sectors and different geographical regions in the computation, the resulting picture is very different. As a matter of fact, more recent trends indicate even greater increases in revenues. For instance, the worldwide PC gaming market is estimated to reach almost 37 billion USD by the end of 2020, and projections further estimate the mobile gaming sector to generate income of over 77 billion USD by the end of the year. In April 2020 alone, for instance, digital games worldwide generated revenues of over 10.54 billion USD, marking the highest total ever. This was undoubtedly a direct consequence of the lockdown made necessary by the COVID-19 threat which forced many people across the globe to stay home and seek means of entertainment compatible with such new lifestyle.[6] However this could not have been possible before the introduction of digital download and the possibility to acquire games directly from home without the need for a physical point of sale.

Incidentally, digital download and online purchases paved the way to another core characteristic of the video game industry as it is known today: the introduction of micro-transactions and predatory dynamics which have been evolving, especially in the mobile gaming market. These dynamics were born from the growing need of gaming companies to generate revenues and to face the increasing costs of the production of video games (factor which indie developers[7] did not have to face), and their implementation was only made possible by the broadening of the internet connection and the development of the new data transfer system represented by digital download. The creation of these micro payments is based on a simple logic: to produce a stable and year round revenue stream and reduce uncertainty in an industry that used to be seasonal. Additionally, micro-transactions reduce

[6] Data gathered by statista.com

[7] By "indie developer" we intend developers of video games which are generally created by a single individual or small teams of developers and usually do not benefit from funding. These small businesses generally experience negligible costs and the technical effort invested into the production of their games may vary from a few days to years, depending on the number of hours spent in the development, on the complexity of the game, on the number of participants etc.

the rising costs of the business (the delivery of in-game services has a much lower cost compared to those associated to the game product itself) and generally maintain the longevity of the interaction between the player and the game, which, in turn, favor the purchase of even more content. One should simply consider that in 2019, the game publisher Activision Blizzard alone reported that almost 5 billion USD (approximately 4.932 billion USD, equal to about 76% of its net revenues) was related to micro-transactions and, the same year, Electronic Arts reported about 2.2 billion USD (approximately 45% of its net revenues) of micro-transactions and additional 1.494 billion USD of other digital revenues (which include full game downloads and the mobile sector), for a total revenue of 4.95 billion USD. If we compare these numbers to those of 2012, when EA's micro-transactions network was not yet radically implemented into its business model, we witness a radical change in the business profile, with 1.16 billion USD of micro-transactions in 2012 which correspond to 28% of its net annual revenue.

2.2. Industry Potential and Perspectives

From the advent of micro-transactions, the business of video games has been expanding steadily. The most prominent reason that led companies to introduce the micro-transaction mechanism within their business models is mainly due to the lucrative options that it can provide. As a matter of fact, this monetization systems allows ample returns with minimal investment (in terms of resources and maintenance costs) on the developer's behalf. Micro-transactions become even more powerful if implemented into franchises which have an already strong position in terms of brand identity/loyalty or which command hefty royalties. Additionally, as mentioned, this system enables the break of the seasonality of the industry, which is generally characterized by releases during the last quarter of the year. This transition to services directly related to the games, which are embedded in micro-transactions, guarantees a more stable revenue stream all year round, satisfying investors and developers. In reality there is little to no brand loyalty in the video game software market due to the extreme diversity and to the fast pace of software development. The only exceptions to this rule are individual titles which are generally called "killer application". Grand Theft Auto, for instance, is one of the best examples of this phenomenon. Grand Theft Auto V, developed by Rockstar, is the best-selling media product of all time, and was still on top of most popular games' charts in 2018, five years after its initial release. This lack of loyalty and the propensity of consumers to easily switch between games, in fact, grants power to the individual buyer which sees its position strengthening. Additionally, the vast market of video games is characterized by a various number of distinct players. There are strong differences between supplier size and very small development studios compete with huge giants leveraging on their higher degree of flexibility and innovative ideas or by partnering with great publishers. These large companies, on the other hand, can rely on a vast portfolio of different game titles and take advantage of the already established business to create economies of scale in every field of production, ranging from the idea conception to the distribution of the products.

In order to understand the scale of the industry, one should reflect upon the idea that the "first generation of gamers", meaning those who were born in the 80's and in the 90's, have now acquired financial stability and have meanwhile significantly increased their spending potential. In fact, video games can no longer be considered merely a product for children and there are studies that prove that gaming is gaining more and more popularity also among parents all over the world. There are, in fact, plenty of statistics that show that, although 24% of adults claim they do not play video games at all, almost 18% state that they spend more than six hours per week playing video games. According to some studies conducted by the Entertainment Software Association (or simply "ESA"), the trade association of the video game industry in the United States, 65% of Americans over the age of 13 play games, and the distribution between male and female is nearing parity.[8] Another study of the ESA claims that 94% of US parents monitor the online activity of their children and some of them (about 45%) also pay attention to the games their children play and to their content.

The highly competitive market creates room for each of its players. On the one hand, development studios, even the smallest ones, could be able to win the market with innovative ideas and cutting-edge proposals and conquer their fair share of the market; on the other hand consumers have the potential to shake the market, moved by a mutual interest in order to achieve a common goal due to the solid and compact community that has developed in the recent years.

The new monetization system of micro-transactions, and of loot boxes in particular, might generate potential damage in terms of image and trust between development companies and consumers. The most prominent example is the video game Destiny 2, a free-to-play (f2p) multiplayer first-person shooter developed by Bungie Studios and released on September 6, 2017. The game first started as a regular title to be purchased for the initial price of 69.99

[8] This estimate includes any type of video game on any platform (smartphone included).

USD which included the complete game with no additional benefit (multiple purchasing packages were developed and they included season pass for future expansions of the game, additional virtual content and even more accessories for the Collector's Edition). It became the highest selling game of 2017 and its huge success caused the developers to decide to distribute the game as a free to play. The game became freely downloadable on October 1, 2019. After the shift to a free to play, Destiny 2 generated lots of criticism, but it is still selling well despite the internet grumbling. This indicates that "freemium[9] games" are probably generating an interest in the consumers. The game is, in fact, one of the best-selling games of 2019.

The proposed analysis and the considerations on the entire market provided in the previous Section 2.1, suggest that micro-transactions will likely remain a constant in the online video game community as long as there are players who are still interested in the additional content such micro-transactions deliver. Developers will strive to push micro-transactions and loot boxes even further so they will probably become increasingly prevalent in the coming years. While the discussion goes wild on the internet, it is more than likely the games will still sell. Unless there are mass consumer boycotts or regulatory interventions, there will not be many choices available to the consumers; they will either pay up or miss out on the opportunity to play. Additionally, the system may also reveal the attempt to exploit of problematic tendencies, leveraging on medium to low income customers and cause people to spend a lot of money on these games. Nowadays, in fact, even games which try not to capitalize on huge "paywalls", in order to open customers to use micro-transactions, become prone to manipulative tendencies and are inclined to create game mechanics around a system which tries to encourage players to spend due to its high profitability.

[9] Freemium is a pricing strategy by means of which a product or service is provided free of charge but which includes additional features at the expense of in-game purchases.

3. Literature Review

This chapter will address the terminology and the characterizations that will be used at a later stage in the study. As a premise, it is important to bear in mind that a lot has been analysed and discussed in recent years with regard to loot box mechanics and gambling habits. There are, in fact, many papers which manage to connect gambling and purchasing habits of gamers to loot box mechanics systems. However, the available literature either lacks extension in terms of in-depth analysis (focusing solely on one aspect of the topic, for instance) or is the result of sociological studies which, instead, lack quantitative insight or consist of a descriptive overview of the topic with demographical analysis and inferences based on observed variables. Nevertheless, the contribution of previous studies and analyses relative on the subject matter at stake has been of great importance to this paper: besides gathering all the available materials developed and summarized in the aforementioned literature, this summary provides an extensive analysis on the topic of the loot box monetization system, with particular focus on Corporate Responsibility and gambling controversies and develops a quantitative study that presents in-depth statistical data analysis to answer the research question, also through self-collected data, interviews and survey data.

Figure 2 The image is a satirical juxtaposition that associates loot boxes with gambling

3.1. Defining the term "loot box"

There is, no general consensus on a precise definition of the term "loot box". Thus, a strict criterion pertaining to the study for the term in question was developed in order to avoid misinterpretation and create a strong base to build upon. The specific and defined criteria that a product needs to satisfy in order to be identified as a loot box can be described as follows:

1. A loot box is a virtual bundle of one or more virtual items. Some games do include “physical loot boxes”, which are shipped and delivered to purchasers and that contain unknown random merchandising products (including T-shirts, labelled gadgets, glasses etc.). However, due to the scarcity of their use and their rare implementation, for the purpose of this paper, physical loot boxes have been excluded from the research study and from the definition of loot box.
2. Upon purchase, the loot box can be opened, and its material must reward the player with contents that can be used for that game in some way or at least has a value for the player. The content may include cosmetic items, in-game features and abilities (power ups, weapons and upgrades), in-game currency, functional items (such as, for instance, new playable characters), or even more loot boxes. From this point on, the content of the loot boxes will be referred to as “rewards” and will characterize all types of content which may be acquired through loot boxes (regardless of the type of the content). The rewards may be consumables, which means that they can be spent in some way and disappear after use, but they must have some degree of concreteness and can be stored and used in subsequent play sessions.
3. Loot boxes do not need to visually resemble a “box” or “crate” but can take many different forms as long as they are characterized as being a container for in-game random rewards. Loot boxes may in fact appear in different forms other than boxes such as packages, cases, chest, crates, spinning wheels, etc.
4. Loot boxes must contain random content (or must at least appear random to the customer).

Additionally, it is important to clarify that at least two different types of loot boxes exist and a brief overview of both will undoubtedly greatly benefit further specific considerations. The two distinct categories of loot boxes mentioned above actually describe two different approaches of selling loot boxes (both of which leverage on the purchasing power of customers) and involve either “closed-loop mechanics” or “cashing-in mechanics”..

Closed-loop loot box mechanics may involve micro-transactions and their key characteristic is that the items obtained from the purchase of the loot box can only be used in the game. In other words, the value of the rewards is intrinsic to the game itself and can, in no way, be transferred to real-world currency (additionally, such rewards cannot be sold to other players). Although these games do not have the intention to create a market of buying and selling rewards, unauthorized markets may emerge and transactions between players may occur. This is a key element if we consider that game developers do not require a specific license to implement loot boxes that are based on closed-loop mechanics. Although rewards contemplated in closed-loop loot box mechanics are not associated with monetary gain, they might be valuable to a player in different ways and can even be used as virtual currency in other on-line activities. In particular, through the years, social motivators, such as prestige, rarity, self-identification, competition, or other non-monetary utility played major roles in the interactions between players in these types of games.

Cashing-in loot box mechanics instead allow gamers to buy and sell loot box items, for real-currency/cash or via items of monetary worth. In many jurisdictions, these mechanics require a license to be implemented (under UK law for instance). There are several unlicensed third-party websites that guarantee real-world value which is attached to the rewards obtained through the purchase of the loot boxes. The transaction may include the sale of single virtual items (in case the rewards could be extrapolated from the game in some way) or the sale of the entire account as a "larger package" which includes all the content purchased over time, up until the moment of sale, as well as any progression in the game (it may include weapons and armors, skins, in-game currency, etc.). These additional transactions present opportunities for tangible monetary gain and additional motivations for gaming. Additionally, the cashing-in loot box mechanics in these systems push loot boxes a little bit further into the "gambling" category. It must be clearly stated that, as the black market dynamics of gaming become more relevant (in certain, specific games in particular), the distinction between the two types of loot box mechanics tend to blur. In fact, if for instance rewards which are designed for games in closed-loop loot box mechanics are sold on third party websites, the "black market effectively presents the opportunity to 'cash-in' virtual goods".[10] (Garrelts, 2010)

3.2. Game as a service

In recent years, the video game industry developed a new model of business which is referred to as games as a service (or "GaaS") whose aim is to provide a video game or game content with a continuing revenue stream. This model was created in order to allow video games to monetize after their sale, or after their download in the case of free-to-play models. The software belonging to this category is typically provided with a continuous stream of new, purchasable content which can be monetized in order to encourage customers to continue

[10] Garrelts, N. (2010). Full reference in Chapter 7 ("References").

buying the additional new features provided. The concept of GaaS was initially developed when the first massively multiplayer online games (or “MMOs”), such as World of Warcraft, were introduced. In these types of games, the players would pay a monthly subscription and ensure the developers with a continuous revenue who, in turn, created new contents frequently.

In addition to game subscription, GaaS were also used for game subscription services such as “Xbox Game Pass” which enables the player to access a large variety of games with no limitations upon download and for Cloud gaming with services such as PlayStation Now, Stadia and GameFly which allows players to play games which are run on remote consoles using the internet connection. Last but not least, GaaS were introduced in the so-called “season passes” which provide one or more downloadable content and updates over the course of the game (during a predetermined period of time called “season”). Players pay the season pass in advance in order to be able to access this additional content at a later stage. In this particular case, the game can still be played by users who did not purchase the pass although they are not granted the additional benefits/contents of the pass. As a consequence, an imbalance within the game is created, especially if the software is heavily PVP (player versus player) oriented.

The reason that led to the implementation of GaaS mechanics is mainly related to money. As a matter of fact, through GaaS, developers are able to capture more revenue than with a single release. However, the objective of this model was to eliminate any legal issues related to software licenses. In fact, in the gaming industry, it is difficult to determine whether a game should be considered a commodity or a service and are protected through End-user license agreements (EULA)s in order to limit post-sale activities. In this sense, GaaS may reduce the probability of unauthorized copies of games and, in certain hosted gaming servers, also the need to install the software in players' computers and consoles.

3.3. Loot boxes and opaque selling

Using an appropriate terminology, loot boxes can be further classified as an example of “opaque selling”. Opaque selling is, in fact, a particular selling methodology that extracts value from certain selected items or features which are not known to the buyer until after the purchase is successfully completed. This selling strategy is particularly popular in the travel industry. In this particular instance, the entire package is sold though websites and the specific details of the package/product (which comprises hotel rental, transportation tickets, car rentals and so on) remains unspecified to the customers until the package is actually purchased. This strategy has great potential in the video game industry since it delivers products that otherwise would remain unsold (such as old and out-dated character skins or below average armor and equipment). This is even more so if we analyze a mechanism that has developed in recent years, regarding micro-transactions; tailor-made prices and products where algorithms are instructed to compute and generate customized prices and packages depending on the spending habits of the customer. In fact, while a video game distributing company would ideally pursue the highest possible margin and, in turn, charge the maximum price a customer is willing to pay, such price is not actually known to the seller who has to opt for alternative methods to identify such value. As described in Kazushige Nojima's paper of 2011 on the Japanese business model applied to video games, three elements of marketing are identified: the hook, the retention (which means keeping the players interested or ensuring they return to the game regularly) and "motivation of players to pay" for features added to the game. There are additional studies, such as the one of King et al. of 2019, which aim at identifying whether there are patents specifically focused on encouraging players to purchase micro-transactions. As mentioned, through the use of artificial intelligence, software is "bent" to more efficient packaging creation over time. These studies found that, in many cases, through the tracking of players' in-game purchases, key demographic information (including time played, previous spending, gender, ethnicity, age, etc.) is collected and used to market loot boxes. Additionally, the information gathered is used to target loot box purchasers and induce new and continued purchases by providing customised packages Further clarifications of this matter reveal that some of these patents are used by game developers with no regard or concern for copyright and intellectual property materials. In many instance gamers may not initially want to purchase loot boxes or additional in-game features but are led to believe that such products are needed in order to have a complete experience of the game.

Recent studies[11], have revealed that the condition that makes opaque selling attractive to firms is that this strategy enables gaming companies to capture a larger portion of the market. In fact, while selling regular products only attracts customers which are either on one side or the other of the spectrum of available products, opaque selling guarantees the elimination of the differentiation and expands market coverage. However, due to the cannibalization effect, the total profit will decrease. Even though this strategy may increase the overall performance of

[11] Such as the one of Marketing Professor Scott Fay and the one of Assistant Professor of Information Systems Yabing Jiang

the firm, it might also lead to self-cannibalization due to customer preferences. If instead opaque selling is implemented in conjunction with regular selling strategies, margins increase considerably. In opaque selling customers' propensity to buy, their maximum sustainable expense and their expectations of the product play a major role in their perception of a good deal when buying an unknown item. The presence of all these different parameters leads to the conclusion that only some industries, especially those which serve relatively differentiated customer groups, may increase revenues by offering a mix of opaque selling and regular selling.

3.4. Predatory dynamics on loot boxes

Loot crates may appear to be just another form of transaction used to create a continuous stream of cash for video game companies and to keep consumers engaged with their products. However, there is a reason why it is currently the most lucrative option available for video game firms to produce revenues. The system utilises the same psychological principles and "hook loop" mechanics which have been described by psychologists as creating some of the most powerful addictive effects and which are the same as those exploited by the slot machine business. These mechanics are deliberately added to games to leverage on the psychological vulnerabilities of individuals for consumer retention and profit. Far worse is that the most popular game titles present some form of these mechanics in a way which is considered of medium to high risk of being predatory and thus likely to cause harm. In certain cases, some of the top-earning game publishers (e.g. Activision and Electronic Arts) have registered patents for micro-transaction systems that incentivize the player to spend money.[12] The recent expansion of micro-transactions in general (and of loot boxes in particular) in modern video games has generated some concerns as to the extent to which certain players (e.g., younger users) may be vulnerable to overspending or impulsive buying on these games. Loot box purchasers may indeed be vulnerable which means that they represent potential targets for the predatory dynamics and the aggressive marketing strategies described above. Loot boxes are indeed perceived by the majority of gamers as a means to enhance the gaming experience and to customize the character in the game. Additionally, many games in which loot boxes are present are "free to play" and only a very small portion of the players make in-game purchases. This leads to a crucial level of artificial balance between players' entertainment,

[12] Marr M. D., Kaplan K. S., Lewis N. T. U.S. Patent no. 9,789,406. 17 October 2017. Available at: https://www.google.com.au/patents/US9789406 (accessed 26 July 2020).

their well-being and the monetization of the product which developers need to create. For this reason, it is important to consider to what degree developers are able to leverage on human phycology (both in term of marketing its loot box mechanics and in terms of ethics) and in which way these predatory mechanics may hinder the players' health and well-being. The literature on the topic has deeply analysed the degree of comparability between loot boxes, other "roulette" style games such as Kinder's surprise eggs or collecting trading sports cards. At a first glance, the different "games" may look similar and it would be quite reasonable to accept the comparability between the two. However, this no longer applies if we consider that the latter benefits from some physical characteristics. These goods are tangible and the company producing them does so only for a limited amount of time before moving on to the next product. In fact, all these goods come in a finished form, meaning that there are only so many limited copies of the same product and this makes them scarce and, consequently, valuable. In turn, the goods gain more value if few copies of the same item are produced. Instead, virtual items carry no value at all and the terms of service of many companies specify that virtual goods are not actually "owned" by users and have no monetary value[13]. Additionally, the argument of similarity fails/falls through if we take into account the predatory monetisation techniques (the customization of loot boxes, based on the different players being considered, which was discussed earlier in this Chapter 3, is the classical example). Lastly, it is important to point out that online gaming is a multi-billion-dollar industry. Since the spending distribution on loot boxes in games which rely on this business model to finance their activities is highly skewed (with many gamers not purchasing loot boxes at all), game developers must find convincing ways to efficiently target the market and monetize in the best possible way in order to survive.

3.5. Loot boxes and gambling: a discussion

The most general characterization of gambling includes at least three elements: stake, chance, and prize which are the components necessary in order for an activity to be classified as a "gambling" activity: the gambler decides to place a stake given a chance, in order to win a prize which has greater value compared to the amount staked. If one of the elements of the gambling process in controversial, for instance if there is no stake, the activity being performed does not constitute a gambling activity and is dealt with accordingly, based on the changing elements. In order for an activity to be considered gambling, the prize must be

[13] See Activision 2019 Annual Report

equivalent to money or something of monetary value, and the probabilities associated to achieving that prize should be made known to the player.

This general definition of gambling is widely shared across many jurisdictions but the interpretation of exactly which activities should be considered gambling vary across different geographical locations. In the particular case of loot boxes, the question appears to be whether or not the convertibility of rewards into some form of monetary compensation is allowed. According to different studies, cashing-in mechanics, which are consistent with the convertibility concept, are believed to be gambling activities in most cases and are thus considered the most harmful. For instance, loot boxes are considered a form of gambling in Belgium, Japan, the Netherlands and several other countries. Some jurisdictions have attempted to regulate the implementation of loot boxes with limited success. This is the case of China, where their commercialization is illegal and their disclosure is mandatory. These attempts to contain the deployment of the loot box mechanism has had limited success with developers who quickly responded by slightly re-writing their loot crate mechanics in order to remain just outside the scope of the law: leveraging on legislation gaps, loot crates were simply offered free of charge through virtual currency purchases.

Several international rating agencies and gambling boards maintain that loot crates do not fall within the definition of gambling activities because the items do not hold real-world value. As a matter of fact, loot boxes are not classified as gambling activities in North America, the U.K. and Australia, presumably because most games do not allow the re-sale of loot box rewards (although, as explained earlier, this occurs anyway through third party websites), and because the link between their use and the consequent harm incurred has not yet been well-established. However, closed-loop mechanics, while not associated with cash gains, may still be considered an abuse, unjust or harmful especially considering the fact that where micro-transactions for loot crates (and the consequent trading of loot crate items between players) are permitted, grey markets and middle-men who exploit the operation of these systems, surface.

Recently the awareness that loot boxes have been associated to gambling and that companies seem to be leveraging on young players (namely children and young adults) has increased greatly and both authorities and the public opinion insist that clarity urgently be made on the subject matter at stake, namely whether the purchase of loot boxes should or should not be considered gambling activities. Additionally, recent studies proved that loot boxes appear to be prevalent in video games that are deemed suitable for children by the PEGI[14], especially on mobile platforms.[15] The most evident example is represented by the game Crash™ Team Racing Nitro-Fueled ("CTRNF") which implemented the loot box system one month after its release, despite the developer's assurance, at the game's launch (which took place on June 20, 2019), that no such mechanic would be involved in the game. CTRNF is a kart racing game, developed by Beenox and published by Activision, which is officially the remastered edition of the very famous Crash Team Racing, originally developed by Naughty Dog for the PlayStation console in 1999. Upon its release, CTRNF seemed to maintain the same "feel" of the original version both in terms of gaming experience and reliability, with major improvements owing to the technological advancements implemented in the game and the capabilities of modern gaming consoles. However, once people started playing the game regularly, they realized that in order to exploit the real possibilities that the game encompassed, too much grinding[16] was needed to acquire "Wumpa coins" (the in-game currency). The only thing that could explain such throttled progression in the game was that additional mechanics were involved. In fact, about one month after its release, "Pit Stop" was introduced as a major, novel feature of CTRNF. The innovation was announced by Activision itself through an official informational blog post of July 30, 2019 on the Activision website, basically one month after the initial release of the game. Besides the introduction of carts, skins and other cosmetic elements to the game, players now had the possibility to spend real money on Wumpa coins which made the progression in the game more steady. The Pit Stop could also be used to purchase featured items which shifted every 24 hours and to which a certain degree of rarity was attached.

There have also been conjectures that the purchase and use of loot boxes may offer a pathway to gambling and/or are designed to resemble slot machines and feature the same basic element of randomness of in-game reward. At the same time, there are concerns that such types of in-game purchasing systems involving randomness may contribute to excessive playing behaviors and psychological overinvestment in video games in general. However, micro-transactions in video games have not generally been subject to the same regulatory controls and player protection measures as gambling. There is increasing interest and discussion at an

[14] The Pan-European Game Information (PEGI) is a European video game content rating system established to help European consumers make informed decisions when buying video games or apps through the use of age recommendations and content descriptors. (source: Wikipedia)
[15] Short Report by Zendle, D., Meyer, R., Cairns, P., Waters, S., and Ballou, N. (2020). See References.
[16] In video games, "grinding" refers to the iteration of repetitive tasks, usually for a gameplay advantage or loot but in some cases for purely aesthetic or cosmetic benefits.

international level on the potential need for regulatory scrutiny of these products, including in relation to the growing popularity of gambling activities on eSports and games of chance using in-game assets and currencies such as “skins”. The loot box reward system actually only provides the illusion of constant winning by exploiting the natural excitement of opening packages. However, the extreme impact of the psychological frustration resulting from not finding the expected or desired “prize” in one of those packages destroys the main aim that a game should have, that is to say to create joy and relieve stress.

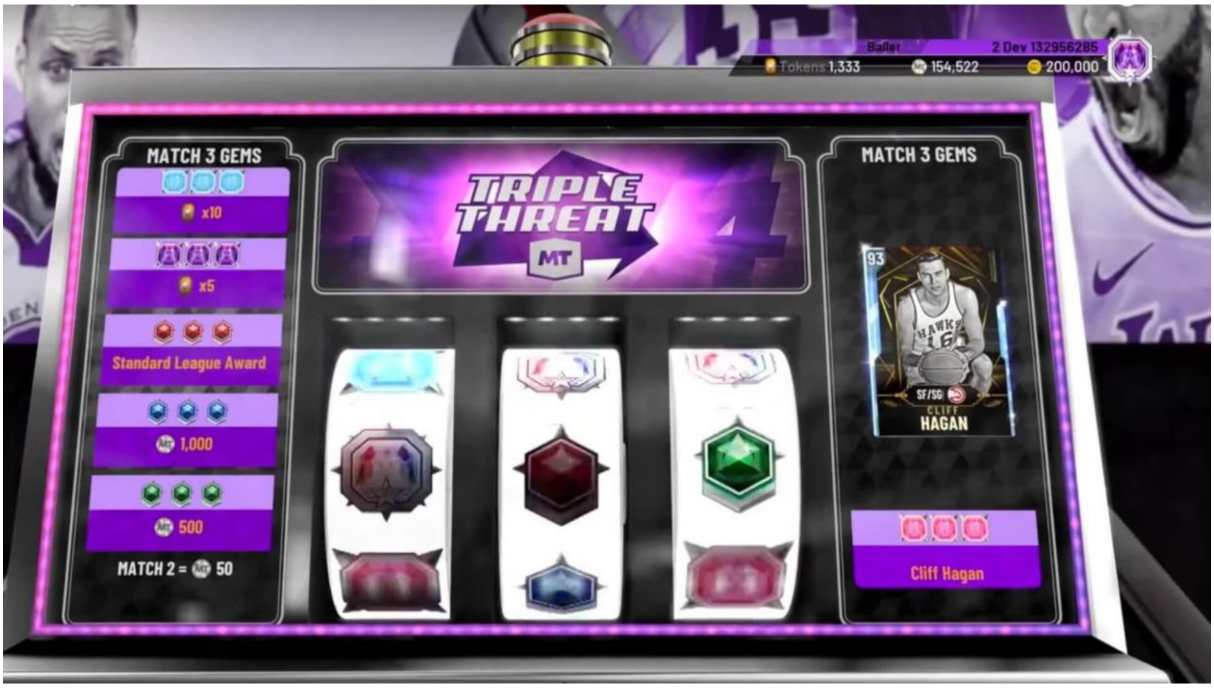


Figure 3 The slot machine showed in a commercial of an NBA video game

The frustration is even more evident if we refer to progression-tied loot boxes (i.e. loot boxes whose rewards are crucial in order to further advance in to the game or to have a balanced experience in multiplayer scenarios). Figure 3 above illustrates the frame of a famous commercial trailer relative to an NBA video game. The commercial was promptly removed from the internet by Electronic Arts after seeing the wave of indignation that it had raised. The most predominant argument that enflamed the discussion was the possibility to spend real-world money to progress into the game. Further indignation followed when the game was attributed a European classification of PEGI 3+, which means that the product had been inspected by the PEGI commission which had deemed it suitable for children of three years or more. Many people, especially influencers and youtubers, pointed out the controversy of the topic for which the PEGI could not provide an exhaustive answer in order not to violate its confidentiality obligations. Another game that raised many eyebrows was Electronic Arts’ Battlefront 2 which created loot crates that were extremely invasive in terms of in-game gameplay and mechanics. These loot boxes in fact did not merely modify the “aspect” (or the “skin”) and the feel of the game but rather created major discrepancies in the game balance system itself, thus increasing the sense of frustration of non-spenders.

The regulation of monetized games and virtual currencies has emerged as a pertinent issue given the global popularity of online gaming products and in-game spending among video gamers. A recent study by Daniel L. King (a researcher at the School of Psychology at the University of Adelaide, Australia) revealed that the majority of video games contain what he refers to as “gambling-like” features. In the criteria used to identify such features, King evaluates the monetary stake which, in return, might guarantee a prize and a given chance of earning it. However, the future outcome is unknown and in order to avoid incurring in losses, players should choose not to engage in loot boxes purchasing activities via micro-transactions at all. As mentioned, gambling can exceptionally be associated with positive outcomes, provided the actor is fully aware and conscious of the type of activity he is performing. Because the line between the two activities, gambling and gaming, is often blurry, a useful distinction between them could be useful. Broadly speaking, the main difference between the two terms is that in the case of gambling, the outcome is achieved by chance, not skill, whereas for gaming, the opposite should be true. This is the reason why the purchase of loot boxes and gambling are probably better characterized in terms of the harm they cause to the user and how their different features may influence the harm they cause (a large or important prize carries a higher stake and, therefore, a more significant risk).

Most studies agree that the purchase of loot boxes might be associated with disordered gambling although there is no clear evidence that loot boxes are directly linked to problem gambling. These studies, in fact, reveal only associative relationships between the two. On the one hand, we have gamers who do not purchase loot boxes, while on the other, gamers who do purchase them. It has been found that this latter category plays video games more often and for longer periods of time and are generally more inclined to online gambling. Games where there is the possibility to cash-out loot boxes or that show what are generally called “near-misses” (i.e., near misses occur when the rewards of the loot box was very close to the desired one), have also been proven to slightly increase the relationship between loot box spending and problem gambling. It is therefore necessary to examine the **direct** effect and consequences of the loot box monetization system and to understand the main reasons that lead players to turn to micro-transactions and the purchase loot boxes in the first place. It is clear that an heterogenous approach is needed in order to properly address the concerns on loot box and their potential harm.

4. Methods

The aim of this research study is to identify to which degree each parameter that was deemed relevant for the purchase of loot boxes influenced peoples' purchase expenditures. In particular, a survey was submitted through different means (details are provided later on in this Chapter 4) directly to video gamers in order to understand their perception on the topic of loot boxes and micro-transactions and to gather material to be analysed together with the data and information collected through previous studies.

4.1. Target sample

In order to select a specific but wide spectrum of respondents, a number of techniques was used. Respondents were gathered through various social media and through relevant channels that are generally used to share information. In particular, a survey in the form of a questionnaire was submitted through different information sharing vehicles such as Survey Circle (a website used to share surveys whereby, upon completion and submission of a complete survey, the respondent acquires "points" to be used to enhance and promote his or her own survey), Facebook groups (details will be addressed later on in this paper), word of mouth, Survey Monkey etc.. Prior to initiating the survey, which was completely anonymous, an agree/disagree policy was submitted to reassure participants of the non-disclosure of their personal data and to gather their consensus to use the information provided albeit for academic purposes only.[17] The questionnaire was submitted to 22 different gaming oriented (and video games related) Facebook groups in order to increase the percentage of subjects coming from the most active sector of the gaming community (i.e. those gamers who are so actively involved in video games that they decide to share content with, and follow posts of, other players around the world). These groups were selected randomly, with no emphasis on any subject or game in particular (that is to say, no importance was attributed to groups focused only on a specific gaming console or video game) in order to gather the most unbiased responses possible. The entire list of the Facebook groups involved in the survey is provided in Appendix A. Additionally, the survey was submitted to 14 subreddits which are listed in Appendix B. These posts had the main aim of capturing less assiduous players or those that are generally referred to as "casual gamers".

[17] The drawback of the anonymous nature of the survey is that by not gathering IP addresses or other forms of identification of the participants, it is impossible to verify if participants took the survey more than once.

The survey was introduced through a brief explanation of the purpose of the research study, a description of the subjects eligible to participate to the survey and expressly stated that the data collected would be used for academic purposes only. Additionally, it was also clearly stated that the completion and submission of the survey represented the acceptance and agreement to the aforementioned conditions. The respondents were deemed eligible to participate to the survey if they had played video games in the 12 months preceding the survey or if they had ever purchased loot boxes in the past. All respondents were included in the research study so as to also consider those players who do not play video games assiduously but nevertheless have an opinion on the matter at stake.

The survey was delivered in English with the possibility to translate it into different languages depending on the web browser used. The survey was posted mainly on English-language websites used by international gaming communities or on exchange survey websites where participants complete and submit questionnaires in exchange for others completing and submitting the surveys they too published on the same website. No particular incentive to participate was given to the respondents, who therefore completed the survey completely for free. Only those who used Survey Circle were rewarded “in app currency” which they, in turn, used to promote their own surveys by using a redeem code which was made available at the end of the questionnaire.

The survey described above was deemed necessary as the data available is incomplete and, in some instances, lacks the magnitude that is essential in order to be characterized in a quantitative study. The decision to collect the relevant data through an online questionnaire was devised after having examined both the target population and the topic. The online submission is, in fact, the most efficient means of reaching the largest number of people (from the specific category of individuals) in the shortest possible time. Additionally, due to the impediments of the pandemic scenario of Covid-19 and the relative, mandatory use of face masks and gloves, hand-to-hand delivery and completion would have been a far less cost-efficient method for reaching our audience. Furthermore, studies have proved that online methodologies of questionnaire submission tend to increase the veracity of responses.

A total of 243 responses were recorded (which constitute our “sample”), of which 140 were fully completed since not all questions were mandatory (due to the fact that some questions were related to the purchase of a loot box and not all respondents had purchased loot boxes).[18]

[18] The data used for the analysis herein can be requested directly to the thesis author.

For the sake of completeness, it should be noted that there is no way to verify the number of people that opened the survey and decided not to participate. Furthermore, a filter question was not provided.

The survey included items which recorded the demographic characteristics of the respondents (age, sex, occupation, income etc.), their general habits in the use of video games (which included satisfaction, average hours spent playing etc.) and broad as well as detailed consideration on loot boxes. Additionally, a distinction was made between those who had purchased loot boxes and those who had not. The former were asked additional questions regarding the subjective importance attributed to loot boxes, their degree of satisfaction, the reason they purchased loot boxes in the past etc. The entire survey is set forth in Appendix C.

The survey was conceived as simple and as short as possible in order to avoid people closing the text before submission due to stress or potential fatigue. For this reason, only 21 questions were envisaged, also with the aim to avoid excessive repetition and the aforementioned escaping behaviour.

Respondents who indicated that they had indeed purchased loot boxes were asked 8 additional questions regarding various aspects of the purchase (their monthly average spending on loot crates and their degree of satisfaction for instance) and their emotional commitment when playing. These respondents are among the most important since, through their answers, we can build a model to predict players' future spending and degree of satisfaction.

4.2. Data presentation and demographics

For the sake of this research study each of the items addressed will be referred to using a specific name in order to avoid confusion in the data management process. Each of the questions in the survey was made to correspond to a certain "Variable". Some examples are:

- "I identify myself as..." corresponds to the Variable "Gender";
- "What's your age range?" corresponds to the Variable "Age".

The complete list of the different Variables is provided in Appendix D.

Variables in the survey are of two types: "Qualitative" and "Quantitative". Qualitative Variables answers are characterized by objects while Quantitative Variables answers are characterized by floats and integers. The **objects** are either a Boolean (that is to say a True/False or Yes/No answer) or can have another form (which may be a single word or the description of a particular phenomenon such as “Continent: Europe”, “ObtainLB: In-game currency only”). Instead, **floats and integers** are simply numbers (which, in this case, range from 1 to 5 with the exception of the values attributed to “HoursAVG” which range from 0.75 to 9). Together, objects and floats and integers will be referred to as “Entries”.

Since the survey included Qualitative Variables, which are characterized by objects and are therefore not suitable for mathematical purposes, additional elaboration was necessary after the gathering phase in order to be allow the processing of the data available. As can easily be understood, data frames which comprise both numerical random variables and objects (that is to say parameters which cannot be associated to numbers) are not easily manageable unless we “translate” the latter into a “workable” format. To do so, we used a coding instrument called “encoding”.[19]

We created a function that replaced each single object of the Qualitative Variables of the sample with a number ranging from 0 to the extent of the Variables. In other words, if a given Qualitative Variable comprised 5 different objects, the function called the first object 0, the second 1, the third 2 etc. up to the last object, without repetition. In doing so however, we introduced a “distance error”, which means that the object number 0 will be considered by the software 5 steps “to the left” compared to the object number 5 thus introducing an artificial distance between objects 0 and 5 which is greater than the distance between the object number 0 and the object number 1. This particular issue will not be addressed in the study, however we suggest to correct such “distance error” in similar, subsequent studies by either not considering objects altogether or by translating objects into integers or floats within the survey so that the artificial distance between two elements is immediately clear to respondents.

The first part of the survey was characterized by the demographic section which included questions about: gender (male, female, other), age (within a range), continent (Asia, Europe, etc.), occupation, income (within a range) and number of people in the household of the respondent. The following graphs, together with a brief description on each of them, will describe the demographics of our sample so that the analysis will be easier to read.

[19] In computing, data storage and data transmission, character encoding is used to represent a repertoire of characters through some kind of encoding system. Depending on the encoding system and the context, a code can be rewritten using patterns, natural numbers etc.

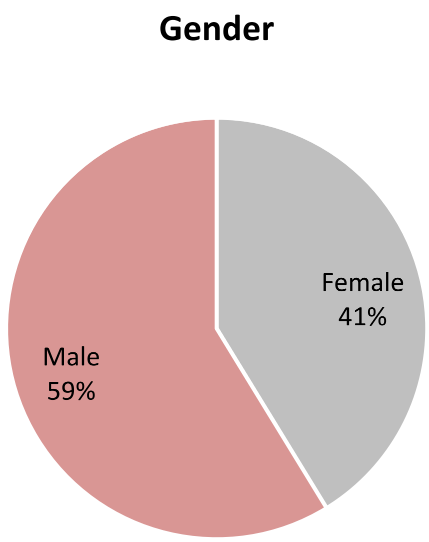


Figure 4 A pie chart of the gender distribution of the sample

It is clear from the graph above that there is a slight predominance of male respondents in the sample. For this reason, it was deemed relevant to run a Variance-Covariance matrix that included all the variables in our analysis, together with a Chi-Square test to check whether the skewed presence of individuals of one sex may have led to a biased result (more about this later). As we will see, Gender is quite an important Variable and has a high correlation with the other Variables within the survey.

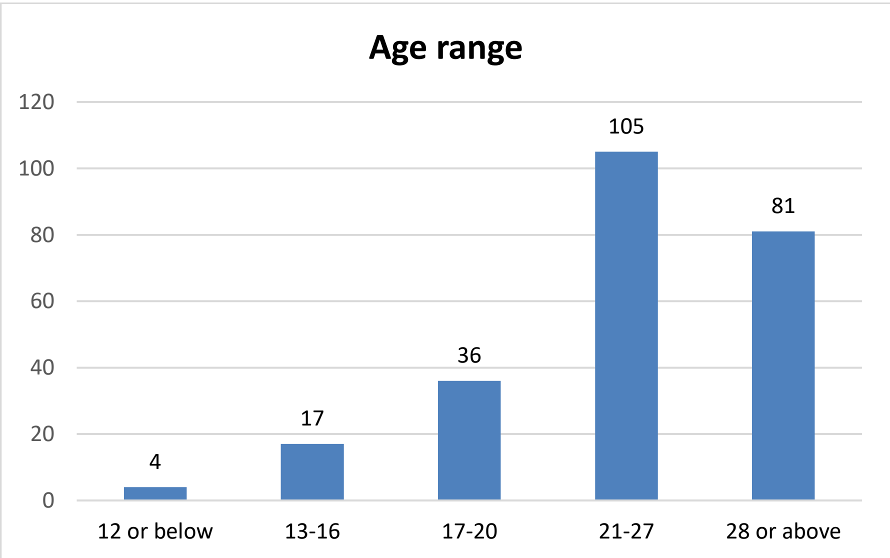


Figure 5 A bar chart that represents the age range of the respondents

The graph above illustrates the Variable related to the age of the respondents. As evidenced by the graph, the sample had quite a skewed selection of respondents in terms of their age being, in 43% of the cases, comprised between 21 and 27. Since the sample should not represent only a small portion of gamers but the entire population itself, we deemed it relevant to compare the data we found through our research study with the data available online regarding the average age of people playing video games. What we found was that 35% of gamers are between 21 and 35 years old.[20] This data is consistent with the findings of our sample and within a statistically relevant confidence interval. If fact after running a Chi-Square test, we found that the p-value is lower than 0.05, hence our sample can be considered statistically significant and thus, representative of the entire population.

Without going into the details of each single Variable related to the demographics, we will leverage on the Variance-Covariance matrix mentioned earlier to understand whether the Variables are somewhat connected and relevant with respect to those related to the purchase and usage of loot boxes. Such matrix is provided in Appendix E. Additionally, in order to exclude any correlation between the Variables that do not strongly affect one another, we filtered the matrix so that only covariances with a value above 0.4 were displayed. Additionally, since the research study specifically addresses the correlation between loot boxes and gamers, we also excluded the covariance among demographic Variables which were considered only in relation to gaming-related variables.

4.3. Data Analysis

In our analysis we used a machine learning method based on linear regression, which is a linear approach to modeling that approximates the variables given a defined data set. In this research study we will refer to both simple linear regression, in case only one explanatory variable is considered, and multiple linear regression for more than one explanatory variable. The linear regression will be used to model our data via a linear predictor function. Linear regression is the simplest method to be used when estimating data. It is one of the first types of regression analyses to be studied rigorously and used extensively in practical applications. This model generally includes unknown parameters which are easier to fit compared to models which are non-linear. A linear regression is the perfect test to determine the extent to which there is a linear relationship between a dependent variable and one or more independent variables. There are two types of linear regression, **simple linear regression** and **multiple linear regression**. In simple linear regression a single independent variable is used to predict the value of a dependent variable. In multiple linear regression two or more independent variables are used to predict the value of a dependent variable. The difference

[20] Data gathered by statista.com

between the two is the number of independent variables. In both cases there is only a single dependent variable. The dependent variable must be measured on a continuous stream while the independent variable(s) can be measured on either a categorical or a continuous measurement scale.

The linear regression that we decided to perform is the multiple linear regression, with the aim to measure the degree of linear relationship between the variables under analysis. To be more precise, the scope of the linear regression model we implemented was to predict and evaluate the value of the dependent variable based on its relation with those of the independent ones.

Generally speaking, a multiple linear regression model is chosen to verify the same assumptions that are commonly attached to a simple linear regression, namely:

- homoscedasticity: which corresponds to the homogeneity of variance (i.e. the error of prediction does not change significantly across all our independent variables);
- independence of observations: all the observations of the sample must be collected in such a way that no hidden relationship between the variables are present in the dataset (in our case there are, in fact, independent variables which have a considerable degree of correlation with one another (with an $R^2 >\sim 0.5$)[21]. Therefore, before performing the regression, we dropped one of them at a time in order to not have biases);
- normality: the data should follow a normal distribution; and
- linearity: the line that best fits the data points must be a straight line.

The general formula of a multiple linear regression is represented by:

$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \varepsilon$

where:

y = represents the predicted value of the dependent variable;
β_0 = represents the y-intercept;
$\beta_1 X_1$ = represents the regression coefficient (β_1) of the first independent variable (X_1);

> *This value can be translated as the increasing effect generated from the variation of the value of the independent variable X_1 on the predicted value y.*

[21] In statistics the R^2 is the coefficient of determination (i.e. the proportion of the variance in the dependent variable that is predictable from the independent variables). From Wikipedia.

… = refers solely to the fact that the procedure must be iterated for as many independent variables as those in the model;

$\beta_n X_n$ = is the regression coefficient of the last independent variable;

ε = represents the model error.

> *The statistical error (aka disturbance) is the amount by which an observation differs from its expected value, the latter being based on the whole population from which the statistical unit was chosen randomly.*[22]

Additionally, because our research study entailed the observation and analysis of more than one outcome variable at a time, we also used **multivariate analysis** (MVA), a valuable type of analysis that is normally performed to address situations in which multiple measurements are made on each experimental unit where the relations among these measurements and their structures are essential.

The analysis of the data comprised a selection of scripts run in a defined programming environment created for the purpose of this paper. The software used in the studies are Anaconda3, an open-source data science toolkit used to perform Python/R data science and machine learning, and Jupyter Notebook, another open-source software easily available through the web. The application allows users to create and share documents which contain codes, equations and also and narrative text. Jupyter Notebook was used to clean and transform data, to simulate and predict expected values and to train a machine using codes.

The first step necessary is to convert into a workable format the data collected directly from the excel sheet used for the online survey. As explained in the previous Section 4.2, each question was transformed into a Variable following the pattern provided in Appendix D. Furthermore, the Variables were re-elaborated so that their order matched the order of the questions. In an attempt to minimize the errors associated with the distance between items, a number of selected Variables was transformed into integers and floats (namely "Hours", "Buy-YN", "CharacterizeLB", "HappyLB" and "ProgressLB"), were manually modified into parameters.

Once the data was ready, we used Python to elaborate statistical information. As a first step, we created the environment. The packages used in the paper are the following:

- NumPy;

[22] Source: Wikipedia.

- Mathplotlib;
- Scikit-learn;
- Xrld;
- Pandas;
- IPython; and
- Pillow.

These packages or "libraries" were all necessary to some extent in the elaboration, characterization and implementation of data that was collected. To be precise, each of them played a specific role in the analysis performed. Below is a brief description of each of these libraries and their role in our research study:

- first on our list is "**NumPy**", a library for the Python programming language, which supports a large variety of operations. Among other functions, NumPy can be used to organize and deal with multi-dimensional arrays and matrices (which is exactly the use we made of it in our research study) and, in some cases, it can facilitate operations with "high-level" mathematical functions while operating with such arrays;
- **Matplotlib** is an essential tool to be used in combination with NumPy when plotting is needed. It deals with numerical mathematics extensions and, in our case, is used to export graphs and materials;
- **Scikit-learn** is the most important tool in our selection of libraries. It is an open-source, machine learning library which can perform regression and clustering algorithms and is meant to work together with NumPy. In our case, we used its regression algorithms to predict values and attributes associated to the Variables of our study;
- **xlrd** is one of the most widely used libraries and it is able to extract data from Microsoft Excel spreadsheet files. The library can extract any form of Excel data from a spreadsheet including ".xls" and ".xlsx" from version 2.0 onwards. It can be used in any platform which is written in pure Python (as is the case of Jupyter Notebook) and it is an essential to our research study;
- **pandas** is an open-source Python library which is used for data manipulation and analysis. It is a very basic tool and provides the necessary instruments to perform basic and advanced operations using data structures and numerical tables and time series;
- **IPython** stands for "Interactive Python". It is mostly used for interactive computing in multiple programming languages. IPython encompasses a variety of tools that are needed to create browser notebook interfaces (which is the main purpose for which we

used this interactive shell) it can support interactive data visualization and is essential to create a flexible and easy to manage environment; and

- **Pillow** is an open-source library which provides support for opening, manipulating, and saving many different image file formats. In this research study it was used together with Mathplotlib to export images and graphs.

We imported all the data from the excel file and created a data frame. We then printed the head of the data frame and extrapolated the necessary the relevant information from the data. In doing so, we were able to easily identify and distinguish objects from floats and integers, and to count the number of responses for each Variable.

The subsequent step was to understand the variation among Quantitative Variables. For this reason, we ran a describe function which printed out the number of items as well as the mean and the standard deviation for each Variable. The function also identified the values corresponding to the minimum, 25%, 50%, 75% and the maximum of the values which, in this case, were not relevant since they coincided with the actual floats and integers of the sample.

	HoursAVG	CosmLB	ProgLB	GambLB	Imp	Satisfy
count	243.000000	243.000000	243.000000	243.000000	162.000000	91.000000
mean	4.425926	3.378601	3.329218	3.670782	1.771605	2.505495
std	3.007681	1.377458	1.520742	1.313727	1.035246	1.186719
min	0.750000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	1.500000	2.000000	2.000000	3.000000	1.000000	1.000000
50%	3.000000	4.000000	4.000000	4.000000	1.000000	3.000000
75%	7.000000	5.000000	5.000000	5.000000	2.000000	3.000000
max	9.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Figure 6 The description of the Quantitative Variables of the sample

The table set forth in Figure 6 above illustrates the activities described in the preceding paragraph. It also provides a general idea of the distribution of the integers and floats through the description of mean and standard deviation. Conversely, the histogram below represents the distribution of the Quantitative Variables of the sample and their frequency.

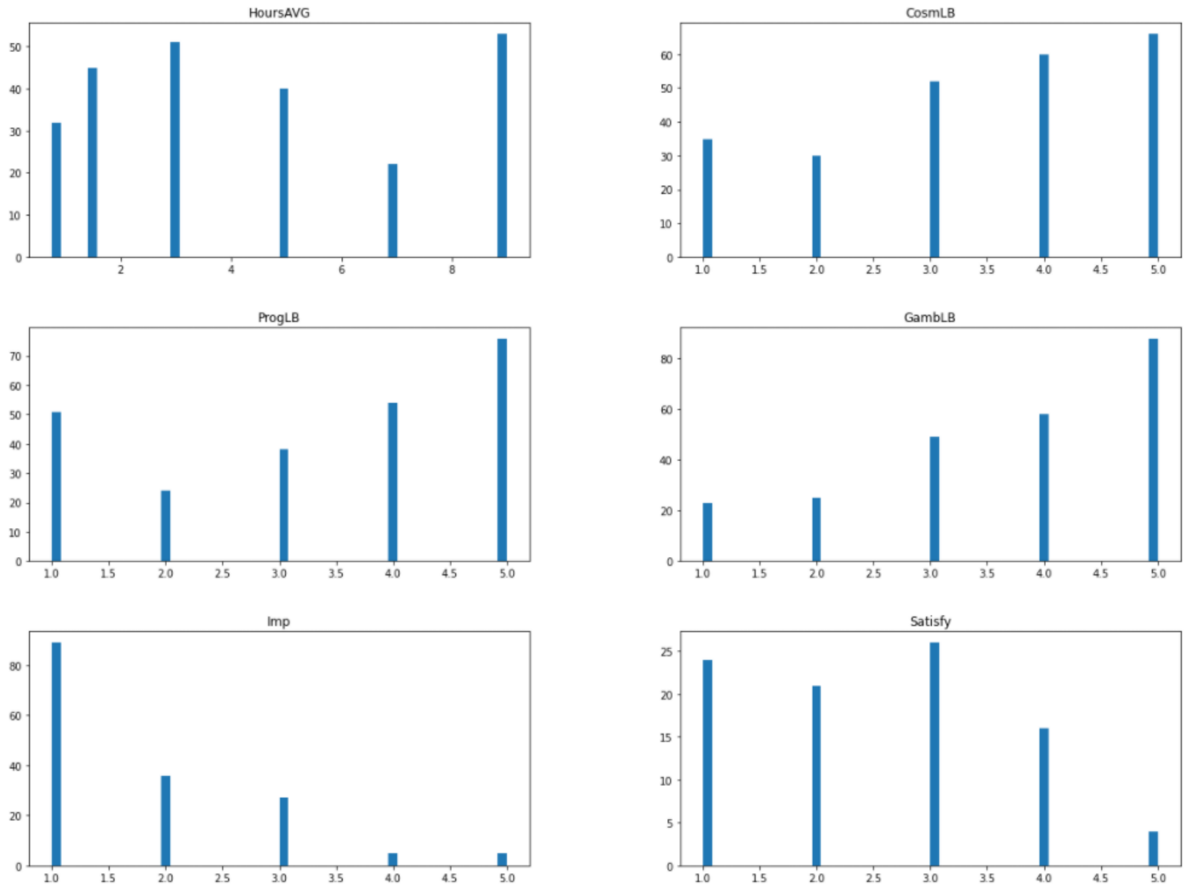


Figure 7 Histograms representing the relevant frequencies of the variables

What Figures 6 and 7 above show is that, except for HoursAVG, the other Variables have a standard deviation which is close to 1. This means that the distribution does not differ greatly from the mean of the sample which, in turn, favours the opportunity to make additional inferences with respect to our sample. To be more specific, we observed that, in the case of the Variables CosmLB and ProgLB, more than 60% of the data are within 1 standard deviation from the mean (3.4 ± 1.4 = [2, 4.8] and 3.3 ± 1.5 = [1.8, 4.8] respectively), while in the case of the Variables GambLB and Imp, we experience a very different distribution: the former is highly skewed to the right (suggesting that our sample has converging ideas regarding the classification of loot boxes as gambling activities) while the latter is highly skewed to the left (which instead suggests that loot boxes are not deemed important for the majority of respondents). As concerns the degree of satisfaction (the "Satisfy" Variable), the situation is quite different: because its mean is slightly skewed to the left and its standard deviation is close to 1, the evident deduction is that the sample is almost indifferent with regard to the degree of satisfaction when purchasing loot boxes. However, an immediately apparent and striking feature is the fact that the number of observations of the said Variable is almost 1/3 of the total number of surveys submitted. For this reason, additional testing was required. In particular, we ran a Chi-square test of independence through which we verified that, due to the many degrees of freedom of the Variable Satisfy compared to the others of the list, it cannot be considered significant since its p-value is greater than 0.05.

The next step of our modelling involved the encoding the objects. First we created a list containing our Qualitative Variables; then we defined a function that was able to transform each object in a number.

We created a function named “encoder” that uses an encoded data frame and a list of variables.

For each entry of the list we repeated the following process:

(i) we created a set of the objects (which managed to extrapolate each object only once, without repetition);
(ii) we transformed the set into a list and created a number of labels equal to the length of the list;
(iii) we then created a dictionary and a loop that, together, were able to match each object to its corresponding label;
(iv) finally we replaced the label with the corresponding item on the list.

At the end of the process, we had expressed each object as a number ranging from 0 to the length of the set of the entries.

It was now possible to create the Variance-Covariance matrix necessary to convey the information regarding the relationship between each pair of variables in terms of covariance.[23] This Variance-Covariance matrix had dimensions 21 x 21 and presented most of the values with little to no covariance. However by implementing a filtering process, thus selecting only those values which are greater than 0.4, we can omit the unnecessary information and focus solely on those parameters that are more relevant.

Incidentally, the Variance-Covariance matrix experiment was intended to serve a greater purpose. As a matter of fact, we were able to feed the relevant information gathered through the analysis of said matrix into a model capable of estimating certain variables of the sample.

[23] The covariance indicates the joint variability of two random variables. In this particular case, the covariance is normalized between -1 and 1. Values close to 0 indicate no meaningful relation between the two variables while values close to 1 indicate a stronger relation.

We decided to model a regression fit and to train a machine to predict the outcome of a given variable through the following process:

1. we defined a list of targets that the model would try to predict;
2. we set a test percentage of the total sample that the model would try to guess and, in doing so, we defined the shape of the train (i.e. how many outcomes does the model see before trying to guess the variable);
3. we passed the model through all the variables with the exception of the targets using a linear regression fit;
4. then the model tried to predict the outcome of the targets and checked whether its guessing was correct (iterating a process called "training");
5. after a number of repetitions (equal to the total number of observations minus the test percentage) the machine no longer corrected itself with the right amount and only provided estimates on the variables;
6. we checked whether the model was correct in its prediction on the test percentage and we counted the correct responses.

Below is an example of the outcome of the function of multi regression fit:

Target variable: ['Spent', 'Reasons']
Test set percentage: 0.15
Train X data shape: (206, 19)
Train y data shape: (206, 2)
Regression score: 0.5977412489808938
Regression Intercept: [0.76960044 6.91630772]
We have for Spent a result of 29 correct test set samples out of 37
We have for Reasons a result of 12 correct test set samples out of 37

This function allowed us to test whether the predictions of the software were correct and, at the same time, to vary the targets, the test percentage and the way the software was selecting the data from the data frame (the so called "seed").

As a first step, we checked which test percentage was most suitable to our case. We performed iterated tests on Relevant Variables (hereinafter "Relevant Variables") only, and in each of these tests, we changed the test percentage of our machine. The range of test percentages we considered was from 0.050 to 0.185 (running the test 15 times, each time increasing the percentage test by 0.015). The output of such tests were recorded and we compared the software's correct guesses on the following Variables:

- GambLB;
- Imp;
- Satisfy;

- Buy-YN;
- Spent;
- Reasons;
- LBInstead;
- CharacterizeLB;
- HappyLB;
- ProgressLB.

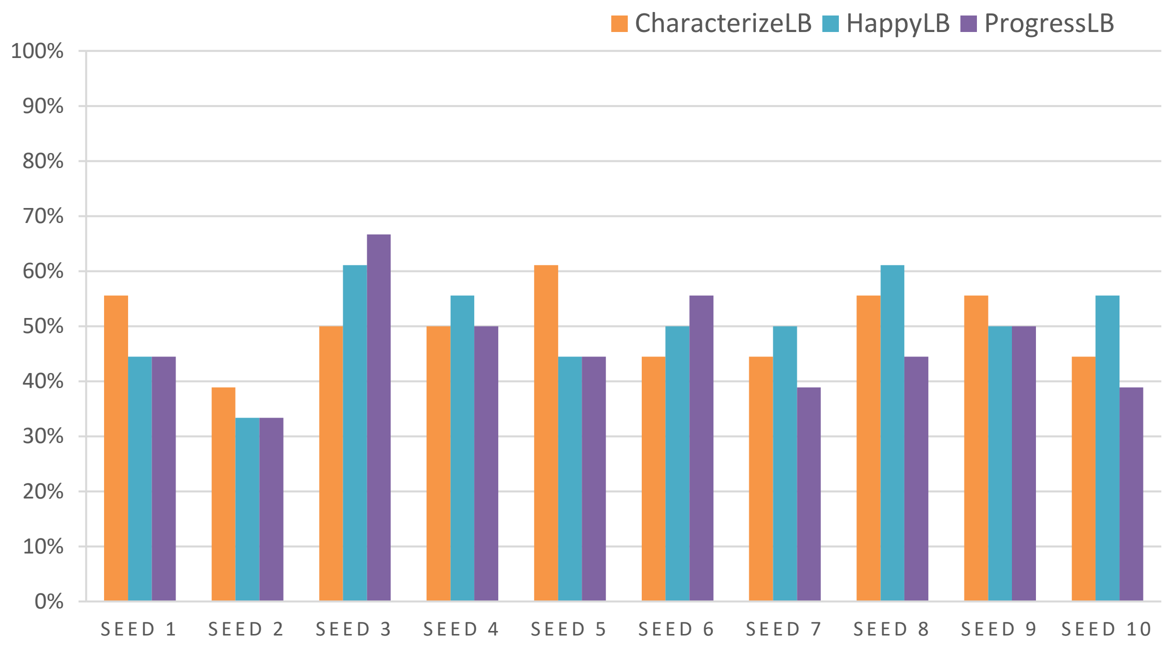
This intermediate step was necessary to determine the best test percentage. In fact, our test needed to be as broad as possible in order to assess whether the predictions of the software were reliable and, at the same time, leave the machine enough room to effectively learn how to predict outcomes.

The benchmark we used to select the appropriate test percentage was that at least 7 out of our 10 variables needed to be predicted correctly at least 65% of the time. The greatest test percentage that met such requirements was 0.070 (which corresponded to a test of 18 guesses).

Having defined the proper test percentage, we then needed to pass the "Relevant Variables" through the model. To test our model, we first checked each variable alone and then together with other variables. We found that three was the maximum acceptable number of variables that we could consider simultaneously and that the test percentages greatly influenced the prediction of the machine in the test phase. More than three variables, in fact, did not yield enough learning space for the machine and led to too many incorrect predictions. The same applies to test percentages (i.e. with a larger test percentage, the software had less material to adjust itself in the learning phase and was unable to yield correct results in the test phase).

Figure 8 Histograms representing the correct responses of the software with different test percentages

The ten histograms represented above illustrate what we explained earlier: different test percentages (considered at 0.05, 0.07 and 0.09) represented on the x-axis in different colours (blue, red and green) yield different degrees of correct answers. Moreover, the histograms show 3 different seeds for each variable in order to confirm the consistency of the findings.



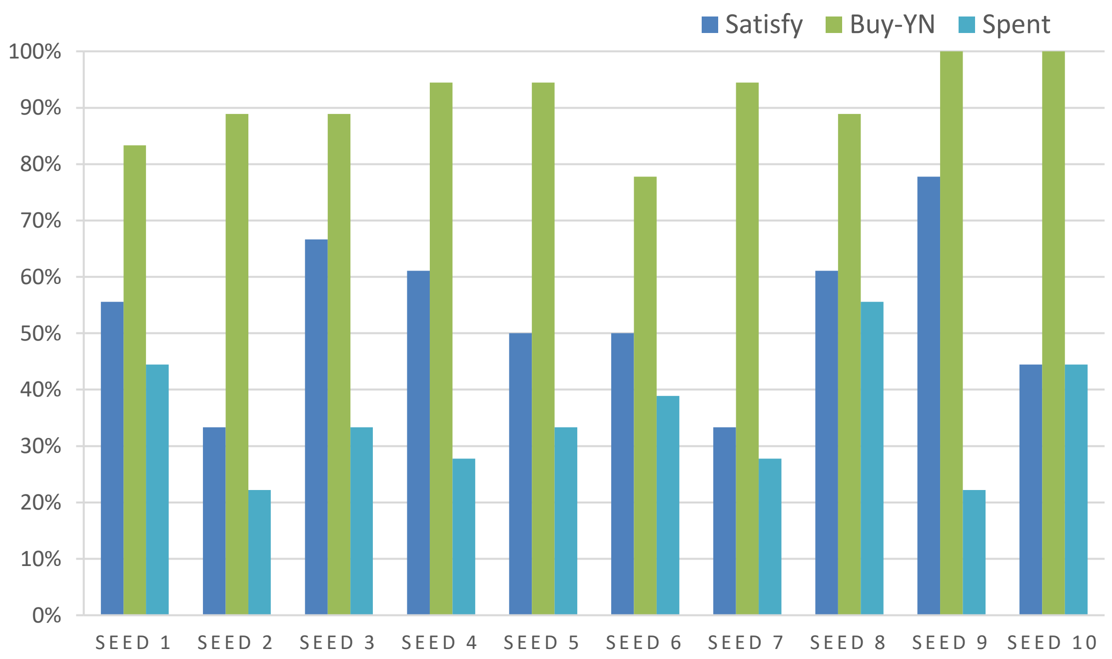


Figure 9 Bar charts representing the total correct guessing of three variables at once

The bar charts above show the model’s correct number of guesses as a percentage of its total guesses. As explained, the greater the covariance of these three variables with those not considered in the guessing tests, the more accurate the predictions of our software will be.

All the computations carried out in this phase of the model are clearly applicable only to the data frame under consideration. Therefore it is worth pointing out that a different data frame could lead to different results. To develop a more accurate analysis, it would be necessary to compare the results of this study with similar analyses produced with respondents with similar characteristics. Additionally, a larger data frame could lead to more precise results and more accurate variable estimates.

5. Results

The research study allowed us to estimate the Relevant Variables in the analysis with an accuracy greater than 70% if we provide a training sample equal to 93% of the entries (which, in our case, amounts to 225 elements). We are satisfied with the results given that the linear regression model we performed is one of the simplest regression fits that can be used and, with the a few tailored adjustments, the model can be improved greatly.[24]

Through the Variance-Covariance matrix we were able to distinguish the variables that proved to be relevant for the study from those that would not have benefited the analysis in any way. The variables that were ultimately selected for our analysis are those listed below:

- GambLB
- Imp;
- Satisfy;
- Buy-YN;
- Spent;
- Reasons;
- LBInstead;
- CharacterizeLB;
- HappyLB;
- ProgressLB.

We used these variables within the linear regression model with the aim of estimating each of them in different circumstances:

- by varying the seeds, we found that the results varied irrelevantly as the sample (using a test percentage of 0.07) was large enough for most of the variables and the software had the necessary data to "learn";
- by varying the tests, we noticed that the percentage of correctly guessed variables decreased as the number of guesses compared to the entire sample increased. This is possibly due to the same reason described in the preceding point above;
- lastly, we tried to estimate more than one variable at the same time. This, in turn, led to different results depending on the correlation between all the variables that we were trying to guess compared to all the variables which we were not trying to guess.

Additional final considerations regarding the model must be made:

- some of the variables had great values of prediction due to the fact that were strictly related to a particular section of the questionnaire, namely the questions specifically

[24] More on this in Chapter 6 ("Conclusions").

addressed to loot box users. Buy-YN and LBInstead, for instance, were easy to predict due to the fact that people who answered in a certain way to one part of the survey gave a certain response to both questions; because the software had learned to recognize this correspondence, it was able to predict the variables with a precision close to 100%;[25]

- certain other variables with significantly lower percentages of correct guesses must be adjusted according to the magnitude of objects they contain. The Variable Reasons, for instance, included a greater number of objects than the other variables (in this specific case, 11 objects, also attributing a value of 0 to the non-responses). The model therefore needs a much larger sample to train the software to predict the results accurately. Instead, in the case of the Variable Spent, the brackets within which the spending ranges for the loot boxes are defined do not correspond precisely to the encoding carried out previously and which was obviously necessary to process the data. In this last case, the problem of the distance error (referred to in Section 4.2 herein) is therefore amplified. Both cases required a correction on the guessing rate. In fact, let's consider the following example:
    - Variabile: Reasons
    - Test percentage = 0.07 → 18 guesses
    - # of objects in Reasons = 11

  Without training, the software model would guess the Variable by random chance and would be correct once every 11 guesses ($1/11 \simeq 9.09\%$ of the times). If we multiply this value by 18, we obtain 1.636 which is the number of correct guesses (out of 18), predicted by the software model by random chance. Instead, our model, adequately trained, was able to correctly predict the results 4 times out of 18 (about 2.44 times better than random chance) which is an extremely important factor.

[25] See Figure 8 in Section 4.3

6. Conclusions

Although the software model we implemented evidenced the ability to predict the outcome of a certain Variable using a **linear regression fit** with an acceptable degree of confidence, we are convinced that there are other types of regression models which are more suitable and potentially able to produce more accurate results for the subject matter under examination. In particular, **radial basis function** (RBF), a popular kernel function used for learning algorithms (especially SVM), and **polynomial regressions**, are much more solid supervised learning models, capable of estimating variables with exceptional precision. Figure 10 below provides a visual comparison between these three different regression models.

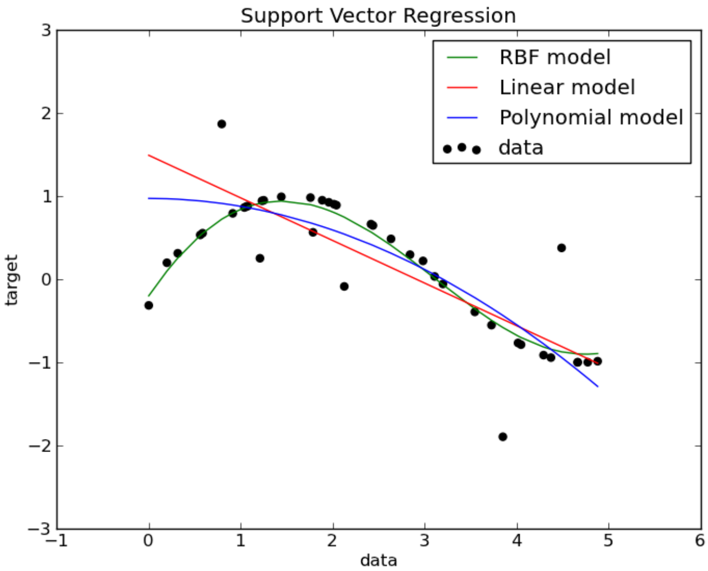


Figure10 Example of regression fit using linear, polynomial and RBF kernels

There are nonetheless many ways to improve any given regression model, including the simple linear regression we performed in this research study. One way our model could be improved is to consider the best 3 guesses of the software. Many algorithms that try to predict the purchasing habits of customers, in fact, rely on this margin of error due to the fact that more than one "package" can be displayed on any given "virtual shelf" (as in the case, for example, of Amazon's user-tailored purchase recommendations or YouTube's advertisements and correlated videos).

The research study we performed can also be improved by increasing the sample size of respondents in order to decrease the probability of biased results and to have a sample that better reflects the population of video gamers. For this reason, we suggest to expand the study to many other different websites and medias and to popularize the survey via any type of incentives (pecuniary or otherwise). Additionally, in order to avoid a participant taking the survey more than once, it would be necessary to implement a system that is able to deny access to the survey from the same IP address or which requires signing in with an account of some kind (we suggest a Google account since it is one of the most common and is possibly the easiest to implement).

Furthermore, one or perhaps two filtering questions may be included in later studies. In our research study, we found some inconsistencies among respondents, suggesting that either they did not read the description of the survey or did not know/understand the general knowledge necessary to participate to the survey (for example, not knowing exactly what a loot box is, may have led to inconsistent answers). Additionally, the encoding factor although important for the accurate codification of most of the variables, did not prove to be beneficial in other instances. In fact, in some cases, the artificial distance we created between the variables influenced the results. What we suggest for later studies is to implement a legend within the survey in order to communicate to the respondent the different "values" attributed to each single variable.

Ultimately, some further consideration about the variables analyzed is due. What our research study in fact found is that the inclusion of optional questions and sections in the survey greatly influenced the prediction of the software, thus creating a bias for the model when predicting certain variables. Subsequent studies should therefore cleverly address this issue by either dropping the Variable altogether (which we would not recommend), or through an alternative solution to the problem which obviously requires more in-depth contemplation.

Finally, while performing our research study, we found that some variables were of minor importance and decided not to consider them at all (namely the variables HouseH and ObtainLB). In order to avoid finding yourself with information that does not benefit your study, we suggest the conduction of a preliminary investigation so that your questionnaire is tailored as best possible to the analysis you intend to perform. In fact, our research study began as an analysis that aimed at including gambling as an important variable. However, during the course of the thesis, the lack of data made it necessary for us to turn to machine learning which is capable of predicting at least the variables for which our understanding was complete.

7. References

Chen, Ningyuan and Elmachtoub, Adam and Hamilton, Michael and Lei, Xiao, Loot Box Pricing and Design (August 1, 2019). Available at SSRN: https://ssrn.com/abstract=3430125 or http://dx.doi.org/10.2139/ssrn.3430125

Alex Knoop, LLB 2018 Loot Crates: Where the Difference Between 'Gaming' and 'Gambling' is Simply Two Letters. Available at https://www.aph.gov.au

Raylu, N., & Oei, T. P. S. (2002). Pathological gambling: A comprehensive review. Clinical Psychology Review, 22, 1009–1061.

Nojima, M. (2011, September 22). Soosharu geemu niokeru nihongata deeta doribun no arikata toha? [How should the Japanese model of data-driven in social games?]. Business Media Makoto. Retrieved from http://bizmakoto.jp/makoto/articles/1109/22/news015.html

King, D.L., Delfabbro, P.H. Video Game Monetization (e.g., 'Loot Boxes'): a Blueprint for Practical Social Responsibility Measures. Int J Ment Health Addiction 17, 166–179 (2019). https://doi.org/10.1007/s11469-018-0009-3

Zendle D, Meyer R, Ballou N (2020) The changing face of desktop video game monetisation: An exploration of exposure to loot boxes, pay to win, and cosmetic microtransactions in the most-played Steam games of 2010-2019. PLOS ONE 15(5): e0232780. https://doi.org/10.1371/journal.pone.0232780

Activision Blizzard, Inc. Activision Blizzard 2019 Annual Report. Available at: https://investor.activision.com/static-files/87075fb9-60bc-4ab9-8604 (accessed 26 July 2020).

Marr M. D., Kaplan K. S., Lewis N. T. U.S. Patent no. 9,789,406. 17 October 2017. Available at: https://www.google.com.au/patents/US9789406 (accessed 26 July 2020).

Xue S., Wu M., Kolen J., Aghdaie N., Zaman K. A. Dynamic difficulty adjustment for maximized engagement in digital games. 5 April 2017. Available at: http://www.webcitation.org/709xobonn.

Fingas, Jon. 2018. South Korea fines game studios over deceptive loot box odds. https://www.engadget.com/2018/04/10/south-korea-fines-game-studios-over-loot-boxes

T.J. Hafer October 26, 2018. Are loot boxes gambling? The jury is still out in most places. https://www.pcgamer.com

Gaming Authority. Study into loot boxes. A treasure or a burden? 10 April 2018. Available at: https://www.kansspelautoriteit.nl/library/study_into_loot_boxes

Zendle, D., Meyer, R., Cairns, P., Waters, S., and Ballou, N. (2020) The prevalence of loot boxes in mobile and desktop games. Addiction, https://doi.org/10.1111/add.14973.

Jamie Coogan, November 2017. Games as a Service. Microtransactions become the new norm in gaming. Available at: www.marketline.com Reference Code: MLAI0002-171

Nenad Tomić, Effects of micro transactions on video games industry. January 2017. Available at: www.researchgate.net/publication/Effects_of_micro_transactions

Prof. Matthew Rockloff, Dr. Alex M T Russell, Ms. Nancy Greer, Dr. Lisa Lolé, Prof. Nerilee Hing, Prof. Matthew Browne, June 2020. Central Queensland University. Loot Boxes: Are they grooming youth for gambling? Prepared for: The NSW Responsible Gambling Fund.

Anderson, C. K., & Xie, X. (2012). Pricing and market segmentation using opaque selling mechanisms [Electronic version]. Retrieved [insert date], from Cornell University, School of Hospitality Administration site: http://scholarship.sha.cornell.edu/articles/374

Elmachtoub, Adam & Wei, Yehua. (2015). Retailing with Opaque Products. SSRN Electronic Journal. 10.2139/ssrn.2659211.

XinGeng, March 2016. School of Business Administration, University of Miami, 5250 University Drive, Coral Gables, FL, 33146, USA. https://doi.org/10.1016/j.orl.2016.09.005

Scott Fay, Jinhong Xie March 2008. Probabilistic Goods: A Creative Way of Selling Products and Services. https://doi.org/10.1287/mksc.1070.0318

Jiang, Y. Price discrimination with opaque products. J Revenue Pricing Manag 6, 118–134 (2007). https://doi.org/10.1057/palgrave.rpm.5160073

Garrelts, N. (2010). I'm just a wizard laboring in a violent and softcore consumer culture: A historical look at the changing culture of consumption in digital games. Bad Subjects, Available at: http://bad.eserver.org/issues/2010/garreltsgames.html

McCaffrey, M. (2019). The macro problem of microtransactions: The self-regulatory challenges of video game loot boxes. In Business Horizons (Vol. 62, Issue 4, pp. 483–495). https://doi.org/10.1016/j.bushor.2019.03.001

McCaffrey, M. (2020). A cautious approach to public policy and loot box regulation. Addictive Behaviors, 102, 106136. https://doi.org/10.1016/j.addbeh.2019.106136

Arvidsson, C. (2018). The Gambling Act 2005 and loot box mechanics in video games. Ent. L.R., 29(4), 112-114. https://ssrn.com/abstract=3374311

Greer, N. (2018). Gambling and video games: Are Esports and Skin Betting pathways to greater youth gambling involvement and harm? ECR Gambling Grant, Victorian Responsible Gambling Foundation, $50K.

Hamari, J., Alha, K., Järvelä, S., Kivikangas, J. M., Koivisto, J., & Paavilainen, J. (2017). Why do players buy in-game content? An empirical study on concrete purchase motivations. Computers in Human Behavior, 68, 538–546. https://doi.org/10.1016/j.chb.2016.11.045

Cerulli-Harms, A. et al., Loot boxes in online games and their effect on consumers, in particular young consumers, Publication for the committee on the Internal Market and Consumer Protection (IMCO), Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020. © Cover image provided by Stefan Coders from Pixabay

Alha, K., Koskinen, E., Paavilainen, J., Hamari, J., & Kinnunen, J. (2014). Free-toplay games: Professionals' perspectives. In Proceedings of DiGRA Nordic 2014, Gotland, Sweden. http://www.digra.org/digital-library/publications/free-toplay-games -professionals-perspectives/

Brooks, G. A., & Clark, L. (2019). Associations between loot box use, problematic gaming and gambling, and gambling-related cognitions. Addictive Behaviors, 96, 26–34. https://doi.org/10.1016/j.addbeh.2019.04.009

Brus, A. (2013). A young people's perspective on computer game addiction. Addiction Research and Theory, 21, 365–375. http://dx.doi.org/10 .3109/16066359.2012.733466

Cleghorn, J., & Griffiths, M. D. (2015). Why do gamers buy 'virtual assets'? An insight into the psychology behind purchase behaviour. Digital Education Review, 27. http://greav.ub.edu/der/

Dreier, M., Wölfling, K., Duven, E., Giralt, S., Beutel, M. E., & Müller, K. W. (2017). Free-to-play: About addicted Whales, at risk Dolphins and healthy Minnows. Monetarization design and Internet Gaming Disorder. Addictive Behaviors, 64, 328–333. https://doi.org/10.1016/j.addbeh.2016.03.008

Drummond, A., & Sauer, J. D. (2018). Video game loot boxes are psychologically akin to gambling. Nature Human Behaviour, 2(8), 530–532. https://doi.org/10.1038/s41562-018-0360-1

Drummond, A., Sauer, J. D., Ferguson, C. J., & Hall, L. C. (2020). The relationship between problem gambling, excessive gaming, psychological distress and spending on loot boxes in Aotearoa New Zealand, Australia, and the United States—A cross-national survey. PLoS One, 15(3), e0230378. https://doi.org/10.1371/journal.pone.0230378

Gainsbury, S. M., Russell, A., & Hing, N. (2014). An investigation of social casino gaming among land-based and Internet gamblers: A comparison of sociodemographic characteristics, gambling and comorbidities. Computers in Human Behavior, 33, 126-135. http://doi.org/10.1016/j.chb.2014.01.031

Griffiths, M. D., & Kuss, D. J. (2015). Online addictions: Gambling, video gaming, and social networking. In S. S. Sundar (Ed.), The handbook of the psychology of communication technology (pp. 247-269). Chichester, UK: Wiley Blackwell. Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/9781118426456

Griffiths, M. D. (2018). Is the buying of loot boxes in video games a form of gambling or gaming? Gaming Law Review, 22, 52–54

Hamari, J., & Keronen, L. (2017). Why do people buy virtual goods: A meta-analysis. Computers in Human Behavior, 71, 59–69. https://doi.org/10.1016/j.chb.2017.01.042

Hamari, J. & Lehdonvirta, V. (2010). Game design as marketing: How game mechanics create demand for virtual goods. International Journal of Business Science and Applied Management, 5(1), 14-29. Handrahan, M. (2018). ESA: We can't go to the "lowest common denominator of government" on loot boxes. GamesIndustry.biz. https://www.gamesindustry.biz/articles/2018-05-25-esa-we-cant-go-to-thelowest-common-denominator-of-government-on-loot-boxes

Wright, M. (2018). Video gamers will be spending $50 billion on gambling-like loot box features by 2022, according to analysts. https://www.telegraph.co.uk/technology/2018/04/17/video-gamers-willspending-50-billion-gambling-like-loot-box/

Appendix A: Facebook groups

Thesis / Survey Questionnaire Filling Group

Student Survey Exchange

Research Participation - Dissertation, Thesis, PhD, Survey Sharing

Videogiochi..che passione, ma non solo!!!!

STEAM ITALIA

THE WORLD GLOBAL GAMES AND PPSSPP GAMES/GAMERS WORLDWIDE

PC GAMING ITALIA 💥

Dissertation Survey Exchange

PC Gamers ITALIA

GameTime - È Tempo di Videogiochi!

Dissertation Survey Exchange – Share Your Research Study, Find Participants

GAME JUNKIES - Buy/Sell/Trade & Everything Video Games

Gaming 🎮

GAMERS AND STREAMERS ZONE

Video Games (🎮🎮🎮♡🎮🎮🎮)

Video Games

La casa dei VIDEOGAMES

100% Videogames Passion

Malati Di PC - Videogames

GameSoul

Anime & Gaming Society ☑

RETROGAMES ARCADE® 🎮

Appendix B: Subreddits

Videogames

Gaming/

Gaming Circlejerk - Home of CD Projekt Red

GamingDetails

Gaming PC

XboxOne

PS4

Nintendo Switch - News, Updates, & Information

Quality Gaming Content and Discussion -- /r/Games

Power to the Readers

Take My Survey

Education

Steam on Reddit

/r/SampleSize: Where your opinions actually matter!

Appendix C: Survey text

Video games & loot boxes (an academic survey)

My name is Lorenzo De Gasperis and I am a student of the university LUISS Guido Carli of Rome in Italy who is preparing a Master Degree Thesis on loot boxes and video games. To this end I have prepared the following survey which will have the sole purpose of gathering material for my thesis with the aim to understand which are the main drivers that move gamers to purchase additional digital contents via micro transactions.

I apologize in advance for any imperfections in the text due to the fact that English is not my mother tongue.

The survey is intended for people who have played video games or purchased loot boxes in the past 12 months. It is completely anonymous and confidential and the data collected will be used for the sole purpose described above.

By submitting this survey you agree to the terms described above and declare that you are eligible to take the survey based on its parameters.

**Answer required*

1. I identify myself as... *

   Choose only one

   - Male
   - Female
   - Altro: ____________

2. What's your age range? *

   Choose only one

   - 12 or below
   - 13-16
   - 17-20
   - 21-27
   - 28 or above

3. Which continent are you from? *

Choose only one

- North America
- Central/South America
- Europe
- Asia
- Africa
- Oceania

4. What is your occupation? *

Choose only one

- Student
- Part time job
- Full time job
- Freelance
- Unemployed
- Retired

5. Which of the following best describes your personal income last year? *

Choose only one

- 0 $ - 5000 $
- 5000 $ - 10 000 $
- 10 000 $ - 20 000 $
- 20 000 $ - 35 000 $
- 35 000 $ - 60 000 $
- more than 60 000 $

6. How many people are in your household? *

Choose only one

- only me
- 2
- 3 - 5
- more than 5

Video games and loot boxes

7. How many hours per week (on average) do you spend playing video games? *

Choose only one

- 30 minutes - 1 hour
- 1 - 2 hours
- 2 - 4 hours
- 4 - 6 hours
- 6 - 8 hours
- more than 8 hours

Loot Box

For the purpose of this study:

1. a loot box is a VIRTUAL bundle of one or more VIRTUAL items (physical loot boxes are not included);

2. the loot box can be opened, and its material must reward the player with contents that can be used for that game in some way or at least has a value for the player. The content may include cosmetic items, in-game features and abilities (power ups, weapons and upgrades), in-game currency, functional items (such as, for instance, new playable characters), or even more loot boxes;

3. loot boxes do not need to visually resemble a “box” or “crate” but can take many different forms as long as they are characterized as being a container for in-game random rewards;

4. loot boxes must contain random content (or must at least appear random to the customer).

8. Which of the following methods for obtaining loot boxes do you agree with? *

Choose only one

- Real money only
- In-game currency only
- Through progression (leveling up rewards)
- Real money and progression
- All of the above
- Don't care

9. On a scale of 1 to 5 how do you consider "cosmetic only" loot boxes acceptable?*

Choose only one

	1	2	3	4	5	
Not acceptable at all						Very acceptable

10. On a scale of 1 to 5 how do you consider "progression tied" loot boxes acceptable? *

Choose only one

	1	2	3	4	5	
Not acceptable at all						Very acceptable

11. How often do you like to bet? (please include any type of bet e.g. football bet, "Scratch and win", Bingo etc) *

Choose only one

- Never
- Once a month
- 2 - 4 times per month
- 5 - 8 times per month
- More than 8 times per month

12. On a scale of 1 to 5 how similar do you consider loot boxes to gambling? *

Choose only one

	1	2	3	4	5	
Not similar at all						Very similar

13. Have you ever spent any money on "in-game purchases" such as loot boxes? *

Choose only one

- Yes
- No

If you answered "no" to the previous question, please leave the rest of the questions blank. In this case you can go to the end of the questionnaire and press "Send"

14. How much have you spent on loot boxes in the last month?

Choose only one

- 10 $ or less
- 10 $ to 50 $
- 50 $ to 100 $
- 100 $ to 500 $
- over 500 $

15. Have you ever decided to buy a loot box instead of something else?

Choose only one

- Yes
- No

16. On a scale of 1 to 5 how important is it for you to buy loot boxes when you play video games?

Choose only one

	1	2	3	4	5	
Non existent						Very strong

17. How much would you say buying loot boxes satisfies you?

Choose only one

	1	2	3	4	5	
Not satisfied at all						Very satisfied

18. Which of the following best describes the reason you purchase loot boxes?

Choose only one

- Just for fun
- To enhance my gaming experience
- To customize my character's appearance
- To possibly gain an advantage
- To ease my progression through the game
- To compete
- To accomplish achievements
- Everyone else buys them
- Addicted to buying loot boxes
- Peer pressure from teammates/friends

19. Loot boxes allow me to better characterize myself in the game.

Choose only one

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

20. Loot boxes make me happier when I play video games.

Choose only one

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

21. Loot boxes allow me to progress faster in my games.
Choose only one

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

For SurveyCircle users (www.surveycircle.com): The Survey Code is: S54G-HDLL-GT91-TK1S

Redeem Survey Code with one click: https://www.surveycircle.com/S54G-HDLL-GT91-TK1S

Appendix D: Variables

Questions	Variables
What's your age range?	Age
Which continent are you from?	Continent
What is your occupation?	Occupation
How many hours per week (on average) do you spend playing video games?	HoursAVG
Which of the following methods for obtaining loot boxes do you agree with?	ObtainLB
On a scale of 1 to 5 how do you consider "cosmetic only" loot boxes acceptable?	CosmLB
On a scale of 1 to 5 how do you consider "progression tied" loot boxes acceptable?	ProgLB
On a scale of 1 to 5 how similar do you consider loot boxes to gambling?	GambLB
Have you ever spent any money on "in-game purchases" such as loot boxes?	Buy-YN
How much have you spent on loot boxes in the last month?	Spent
On a scale of 1 to 5 how important is it for you to buy loot boxes when you play video games?	Imp
Which of the following best describes the reason you purchase loot boxes?	Reasons
I identify myself as...	Gender
Which of the following best describes your personal income last year?	Income
How many people are in your household?	HouseH
How much would you say buying loot boxes satisfies you?	Satisfy
Have you ever decided to buy a loot box instead of something else?	LBInstead
Loot boxes allow me to better characterize myself in the game.	CharacterizeLB
Loot boxes make me happier when I play video games.	HappyLB
Loot boxes allow me to progress faster in my games.	ProgressLB
How often do you like to bet? (please include any type of bet e.g. football bet, "Scratch and win", Bingo etc)	BetOften

Appendix E: Matrix

	Gender	Age	Continent	Occupation	Income	HouseH	HoursAVG	ObtainLB	CosmLB	ProgLB
Gender	1	-0.149	0.172	0.002	0.766	0.633	0.292	0.147	0.052	0.084
Age	-0.149	1	-0.276	0.205	-0.204	-0.17	-0.205	0.069	-0.165	-0.079
Continent	0.172	-0.276	1	-0.045	0.178	0.239	0.193	0.062	0.123	0.089
Occupation	0.002	0.205	-0.045	1	0.055	-0.027	-0.076	0.035	0.019	-0.057
Income	0.766	-0.204	0.178	0.055	1	0.536	0.252	0.161	0.055	0.091
HouseH	0.633	-0.17	0.239	-0.027	0.536	1	0.227	0.119	-0.038	0.109
HoursAVG	0.292	-0.205	0.193	-0.076	0.252	0.227	1	0.065	0.085	0.115
ObtainLB	0.147	0.069	0.062	0.035	0.161	0.119	0.065	1	0.106	0.002
CosmLB	0.052	-0.165	0.123	0.019	0.055	-0.038	0.085	0.106	1	-0.024
ProgLB	0.084	-0.079	0.089	-0.057	0.091	0.109	0.115	0.002	-0.024	1
BetOften	0.758	-0.138	0.225	0.103	0.755	0.648	0.308	0.115	0.051	0.176
GambLB	-0.06	-0.024	0.039	-0.028	-0.063	-0.084	0.129	-0.098	0.028	-0.185
Buy-YN	0.009	-0.076	0.093	-0.148	0.026	-0.079	0.185	0.164	0.248	0.081
Spent	-0.004	-0.115	0.034	-0.108	0.024	-0.122	0.099	0.16	0.225	0.03
LBInstead	0.197	-0.018	0.046	0.139	0.229	0.157	0.136	0.161	0.075	0.143
Imp	-0.102	0.011	0.044	0.015	-0.067	-0.134	0.06	0.21	0.203	0.056
Satisfy	0.423	-0.11	0.21	0.004	0.437	0.293	0.222	0.313	0.11	0.217
Reasons	-0.088	-0.001	0.023	-0.08	-0.085	-0.1	0.07	0.123	0.202	0.083
CharacterizeLE	0.472	-0.072	0.18	-0.041	0.447	0.313	0.281	0.254	0.109	0.111
HappyLB	0.468	-0.136	0.197	-0.037	0.493	0.306	0.292	0.183	0.053	0.06
ProgressLB	0.466	-0.114	0.17	-0.059	0.48	0.32	0.207	0.232	0.072	0.062

BetOften	GambLB	Buy-YN	Spent	LBInstead	Imp	Satisfy	Reasons	CharacterizeLB	HappyLB	ProgressLB	
0.758	-0.06	0.009	-0.004	0.197	-0.102	0.423	-0.088	0.472	0.468	0.466	**Gender**
-0.138	-0.024	-0.076	-0.115	-0.018	0.011	-0.11	-0.001	-0.072	-0.136	-0.114	**Age**
0.225	0.039	0.093	0.034	0.046	0.044	0.21	0.023	0.18	0.197	0.17	**Continen**
0.103	-0.028	-0.148	-0.108	0.139	0.015	0.004	-0.08	-0.041	-0.037	-0.059	**Occupati**
0.755	-0.063	0.026	0.024	0.229	-0.067	0.437	-0.085	0.447	0.493	0.48	**Income**
0.648	-0.084	-0.079	-0.122	0.157	-0.134	0.293	-0.1	0.313	0.306	0.32	**HouseH**
0.308	0.129	0.185	0.099	0.136	0.06	0.222	0.07	0.281	0.292	0.207	**HoursAV**
0.115	-0.098	0.164	0.16	0.161	0.21	0.313	0.123	0.254	0.183	0.232	**ObtainLB**
0.051	0.028	0.248	0.225	0.075	0.203	0.11	0.202	0.109	0.053	0.072	**CosmLB**
0.176	-0.185	0.081	0.03	0.143	0.056	0.217	0.083	0.111	0.06	0.062	**ProgLB**
1	-0.045	-0.121	-0.132	0.186	-0.142	0.343	-0.162	0.378	0.339	0.349	**BetOften**
-0.045	1	-0.025	0.016	-0.005	-0.089	-0.162	-0.105	-0.114	-0.075	-0.1	**GambLB**
-0.121	-0.025	1	0.807	0.263	0.64	0.559	0.743	0.527	0.544	0.522	**Buy-YN**
-0.132	0.016	0.807	1	0.219	0.424	0.419	0.653	0.463	0.474	0.459	**Spent**
0.186	-0.005	0.263	0.219	1	0.318	0.512	0.258	0.339	0.254	0.347	**LBInsteac**
-0.142	-0.089	0.64	0.424	0.318	1	0.547	0.67	0.382	0.333	0.361	**Imp**
0.343	-0.162	0.559	0.419	0.512	0.547	1	0.471	0.743	0.675	0.713	**Satisfy**
-0.162	-0.105	0.743	0.653	0.258	0.67	0.471	1	0.453	0.438	0.425	**Reasons**
0.378	-0.114	0.527	0.463	0.339	0.382	0.743	0.453	1	0.822	0.803	**Characte**
0.339	-0.075	0.544	0.474	0.254	0.333	0.675	0.438	0.822	1	0.824	**HappyLB**
0.349	-0.1	0.522	0.459	0.347	0.361	0.713	0.425	0.803	0.824	1	**Progressl**

8. Executive Summary

The topic I decided to address in my thesis, leveraging on the knowledge gained during my course of quantitative methods, regards the video games industry. Considering the research, investments and resources involved in the industry, video games rarely receive the attention they deserve. Economists rarely even treat the subject of video games and seem to have a blurred vision of the vast world that comprises the industry. It is undeniable, however, that video gaming have not only become very widespread in recent years, but that the industry scenario in which developers and gamers navigate is ever more complex and articulated. Competition is fierce and costs for developers are constantly rising. For this reason, game developers tried for years to elaborate alternative ways to increase returns in order to create a sustainable stream of cash and revenues.

The industry of video games has developed rapidly and has expanded greatly in recent years. Some of the most important changes that the industry observed come from the introduction and development of mobile gaming and e-sports, both of which constitute nowadays an important portion of revenues for the market. In June 2004 we witnessed for the first time an unprecedented phenomenon; the introduction of an item called *Gachapon ticket* included in the Japanese version of “MapleStory”. These tickets allowed players to obtain random virtual items in the game and were sold at the price of 100 JPY (approximately 0.95 USD) each. Despite all the conditions necessary to exploit this new payment system were already available, no one would ever have expected that the creation and development of this new form of payment would eventually be able to boost the video game industry and guarantee unprecedented revenues in the years that followed.

During the past few years, the video game industry has begun perfecting this articulated and complex system also through the introduction of micro-transactions. In the case of video games, through the payment of a nominal amount, certain virtual items to be used in the game can be purchased. This phenomenon has become so relevant that many refer to the design, development, and implementation of micro-transactions as constituting an entirely new “Business Model”.

As explained, through micro-transactions players can purchase additional game content. This constitutes the highest percentage of revenues for most games that, initially. do not require any paying commitment to access the game and begin playing (and belong the so-called “free-to-play” category).

Micro-transactions have been studied for years in order to identify the proper optimization method to encourage purchases and enhance the experience of players. In other words, game developers seek a clear way to implement such monetization strategies in their games in manner that prevents them from appearing redundant, obtrusive or useless to the customers. Additionally, competition has become so fierce that developers are obliged to attract players that are inclined to spend as well as those which are not. Optimally selecting the pricing strategy that could meet the purchasing capabilities of the customers and create a tailored product that attracts multiple types of players is essential in order to achieve a revenue-maximizing strategy for video game companies. For this reason developers have started analysing consumers' reactions and their propensity to spend according to certain selling mechanisms. This is the very reason we concluded that the activities of design, development, and implementation of micro-transactions has become increasingly aligned to what is generally referred to as a "Business Model".

Among the methods used to customize and tailor micro-transactions to the expectations of the players is the presentation of certain specific products and features, for example through a so-called loot box (also known as "loot prize/crate") which is a virtual item that can be redeemed in order to receive a selection of additional virtual items to enhance the gaming experience.

This is where my research study steps in: designing loot boxes has become so important that firms have started discussing ways and developing theories on how to maximise the probability that a given package be purchased. This study began as a research aimed at identifying the most common reasons that encourage players to purchase loot boxes and to turn to micro-transactions in general, in order to identify the most suitable and appropriate "package" which is tailored to their purchasing habits and therefore able to satisfy their needs. The aim of this research study is, in fact, to provide an extensive analysis of the background regarding the topic and to develop a quantitative study that presents an in-depth statistical data analysis in order to better comprehend the potentially related variables among those analysed. The information provided will then be necessary to understand the extent of the analysis developed as well as the rationale underlying the answer to the thesis question: to what extent and with which degree of accuracy can the algorithms developed by the software houses predict the spending habits of video games players?

The results of this paper are based on the responses to a questionnaire submitted to video gamers and shared online from the 1st of June until the 10th of September 2020. It sampled 243 video gamers and 134 loot box purchasers using the following criterion:

- the survey was uploaded on 22 Facebook pages and 14 subreddits, most of which were related to either general gaming or to video gaming in particular;
- the questionnaire was submitted in English (with the possibility of translating it into different languages depending on the browser used in the filing) in order to maximize the accessible audience;
- in order to be eligible to take the survey, it was necessary that respondents had played video games in the 12 months preceding their participation to the survey or had purchased loot boxes (or similar in-app purchases) in the past;
- prior to initiating the survey an agree/disagree policy was submitted to reassure participants of the non-disclosure of their personal data and to gather their consensus to use the information provided albeit for academic purposes only.

The entire text of the survey is set forth in Appendix C. The survey was conceived to be as simple and as short as possible in order to avoid people closing the text before submission due to stress or potential fatigue. For this reason, only 21 questions were envisaged, 8 of which were exclusive only to respondents who indicated that they had indeed purchased loot boxes on some occasion, since the questions regarded various aspects of the purchase as well as the emotional commitment when playing. These respondents are among the most important since, through their answers, we can build a model to predict players' future spending and degree of satisfaction.

After the gathering phase, the data collected was re-elaborated to make it easier to manage and "translated" for the envisaged purpose. The data collected was organized in Variables.

Variables in the survey are of two types: "Qualitative" and "Quantitative". Qualitative Variables answers are characterized by objects while Quantitative Variables answers are characterized by floats and integers. The objects are either a Boolean or can have another form. Instead, floats and integers are simply numbers. Together, objects and floats and integers will be referred to as "Entries".

The analysis of the data comprised a selection of scripts run in a defined programming environment created for the purpose of this research paper. The software used in the studies are Anaconda3, an open-source data science toolkit used to perform Python/R data science and machine learning, and Jupyter Notebook, another open-source software easily available through the web.

We installed and imported all the packages necessary for the study (the complete list, together with the description of these packages, can be found in Section 4.3). We imported all the data from the excel file, we created a data frame and then printed the head of the data frame. At this point could easily identify and distinguish objects from floats and integers, and to count the number of responses for each Variable.

As explained earlier, Qualitative Variables are not suitable for mathematical purposes and require additional elaboration. We therefore needed to "translate" them into a "workable" format. We used a coding instrument called "encoding" to create a function that replaced each single object of the Qualitative Variables of the sample with a number ranging from 0 to the extent of the Variables. In other words, if a given Qualitative Variable comprised 5 different objects, the function called the first object 0, the second 1, the third 2 etc. up to the last object, without repetition. To do so we created a list containing the Qualitative Variables we needed to encode. Then we defined a function that was able to transform each object into a number. The function (which we will call "encoder") used an encoded data frame and a list of Variables. For each entry of the list it repeated the following process:

(i) it created a set of the objects (extrapolating each object only once, without repetition);
(ii) it transformed the set into a list and created a number of labels equal to the length of the list;
(iii) it created a dictionary and a loop that was able to match each object to its corresponding label;
(iv) finally it replaced the label with the corresponding item on the list.

Quantitative Variables needed no encoding procedure.

We then performed a series of investigations in order to infer the statistically significance of the data collected. We ran Chi-Square tests of independence and portrayed a Variance-Covariance matrix to elaborate on Relevant Variables (as defined in Section 4.3). The matrix had dimensions 21 x 21 and presented most of the values with little to no covariance. However by implementing a filtering process, thus selecting only those values which are greater than 0.4, we could omit the unnecessary information and focus solely on the more relevant parameters.

Leveraging on this Variance-Covariance matrix we understood whether the Variables are somewhat connected and hence "relevant" with respect to those related to the purchase and usage of loot boxes. Such matrix is provided in Appendix E. We also filtered the matrix to avoid uncorrelated elements and displayed only those covariances with a value above 0.4. Additionally, since the research study specifically addresses the correlation between loot boxes and gamers, we also excluded the covariance among demographic Variables which were considered only in relation to gaming-related Variables.

We then used a machine learning method based on linear regression, which is a linear approach to modeling that approximates the variables given a defined data set. In this research study we will refer to both simple linear regression, in case only one explanatory variable is considered, and multiple linear regression for more than one explanatory variable. The linear regression will be used to model our data via a linear predictor function.

As a first step, we checked which test percentage was most suitable to our case. We performed iterated tests on Relevant Variables only, and in each of these tests, we changed the test percentage of our machine. The range of test percentages we considered was from 0.050 to 0.185 (running the test 15 times, each time increasing the percentage test by 0.015). The output of such tests were recorded and we compared the software's correct guesses on Relevant Variables.

This intermediate step was necessary to determine the best test percentage. In fact, our test needed to be as broad as possible in order to assess whether the predictions of the software were reliable and, at the same time, leave the machine enough room to effectively learn how to predict outcomes.

The benchmark we used to select the appropriate test percentage was that at least 7 out of our 10 variables needed to be predicted correctly at least 65% of the time. The greatest test percentage that met such requirements was 0.070 (which corresponded to a test of 18 guesses).

Having defined the proper test percentage, we then needed to pass the Relevant Variables through the model. To test our model, we first checked each variable alone and then together with other variables. We found that three was the maximum acceptable number of variables that we could consider simultaneously and that the test percentages greatly influenced the prediction of the machine in the test phase. More than three variables, in fact, did not yield enough learning space for the machine and led to too many incorrect predictions. The same applies to test percentages.

The software we trained predicted the outcome of a given variable through the following process:

1. we defined a list of targets that the model would try to predict;
2. we set a test percentage of the total sample that the model would try to guess and, in doing so, we defined the shape of the train (i.e. how many outcomes does the model see before trying to guess the variable);
3. we passed the model through all the variables with the exception of the targets using a linear regression fit;
4. then the model tried to predict the outcome of the targets and checked whether its guessing was correct (iterating a process called “training”);
5. after a number of repetitions (equal to the total number of observations minus the test percentage) the machine no longer corrected itself with the right amount and only provided estimates on the variables;
6. we checked whether the model was correct in its prediction on the test percentage and we counted the correct responses.

This function allowed us to test whether the predictions of the software were correct and, at the same time, to vary the targets, the test percentage and the seed.

The research study also allowed us to estimate the Relevant Variables in the analysis with an accuracy greater than 70% if we provide a training sample equal to 93% of the entries. We are satisfied with the results given that the linear regression model we performed is one of the simplest regression fits that can be used and, with the a few tailored adjustments, the model can be improved greatly.

We used these variables within the linear regression model with the aim of estimating each of them in different circumstances:

- by varying the seeds, we found that the results varied irrelevantly as the sample (using a test percentage of 0.07) was large enough for most of the variables and the software had the necessary data to “learn”;
- by varying the tests, we noticed that the percentage of correctly guessed variables decreased as the number of guesses compared to the entire sample increased. This is possibly due to the same reason described in the preceding point above;
- lastly, we tried to estimate more than one variable at the same time. This, in turn, led to different results depending on the correlation between all the variables that we were trying to guess compared to all the variables which we were not trying to guess.

Additional final considerations regarding the model must be made:

- some of the variables had great values of prediction due to the fact that were strictly related to a particular section of the questionnaire, namely the questions specifically addressed to loot box users;
- certain other variables with significantly lower percentages of correct guesses must be adjusted according to the magnitude of objects they contain.

The research study we performed has huge margins of improvement, for example by considering the best 3 guesses of the software. Many algorithms that try to predict the purchasing habits of customers, in fact, rely on this margin of error due to the fact that more than one "package" can be displayed on any given "virtual shelf".

The research study we performed can also be improved by increasing the sample size of respondents in order to decrease the probability of biased results and to have a sample that better reflects the population of video gamers. For this reason, we suggest to expand the study to many other different websites and medias and to popularize the survey via any type of incentives (pecuniary or otherwise). Additionally, in order to avoid a participant taking the survey more than once, it would be necessary to implement a system that is able to deny access to the survey from the same IP address or which requires signing in with an account of some kind. Filtering questions may also be included in later studies in order to avoid some inconsistencies among answers.

Additionally, the encoding factor although important for the accurate codification of most of the variables, did not prove to be beneficial in other instances. In fact, in some cases, the artificial distance we created between the variables influenced the results. What we suggest for later studies is to implement a legend within the survey in order to communicate to the respondent the different "values" attributed to each single variable.

Ultimately, some further consideration about the variables analyzed is due. What our research study in fact found is that the inclusion of optional questions and sections in the survey greatly influenced the prediction of the software, thus creating a bias for the model when predicting certain variables. Subsequent studies should therefore cleverly address this issue by either dropping the Variable altogether, or through an alternative solution to the problem which obviously requires more in-depth contemplation. Finally we found some variables which were of minor importance and decided not to consider them at all.

The results of our study reveal that it is indeed possible to map video gamers' preferences and, by duly characterising the products (whether loot boxes or other product that rely on the micro-transaction mechanism), one can tailor the sale of certain, selected loot boxes by matching them to specific consumers. As a matter of fact, although the model used in our research study is relatively simply, it was able to effectively predict the preferences of video gamers with a reasonably satisfactory degree of reliability, also in consideration of the defects and flaws of the sample. The software model implemented was indeed able to predict the outcome of a certain Variable with an acceptable degree of confidence but we are convinced that there are other types of regression models which are more suitable and potentially able to produce more accurate results. In particular, radial basis function (RBF) and polynomial regressions are much more solid supervised learning models, capable of estimating variables with exceptional precision.

The information gathered constitutes an interesting and undoubtedly convenient piece of information for game developers which, if properly managed, can be used to propel cash flow and revenues significantly.

Main References:

Chen, Ningyuan and Elmachtoub, Adam and Hamilton, Michael and Lei, Xiao, Loot Box Pricing and Design (Aug 1, 2019).

King, D.L., Delfabbro, P.H. Video Game Monetization (e.g., ‘Loot Boxes’): a Blueprint for Practical Social Responsibility Measures. Int J Ment Health Addiction 17, 166–179 (2019).

Zendle D, Meyer R, Ballou N (2020) The changing face of desktop video game monetisation: An exploration of exposure to loot boxes, pay to win, and cosmetic microtransactions in the most-played Steam games of 2010-2019.