Identifying clusters of game play and social behaviors of heavy gamers in the Steam platform

Student Name: Sarthak Dey
Supervisor Name: Pratik Agarwal

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Master's in data science

FEBRUARY 2020

ABSTRACT

This study presents an in-depth investigation the virtual world of addicted gamers. These gamers are part of the popular digital distribution service Steam, which is one of the biggest gaming platforms on personal computers. This study on heavy gamers, uses data from the first overall examination on Steam which collected over 108.7 million user accounts and 384.3 million owned games. This previous study gave us an overview of the behavior of players in the gaming network. However, that study focused on the entire Steam community and found that 1% of the players played more than 5 hours. This top 1% amounts to around a million players. Analysis on these heavy gamers were out of scope for the study. The focus of the current study is investigating types of these heavy gamers and their behavior such as monetary expenditure, game ownership and play time statistics in specific video game genres. The current study finds these heavy gamers to have significant amounts of resource expenditure in terms of time and money. However, most of these heavy gamers showcase similar levels of social structure in terms of number of friends. The study finds two distinguishable clusters of heavy gamers with respect to their play times, resource expenditure and social structure. The first group can be deemed to be more addicted to buying video games than the second group. The second group despite playing around the same length of time as the first group only buys the best rated games and plays mostly free games. The first group seems to be more satisfied in acquiring the games themselves while the other seems to be either financially incapable of buying games or aims to master the game itself. The second group however can also be spending on in-game items such as skins and other digital items in free and multiplayer games to be masters in the game or look the part. The current study also finds both groups of the heavy gamers to be more engaged in games with a multiplayer component. The study recommends ways to help these gamers through connecting with them or their friends using the Steam ID provided.

ACKNOWLEDGEMENTS

I would like to thank my parents for supporting me throughout my studies and help me complete my thesis in the limited timeframe.

I would like to thank UPGRAD and Liverpool John Moores University for allowing me to pursue a master’s degree in data science.

In addition, I would like to thank google for allowing free and trial services such as Google Colaboratory and Google Cloud which greatly helps make research more accessible for students such as myself.

I would also like to thank my supervisor Professor Pratik Agarwal for guiding me throughout my thesis.

CONTENTS

LIST OF ABBREVIATIONS

ID – ‘Identity Document’. Meaning a unique identifier.

API – ‘Application Program Interface’.

SQL – ‘Structured Query Language’

DLC – ‘Downloadable Content’

IQR – ‘Interquartile range’

RPG – ‘Role Playing Game’

1. INTRODUCTION

1.1 Gaming Industry

Video game industry has been on the rise like never before. In 2019, the industry is worth around 130 billion dollars[1] and is bigger than the movie, music, and television industry[2]. In terms of the market, market capitalization of the largest 30 video game publishers had risen to a trillion U.S. dollars in 2018[1]. The immense growth of the video gaming industry over the years can be attributed to the rise of mobile games. Despite the biggest games only being released on computers and consoles, the mobile gaming industry makes gaming much more accessible than before. Big budget games like the video game ‘Red dead redemption 2’, according to the announcement by their developers, made a revenue of over 725 million U.S. dollars in just three days[3]. The first ever commercial video game and also the highest grossing video game ‘Space Invaders’ has made around 13.9 billion U.S. dollars since their launch in 1978 [3]. On another note, ‘Grand Theft Auto V’, in 2013, made over $800 million U.S. dollars[4] in worldwide revenue within just a day of its release. This game has 7 Guinness World Records[4], some of which are ‘Bestselling video game in 24 hours’ and ‘The fastest video game to gross 1 billion U.S. dollars’. Comparing this with the movie industry, the new movie Avengers Endgame in the whole of 2019 made around 2.8 billion U.S. dollars[5]. These figures showcase the sheer size of the gaming industry and the overall monetary expenditure of all the customers who buy the digital products. In addition to buying digital products there are also the consumption of video game related content such as streaming, gaming articles and esports. Video game testing for the gaming corporations, streaming live or recorded video game footage are also very popular in terms of earning money by garnering popularity among the viewership. Esports which is a form of competitive sport, usually in the form of gaming tournaments, is also a way of making a living out of video gaming.

---

[1] Nakamura, Y. (2020). Peak Video Game? Top Analyst Sees Industry Slumping in 2019. [online] Bloomberg.com. Available at: https://www.bloomberg.com/news/articles/2019-01-23/peak-video-game-top-analyst-sees-industry-slumping-in-2019 [Accessed 14 Feb. 2020].

[2] Reuters. (2020). Investing in the Soaring Popularity of Gaming. [online] Available at: https://www.reuters.com/sponsored/article/popularity-of-gaming?utm_source=reddit.com [Accessed 14 Feb. 2020].

[3] Sarkar, S. (2020). Red Dead Redemption 2 tops 17 million copies shipped. [online] Polygon. Available at: https://www.polygon.com/2018/11/7/18073314/red-dead-redemption-2-sales-17-million-units [Accessed 14 Feb. 2020].

[4] Karmali, L. (2013) GTA 5 Currently Holds Seven Guinness World Records, IGN. IGN. Available at: http://www.ign.com/articles/2013/10/09/gta-5-currently-holds-seven-guinness-world-records (Accessed: February 14, 2020).

[5] Avengers: Endgame (2019) - Box Office Mojo, web.archive.org/web/20191015150428/www.boxofficemojo.com/movies/?id=marvel2019.htm

Steam[6], a digital distribution service, was launched in 2003 by Valve corporation and has gained popularity ever since. Steam has its own digital store where people can buy games and software. In 2017, Steam made an estimated revenue of 4.3 billion U.S. dollars[7]. Steam runs its business by selling digital games and other digital components while also allowing social network type features for the players such as co-operative and other online multiplayer modes where the players can team up and play together or against each other. This however, not unique to Steam, is a compelling feature of the platform. At the time of data collection in 2013, Steam held 75% of the PC gaming market space[8]. In 2019, Steam had around 34000 games and over 95 million active users.

1.2 Background

Digital addiction has been defined as:

'a significant degree of dependent behavior that is triggered and facilitated by software products. It can lead to both pleasure and relief of discomfort, but unfortunately, in a way that can harm a person socially, physically and psychologically' [4].

A study on digital addiction done by Erik Peper and Richard Harvey [5] found that higher levels of usage of smartphones correlate with higher levels of isolation, depression and anxiety. Video gaming addiction, similar to internet addiction, is behavioral addiction. The brain activation of players in online gaming addiction was found to be closely similar to that of substance addiction [6]. The World Health Organization in May 2019 has added gaming disorder in the International Classification of Diseases (ICD-11) [1,2]. The World Health Organization has now officially classified gaming disorder as a condition and defined it as 'persistent and recurrent gaming behavior along with negative consequences in social life'.

Coming to real world cases of addiction, multiplayer mobile game, Player Unknown's Battlegrounds mobile has been on the receiving end of a lot of backlash, including a ban in Nepal[9][10]. Video games are very popular in China and hence addiction is also assumed to be

---

[6] Store.steampowered.com. (2020). Welcome to Steam. [online] Available at: https://store.steampowered.com/ [Accessed 14 Feb. 2020].

[7] Gough, C. (2019) Steam sales revenue 2017, Statista. Available at: https://www.statista.com/statistics/547025/steam-game-sales-revenue/ (Accessed: February 14, 2020).

[8] Edwards, C. (2013) Valve Lines Up Console Partners in Challenge to Microsoft, Sony, Bloomberg. Available at: https://web.archive.org/web/20141024081126/https://www.bloomberg.com/news/2013-11-04/valve-lines-up-console-partners-in-challenge-to-microsoft-sony.html (Accessed: February 14, 2020).

[9] Sharma, G. (2020). Nepal bans online game PUBG citing negative impact on children. [online] Reuters. Available at: https://www.reuters.com/article/us-nepal-pubg/nepal-bans-online-game-pubg-citing-negative-impact-on-children-idUSKCN1RN2FQ [Accessed 14 Feb. 2020].

[10] Mekaad, S. (2020). MP: Boy playing PUBG dies of heart attack in Neemuch | Indore News - Times of India. [online] The Times of India. Available at: https://timesofindia.indiatimes.com/city/indore/mp-boy-playing-pubg-dies-of-heart-attack-in-neemuch/articleshow/69577299.cms [Accessed 14 Feb. 2020].

worse in the country [14]. This assumption was made stronger when the Chinese government, in Nov 2019, has set playtime and monetary limits for its young gamers[11]. Some of these limits included no video gaming sessions between 22:00 and 8:00 and playtime limits on weekdays and weekends. In the past, the South Korean government reportedly started 140 counselling centers to treat online addiction [7]. In western countries, such as United Kingdom and Holland there have been gaming addiction clinics to help with the rehabilitation of addicted gamers [8, 9].

Looking at the history, video games have been around since 1958 when the first video game 'Tennis for two' was revealed. In 2019, with the rise of mobile games, gaming has become more accessible than ever before. Steam, being one of, if not the most famous gaming platform on personal computers has millions of players. As a result, overuse of video games was also found to be prevalent in the platform. In the study by O'Neil et al. [10], it was found that the top 1% of the players play over 5 hours a day. In addition to this, smartphones have been shown to be addictive in nature alongside the internet itself. One study [15] on smartphone addiction found higher smartphone use positively correlated with higher amounts of stress and negatively correlated with academic performance. However, it would be important to note that video gaming addiction may not be a standalone phenomenon but be accompanied by other forms of digital addiction such as smartphone and internet addiction as well.

1.3 Aim and Objectives

The study investigates the virtual world of gamers like no other study before to get a more in-depth look into the world of addicted gamers. The millions of players on the Steam platform ensures thousands of heavy gamers to be studied upon in hopes to understand and recommend possible ways to decrease their addictive behavior.

The study tries to find groups among the heavy gamers that have distinct diversity among them. This diversity can be in terms of gaming behavior, social behavior and monetary expenditure. For example, the study will take into consideration the amount of time the gamers in question have spent playing, the amount of money they spend playing video games, the genres of games they like, the amount of friends they have and the groups they are a part of. These attributes are used to run a clustering algorithm to segregate the groups.

The objective of the study will therefore be to reveal the types of addictive behavior of the gamers and recommend techniques to reduce their addiction. This technique can be in terms of

[11] 'Video Game Addiction: China Imposes Gaming Curfew for Minors.' BBC News, BBC, 6 Nov. 2019, www.bbc.com/news/world-asia-50315960.

rehabilitation methods for the gamers. The study hence can prove vital for the overall gaming community as the gaming corporations could help put a stop to gaming addiction itself. The platform like Steam, despite making a profit out of the addicted players, might have a bigger interest in helping them to avoid government regulations. This knowledge can also help the platform earn more revenue from the heavy gamers. Steam, being possibly the biggest digital distribution service in personal computers may find it best fit to help the gamers as these gamers which will help the platform garner even more popularity in the gaming community. The clusters found at the end of the current study may help with the treatment of addiction since one method to treat addiction may not work for all the individuals suffering from addiction as their behavior may be quite different from each other. Henceforth, understanding the different aspects of the clusters found will also help us to become more aware of the nuances of gaming addiction itself.

1.4 Research Questions

The research question revolves around the heavy gamers in the Steam platform. The dataset[12] was collected in 2013, using data collected through Steam API [13] during the study on Steam by O'Neil et al. [10]. The dataset shows various game play, social and/or economic behavior for different genres and types of games. The research questions in the current study are the following:

Do all heavy gamers on the Steam platform show the same type of behavior or are there groups of heavy gamers with enough variation between them? If such is the case, how many clusters and what sort of variation is there between the clusters? What sort of diversity do the clusters showcase in terms of the behavioral information collected from the Steam API? What sort of techniques can be suggested to help these heavy gamers spend more time away from the keyboard?

1.5 Scope and Significance of the Study

The dataset[12] was collected in 2013 as part of the complete examination of Steam. Information on around 109 million players were captured using the open Steam API and examined upon. For a Steam user, the dataset has two-week playtime, in minutes, over every game the user owns. This also includes games the user has not played. The current study aggregated the total

[12]Steam.internet.byu.edu. (2020). Steam Dataset. [online] Available at: https://steam.internet.byu.edu/ [Accessed 14 Feb. 2020].

[13]Steam Community: Steam Web API Documentation. Available at: https://steamcommunity.com/dev (Accessed: February 15, 2020).

playtime over two weeks and extracted all the information only on players that appear to be addicted to games. Selecting a cutoff for the playtime over two weeks can be challenging as the reported rates of gaming addiction varies a lot. A literature review by Kuss and Griffiths [16] on the internet gaming addiction suggested 16 hours a day would be considered addictive while another study on players of the massive multiplayer game World of Warcraft [29] showed an average of 9 hours every day for a week (or 63 hours weekly) can have negative impacts on an individual. A study on Dutch adolescents [11] found even 55 hours a week (or less than 8 hours a day on average every day for a week) can be considered addictive. For this study, the cutoff has been chosen to be 6 hours a day on average for two weeks which results into 5040 minutes every two weeks. This appears to be a very safe estimate considering the previous research on video game addiction has reported similar levels of heavy gaming in terms of weekly play times [11]. The study sampled a large enough portion of the players from the extracted dataset since the entire dataset can be resource heavy to handle. The sampled player information was combined, cleaned and then and a clustering algorithm was run to find potential groups of heavy gamers.

There have been many previous studies which found addictive behavior amongst gamers. However, most of the studies have been solely based on surveys where the data has been self-reported. Hence, the samples were smaller in nature and the datasets, being self-reported were open to bias. Steam, being a digital entity records player data on its own accord and hence can be more accurate. This study manages to extract and process information on around 60 thousand players. The clusters found at the end will give us a look into the virtual world of addiction. The study hopes to find potential variation in behavior of the gamers in the clusters. This can be used to prescribe in game and even real-world methods to reduce gaming addiction of the players. This is important as one method may not work on all the gamers as their motives and in game behavior could be different. Many mobile games nowadays have a playtime limit to stop the user from overspending their time on the game itself. This can be enhanced once we have a clearer idea on addiction. Big gaming corporations can use this information to better understand their customers which could help them earn more revenue or help the players reduce their addictive nature.

Steam, like other gaming platforms, may also be able to notify the friends of the heavy gamers about their addictive nature. In one instance, Electronic Arts faced massive backlash for having lottery based microtransactions in its game which caused the company to lose 3 billion U.S.

dollars in stock[14] value and even face investigations[15]. This controversy even caused the Chinese government to impose new rules[16]. The corporations may hence find it best fit for them to help the players to avoid government regulations on the games or garner popularity by helping the addicted individuals.

[14] Kim, T. (2017) EA's day of reckoning is here after 'Star Wars' game uproar, $3 billion in stock value wiped out, CNBC. CNBC. Available at: https://www.cnbc.com/2017/11/28/eas-day-of-reckoning-is-here-after-star-wars-game-uproar.html (Accessed: February 14, 2020).

[15] Bailey, D. (2017) Sadly, the Belgium Government has not yet declared loot boxes gambling, PCGamesN. Available at: https://www.pcgamesn.com/star-wars-battlefront-2/battlefront-2-loot-box-gambling-belgium-gaming-commission (Accessed: February 14, 2020).

[16] McAloon, A. Online games will be required to disclose random loot box odds in China, Gamasutra. Available at:https://www.gamasutra.com/view/news/287258/Online_games_will_be_required_to_disclose_random_loot_box_odds_in_China.php (Accessed: February 14, 2020).

2. LITERATURE REVIEW

2.1 Introduction

Gaming addiction has been studied multiple times in the past with the first incident being reported in 1980. This is a few years after the launch of the first commercial video game 'Space Invaders'. In 1980, three cases of addiction on this game were reported [17]. The first instance of the term 'video game addiction' was used in 1983 by school counselors Soper and Miller [18] who found video game addiction to be similar to behavioral addiction where the person suffered from lack of interest in other activities, compulsive behavior and other physical and mental symptoms. The first study on gaming addiction was done in 1989 by Shotton [21], who found the addicts to be completely engaged in computer games for 5 years. However, there was no use of any measure of addiction and the data was self-reported. Self-reported data can be biased however a research in 2011 by Widyanto, Griffiths and Brunsden [22] revealed that self-diagnosis of addiction (or no addiction) had high correlation with standardized measures of internet addiction. Later in the 2000's, many studies were published on gaming addiction. This is around the same time when the online component was starting to be added in video games and gaming's growth was substantial. Some of these studies collected online data while other relied upon medical examinations such as visual and verbal memory tests [23]. There have been previous studies with focus on game data on particular titles such as Everquest [13], World of Warcraft [29]. Large scale game data analysis is few and far between.

A study on 3389 gamers randomly selected from the National Population Registry of Norway [24] resulted in 1.4% of gamers to be addicted to video games. It also showed 7.3% of the gamers to be heavy gamers, 3.9% to be engaged gamers and 87.4% of the gamers to be normal gamers. The study found many factors that contribute to video game addiction which includes gender, age, lifestyle, country of origin and health. The study [24] found the following factors to be associated to gaming addiciton:

'being of male gender, being young in age, living alone, being born in Africa, Asia, South America or Central America, scoring low on conscientiousness, scoring high on neuroticism, and having poor psychosomatic health'.

Other studies have reported from 0.2% addicted gamers in Germany [31] to 46% in Taiwan [32]. The study [31] in Germany included 4382 individuals aged 14 to even 40 years and found 0.2% individuals to be addicted to games while 3.7% of the individuals were considered to be problematic gamers. This study used 'Gaming Addiction Short Scale (GAS)' which measures salience, withdrawal and conflicts in the individuals to name a few. Other forms of game addiction scale have also been studied [12].

Rooij et al. in their study [11] through a survey on about 1.5% of children aged 13-16 in Netherlands confirmed that there was a small subgroup of children who showed potential addiction like symptoms. The same study through cross-sectional surveys done in 2008 and 2009, found that 3% of the surveyed individuals were addicted to gamers. Along with that, it was found that there were two distinct groups of addicted gamers where one would be addicted heavy gamers, and another would be non-addicted but also heavy gamers. They varied only slightly from each other in terms of psychosocial health and persistence over time. The addicted gamers (aged 13-16 years) reported 55 hours of video game playtime in a week on average. These gamers were also found to be more depressed than the non-addicted gamers which could be caused by relying heavily on online relationships than in real life. Another survey [27] was done on 1178 individuals aged 8 to 18 living in the United States. Addiction like symptoms were found in around 8.5% gamers. According to the study, these individuals showcased at least 6 out of 11 psychosocial problems in real life. This is in conjunction to the study by Tejeiro Salguero et al. [25] where the prevalence rate of video gaming addiction among Spanish adolescents was found to be 9.9%. Another research exhibited that out of 7069 gamers who answered online questionnaires, 11.9% gamers satisfied diagnostic criteria of video game addiction [28]. An online survey was done on players of the famous massively multiplayer game World of Warcraft. This survey on 206 players aged 14 to 65 demonstrated one group of 20 players who were engaged in videogames for around 45 to 82 hours a week [29].

Study [26] concluded that this difference in prevalence rates can be accounted for partly due to the difference in the methodologies assessing video game addiction such as not assessing lifetime prevalence of addiction or other prior problems. This same study also pointed out the lack of a ‘control group’ for between group comparisons. Despite the prevalence rates, many studies now show that excessive gaming can have negative psychosocial consequences such as lack of attention [30,33], increased aggression [34], loneliness [35], stress [30] and even suicidal tendencies [36]. There are long lists of studies that show the negative psychosocial and physical effects excessive gaming can have. A survey done on virtual reality games [20] also concluded that most popular virtual reality games can be addictive due to their immersive nature. Even though most of the studies on addiction have been surveys and based on user reported data, most of the earlier studies have always exhibited small subgroups of gamers that are seemingly addicted to video games.

2.2 Previous Analysis on Steam:

The gaming industry has been the focus of many studies in the past. Some were financial in nature while others revolved around gaming behavior in specific games [37]. Blackburn et al. [38] studied cheaters in Steam and their friendship connections. Steam, as of 2018, had over 150 million players [39]. Orland[17], in 2014, analyzed 250000 user accounts in Steam and gave a descriptive idea on the popularity of games. Becker et al. [40], in 2012, analyzed friendships on Steam using partial data on 9 million users, 1824 games and 88.5 million friendships. They studied friendships and how the network changed over time. Gamer behavior has also been studied extensively with regard to their play times. [41,19].

Condensing Steam: Distilling the Diversity of Gamer Behavior [10], was the first complete examination of any major gaming network, which here is Steam. The study consisted of 108.7 million user accounts and 384.3 million owned games at the time of the research (2013) when the dataset was collected. Thanks to this previous work [9] on Steam the dataset[12] is publicly available for non-commercial use. The dataset was collected from the Steam API service and hence was not self-reported. It could provide a broader picture because the dataset has information on types of games played, the price of games, gamer age in the platform to name a few. This analysis delved into overall gamer behavior and gamer stereotypes. The results of the analysis showed that even though most gamers are casual gamers, the outliers have extreme and abnormal behavior. Along with that, the overall gamer behavior was found to be very diverse among the millions of gamers who have been studied. Gaming addiction was not an objective of the study. However, the study found that the top 1% of the gamers, which happen to be over a million players, played more than 5 hours of day.

2.3 Summary:

There has been a significant increase in research on gaming addiction over the years. In the 1980s, the research was done on arcade games, in 1990s the research focused on offline video games and since the 2000s most research focused on massive multiplayer games. Almost all the previous work done on video game addiction focused on the individual's social life. The datasets collected in such cases were self-reported and hence were smaller. Since the data was self-reported, the observations were open to bias and did not represent the overall population. Along with this, the methods used to identify video game addiction were inconsistent and non-

[17] Orland, K. (2014) Introducing Steam Gauge: Ars reveals Steam's most popular games, Ars Technica. Available at: https://arstechnica.com/gaming/2014/04/introducing-steam-gauge-ars-reveals-steams-most-popular-games/ (Accessed: February 14, 2020).

standardized and described the gaming addiction from the perspective of the individuals themselves. The current study intends to look into gaming data captured on the Steam platform which would not be biased since it is captured by the platform itself. Steam requires the internet and monitors the user's activity. This information is sent over the internet to Steam where they store the information. If the user is offline, the data is sent when the user connects back to the internet. This data would give a better understanding on video game addiction than self-reported surveys.

3. ANALYSIS AND DESIGN

3.1 Introduction

Data from the previous work [14] is in a compressed SQL file of 17GB. When uncompressed the size of the file is over 170 GB. This data consists of data of over 109 million players including their game play information. Data also contains information on the games and players on steam.

3.2 Dataset Description:

A summary of the detailed description of the data is mentioned below:

1. ACHIEVEMENT_PERCENTAGES: Contains achievement information of all the applications/games. Also contains achievement names and the percentage completion of each achievement for each of the games by the players in the Steam network.
2. APP_ID_INFO: Contains information on the Applications such as games. This includes the name of the application, price, release date of the application, required age to install as well as its rating and whether it has multiplayer capabilities or not.
3. FRIENDS: Contains a reciprocal list of friendship of the steam users, along with the time the users became friends in the network.
4. GAMES_1, GAMES_2: These two contain game data on the user's gameplay such as play times for each game at the time of the collection. The study would use GAMES_1 which was collected in 2013.
5. GAMES_GENRES: Genres of each game. One game can have multiple genres.
6. GROUPS: Lists the group memberships of the users. Contains list of groups each user is part of.
7. PLAYER_SUMMARIES: Contains profile information on the players such as account creation date, location (if given) and last logoff time.

3.3 System and Data Preparation:

The dataset is of considerable size (over 170 gigabytes when uncompressed) and requires a machine with high specifications or a cloud subscription. The data being a compressed file (structured query language file) needs to be extracted and imported into a MySQL database. Once the data is imported into a MySQL database, the tables containing the dataset can be accessed using the database shell.

To start the dataset preparation the uncompressed steam.sql.gz file which was downloaded from the link was uploaded into Google Cloud Storage[19] which is a service provided by the cloud subscription service Google Cloud[18]. Access to Google Cloud requires a subscription. This study used the trial services provided by the service. The data from Google Cloud Storage[19] was subsequently loaded into a MySQL database provided by the Google Cloud SQL service using the user interface provided by the web application. The firewalls and ports were modified to allow seamless interaction with the database. This process can be imitated in a local machine provided the resources are sufficient to handle a data this large in size.

3.3.1 Data Extraction:

Data Extraction was done using a Google Compute Engine[20] virtual machine instance which was prepared by installing all the necessary software such as Python, Jupyter Notebook and the required libraries to connect to a MySQL instance. The firewall configurations were modified to connect the Jupyter notebook (from the compute engine) with the MySQL database.

Since the research focused on heavy gamers, only the information on the heavy gamers were extracted. These gamers have shown high playtimes on average over 2 weeks leading up to when the data was originally collected during the previous research work [9]. The cutoff for which was selected to be over 6 hours a day, which would be in accordance with other studies that exhibit addictive behavior in the range of 45-80 hours a week. The 6-hour cutoff for two weeks comes in at 5040 minutes in two weeks.

All the information collected on the users by the previous research is publicly available for non-commercial use and collected over an open API provided by Steam. The information does not collect any personally identifiable information other than real name and country of location, both of which are optional and unverified. The users who showcased over 5040 minutes of playtime in two weeks, were extracted. This included their game play data, friends and group information as mentioned in the data dictionary above. The entire information on APP_ID_INFO, ACHIEVEMENT_PERCENTAGES and GAMES_GENRES were extracted as it holds application related data and not player level data. We would also extract the count of players in each group for analysis.

---

[18] Cloud Computing Services. Google. Available at: https://cloud.google.com/ (Accessed: February 14, 2020).
[19] Cloud Storage: Object Storage | Google Cloud. Google. Available at: https://cloud.google.com/storage (Accessed: February 14, 2020).
[20] Compute Engine: Virtual Machines (VMs) | Google Cloud. Google. Available at: https://cloud.google.com/compute (Accessed: February 14, 2020).

3.3.2 Data Preparation:

The extracted datasets were kept in Google Drive[21]. The drive was mounted in Google Colaboratory[22] notebook using Python and the required libraries. This allows the notebook to have direct access to the files when the drive is mounted.

The following steps were taken to prepare the Master dataset. All the code is available on a repository in GitHub[23].

3.3.2.1 Sampling

The extracted data contained game play information on over 250000 accounts such as play times on every game owned by each of those accounts, their friends, group participations etc. Due to the sheer size of the data the current study used sampling to reduce the volume of the data. This was also done as studying the entire population may be impractical. Sampling will also significantly reduce the resource requirements to process the information.

The study decided to use simple random sampling over the 250000 accounts where each user will have an equal probability of getting chosen. The extract of the game play information in GAMES_1 table was ideal to run the sampling method on as the game play information was our primary source of information. The study decided to pick 125000 accounts using the simple random sampling method, however all the datasets do not contain information on all the players at the same time. After merging the datasets, the final total number of accounts were close to 60000 players. This was due to the same player information not being present in all the tables at the same time. The sampling was not done after the merging of the datasets as the preprocessing of the data will consume a significant amount of processing power.

3.3.2.2 Validation Checks

Basic validation checks included duplicate checks of all the datasets. Duplicates were also checked by temporarily removing date columns and other columns depending on the dataset being validated. Overall data quality was validated by checking the data types of the columns of the datasets. For example, the primary key ‘SteamID’ can only be numeric in nature. Other checks included, whether the datasets are from a single timeframe. The total playtime in two

[21] Google Drive. Google. Available at: http://drive.google.com/ (Accessed: February 15, 2020).
[22] Google Colaboratory. Google. Available at: https://colab.research.google.com/ (Accessed: February 15, 2020).
[23] Thesis Code, GitHub. Available at: https://github.com/in6tinct/thesis (Accessed: February 15, 2020).

weeks was also validated to confirm whether the extraction was done properly. Below is the list of basic validation checks on the following datasets:

1. ACHIEVEMENT_PERCENTAGES: Overall duplicate checks and duplicates excluding ‘Percentage’ to check if duplicate achievements exist. Null checks were also done.
2. APP_ID_INFO: Duplicate checks and checking if the same app id exists in more than one row. Null checks in the columns were done.
3. FRIENDS: Duplicate checks overall and checking if the same relationship is present more than once. Checking if ‘steamid_a’ is the same as ‘steamid’ and also overall null checks. Friendships established before 2009 were given a default date (1969-12-31 17:00:00).
4. GAMES_1, GAMES_2: Duplicate and null checks. Checking unique date retrieved values. Verifying if ‘Playtime_2weeks’ is over 5040 as extracted and also if two week playtime is greater than forever playtime. All the SteamID showing two weeks play times greater than their forever playtimes were discarded.
5. GAMES_GENRES: Duplicate checks and null checks. Checking for multi genre games. Many games were found to have multiple genres.
6. GROUPS: Overall duplicates and checking if a SteamID for a single group has been recorded more than once or not.
7. PLAYER_SUMMARIES: Duplicates and null checks. Each row should represent one unique SteamID.

3.3.2.3 Elimination and Transformation of Variables

A few of the columns were removed from the extracted data sets due to them not being important for analysis, such as avatar links, their current activity state (online or offline) etc. to name a few. Other column elimination was done as they contained too many nulls. Below is the list of eliminations and transformations:

1. ACHIEVEMENT_PERCENTAGES: Achievement ‘Name’ was not needed. ‘Percentage’ was averaged out for each app id so each row consists of the app id and average percentage completion of all the achievements of the application which would be an estimate of how difficult the game is.
2. APP_ID_INFO: ‘Title’ name is not needed. ‘Rating’ was absent (-1) for a lot of applications to indicate these applications have not been rated. These negative ratings were transformed into the median value of the non-negative rated games.

3. FRIENDS: ‘steamid_b’ is not needed as a friend's steamid is not important. ‘Relationship’ is only ‘friend’ hence not needed. ‘Lcctag’ is null and ‘steamid’ is the same as ‘steamid_a’ hence both are eliminated.
4. GAMES_1, GAMES_2: ‘Dateretrieved’ not needed. ‘Playtime_2weeks’ and ‘Playtime_forever’ values were transformed to 0 if it was Null to avoid aggregation issues.
5. GAMES_GENRES: One application can have multiple genres. Hence, genres were transformed into dummy columns and the dataset was brought into the app_id level where each row represents a single app_id.
6. GROUPS: ‘Dateretrieved’ not needed.
7. PLAYER_SUMMARIES: ‘personaname’, ‘profileurl’, ‘avatar’, ‘avatarmedium’, ‘avatarfull’, personastate’, ’communityvisibilitystate’, ’profilestate’, ‘primaryclanid’, ‘gameid','gameserverip','gameextrainfo','cityid','loccountrycode','locstatecode','loccityid','dateretrieved’ not needed as these do not provide much useful data. ‘realname’ was transformed into 0 if users real name is not given and 1 if given.

3.3.2.4 New Features

New features were added to create multiple more stable KPIs in conjunction with elimination of the variables.

1. ACHIEVEMENT_PERCENTAGES: New feature ‘total_achievements’ was added which gives an idea on how many achievements a game has, which estimates the depth of a game.
2. APP_ID_INFO: ‘Is_Adult’ feature was added if the required age to play the game was 18 years old. ‘Game_age’ was created by deducting the retrieval date with the release date to show how old the game is.
3. FRIENDS: ‘friend_age’ feature by deducting ‘dateretrieved’ from ‘friend_since’ to show how old a friend is. The data was brought into the ‘steamid_a’ grain by averaging the ‘friend_age’ for each player and the ‘no_of_friends’ feature was also added to show how many friends a player has
4. GAMES_1, GAMES_2: ‘no_of_games’ to show how many games a player has. ‘Unplayed’ to show if a game has been played or not (‘playtime_forever’ would be null or 0 for unplayed games).
5. GAMES_GENRES: ‘free_to_play’ was added if an app id had a price set to 0.

6. GROUPS: Count of groups and 'total_group_connections' were added by grouping with steamid and merging with the 'groups' dataset to indicate the reach of the groups. Total group connections would indicate how many people are in the groups the player is part of.
7. PLAYER_SUMMARIES: 'logoff_duration' in days calculated by deducting dateretrieved and 'lastloggoff' to show how many days ago the player last logged off.

3.3.2.5 Merging of Datasets

All the datasets have either SteamID as a unique identifier for the players or AppID as a unique identifier for the Apps (games in this case). Playtime dataset has both. All the datasets were merged accordingly to bring the entire dataset into the SteamID grain i.e. each row in the final dataset would represent the entire information on a single player only. The genre-based information was also collected for each user. The final dataset had the following information:

- Steamid, which is the Primary key that identifies each user
- Commentpermission, 0 or 1 to indicate whether the profile allows public comments.
- Realname, 1 if player has provided a realname else 0.
- Playtime_2weeks, two-week total playtime in minutes for all games for each player which is greater than 5040 minutes in this study.
- Playtime_forever, total forever playtime in minutes for each player considering all games.
- No_of_games: Number of games owned.
- Free_to_play: Number of free to play games.
- Price: Overall money spent for each player in dollars. May indicate the players financial status.
- Rating: Average rating of games owned (A lot of the missing information was modified by the median of the games that are rated on the steam platform)
- Required_Age: Average required age for the owned games by each player in years. May be an indicator of the age of the gamer.
- Is_Multiplayer: Number of multiplayer games owned.
- Is_Adult: Number of adult games owned.
- Unplayed: Number of unplayed games.
- Game_age: Average age, in days, of the games owned by each player.
- No_of_groups: Number of groups the player is a part of.

- Avg_friend_duration: Average age in days of friendship for all the friends a player has. This will indicate if player makes too many new friends.
- Avg_percentage_completion: Average percentage completion of all the achievements for all the games owned by a player. Indicates overall difficulty of the games owned.
- Total_group_connections: Total players in the all groups a player is part of. This will indicate if player joins the biggest groups.
- Total_friend_connections: Number of friends a gamer has.
- Total_achievements: Total achievements for all the games owned by each player.
- Loggoff_duration: Indicates how long ago the player last logged off, in days.
- Account_age: Age of the player account in days.
- Type_demo: Count of demo games owned (these are normally free).
- Type_dlc: Count of DLCs (downloadable content) owned. These can be free or priced but is usually an add on over the current game.
- Type_game: Count of games.
- Type_mod: Count of mods, mods are also add on packs but they modify the experience of the game and do not add any extra content like DLC does.
- Type_Unknown: Count of games where the type is not known.

The following information was derived from the Genres database. Since one game can have multiple genres, the total of the following Genre based dummy variables do not equal to the total. However, they are a good indicator of the players inclination in each gaming genre.

Genre based dummy variables were created such as Genre_Action, Genre_Adventure etc to indicate count of games in each genre. Other dummy variables were also created to echibit two-week playtime for each genre in minutes and all time total playtime for each genre in minutes.

3.3.2.6 Outlier detection and removal

The study found the presence of many outliers. This could be the result of bugs in the Steam API or significantly addicted individuals that may skew the information. A boxplot of the Playtime forever information when aggregated shows many outliers. This could be a result of outliers such as very old players in the account who have been accumulating gameplay times since the launch of Steam.

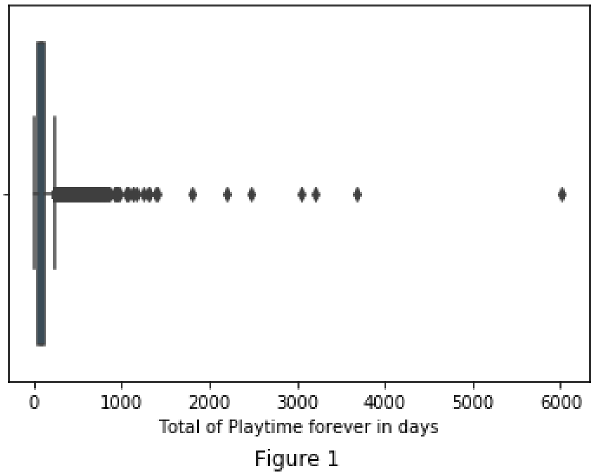

Figure 1: Boxplot of total overall playtime in days

This figure (1) shows many outliers with respect to total aggregate play times for the players.

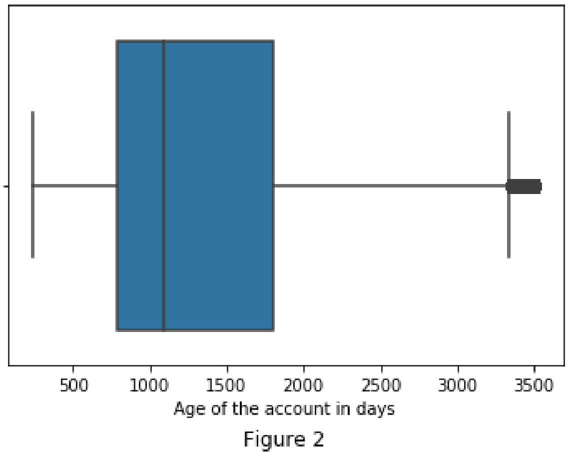

Figure 2: Boxplot of age of the players account in days

This figure (2) shows some outliers however the total time spent in the steam platform by the heavy gamers which varies from 500 days to all the way to less than 10 year. Perhaps due to the same reason, there was the existence of outliers who buy huge amounts of games and as a result had significantly higher monetary expenditures as shown in the following figures 3 and 4.

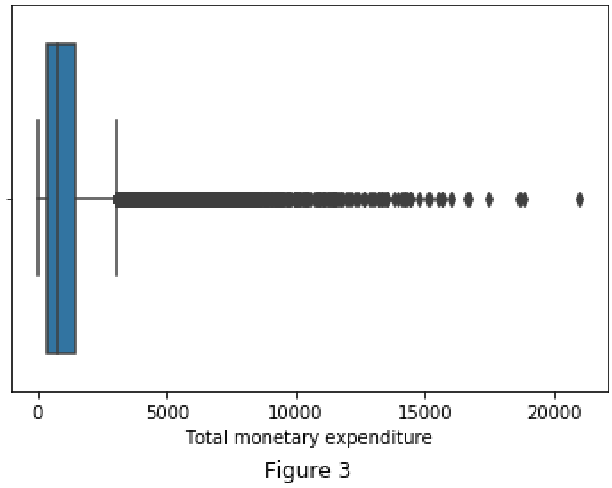

Figure 3: Boxplots of total monetary expenditure in U.S. dollars by the heavy gamers

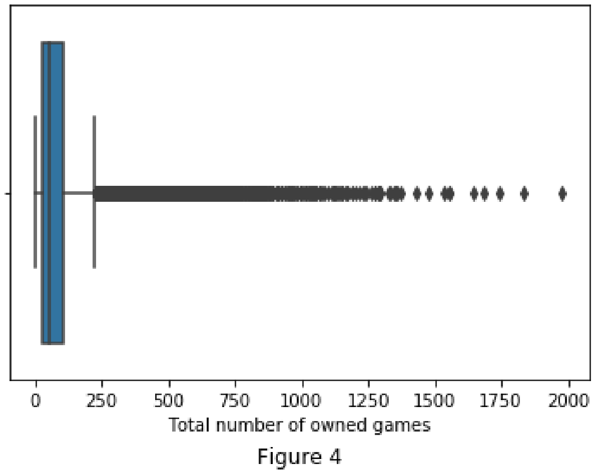

Figure 4: Count of owned games by the heavy gamers

The study also found players with very high (and low) social connections on the platform. Typically, a user cannot have more than 250 friends. This limit is increased to 300 friends if the user connects their Facebook account. However, we found over 500 players for whom the limit was over 300 which is shown in figure (5).

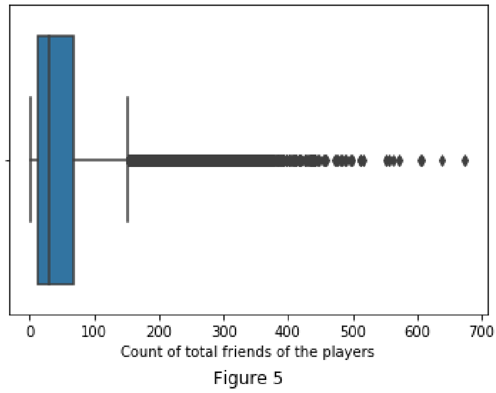

Figure 5: Boxplots of total count of friends of the heavy gamers

Along with friendships, the number of group participations also showed many outliers as shown in the following figure (6).

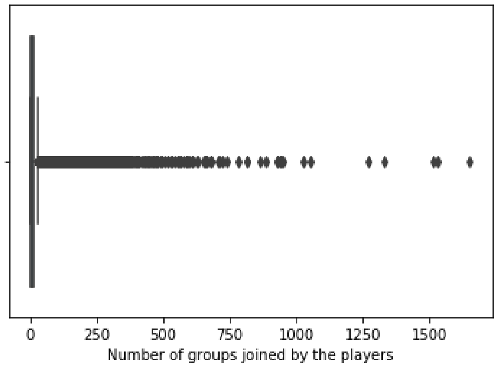

Figure 6: Total count of participated groups

The study found the distribution of the attributes to be mostly similar to a normal distribution. The distributions hence had most the values to be clustered around the median. The outliers are present away from the median. To detect these outliers, the interquartile range (IQR) rule was applied. This is a simple and useful way of detecting outliers. IQR can tell us how spread out the values are, and also tell us which values are far enough from the median to call it an outlier. Interquartile range (IQR) is the subtraction of the third quartile (Q3) and the first quartile (Q1). The interquartile range (Q3-Q1) is then multiplied by 1.5 which is a constant used to detect outliers. This multiplied value (1.5 times IQR) is then added to the third quartile to get the upper level of outliers and it is subtracted from the first quartile to get the lower level of outliers. Any value outside these limits are considered as outliers in this study.

For example, the account age and monetary expenditure of the players showcase a normal distribution with outliers in the following figures (7) and (8).

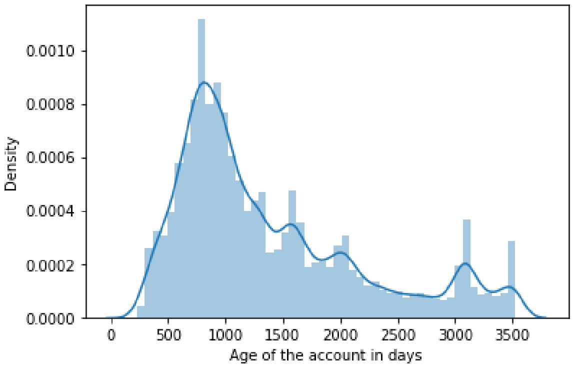

Figure 7: Distribution plot of account age in days

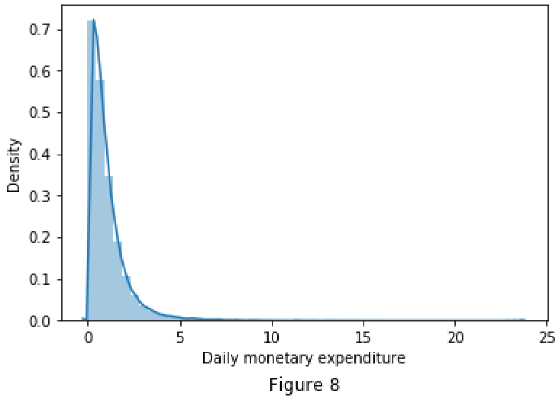

Figure 8: Daily monetary expenditure in U.S. dollars

The outliers, rather than removing them, are transformed into the closest outlier limit. Hence, any value greater than the upper limit is assigned the upper limit value and any value lesser than the lower limit is assigned the lower limit value. Post outlier treatment the boxplots do not show any outliers on the columns. The figure (9) and (10) below shows a boxplot and a normal distribution of the forever play time post the outlier treatment.

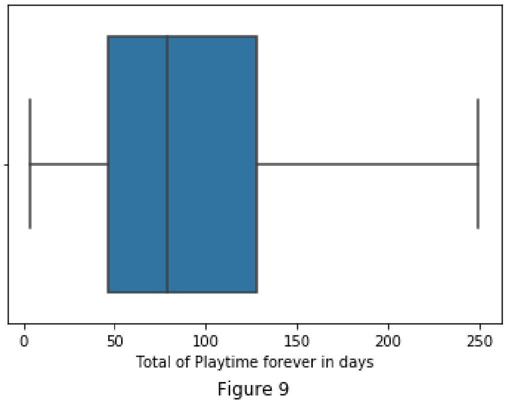

Figure 9: Boxplot plot of total overall playtime in days after the outlier treatment

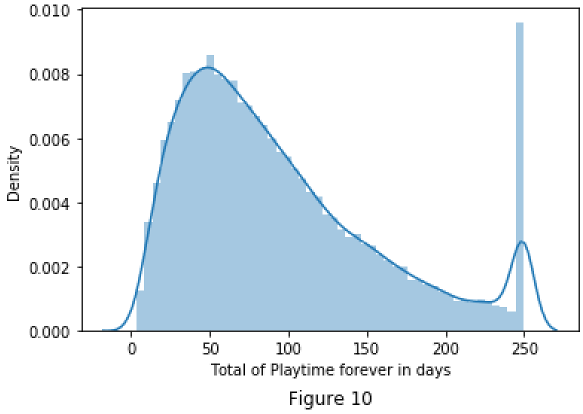

Figure 10: Distribution plot of total overall playtime in days after the outlier treatment

3.3.2.7 Final Dataset Preparation:

Some new features were created post outlier treatment. Also, some transformations were made to help with the clustering. This will help stabilize the clustering model. To prepare the final dataset for the clustering, we chose the following attributes:

Attributes: These attributes represented overall information on the players such as,

- 0 or 1 to denote whether the player has chosen to provide a name in the realname field.
- Two-week playtimes.
- Number of games of the players and the average rating of these games the player owns.
- Average age of the owned games and average required age to play them.
- Social connections such as number of friends, average age of the friendship and total number of groups the player is part of.
- The average number of achievements available in the owned games to denote the depth of the game, and the average overall percentage completion of these games (by the players of the steam network) to denote the difficulty of the games.
- Age of the players account and how long the player has been logged off up to the time of the data collection.
- Average daily monetary expenditure of the player and the number of unplayed games of these players.
- Out of the total number of games the player owns, what percentage of the games are free, multiplayer, adult (requires age of 18 or older to play) and unplayed.

3.4 Data Analysis:

3.4.1 Introduction

An overall exploratory data analysis was done on the dataset to get basic idea of the dataset.

Comparing the games, it looks like Action/Adventure games are priced average while Sports games are the costliest.

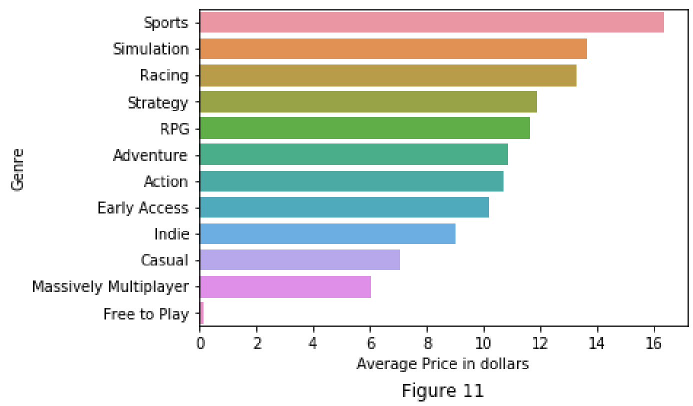

Figure 11: Bar chart of Average price of games in each genre

Massively multiplayer games cost the least as it may have many in game items for purchase. They are also mostly free to play. The dataset contains ‘Free to play’ as a genre hence that is also appearing in the below chart.

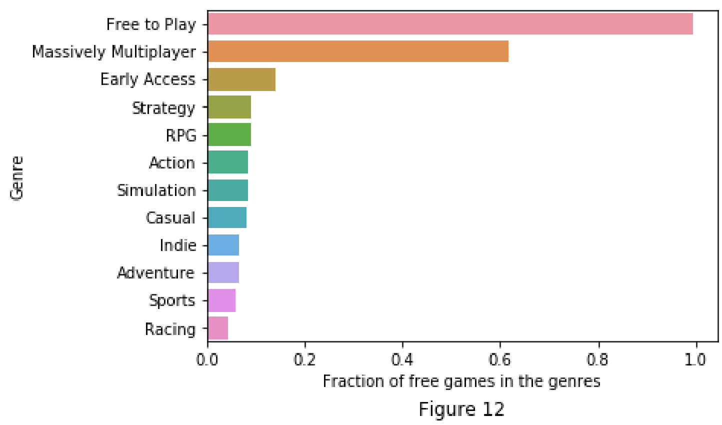

Figure 12: Bar chart of fraction of free games in each genre

As for playtimes, according to the previous study [10], Action games by far the most popular followed by Strategy, Free to play and RPG games. Similar results were obtained for all time and two-week total play times for each genre. However, the study found that these addicted gamers play more RPG games than the overall community with RPG being the third most played game for addicted gamers than the overall community where RPG was in the 4th place.

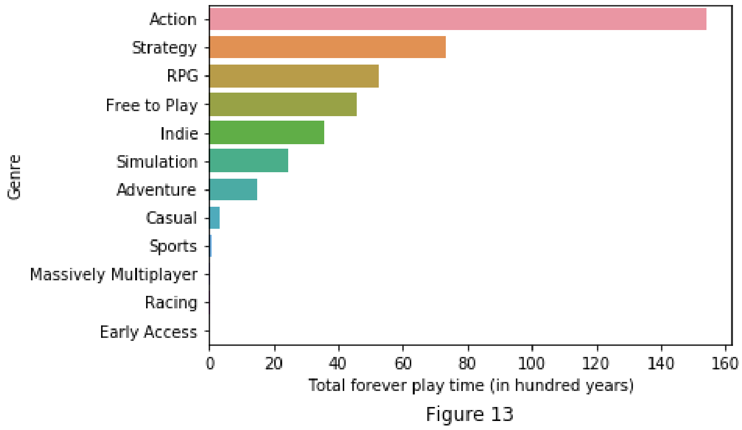

Figure 13: Bar chart of total time in 100 years for each genre

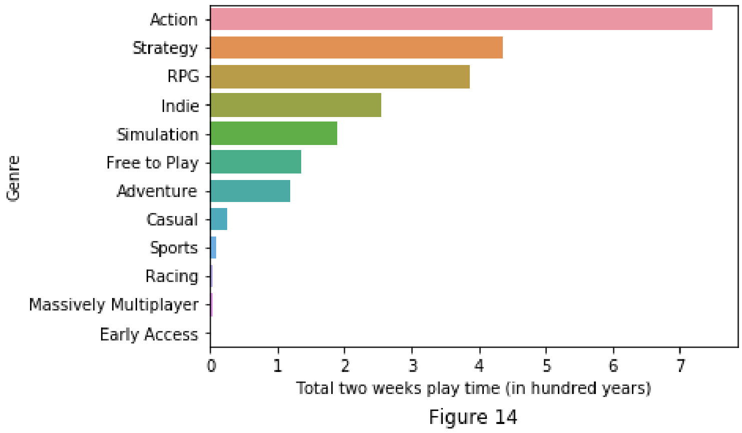

Figure 14: Bar chart of two-week playtimes in 100 years for each genre

As for the Achievement information, Casual games have the most completions, as easier games would. Massively multiplayer games would be the hardest due to competition from real players and not in game characters.

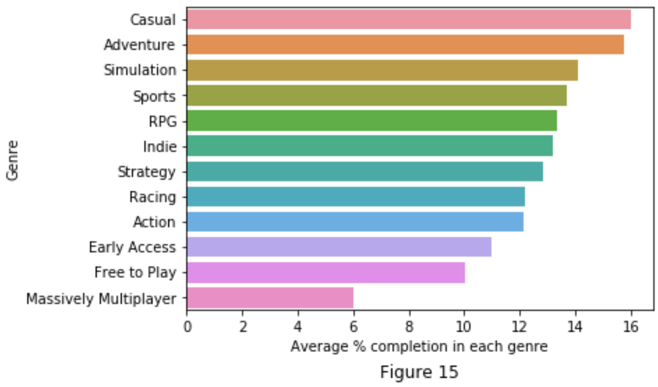

Figure 15: Bar chart of average percentage completion of the achievements for each genre

Massively multiplayer games, being mostly free or priced lower with many in game purchasable items, would also have the most achievements as they require the players to stay engaged.

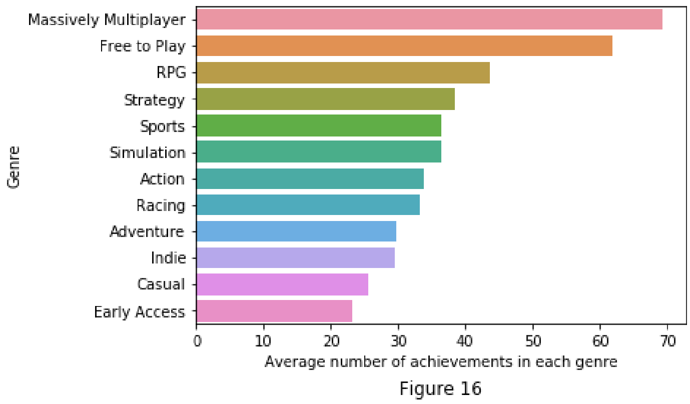

Figure 16: Bar chart of average number of achievements for each genre

3.4.2 Player profiling:

The second form of data analysis on the players was chosen to be random analysis on a few players based on social and economic attributes like monetary expenditure, two weeks or overall playtime, number of unplayed games and social connections (friends and group connections). This process was followed to get some basic ideas on the players to manually check for types of addicted gamers.

3.4.2.1 Money Spent

Two random players were analyzed who have high and low overall money spent on their profiles. Comparing their overall data, it was found that the player who had spent over 14000 dollars over the lifetime had a few significant differences from the player who had spent less than 300 dollars despite them having the Steam account for around the same amount of time.

The player who had spent more money when compared to the later had:

- More paid games and even more free to play games. The high spender had over 1200 games with 19 free games. The low spender had 20 games and 2 free games.
- Longer two weeks play times but lower overall playtimes which was almost half of the low spender.
- Played games that are of lesser quality, lesser age limits and older games. Also has a lot of unplayed games. This player had not played around 83% of the owned games. This player could be a ‘collector’ who loves to collect the games and has high financial capabilities.
- Played mostly RPG and Action games. Had very low play times on free games.
- Is part of fewer groups and has fewer group connections. The player also has overall older and fewer friends in the network.

- Plays more diverse games.

3.4.2.2 Daily average minutes played

Next, we analyze two players who have overall spent high and lower amounts of time playing games on the platform while having the account for about the same amount of time. High being 400 minutes on average daily and lower being 30 minutes on average daily. Interestingly, these behavioral differences between the heavier (with respect to daily play times) gamer and the comparatively more casual gamer showcased similar difference in attributes to the players with high and low monetary expenditure. Just like the high spender in the previous analysis the player with higher average daily play time was:

- Higher monetary expenditure. (over 2300 dollars vs less than 300 dollars)
- More owned games and around 34% unplayed games.
- Had fewer multiplayer games. 37% of the games were multiplayer games while the more casual player had 70% multiplayer games.
- Played more free games and lesser quality games.
- Had more friends and they were also older friends despite both having the account for a similar amount of time.

3.4.2.3 Unplayed percentage

Keeping the number of games more or less in similar scale (180-250 games) and also keeping their account age almost similar, two players were analyzed who had high and low unplayed percentages. Player 1 had around 90% of the games unplayed while Player 2 had around only 12% of the games unplayed. Comparing Player 1 and Player 2 it was found that,

- Player 1 had much higher two-week playtimes (slightly over 12000 minutes vs Player 2's over 6000 minutes). Player1 had played an hour of games every day on average while Player 2 had played 90 minutes average every day. This is significant considering the players had the account for around 10 years (more or less since the creation of Steam)
- Player 1 is part of lesser number groups (2 vs 27) however has more friends most of which are newer friends.
- Player 1 plays mostly Action games while Player 2 had more diversity in terms of their choice of games to play.

3.4.2.4 Social connections

Two SteamID were selected where Player1 had high number social connections i.e. Player 1 was part of a lot of groups and had more friends than the 2nd SteamID (Player 2).

Keeping the account age, number of games, unplayed games percentage similar, it was seen that.

- Player1 had more games and had spent more money.
- Player1 plays more free games and multiplayer games and also easier games.
- Player1 had played less amount of time daily on average (around half an hour vs Player 2's close to an hour daily)

3.4.3 Summary:

It was found through the overall data analysis and the basic profiling that there are mostly two types of addictive behavior. One group of players are more consistently active than the other group. They spend more money yet do not play a good portion of the games they have bought. These players also try out or play various types of games, have more social connections.

3.5 Clustering:

Clustering is a popular way to analyze and get some idea on the structure of the datasets. Clustering helps in identifying subgroups in the data where the data points in the subgroup are very similar while the subgroups show diverse differences among them. Clustering is an unsupervised technique used on non-labelled data points. The study will apply a clustering algorithm to find potential groups of gamers showing distinct diversity among them. Since the groups are not known in advance an unsupervised technique such as clustering will pick the groups. These groups will show significant variation between them in terms of the attributes in the dataset such as monetary spend, playtimes, social connections to name a few.

3.5.1 K-Means Clustering:

This clustering algorithm was chosen to be K-Means algorithm.

K means clustering is a popular and simple clustering analysis algorithm that uses the partitioning method of clustering. K-Means Algorithm aims to partition the observations into K clusters with the nearest mean. K means aims to decrease inter cluster Euclidean distance between observations while increasing intra cluster Euclidean distances. The parameter 'K' needs to be tuned where the value of 'K' needs to be determined. Other clustering algorithm techniques can use density based, hierarchical based or grid-based methods. The current study decided to go with K means clustering as it is inherently simple, and the number of clusters can be manually entered into the code. This helps us to check the clusters before selecting the ideal number clusters.

3.5.2 Hyperparameter Tuning (K):

For finding the optimal value of K, the elbow method was chosen. The elbow method is the common method for identifying the right value of K i.e. the number of clusters. The method is followed by plotting the sum of squared distances for the values of K in a certain range. The elbow, which would be a point in the graph where the decrease in the sum of squared distance with respect to the increase in the number of K would not be large enough compared to the decrease in sum of squared distance in the previous instances of increase in K. This would be a point in the plot where increasing the number of clusters would not provide enough increase in heterogeneity among them to justify the increase in the number of clusters. For the elbow method, taking a range of 1 to 21, the sum of squared distances was plotted in a graph for each value of K in the figure (17) below.

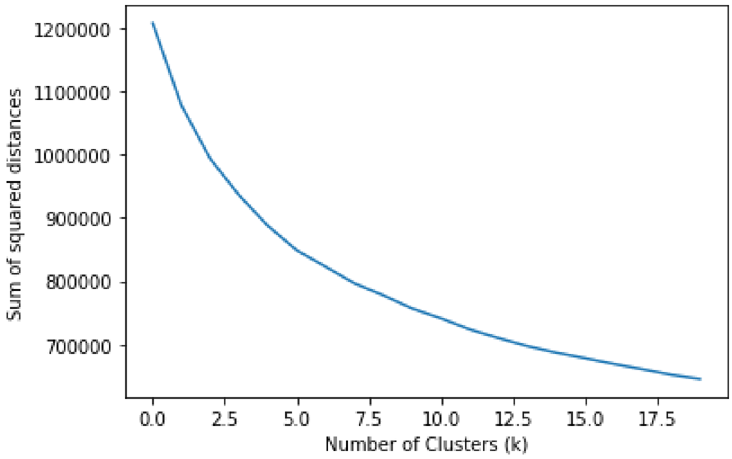

Figure 17: Line plot of Sum of squared distances of the points vs the values of K

Although the ideal value of K using the elbow method looks to be somewhere around the 5 to 7 mark, the elbow method does not always work well when the data is not very well clustered to begin with. The study analyzed 4 clusters and found the clusters to be somewhat insignificant for the real world in terms of diversity among them. These clusters despite showing enough variation among them would not translate well into the real world of addiction. Hence, we decided to pick K=2 as the cluster means showed enough variation among them to justify the two cluster’s existence. This will also help the study recommend easier rehabilitation techniques for the gamers.

4. RESULTS

4.1 Introduction

The clustering done on ~60 thousand heavy gamers revealed two groups of heavy gamers. The clusters, going forward to be referred to as Cluster 1 and Cluster 2, have ~38 thousand and ~22 thousand players respectively. The study dives deeper into the clusters by looking at the social and gameplay behaviors of the heavy gamers found in each cluster. The current study looks at the results and compares it with the previous studies done on Steam. This helps to compare the heavy or addicted gamers with the overall population. This is also accompanied by the comparison of the clusters and their overall behavior which helps us understand what sets both the clusters apart.

4.2 Analysis

4.2.1 Game Ownership

Steam being a digital distribution service mostly sells digital games and other content. The gameplay dataset extracted using the Steam API from the previous study contains play time for every game a user owns for two weeks and all time. According to the previous study done on the entire Steam community it was found that the 80th percentile of ownership is 10 games owned and 7 games played. Here the community consisted of casual all the way to heavy gamers. This current study focused on heavy gamers when compared directly with the previous study, found the 80th percentile is 46 games owned and 32 games played. Even the 50th percentile to be more than the 80th percentile of overall ownership, with the median being 11 games owned and 8 games played. This is a massive increase over the casual players on the platform.

The current study also looked at the percentage of unplayed games for the users. 50 percent of the heavy gamers had not played over 16 percent of the owned games, while the top 10% had not played over 50% of the games. This is in accordance with the previous study where it was believed that some players may have behavior similar to that of a ‘collector’ where the goal of the player is the acquisition of the game only. Other aspects such as game bundles may also account for the unplayed games in the users account.

Comparing the clusters, Cluster 1 and Cluster 2 had a massive difference in terms of game ownership. Cluster 2 looks to be more addicted to owning games with the median of ownership being 128 games owned and 77 games played while Cluster 1 has the median of ownership at 34 games owned and 24 games played. Cluster 2 seems to be much more addicted to owning and playing video games and appears to be more of a 'collector'. This can be seen in the following figure where the overall percentage of unplayed games with respect to year of release of the games were plotted. Here the difference between the groups is more obvious. Post 2005, the difference in unplayed games percentage is increasing as Cluster 2 is buying more of the latest games but not playing them. The difference is also seen in 1970 which can be the result of acquiring vintage games by players in Cluster 2.

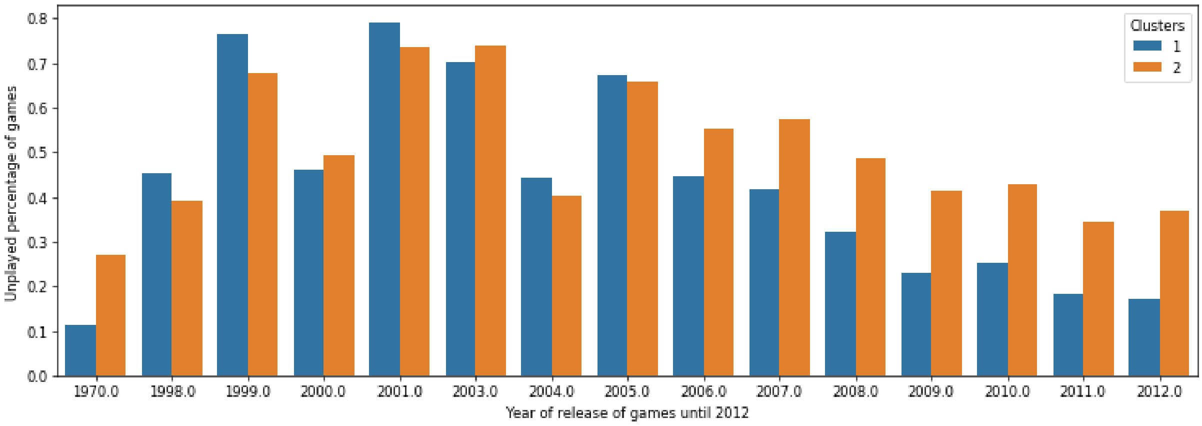

Figure 18: Fraction of unplayed games with respect to the year of the release of the games for each Cluster

The study found a very strong correlation between the total number of owned games and the number of unplayed games. 50 percent of the users in Cluster 2 had not played over 35 percent of their total owned games. This number in Cluster 1 is 24 percent, which is also significant considering the total number of ownerships is much less in Cluster 1 than Cluster 2.

The study looked into the genres of the games owned by the heavy gamers. One game or application in Steam can have multiple genres. Showing genre affinities as a fraction will give us an accurate overall look into the ownership of games by the clusters. Looking at fraction of owned games for each Genre, we found Cluster 1 has a bigger fraction of free games in their overall games catalogue. This is understandable as Cluster 1 also buys fewer games and leaves less games unplayed.

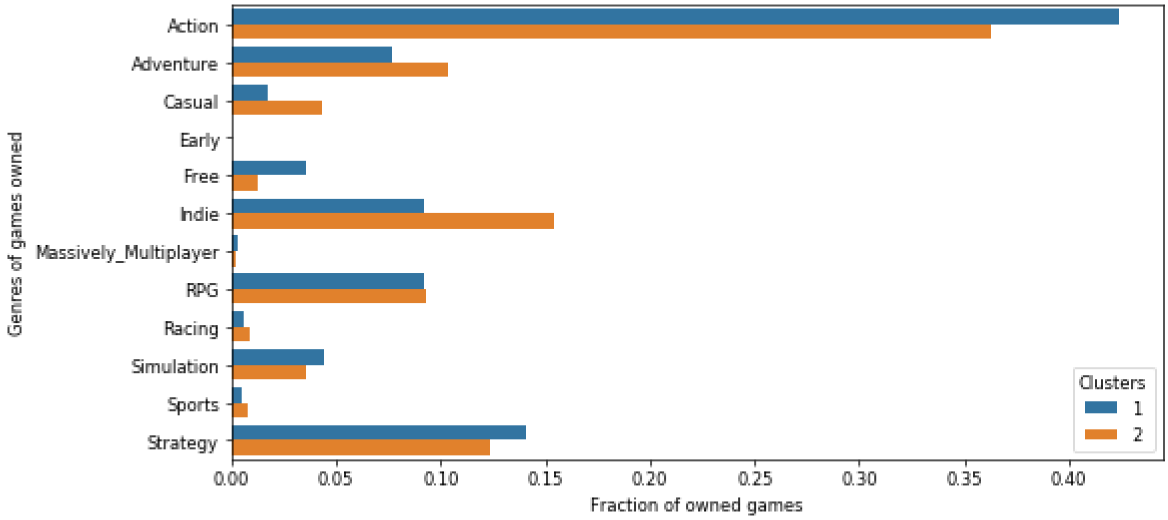

Figure 19: Fraction of owned games in each genre for both the clusters

4.2.5 Monetary Expenditure

Monetary expenditure of the heavy gamers is an important aspect of their addictive behavior. Steam API does not provide transactional data of the games bought. The dataset on the applications has only the market value of the games instead. This will not give a precise idea on the amount spent by the heavy gamers (or any user in the steam) since the player may have bought the game during a sale or during any other price reduction.

Looking at the overall monetary spend of the heavy gamers, the study decided to exhibit daily spending of the players. The daily spending or daily average monetary expenditure considered the total money spent with respect to the market prices of the games owned and the account age of the players. The 70th percentile of all the sampled gamers were found to be just over 1 U.S. dollar a day.

The study compared the clusters and their daily monetary expenditure and Cluster 2 was shown to be more addicted to buying new games. This group had accounts that were over 8 months older on average than Cluster 1 and had spent more money daily on average than Cluster 1. In this study, Cluster 1 was seen to spend around half a dollar every day on average since their account was created on the platform whereas Cluster 2 had spent a dollar more than that. As a result, the total average monetary expenditure of Cluster 2 was much greater than Cluster 1.

The fraction of the money spent in each genre of games for the individual clusters were plotted. This figure (20) shown below only considered the unplayed games to give us an idea of the fraction of money unutilized (with respect to total money spent in each genre) by the player in each genre.

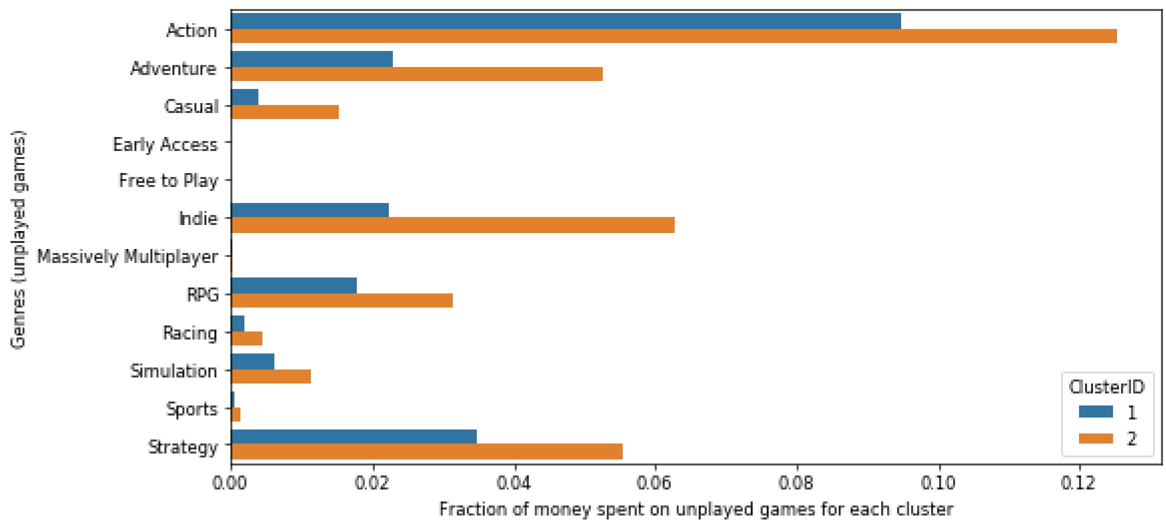

Figure 20: Fraction of money spent (with respect to total money spent in each genre) on unplayed games overall for each cluster in each genre

Comparing this with the previous figure (19) it is seen that Cluster 1 owns a greater fraction of Action games than Cluster 2 but ends up wasting less amount of money on the same genre by not playing them. Similar figures can be seen in the Strategy genre. Cluster 2 ends up spending more money in unplayed games in each genre of games.

4.2.2 Friendships

Steam allows the users to befriend each other like a social network. This feature allows the users to connect to each other. The friends can engage in text chats and also play games together or against each other depending on the multiplayer components that are developed in the game. The previous study on steam found 196.37 million friendships [10]. Steam did not record friendships before September 2008 hence the dataset does not record friendships from 2003 to 2008.

The current study focused on heavy gamers contained around 10 million friendships. The friendships are bidirectional as the dataset contains relationships between the players.

According to the previous study [10], 'The average number of friends in the entire network is four. However, on average, 9 million gamers (88.06%) add ten or less friends per year and only 2,500 gamers (0.02%) add more than two hundred friends per year'.

In this study we looked at the number of new friends being added each year by the heavy gamers. It was found that over 50 percent of the heavy gamers add ten or less friends per year and 0.02 percent of the gamers add around ~100 friends every year. The study found the number of friendships increasing every year until the time of the data collection. This corresponds with the new accounts created each year in the figure (21) below. This figure also shows the increase in the number of heavy gamers joining the platform.

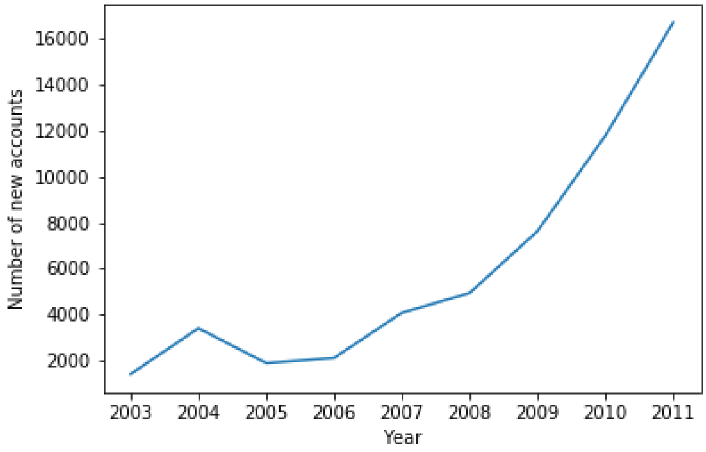

Figure 21: Lineplot of total number of new accounts of the heavy gamers every year

Cluster 1 and Cluster 2 showed significant variance in terms of adding new friends. For the year 2012 the median number of new friendships were found to be 6 and 15 respectively. As a result, Cluster 1 has less friends than Cluster 2. With this, the study can conclude that the Cluster 2 is more socially active than Cluster 1 in terms of adding friends. It is important to note that Cluster 2 is also older in the platform with over 250 days older on average than Cluster 1.

4.2.3 Group Participation

Groups in the Steam platform is a way for the players to connect to other players with similar interests. A group can be made with a goal to attract players with very specific interest such as a particular game only or a complete genre. A player will be part of the group to make new friends with similar interests. Hence like friendships, Cluster 2 shows increased social behavior in terms of being part of groups. On average, 50 percent of the individuals in Cluster 1 were part of 4 groups or less and only 10 percent were part of over 16 groups. Whereas, for Cluster 2, 50 percent of the individuals shown were part of 12 groups or less and 10 percent were part of over 61 groups. This is understandable as the number of distinct games for Cluster 2 is significantly higher than Cluster 1. The gamers in Cluster 2, having many games also tend to be part of more groups to find more gamers with the same interest to play with and connect. Hence Cluster 2 also has much more friends as they would likely tend to befriend those fellow group members due to having similar interests.

4.2.4 Play Time Duration

Steam collects play times for players through the Steam client installed on the user’s computer which would be connected to the internet. Play times are collected two ways, one would be all time aggregate play times for each game in the players game library since the time it was added. Another would be a rolling two week play time up until the data is collected through the Steam API. This information would be very accurate as steam directly monitors the play times and sends the information back to its servers once the user is online. However, Steam only monitors how long the game is open and not how long the player is active in the game itself.

The previous study on Steam [10] found evidence of the Pareto principle where the top 20 percent of the steam users accounted for over 82 percent of all the gameplay time on steam. As suggested in the previous study the top 20% can contain players who purposefully keep the game open to increase the overall play times for achievements or bragging rights. However, this might be few and far between as the resource consumption will not lead to significant results.

Looking at the analysis from the previous study [10], for two-week playtimes of all gamers, the 80th percentile was found to be around 16 hours a week on average or just over 2 hours a day on average for a week. Since the current study only extracted information on the players who portrayed an average of 6 hours of gameplay every day for two weeks, the median for two-week playtime was found to be 7.5 hours every day on average. Both the clusters showcased similar play times for two weeks.

As for the all-time total play time information, this study decided to look at the average daily gameplay time of the players. This would help because the user can be part of steam for any period of time which would make the overall forever play time for an individual irrelevant. It was found that 50 percent of the gamers had created the steam account around 3 years ago and the same percentage of players had racked up over 100 minutes of gameplay every day. Since the data only has information on the past two week playtimes and not all the two weeks post time of creation of the account, it will not be possible to understand whether the player had long running but infrequent gaming sessions or smaller but more consistent ones. For this reason, it would be difficult to contemplate the clusters and their forever playtimes. However, looking at the data, Cluster 2 was found to play around 2 hours and 30 minutes more everyday which is 30 minutes more than Cluster 1. Cluster 2 however consisted of newer accounts on average of around 300 days younger than Cluster 1.

The study decided to investigate the aspect of the online component of the games and its effect on the heavy gamers. The previous study [10] found, '67.7% of two-week playtimes as well as 57.7% of total playtimes are devoted entirely to playing multiplayer games, even though only 48.7% of games on Steam are multiplayer games (i.e. have a multiplayer component)'. In the current study of heavy gamers, over 83% of the total two-week playtimes as well as the total overall playtimes accounted for multiplayer games (or at least games that have a multiplayer component). This is quite a significant rise over the overall population, indicating that the heavy gamers are more engaged in games with an online component than the entire community. Cluster 2 had spent 76% of their total playtime on games with multiplayer component while Cluster 1 spent over 80%. This is because Cluster 2 also spends more money buying the offline games (games without a multiplayer component).

5. CONCLUSIONS AND RECOMMENDATIONS

The study found the existence of subgroups of heavy gamers with respect to their monetary social and gameplay behaviors. One group exhibited an extreme case of addiction wherein the players of said group were playing for excessive amounts of time since their account creation. These players also overspent and bought many games, much of which remained unplayed. The other group can be assumed to be incapable of such monetary expenditure as their gameplay time was spent playing free games and mostly the higher rated games. Free multiplayer games most often have in game microtransactions for digital items like ‘character skins’ or other items that improve the look and feel of the game. These items can also sometimes even provide the buyer with an advantage over the competition. These sorts of purchases are not available in the dataset and hence it remains a possibility that both the groups have similar monetary expenditure where one group mostly spends on buying the games while the other spends mostly on buying in game items. The addiction in this sense can be of owning games and or mastering them. The study also found that heavy gamers are significantly more engaged to games with an online component than the overall community. The overall heavy gamers showcased social behavior such as friendships to be similar to that of the casual gamers. The clusters were however different to each other in terms of social structure with one group being part of more groups and adding more friends than the other cluster and the overall community.

To reduce addiction in this groups, the individuals can be treated the same way individuals suffering from other types of behavioral addiction. However, connecting to these individuals would be the first step in that journey. Fortunately, the Steam ID provided in the datasets can be used to go to the profiles using an URL[24]. If the individuals on this study are still present in the platform and have an active account, their Steam ID can be used to connect with them through steam.

[24] http://steamcommunity.com/profiles/<steamid>

Despite the methods to treat gaming addiction virtually is out of scope for this study we can recommend the following methods:

Cluster 1 is clearly more open to buying new games and seems addicted to buying newer games more than playing them. Steam or an individual can choose to point this out to the gamers in hopes of causing ‘buyer’s remorse’ where the buyer may feel guilty of buying a product. Steam, despite getting a profit from these players may find it better fit to help these players out to get more popularity in the gaming community. This may raise their revenue long term for being more careful with the players money since these heavy gamers will surely add to a very small portion to their overall revenue. If the player continues to buy these games however, Steam or the individual can ask for help from the many friends the user has in an attempt to put peer pressure on the heavy gamers.

Cluster 2 is either unable to afford new games, satisfied with playing extended hours of free and free multiplayer games or is buying in game add-ons that improve their experience in the game. These gamers will be much more hooked on to one single game at a time. Since the transactional data is private Steam can investigate the data and check if this cluster also spends high amounts of money and in game items. If such is the case Steam can offer discounts on other games in an attempt to get the players to diverge from the current game. Individuals can also connect with these gamers and survey them to better understand their motives and recommend techniques to help these addicts.

6. FUTURE WORK

With the rise of technology and gaming becoming more available every year, it can safely be assumed that gaming addiction will only be on the rise. The dataset collected in 2013 gives us an overall idea on the world of addicted gamers. Future work can look into new data using the Steam API which continues to be offered by Steam at the time of this study in early 2020. Specific gamers that show addictive behavior can be tracked for a longer period of time than just two weeks to better understand their behavior. The Steam API has also been updated since the dataset collection and currently allows player wise achievements which was not collected during the previous study and hence was not available for the current study. Further work on newer data can reinforce the findings in this study and also look into other behavioral aspects such as two-week playtimes throughout a longer period of time.

It is important to note that the Steam ID present in the dataset is the actual Steam ID of the player. Hence, if these individuals found in the cluster are still present in the Steam platform their information may prove vital in understanding video game addiction in the long term. Third party sites already allow to lookup players using a Steam ID and some of their information. Future work can utilize the API and work on understanding long term video game addiction. The heavy gamers can also be surveyed to get a better understanding of their addiction and recommend better ways to help the gamers out. In addition to this, future studies can also enhance the recommendations or provide newer recommendations to help these addicted gamers or prevent newer gamers from becoming seemingly like the addicts in this study.

REFERENCES

1. World Health Organization. (2020). Gaming disorder. [online] Available at: https://www.who.int/features/qa/gaming-disorder/en/ [Accessed 14 Feb. 2020].
2. Icd.who.int. (2020). ICD-11 - Mortality and Morbidity Statistics. [online] Available at: https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1448597234 [Accessed 14 Feb. 2020].
3. Hansen, D. (2016) Game On!: Video Game History from Pong to Pac-Man to Mario, Minecraft, and Beyond. Feiwel & Friends.
4. Alrobai A., Phalp K., Ali R. (2014) Digital Addiction: A Requirements Engineering Perspective. In: Salinesi C., van de Weerd I. (eds) Requirements Engineering: Foundation for Software Quality. REFSQ 2014. Lecture Notes in Computer Science, vol 8396. Springer, Cham
5. Peper, E., & Harvey, R. (2018). Digital addiction: Increased loneliness, anxiety,and depression. NeuroRegulation, 5(1), 3–8. http://dx.doi.org/10.15540/nr.5.1.3
6. Ko, C.-H., Liu, G.-C., Hsiao, S., Yen, J.-Y., Yang, M.-J., Lin, W.-C., Yen, C.-F. and Chen, C.-S. (2009) ‘Brain activities associated with gaming urge of online gaming addiction,’ Journal of Psychiatric Research, 43(7), pp. 739–747. doi: 10.1016/j.jpsychires.2008.09.012.
7. Kim, Jong-Un. (2008). The Effect of a R/T Group Counseling Program on The Internet Addiction Level and Self-Esteem of Internet Addiction University Students. International Journal of Reality Therapy. 17.
8. Griffiths, M. D. and Meredith, A. (2009) ‘Videogame Addiction and its Treatment,’ Journal of Contemporary Psychotherapy, 39(4), pp. 247–253. doi: 10.1007/s10879-009-9118-4.
9. King, D. L., Delfabbro, P. H., Griffiths, M. D. and Gradisar, M. (2011) ‘Assessing clinical trials of Internet addiction treatment: A systematic review and CONSORT evaluation,’ Clinical Psychology Review, 31(7), pp. 1110–1116. doi: 10.1016/j.cpr.2011.06.009.
10. O'Neill, M., Vaziripour, E., Wu, J. and Zappala, D. (2016). Condensing Steam: Distilling the Diversity of Gamer Behavior. Proceedings of the 2016 ACM on Internet Measurement Conference - IMC '16.

11. Van Rooij, A., Schoenmakers, T., Vermulst, A., Van Den Eijnden, R. and Van De Mheen, D. (2010). Online video game addiction: identification of addicted adolescent gamers. Addiction, 106(1), pp.205-212.
12. Lemmens, J., Valkenburg, P. and Peter, J. (2009). Development and Validation of a Game Addiction Scale for Adolescents. Media Psychology, 12(1), pp.77-95.
13. Chappell, D., Eatough, V., Davies, M. N. O. and Griffiths, M. (2006) 'EverQuest—It's Just a Computer Game Right? An Interpretative Phenomenological Analysis of Online Gaming Addiction,' International Journal of Mental Health and Addiction, 4(3), pp. 205–216. doi: 10.1007/s11469-006-9028-6.
14. Lin-Liu J. China's e-Junkies Head for Rehab. Beijing hospital ward treats Internet addicts (Internet). IEEE Spectrum Online 2006. Available at: http://spectrum.ieee.org/telecom/ internet/chinas-ejunkies-head-for-rehab
15. Samaha, M. and Hawi, N. S. (2016) 'Relationships among smartphone addiction, stress, academic performance, and satisfaction with life,' Computers in Human Behavior, 57, pp. 321–325. doi: 10.1016/j.chb.2015.12.045.
16. Kuss, D. J., & Griffiths, M. D. (2012). Internet gaming addiction: A systematic review of empirical research. International Journal of Mental Health and Addiction, 10(2), 278-296.
17. Ross, D. R. (1982) 'Space Invaders Obsession,' JAMA: The Journal of the American Medical Association, 248(10), p. 1177. doi: 10.1001/jama.1982.03330100017009.
18. Soper, W. B., & Miller, M. J. (1983). Junk-time junkies: An emerging addiction among students. School Counselor, 31(1), 40–43.
19. Bauckhage, C., Kersting, K., Sifa, R., Thurau, C., Drachen, A. and Canossa, A. (2012) 'How players lose interest in playing a game: An empirical study based on distributions of total playing times,' 2012 IEEE Conference on Computational Intelligence and Games (CIG). doi: 10.1109/cig.2012.6374148.
20. A. V. Rajan et al., 'Virtual Reality Gaming Addiction,' 2018 Fifth HCT Information Technology Trends (ITT), Dubai, United Arab Emirates, 2018, pp. 358-363.
21. Shotton, M. (1989). Computer Addiction Pb. London: CRC Press, https://doi.org/10.1201/b12555
22. Widyanto, L., Griffiths, M. D. and Brunsden, V. (2011) 'A Psychometric Comparison of the Internet Addiction Test, the Internet-Related Problem Scale, and Self-Diagnosis,' Cyberpsychology, Behavior, and Social Networking, 14(3), pp. 141–149. doi: 10.1089/cyber.2010.0151.

23. Dworak, M., Schierl, T., Bruns, T. and Struder, H. K. (2007) ‘Impact of Singular Excessive Computer Game and Television Exposure on Sleep Patterns and Memory Performance of School-aged Children,’ Pediatrics, 120(5), pp. 978–985. doi: 10.1542/peds.2007-0476.
24. Wittek, C. T., Finserås, T. R., Pallesen, S., Mentzoni, R. A., Hanss, D., Griffiths, M. D. and Molde, H. (2015) ‘Prevalence and Predictors of Video Game Addiction: A Study Based on a National Representative Sample of Gamers,’ International Journal of Mental Health and Addiction, 14(5), pp. 672–686. doi: 10.1007/s11469-015-9592-8.
25. Salguero, R. A. T. and Morán, R. M. B. (2002) ‘Measuring problem video game playing in adolescents,’ Addiction, 97(12), pp. 1601–1606. doi: 10.1046/j.1360-0443.2002.00218.x.
26. King, D. L., Delfabbro, P. H. and Griffiths, M. D. (2012) ‘Clinical Interventions for Technology-Based Problems: Excessive Internet and Video Game Use,’ Journal of Cognitive Psychotherapy, 26(1), pp. 43–56. doi: 10.1891/0889-8391.26.1.43.
27. Gentile, D. (2009) ‘Pathological Video-Game Use Among Youth Ages 8 to 18,’ Psychological Science, 20(5), pp. 594–602. doi: 10.1111/j.1467-9280.2009.02340.x.
28. Grüsser, S., Thalemann, R. and Griffiths, M. (2007) ‘Excessive Computer Game Playing: Evidence for Addiction and Aggression?,’ CyberPsychology & Behavior, 10(2), pp. 290–292. doi: 10.1089/cpb.2006.9956.
29. Longman, H., Oconnor, E. and Obst, P. (2009) ‘The Effect of Social Support Derived from World of Warcraft on Negative Psychological Symptoms,’ CyberPsychology & Behavior, 12(5), pp. 563–566. doi: 10.1089/cpb.2009.0001.
30. Batthyány, D., Müller, K. W., Benker, F. and Wölfling, K. (2009) ‘Computerspielverhalten: Klinische Merkmale von Abhängigkeit und Missbrauch bei Jugendlichen,’ Wiener klinische Wochenschrift, 121(15-16), pp. 502–509. doi: 10.1007/s00508-009-1198-3.
31. Festl, R., Scharkow, M. and Quandt, T. (2012) ‘Problematic computer game use among adolescents, younger and older adults,’ Addiction, 108(3), pp. 592–599. doi: 10.1111/add.12016.
32. Wan, C.-S. and Chiou, W.-B. (2006) ‘Psychological Motives and Online Games Addiction: ATest of Flow Theory and Humanistic Needs Theory for Taiwanese Adolescents,’ CyberPsychology & Behavior, 9(3), pp. 317–324. doi: 10.1089/cpb.2006.9.317.

33. Chan, P. A. and Rabinowitz, T. (2006) ‘A cross-sectional analysis of video games and attention deficit hyperactivity disorder symptoms in adolescents,’ Annals of General Psychiatry, 5(1). doi: 10.1186/1744-859x-5-16.
34. Chiu, S.-I., Lee, J.-Z. and Huang, D.-H. (2004) ‘Video Game Addiction in Children and Teenagers in Taiwan,’ CyberPsychology & Behavior, 7(5), pp. 571–581. doi: 10.1089/cpb.2004.7.571.
35. Lemmens, J. S., Valkenburg, P. M. and Peter, J. (2011) ‘Psychosocial causes and consequences of pathological gaming,’ Computers in Human Behavior, 27(1), pp. 144–152. doi: 10.1016/j.chb.2010.07.015.
36. Rehbein, F., Psych, G., Kleimann, M., Mediasci, G. and Mößle, T. (2010) ‘Prevalence and Risk Factors of Video Game Dependency in Adolescence: Results of a German Nationwide Survey,’ Cyberpsychology, Behavior, and Social Networking, 13(3), pp. 269–277. doi: 10.1089/cyber.2009.0227.
37. Moore, C. (2011) ‘Hats of affect : a study of affect, achievements and hats in Team Fortress 2,’ Game Studies, 11(1), pp. 1–14. Available at: http://hdl.handle.net/10536/DRO/DU:30041568.
38. Blackburn, J., Simha, R., Kourtellis, N., Zuo, X., Long, C., Ripeanu, M., Skvoretz, J., & Iamnitchi, A. (2011). Cheaters in the Steam Community Gaming Social Network. ArXiv, abs/1112.4915.
39. Toy, E. J., Kummaragunta, J. V. H. H. and Yoo, J. S. (2018) ‘Large-Scale Cross-Country Analysis of Steam Popularity,’ 2018 International Conference on Computational Science and Computational Intelligence (CSCI). doi: 10.1109/csci46756.2018.00205.
40. Becker, R., Chernihov, Y., Shavitt, Y. and Zilberman, N. (2012) ‘An analysis of the Steam community network evolution,’ 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel. doi: 10.1109/eeei.2012.6377133.
41. Sifa, R, Drachen, A & Bauckhage, C 2015, Large-Scale Cross-Game Player Behavior Analysis on Steam. in PROCEEDINGS OF THE ELEVENTH ARTIFICIAL INTELLIGENCE AND INTERACTIVE DIGITAL ENTERTAINMENT INTERNATIONAL CONFERENCE (AIIDE 2015). AAAI Press, pp. 198-204, Artificial Intelligence and Interactive Digital Entertainment International Conference , California, United States, 15/11/2015.